Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Scuola di Dottorato di Ricerca in Scienze Linguistiche, Filologiche e Letterarie

Indirizzo di Linguistica, Filologia e Letterature Anglo-Germaniche

XXV Ciclo - Anno 2013

The Pitch Range of Italians and Americans. A Comparative Study

Direttore della Scuola: Ch.ma Prof.ssa ROSANNA BENACCHIO

Coordinatore d'indirizzo: Ch.ma Prof.ssa ANNALISA OBOE

Supervisore: Ch.ma Prof.ssa MARIA GRAZIA BUSÀ

Dottoranda: MARTINA URBANI

Contents

Dedication 5
Acknowledgements 5
Abstract (in English) 8
Summary (in Italian) 10
List of figures 19
List of Tables 23
List of Abbreviations 26

CHAPTER 1: INTRODUCTION 27

1.1 Suprasegmental features 27
1.1.1 The role of prosody 28
1.1.2 The functions of intonation 30
1.2 The dichotomies of pitch range 33
1.2.1 Pitch range in L2 35
1.2.2 Production and Perception 38
1.3 Motivation 39
1.4 Research outline 43

CHAPTER 2: ACOUSTIC ANALYSES OF PITCH RANGE 47

2.1 F0 and pitch 47
2.1.1 Pitch range: level and span 50
2.1.2 Pitch perception 53
2.1.3 Linear and non-linear frequency scales 55
2.1.4 Speech rate and length 58
2.1.5 The effects of segmental factors on F0 61
2.2 Impressionistic descriptions and acoustic studies 62
2.2.1 The effects of gender 63
2.2.2 The effects of age 64
2.2.3 The effects of body size and health history 66
2.2.4 The effects of socio-cultural factors 67
2.2.5 The effects of speech task 69
2.3 Pitch range and biological codes 71
2.3.1 The frequency code 73
2.3.2 The effort code 75
2.3.3 The production code 76
2.4 New methods in the analysis of pitch range 78
2.4.1 Ladd (1985-2003) 78
2.4.2 Patterson (2000) 81
2.4.3 Keating (2010) 82
2.4.4 Mennen (2007-2012) 85
2.5 Summary 90

CHAPTER 3: CROSS-LINGUISTIC RESEARCH ON L2 PITCH 93

3.1 The Auto-segmental Metrical Approach 93
3.1.1 The tone system 97
3.1.2 The American and British approaches 98
3.2 ToBI 102
3.2.1 Pierrehumbert (1980) 106
3.2.2 Silverman et al. (1992) 107
3.2.3 Beckman et al. (1994, 2005) 110
3.2.4 Veilleux at el. (2006) 113
3.3 Review of previous pitch range analyses across different languages 122
3.3.1 Italian and other Romance languages 123
3.3.2 English and other Germanic languages 133
3.4 L2 acquisition models 144
3.4.1 Speech Learning Model 145
3.4.2 Perceptual Assimilation Model 148
3.4.3 Native language Magnet Model 150
3.4.4 Feature Competition Model 151
3.5 Summary 152

CHAPTER 4: PRODUCTION STUDY 155

4.1 Introduction 155
4.2 Rationale behind the experiments 156
4.3 First Experiment 157
4.3.1 Research Questions 158
4.3.2 Subjects 159
4.3.3 Materials 160
4.3.4 Procedure 161
4.3.5 Method 162
4.3.6 Results 168
4.3.7 Discussion 179
4.4 Second experiment 181
4.4.1 Yes-no questions (YNQ) 183
4.4.2 Wh-questions (WHQ) 186
4.4.3 Statements (STM) 188
4.4.4 Research Questions 191
4.4.5 Subjects 192
4.4.6 Materials 196
4.4.7 Procedure 197
4.4.8 Method 198
4.4.9 Results 199
4.4.10 Discussion 219
4.5 Summary 220

CHAPTER 5: PERCEPTION STUDY 223
5.1 Introduction 223
5.2 The effects of pitch variation 226
5.3 Third experiment 228
5.3.1 Research questions 230
5.3.2 Subjects 233
5.3.3 Materials 234
5.3.4 Procedure 237
5.3.5 Results 238
5.3.6 Discussion 248
5.4 Fourth experiment 250
5.4.1 Manipulation techniques 251
5.4.2 Research questions 254
5.4.3 Subjects 258
5.4.4 Materials 259
5.4.5 Procedure 262
5.4.6 Results 268
5.4.7 Discussion 273
5.5 Summary 275

CHAPTER 6: CONCLUSION 277
6.1 Summary of the findings 277
6.2 Production studies 277
6.2.1 First experiment 278
6.2.2 Second experiment 279
6.3 Perception Studies 280
6.3.1 Third experiment 280
6.3.2 Fourth experiment 282
6.4 General Discussion 283
6.5 Future Directions 283

References 285

Appendix A 307
Appendix B 333
Appendix C 343

Dedication

To Angelo and Francesca,
for being such great examples
of wisdom, patience and courage.

Acknowledgements

First of all I would like to express my deepest gratitude to my advisor, Professor Maria Grazia Busà, for her help in bringing this dissertation into being and for her technical expertise. She provided me with enormous support and encouragement, by inciting me to develop an aptitude for critical analysis and originality. When I was running low on ideas, she always helped me with her useful comments. With her invaluable suggestions, she spurred me to navigate past the many research obstacles I faced. I thank her also for initiating me into the perils and gratifications of experimental research, for her inspirational guidance, and for the enthusiasm she puts into everything she does.

Throughout the different stages of my work, I received constructive comments and precious feedback from a number of people who helped me to improve the quality of my thesis. I would like to thank Professor Ineke Mennen for sharing her vast knowledge on pitch range analysis with me and providing me with brilliant insights into the complexities of pitch range. During her visits in Padova and my visit in Bangor, she generously taught me her techniques for calculating F0 and I greatly benefitted from stimulating discussions.

At the Department of Linguistic and Literary Studies in Padova I spent most of my academic life. I greatly appreciate the support of my advisors and mentors, who influenced my career with their ideas and tutoring. Professor Mario Melchionda supported me since my undergraduate years and encouraged me to pursue advanced studies in my exchange programs at Boston University and Aberystwyth University. I thank the thriving group of researchers at the Language and Communication Lab (LCL), Carla Xavier, Luca Rognoni and Antonio Stella, for their valuable collaboration and for providing me the most enjoyable working environment. I feel happy for having shared my Ph.D. studies with Raffaella, Marta, Giulia and Elena.

I thank Professor Sun-Ah Jun for sponsoring my visit at the University of California – Los Angeles and for her warm welcome. At UCLA, Professor Patricia Keating and Marc Garellek introduced me to statistical methods for linguistics. I was incredibly privileged to have the opportunity to sit in their classes and to receive their precious feedback on my work. I greatly benefitted from their committed mentoring and supervision in the designing and recording of speech materials for experimental analysis. I thank my office-mates Adam Chong, Sadie Martin, Elyssa Dudley, Victoria Mateu, Jos Tellings, Yu Tanaka, and Jinni Kang. I really miss my friends Damiano Pasetto and Lorena Cayre Baito, unique fellow Ph.D. students. I thank the members of the UCLA phonetic lab, including Professors Bruce Hayes, Kie Zuraw, and Robert Daland, for providing me with excellent laboratory facilities and giving me the opportunity to work in such a collaborative and stimulating environment.

Many thanks to the students in Communication at the University of Padova and also to the students from the UCLA Department of Linguistics who volunteered to be my subjects for the experiments. Much appreciation goes to the director of the Boston University – Padua Program, Dr. Elisabetta Convento, and the student affair coordinator, Bonnie Ponthier, for their help with recruiting participants. I owe a debt of thanks to Suzanne Cloke for letting me conduct my experiments during her lessons.

A very special thanks to my parents, Angelo and Francesca, for sharing my joy in the most exciting achievements and offering words of encouragements in my occasional disappointments. My parents sustained me every day with their love, trust and continuous support. This thesis would not exist were it not for the inspiration and the encouragement of so many people. I wish to express my gratitude to all the people who have encouraged me along the way, by giving me strength, motivation and much more. I feel grateful to my family and my friends who helped me when I needed them most.

Padova, January 30th 2013

Thanks to my subjects for the experiments (in alphabetical order):

Alberghini Greta
Alpert Isaac
Andreose Luca
Ardeni Viola
Asquer Raffaele
Bagatella Silvia
Baggio Michele
Baldi Luca
Barbuio Chiara
Baruffaldi Elisa
Basso Tobia
Beghin Jenni
Bergamin Silvia
Bianchin Samuele
Bodo Michele
Bordignon Marco
Boroda Cristina
Bottelli Anna
Bovo Giulia
Calore Giulia
Canzian Luca
Carraro Elena
Castaldo Francesco
Cavaliere Roberta
Cei Erica
Conte Stefano
Corazzin Annachiara
Costa Riccardo
D'Agostini Giulia
Dai Pra Aurora
Damesin Massimo
D'Angelo Catherine
De Francesco Meghan
De Marchi Alice
De Marchi Denise
Del Giudice Pierluigi
Di Renzo Tommaso
Dudley Elyssa
Fasoli Sonia
Fattore Giada
Franceschi Paola
Franzutti Valentina
Furlan Chiara
Gasparini Jacopo
Gelonesi Laura
Giacometti Alessio
Giannina Matteo
Girotto Manuela
Grigolo Amedeo
Guarda Marta
Hayden Aaron
Labra Natascia
Lauzon Rebecca
Lo Re Riccardo
Loveridge Emily
Marcon Gaia
Marin Eleonora
Martin Sadie
Martini Giada
Mazzaro Lorenzo
Mazzuccato Davide
Mezzalana Ilaria
Nunziata Valeria
Paggiaro Luca
Panizzon Raffaella
Pasetto Damiano
Pastorello Elisa
Perin Luca
Perseghin Erica
Peruzzo Riccardo
Pilotto Silvia
Ranzato Giulia
Resmini Fabio
Rhyme Kayleah
Robinson Nicole
Sabrina Bertollo
Sandrin Sabrina
Saviolo Veronica
Scatà Gianmaria
Signori Laura
Simonato Emma
Simoni Alessandro
Smith Brooks
Soro Federica
Stranieri Marissa
Streifer Adriana
Streifer Monica
Terenzi Elsa
Trabacchin Francesca
Trento Eleonora
Uroni Cristina
Vagrotelli Francesca
Venturini Michele
Zamboni Camilla
Zancato Alberto
Zannoni Thomas
Zanuso Eleonora
Zimmerman Nina
Zorzo Giulia

Abstract (in English)

Linguistic experiments have investigated the nature of F0 span and level in cross-linguistic comparisons. However, only few studies have focused on the elaboration of a general-agreed methodology that may provide a unifying approach to the analysis of pitch range (Ladd, 1996; Patterson and Ladd, 1999; Daly and Warren, 2001; Bishop and Keating, 2010; Mennen et al. 2012).

Pitch variation is used in different languages to convey different linguistic and paralinguistic meanings that may range from the expression of sentence modality to the marking of emotional and attitudinal nuances (Grice and Baumann, 2007). A number of factors have to be taken into consideration when determining the existence of measurable and reliable differences in pitch values. Daly and Warren (2001) demonstrated the importance of some independent variables such as language, age, body size, speaker sex (female vs. male), socio-cultural background, regional accents, speech task (read sentences vs. spontaneous dialogues), sentence type (questions vs. statements) and measure scales (Hertz, semitones, ERB etc.). Coherently with the model proposed by Mennen et al. (2012), my analysis of pitch range is based on the investigation of LTD (long-term distributional) and linguistic measures. LTD measures deal with the F0 distribution within a speaker's contour (e.g. F0 minimum, F0 maximum, F0 mean, F0 median, standard deviation, F0 span) while linguistic measures are linked to specific targets within the contour, such as peaks and valleys (e.g. high and low landmarks) and preserve the temporal sequences of pitch contours.

This investigation analyzed the characteristics of pitch range production and perception in English sentences uttered by Americans and Italians. Four experiments were conducted to examine different phenomena: i) the contrast between measures of F0 level and span in utterances produced by Americans and Italians (experiments 1-2); ii) the contrast between the pitch range produced by males and females in L1 and L2 (experiment 1); iii) the F0 patterns in different sentence types, that is, yes-no questions, wh-questions, and exclamations (experiment 2); iv) listeners' evaluations of pitch span in terms of ±interesting, ±excited, ±credible, ±friendly ratings of different sentence types (experiments 3-4); v) the correlation between pitch span of the sentences and the evaluations given by American and Italian listeners (experiment 3); vi) the listeners' evaluations of pitch span values in manipulated stimuli, whose F0 span was re-synthesized under three conditions: narrow span, original span, and wide span (experiment 4); vii) the different evaluations given to the sentences by male and female listeners.

The results of this investigation supported the following generalizations. First, pitch span more than level was found to be a cue for non-nativeness, because L2 speakers of English used a narrower span, compared to the native norm. What is more, the experimental data in the production studies indicated that the mode of sentences was better captured by F0 span than level. Second, the Italian learners of English were influenced by their L1 and transferred L1 pitch range variation into their L2. The English sentences produced by the Italians had overall higher pitch levels and narrower pitch span than those produced by the Americans. In addition, the Italians used overall higher pitch levels when speaking Italian and lower levels when speaking English. Conversely, their pitch span was generally higher in English and lower in Italian. When comparing productions in English, the Italian females used higher F0 levels than the American females; vice versa, the Italian males showed slightly lower F0 levels than the American males. Third, there was a systematic relation between pitch span values and the listeners' evaluations of the sentences. The two groups of listeners (the Americans and the Italians) rated the stimuli with larger pitch span as more interesting, exciting and credible than the stimuli with narrower pitch span. Thus, the listeners relied on the perceived pitch span to differentiate among the stimuli. Fourth, both the American and the Italian speakers were considered more friendly when the pitch span of their sentences was widened (wide span manipulation) and less friendly when the pitch span was narrowed (narrow span manipulation). This happened in all the stimuli regardless of the native language of the speakers (American vs. Italian).

Summary (in Italian)

Linee di ricerca

Questa ricerca si basa sullo studio delle melodie intonative prodotte degli americani e dagli italiani ed esamina le differenze tonali dell'inglese e dell'italiano. In particolare, è interessante indagare quali siano le reazioni dei parlanti madrelingua rispetto alle produzioni orali realizzate dai parlanti di una seconda lingua. Nell'analizzare i contorni melodici, si tratta di misurare il livello (level) e l'ampiezza (span) dei movimenti intonativi.

Da un punto di vista acustico, l'oggetto dell'indagine sperimentale è la frequenza fondamentale (F0) delle produzioni vocali dei parlanti. F0, ovvero la prima armonica della forma d'onda, è una caratteristica fisica del segnale acustico che può essere misurata in hertz (Hz) o semitoni (ST). Mentre il livello generale di F0 viene misurato solitamente in Hz, l'ampiezza del movimento (calcolata come la differenza tra il picco massimo e il picco minimo) viene di norma calcolata in ST.

Da un punto di vista percettivo, l'orecchio umano ha un campo di udibilità che oscilla dai 16 Hz ai 20.000 Hz. Inoltre è molto più facile percepire un innalzamento o abbassamento della F0 a basse frequenze che ad alte frequenze. Questo è dovuto alla particolare conformazione della coclea, all'interno dell'appartato uditivo. Dal momento che anche un minimo scarto di F0 può essere agilmente individuato alle basse frequenze mentre risulta magari impercettibile alle alte frequenze, è necessario considerare le variazioni di F0 a livello percettivo. Per questo, F0 acquisisce anche una valenza percettiva e viene indicata con il termine altezza tonale. L'altezza tonale identifica quindi una caratteristica percettiva del suono che viene valutata attraverso un giudizio uditivo e non acustico.

Data la complementarietà del fattore acustico e percettivo nello studio di F0, il tipo di analisi condotto in questo studio si articola su due livelli. Nel primo livello, vengono esaminate le differenze di tipo acustico nel parlato letto in inglese e in italiano da italiani e da americani. Le produzioni orali sono divise per tipologia di frase e vengono registrate da un gruppo di maschi e un gruppo di femmine. Lo scopo dei due studi produttivi è quello di riscontrare se ci sono differenze di F0 riguardanti: 1) la lingua in esame (inglese o italiano, 2) il fattore prima o seconda lingua (inglese parlato dagli americani e inglese parlato dagli italiani), 3) il sesso dei soggetti (valutazione delle differenze di F0 tra maschi e femmine), 4) la natura del materiale utilizzato e la tipologia di frase (le frasi vengono analizzate secondo la loro diversa tipologia: domande polari, domande k, e affermazioni).

Per quanto riguarda il secondo livello, i dati raccolti negli studi di produzione vengono testati con uno studio percettivo. Ad un gruppo di studenti americani ed italiani, viene chiesto di valutare una selezione di frasi prodotte in inglese da parlanti americani e italiani. Lo scopo di questa indagine è di esaminare la correlazione tra livello/ampiezza dell'altezza tonale e la valutazione positiva/negativa degli enunciati. Le valutazioni si articolano su giudizi di tipo qualitativo e quantitativo, in riferimento a diversi gradi di interesse, credibilità, e coinvolgimento.

Struttura della tesi

Il primo capitolo definisce gli obiettivi del mio studio sperimentale alla luce degli studi linguistici svolti su altezza tonale e contorni intonativi. F0 costituisce una delle strutture prosodiche fondamentali per la ricostruzione dei contorni intonativi. In particolare, F0 è determinante nella percezione del focus, dello stress e dell'accento straniero.

Il secondo capitolo passa in rassegna alcuni studi acustici di tipo impressionistico ed altri di carattere sperimentale che hanno identificato i fattori principali che influiscono sulle variazioni di F0. La frequenza fondamentale infatti può subire modifiche rilevanti dovute al sesso e all'età dei parlanti. Inoltre anche la corporatura e la fisiologia delle corde vocali possono influire sulle escursioni di F0. E' noto che il fumo o gravi patologie alla laringe producono delle modifiche permanenti al cavo faringeo e alle corde vocali. Oltre a fattori di natura prettamente fisiologica, pressioni di tipo socio-culturale possono indurre alcuni parlanti a modificare l'ampiezza e il livello generale di F0. Per esempio, diversi studi sull'altezza tonale delle donne giapponesi (van Bezooijen, 1995 e Yuasa, 2008) dimostrano che la maggior parte di esse utilizza un livello di altezza tonale maggiore rispetto a quello naturale per dare un'immagine di ricercata gentilezza e femminilità. Anche la tipologia dei materiali analizzati negli studi su F0 può avere degli effetti sui risultati ottenuti. Alcuni studiosi hanno avanzato l'ipotesi che F0 presenti delle differenze nel parlato letto e nel parlato spontaneo. Tuttavia, altri studi hanno dimostrato che queste differenze sono molto limitate e pressoché ininfluenti nella realizzazione dei contorni intonativi. Il capitolo passa quindi in rassegna le metodologie più accreditate per l'analisi di F0 e altezza tonale. In particolare quattro modelli vengono presi in analisi: Ladd (1985-2003), Patterson (2000), Keating (2010) e Mennen (2007, 2012). Vengono messe in risalto le intuizioni più interessanti proposte in questi quattro modelli e le teorie biologiche alla base dello sviluppo dell'altezza tonale: frequency code, effort code e production code (Ohala, 1984, 1994; Gussenhoven, 2004).

Il terzo capitolo tratteggia gli studi riguardanti l'analisi dell'altezza tonale nelle lingue romanze e germaniche. La descrizione dei modelli intonativi delle varie lingue si basa su un approccio di tipo auto-segmentale, dove i vari contorni intonativi vengono descritti con un'annotazione bitonale: H (high) per i picchi alti e L (low) per i picchi bassi. Il modello ToBI è riconosciuto all'interno della comunità scientifica come uno dei sistemi di annotazione dell'intonazione più efficace, poiché consente di decodificare e mettere a confronto movimenti dell'altezza tonale e rende conto di una quantità di informazioni prosodiche, in modo accurato e relativamente semplice. ToBI può essere definito un sistema completo perché non si limita ad indicare il livello di salita e discesa della F0 tramite l'uso di segni H e L, ma fa ampio uso di segni diacritici per indicare l'accento primario (*), le interruzioni o pause (%), i valori di F0 in corrispondenza del phrase accent (-) e i movimenti bitonali (L+H).

Ad oggi, non è ancora stato proposto un modello per l'acquisizione dell'intonazione. La maggior parte degli studi sull'acquisizione del parlato si focalizzano quasi esclusivamente sull'acquisizione dei segmenti. Il modello Speech Learning di Flege, il modello Perceptual Assimilation di Best, il modello Native Language Magnet di Kuhl, e il modello Feature Competition di Hancin-Bhatt sono incentrati sulle modalità di acquisizione dei tratti segmentali di una lingua. Anche se nessuno di questi modelli tratta l'acquisizione prosodica, essi fungono da quadro di riferimento per le dinamiche che entrano in gioco nel processo di acquisizione di una seconda lingua.

Il quarto capitolo, entra nel vivo dell'indagine sperimentale sull'escursione di F0 in italiano ed in inglese. Negli studi produttivi, l'analisi di F0 si articola in quattro direzioni: 1) due diverse lingue in esame: inglese e italiano, 2) influenza della prima lingua nei gruppi di parlanti americani e italiani, 3) resa di F0 nei maschi e nelle femmine, 4) tipologia di frasi analizzate: domande polari vs. domande wh vs. affermazioni. Gli studi di produzione vengono realizzati adottando due linee di indagine: il calcolo delle misure linguistiche e quello delle misure LTD (distribuzione a lungo termine). Le misure linguistiche vengono calcolate sulla base di un'etichettatura delle frasi. Ogni frase viene annotata manualmente con un sistema di siglatura proposto da Mennen et al. (2012). Al termine del processo di annotazione, tutti i valori di F0 corrispondenti a ciascuna etichetta vengono suddivisi e mediati. Questo tipo di annotazione permette di calcolare i picchi e le valli dei movimenti intonativi, restituendo un'immagine dell'andamento delle frase. Le misure LTD, al contrario di quelle linguistiche, non preservano l'ordine cronologico delle sequenze intonative, ma consentono di calcolare i valori medi di: F0 massima, F0 media e F0 minima. Mentre queste misure vengono calcolate in Hz, lo scarto tra F0 massima e F0 minima (detto anche span) viene misurato in semitoni. Grazie alle misurazioni dei valori di F0, le misure LTD offrono un resoconto dettagliato dei movimenti del contorno intonativo. Il calcolo delle misure LTD agevola la comparazione dei dati acustici e la valutazione della significatività dei risultati ottenuti.

Il quinto capitolo ha lo scopo di determinare in quale misura l'altezza tonale influisca nella percezione dell'intonazione in L1 e L2. L'indagine percettiva si articola in due fasi. La prima fase prevede la somministrazione di file audio non modificati all'ascolto di dieci studenti americani e dieci studenti italiani. Ai due gruppi di studenti viene richiesto di valutare ed esprimere giudizi qualitativamente diversi sulle frasi, in una scala da 1 a 5. Nella seconda fase, vengono selezionate alcune frasi in cui l'ampiezza di F0 è stata manipolata per aumentarne o diminuirne l'escursione e per creare zone piatte e zone con picchi elevati. Poichè la differenza tra F0 massima e F0 minima è molto maggiore in inglese che in italiano, è ipotizzabile che forti escursioni siano associate alla lingua inglese, mentre escursioni più deboli siano associate alla lingua italiana. Inoltre, è possibile che un'escursione forte sia percepita positivamente ed associata ad un maggior livello di interesse, credibilità, e coinvolgimento. Al contrario, frasi con una minima ampiezza di F0 possono risultare monotone e possono suscitare negli ascoltatori un senso di noia, distacco e disinteresse.

Materiali e Strumenti

La raccolta dei materiali si è svolta in un laboratorio fornito di una cabina insonorizzata, attrezzata con strumenti per la digitalizzazione del segnale acustico (microfono, DAT, etc.). Il microfono utilizzato era un Shure SM1, i file audio sono stati registrati direttamente su disco in un computer posizionato all'esterno della cabina di registrazione. Il segnale audio è stato digitalizzato a una frequenza di campionamento di 44.1 KHz e a un indice di quantizzazione di 32 bit, usando un AudioBox. I dati sono stati raccolti e processati con il software *Praat*. I soggetti che hanno partecipato agli esperimenti sono stati reclutati tra la popolazione studentesca e si sono prestati volontariamente alla registrazione di alcune brevi frasi e testi. Le analisi delle registrazioni sono state fatte con Praat. Per testare la significatività dei dati raccolti, è stato utilizzato il software per l'analisi statistica *SPSS* (Statistical Package for the Social Sciences). Per quanto riguarda lo studio percettivo, è stato creato un test online con l'ausilio del software opensource *surveymonkey*.

Risultati

I risultati del primo studio produttivo dimostrano che gli italiani usano livelli medi di F0 superiori a quelli degli americani, questo succede sia quando parlano in inglese che quando parlano in italiano (con valori più alti per l'italiano). La tendenza generale è quella di avere livelli di F0 massima, F0 media e F0 minima molto più alti in italiano che in inglese. Tuttavia, è emersa una significativa differenza nelle produzioni orali realizzate dai maschi e quelle realizzate dalle femmine. Mentre i soggetti maschi italiani hanno utilizzato livelli molto simili di F0 sia in italiano che in inglese; i soggetti femmine dello studio hanno notevolmente modificato la loro F0 a seconda della lingua in oggetto. Quando le informatrici italiane parlavano in italiano raggiungevano dei livelli di F0 piuttosto alti, quando invece parlavano in inglese cercavano di adeguare l'andamento di F0 al modello delle americane, ovvero diminuivano drasticamente F0. Per questo motivo, i maschi italiani di questo esperimento hanno dimostrato di subire una forte influenza dell'italiano (prima lingua) nelle loro produzioni in inglese (seconda lingua). Al contrario, le femmine italiane si sono rivelate molto più in grado di approcciarsi al livello di F0 tipico dell'inglese, subendo quindi una minore influenza della loro lingua nativa.

Sulla scorta dei risultati ottenuti in questo esperimento è stato creato un secondo esperimento, per testare solamente i dati prodotti dai soggetti femminili: dieci americane e dieci italiane. I dati raccolti sono stati suddivisi per tipologia di frase: interrogative totali (o domande polari), interrogative parziali (o domande k), e affermazioni. I risultati dimostrano che, per quanto riguarda il livello di F0, i valori medi sono superiori nelle frasi prodotte delle italiane rispetto a quelle delle americane. Questo si verifica per tutte e tre le tipologie di frasi (domande polari, domande k e affermazioni). Per quanto riguarda l'ampiezza di F0, il rapporto tra i valori delle italiane e quelli delle americane è rovesciato. Infatti, in tutte le tipologie di frasi, i valori di ampiezza di F0 delle americane sono sempre maggiori rispetto a quelli delle italiane. Le diverse frasi sono e poi suddivise in gruppi a seconda della tipologia e analizzate singolarmente, una ad una, soggetto per soggetto. Dall'analisi dei dati è emerso che le italiane e le americane realizzano in modo molto simile le domande polari, le quali raggiungono valori di F0 quasi sovrapponibili (sia per livello che per ampiezza). La differenza tra le produzioni delle americane e quelle delle italiane è invece molto più evidente nelle domande wh e nelle affermazioni. Mentre il livello di F0 è maggiore nelle frasi pronunciate dalle italiane, l'ampiezza di F0 è maggiore nelle frasi pronunciate dalle americane. Questo implica che, nonostante il livello di F0 sia maggiore nelle frasi prodotte dalle italiane, l'ampiezza risulta piuttosto bassa. L'ampiezza tonale è pressoché piatta nelle italiane a fronte di una grande variabilità dell'ampiezza tonale delle americane.

Le differenze di F0 (livello e ampiezza) tra l'inglese parlato dalle americane e quello parlato dalle italiane sono state esaminate anche negli studi percettivi. Una curva melodica relativamente piatta viene riconosciuta e percepita generalmente in modo negativo, come riportato in uno studio di Holub (2010). Le frasi inglesi prodotte dalle italiane, essendo più piatte rispetto a quelle delle americane, possono risultare monotone e venire quindi percepite in maniera più negativa rispetto a quelle delle americane. Infatti, i risultati raccolti nello studio percettivo dimostrano che: 1) la valutazione dell'ascoltatore è direttamente proporzionale all'ampiezza di F0, 2) il livello di F0 ha un ruolo decisamente minore rispetto all'ampiezza di F0 nell'interpretazione positiva o negativa del messaggio trasmesso, 3) le variazioni di F0 non solo veicolano informazioni riguardanti l'atteggiamento e lo stato d'animo del parlante ma hanno un'influenza anche sull'attendibilità di un messaggio trasmesso. Più ampia è l'escursione di F0, più il parlante viene percepito come interessato, credibile, e partecipativo. Infatti, nel caso di uno stimolo riguardante un invito a cena, le frasi formulate con una maggiore ampiezza di F0 sono state percepite come più allettanti e coinvolgenti rispetto ad altre.

In conclusione, i risultati ottenuti negli studi produttivi e percettivi dimostrano che F0 e altezza tonale hanno una doppia valenza: da un lato, registrano dei valori medi specifici per ogni lingua (in questo caso inglese e italiano) che rispecchiano le peculiari caratteristiche di lingue differenti; dall'altro lato, sottostanno a delle regole o principi universali che sembrano essere caratteristiche globalmente presenti nella maggior parte delle lingue. In effetti, pur avendo identificato delle tendenze generali diverse per l'italiano e per l'inglese, i soggetti analizzati negli esperimenti sono stati percepiti dagli ascoltatori come più o meno socievoli, interessati, e credibili a seconda della variazione dell'ampiezza di F0. Questo dimostra, che F0 è un fattore tipicamente distintivo per ogni parlante, qualunque sia la sua lingua nativa. Si auspica che, grazie allo sviluppo di nuove teorie e modelli sull'intonazione, sia possibile realizzare studi sempre più approfonditi e sistematici in questo campo, per ampliare la trattazione dei modelli intonativi nelle diversi lingue.

Conclusione

I risultati descritti in questo studio forniscono una descrizione accurata dei movimenti dell'altezza tonale in inglese e in italiano. In particolare, sono state prese in esame diverse condizioni attinenti l'interferenza delle prima e seconda lingua nelle produzioni orali in inglese e in italiano. I picchi intonativi sono stati analizzati sia a livello globale che locale, in quanto l'analisi è stata svolta sia in gruppi di frasi pronunciate all'interno di una conversazione letta in modo semi-naturale, sia in una sequenza di frasi isolate e divise per tipologia (domande polari, domande k, affermazioni).

Questo lavoro offre notevoli spunti per approfondimenti futuri. Innanzitutto, sarebbe interessante confrontare un maggiore numero di dati provenienti da un corpus di produzioni orali elicitate attraverso un map-task e non lette. Inoltre, questo studio prende in considerazione gli aspetti teorici e metodologici dei modelli di analisi dell'altezza tonale per poi condurre esperimenti di produzione e percezione sulla variazione di F0 in inglese e in italiano.

Infine, il sistema integrato di misure linguistiche e LTD che è stato adottato in questo studio, può fungere da modello di riferimento per studi futuri. Questo modello è ispirato a un sistema di codifica dell'analisi prosodica che verte sul connubio di aspetti fonetici e fonologici. Le ricadute applicative di un tale modello riscuotono un notevole interesse non solo per la teoria linguistica ma anche per la sintesi e il riconoscimento vocale. La creazione di un sistema integrato per l'analisi dell'altezza tonale rimane un tema centrale nella trattazione teorica ed empirica dell'analisi prosodica.

List of figures

Figure 1. The prosodic hierarchy. (Adapted from Keating et al., 2003: 5)..................................29
Figure 2. Normalized pitch contours of stressed vowels showing L1 influence on L2 English. (From Yuan, 2010: 1848). 36
Figure 3. A harmonic spectrum of the vocal fold vibration showing the F0 (first harmonic) and the 10th harmonic. (This figure has been taken from Johnson, 2003: 80)..48
Figure 4. F0 contours of the same sentence uttered with different pitch ranges. (From Lieberman and Pierrehumbert, 1984; in Ladd and Morton, 1997: 314). 52
Figure 5. Analysis of sound frequencies by the basilar membrane (from http://www.britannica.com/EBchecked/media/537/The-analysis-of-sound-frequencies-by-the-basilar-membrane, Encyclopedia Britannica, 1997) 54
Figure 6. Comparison of an auditory frequency scale (the Bark scale) and an acoustic frequency scale (in kHz). (From Johnson, 2003: 52)..56
Figure 7. General trend of F0 across male and female life span. (From Baken and Orlikoff, 2000: 174). 65
Figure 8. How H and L signal peaks and valleys. 98
Figure 9. Difference in the annotation system of pitch movements: Auto-segmental Metrical approach (left) and British Nuclear Tone approach (right). 100
Figure 10. The segmental, phonetic and phonological representations. (adapted from Silverman, 1987)..109
Figure 11. Relation between tonal target (pitch accents, phrase accents, and boundary tones) and phrase constituents divided into intermediate and intonation phrases. (From Liu, 2009: 2)..114
Figure 12. F0 contours for 'Marianna made the marmalade'. The same sentence is uttered in the top panel as a statement (H*H* L-L%,) and in the bottom panel as a yes-no question (L*L* H-H%). The recordings are from Beckman and Ayers (1993). 116
Figure 13. Marianna made the marmalade. The same sentence is uttered in the top panel as a default yes-no question (see example 2) and in the bottom panel with focus on the word marmalade (see example 3c). The recordings are from Beckman and Ayers (1993). 118
Figure 14. Straight-line approximation of possible pitch contours consisting of a combination of pitch accents, phrase accents and boundary tones. From Veilleux et al. (2006). 121
Figure 15. Distribution of English tonal patterns across declaratives and inversion questions in Cambridge and Belfast English. (From Nolan, 2006: 449)......................138
Figure 16. Standard settings for the calculation of LTD measures in Praat. 163
Figure 17. Screen view of the pitch line in the sound file of the sentence 'Do you need any money?', uttered by the female American speaker 1 165
Figure 18. Screen view of the pitch line in the pitch object of the sentence 'Do you need any money?', uttered by the female American speaker 1...165

Figure 19. Screen view of the pitch line in the pitch object of the sentence 'Do you need any money?', uttered by the female American speaker 1. 166
Figure 20. Screen view of the F0 stylization process where every local F0 maximum and minimum is signaled by a pitch point. Every pitch point receives a label. The manipulated sentence, 'You dropped down from the sky?', was extracted from the first passage of my corpus of sentences. 167
Figure 21. Linguistic measures by male speakers divided into three groups: American males speaking English (Am), Italian males speaking English (It-En), Italian males speaking Italian (It-It). 170
Figure 22. Linguistic measures by female speakers divided into three groups: American females speaking English (Am), Italian females speaking English (It-En), Italian females speaking Italian (It-It). 170
Figure 23. Estimated marginal Means of F0 mean distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects............................. 172
Figure 24. Estimated marginal Means of F0 max distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects............................. 173
Figure 25. Estimated marginal Means of F0 max distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects............................. 174
Figure 26. Estimated marginal Means of HZrange, that is F0 span calculated in Hz. Values are distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3).. 176
Figure 27. Estimated marginal Means of STrange, that is F0 span calculated in ST. Values are distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3).. 177
Figure 28. Patterns of yes-no questions (YNQ). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Do you need any money?' are compared. 185
Figure 29. Patterns of wh-questions (WHQ). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Where were you when the money ran out?' are compared....................................... 187
Figure 30. Patterns of statements (STM). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Now you are going away' are compared. 189
Figure 31. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along YNQ patterns individuated by pitch points corresponding to linguistic landmarks. 202

Figure 32. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along WHQ patterns individuated by pitch points corresponding to linguistic landmarks. 202
Figure 33. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along STM patterns individuated by pitch points corresponding to linguistic landmarks. 203
Figure 34. F0mean values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots)....210
Figure 35. F0max values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots)....212
Figure 36. F0min values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots)....213
Figure 37. STrange values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots)....214
Figure 38. Ratings for 'friendly' scores calculated for 64 listeners in 5 pitch level (i.e. register) conditions and 5 pitch span conditions. (From Chen, 2005: 83)....228
Figure 39. Values for F0 span across 30 English sentences (15 of them were produced by the Italians and the other 15 by the Americans)....235
Figure 40. Ratings of the stimuli obtained from the American raters (blue diamond) and the Italian raters (red square). The rating scores obtained for the stimuli (on the y-axis) are plotted against the values of F0 span (on the x-axis). 239
Figure 41. Ratings along the '± interesting' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners....242
Figure 42. Ratings along the '± exciting' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners. 243
Figure 43. Ratings along the '± credible' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners. 244
Figure 44. Ratings of YNQ. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel)....246
Figure 45. Ratings of STM. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel). 247
Figure 46. Ratings of EXM. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel). 248

Figure 47. F0 levels (F0 max, F0 mean, and F0 min) across the YNQ produced by American and Italian native speaker. ... 260
Figure 48. F0 span across YNQ produced by American and Italian native speakers. ... 261
Figure 49. Manipulated versions: original (standard), wide and narrow pitch span of the YNQ 'Do you need money?' produced by two American NSs (speaker 1 and 2). ... 263
Figure 50. Manipulated versions: original (standard), wide and narrow pitch span of the YNQ 'Do you need money?' produced by two Italian NSs (speaker 3 and 4). ... 264
Figure 51. F0 level across YNQ produced by an American (blue diamond) and an Italian (red square) native-speaker, in three pitch span conditions (narrow vs. original vs. wide pitch span). ... 265
Figure 52. F0 span across YNQ produced by an American and an Italian native-speaker, in three manipulated versions (narrow vs. original vs. wide pitch span). ... 266
Figure 53. Ratings of stimuli with manipulated pitch span. The stimuli are compared across native the language of the speakers (American vs. Italian) and pitch span variations (narrow vs. original vs. wide pitch span). ... 269
Figure 54. Ratings of males and females. The scores given by the male and the female listeners are compared across the native languages of the speakers (American vs. Italian) and the pitch span variation (narrow vs. original vs. wide pitch span). ... 272

List of Tables

Table 1. F0 level and span. 53
Table 2. Dimensions of tone and quantity within the perception and acoustic analysis (see Lehiste, 1970). 59
Table 3. Uncertainty and certainty model, based on Halliday and Greaves (2008: 62). 74
Table 4. Graphic representation of the frequency code that describes the directly proportional relation between degree of effort and F0 level and span. 76
Table 5. Graphic representation of the production code that describes the directly proportional 77
Table 6. Inventory of F0 measures calculated in Keating and Kuo's STRAIGHT method (2010: 172). 85
Table 7. Inventory of LTD and linguistic measures calculated for F0 level and span in Mennen et al. (2012: 2254). 89
Table 8. Correspondences between Pierrehumbert's annotation system and British-style nuclear tones (from Ladd, 1996: 82). 101
Table 9. *ToBI Contours for Standard American English.* Standard schematic representation of all the possible combinations which can occur in Standard American English *(*adapted from Hirschberg, 2002 and Ladefoged, 2006). 104
Table 10. Principles of the ToBI system (from http://www.ling.ohio-state.edu/~tobi/). 111
Table 11. Pitch accents and their typical meaning/mode. (From Pitrelli et al., 1994) 112
Table 12. 'Marianna'. The same word is uttered with four different tunes to convey different meanings in context. The drawings are from Harrington (2008), and the interpretations are based on Ladefoged (2006), Veilleux et al. (2006), Halliday and Greaves (2008), O'Connor (2008), and Truckenbrodt (2012). 120
Table 13. Typical intonation patterns in yes-no questions across Italian varieties. (From Grice, D'Imperio, Savino and Avesani, 2005: 367) 125
Table 14. Occurrences of American English yes-no questions tonal patterns calculated over 410 yes-no questions. (From Hedberg and Sosa, 2011: 847). 134
Table 15. Occurrences of American English wh-questions tonal patterns calculated over 200 wh-questions. (From Hedberg et al., 2010: 2). 135
Table 16. Personal data of the eight American participants (Sp) in the experiment. Information is about their native language, gender, age, birthplace, and major. 159
Table 17. Personal data of the eight Italian participants (Sp) in the experiment. Information is about their native language, gender, age, birthplace, and major. 159
Table 18. English and Italian version of the sentences selected from the second passage for the analysis of pitch range. 161
Table 19. Description of labels used to annotate the corpus. The method and guidelines proposed by Mennen et al. (2012) contain additional labels (!H*, D, $L*, U) that were not used for the analysis of the present corpus. 167
Table 20. Overview of linguistic measures for level. Mean values for each landmark were calculated in Hz for male (M) and female (F) subjects divided into three

groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It). 168
Table 21. Overview of linguistic measures for span. Mean values for each landmark were calculated in Hz for male (M) and female (F) subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It). 169
Table 22. Graph containing F0 max, mean and min values in Hz by male and female subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It). 171
Table 23. Span values in male and female subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It). 178
Table 24. Values for F0max, F0min, span, and duration of the STM 'Now you are going away', compared across language groups: American speakers (A7, A8, A9) and Italian speakers (I7, I8, I9). 190
Table 25. Personal data of the ten American participants (Sp) in the experiment. Information is on their native language, their sex, their age, status of smoker/non smoker, presence of any reported language disorder or impairments, and their major. 192
Table 26. Personal data of the ten Italian participants (Sp) in the experiment. Information is on their native language, their sex, their age, status of smoker/non smoker, presence of any reported language disorder or impairments, and their major. 193
Table 27. Personal data of the ten American speakers (Sp) in the experiment. Information is on their native language, years of L2 learning (Italian as an L2 for the Americans), languages learned and spoken at different levels, length of residency in a L2 country (Italy for the Americans). 194
Table 28. Personal data of the ten Italian speakers (Sp) in the experiment. Information is on their native language, years of L2 learning (English as an L2 for the Italians), languages learned and spoken at different levels, length of residency in a L2 country (the United States for the Italians). 195
Table 29. English sentences for the analysis of pitch range grouped according to sentence type: yes/no questions, wh-questions, and statements. 196
Table 30. Overview of linguistic measures for level. Mean values for each landmark were calculated in Hz for the American English (AmE) and the Italian (It) female subjects. Data were divided for each label into three groups depending on the sentence types: YNQ, WHQ and STM. 200
Table 31. Overview of linguistic measures for span. Mean values for each landmark were calculated in Hz for the American English (AmE) and the Italian (It) female subjects. Data were divided for each label into three groups depending on the sentence types: YNQ, WHQ and STM. 201
Table 32. F0 max, mean and min values in Hz by American and Italian subjects obtained in three sentence types: YNQ, WHQ, and STM. 205
Table 33. Span values across sentences (YNQ, WHQ, and STM) by the American and the Italian subjects. 206
Table. 34. Median values for F0mean, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). 211

Table 35. Median values for F0max, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). 212
Table 36. Median values for F0min, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). 214
Table 37. Median values for STrange, grouped depending on native language (Americans vs. Italian) and sentence type (YNQ vs. WHQ vs. STM)........................215
Table 38. Personal data of the 10 American participants in the third experiment. Information is about their native language, age, town, state, major, and gender..........233
Table 39. Personal data of the 26 Italian participants in the third experiment. Information is about their native language, age, town, area, class, major, and gender. 234
Table 40. Values for F0 span across YNQ, STM, and EXM. Mean F0 span values are calculated for utterances produced by Italians vs. Americans...................................236
Table 41. Rating scores of English stimuli produced by some Americans and Italians. The stimuli were rated by the Americans and the Italians across sentence types. 240
Table 42. Three versions of the same sentence: original version (non manipulated sentence), low-pass filtered version, and re-synthesized version with flat intonation. The manipulated sentence is ‘Do you wanna come for dinner?’ uttered by an American native speaker. 252
Table 43. Personal data of the 22 Italian participants in the fourth experiment. Information is about their native language, age, city, area, class, major, year, and gender. 258
Table 44. Ratings of YNQ with manipulated pitch span. The stimuli are compared across the native language of the speakers (American vs. Italian), pitch span variation (narrow vs. original vs. wide pitch span), the listening session (first and second repetition)...270

List of Abbreviations

AM	Auto-segmental Metrical
AmE	American English
ANOVA	Analysis of variance
AR	Articulation rate
ASD	Average syllable duration
BrE	British English
Dur	Duration
ERB	Equivalent rectangular bandwidth
EXM	Exclamations
F0	Fundamental frequency
F0max	Maximum fundamental frequency
F0mean	Mean fundamental frequency
F0min	Minimum fundamental frequency
FCM	Feature competition model
HRT	High rising terminal
Hz	Herz
Hzrange	Span measured in Herz
It	Italian
IP	Intonation phrase
iP	Intermediate phrase
L1	First language
L2	Second language
NLM	Native language magnet model
NNS	Non-native speaker
npause	Number of pauses
NS	Native speaker
nsyll	Number of syllables
PAM	Feature competition model
PhT	Phonation time
SLM	Speech learning model
SR	Speaking rate
SSB	Standard Southern British
ST	Semitone
STM	Statements
STrange	Span measured in semitones
ToBI	Tones and Break Indices
WHQ	Wh-questions
YNQ	Yes/no questions

Chapter 1:
Introduction

1.1 Suprasegmental features

In human communication, intonation plays an important role in providing speakers with a mechanism to arrange information, to disambiguate the meaning of sentences, to emphasize specific parts of discourse or to convey emotional message, such as a sense of participation, interest, detachment etc. (Ladd, 1996; Ladefoged, 1996, 2006; Jenkins, 2000; Chun, 2002; Wells, 2006; Levis and Pickering, 2008). Moreover, spoken language conveys meaning not only through words, but also through a wide range of other features, such as intonation, melody, prominence, phrasing, voice quality, speaking rate, rhythm, and timing. Several prosodic aspects such as pitch, stress and length (and their acoustic correlates F0, intensity and duration) are used by speakers to communicate this structural information (Lehiste, 1970; Kent and Read, 1992; Ladefoged, 1996). In fact, prosodic aspects of speech, including sentences, phrases, words, and syllables constitute a common core of factors that contribute to conveying different linguistic and paralinguistic meanings. Even though ‘it is known that certain prosodic aspects of speech play a role in the expression of paralinguistic meaning, yet the concrete mechanisms of how this is implemented have not yet been fleshed out’ (Prieto, 2011: 841).

Pitch, stress and length are prosodic features that combine together to perceptually make up intonation. These features are called suprasegmentals because ‘they can span regions larger than a single phonemic segment (consonant or vowel)’ (Veilleux et al., 2006: 1). The pitch range of an utterance yields variations of the fundamental frequency (F0), characterizing the different intonation patterns of an utterance (for a discussion on F0 and pitch, see § 2.1); the stress/unstressed status of a sound depends on its degree of intensity; while the length of a constituent refers to the extension of time of a specific syllable, word, phrase or utterance.

The term ‘suprasegmental’ refers to all those prosodic aspects that characterize speech. Unlike segmental features, suprasegmental features ‘are properties of speech sounds or their sequences that are simultaneously present, that do not change the distinctive phonetic quality of the speech sounds, but do modify the sounds in a way that may change the meaning of the utterance’ (Lehiste, 1996: 227). In particular,

> ‘in contrast to segmental features, suprasegmental features are established by a comparison of items in a sequence, whereas segmental features are indentifiable by inspection of the segment itself. For example, the rounding of a vowel in a sequence of rounded vowels can be established without reference to preceding or following vowels; however, the stressedness of a vowel cannot be established without reference to other vowels that carry relative weaker stress’
> (Lehiste, 1996: 227).

Suprasegmental aspects of spoken language convey both linguistic and paralinguistic meanings (Chun, 2002; Hirschberg, 2002; Ladefoged, 2006; Grice and Baumann, 2007), including distinctive semantic, syntactic or even morphological events. They can be helpful in a number of tasks, such as ‘mapping prominence and grouping patterns to meaning differences, understanding the effects of prominence and grouping on the pronunciation of words, and synthesizing prosodically natural-sounding speech’ (Veilleux et al., 2006: 1). In sum, segmental features refer to the vowel and consonant inventories of a language whereas suprasegmental features refer to prosodic and intonation aspects that are larger than and superimposed to segments.

1.1.1 The role of prosody

The term ‘prosody’ broadly refers to intonation, rhythm, timing, phrasing and stress in speech (Hirst, 2004, 2005; Veilleux et al., 2006; Busà, 2008; Shue, 2010). Prosody has two main functions: the groupings of words into distinct constituents and a prominence-marking functions due to pitch accent placement (Shue, 2010). Since prosody has to do with the ‘phrasal and tonal organization of speech’ (Keating et al., 2013: 1), it focuses on how phrasal and tonal constituents contribute to meaning.

Prosody is considered to have a great impact in communication. In fact, it affects the transmission of meaning by enhancing intelligibility in multi-language contexts (Nelson, 1992; Derwing and Munro, 1997; Hinkel, 2005; Pickering 2006; Busà 2008;

Kennedy and Trofimovich 2008). It also contributes to speech intelligibility and successful communication.

> 'Prosody is used in the disambiguation of structurally ambiguous sentences; it signals the information status of an utterance by highlighting given vs. new information, emphasis or contrast, etc.; it may define the speech function of an utterance […]; and finally it many convey paralinguistic information, for example with regard to the emotional state of the speaker […], the truth value of the proposition (e.g., certainty vs. uncertainty) or the level of the speakers' engagement.
> (Busà, 2008: 115).

Prosodic components are useful to emphasize prominent information, to distinguish between word meanings, to identify different sentence modes and to signal the timing of turn-taking in conversational speech.

They are realized within a hierarchical structure. Syllables, words, phrases and utterances are separated by prosodic boundaries that divide portions of speech into units. Fig. 1 shows the hierarchical representation of different prosodic levels: the intonation phrase (IP); the intonation phrase (ip); the word (w); the syllable (σ).

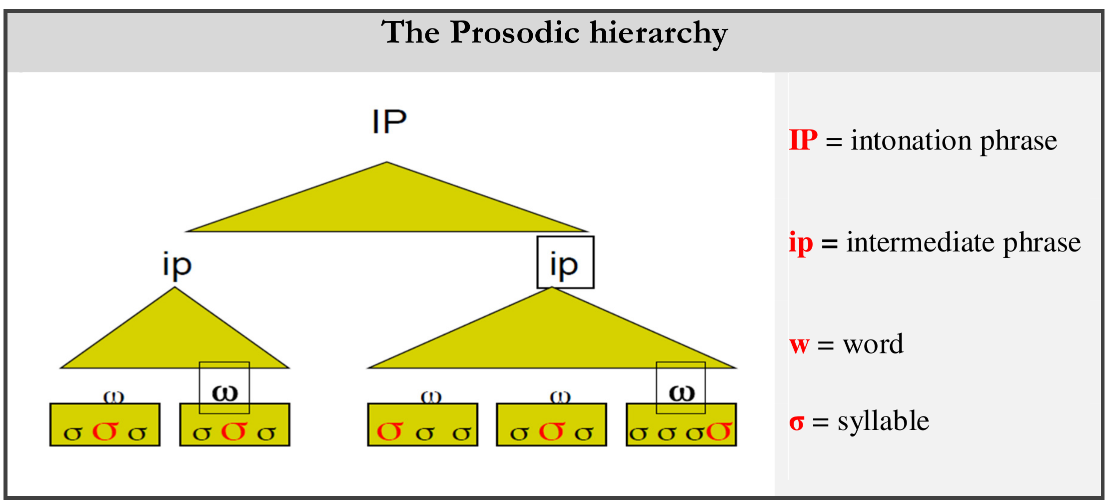


Figure 1 The prosodic hierarchy. (Adapted from Keating et al., 2003: 5).

Within the prosodic hierarchy, at least four different levels can be identified and placed in a schema (fig. 1). From the larger to the smaller units, the prosodic components are described as: the intonation phrase (IP); the intonation phrase (ip); the word (w); the syllable (σ).

The highest level, the intonation phrase, corresponds to the entire development of an utterance. Usually, intonation phrases are prosodically separated by boundary tones that are the tonal correlations of full stops. Intonation phrases contains one or more intermediate phrases, that work as smaller portions of information characterized by specific tonal movements (rises vs. falls). Words are the units that build up intermediate phrases. They can contain one or more syllables, that can be stressed or unstressed. The stressed syllables are more prominent than the unstressed ones (Tench, 1996; Chun, 2002; Halliday and Greaves, 2008) and they are often referred to as primary or secondary stressed syllables (depending on their relative prominence).

1.1.2 The functions of intonation

Intonation is a combination of multiple aspects of speech and its functions largely depend on the language, the style, the mood and the attitude of the speaker. According to Levis and Pickering 'the intricate modulation of the voice, with its ranges and movements of pitch, its subtle nuances of voice quality, and its expressiveness of staccato or lengthened syllables have often seemed to hold the key to language meaning' (2008: 506). As a result, listeners use intonation to predict utterances as they unfold (Féry et al., 2009) due to the fact that intonation is a key element to express a number of meanings that may vary across languages, cultures, and speaking styles.

Intonation refers to 'the pitch contour of an utterance' (Ahrens, 2005: 53), in which the modulation of the whole melodic pattern is characterized by changes in pitch level and span, intensity and duration. Generally speaking, a wider span on a rising pitch movement can sound as more questioning and more polite than a falling movement that may signal assertion or a less polite attitude (Ladd, 1996; Wichmann, 2000; Post et al., 2007; Halliday and Greaves, 2008). Thus the general movement of a sentence, that can be either rising or falling, contributes to project onto the listeners' imagination some expectations about the nature of the message that is going to be transmitted. For example,

> 'English has two basic intonation patterns: *rising* and *falling*. "Is Mr. Jones in?" has rising intonation. The pitch of the voice goes up at the end of the utterance. The speaker is asking a question. "No, he's not in" has falling intonation. The pitch of the voice goes down at the end of the utterance. The speaker is answering a question'.
> (Orion, 1996: 62–63)

Under this perspective, questions are likely to be uttered with rising patterns whereas statements are likely to be uttered with falling patterns. Despite this oversimplification proposed by Orion (1996: 62-63), the study of intonation is much more complex and challenging. A number of different phenomena may affect and modify intonation patterns, so the tune of a sentence is not based on a one-to-one correspondence with its grammatical function and mode. This means that not all questions are produced with rises and not all statements are produced with falls. Several theories and approaches (see the differences between the British and the American school, § 3.1.2) have been put forward to describe the nature and the use of intonation patterns. These different approaches have been pervaded by controversies and criticisms. According to Haan (2002) this is in part due to the fact that

> it hard to decide on appropriate unit(s) of description. Should one make inventories of intonation melodies covering entire utterances ('holistic tunes')? Or, rather, are tunes to be analysed as consisting of smaller component parts? At the same time, establishing the existence of smaller descriptive units seems to require at least some knowledge of which elements can be considered meaningful, and which cannot.
> (Haan, 2002: 24)

In order to better individuate and shape a framework for the analysis of intonation, a critical examination of its main components and functions must be considered. At least six major functions of intonations have been identified by Wells (2006): attitudinal, grammatical, informational, discourse, psychological, and indexal functions. However, 'the functions of intonation cannot be divided into neat, clear-cut categories since they typically involve the grammatical, attitudinal, information-structural, illocutionary, pragmatic, and sociolinguistic domains of conversations and discourses with much potential overlap' (Chun, 2002: 75). Thus, the functions of intonation relate to different linguistic, paralinguistic and even extra-linguistic domains.

From a strictly linguistic perspective, intonation has an informative function because it helps to distinguish new vs. old information, prominent vs. non-prominent constituents, broad vs. narrow focus, theme vs. rheme, etc. The effect of segmenting portion of speech is partly achieved by intonation movements that signal the beginning and the end of someone's speech in the regulation of the speaker-listener interaction (discourse function). What is more, different patterns are used to mark sentence types or grammatical units (grammatical function). Intonation has also a syntactic function because it separates different propositions and syntagma through tonal changes in the intonation tunes.

From a paralinguistic perspective, intonation is useful to express different emotions and reactions to what is being said, thus it has an attitudinal function. The interlocutors' attitude is expressed by conveying different emotional states through intonation variation (Mozziconacci and Hermes, 1999; Mozziconacci, 2002; Liscombe et al., 2003). Intonation has an indexal function, because speakers tend to mark their personal or social identity by controlling a number of variants within similar intonational patterns (consider the case of the high-terminal rises (HTR) in Australian and New Zealand English, § 3.3.2). Finally, intonation has also a psychological function because it helps organizing 'speech into units that are easy to perceive, memorize and perform' (Wells, 2006: 11).

From an extra-linguistic perspective, intonation may be influenced by physiological factors, such as the gender and age of the speakers, that determine a change in the characteristics of pitch (see the effect of gender and age on pitch variation, § 2.2.1 and § 2.2.2). Also problems related to breath control have an effect on intonation, because a muscular effort is required to modify the intonation melody. At every inhalation corresponds a resetting of sub-glottal pressure and F0 values (Hanson and Chuang, 1999; Vaissière, 2005; Keating et al., 2012). Also phonation contrasts across languages (Gordon and Ladefoged, 2001; Keating et al., 2011) are likely to have an influence on the intonation patterns used in different languages.

In sum, intonation is used by speakers to segment continuous speech into informational units (grammatical, syntactic, lexical, pragmatic) and to express their attitudes toward listeners and towards what is being said. In addition to the attitudinal, grammatical, informational, discourse, psychological, and indexal functions already mentioned, Vaissière (2005) suggests that intonation and particularly pitch modulation contribute to provide a global representation of the characteristics of speakers, thus defining their 'identity, sex, age, physiological state, regional varieties, stylistic variations, socio-cultural background, etc.' (Vaissière, 2005: 239).

Indeed, intonation conveys meaning through different channels and performs a number of functions. The meanings expressed by intonation patterns are believed to be 'universal and language-specific at the same time for any language' (Chen 2001: 43). According to Gussenhoven (2002), the universality of intonation is found in the ability to signal emotions and attitude (universal paralinguistic meaning) by modifying the phonetic detail (that is, the pitch range). For example, a high pitch level is cross-linguistically perceived as more uncertain and questioning than a low pitch level. The intonational meaning is also language-specific because cross-linguistic differences are found in the realizations of questions and statements through different rising vs. falling patterns. Phonologically distinct intonation contours are used in different languages to express the same meanings.

Intonation and particularly pitch excursions appear 'to be relevant to capture certain intonational variation between dialects, yet more research on pitch excursion is needed to fully understand its role with respect to pitch accent realization' (Kügler, 2009: 408). Sometimes it is difficult to decide which variations in frequency and intonation patterns are intended by the speakers. Not only phonological changes but also phonetic changes can have significance (Ladefoged, 2006). Recently, the phonological role of intonation has been investigated cross-linguistically. However, little is known about the phonetic meaning of intonation.

1.2 The dichotomies of pitch range

Pitch modulation carries crucial information for the phonetic analysis of intonation. In particular, pitch range is considered the main correlate of intonation (Bolinger, 1986; Beckman, 1996; Cruttenden, 1997; Haan, 2002; Vaissière, 2005). The characteristics and uses of pitch range vary across languages and cultures. However, some major trends have been recognized and described by Ladd (1996). He called them *Intonational Universal* and formalized them as follows:

1. 'The tendency of pitch to drop at the end of an utterance, and to rise (or at least not to drop) at major breaks where the utterance remains incomplete;
2. The use of higher pitch in questions, since in questions the speaker expresses interest, and since the exchange is incomplete until the addressee answers;
3. The use of local pitch peaks (e.g. pitch accents) on words of special importance or newsworthiness in an utterance'. (Ladd, 1996: 114)

The analysis of pitch range is intricate and captivating for at least four reasons. First, pitch range is considered the perceptual correlate of an acoustic property of sounds called fundamental frequency (henceforth F0). F0 reflects changes in the vocal fold vibrations while pitch accounts for these variations as they are perceived by listeners (Ladd, 1996; Baken and Orlikoff, 2000; Johnson, 2003; Vaissière, 2005; Halliday and Greaves, 2008; Coleman, 2011; Keating et al., 2011). Thus, it is crucial to understand the subtle but fundamental difference between F0 and pitch range (see §2.1 for a detailed examination of the differences between F0 and pitch).

Second, pitch range has been erroneously considered as a single unitary measure, while, actually, it is the result of two different dimensions: level and span (see § 2.2.1). While pitch level is a sort of reference line calculated over the rises and falls within each intonation contour, pitch span is a measure of the distance between the highest and the lowest F0 value in the contour (Ladd, 1996; Cruttenden, 1997; Gussenhoven, 2004). Hence, the description of pitch range variation is given by the sum of the changes in pitch level and span.

Third, previous research on pitch range suggested that the perception and the production of pitch variation have proper and distinctive features (for an overview see t' Hart et al., 1990). Thus, 'the non-trivial relation between perception and production should be treated seriously' (Li, 2010: 3), when analyzing the characteristics of pitch range across languages and population. Some acoustic features of pitch range may be relevant in a study on production data and totally irrelevant in an study on perception data. Changes in the modulation of pitch patterns may be linguistically or communicatively relevant, depending on the kind of measures investigated. For example, purely acoustic studies can show dramatic changes of pitch range that may not be noticed and perceived by listeners. Thus, 'some apparently major pitch event may play a negligible role in perception while, conversely, a seemingly minor phonetic detail may prove indispensible' (Haan, 2000: 24).

Fourth, pitch range has a substantial impact on distinguishing native speakers from non-native speakers of different languages. Thus, pitch range might also have a role in the perception of foreign accent. However, unlike other prosodic cues such as stress, rhythm, speech rate, etc., it is not clear to what extent pitch range may contribute to the detection of accented speech produced by L2 speakers. What is evident is that the speakers of different languages seem to have distinctive characteristics of pitch range (Gussenhoven, 2002; Chen, 2005; Mennen et al., 2012).

This study is aimed at comparing the pitch range of Italians and Americans in order to provide evidence for cross-linguistic differences between L1 and L2 speakers. The focus of the analysis is based on the double nature of pitch range (i.e. pitch level and span) that is examined with an experimental approach relying on production and perception data.

1.2.1 Pitch range in L2

A number of studies have addressed the problem of measuring pitch range. However, documenting cross-linguistic differences in pitch range has proven to be difficult, because pitch is subjected to a wide range of inter-speaker and within-speaker variability, with data of speakers from different languages often overlapping. Studies in L2 intonation have shown that 'L2 learners tend to use the standard pitch range of their native language in their L2' (Chen, 2004: 38). What is more, L2 language learners have consistently been reported as speaking with a narrower pitch range and less pitch variation than L1 language speakers (Mennen, 2006).

Cross-linguistic differences in pitch range have been reported especially in relation to second-language speech. For example, it has been claimed that American men speak at a lower pitch than German men (Scherer, 1979), Japanese women speak at a higher pitch than American women (Ohara, 1992), and Dutch spoken by Belgian women has an higher pitch than Dutch spoken by Dutch women (van Bezooijen, 1993). However, generalizations cannot be drawn on the basis of few studies carried out on data that are not easily comparable. In particular, the focus of the majority of works on pitch range is on L1 rather than L2. In this perspective, researchers have tried to give a more comprehensive overview of pitch range in L2 English.

As shown in fig. 2, the influence of L1 on L2 English pitch range has been documented in a study on vowel F0 and duration. The pitch contours of L2 English produced by German and Russian speakers have been found to be much more similar to L1 English than to L2 English by French, Italian and Spanish speakers. This shows about the fact that Germans and Russians are more capable than French, Italians and Spanish people to approach the pitch range model of the British native speakers. One of the reasons why there is this striking difference may be that while English, Russian and German are stressed-time languages, Italian, Spanish and French are syllable-time languages.

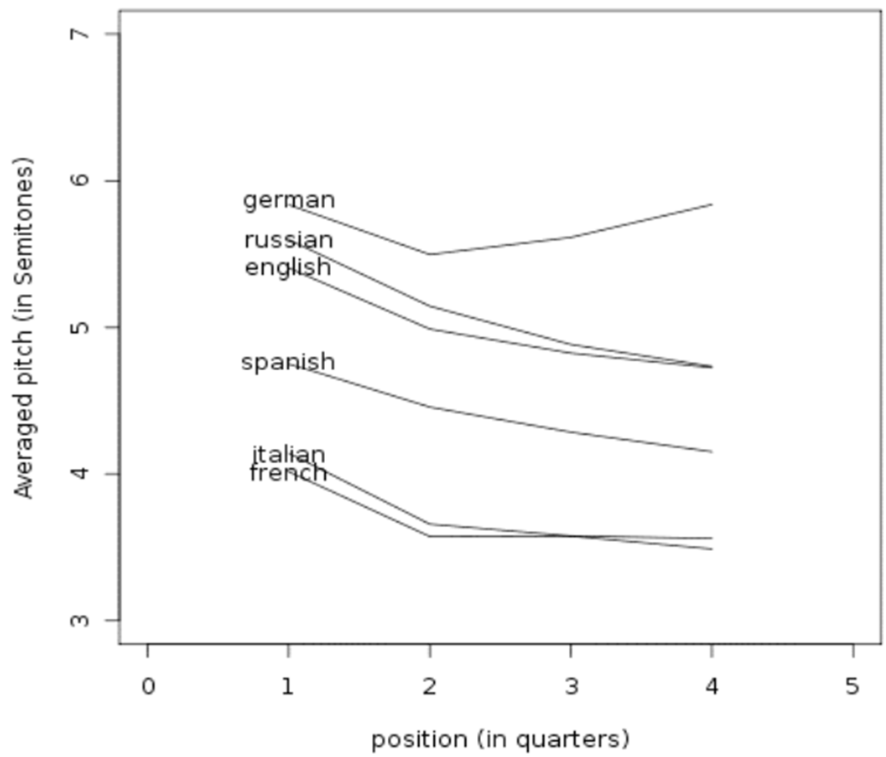


Figure 2. Normalized pitch contours of stressed vowels showing L1 influence on L2 English. (From Yuan, 2010: 1848).

Even though German and Russian learners of English have an average pitch similar to the one of English native speakers, they are far from being able to reproduce English pitch range. This has been shown in a study by Mennen et al. (2008b, 2012) who compared the production of pitch range by Southern Standard British English speakers (SSBE) and Northern Standard German speakers (NSG), by proposing a new methodological approach in order to quantify pitch range across languages.

The study found that SSBE have higher and more varied pitch range than the NSG. It is claimed that British high-pitched voices (especially female) may be perceived as "over-excited" or even "aggressive" by German listeners. On the contrary, German low-pitched voices may sound "bored" or "unfriendly" to British listeners (Mennen et al. 2008b, 2012). By contrasts, in a study aimed at comparing the production of pitch range by British and Dutch speakers, Chen (2009) showed that at identical pitch ranges, British

English speakers are perceived as more confident and friendlier than Dutch speakers. This may be due to the fact that Dutch speakers to use a narrower pitch range than British English speakers when reading sentences in English (Willems, 1982).

A further indication of pitch range effects is found in cross-linguistic studies on the pitch range of English and Japanese. The results of a study by Aoyama et al. (2007) support the hypothesis that Japanese speakers rely on pitch differences to indicate stress in English. In addition, van Bezooijen (1995) hypothesizes that pitch differences between Japanese and Dutch women are due to socio-cultural aspects and she reports ‘a preference for high pitch in women in Japan and for medium or low pitch in women in the Netherlands’. As for Finnish learners of Russian, Ullakonoja (2007, 2010) shows that Finnish speakers of Russian L2 use a narrower pitch range and a less variable pitch than Russian L1 speakers, though they show more Russian-like patterns in advanced stages of language learning.

As far as Italian is concerned, Busà (2008a, 2008b) suggests that Italian speakers' intonation in English is characterized by rather level and unvaried contours in different sentence types (e.g. yes-no questions, wh-questions, statements etc.). By contrast, English NSs use a variety of intonation contours and modulate their pitch range depending on sentence types. Since in English a level contour is typically associated with boredom or lack of interest (Hirschberg, 2002 and Wells, 2006), Italian speakers of English are likely to be perceived as unengaged in the conversation or lacking of interest. In addition, the extensive use of the ‘default’ level contour may contribute to creating a distorted image of the Italian speakers' attitude or emotional state in specific communicative contexts. In line with the reviewed research, one may infer that no matter which languages we deal with L2 speech may be characterised by a narrower pitch range than L1 speech.

These data suggest that pitch level and range may represent a language-specific meaning component which is affected by the principle of the ‘Frequency Code’ formulated by Ohala (1983, 1994). According to the principle of ‘Frequency Code’ (for a review see § 2.4.1), high pitch is associated with smallness, politeness and lack of threat while high pitch is associated with largeness, assertiveness and threatening intent (1994). Despite the universal components of the frequency code, some differences in pitch range that were reported cross-linguistically might be due either to inherent characteristics of the L2 learners or to the specificity of each language.

> For example, the fact that both Dutch and Spanish speakers acquiring English intonation produce a smaller pitch range compared to native English speakers does not necessarily indicate that a reduction of pitch range is a universal tendency in L2 acquisition. The smaller pitch range in the data of the learners could simply be a case of transfer, since both Dutch (Jenner, 1976) and Spanish (Stockwell and Bowen, 1965) are reported to have a smaller pitch range than English. It is therefore more likely that there is more than one process involved in the acquisition of L2 intonation, a conclusion which has also been reached in other fields of L2 acquisition.
> (Mennen, 1999: 42)

1.2.2 Production and Perception

A rigorous and systematic investigation of pitch range should be approached with a methodology based on production and perceptual tests. This is in line with the idea that 'it does not seem possible to infer production abilities from perceptual ones and vice-versa' (Llisterri, 1995: 98). In fact, it has been shown in current research that the natural acoustic realization of an underlying target can be realized in different phonetic forms whose impact on the listeners in not identical (Li, 2010). Generally speaking,

> 'stage in the acquisition of L2, experience with the language, degree of exposure, and age of acquisition seem to play a major role in the interaction between production and perception in L2. The relation between production and perception might differ according to the class of sounds, to the acoustic and perceptual correlates of these classes and to contextual effects. Similarity between L1 and L2 sounds might also have an effect on the interplay between production and perception'.
> (Llisterri, 1995: 98)

The study of pitch range can be approached from various angles and through various methods. As for production, measurements of the fundamental frequency (F0) are done on the assumption that pitch data are objectively comparable and reveal robust acoustic differences. However, the plain analysis of acoustic measures overlooks and fails to notice whether or not differences in pitch range may be perceptually and communicatively relevant. The acoustic characteristics of pitch range do not always correlate with the perceptual ones. To grasp this double dimension of pitch range, it is necessary to map both the acoustic and perceptual features in a series of production and perception studies.

For this reason, I tested the potential differences of pitch range across a number of measures by examining production and perception data. Previous research ('t Hart et al., 1990; Cruttenden, 1992; Grice 1995; Llisterri, 1995; Ladd, 1996; Ladd and Morton, 1997) showed that a combination of production and perception approaches permits to obtain reliable results that are calculated by measuring different parameters and factors. The fact that pitch range variation is determined not only by acoustic measures but also by perception judgments given by listeners should not be underestimated. The idea is that 'not all phonetic detail carried by the speech signal has perceptual relevance; from part of it, the listener abstracts away' (Haan, 2002: 25). The advantage of validating the pitch range differences found in production tests with perceptual experiments is that 'only listening tests can provide reliable behavioral data on perceptual equivalence, on the one hand, and functional equivalence (within one language, and cross-linguistically) on the other hand' (Vaissière, 2005: 253).

Based on a collection of production and perception data, this work provides an overview of F0 modulation from an acoustic and perceptual perspective. I analyze two dimensions of pitch range (pitch level and pitch span) by comparing F0 measures in two perception studies whose goal is to determine the realization of pitch excursions. The results obtained from these studies is implemented by designing perception tests in which listeners have to indicate how they perceive original and manipulated sentences. It is expected that the results of the production and perception experiments underpin the basic features of pitch range variation across American and Italian speakers.

1.3 Motivation

This dissertation investigates the characteristics of pitch range in English, as spoken by American native speakers and Italian learners of English. More specifically, the goal of this study is to unfold the specific features of pitch range from an acoustic (see production studies, chapter 4) and a perceptual point of view (see perception studies, chapter 5). Since pitch range has a twofold dimension (acoustic and perceptual), a study focusing on only one of these aspects would only be partial. Thus, the objectives set for these experiments are quite ambitious. On the one hand, the pitch range of Italians and Americans is compared in order to shed light on the acoustic differences between utterances produced by native speakers and non-native speakers; on the other hand, the perceptual effect of the differences found in the Americans' and Italians' productions is tested to see whether or not pitch range may be a factor contributing to the stereotyping of L2 speakers.

Indeed, the motivations for studying pitch range across intonation contours are multiple. In multilingual contexts, namely in interactions between speakers who have different mother-tongues, speakers can exploit pitch range patterns in different ways, sometimes not obvious at first sight. Pitch range varies highly from speaker to speaker depending on prosodic elements, emotional states, physiological reasons, individual characteristics of the voice quality, and specific features of the spoken regional variety. Thus the dynamic nature of pitch range is correlated to different physiological, psychological and sociological dimensions. The complexity of pitch variation and the difficulty to discuss it empirically makes the investigation of pitch range a challenging issue. The choice to study pitch range is thus motivated by the fact that this topic appears to be still elusive and an open debate has been raised within the linguistic community.

Recently, different models for measuring pitch range variation have been proposed by scholars who attempted to find consistent correlations between pitch level, pitch span and pitch variability (Ladd et al., 1985; Ladd, 1996; Shriberg et al., 1996; Patterson and Ladd, 1999; Patterson, 2000; Clark, 2003; Ladd and Schepman, 2003; Mennen et al., 2007, 2008, 2012; Bishop and Keating, 2010, 2012; Keating and Kuo, 2010, 2012). Even though every method proposed is aimed at fully capturing the complex realization of pitch range, within the scientific community there is no agreement on a standard model for analyzing pitch range.

To that end, there is a need for large-scale comparable data investigating the extent to which pitch varies across genders and languages. The lack of a generally recognized system for the analysis of pitch leaves a considerable gap in the literature on pitch range, though various models for the analysis of pitch range have been proposed. For this reason, I approached the study of pitch range by examining previous theoretical models of L2 speech learning and acquisition (see SLM in Flege, 1995; PAM in Best, 1995; NLM in Kuhl and Iverson, 1995; FCM in Hancin-Bhatt, 1994) and intonational phonology description (see Bruce, 1977; Pierrehumbert, 1980; Pierrehumbert and Hirschberg, 1990; Silverman et al., 1992; Beckman et al., 1994, 1995; Pitrelli et al., 1994; Ladd, 1996; Beckman and Ayers, 1997; Cruttenden, 1997; Ladefoged, 2006; Veilleux et al., 2006; Harrington, 2008).

The complexity of pitch variation is determined by the number of factors that influence changes in pitch production and perception. Pitch range is affected by the physical or anatomical characteristics of different speakers (Kuwabara and Sagisaka, 1995; Ho, 2001). Physical characteristics are based on the gender of speakers (Henton, 1989; Traunmüller and Eriksson, 1995; Cruttenden, 1997; Daly and Warren, 2001; Honorof and Whalen, 2005; Dilley and Brown, 2007; Hillenbrand and Clark, 2009; Lee, 2009; Shue, 2010) and other factors that may be altered along the years due to aging process (Hollien and Ship, 1972; Cooper and Sorensen, 1981; Cruttenden, 1997; Baken and Orlikoff, 2000), body size (Graddol and Swann 1989; Titze, 1989; Shutter et al., 1996; Yuasa, 2008), health conditions including a reported history for smoking habits (Gilbert and Weismer, 1974) and incidence of language impairments or illnesses such as, aphasia (Danly et al., 1979; Cooper and Zurif, 1981), hearing-impairment (Nickerson, 1975), autism (Paccia-Cooper and Curcio, 1980) and depression (Nilsonne et al. 1988). Also socio-cultural factors affect the speaking styles of different speakers. Socio-cultural effects may be caused by semi-permanent factors, such as residential area, education and social expectations (Ohara, 1992; Dolson, 1994; van Bezooijen, 1995; Grabe, 1998; Ho, 2001; Kroløkke and Søresen, 2005; Yuasa, 2008; Mennen et al., 2012) and temporary-factors, such as emotional attitude or speaking purpose (Traunmüller and Eriksson, 1995; Hincks, 2005; Rosenberg and Hirschberg, 2005; Strangert and Gustafson, 2008).

Recent experiments have been aimed at investigating the characteristics and perceptual effects of pitch level and span, as well as their differences across languages (for a review see § 2.2.1). However, it is still unclear to what extent pitch range can be a relevant prosodic cue to identify cross-linguistic differences across speakers. In my opinion, the study of pitch range is not only a fundamental contribution to pure linguistic research but also it has an impact on the treatment of language impairments, in the domain of forensic phonetics, language teaching, speech synthesis and speech recognition. What is more, a systematic analysis of measures of pitch range results in the development of innovative speech analysis software. Programs for the analysis of pitch are not only the outputs of linguistic research but also the conditio sine qua non this kind of analysis can be pursued. In fact, an experimental investigation of pitch cannot be carried out without the help of speech analysis software and sophisticated equipment. In the teaching domain, 'it has been found that suprasegmentals can be most effectively taught through the use of equipment which extracts pitch and intensity from the speech signal and presents the information on a video screen in real time, providing instantaneous visual feedback on stress, rhythm, and intonation' (Anderson-Hsieh, 1994: 6).

Generally, pitch level and pitch span are correlated and covary to a large extent: the higher the pitch level, the higher the pitch span (Rietveld and Vermillon, 2003). Thus

> 'pitch can be thought of as a continuous variable; it occupies a sliding scale from low to high and is linked to many paralinguistic phenomena, such as gender and emotion. Independent of any prosodic function pitch may have, it is know that pitch can be used to form phonological categories in lexical tone languages. […] In languages without lexical tone, such as English, pitch is used to mark prosodic prominence and boundaries'.
> (Epstein, 2002 : 4)

It is clear to me that despite the large number of studies conducted on pitch range, very few analyses have shed light on the double dimension (pitch level and span) of pitch modulation. This might be due to the fact that most studies have focused on a phonological approach to the description of pitch and not a phonetic one. It is important to extend any current phonological model of pitch variation (see the Auto-segmental model, § 3.1) to the theoretical examination of the phonetic variation in pitch range, as intonation patterns are marked by both phonological and phonetic factors.

> 'An example of *phonological* influence is the use of rises where native speakers would use falls and vice versa, found in many studies of L2 intonation (e.g. Adams and Munro 1978; Backman 1979; Jenner 1976; Lepetit 1989; Willems 1982). An example of *phonetic* influence is the finding of a different pitch range (e.g. Mennen, this paper) or a different slope of a rise (e.g. Ueyama 1997) compared to the monolingual norm'.
> (Mennen, 2006: 4).

A model for the analysis of pitch should be primarily focused on the phonological and phonetic realization of pitch variation. This consideration is based on the idea that a phonological contrast manifests differences in the inventory of pitch accents (see the ToBI notation, § 3.2) while a phonetic contrast results in a difference in the phonetic realization of similar pitch patterns (Ladd, 1996). The importance of the phonetic approach has been pointed out by one of the most authoritative theorizers of the Auto-segmental and Metrical approach, who claimed that 'a continuous phonetic representation can never be replaced by even the most detailed symbolic encoding of pitch events'

(Beckman et al., 2005: 37). For this reason, the present study of pitch variation yields to the phonetic representation of pitch across languages.

1.4 Research outline

The goal of this work is to investigate how pitch range contributes to prosodic differences in American and Italian. Experiments are carried out on the production and perception of pitch range by American and Italian speakers.

Chapter 2 begins the discussion with an introduction to the acoustic analyses of pitch range. The notion of fundamental frequency (F0) and pitch are defined from an acoustical and perceptual perspective. Pitch level and span are identified as the two major components of pitch range which can be measured with linear or non-linear frequency scales and is affected by segmental features. Also speaking rate and the duration of utterances plays a role in determining higher or lower F0 slopes. Despite the lack of a generally acknowledged methodology for the analysis of pitch range, four of the most accredited methods proposed in recent years (Ladd, 1996; Patterson, 2000; Keating, 2010; Mennen, 2012) are outlined and discussed.

Chapter 3 presents the outcome of cross-linguistic research on pitch variation across different languages. A summary and critical overview of previous studies conducted on first and second languages are provided. This is done by taking into consideration the inventories of intonation patterns reported for Italian varieties and Romance languages (i.e. Spanish, Catalan, French, European and Brazilian Portuguese) compared to English varieties (i.e. American English, British English, Australian and New Zealand English) and other Germanic languages (i.e. German and Dutch). The examination of previous studies on pitch range variation is motivated by the necessity to critically investigate the pitch differences envisaged in Italian and English. The theoretical framework used for the phonological analysis of pitch range is based on the Auto-segmental Metrical approach. Four L2 acquisition models are critically analyzed: the speech learning model (SLM); the perceptual assimilation model (PAM); the native language magnet model (NLM); the feature competition model (FCM).

Chapter 4 describes the goals and the achievements of the two production tests devised for this study. The first comparative study explores the pitch range of men and women in American English (L1), English (L2), and Italian (L1). The pitch range values produced by the Italians speaking English are shown to be significantly different from those produced by the American native speakers and the Italian native speakers. Generally, the English sentences produced by the Italians are found to have a narrower pitch span than those produced by the Americans. This is due to two possible reasons. First, regardless of the nature of the languages investigated, L2 speech is always characterised by a narrower pitch range than L1 speech. Second, Italian learners of English are influenced by their L1 and transfer their L1 pitch range variation into their L2. The second comparative study presents and discusses different realizations of pitch patterns across three sentence types: yes-no questions, wh-questions and statements. The main goal of this experiment is to identify whether or not American and Italian females adopt a similar pitch range (that represents phonetic details) across different phonological patterns. Hence, the development of pitch analysis is considered across a phonetic and a phonological interface. In order to grasp the whole aggregate of aspects influencing pitch variation, data on the duration, phonation time, speech and articulation rate, and number of syllables in the utterances are provided, too. Despite the similarities in phonological patterns produced by the American and the Italian subjects, meaningful phonetic differences are found across sentence types.

Chapter 5 examines the perceptual impact of the differences observed in the production studies. The general hypothesis at the basis of these experiments is that both American and Italian listeners may display a similar understanding and evaluation of pitch range variation. The third comparative study examines the correlation between pitch span variation and the perception of different sentence types, yes-no questions, statements, and exclamations by American and Italian listeners. The results show that the stimuli with larger pitch span are perceived as more interesting, exciting, credible and friendly than the stimuli with narrower pitch span. The fourth comparative study verifies whether or not there is a correlation between the evaluation of listeners and pitch span variation. Original stimuli uttered by native and non-native speakers of English are manipulated and re-synthesised to create three pitch span conditions: narrow span, original span, and wide span. Variation in pitch span is found to be a robust cue for the perception of friendly/unfriendly questions. Regardless of their native language, both the American and the Italian speakers were judged to be more friendly when the pitch span of their sentences was widened (wide span condition) and less friendly when the pitch span was narrowed (narrow span condition).

Chapter 6 provides concluding remarks and a general summary of the results obtained in the production and perception studies. The data analyzed in the production studies (chapter 4) are interpreted in the light of the findings obtained from the perception studies (chapter 5). Pitch range variation is acoustically examined and perceptually determined, because the acoustic differences observed in the inspection of LTD and linguistic measures are perceptually validated by the judgments of listeners. Despite the complex, elusive and problematic nature of pitch range analysis, the studies proposed in this investigation contribute to the understanding of how pitch varies across languages, cultures, and genders.

Chapter 2:
Acoustic analyses of pitch range

2.1 F0 and pitch

The notion of F0 is strictly connected with the notion of pitch. In the literature, these terms have assumed different meanings or connotations due to controversial descriptions or characterizations. For this reason, a comprehensive picture is presented in the next paragraphs which critically examine F0/pitch definitions and measures.

In speech, 'fundamental frequency (F0) conveys linguistic information (e.g. tone and intonation), but it does so simultaneously with paralinguistic (e.g., emotion, emphasis) and non-linguistic (physiology) information about the speaker' (Bishop and Keating, 2012). Acoustically, F0 depends on the number of vocal fold vibrations in the larynx and can be expressed as cycles per second. It is measured acoustically in Hertz (Hz). 'The more quickly the vocal folds vibrate, the higher the F0' (Ladefoged, 2003: 168). The way to calculate F0 is to divide one second by the duration of one cycle (repetition of a sinusoidal pattern). Technically speaking, F0 is expressed by the following formula:

(1) $F0 = 1/T$

where T is the period in seconds (Johnson, 2003).

F0 is an acoustic property of complex periodic waves that can be measured via the auto-correlation technique. This technique is based on a method that computes fundamental cycles in the waveform (see Ladefoged, 1996: 148-151 and Ladefoged, 2003, chapter 4)[1].

[1] For a description of the calculation of F0 explained in Italian, I would recommend Giannini and Pettorino, 1981 (chapter 5).

The pitch track function of several speech analyses packages 'shows estimates of F0 of vocal folds vibration as a function of time' (Johnson, 2003: 30).

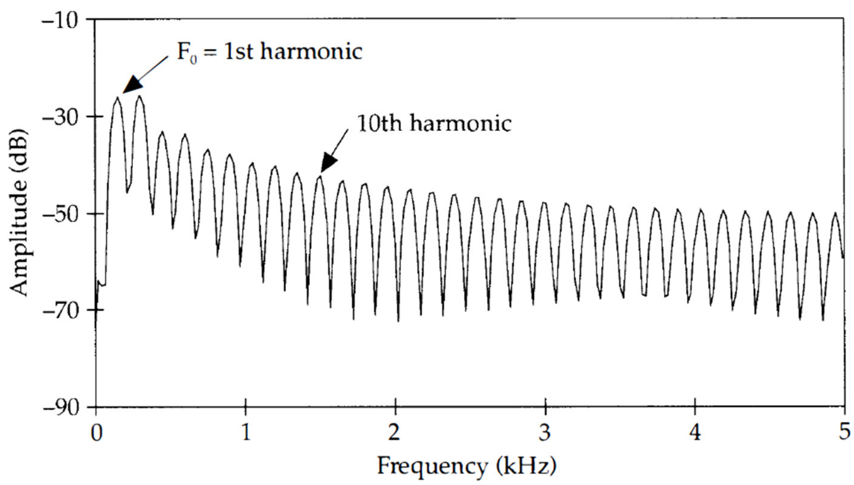


Figure 3. A harmonic spectrum of the vocal fold vibration showing the F0 (first harmonic) and the 10^{th} harmonic. (This figure has been taken from Johnson, 2003: 80).

As fig. 2 illustrates, in a harmonic spectrum of a voiced sound, the lowest frequency component (peak in the spectrum) is F0 and higher frequency components (peaks) are associated to higher harmonics. All are equally spaced. As Johnson asserted 'F0 is 'the first (lowest-frequency) peak in the power spectrum, and each of the other peaks in the spectrum is at a multiple of the fundamental frequency' (2003: 79). Thus, the lowest harmonic is the fundamental (F0), heard as pitch of the voice.

Variation in F0 depends on a number of physical and paralinguistic factors. In her class on Experimental Phonetics (Fall, 2012), Prof. Keating explained that the rate of vocal fold vibration depends on: (i) overall size of the vocal folds (the bigger the folds, the slower the vibration), (ii) stiffness of the vocal folds (the stiffer the folds, the faster the vibration), (iii) the amount of air going through the glottis (the more air passes through the glottis, the faster they vibrate). Technically speaking, F0 varies with individual anatomy, vocal folds state and position, and airflow conditions. As a result, speakers have different individual F0 and they are comfortable at producing a particular F0 range depending on their vocal folds size. Variation within a F0 range is achieved by diminishing or increasing airflow in the glottis, or just by stretching or tensing the vocal folds. Strategies for measuring F0 include: zero-cross detection, peak-picking, waveform matching etc. (see Baken and Orlikoff, 2000: 154-157).

Halliday and Greaves (2008: 29) have argued that F0 and pitch are ‘closely related terms dealing with the same phenomenon, the transmission of sound in air, but [are analyzed within] different frameworks’. Ladd (1996: 7) further discussed these different frameworks, or better dimensions, by underlying that ‘F0 is a physical property and pitch is its psychophysical correlate’. Thus, while F0 deals with the physical aspect of speech, pitch focuses on the way it is perceived by humans. Baken and Orlikoff argued that a correlation is found between F0 and pitch as ‘pitch (a perceptual attribute of sound) increases with fundamental frequency (a physical parameter of vibration)’ (2000: 148). In other words, pitch is a psychological percept based largely on physical F0 and it represents the perceived fundamental frequency of a sound. Even though F0 and pitch are two distinct concepts, most of the time these are used interchangeably.

F0 variation is particularly interesting in tonal descriptions, because it gives an accurate and precise description of the fall and rise of tones. Since speakers have different individual pitch ranges, it is not possible to establish an absolute value of F0 on a given tone. The incidence of rising or falling tones can be measured acoustically by indicating how sharply the rise or the fall occurs. When a sound is perceived as high-pitched, it means that it contains a high frequency of vibration. By contrast, a sound is perceived as low-pitched when it contains a low frequency of vibration. Thus, pitch is a property of a sound that enables it to be ordered on a scale going from low to high or vice versa (Halliday and Greaves, 2008).

According to Coleman (2011), the basic properties of F0 can be listed as follows:

a) F0 corresponds to rate of vibration of the vocal cords;

b) During unvoiced speech (during voiceless consonants as well as pauses), F0 = 0;

c) F0 is discontinuous;

d) The overall shape of the F0 contour is under the conscious control of the speaker, but some speech sounds introduce fine-scale micro-prosodic perturbations, often due to aerodynamic factors;

e) Speakers do not usually use their full pitch range in speech. The actual range may vary e.g. be larger in more animated speech..

f) A speaker's pitch range may fall or rise during speech, independently of the falls and rises of F0.

(From Coleman, 2011 see http://www.phon.ox.ac.uk/jcoleman/intonation_f0.htm).

In addition to the basic properties of F0 outlined by Coleman (2011), also phonation types set some specific limits to F0 range variation. In fact, modal, creaky and breathy voicing are the three main types of voicing occurring cross-linguistically (Hollien, 1974; Titze, 1988; Baken and Orlikoff, 2000; Johnson, 2003). Phonation contrasts across languages are calculated over a number of measures (Keating et al., 2011). The modal phonation, has been limited to a specific pitch range. F0 values occurring under this range are typical of creaky voice while values occurring over this range are typical of falsetto voice. The F0 baseline for modal phonation is set at 70 Hz, with voicing under 70 Hz defined as creaky phonation (or vocal fry). The F0 topline of modal phonation is set at 175 Hz for males and 275 Hz for females, with voicing over those values defined as falsetto or breathy phonation (Johnson, 2003).

2.1.1 Pitch range: level and span

As languages are different in terms of prosodic properties, they also differ in their pitch range. Pitch range has been defined as 'the range of pitch employed by a particular speaker at a particular time and can be specified by a minimum and maximum pitch' (Clark, 2003: 9). Being such a crucial concept in acoustic analysis, pitch range has been largely investigated and reviewed by Ladd (1996, chapter 7). He claimed that pitch range has been erroneously considered as a single unitary measure, while, actually, it is the result of two different dimensions: level and span. While pitch level is a sort of reference line calculated over the rises and falls within each contour, pitch span is a measure of the distance between the highest and lowest F0 value in the contour (Ladd, 1996; Cruttenden, 1997; Gussenhoven, 2004)[2].

Most studies have focused on the acoustic analysis of F0 range and on the perceptual correlations of F0 distribution (that is to say 'pitch distribution'). Nonetheless, as Mennen et al. (2012: 2249) emphasized, 'pitch range is used as a term for what is probably best referred to as F0 range'. A speaker's F0 range must be distinguished by a speaker's vocal range. While F0 range describes the level and span of F0 within a speech performance, vocal range is used to measure the range of F0 that 'is physically possible for a speaker to produce' (Mennen et al., 2012: 2249).

[2] Cruttenden (1997) introduced the concepts of pitch level and span by calling them 'register' and 'key', respectively. Register (also called level) 'involves a rising of the baseline, the range itself within which a speaker is operating is raised […] key (also called span) 'involves the width of the pitch range over whole intonation groups' (1997: 124).

Pitch range contributes to project onto listeners the speakers' emotions and mood. Generally speaking, at a wide pitch range corresponds a more excited and involved participation of the speaker in a conversation; by contrast, a narrow pitch range may be used to indicate lack of commitment and interest. Thus, pitch range is used in different contexts to emphasize specific parts of discourse or to convey emotional messages, such as a sense of participation, interest, detachment etc.

> 'There are both between-speaker effects, for example each speaker's natural range or tessitura, and within-speaker effects, ways in which a speaker may use pitch range for communication or unintentional effects due to physical or emotional states'.
> (Knight, 2003: 71)

Speakers using narrow pitch ranges may sound detached, cool, dispassionate, uninterested, unemotional or even unfriendly. However, it is not clear to what extent narrow pitch range plays a role in giving this impression of detachment (Mennen, 1997). Narrow pitch range has been found to be simply an indicator of low competence in L2 (Mennen, 1997) or to be associated with the phenomenon of declination (namely, F0 range usually decreases over the course of a sentence).

In fig. 3, a visual description of F0 range has been taken from Lieberman and Pierrehumbert (1984) (in Ladd and Morton, 1997: 314). The range of peaks covers the frequencies set between 160 Hz and 300 Hz. The range of lows, instead, is set at about 100 Hz.

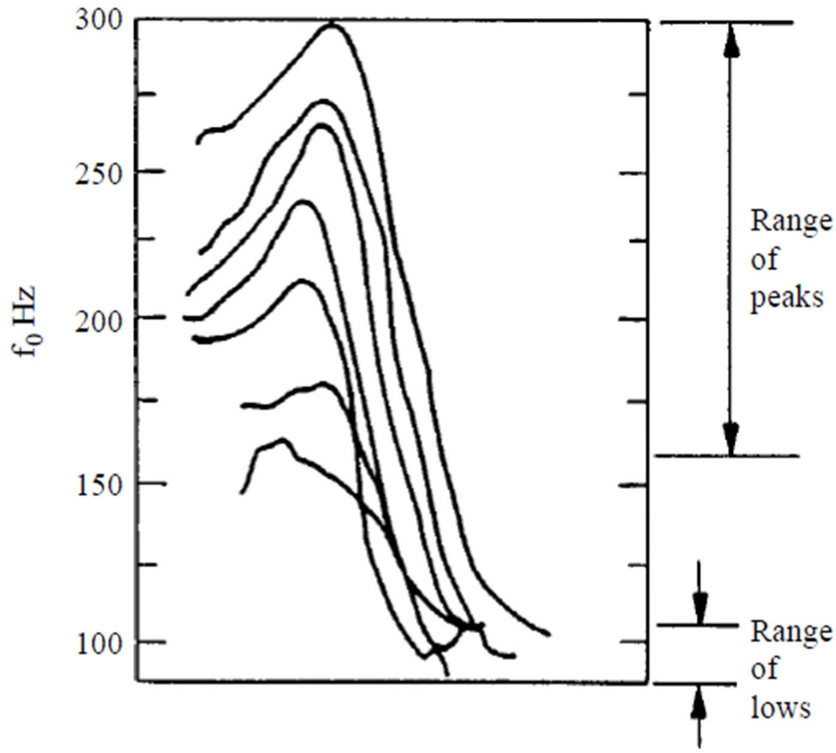


Figure 4. F0 contours of the same sentence uttered with different pitch ranges. (From Lieberman and Pierrehumbert, 1984; in Ladd and Morton, 1997: 314).

The two dimensions which describe variation in pitch range are: pitch level and pitch span. If we had to calculate the pitch level of the contour in fig. 3, we should average all F0 values and get the measure of the F0 mean. As far as pitch span is concerned, it is calculated by subtracting the F0 minimum (lowest F0 value) from the F0 maximum (highest F0 value). Generally speaking, pitch level and pitch span are correlated and co-vary to a large extent: the higher the pitch level, the higher the pitch span (Rietveld and Vermillon, 2003). Even though pitch level and span are very closely related measures and they often co-vary, it is fundamental to keep them separate and analyse them as two distinct measures. Indeed, equal F0 means can correspond to completely different spans.

To sum up, in tab. 1, level and span are compared by paralleling high and low level with wide and narrow span. On the one hand, level is measured on the overall F0 mean value of the F0 contour; on the other hand, span is measured as the difference between F0 maximum and F0 minimum. The F0 contours in the four boxes represent variants of F0 level and span: in 1) a high F0 level; in 2) a low F0 level; in 3) a wide F0 span; in 4) a narrow F0 span.

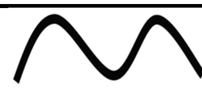
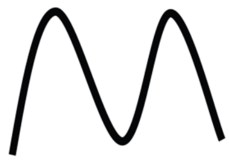
LEVEL	 1. high	 2. low
SPAN	 3. wide	 4. narrow

Table 1 F0 level and span.

Indeed, level and span are two measures of pitch range that account for independent measures of variation. While pitch level describes the overall height of a speaker's voice, pitch span describes the range of frequencies covered by that speaker (Mennen et al. 2012).

2.1.2 Pitch perception

F0 reflects changes in the vocal fold vibrations while pitch accounts for these variations as they are perceived by listeners. Pitch movements are perceptually detected and provide auditory and visual information. Nowadays, speech analysis packages allow researchers to visually inspect and automatically calculate F0 movements. Instead, musically trained listeners are capable of capturing differences in pitch range by relying just on their ears. The ability of these listeners allows them to perceive any slight change in F0. It has been found by Klatt (1972) that listeners can detect a F0 change up to 2.0 Hz. This parameter is called JND (just-noticeable difference) and defines the smallest F0 change that is feasible for a listener to be acoustically perceived. Klatt's results were based on synthetically-generated speech segments. Therefore there is no study, to date, on natural speech. It is expected that the JND in natural speech is larger than the JND measured by Klatt (1972) in his study. According to Silverman (2003), the JND of pitch in lexical tones is about 9 Hz. However, a phonological contrast between non-lexical tones requires a larger amount of Hz to be effectively perceived. For example, even though in Cantonese tones 22 and 33 have a 20-30 Hz difference of tonal contrast, these two tones are reported to be very confusable and merging (Mok and Wong, 2011). It is generally agreed upon that,

> ‘to maintain a multi-level pitch contrast is extremely hard for people because of the limitations of production and perception. On one hand, the pitch range used in normal speech is fairly small, usually less than 100 Hz (Baken and Orlikoff, 2000); on the other hand, the JND for tone is not less than 9 Hz (Bent et al., 2006), and languages usually require a much larger difference than the JND to maintain a phonological contrast’.
> (Kuang, 2012: 1)

In order to understand how our auditory system manages to perceive JND within pitch range and discriminates F0 variation across pitch movements, one has to consider the complexity of the human auditory system. The auditory system is based on six main components: the auditory nerve, the cochlea, the bone chain, the eardrum, the ear canal and the outer ear. It is the cochlea that plays a major role in pitch perception. The cochlea is an auditory organ that has the form of a spiral-shaped cavity. It contains two fluid chambers separated by the basilar membrane, a membrane that determines the mechanical wave propagation (Johnson, 2003: 46-57).

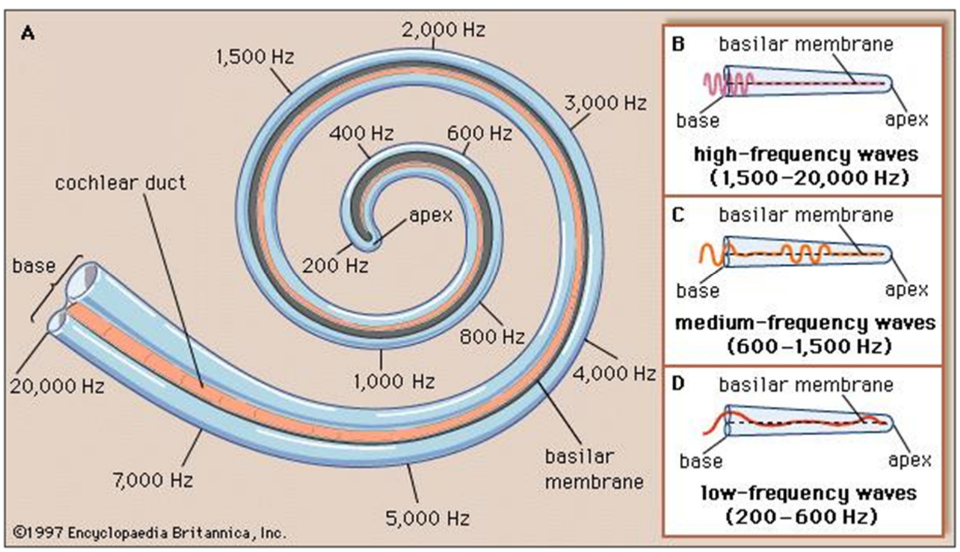


Figure 5. Analysis of sound frequencies by the basilar membrane (from http://www.britannica.com/EBchecked/media/537/The-analysis-of-sound-frequencies-by-the-basilar-membrane. Encyclopedia Britannica, 1997)

Why is the basilar membrane important in the perception of F0? Because the basilar membrane devotes more real estate to lower frequencies (60% for under 4,000 Hz) than to higher frequencies. Fig. 4 illustrates a map of frequencies that resonate preferentially at different levels of the basilar membrane, from the base to the apex. ‘The basilar membrane is thin at its base and thick at its apex’ (Johnson, 2003: 52). As a result, higher frequency waves resonate at the base level, medium frequencies stimulate the middle part and lower frequencies affect the apex (Britannica, 1997). Thus, one can infer that pitch resolution is greater in lower than higher F0. This is confirmed by Johnson (2003: 53) who claims that 'small changes in frequency below 1,000 Hz are more easily detected than are small changes in frequency above 12,000 Hz'.

The pitch of human voices is well perceived within the limits of 40 and 4,000 Hz ('t Hart, Collier and Cohen, 1990). So, humans are relatively insensitive to F0 differences at high frequencies. Pitch excursions can be expressed either by linear or non-linear scales. Pitch has been described as 'that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale from low to high' (Crystal, 1997: 294). Musical, acoustical and psychoacoustic scales have been designed in order to better capture differences in pitch scaling.

2.1.3 Linear and non-linear frequency scales

The perceptual representation of pitch is different from the acoustic description of F0. This is due to the fact that, as shown in the previous paragraph, F0 is analyzed acoustically while pitch is retained auditorily and perceptually.

> 'Acoustically, speech sounds are sound waves that are produced by vocal tracts in motion. Auditorily, they are shaped by nonlinearities of the auditory system. Perceptually, speech sounds are no longer even limited to speech audition […] and perception takes place in the context of a lifetime of experience with language' (Johnson, 2003: 59).

When investigating the functions of pitch in three different branches such as the *acoustic*, *auditory* and *perceptual* domains, it is necessary to adopt different frequency scales depending on the research purposes. An acoustic linear scale, measured in Hz, better captures acoustic differences and phenomena. For example, in a study by Rietveld and Gussenhoven (1985) on prominence-lending pitch movements, it was argued that prominence judgments given by the participants in the experiment were better in line with the results obtained by a Hz scale rather than a semitone (ST) scale. Most acoustic studies calculate F0 intervals in Hz, because Hz critically permits to compare frequencies linearly and to express ratios on measures such as F0 mean, median, maximum, minimum, span etc.

An auditory frequency scale, such as the Bark scale, is 'proportional to a scale of perceived pitch […] and to distance along the basilar membrane (Johnson, 2003: 51). Thus, the Bark scale is based on the basilar membrane as a filter bank of 24 critical bands.

In fig. 6, the acoustic frequency scale (in Hz) and the auditory frequency scale (in Bark) are compared. The graph illustrates that 'the auditory system is more sensitive [than the acoustic system] to small changes in frequency at the low end of the audible range' (Johnson, 2003: 52).

The ST scale, a non-linear logarithmic scale, is to be preferred in perceptual studies when pitch intervals are compared at different frequencies. Indeed, semitones make an appropriate normalization for the non-linearity of speech perception. The ST scale, also called musical scale, is suitable for studies comparing pitch intervals in males vs. females. 'A large change in frequency at the higher absolute pitch range of a female voice is needed to produce the same perceptual effect as a smaller change in the frequency of a male voice' (Daly and Warren, 2001: 86).

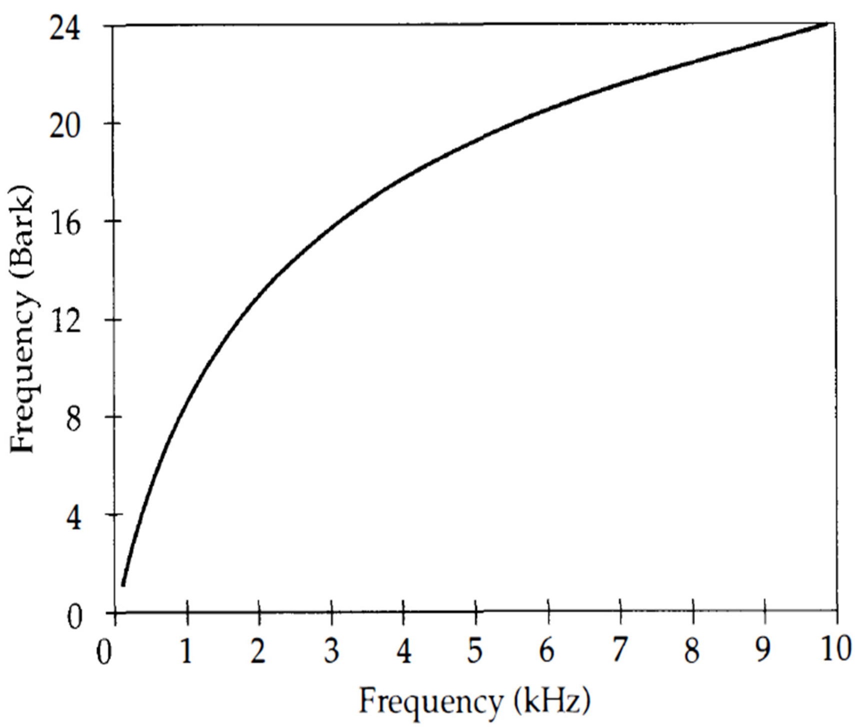


Figure 6. Comparison of an auditory frequency scale (the Bark scale) and an acoustic frequency scale (in kHz). (From Johnson, 2003: 52).

Recent perceptual analyses of pitch range have adopted another scale that is based on 'the frequency selectivity of the human auditory system' (Daly and Warren, 2001: 86; Yuasa, 2008: 79). The ERB (Equivalent Rectangular Bandwidth) scale, also called psychoacoustic scale, combines the features of linear and non-linear scales by applying the physical property of pitch to the constraints of the human auditory system. In particular, Hermes and van Gestel (1991) found that the ERB scale appears to be the most appropriate scale because it successfully matches the perceived prominence of movements within pitch intervals. Despite the fact that the accuracy of this kind of scale is extolled in many studies (Glasberg and Moore, 1990; Hermes and van Gestel, 1991; Nooteboom, 1997; Daly and Warren, 2001; Yuasa, 2008), ERB is a quite complex measure. Its accuracy is comparable to the accuracy of an acoustic scale such as Hz, in certain conditions. For instance, Mennen et al. (2012) reported that if non-parametric tests based on ranks are used for statistical analyses, the results obtained from the Hz scale will be the same as those obtained by the ERB scale.

To test the accuracy of acoustic (Hz), musical (ST) and psycholinguistic (ERB) scales, several experimental-based evaluations of pitch scales were carried out. Rietveld and Gussenhoven (1985) found some evidence of the Hz scale being superior to ST scale in expressing differences between F0 peaks. The judgments on F0 given by the subjects of the experiment were better correlated with pitch intervals when they were expressed in Hz rather than ST. These data are in contrast with results obtained from Nolan (2003) who claimed that speakers' intuitions of pitch span are better modeled and captured by a logarithmic scale where measures are calculated in ST. Hermes and van Gestel (1991) rejected the Hz and ST scales in favor of the ERB scale, again by comparing speakers' judgments on prominence level.

For the purpose of my study, I will employ the acoustic Hz scale for comparing linear measures such as F0 mean, F0 median, F0 max, F0 min etc. The reason of this choice is based on the fact that if 'ST were to be applied to level measures, an arbitrary reference point would have to be defined' (Mennen et al. 2012: 2255). Thus, ST scale is reported to be suitable especially for frequency differences (span) and not for level pitch estimation. At present, there is no consensus on which scale better captures speakers' intuitions about span. In line with the indications proposed in previous literature, I will calculate it in Hz and ST, in order to check whether or not large effect sizes are successfully captured by a specific scale.

2.1.4 Speech rate and length

Speech rate (henceforth SR) is calculated as the number of words, syllables, or phonemes per minute. To date, there is no agreement on which unit (words vs. syllables vs. phonemes) may be the best candidate to account for SR (Pellegrino et al. 2004). According to Kormos and Denes (2004), who follows Riggenbach (1991), SR is calculated by dividing 'the total number of syllables produced in a given speech sample [...] by the amount of total time required to produce the speech sample, (including pause time) expressed in seconds' (2004: 155).

Generally speaking, the standard SR for adult population falls within 140-160 words per minute (Conture and Curlee, 2007). However, SR values for young children and old people are considerably lower than those of adults, due to the fact that children and old people speak more slowly than adults. On average, fast-speaking people are reported to be less intelligible than slowly-speaking people. However, there is anecdotal evidence that people speaking with a very low speech rate may be negatively perceived as tired, lacking concentration or even depressed (Covington et al., 2005). In a study on the effects of pitch and speech rate on personal attribution, 'slow-talking speakers were judged less truthful, less fluent, and less persuasive and were seen as more "passive" (slower, colder, passive, weaker)' that fast-talking speakers (Apple et al. 1979: 715).

Within bilingualism research, SR had been used as a parameter to measure fluency in L2 speech (de Jong and Wempe, 2007; 2009). In particular, English native speakers are likely to speak quickly, as compared to L2 learners who are not so fluent in their L2. Therefore, a lowering of SR in speakers of L2 is due to difficulties related to mastering a foreign language, such as pauses to find the right words given the small vocabulary, hesitations on the pronunciation of words etc. When a L2 speaker has a dramatically low SR, speech is broken up and is difficult to listen to. By contrast, when the SR is very high, it is hard for the listener to understand his or her interlocutor. For the purpose of the present study, it is relevant to determine whether native speakers (NSs) have higher SR values than (NNSs), in order to test the effects that SR can have on the perception of pitch.

The interaction between SR and F0 level/span is crucial for contour analysis because strong correlation had been found between SR and F0 values. In a study on the speaking rate effects on discourse prosody in standard Chinese (Li and Zu, 2008), results showed that to fast speech corresponds an increase in F0 mean, to slow speech corresponds a decrease in F0 mean. Thus SR variation affects F0 level and these two dimensions covary in a directly proportional way. As for F0 span, Ladd et al. (1999) found that F0 span gets wider as SP slows down. This can be due to the fact that lower SP implies more time, and thus effort, devoted to the articulation of the utterance.

The notion of SR has always been related to the more familiar notion of length. ‘Phonetically, length refers to the physical duration of a sound, but phonologically, it refers to the relative duration of sounds and syllables when these are linguistically contrastive’ (Chun, 2002: 5). The length of phonemes, syllables or utterances is measured on the basis of time elapsed in milliseconds (ms). The relative length of chunks of speech can contribute to project an effect of fast vs. slow speech. Indeed, the dimension of duration plays a role in the analysis of pitch patterns by highlighting differences in long vs. short utterances, prominent vs. non-prominent syllables, temporal alignment of pitch accents, pitch excursion size and overall pitch range.

Just as pitch is the perceptual correlate of F0, length is the perceptual correlate of duration:

Linguistic	**Perception**	**Acoustic**
Tone	Pitch	F0
Quantity	length	duration

Table 2. Dimensions of tone and quantity within the perception and acoustic analysis (see Lehiste, 1970).

The duration of phonemes, syllables or utterances is measured on the basis of time elapsed in milliseconds (ms). The relative duration of chunks of speech can contribute to project an effect of fast vs. slow speech. Indeed, the dimension of duration plays a role in the analysis of pitch patterns by highlighting differences in long vs. short utterances, prominent vs. non-prominent syllables, temporal alignment of pitch accents, pitch excursion size and overall pitch range. In addition to SR and duration, other measures can be calculated to determine speech fluency: the number of syllables, the phonation time, and the articulation rate. The number of syllables (henceforth nsyll) is a measure that simply count the number of syllable in every utterances.

Phonation time (PhT) is defined as ‘the total time spent speaking divided by total time to produce speech sample, ST/TR’ (Wang, 2008: 21), in which ST indicates the speech time while TR indicates the total response time (Wang, 2008). Phonation time 'accounts for the percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample' (Kormos and Dénes, 2004: 151). This measure was found to be a good predictor of fluency (Towell et al., 1996; Lennon, 1990; van Gelderen, 1994; Cucchiarini et al., 2002; Kormos and Dénes, 2004).

Articulation rate (AR) is defined as 'the total syllables produced in speech sample divided by total time required to produce those syllables' (Wang, 2008: 21). Articulation rate measures the rate of speaking in which all silent pauses are excluded from the calculation and is defined as the number of production units (often phones but also syllables and words) per unit time (Crystal and House, 1990; Dankovičovà, 1999; Kormos and Dénes, 2004; DanWang, 2008).

The length of words and sentences has been interpreted differently, depending on the expression of durations at the segmental and suprasegmental levels. In fact,

> 'the function of duration at the sentence level is quite different from its function at the word level. Changes in the relative durations of linguistic units within a sentence do not change the meanings of individual words, as they may do when quantity functions at the word level. However, significant changes in tempo – changes from a neutral rate of articulation – may convey something about the mood of the speaker or about the circumstances under which the utterance is made' (Lehiste, 1996: 231).

It has been claimed that 'trends in pitch variation are believed to appear at different time-scales – such as microprosody, accent, phrase and discourse levels – making wavelet analysis of the F0 contour a suitable choice for investigating the corresponding pitch patterns' (Farahani et al., 2004: 1).

F0 patterns play a significant role in distinguishing between two different durations and in providing characteristic features for overlength (Lehiste, 1989). High pitch, a long duration and a considerable intensity make a word more prominent than the other words of a sentence. Differences in tonal alignment or slope jumps are determined by the timing of F0 rise or fall. In tonal languages, lexical tones are 'contrastive in terms of F0 height, contour and duration' (Pan, 2007: 196). For example, 'in Mandarin, with four distinctive f0 contours for each lexical tone, F0 range expansion is used as the major cue for signaling narrow focus. In Taiwanese, duration lengthening is a more consistent cue for narrow focus'. (Pan, 2007: 211). Declination and final lowering are two phenomena that are influenced by F0 range and duration, and to some extent also by SR. Declination has to do with the gradual decrease and narrowing range of F0 over the whole contour while final lowering focuses only on the last part of an utterance (Liberman and Pierrehumbert, 1984; Pierrehumbert and Beckman, 1986).

2.1.5 The effects of segmental factors on F0

As a result of their glottal and oral configurations, segments (vowels and consonants) have their own typical F0, also called intrinsic F0 (IF0). Calculations of F0 in vowels have shown a 'tendency for the high vowels, such as [u] and [i], to have a higher fundamental frequency than the low vowels, such as [a] and [æ]' (Whalen and Levitt, 1995: 349). This phenomenon is thought to be a consequence of the pulling of the tongue on the larynx, and thus it has an articulation origin (Steele, 1986).

In order to test the impact of IF0 of vowels in F0 analysis, Ladd and Silverman (1984) carried out a study on the IF0 of vowels compared in different experimental conditions (carrier sentence vs. paragraph reading). It was found that the effect of IF0 is smaller in connected speech (i.e. paragraph reading) than in carrier sentences. According to Ladd and Silverman (1984), this attenuation of the IF0 effect may be due to the influence of declination. The results obtained in this study support the idea that connected speech better than isolated sentences provides data suitable for pitch analysis. Even though IF0 is considered to have a fairly insignificant effect on the overall F0 contour, analyses on connected speech consent to exclude any interference of IF0.

Not just vowels but also consonants are reported to have an impact in the lowering or raising of F0. According to Lehiste (1970), F0 is high and falling after voiceless stops and fricatives, low and rising after voiced stops and fricatives. This has been confirmed by studies in which voiced obstruents have been shown to have a lower F0 than voiceless obstruents (Umeda, 1981; Cruttenden, 1986). It may be due to the fact that the build-up of air pressure behind a consonant constriction during voicing results in less airflow through the glottis. Less airflow results in slower vibration so that glottal tensing for voicelessness raises F0 (Lehiste and Peterson, 1961). According to Hombert (1978), aspirated or tense stops cause a raise in the F0 of the following vowel onset. Languages seem to differ in how long any consonant effect persists into the following vowel. Such consonant perturbations or micro-F0 movements can make F0 tracks look jagged, and are often ignored or smoothed away when looking at larger-scale F0 patterns. (Keating, lectures on Experimental Phonetics: Fall 2012).

2.2 Impressionistic descriptions and acoustic studies

What are the effects of pitch range in the perception of L2 speech? A number of studies have investigated the nature of F0 span and pitch level in cross-linguistic comparisons. However, only few experiments have focused on the real necessity to work on the elaboration of a generally-agreed-upon methodology. To shed light on several linguistic dimensions of pitch range, recent studies (Ladd, 1996; Patterson, 2000; Daly and Warren, 2001; Mennen et al. 2007, 2008a, 2008b, 2012; Bishop and Keating, 2010) have developed and tested different strategies for pitch analysis, in order to examine critically which measures are more suitable for the description of cross-linguistic differences in pitch patterns. A series of factors has to be taken into consideration when determining the existence of measurable and reliable differences in pitch values. This was shown by Daly and Warren (2001), who demonstrated the impact on pitch of some independent variables such as language, age, body size, speaker sex (female vs. male), speech task (read sentences vs. spontaneous dialogues), sentence type (questions vs. statements) and measure scale (Hertz, semitones, ERB etc.).

The first works on pitch range were conducted in the early twentieth century in the United States. These studies yielded consistent results, based especially on impressionistic considerations of the quality of male vs. female voices. Along the years, acoustic and phonetic investigations aimed at providing a more thorough examination of those phenomena which had largely escaped detailed description. In particular, different approaches were used to determine and quantify the factors that may affect pitch range and cause inter-speaker variability. Pitch variation is the result of a combination of linguistic and paralinguistic factors. Indeed, there is an interaction among F0 and other variables, such as speakers' personal characteristics; types of materials; socio-cultural aspects; emotional context; speaking styles. Results obtained from experiments on pitch range should be comparable in terms of influence of linguistic and paralinguistic variables on the data. Factors influencing F0 variation relate to: speakers' personal characteristics (i.e. age, sex, body size, native language, language impairments); types of materials (i.e. speech task, read passages vs. elicited sentences vs. spontaneous dialogues); socio-cultural aspects (i.e. in the Japanese society, the use of a high-pitched voice by women is, to some extent, due to persistent expectations of femininity, related to the socio-cultural roles of women); emotional contexts (i.e. pitch modulation is used to express emotion and to connote the presence of emotional involvement); speaking styles (i.e. high-pitched modulation is generally associated to an higher degree of politeness).

In order to understand how F0 varies across languages and speakers, it is important to tell exactly to what extent F0 change across a population. F0 is affected by a series of physical and sociolinguistic factors such as gender, age, body size, etc. Thus, it is crucial to control for all these variables when comparing F0 mean. In the following paragraphs, I will review previous findings on the effects on F0 variation given by: gender; age; body size; health history; socio-cultural factors; speech task.

2.2.1 The effects of gender

The gender of subjects is a fundamental factor that yields significant main effects in pitch analysis and perception. Indeed, F0 has been identified as one of the 'most obvious and heavily studied candidates for conveying speaker sex information' (Hillenbrand and Clark, 2009: 1150).

From past to recent research, female subjects have been reported to have significant higher F0 level and span than males. Anecdotic evidence shows that, on a scale, F0 values are considerably low for males, quite higher for females and very high for children. Cruttenden (1997: 3) claimed that, 'the average fundamental frequency for men is approximately 120 Hz, for women 225 Hz, and for children 265 Hz'. In an earlier study, F0 mean had been found to be 131 Hz for men and 220 Hz for women (Hillenbrand et al., 1995). For Peterson and Barney (1952), the F0 mean of men was 132 Hz while for women it was 224 Hz. Therefore, even though F0 means in men, women and children cannot be considered as absolute values, it is clear that the results obtained in the three reported studies are almost identical. Males speak not only with a lower F0, but also with less pitch variation than females. Studies on gender stereotypes (Loveday, 1981; Henton, 1989, 1995) characterized female speech as 'high-pitched, shrill, over-emotional, and swoopy' (Daly and Warren, 2001: 85).

In one of the first studies on pitch range, Weaver (1924) investigated F0 mean values of 43 university students at the University of Wisconsin. His results showed that, for the population he studied, the F0 mean values were: 318 Hz for females and 151 Hz for males. Only a few years later, Cowan (1936) recorded ten professional actors and actresses in New York and he found out that the F0 median value for females was 233 Hz (ranging from 199 Hz to 295 Hz); for males it was 141 Hz (ranging from 134 Hz to 146 Hz). In a later study conducted by Fitch and Holbrook (1970), a large-scale experiment carried out on 200 students of Florida State University, it was shown that the F0 mean of the female students was 217 Hz and that of the males was 116 Hz. It is clear from these results that there is no overlap between F0 mean of males and females. However, values sensibly changed from study to study (Henton and Bladon, 1985; Klatt and Klatt, 1990; Traunmüller and Eriksson, 1995; Hanson and Chuang, 1999; Daly and Warren, 2001; Honorof and Whalen, 2005; Dilley and Brown, 2007; Lee, 2009; Shue, 2010). This is due to the fact that researchers used different methods, materials and acoustic instruments.

In order to prevent gender factors from affecting the results, researchers started to focus their interest either on male or female pitch range, by recruiting only male participants in their experiments (Provonost, 1942; Snidercor, 1943; Mysak, 1959; McGlone and Hollie, 1963; Terango, 1966; Hollien and Shipp, 1972; Hollien and Jackson, 1973); or female subjects (Snidercor, 1951; Saxman and Burke, 1967; Linke, 1973; Stoicheff, 1981; Kasuja; 1996; Mennen, 2008b, 2012; Busà and Urbani, 2011). The advantage of working with either male or female subjects is that the gender variable can be excluded from the analysis of the results. This permits to give more emphasis to other factors that may affect pitch, such as language, age, size, state of health, regional accent.

2.2.2 The effects of age

Apart from gender, one of the biggest issues affecting pitch range is age. Longitudinal studies were carried out to provide some evidence that physiological changes occurring during the aging process considerably affect F0 values. In order to test this hypothesis, Stoicheff (1981) recruited female subjects, who, at the time, were from 20 to 82 years old. Results showed that while females in their 20s had a F0 mean of 224.3 Hz (ranging from 192.2 Hz to 275.4 Hz) females in their 30s had a F0 mean of 213.3 Hz (ranging from 181 Hz to 240.6 Hz). A decrease in the F0 values was observed in females in their 50s due to changes in the vocal fold mass. An increase in intra-subject variability of fundamental frequency was observed in subjects older than 50-60 years old. This was interpreted as 'indicating decreased laryngeal control over fundamental frequency in postmenopausal adults' (Stoicheff, 1981: 437).

Comparable results were obtained by de Pinto and Hollien (1982) who recorded their subjects twice (at a distance of 35 years) and obtained different results for the two recordings. They recorded 11 Australian women at the age of 18-25 years and then they recorded the same subjects 35 years later. The results of this study showed that the same subjects had a F0 mean of 224 Hz when they were recorded in their 20s and 180 Hz in their 50s. A similar study was conducted on male speakers grouped by age (see Hollien and Ship, 1972). In this experiment, F0 mean varied across subjects depending on their age: 119 Hz for males in their 20s, 107.1 Hz for males in their 40s, 146.3 Hz for males in their 80s.

Fig. 7 illustrates the way in which F0 changes across male and female life span:

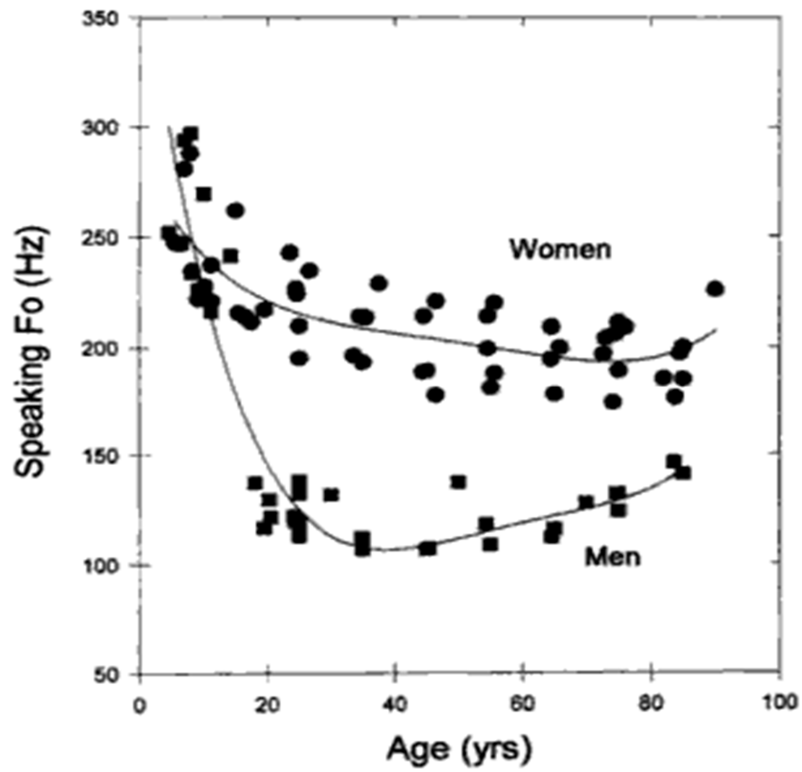


Figure 7. General trend of F0 across male and female life span.
(From Baken and Orlikoff, 2000: 174).

Changes in F0, occurring over the years, are expected to be broadly predictable (see Baken and Orlikoff, 2000: 145-296). Indeed, changes in a speaker's F0 have been found to be systematic and they are due mostly to 'maturational development of the vocal apparatus as well as the development of cognitive and linguistic skills' (Cooper and Sorensen, 1981: 164). In early childhood the F0 of males and females is comparable (consider the clustering around 240 Hz in fig. 7). Infants produce cries containing extremely high F0 values and their vocalizations are based on an increasing number of patterns, as children step forward in their language acquisition process. The average F0 mean for children between the age of 5 and the onset of puberty is calculated at about 270 Hz (Cooper and Sorensen, 1981; Cruttenden, 1997). F0 values decrease over the years with the most considerable drop accompanying puberty. Menopause is a contributing factor for the lowering of F0 in women; physical modifications of the larynx are the main cause for the lowering of F0 in men. It has been found that F0 mean continues to decrease till age 40-50; from 60 to 80 it slightly raises due to changes in the musculature of the vocal tract (Cooper and Sorensen, 1981).

2.2.3 The effects of body size and health history

Some studies have provided evidence that F0 highly correlates with body size, elastic properties of vocal tract tissue and folds length (Yuasa, 2008: 6). However, Graddol and Swann (1989) found no statistically significant correlation between body size and F0 excursion. The height or weight of a speaker has from slight to no influence on the F0 of his or her voice. On the contrary, it is generally agreed upon that vocal fold length significantly influences gender differences in average F0.

The vocal folds of men were reported to be significantly longer than those of women (Ellis, 1929; Kahane, 1978; Hirano, 1983). In particular, they were found to be approximately 60 percent longer than those of males (Titze, 1989: 1699). These results showed that F0 is inversely related to the length of vocal folds; that is, the longer the vocal folds, the lower the F0 (and vice versa). Moreover, Hollien (1960) demonstrated that also the thickness of the vocal folds correlates with F0 excursions. In fact, the thicker the vocal folds, the lower the F0. These findings have been disproved by Titze (1989: 1706) who claimed that vocal folds length accounts almost entirely for gender differences in F0 mean. As the length of the vocal folds increases, their thickness decreases due to the fact that the vocal folds become thinner and thinner. However, a recent study by Shutter et al. (1996) confirmed the thesis of Hollien (1960). 'With an α level of $p = 0.005$, men differed from women with respect to laryngeal appearance […] and their vocal folds were rated thicker in the vertical dimension, smaller in the lateral dimension, longer, and more tense, with smaller amplitudes of excursion during vibration' (Shutter et al., 1996: 175).

Factors such as particular speaking styles, socio-phonetic situations, emotions or speech pathologies have a great influence on F0. An important issue is smoking history and clinical problems, such as depression or psychological disturbs. A study on female smokers and non-smokers (Gilbert and Weismer, 1974) showed that smokers have a significantly lower mean F0 (163.7 Hz) than non-smokers (182.8 Hz). In addition, cries of babies affected by Sudden Infants Death Syndrome (SIDS) were found to have distinctive patterns as compared to healthy children (Golub, 1979). Thus, the F0 analysis of infant cries and children speech patterns was used to carry out diagnoses and cure several impairments, such as, among others, aphasia (Danly et al., 1979; Cooper and Zurif, 1981), hearing-impairment (Nickerson, 1975) and autism (Paccia-Cooper and Curcio, 1980).

Abnormal F0 values are not mere an indication of physical pathologies but also indicators of more global disorders. Depression, investigated in a study by Nilsonne and his colleagues (Nilsonne et al. 1988), has been found to have a great effect on pitch range. Schizophrenic adult females have been found to have higher F0 mean level and larger F0 variation than other healthy women, used as a control group (Saxman and Burk, 1968). Even in the absence of specific pathologies, speakers with no impairment are capable, to different extents, of producing a wide range of pitch excursions, depending on their mood, motivation and effort (Texeira et al. 2008). Personality and emotion play a role in the production of F0 patterns. So, listeners are capable of determining the particular emotional states and personalities of their interlocutors, just by grasping minimal pitch variation in the pitch patterns.

2.2.4 The effects of socio-cultural factors

Pitch variation may be influenced by socio-cultural contexts. Within a linguistic community, it has been argued that 'individuals may shape the pitch range of their speech to fit the prevailing pitch range in a given linguistic community' (Dolson, 1994: 323). What is more, socio-cultural aspects and environmental factors play a role in determining specific correlations in parameters affecting pitch range. In their survey on gender communication, Kroløkke and Søresen (2005) reviewed previous studies on gender and they pointed out those habits that seem to be influenced by socio-cultural factors:

> 'Lee et al. (1995) have documented how young girls tend to have more extremes of frequency and a greater intonation variety than boys. Such variations can also be found to be national/regional. For instance, American women have been reported to use a larger vocal range than British women, whereas American men have been reported to employ a more limited vocal range than British men (Graddol & Swann, 1989)'.
> (Kroløkke and Søresen, 2005: 91)

Despite the differences between British and American English, the United Kingdom and the United States share, to some extent, a common cultural heritage, social expectations and lifestyle. By contrast, Japan and the United Sates are considered to 'exemplify completely opposite sets of cultural norms and consequently offer excellent sample pools by which to compare and contrast abiding social variables' (Yuasa, 2008: 8). If the American society stands for values such as individualism, leadership and high competition, the Japanese society is very attached to tradition, honor and respectability. In particular, women are asked to respect a code of behavior that prevents them from being perceived as independent, determined or aggressive as some American women are.

Studies on pitch range have shown than the F0 mean of Japanese females is consistently higher than those of British women (Loveday, 1981), American women (Ohara, 1992; Kasuya, 1996) and Dutch women (van Bezooijen, 1995). In addition, Loveday (1981) showed that Japanese males used considerably lower pitch values as compared to British males, while Japanese females had fairly higher F0 values than British women. These results were supported by the idea that men want to project an idea of unemotional and self-restrained appearance, while women present their attitude as innocent, feminine and helpless. These considerations were reiterated in a study on F0 values of Japanese and Dutch females, in which it was claimed that 'Japanese women would raise their pitch to conform to gender stereotypes and project the desired feminine attributes of powerlessness' (van Bezooijen, 1995: 254). The social role of Japanese women leads them to use, probably subconsciously, a pitch range that reflects this image. This hypothesis was further tested and confirmed by the results of a perception study by van Bezooijen (1995). The results of her perceptual study showed that Japanese listeners (males and females) perceived high-pitch voice in female as being more attractive than Dutch listeners did.

If pitch differences across the Western (i.e. British, American, and Dutch) and the Japanese cultures are justified by socio-cultural reasons, how can we explain pitch range differences across European languages? In the Mediterranean cultures, languages such as Italian, Spanish, French and Portuguese share a common linguistic heritage (i.e. Latin) and have similar cultural background. The social representation of men and women is quite similar among these societies. Nonetheless, pitch values significantly differ across genders and languages. This is true also for Germanic languages such as Dutch, English and German which have similar prosodic systems but quite different pitch ranges (Gibbon, 1998).

For example, in a study on British and German females, it was found that British females are often perceived as over-excited or aggressive while German females may sound bored or even unfriendly (Mennen et al., 2012). This is due to the fact that in English there is more pitch variation than in German and overall F0 level is reported to be higher in English than in German (Grabe, 1998; Mennen et al., 2012). Not only females, but also males are sensitive to socio-cultural effects that influence their pitch range. For example, Majewski et al. (1972) reported that the F0 mean of Polish males is significantly higher than the F0 mean of American males. No correlation with body size, vocal folds length or other physical factors. Thus, the authors of this study hypothesized that the main effect of difference across languages was socio-cultural.

2.2.5 The effects of speech task

As for the data set used in the experiments, different F0 values have been obtained depending on the materials selected and on the speech tasks created for the experiments. In particular, important variables such as conversational dialogues vs. monologues, spontaneous vs. elicited vs. read speech, formal vs. informal register etc. appear to influence pitch level and span. The relevance of the differences reported between read vs. spontaneous speech has dramatically increased in recent years, due to the growing interest in automatic speech recognition and models for synthetic speech (see Klatt and Klatt, 1990). According to Silverman et al. (1992a), marked prosodic differences are reported in read versus spontaneous speech'. In fact,

> 'Read speech differs from spontaneous speech in some important ways: (i) although the tuned on focused words are selected from the same inventory in both read and spontaneous speech, the prior probabilities of the tunes differ greatly – spontaneous speech predominantly contains rises, while read speech predominantly contains falls, (ii) pauses in read speech are shorter than in spontaneous speech, and they predominantly are located at structurally predictable positions (grammatical boundaries), whereas in spontaneous speech this generalization hardly holds true at all, (iii) read speech tends to not contained filled pauses'.
> (Silverman et al., 1992: 1302)

Even though informal speech is reported to display higher F0 variability than formal speech, it is not easy to obtain comparable data from spontaneous speech, extracted from real everyday talking. For this reason, for experimental research, it is common practice to use read materials. The advantage of read materials is twofold: recordings of read data permit comparability and repeatability. Read materials are easily created and controlled by the experimenter in relation to the purpose of the experiment. What is more, the same sentence can be recorded many times, in isolation, inserted into a passage or in different contextual positions.

In one of the first studies on spontaneous vs. read speech (Snidercor, 1943), subjects were asked to speak about a topic they were familiar with (i.e., ‘my future job’). Then, they were required to read out a transcript of their spontaneous speech in the same experimental conditions. By comparing the spontaneous speech to the typewritten texts of the original speeches, it was clear that F0 values co-varied with the spoken or read condition. Incidentally, values for F0 mean were 120 Hz for spontaneous speech and 132 Hz for read transcripts. Also, F0 median was higher for spontaneous than read materials. These findings were confirmed also by following studies which showed that higher pitch level and larger pitch span are displayed in spontaneous speech, as compared to read texts (Daly and Warren, 2001; Zipp and Dellwo, 2012). Differences in F0 range are reported even within read materials, between isolated and contextualized sentences (Garrido et al., 1993).

Zipp and Dellwo (2011) compared sentences elicited with a map task in two different experimental conditions. First, subjects were recorded while speaking to a friend and unaware of the fact that they were being recorded (informal setting); then, subjects were recorded while speaking to the experimenter who was wearing a formal work outfit and maintained a formal interaction style by projecting social distance (see Zipp and Dellwo, 2011). The results showed that ‘the variability of intonation is higher in informal than formal speech compared to read speech’ (Zipp and Dellwo, 2011: 2331).

However, there are some studies which came to opposite conclusions, diverging from this trend. Baken and Orlikoff (2000) found out that mean F0 is slightly higher in read than in spontaneous speech; Torgerson (2005) found no differences between read or spoken data in read and spontaneous Mandarin speech.

2.3 Pitch range and biological codes

Principles based on physiological and emotional conditions have been envisaged by scholars in order to account for similar pitch patterns used across languages. Pitch range variation may signal a number of meanings, some of them are biologically encoded. To this conclusion have arrived a number of scholars who have formulated several theories on the correlation between pitch range and biology. High pitch, associated with an extensive tension of the vocal folds, and a fast speaking rate reflects a high level of excitement and involvement of the speaker. By contrast, lower pitch characterizes a certain detachment and lack of emotion. Thus, pitch range is believed to correlate with the expression of emotion and involvement (Fonagy, 1981; 1983; Bolinger, 1989; Hirschberg, 1992).

One of the most influential principles has been theorized by Ohala (1983, 1984, 1996) who described the cross-language use of F0 from a, so-called, ethological perspective. He was the first to point out that certain variations of pitch range are determined by physiological and emotional causes; what is more, these changes occur depending on innate and universal principals that have been shown to be valid for a number of languages and even non-human beings (such as gorillas). Ohala (1983: 13) claimed that what he called ‘frequency code’ is ‘an inherent part of the human vocal communication system’. In fact, sound-meaning correlations have shown cross-language consistency that ‘can be explained by reference to the factors which have influenced the shape of the acoustic component of agonistic displays in virtually all vocalizing species’ (Ohala, 1983: 14). For example, high-pitch is associated with the primary meaning of small vocalizer while low-pitch is associated with the primary meaning of large vocalizer. This was attested widely by Morton (1977) who studied the vocalizations of 28 avian and 28 mammalian species and documented significant similarities in the acoustic properties of sounds produced in competitive situations.

The frequency code is a universal principle that governs sound-meaning correlations. It is based on an assumption, empirically tested, by which ‘high F0 signifies (broadly) smallness, non-threatening attitude, desirous of the goodwill of the receiver, etc., and low F0 conveys largeness, threat, self-confidence and self-sufficiency’ (Ohala, 1996: 343). This way, the ‘frequency code’ underlies the sound-symbolic use of pitch. Body size and physiological conditions have an influence on F0 habitual range (as already mentioned in § 2.2.2). To some extent, body size is both the cause and the effect of F0 variation. Namely, not only F0 is affected by body size but it also plays a role in conveying an impression on the listener about the size and attitude of the speaker (Morton, 1977; Ohala, 1996). For example,

> 'to give the impression of being large and dangerous, then, an antagonist should produce a vocalization as rough and as low in F0 as possible. On the other hand, to seem small and non-threatening a vocalization which is tone-like and high in F0 is called for' (Ohala, 1996: 331).

This confirms the idea that humans are naturally inclined to associate determined pitch values to body size, speakers' emotions and intents. F0 is systematically used to signal emotions, attitude and intent; it is also accompanied and enhanced by the use of gestures. For instance, kinesic signals, namely body language, emphasize the message transmitted by prosodic features, such as pitch and intonation. Some speakers have been found to occasionally parallel the rise and fall of F0 with the raise and lower of eyebrows (Ohala, 1996).

The concepts outlined in Ohala's works (1983, 1984, 1996) were implemented and reinterpreted by Gussenhoven, who elaborated a theory on the universal meaning of pitch range, based on three biological codes: the frequency code, the effort code, and the production code. In his survey on the interaction between biological conditions and pitch, he came to the conclusion that 'the exploitation of the biological codes is a controlled use of pitch variation' (Gussenhoven, 2004: 94). In fact, many languages are thought to encode meanings by means of pitch variation so that these meanings are believed to be universal. For example, 'speakers with high-pitched voices were judged less truthful, less emphatic, less potent (smaller, thinner, faster), and more nervous' than speakers with low-pitched voices (Apple et al. 1979: 715). This implies that by exploiting pitch excursions, speakers signal several meanings (e.g. emphasis, surprise, fear, etc.) that are, to some extent, universally understood across languages. However, the impact of the speaker's message is modulated also on the some physiological conditions that cannot be modified (e.g. body mass, vocal tract size, vocal fold texture). Thus, 'biological are form-meaning relations which are based on effects of physiological properties of the production process on the signal' (Gussenhoven, 2004: 80). In particular,

> ‘the implementation of form-meaning relations embodied in the biological codes is not restricted to speech production. In speech perception, this can be reflected in that speakers are capable of interpreting pitch variations in others’ speech (in known or unknown languages) in accordance with the biological codes’.
> (Chen, 2004: 36).

In the following paragraph, the concepts of the frequency code, the effort code, and the production code and will be explained and discussed in order to unravel the role of pitch variation in the linguistic interpretations of biological codes.

2.3.1 The frequency code

The frequency code had been defined as a ‘size code’ because it the universal meaning at issue is based on the fact that the men larynx is almost twice the size of the women larynx (Hollien, 1960; Titze, 1989; Shutter et al., 1996; Gussenhoven, 2004). Not just the larynx but also the vocal folds have been reported to be significantly longer in males than thin females (Ellis, 1929; Kahane, 1978; Hirano, 1983). It is generally agreed upon that larger larynxes produce lower notes than smaller ones (Chen et al., 2004). Thus, larynx size and vocal fold length significantly influence differences in average F0.

Three different interpretations have been associated and influenced by the frequency code: an affective property, an informational value and a linguistic implication. The affective property of the Frequency code has to do with the way in which speakers control their use of pitch to express to states or attitudes. Even though it is physically impossible to modify the size of the vocal tract, speakers can manipulate their pitch in order to sound more or less attractive, confident, friendly, etc.

> ‘The frequency code transmits the idea that low pitch sounds protective and dominant while high pitch sounds vulnerable and submissive. Thus, ‘high, and particularly high-ending utterances sound dependent, uncertain, appealing etc.; low and low-ending utterances sound powerful, assertive and authoritative’ (Chen et al. 2004).

Also pitch variation plays a role in perception. Vaissière (2005: 241) asserted that ‘small pitch variation is found to be associated with *disgust, anger, fear, boredom*, and large pitch variation with *happiness, pleasantness, activity, surprise* (Bolinger, 1989; Hirschberg, 1992) and *benevolence* (Brown et al. 1973)’. What is more, the pitch range used to convey meaning about levels of politeness (Ito, 2002; Ofuka et al., 2000) and levels of certainty-uncertainty (Hirschberg and Ward, 1992; Ward and Hirschberg, 1988; Ramirez Verdugo, 2005).

Raised F0 is heard as more polite than lowered F0. What is more, ‘a falling contour conveys certainty or dominance with regard to our knowledge of the polarity of the message. A rising tone, on the contrary, denotes uncertainty, deference, vulnerability or polarity unknown’ (Ramirez Verdugo, 2005: 2089). The information value conveyed by the frequency code is that higher pitch is associated to the expression of uncertainty while low pitch is associated to the idea of uncertainty (Gussenhoven, 2004). These kinds of judgments can be linguistically interpreted in the light of sentence type and communicative functions. More specifically, ‘the falling tone realizes a lexico-grammatical category declarative which in turn realizes a semantic category of *statement*, while the rising tone realizes a lexico-grammatical category of interrogative which in turn realizes a semantic category of *question*.’ (Halliday and Greaves 2008: 63). In table 8, the logical connections between tone type (rise vs. fall), lexico-grammatical function (interrogativity vs. assertivity), and semantic categories (question vs. statement) are graphically represented. While uncertainty is conveyed by questions, interrogatives and rising tones; certainty is conveyed by statements, declaratives and falling tones.

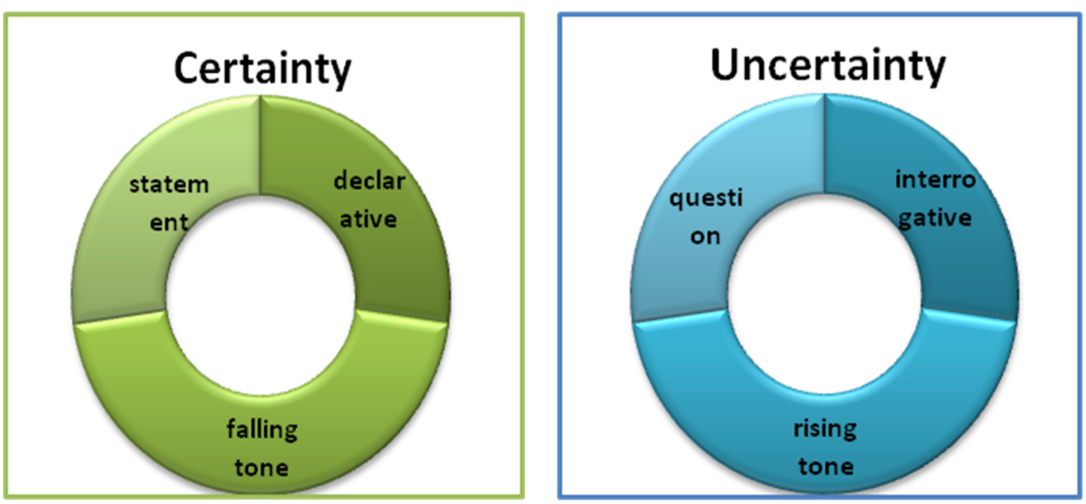


Table 3. Uncertainty and certainty model, based on Halliday and Greaves (2008: 62).

The linguistic implication of the frequency code had already been explored by Ohala (1983) who used the rising intonation (questions) and falling intonation (statements) to explain the similarities across languages in the expression of universal meanings. In a study on the perception of question vs. statement status of some sentences it was found that high peaks and high end pitch attracted judgements of question-status while low peaks and low end pitch are more likely to lead to judgements of statement-status (Hadding-Koch and Studdert-Kennedy, 1964). This proves that 'the variables peak height and end pitch are rather salient ways of manipulating the frequency code' (Gussenhoven, 2004: 82). In addition, it has been claimed that L1 may play a role in determining the contrast question vs. statement (Grice, 1995; D'Imperio, 1997). Even though over 70% of world languages are estimated to make use of rising contours to mark interrogativity (Bolinger, 1972), little is known about language-specific cues to produce questions. For example, English and many other languages have final rises as a cue to signal questions while Swedish does not. This has been thought to be the reason why Swedish listeners are more sensitive than Americans listeners in detecting final rises in both questions and statements (Hadding-Koch and Studdert-Kennedy, 1964).

2.3.2 The effort code

Depending on the energy expended to convey a certain articulatory precision, the effort code determines variation in meaning derived from the expenditure of different degree of effort. Greater articulatory precision is a consequence of an increased articulatory effort (de Jong, 1995). In particular, great articulatory precision had to do with less slurring together of pitch movements and widening of pitch excursions (Chen, 2004).

High pitch level and wide pitch span are believed to be cues for prominence (Ladd, 1996; Ladd and Morton, 1997). It is easy to find evidence for this assumption in news report. For example, in a study on the British English radio news bulletins, it was found that the overall pitch range of utterances was higher and wider to signal informational salience and prominence. In this case, it was obvious that radio speakers, concerned about their messages coming across, increased their effort level by raising the pitch level and widening the pitch span (Wichmann, 2000). Several scholars (Ladd, 1996; Ladd and Morton, 1997; Patterson, 2000; Chen, 2004; Gussenhoven, 2004; Mennen; 2012) agreed upon the idea that the perception of prominence has to do with the interpretation of pitch span, not just with the pitch level (see § 2.2.1 and § 2.2.2). Thus the effort code has an influence on pitch excursions and variation.

According to Gussenhoven (2004), affective interpretations of a speaker effort are perceived as authoritative, insistent and enthusiastic when pitch is realized with wide excursions and span. By contrast, a rather flat pitch span is somehow related to a lack in commitment or interest. Thus, emphatic and relevant information is likely to be reported with a wide pitch span. This implies that a greater articulatory effort is required when communicating significant information. The effort code proposes a correlation between values of pitch range (level and span) and expenditure of effort.

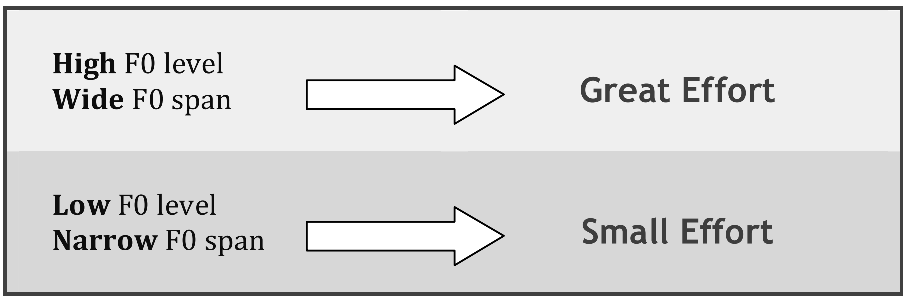


Table 4. Graphic representation of the frequency code that describes the directly proportional relation between degree of effort and F0 level and span.

As shown in tab.4, to a greater effort correspond a raised F0 level and a widened F0 span. By contrast, low F0 level and narrow F0 span require a smaller effort. No specific physiological conditions are needed to create an effect of great or small effort. To signal a higher degree of commitment and interest, speakers do not put more effort on their speech production, they just select a higher F0 level and wider span from their inventories of pitch range (Chen, 2004). In addition to pitch range, the effort code has an impact also in the realization of broad vs. narrow focus. In a broad-focus sentence, the focus constituent corresponds to the whole sentence; in a narrow-focus sentence, the focus constituent corresponds to a specific part of the sentence (Ladd, 1980). In line with the predictions of the effort code, broad focus is likely to be realized with a smaller pitch span, as compared to narrow focus that requires a sharp rise in span on the constituent in focus position (Gussenhoven, 2004).

2.3.3 The production code

The third biological principle, the production code, is based on the physiological phenomenon of diminishing of energy in exhalation of air produced through phonation. Namely, subglottal air pressure is higher at the beginning of a sentence and lower towards its end (Gussenhoven, 2004). Assuming there is a correlation between utterances and exhalation phases, the production code associates high pitch with the utterance beginnings and low pitch with utterance ends' (Chen, 2004; Gussenhoven, 2004).

As shown in tab.5, as a global tendency, the beginnings of utterances are produced with a high pitch while the endings of utterances are produced with a low pitch. However, in certain contexts, speakers use different pitch variation in order to signal the intended meaning. When speakers are introducing a new topic in the conversation, they are likely to use a high-pitch utterance beginning; by contrast, when they are continuing a topic already introduced, they will start a new sentence with a low-pitch utterance beginning. A reverse relation holds in the case of utterance endings. High utterance endings signal that speakers intend to continue their argumentations (e.g. high utterance endings are used in enumerations of items, listing of propositions etc.). On the contrary, low utterance endings evoke a meaning of finality and they are specifically used to mark end of a speaker's turn.

Global tendency	Local tendencies	
High pitch – utterance **beginnings**	High beginnings →	new topics
	Low beginnings →	continued topic
Low pitch – utterance **endings**	High endings →	continuation
	Low endings →	finality, end of turn

Table 5. Graphic representation of the production code that describes the directly proportional relation between degree of effort and F0 level and span.

The production code mainly involves pitch variation at the utterance beginning or endings. However, it is also responsible for the linguistic phenomenon of declination. Declination, that is the gradual lowering of F0 throughout a sentence (Lieberman, 1967; Ladd, 1996), is produced by a fall-off in energy and a gradual drop in F0 and intensity. Variation in the declination effect has been reported to be affected by the production code (Gussenhoven, 2004).

2.4 New methods in the analysis of pitch range

There is a need for large-scale comparable data investigating the extent to which pitch varies across gender and languages. At a glance, many researchers have long attempted to find consistent correlations between pitch level, pitch span and pitch variability. However, there is a considerable gap in the acoustic analyses of pitch range. This gap is given by the lack of a standard model able to capture pitch differences effectively. The goal of Lieberman and Pierrehumbert was to create 'a coherent model of F0 realization that (a) interprets the categories of intonational description as we understand them, (b) incorporates the qualitative characteristics of F0 measurements, and (c) shows good quantitative agreement as well (1984: 177). To my mind, a successful model should meet also the following criteria: (1) cohesion and coherence of each component within the whole structure, (2) reliability of measures used, (3) capability of being learned in a relative easy way in order to be applied to different studies, (4) versatility of application to different languages, and (5) possibility for implementation of additional resources and parameters.

Despite the number of studies and more or less efficient methodologies used in later years, there is a dearth of research on comparable analysis of pitch range. As Dolson (1994) highlighted, just a restricted number of studies 'have been similar enough in design and instrumentation to allow direct comparison'. In order to overcome this problem, new methods have been proposed in recent years.

2.4.1 Ladd (1985-2003)

Ladd and his colleagues elaborated in their investigations a well-agreed and integrated model for pitch analysis (Ladd et al., 1985; Ladd, 1996; Shriberg et al., 1996; Ladd and Schepman, 2003). The first thing to point out is that Ladd (1996) adopted two kinds of approach to pitch analysis: the initializing and the normalizing approach.

The initializing approach has to do with the somehow relational or syntagmatic feature of pitch, which can be interpreted only in relation to other parts of the utterance. For its inherent nature, pitch has to do with a movement or a change in the intonation contour. This kind of approach can be used to describe local modifications of pitch in relation to what is immediately preceding (i.e. the starting point of an utterance). Ladd stated that 'the only thing such a model requires in order to derive actual F0 values is an initial state for each utterance. It does not even need to refer to characteristics of a speaker's range; all it needs is a starting point' (Ladd, 1996: 253). Thus, the initializing approach focuses on the assumption that a low rising movement can be distinguished from a high-rise movement with respect to the starting point, the initial F0 value.

The normalizing approach is based on ‘speaker-specific reference points, such as upper and lower F0 values’ (Ladd, 1996: 256). Ladd elaborated a sort of phonology of pitch by giving a quantitative definition of pitch scaling. He did so in a model based on three targets: H (for high), M (for mid) and L (for low) pitch. The idea is that these labels (based to some extent on the Autosegmental-Metrical system) can describe pitch movements in any language (see also the recent detailed analysis of tonal autosegments and features of pitch in Hayes, 2009: 291-312). Indeed, by combining H, M, and L targets, it is possible to describe pitch movements without making any reference to F0 values. This is true only at an abstract level; at a practical level, a speaker’s overall speaking range is determinant in the identification of targets. For instance, it has been argued that

> ‘the actual F0 values corresponding to H tone and M tone will depend on whether they are spoken by a man or a woman, by a person with a monotonous voice or a person with a lively voice; that is, the acoustic realization of the pitch scale depends crucially on the speaker and the paralinguistic context’ (Ladd, 1996: 270).

In order to create a normalizing model of the phonetics of pitch, Ladd elaborated a model capable of capturing quantitative properties of any speaker-specific scale. In line with Lieberman and Pierrehumbert (1984), Ladd observed that the utterance-final low of the speaking range could be considered as a reference frequency (Fr). This is based on the premise that, within a F0 contour, while peak dramatically raises and falls creating some valleys, low values remain nearly constant (Lieberman and Pierrehumbert, 1984). In addition, Ladd claimed that ‘the bottom of the speaking range is a fairly constant feature of an individual’s voice’ (1996: 267). F0 of a specific point within the utterance is calculated by this formula:

(1) $F0 = Fr \cdot T \cdot r$

where *Fr* is the zero level or reference frequency, *T* is an invariant abstract pitch value and *r* is a range multiplier whose value is 1 for normal range (for more mathematical insight on the application of this formula, see Ladd, 1996: 267). Despite the adequacy of the model described above, Ladd himself admitted that this model ‘does not work, in the very basic sense that it fails to make accurate predictions about the quantitative regularities that have been observed in range-modifications and range-comparisons' (Ladd, 1996: 269).

Ladd studied in detail also other phenomena related to pitch range variation within and across speakers, such as the 'segmental anchoring of F0' and the prediction of F0 targets when 'speaking up'. Segmental anchoring of F0 has to do with tonal alignment and, in particular, with the temporal coordination between F0 and phonetic segments. Local F0 maximum and minimum are claimed to be aligned in predictable ways along the segmental string (Atterer and Ladd, 2004). Thus, F0 rises and falls are aligned within specific landmarks within the segmental strings called 'anchor points' (Arvaniti et al., 1998). Since the association between tonal and segmental elements have an effect on the alignment of pitch targets, this has implications for the phonological description of intonation (Ladd et al., 1999; Atterer and Ladd, 2004). Ladd and his colleagues showed that F0 is related to the segmental structure by means of the 'segmental anchoring' phenomenon. In addition, F0 variation is influenced also by other correlates, such as speakers' emotional states and pragmatic intentions. The prediction of F0 when 'speaking up' has to do with the F0 movements that occur when speakers decide to deliberately raise their voices in noisy or high-emotional contexts (Shriberg et al., 1996). This linguistic phenomenon deals with the idea that it exists 'a raising function by which to relate F0 targets in the raised mode to corresponding targets in the same sentences spoken in the normal mood' (Shriberg et al., 1996: 1). The analysis of speech data from 15 Dutch native speakers (7 males and 8 females) showed that subjects produced higher pitch levels when speaking over a (simulated) noisy telephone channels, as compared to normal face-to-face communication conditions.

It is still an open question whether the variability of pitch range should be described as gradient or categorical. Ladd has argued that, contrary to the view of Bolinger (1989) and Pierrehumbert (1994) who consider pitch as gradient, 'it is theoretically coherent to recognize the existence of factors that are categorical and linguistic' (Ladd, 1996: 282) such as the variability of pitch range. Even though it is clear that F0 variation is continuous along a pitch contour, Ladd and Morton (1997) found some evidence of the fact that pitch movements were categorically interpreted by speakers in their study. These considerations on the categorical nature of pitch by Ladd and Morton were further validated also in a study by Kohler (2004). The categorical essence of pitch is based on the fact that a speaker who has no knowledge of a specific language is however able to perceive F0 changes in the contour. This means, that pitch movements are essentially categorical in their nature.

2.4.2 Patterson (2000)

Various attempts have been made by Ladd and his colleagues to create a sophisticated quantitative model for the analysis of pitch. One of their most felicitous outcomes has been the model proposed by Ladd and one of his collaborators, David Patterson. The model described by Patterson in his Ph.D. thesis was welcomed by the linguistic community as ‘one of very few studies to have treated F0 range as the central object of study’ (Mennen et al., 2012: 2250); Patterson is thought to have proposed ‘a unified model of F0 range variation capable of being applied across a number of domains’ (2012: 2250).

In his Ph.D. thesis, Patterson (2000) argued on the benefits of combining the analysis of LTD (long-term distributional measures) and linguistic measures. On the one hand, LTD measures deal with F0 distribution within a speaker’s contour; on the other hand, linguistic measures are linked to specific targets within the contour, such as peaks and valleys, which LTD measures fail to capture. In a study by Patterson and Ladd (1999), it was emphasized that the advantage of using linguistic measures is that they account for those non-normal distributions along the F0 contours which are eluded in the analysis of LTD. Linguistic measures are calculated by manually labeling the speech corpus and by assigning labels in correspondence to specific target points. The targets Patterson and Ladd (1999) used were: H (on initial peaks), M (on other accent peaks), L (on valleys) and F (on sentence final lows). For F0 level, L and F were counted as the measures of the effective bottom of the range; for F0 span, H-F, H-L, M-F and M-L were measured (Patterson and Ladd, 1999).

A similar method and comparable targets were used by Clark, in his Ph.D. thesis (2003). The targets he used were: start_f0 (calculated at the starting point of a tonal group), end_f0 (calculated at the ending point of a tonal group), min_f0, max_f0, mean_f0, sd_f0 (standard deviation) and delta_f0 (a measure for max f0-min f0). These targets were used in text to speech synthesis to test the validity of measures in capturing pitch movements (Clark, 1999).

The impulse that Clark’s and especially Patterson’s method gave to the analysis of pitch range is based on the intuition of Ladd (1996) about the duality of pitch range (i.e. pitch level vs. pitch span). What is more, Patterson (2000) succeeded in decoding a series of measures, called linguistic measures, able to capture pitch range variation across speakers. Linguistic measures were found to correlate also with judgments given by listeners on a scale of 12 rating parameters: confident, tense, harsh, expressive, deep, weak, irritated, happy, afraid, relaxed, emphatic, and bored (Patterson and Ladd, 1999). The M-L measure (difference between non-initial accent peak and post-accent valley) was found to be the most effective measure for pitch span. The best measure of level was the average of final lows.

2.4.3 Keating (2010)

The model proposed by Keating and her collaborators, Jason Bishop and Grace Kuo, is based on the assumption that voice quality is a cue to location in F0 range.

> 'If voice quality is useful for recovering the location of an F0 in an individual speaker's own range, it means there is a sufficiently salient relationship between a value on acoustic parameter X and a speaker's location in her own individual F0 range, such that value Y on acoustic parameter X indicates location Z in range'
> (Bishop and Keating, 2010: 115).

Voice quality can be acoustically measured from a spectrum where relative amplitude of harmonics in the source has to be calculated. In order to do this, one needs to get estimates of the formant frequencies and correct the harmonic amplitudes. Harmonics are numbered from the first harmonic (H1), which is equal to F0, to the second harmonic (H2), the third harmonic (H3) and so on. Harmonics nearest the formants are called A1, A2, A3 and so forth. Bishop and Keating (2010) make a direct and indirect use of voice quality to determine a speaker's specific F0 range.

The direct method consists on the direct correlation between F0 and H1*-H2*[3]. In a study by Iseli et al. (2007), mentioned also by Bishop and Keating (2010), it is shown that to a low value of H1*-H2* corresponds a low F0 in the sample. Thus, a correlation can be established between F0 and H1*-H2* values. This also implies that any parameter of voice quality 'varies along a speaker's F0 range, and it does so more reliably than does with F0 across speakers' (Bishop and Keating, 2010: 115).

The indirect method is based on the idea that voice quality is a cue for listeners, who compare the F0 range of a speaker to the reference level they are accustomed to. It is

[3] Formants boost harmonic amplitudes so that, to obtain reliable values, it is necessary to correct the harmonic amplitudes. Harmonic amplitudes are corrected for formant frequency and estimated bandwidth, as indicated by an asterisk, e.g. H1*, A1* etc. (Bishop and Keating, 2010).

probable that subjects make decisions about F0 range 'not directly by way of indicating the location in a given speaker's own range per se, but more indirectly by helping to identify the sex of the speakers' (Bishop and Keating, 2010: 116). For example, H1-H2 values are reported to be higher for females than males (Henton and Bladen, 1985; Klatt and Klatt, 1990). Thus, the identification of male vs. female voices is based on a series of factors including F0 range, formant frequencies and other voice properties (Kreiman and Sidtis, 2011). Bishop and Keating (2012) aim at estimating listener's skills in locating F0 changes with a speaker-specific F0 range and examining what factors contribute to identify speaker sex. In order to do this, they calculate F0; cepstral peak prominence (CPP); difference of amplitude between the first and second harmonics, also called open quotient – OQ (H1*-H*2); first, second and third formants (F1, F2 and F3); spectral tilt (H*1-A1* and H*1-A3*); a measure for high-pitched voice quality, characteristic of falsetto (H2*-H4*). In the perceptual experiment, listeners were asked to locate a token in a speaker's individual range. The results were obtained by comparing the following parameters: (a) listener's language; (b) speaker's sex; (c) F0 of the token; (d) measures of voice quality. It was found that 'the greatest predictor of listeners' judgment of F0 location was F0 itself' and 'listeners have separate expectations about F0 ranges for each of the sexes' (Bishop and Keating, 2010: 137). Being F0 the primary predictor of F0 range, voice quality measures were found to be not as significant as previously thought, with the exception of the H2*-H4* parameter considered to be crucial to distinguish male from female voices.

The approach presented in Keating and Kuo (2010) is quite different from the method used in Bishop and Keating (2010). In their comparison of F0 in English and Mandarin, Keating and Kuo (2010) neglected the analysis of voice quality measures in favor of a very detailed study of F0 properties. They based their experiment on two different methods: (1) the cepstral plus manual method and (2) the semi-automated STRAIGHT method.

The cepstral plus manual method used the cepstral pitchtracker in the PCQuirer/Pitchworks program. Since this program occasionally makes some mistakes in correspondence with unvoiced sounds, creaky intervals or missing values, a substantial revision of pitch tracking errors was needed. Pitch setting parameters had to be adjusted and some F0 values were calculated directly from the waveform using the formula for F0= 1/T (see the paragraph on pitchtracking in Keating and Kuo, 2010: 170-171). This method was found to be reliable but definitely time-consuming. For this reason, only a small selection of the corpus materials was analyzed with the cepstral plus manual method.

The semi-automated STRAIGHT method focuses on an algorithm used in a new application for voice analysis called VoiceSauce (Shue et al., 2009, 2011). VoiceSauce has been developed by a group of linguistic and electrical engineering researchers at University of California, Los Angeles. This program gives automated voice measurements over time from audio recordings and computes a number of voice measures[4] including automatic corrections for formant frequencies and bandwidths. In particular, VoiceSauce allows running entirely automatic measures of F0 values at 1 ms intervals. The algorithm and specific corrections incorporated in VoiceSauce permits to minimize pitch tracking errors since STRAIGTH finds 'very low F0 values directly from the waveform' (Keating and Kuo, 2010).

In the experiment, some measures reviewed by Baken and Orlikoff (2000) were adopted. Also the most extreme F0 values for each speaker were included. Table 1 lists all the values used in the study by Keating and Kuo (2010).

[4] Measures computed by VoiceSauce include: F0 from STRAIGHT (Kawahara et al., 1999), Snack Sound Toolkit (Sjölander, 2004) and Praat (Boersma and Weenink, 2008) algorithms; harmonics measures both corrected (*) and uncorrected, H1-H2 (*), H1-A1 (*), H1-A2 (*), H1-A3 (*), H2-H4 (*); formants and bandwidths (F1, F2, F3, F4, B1, B2, B3, B4); energy; subharmonic to harmonic ratio, cepstral peak prominence; harmonic to noise ratios.

Measure	Abbreviation	Unit	Definition
Minimum F0	Min	Hz	Lowest F0 value in a token
Maximum F0	Max	Hz	Highest F0 value in a token
Mean Minimum F0	MeanMin	Hz	Average Min across tokens
Extreme Minimum F0	XMin	Hz	Lowest Min across tokens
Mean Maximum F0	MeanMax	Hz	Average Max across tokens
Extreme Maximum F0	XMax	Hz	Highest Max across tokens
Mean F0 Range	MeanRange	Hz	MeanMax – MeanMin
Mean F0 Range in semi-tones	MeanRangeST	Semi-tone	39.863 * log(MeanMax/MeanMin)
Extreme F0 Range	XRange	Hz	XMax - XMin
Extreme F0 Range in semi-tones	XRangeST	Semi-tone	39.863 * log(XMax/XMin)
Mean F0	Mean	Hz	Average of F0 values in a token
Mean of Mean F0	MeanMean	Hz	Average Mean across tokens
Standard deviation of F0	SD	Hz	SD of F0 values in a token
Mean of Standard deviations of F0	MeanSD	Hz	Average SD across tokens

Table 6. Inventory of F0 measures calculated in Keating and Kuo's STRAIGHT method (2010: 172).

The results of the Keating and Kuo (2010) study, obtained with the semi-automated STRAIGHT method, showed a significant size effect for sex in English vs. Mandarin. By contrast, no significant effect was found with the cepstral plus manual method.

2.4.4 Mennen (2007-2012)

In line with the model elaborated by Patterson (2000), Mennen and her collaborators tackled the problem of conveying information about the distribution of F0 values along the pitch contours. In line with previous findings of several studies (Clark, 1999; Patterson and Ladd, 1999; Patterson, 2000; Clark, 2003), they came to the conclusion that a two-dimensional model of pitch range, based on (a) two distinct dimensions (i.e. pitch level vs. pitch span) and (b) two distinct approaches (i.e. LTD vs. linguistic measures) is needed.

LTD measures such as, F0 maximum, F0 minimum, F0 mean, F0 median are combined to calculate F0 level; measures such as F0 maximum – F0 minimum, 1/4 standard deviations around mean, 95^{th}-5^{th} percentile, 90^{th}-10^{th} percentile, skew and kurtosis are calculated to measure F0 span. The method proposed by Mennen et al. (2007, 2012) added to the measures listed above (also called LTD measures) the so called 'linguistic measures', which were found to be better predictors of difference in pitch range across speakers and languages. The idea of using linguistic measures for pitch analysis was based on the assumption that pitch can best be described in terms of pitch movements than pitch levels.

With this idea in mind, Mennen and her collaborators analyzed F0 range (level vs. span) by manipulating the speech signal and adding F0 landmarks at specific target points. By creating a 'manipulation object' in Praat, two different panels were compared: the original pitch contour and the new manipulated pitch contour. Manipulation was obtained by editing pitch points in the F0 contours. Pitch points, corresponding to F0 landmarks, were placed only at specific target points (usually F0 maximum and F0 minimum). They could be inserted, deleted or shifted in time and frequency, depending on the F0 movements. F0 landmarks were derived via visual, auditory and linguistic inspection. All the procedure was articulated in four steps:

1) IP (intonation phrase) received an initial and a final pitch point (labeled respectively as 'I' for initial point and 'F' for final point);
2) Every F0 maximum and minimum was marked by a pitch point, also called F0 landmark;
3) Additional landmarks were placed wherever there were changes in slope that deviated from the original F0 contour;
4) Every landmark was annotated with specific labels, giving indications on the time of occurrence (initial position or final position), the F0 level (high or low value) and the prominence of syllables corresponding to pitch points (unaccented or unaccented syllables).

By following these four steps, 'pitch contour was interpreted as a combination of pitch targets and linear interpolations between targets' (Mennen et al., 2008). Thus, a series of different measures could be calculated at specific pitch targets located on initial peaks, on other accent peaks, on valleys and on sentence final lows or highs. The F0 movements determining rises or falls in upward and downward slopes were marked by assigning F0 landmarks.

The primary difficulty in describing the F0 contour is the extent to which the F0 continuous can be described as distinct pitch points. According to D'Imperio (2002:103), 'there are cases in which only stipulative decisions can be made regarding tonal target position'. As pitch range varies considerably both within and between speakers, pitch range is not a single determined unit (set at certain Hertz values). The peculiarity of the analysis system created by Patterson (2000) and implemented by Mennen and her colleagues (2006, 2008a, 2008b, 2012) is that this method is able to abstract away from a surface F0 contour, that is unavoidably full of irregularities, and synthesize the main pattern of a sentence by processing data that produce comparable, simple and reliable results. This is possible only by means of a two-level analysis of pitch range where relative height is differed by H or L values. However, the identification of peaks and valleys within a pitch contour is not always a simple target.

> 'In identifying peaks (H) and troughs (L) in the F0 contour itself, in many cases the physical evidence is not clear cut. For example, the high or low value may be sustained over a period of time, and analysts must make a principled decision: should they choose a point at one or other edge of this sustained frequency, or select a mid-point?'
> House and Wichmann (1996: 3).

Pitch range should be measured as a sequence of target points that give account of F0 movements; a two-level analysis fails to capture the great variability within a pitch contour. For this reason, a more sophisticated system of labels was elaborated in order to specify the direction, the timing and the prominence of F0 movements. The labels used in Mennen et al.'s approach were somehow borrowed from principles of the Autosegmental-Metrical approach of intonation analysis (as exemplified by Pierrehumbert, 1980) and rearranged to define local pitch excursions within F0 contours. The labels used by Mennen and colleagues were:

- **I** Phrase initial value
- **H*-H** Local peak, (non)-prominent syllable
- **L*- L** Local valley, (non)-prominent syllable
- **!H*** Change in downward slope (accented)
- **D** Change in downward slope (unaccented)
- **$L*** Change in upward slope (accented)
- **U** Change in upward slope (unaccented)
- **FH** Final local maximum
- **FL** Final local maximum or minimum

The phrase initial value, transcribed as I, works as the reference line for the starting point of the sentence. To the phrase final value F, a label indicating the F0 location of the pitch point is added (either H or L) to provide information about the direction of movement. This is due to the fact that pitch points at the end of the phrase are phonologically associated to boundary tones (see Autosegmental-Metrical theory).

Indications about a final downward or upward slope are needed in order to distinguish F0 final movements (e.g. boundary tones placed at the end of a phrase distinguish declaratives from questions). Local peaks and valleys are transcribed as H and L, prominence-lending landmarks are signaled with an asterisk (e.g. H* and L*) to distinguish them from landmarks appearing in non-prominent syllables (e.g. H and L). The status of prominent vs. non-prominent peaks or valleys is based solely on whether H and L are placed on prominent vs. non-prominent syllables.

Changes in upward or downward slopes that do not appear in correspondence to peaks or valleys are described as D (for 'down') and U (for 'up') when changes occur in unaccented syllables. In the case of accented syllables, two different labels are used to indicate upward ($L*) or downward (!H*) slopes. It has been claimed that that the 'decision to assume a direct relationship between turning points and phonological tones was driven by practical reasons so as to ensure consistency' in the analysis and labeling procedure' (Mennen et al., 2012: 2259). This implies that there is no one-to-one correspondence between the turning points and the phonological structure underlying.

The most recent version of Mennen et al.'s method was tested in a study on the pitch range of English and German females, published in March 2012. Pitch contours were manipulated in Praat and re-synthesized. After placing all the landmarks described above, a script in Praat was used to calculate the F0 of each pitch point. Then, values were averaged across speakers (female only) and language (English vs. German). In table 2, all measures calculated in the production study were divided according to pitch dimension (level vs. span) and methodological approach (LTD vs. linguistic measures).

Pitch level was calculated by measuring mean F0, median F0, maximum and minimum F0 (for LTD measures); first and final landmarks, transcribed as I, FL, FH, and measures for peaks and valleys, transcribed as H and L (for linguistic measures). Pitch span was calculated with LTD measures by averaging the difference between highest and lowest F0 values (e.g. max-min F0) in Hz, ST, and ERB; 1/4 standard deviations around mean; 90% span; 80% span; skew and kurtosis. As for linguistic measures, the difference between highest and lowest F0 values was calculated on the difference between peaks targets (first peak, H*i, H*) and valleys (L, FL) in Hz, ST, and ERB.

	LTD	Linguistic
Level	Mean f0	First peak
	Median f0	H*i
	Maximum	H*
	Minimum	(Hi)
		(H)
		(L*)
		L
		FL
Span	SD4	H*i-L
	SD	H*i-FL
	Max-min f0	H*-L
	Max – min ST	H*-FL
	Max-min ERB	H*-FL ST
	90% span	H*-FL ERB
	80% span	First peak–L
	Skew	First peak–L ST
	Kurtosis	First peak–L ERB
		First peak–FL

Table 7. Inventory of LTD and linguistic measures calculated for F0 level and span in Mennen et al. (2012: 2254).

While working on their data, Mennen et al. took the decision of excluding target points for changes in upward and downward slopes, transcribed as !H*, $L*, D, U, 'as they did not constitute local minima or maxima and were therefore considered of minor importance for f0 range assessments' (2012: 2253). Indeed, those landmarks were discarded from the analysis because preliminary results showed that values for !H*, $L*, D, U reached a very small impact on the size of pitch movements. This was probably due to the fact that upward or downward changes in the slope had a minor effect than peaks and valleys (based on personal communication with Mennen).

In order to test the reliability of this labeling system, inter-rater agreement was calculated for a subset of data in the corpus. The results obtained by the Cohen's kappa test showed that the annotation data were highly reliable (with a kappa value being 0.67). Interestingly, the results from the production study showed that F0 range is based on 'the phonological and/or phonetic conventions of the language being spoken and is not solely an artifact of physiological factors or cultural differences, as often assumed' (Mennen et al., 2012: 2258). In particular, the data obtained in this study are in line with reported stereotypical beliefs that English females have higher F0 level and wider F0 span than German females. This has implications on cross-linguistic studies and L2 learning.

As a follow up of the production study, Mennen et al. (to appear) conducted a perception study aiming at identifying 'the nature of the cross-language differences in performance which underpin the differences that people perceive' (Mennen et al., 2008a: 16). This perception study shows that:

> 'methods developed for capturing within-language pitch range variation (such as the one developed by Patterson, 2000) cannot be applied without adaptation to cross-language or cross-regional comparisons. For example, Patterson's best measure for level (final low) was not a good predictor of language membership'
> (Mennen et al., 2008a: 17).

These considerations shed some light on the issue of creating a standardized but versatile model, suitable for different languages. The goal of the investigations by Mennen et al. (2007, 2008a, 2008b, 2012) was to assess the validity of LTD vs. linguistic measures in order to test, whether or not, a combination of these two strategies could account for differences in pitch range. The results showed that 'a pitch range model that is based on linguistic dimensions of variation better captures cross-language variation' (Mennen et al., 2008a: 15).

2.5 Summary

Chapter 2 begins the discussion with an introduction to the acoustic analyses of pitch range. Before tackling the problems connected to the lack of generally agreed-upon methodologies and measures through which to carry out the analysis, the notion of fundamental frequency (F0) and pitch are defined from an acoustical and perceptual perspective. What is more, pitch level and span are identified as the two major components of pitch range which can be measured with linear or non-linear frequency scales and is affected by segmental features. Also speaking rate and the standard duration of utterances plays a role in determining higher or lower F0 slopes.

The chapter also explains how pitch range is differently modulated depending on the emotions expressed by the speakers. In fact, pitch modulation reflects the speakers' attitude and intent by conveying meaning through the expression of emotional states. A number of anatomical and physical factors such as the gender, the age, the body size and the heath history of the speakers are shown to have an impact on the realization and perception of pitch. In addition, socio-cultural factors determine different realizations of pitch across languages and cultures. Observations on people’s voice pitch have been decoded into three universal principles called biological codes (Ohala, 1984; 1994). These codes provide a comprehensive picture of the global and local characteristics of pitch range. Despite the lack of a generally acknowledged methodology for the analysis of pitch range, four of the most valid methods proposed in recent years (Ladd, 1996; Patterson, 2000; Keating, 2010; Mennen, 2012) are outlined and discussed.

Chapter 3:
Cross-linguistic research on L2 pitch

3.1 The Auto-segmental Metrical Approach

The Auto-segmental Metrical (AM) approach is the method adopted in the vast majority of current studies to analyze intonation patterns. The first theorizers of the AM system (Liberman, 1975; Bruce, 1977; Pierrehumbert, 1980; Gussenhoven, 1984; Liberman and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Pierrehumbert and Hirschberg, 1990; Ladd, 1996;) based their studies (and in some cases, their Ph.D. theses) on the study of intonation systems and models. From their theories, the current AM model has been shaped as a standard and unified model to describe intonation across languages. This approach is characterized by a peculiar meaning-form connection, based on the 'decomposition of intonation contours into meaningful parts' (Warren, 2005: 225). These parts work as whole, divided into components of melodic patterns, analyzed across different tiers. Among a number of tiers (segmental, timing and stress tier), the tonal tier represents sequences of tonal events where pitch movements move from high to low and vice versa.

The AM framework treats intonation in terms of pitch accents, phrase accents, and boundary tones. Each of these tonal features is prosodically marked and transcribed with the symbol H (high) or L (low), depending on the relative height of pitch along the pitch pattern. Pitch accents are associated with those syllables that carry metrical prominence in the utterance, realize a significant pitch excursion, and change their movement either upwards or downwards. Generally,

> 'an auto-segmental tonal analysis implies that the phonetic contour is constructed from 'levels', a traditional term for 'pitch points', rather than 'contours', a traditional term for 'pitch movements'. The representations presented so far are in the auto-segmental tonal tradition: tone is

> represented by discrete elements (i.e. H and L) that are interpreted as pitch levels. In this 'levels' approach, a rising pitch contour is represented as a L followed by a H, and a falling contour as a H followed by a L. The pitch between the H's and L's must be filled in by the phonetics'.
> (Gussenhoven, 2002: 5)

Thus, pitch movements are described and analyzed in terms of categorically distinct elements that provide strings of H and L markers. These markers are associated to pitch accents, phrase accents and boundary tones; all these elements determinate the intonation contour. 'The meanings of intonation contours are said to be compositional in the sense that each tone in any sequence contributes separately to the overall meaning and the meaning of the whole is equal to the sum of the parts' (Cruttenden, 1997: 64). Therefore, sequences of H and L markers describe the specific contours of intonation patterns by providing linguistic and paralinguistic meaning.

The essential rules and principles outlined in the AM approach have been defined by Ladd (1996: 42-43) as follows:

1. *Linearity of tonal structure:*
   A linear tonal structure consisting of sequences of rises and falls can be represented by pitch movements analyzed as strings of tonal events. Specifically, contours appear to be broken down into smaller constituents that identify peaks vs. valleys.
2. *Distinction between pitch accent and stress:*
   Pitch accents are those tonal features that identify pitch movements by distinguishing between prominent vs. non-prominent syllables. Stress is somehow independent from the realization of pitch accents, since pitch accents may be realized in syllables that are not necessarily stressed.
3. *Analysis of pitch accents in terms of level tones:*
   Combinations of pitch targets, H (high) and L (low), are used as constituents of melodic patterns. Pitch targets describe distinct pitch configurations, based on the phonetic evidence provided by the distinction between high and low pitch levels.
4. *Local sources of global trends:*
   Sequences of phonological events describe a number of global trends, such as the declination phenomenon (i.e. the gradual lowering of pitch along the tonal space) and the low pitch at the end of a sentence-final fall (i.e. the

level of low pitch at the end of falling contours varies only slightly across speakers).

These paradigms focus on the description of the main tendencies of pitch range variation, described according to the AM phonological framework. Even though this system does not make predictions on the phonetic variation of pitch range, it gives a broad description of the regularities expressed by the registered variation across pitch level and span. In particular, the relation between excursion size (pitch span) and overall pitch (pitch level) is much more regular than the relation between peaks in a sentence (see Lieberman and Pierrehumbert, 1984: 210-215).

Ladd (1996) also elaborated a taxonomy of intonation differences across languages, where the intonation variation can be categorized into four different dimensions: the semantic, systematic, realizational and phonotactic variation.

Semantic variation captures ‘differences in the meaning or use of phonologically identical tunes’ (Ladd, 1996: 119). Variation among language varieties occurs when the same intonation pattern is used to convey distinct meanings and functions. An example of semantic difference between language varieties is the case of the pitch patters of questions. For instance, questions produced with a high rise convey a meaning of wheedling or insistent request in British English. By contrast, a high rise question in ‘American varieties of English is frequently used to express a kind of statement while at the same time asking for feedback from the interlocutor’ (Kügler et al., 2009: 12). Also the Australian and New Zealand English intonation of questions is different from the British English pattern. In fact, in Australian and New Zealand English phonetically identical high-rising patterns can be used to signify both yes-no questions and statements (Fletcher et al., 1999).

Systematic variation considers ‘differences in the inventory of phonologically distinct tune types, irrespective of semantic differences’ (Ladd, 1996: 119). It is realized when speakers use different intonation patterns to express the same function. For example, two or more languages may use variation of the same patterns (e.g. low and high rises) to express the same meaning: a request for information. Systematic variation mainly occurs within four dimensions: 1) *phrasing*, the separation of intonation contours into chunks by means of boundary tones; 2) *pitch accent*, the collocation of prominence on accented syllables; 3) *tune*, the compositional meaning of intonation contours; 4) *pitch range*, the systematic variation within melodic patterns, across level and span (Ladd, 1996). A specific case of the systemic variation is the phenomenon of downstep in declarative intonation. Downstep, that is ‘a stepwise lowering of high F0 targets at well defined points in the utterance’ (Docherty and Ladd, 1992: 331), is frequently realized in declarative sentences. The meaning of a sentence is not affected by the presence or absence of downstep. In fact, the downstep ‘adds a nuance of greater finality, but does not otherwise seem to affect the meaning of the contour’ (Ladd, 1996: 126) in a number of languages. However, in European Portuguese, downstep is specifically associated to the conditions of broad vs. narrow focus (Frota, 1995). Portuguese sentences in broad focus condition are usually realized with a final downstep while the sentences in narrow focus condition do not.

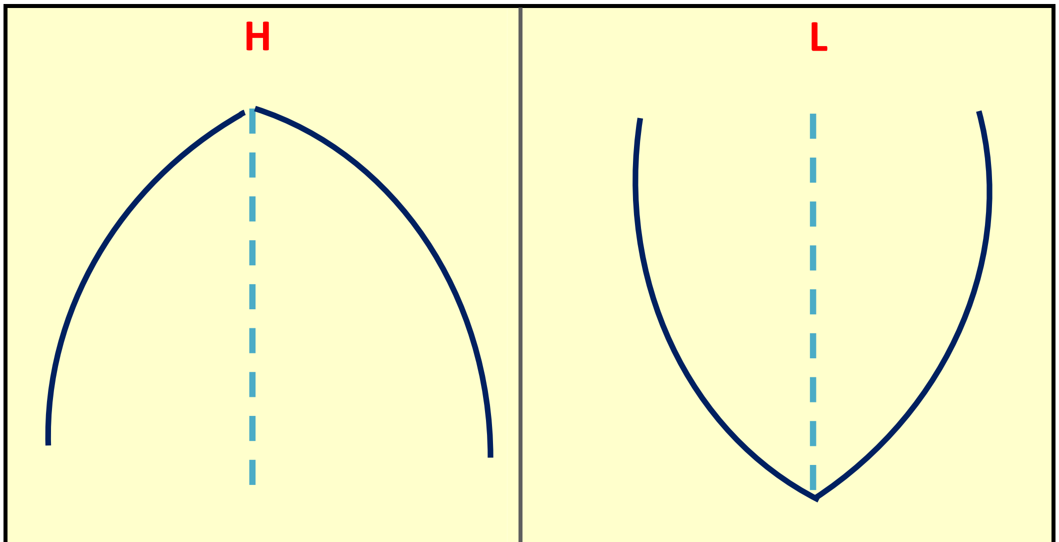
Realizational variation in the intonation of languages analyzes the ‘differences of detail in the phonetic realization of what may be regarded phonologically as the same tune’ (Ladd, 1996: 119). For instance, the alignment of tones relative to segments, syllables or tone bearing units depends on different factors such as syllable structure, proximity of other prosodic events (accent tones or prosodic boundaries), information structure, and discourse structure. However, dialect varieties differ in the exact segmental anchoring of tones, that is, in tonal alignment (Arvaniti et al., 1998, 2006; Ladd et al., 2000; Atterer and Ladd 2004; Gilles 2005). Tonal alignment is an important phenomenon that captures the realizational variation across languages. Tonal alignment occurs in both English and Italian; however, its realization is different. In English the peak on the stressed syllable is realized later than in Italian. Ladd (1996) mentioned the word Mantova as an example. Even though most L2 speakers of Italian manage to correctly place prominence on the first syllable, their tonal alignment is usually delayed. In fact, the word Mantova is pronounced by L2 speakers of Italian with a late peak in the falling contour. A delay in the realization of the peak ‘may be interpreted by native Italians as a mistake in the placement of word stress, i.e. they may perceive this as stressed on the penultimate, rather than on the antepenultimate syllable’ (Mennen, 2006: 5). Thus, non-native speakers of Italian fail to produce the correct realization of tonal alignment because they place the peak rather late, at the end of the stressed syllable (Ladd, 1996).

Phonotactic variation focuses on the ‘differences in tune-text association and in the permitted structure of the tunes’ (Ladd, 1996: 119). Thus, it evidences restrictions in the way phonological components are combined. Given the fact that phonotactics is based on phonological constraints (Hayes and Wilson, 2008), every language has specific phonological rules that cannot be violated. For example, some phonological constraints relate to the structural distribution of phonemes. Unlike in Italian, English and other languages, in Japanese ‘a liquid (r) can never follow a stop consonant (p, b, k…)’ (Ramus and Mehler, 1999: 512). When considering the tone system of a language variety, it is fundamental to codify the phonotactic variation of certain phonological phenomena by getting phonological associations right. According to the AM theory, pitch accents are placed on stressed syllables. However, in Italian, an additional emphatic falling declarative accent may be located on final unstressed syllables to add emphasis or convey emotional involvement (Ladd, 1996). In this case, it is appropriate to place a pitch accent on an unstressed syllable, because the speaker’s intent is to increase the emphasis on a particular word.

3.1.1 The tone system

Tone has been defined as ‘the contrastive pitch movement on the tonic syllable’ (Tench 1996: 73) and it represents the auto-segmental property *par excellence* (Hyman, 2007). In fact, it describes fluctuations within intonation contours by signaling changes in pitch movements. As it is known, English and Italian are not tone languages because they do not exploit their tone systems to convey and distinguish among lexical meanings. Differences in pitch do not transmit different meanings in lexical words; however, pitch has a substantial role in providing grammatical, informational and affective meanings (Gussenhoven, 2002). Tones work as prosodic components that, while interacting among themselves, create a unitary picture of an utterance. Specifically, ‘tone sequences are much more likely to be treated as contours which can be manipulated as units or as “melodies” which can be mapped’ (Hyman, 2007: 518).

Within the AM framework, a two-level tone system was proposed. Tone-bearing units constitute the primary components of the AM approach, because H and L targets indicate peaks and valleys along the intonation melodies. As shown in fig. 8, H represents the peak of a stressed syllable while L represents a valley on the stressed syllable.



Peak: A H* mark signals the presence of a peak on the stressed syllable of the accented word. Sometimes a peak can be labeled as L+H* when there is a long rise to the peak.	**Valley:** A L* mark signals the presence of a valley on the stressed syllable of the accented word. A boundary tone, L%, is characterized only by the first part of the curve.

Figure 8. How H and L signal peaks and valleys.

The incidence of peaks and valleys within a sentence is an effect of pitch level and span variation. Thus, the more F0 excursions occur along a pitch pattern the more varied the pattern of peaks and valleys results.

3.1.2 The American and British approaches

The AM approach analyzes pitch movements as sequences of pitch features, rather than as contours. In particular, in the AM approach, pitch patterns are considered in terms of levels (i.e. high and low) and not in terms of configurations of rises and falls. The AM approach, developed in the U.S., differs from the approach to the analysis of intonation developed in Great Britain.

The British nuclear tone framework is based on the assumption that information is provided by spoken material divided into chunks. Intonation phrases, called IPs, are the basic constituents of sentences and are characterized by their own tonal structure (Wells, 2006). The specific tone associated to each IP determines the representation of the contours. It is clear that when an IP performs an abrupt rise, the contour is described as a high rise; when the IP performs a fall followed by a rise, the resulting contour is a fall-rise. Typically, an IP can be composed of four different elements: the pre-head, the head, the nucleus, and the tail (Ladefoged, 2006). Following Wells (2006: 7-9), the head and the nucleus of an IP always carry stress. While the head is the most prominent stressed word, usually the nucleus is placed on the last stressed syllable of an utterance. The tail is, by definition, the last part of the IP and contains no stressed syllables. Also the pre-head, the first part of the IP, is never stressed.

The British nuclear tone system was compared to the AM system. In recent years, the AM system has been one of the most appraised (but also debated) systems. It is still difficult to decide which approach is superior to the other because both systems, based either on the British or on the American traditional phonology, present positive and negative aspects. Cruttenden argued (1997: 66) that 'the transcription of text using H's and L's is an altogether much more difficult affair than using the tone marks most widely used in nuclear tone analysis.' However, despite the relative complexity of the AM transcription system, a system based on simple sequences of H and L targets permits to annotate larger corpora of speech materials and compare them. Cruttenden himself admitted that the AM system better than the British nuclear analysis captures 'the relationship between level and non-level contours' (1997: 64). Thus, all in all, the AM approach is a largely agreed system that provides a better analysis of intonation patterns, as compared to the British nuclear tone system. This is due to the fact that the AM approach 'views pitch contours as composed of a sparse linear sequence of accents and tones selected from a relatively small inventory' (Silverman et al., 1992a: 436).

In order to better examine and evaluate the approaches proposed within the AM and the British nuclear tone frameworks, fig. 9 compares how intonation patterns would be analyzed following the two approaches. As it is clear from the picture, the AM and the British nuclear tone frameworks have different ways of transcribing upwards and downwards pitch movements. The AM approach (left box) analyzes pitch patterns as sequences of highs and lows, while the British tradition (right box) considers the development of the sentence as sequences of rises and falls.

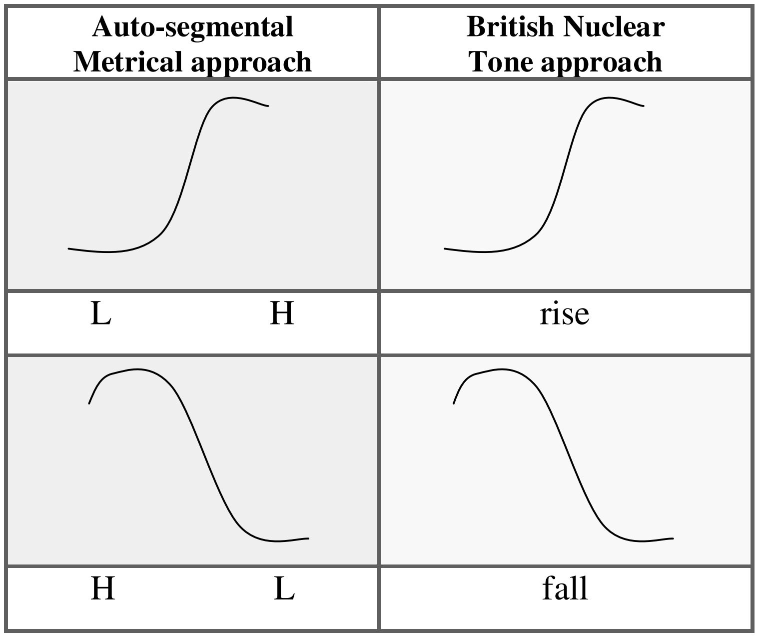


Figure 9. Difference in the annotation system of pitch movements: Auto-segmental Metrical approach (left) and British Nuclear Tone approach (right).

In the AM approach, an upward pitch movement is described as a sequence of two tones, labeled respectively as L and H. By contrast, the British nuclear tone approach simply identifies an upward movement with a rise. As for downward pitch movements, in the AM theory, they are annotated as a sequence of H and L while, in the British nuclear tone approach, they are transcribed as falling contours.

While the British nuclear tone system provides a general indication on the direction of the movement and on the entity of the rise or fall (low or high), the AM approach offers a number of other ways to provide information on the nature of the rise and falls. For example, the AM approach gives indications on the nature of the pitch movements by signaling the status of pitch accents, phrase and boundary tones (see the description of ToBI, §3.2). Indeed, the AM framework addresses the connection between the segmental and suprasegmental levels of speech by considering the tonal composition of a sentence, the link between prosodic and segmental context and a number of events related to the tonal environment (such as, pitch and phrase accents, boundary tones, rhythmical organization of a sentence, focus, pitch range, speaking rate and intonation mode).

Despite the differences between the two approaches, many correspondences can be found among the transcriptions of tonal patterns analyzed according to the AM approach and the traditional British nuclear tone system. The table below (tab. 8) shows twenty-two combinations of patterns following the annotation code of the AM system elaborated by Pierrehumbert (1980) and the British nuclear tone framework.

Auto-segmental Metrical		British nuclear tone
H*	L-L%	Fall
H*	L-H%	Fall-rise
H*	H-L%	Stylized high rise
H*	H-H%	High rise
L*	L-L%	Low fall
L*	L-H%	Low rise (narrow pitch range)
L*	H-L%	Stylized low rise
L*	H-H%	Low rise
L+H*	L-L%	Rise-fall
L+H*	L-H%	Rise-fall-rise
L+H*	H-L%	Stylized high rise (with low head)
L+H*	H-H%	High rise (with low head)
L*+H	L-L%	Rise-fall (emphatic)
L*+H	L-H%	Rise-fall-rise (emphatic)
L*+H	H-L%	Stylized low rise
L*+H	H-H%	Low rise
H+L*	L-L%	Low fall (with high head)
H+L*	L-H%	Low fall-rise (with high head)
H+L*	H-L%	Stylized high rise (with high head)
H+L*	H-H%	Low rise (high range)
H*+L	H-L%	Stylized fall-rise
H*+L	H-H%	Fall-rise (high range)

Table 8. Correspondences between Pierrehumbert's annotation system and British-style nuclear tones (from Ladd, 1996: 82).

Almost perfect correspondences are evidenced between the two systems. Even though the AM and British nuclear tone systems are based on different descriptive criteria, the parallelism between the two is clear. As Toivanen (2005:2) argued, the main examples of correspondences between the two systems can be 'represented as follows: high-fall (H* L-L%); low-fall (L* L-L%); high-rise (L* H-H%, H* H-H%); low-rise (L* H-L%); level tone (H* H-L%)'. In his comparison of the AM and British-style labels (tab. 8), Ladd (1996) pointed out that the grouping based on the AM system (that is the analysis by Pierrehumbert, 1990) shows five completely parallel sets of pitch accents (i.e. H*, L*, L*+H, L+H*, H+L*) found in four sequence types plus one additional pitch accent (i.e. H*+L) collocated in two sequence types. From the point of view of the British tradition, some 'transcriptions like 'low rise' and 'high rise' show up rather unpredictability at several different places in the table, and references to pitch range or to the preceding head are required here and there to describe certain distinctions' (Ladd, 1997: 86).

3.2 ToBI

The Tones and Break Indices annotation system (henceforth ToBI) is a methodology for transcribing intonation patterns and other prosodic elements, such as stress, pauses etc. It was originally created by Pierrehumbert (1980), but it has been re-elaborated and modified in later studies (Pierrehumbert and Hirschberg, 1990; Silverman et al.,1992; Syrdal et al., 2001, Beckman et al., 1994, 2005; Veilleux, et al., 2006). ToBI has been conceived as 'an agreed system for transcribing prosodic structures which could be used consistently by researchers in various fields' (Halliday and Greaves, 2008: 12). Undoubtedly, it is one of the most popular annotation systems of intonation[5].

This model of intonation description is used to talk about prosodic phenomena, 'to allow researchers to compare their findings more easily, within and across languages, and to facilitate the construction of very large speech corpora, especially for learning associations between prosodic features and other aspects of texts' (Hirschberg, 2002: 31). Thus, ToBI represents a standard system for prosodic transcription and it has been developed within and adapted to different scientific disciplines (e.g. speech transcription and recognition, phonetics and phonology, speech therapy and psychology, etc.). Compared to other systems in use, the main innovation proposed by ToBI is that 'rather than analyzing intonation patterns in terms of pitch contours (rise, fall, fall-rise, etc.), [it] breaks them down into components, basically High and Low in various combinations' (Wells, 2006: 261).

Beckman and Elam (1997) outlined the basic characteristic of ToBI as consisting of a four-level analysis based on four tiers: 'an orthographic tier, a break index tier, a tonal tier and a miscellaneous tier' (see also Cruttenden, 1997: 59; Hirschberg 2002: 33, 2004: 2). The orthographic tier contains segmental transcriptions of time-aligned words where words are described as strings of vowels and consonants (in some cases, the transcriptions are based on the IPA alphabet). The break index tier indicates the entity of pauses among strings of words, that is the presence/absence of boundaries. Levels of juncture are rated on a five point scale, where 0 stays for 'no word boundary and 4 for

[5] A comprehensive description of the ToBI system can be found in the ToBI conventions document and the training materials are available at http://ling.ohio-state.edu/tobi, (written by Beckman and Elam, 1997). This model is based on earlier work by Pierrehumbert (1980,1987) and Pierrehumbert and Hirschberg (1990).

boundary tones (see Price et al., 1990). The miscellaneous tier signals additional prosodic phenomena that may be optionally marked, such as disfluencies, creaky or breathy voice, etc. (Hirschberg, 2002).

The center of the linguistic analysis lays especially on the tonal tier, ‘where pitch accents, phrase accent, and boundary tones describing targets in the fundamental frequency (F0) define intonational phrases’ (Hirschberg, 2002: 33). The tonal tier is mainly concerned with the pitch movements of pitch accents, phrase accents and boundary tones. These movements or pitch tendencies are identified as either H or L.

H and L respectively represent high or low pitch. Depending on the location and role of pitch movements within an utterance, they are defined by additional diacritics (e.g. *,-, !, %) indicating specific tonal events. High and low pitch accents, placed on prominent syllables, are labeled respectively as H* (read ‘H star’) and L* (read ‘L star’). The asterisk (that stays for prominent syllable) indicates the central part of the pitch accent and it is never positioned on a phrase accent or on a boundary tone (Ladefoged, 2006). Unlike pitch accents (that are identified by the asterisk on the prominent syllable), phrase accents indicate the general tendency of an utterance, that is upwards vs. downwards. The phrase accent is that tonal component placed ‘between the last pitch accent and the boundary tone and it is represented by H- or L- without any diacritic’ (Halliday and Greaves, 2008: 12). The symbol ‘–’ indicates that the L or H targets describe a pitch movement at the level of phrase accent. As for boundary tones, in order to differentiate them from other kinds of pitch and phrase accents, they are signaled by a percentage sign (%). Since this symbol identifies the boundary tone as the last component of an utterance, it is usually placed at the end of a sentence. The boundary tone is marked as H% or L% depending on whether the utterance ends with a rising or falling pitch (Ladefoged, 2006).

The distribution of pitch accents (e.g. H*, L*, L*+H, L+H*, H+!H*, H*+!H), phrase accents (e.g. H- and L-) and boundary tones (e.g. H% and L%) has been graphically schematized in tab. 9. This model is based on the system elaborated by Hirschberg (2002) and Ladefoged (2006). Previous and later versions of ToBI have been proposed by several other scholars (among others, Pierrehumbert and Hirschberg, 1990; Silverman et al.,1992; Syrdal et al., 2001, Beckman et al., 1994, 2005 and Veilleux, 2006).

Pitch Accent		Phrase Accent	Boundary tones
H*			H %
L*		L -	L %
L* + H			
L + H*			
H +!H*		H -	H %
H* !H*			L%

Table 9. *ToBI Contours for Standard American English.* Standard schematic representation of all the possible combinations which can occur in Standard American English (adapted from Hirschberg, 2002 and Ladefoged, 2006).

It is generally agreed upon that a standard declarative intonation pattern ends in a low phrase accent and a low boundary tone (L-L%) while a standard yes-no question ends in a high phrase accent and high boundary tone (H-H%) (Hirschberg, 2004 and Ladefoged 2006). However, it is difficult to predict which sequences of tonal events are produced to realize specific sentences because of a virtually endless set of possible combinations. An unpredictable number of pitch accents can precede the realization of a phrase accent and a boundary tones. Thus, even though the occurrence of tonal targets is not fixed, their role is well-defined and predetermined. For instance, the single tones H* and L* indicate a single high or low peak accents. The combined tones L*+H and L+H* designate respectively a 'scooped' late rise and peak realized with a gradual rise from a low pitch (Pitrelli et al., 1994). As for the broad pitch tendency, the phrase accents L- and H- suggest the tonal development of the sentence. The low sentence final is indicated by L%, while the high sentence final is signaled by a H% (Martin, 2004: 2)

While ToBI has been specifically tailored to the analysis of American English, the British nuclear tone system has been modeled on the intonation characteristics of British English. Even though ToBI has been developed for American English, in recent years, it has been adapted also to other languages. More than fourteen versions of ToBI have been developed to describe different language varieties. To date, the models available are:

1. **MAE_ToBI**: model of Mainstream *American*
   (Beckman, Hirschberg and Shattuck-Hufnagel, 2005)
2. **CatToBI**: model of *Catalan*
   (Escudero-Mancebo and Aguilar, 2010; Priteto et al. 2007; Prieto, 2002, 2013)
3. **CToBI**: model of *Cantonese*
   (Wong, Chan and Beckman, 2005)
4. **ToDI**: model of standard *Dutch*
   (Gussenhoven, 2002a, 2002b, 2005; Gussenhoven, Rietviled, Kekhoff, and Terken, 2003)
5. **FToBI**: model of standard *French*
   (Jun and Fougeron, 2000; 2002)
6. **GToBI**: model of standard *German*
   (Grice and Benzmüller, 1995; Grice, Baumann and Benzmüller, 2005)
7. **GRToBI**: model of Athens *Greek*
   (Arvaniti and Baltazani, 2000, 2005; Arvaniti, 2009; Arvaniti and Ladd, 2009)
8. **IToBI**: model of Neapolitan, Bari, Palermo and Florentine *Italian*
   (Grice, D'Imperio, Savino and Avesani, 2005)
9. **JToBI**: model of Tokyo *Japanese*
   (Venditti, 1997; 2005)
10. **KToBI**: model of Seoul *Korean*
    (Jun, 2000; 2005)
11. **PToBI:** model of *Portuguese*
    (Frota, 2000)
12. **MToBI**: model of *Mandarin*
    (Peng, Chan, Tseng, Huang, Lee and Beckman, 2005)
13. **SCToBI**: model of *Serbo-Croatian*
    (Godjevac, 2005)
14. **SpToBI**: model of *Spanish*
    (Beckman, Díaz-Campos, McGory and Morgan, 2002; Face and Prieto 2007; Estebas and Prieto 2010)

These versions of ToBI use the same inventory of labels, combined together to describe a number of prosodic events occurring in different languages. ToBI is considered by far 'one of the best annotation systems available because of the complete correspondence of graphic representation to phonetic reality and to semantics and pragmatics' (Cruttenden 1997: 64). Thus ToBI is a quite versatile system to give account of both global and local characteristics across languages, since it captures recurrent patterns and specific phenomena of different language varieties.

In the following paragraphs (from § 3.2.1 to § 3.2.4), the basic stages in the development of the ToBI annotation system are discussed in order to shed light on the theories proposed and debated for this approach.

3.2.1 Pierrehumbert (1980)

Scholars had discussed and criticized for decades the opportunity of describing intonation as an alternation of rises and falls (British approach). Pierrehumbert and her colleagues addressed this issue by proposing an innovative method for the analysis of melodic patterns conceived as sequences or highs and lows (American approach).

Pierrehumbert (1980) created a system based on a two-levels description: the metrical framework is built in a series of high and low values of F0. Any utterance is labeled as a string of H (high) and L (low). Under this perspective, a system based on a bi-tonal (H and L) description of pitch movements overcomes and solves the problems presented by the arbitrariness of different typologies of rising and falling pitch movements (see the British approach). In fact, as it has been claimed by Bruce (1977: 132), 'reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself'.

According to the model proposed by Pierrehumbert and Hirschberg, intonation should be treated as a compositional combination of pitch accents, phrase accents and boundary tones (Pierrehumbert and Hirschberg, 1990). In her Ph.D. thesis, 'The Phonetics and Phonology of English', Pierrehumbert (1980) clarified the intent of her work: to describe those rules that 'map the underlying [phonological] representations into phonetic realizations' (Pierrehumbert, 1980: 2). She did so by creating a system where:

1. pitch movements are no longer considered as contours but as strings of highs and lows.
2. the analysis of intonation is structured on different components (e.g. pitch accent, phrase accents and boundary tones).

Much of the theories proposed by Pierrehumbert (1980) and re-elaborated by Pierrehumbert and Beckman (1986; 1988) is based on previous works of some other scholars. Pierrehumbert's annotation system clearly draws upon ideas from previous theories on pitch accent (see Bolinger, 1972); metrical structure (see Liberman, 1975); and the two-level pitch phonology (Bruce, 1977).

> 'Specifically, Pierrehumbert's thesis draws together three key ideas and weaves them into a coherent whole that has dominated research on intonation ever since. Those ideas are: the notion of "pitch accent" from Dwight Bolinger, the notion of metrical structure from Mark Liberman, and the notions of the phrase accent and of two-level pitch phonology from Gösta Bruce'. (Ladd, 2007: 2)

Liberman theorized a metrical system 'which assigns metrical patterns to text and tune, establishes a congruence between these patterns in any given case, and specifies possible alignments of the congruent patterns with a metrical grid' (1975: 2). Within this specific interaction between metrical patterns (tunes) and text, the analysis of intonation is carried out through a bi-tonal system based on H and L. In Ladd's words, the original and innovative intuition proposed by Bruce is described as follows,

> 'If we want to capture the linguistically significant phonetic parameters of pitch in our description, we must avoid a syllable-by-syllable segmentation of the overall pitch contour, and avoid talking about the rises and falls of pitch that happen to result from such a segmentation. Instead, we must identify locally significant points in the pitch contour – these are often local minima and maxima, the Highs and Lows of Gösta [Bruce]'s phonological description'.
> (Ladd, 2007: 4).

Pierrehumbert (1980) grasped the intuitions given by Bruce and managed to create a coherent and productive system for prosodic annotation that has been later modified and implemented by several scholars. The model provided by Pierrehumbert presents sequences of targets interpolated by connection lines that are linked together and provide differentiation between strong and weak targets (namely, accented or not accented) in a sentence (Gili Fivela, 2010).

In particular, this annotation standard to label the prosody of English was recognized as a general method to annotate large corpora of speech, in speech recognition and speech synthesis models. Its application to standard American English represents the basis of an innovative transcription system that was later codified by Silverman et al. (1992), called ToBI (see §3.2).

3.2.2 Silverman et al. (1992)

It is worth noting that 'the AM framework became the dominant paradigm in intonational research under the influence of Pierrehumbert (1980) and subsequent work' (Nolan, 2006) of scholars working together at MIT (such as Silverman, Beckman and others). Silverman (1987) wrote his doctoral thesis on the structure and processing of F0 contours. Then, he collaborated with Pierrehumbert and many other scholars to create an innovative approach to the analysis of intonation. Specifically, Silverman and Pierrehumbert (1990) worked in synergy to give account of the surface phonetic realizations of underlying phonological forms. They stated clearly their basic assumptions and presuppositions:

> ‘Speaking English means doing two things at once: saying the words, and saying the melody. The coordination between these two activities is far from arbitrary. Rather, it is determined by the stress pattern and phrasing of the utterance. Some features of the melody fall on certain stressed syllables, while others fall at the boundaries of prosodic phrases’. (Silverman and Pierrehumbert, 1990: 72)

The coordination between words (that are the segmental tier) and melody (that is the tonal tier) is conveyed through the syllable level in the prosodic structure. An example of this level transcription is given in (1) and quoted from Ladefoged (2006: 125).

(1)	[Amelia] - VCVCVV	SEGMENTAL TIER
	/ə m i: l i: ə/	PHONEMIC TIER
	σ σ σ	SYLLABLE TIER
	H* L - L%	TONAL TIER

This sample transcription is aligned on four tiers: segmental, phonemic, syllable and tonal tier. Within the word ‘Amelia’, the third segment [i] corresponds to the third phoneme /i/. In turn, the phoneme /i/ constitutes a syllable that is realized prosodically by the tonal constituent H*. The combination of these tiers connects the tonal representation of intonation to the order of syllables, phonemes and segments.

As a matter of fact, the tonal tier is strictly anchored to the segmental tier (i.e. alternation of consonants and vowels) and consists of a phonetic and a phonological representation, as evidenced in fig. 10.

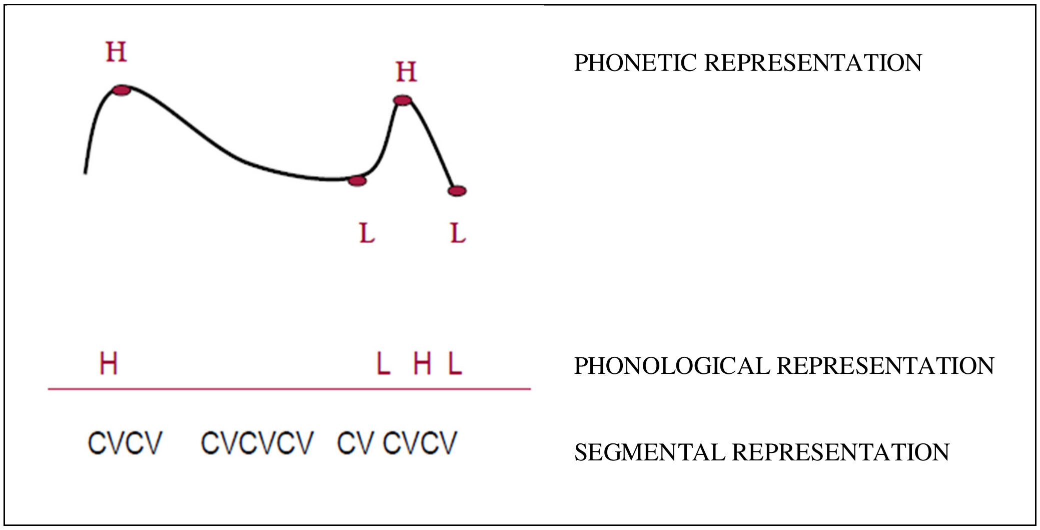


Figure 10. The segmental, phonetic and phonological representations. (adapted from Silverman, 1987; Avesani, 2011).

Fig. 10 illustrates the alternation of two types of representations rhythmically coordinated to the segmental string: the phonetic and phonological representation. At the phonetic level, target landmarks display a link with the F0 values reached at the red points (see fig. 10); at the phonological level, landmarks are considered as abstract and discrete representations of the pitch movements, without any specific reference to the values of F0.

Silverman and his colleagues adapted the model elaborated by Pierrehumbert (1980) and they called it ToBI, a short form for Tones and Break Indices. The name ToBI refers to the two kinds of components of analysis in this method: tonal patterns and breaks (e.g. boundaries) in intonation phrases. Silverman et al. (1992) tailored the ToBI system to the specific needs required by speech recognition and synthesis. Thus, in addition to intonation description, ToBI has applications also to sentence processing in new technologies and clinical studies.

Under this perspective, ToBI constitutes a standard ‘intended to be extensively used with good inter-transcriber agreement and it relays on a machine readable notation in order to favor the sharing of corpora recourses (Gili Fivela, 2010: 10). These objectives were achieved and tested in a study (Silverman et al., 1992) where twenty researchers were asked to transcribe sentences for a total of 20,000 targets. Data confirmed a very high inter-transcriber reliability of notation, thus evidencing and documenting the reliability of ToBI as a system for prosodic transcription of large amounts of speech materials. Good inter-labeler reliability for expert and naive labelers was calculated as follows: 88% agreement on presence/absence of tonal category, 81% agreement on category label, 91% agreement on break indices (Silverman et al., 1992 and Pitrelli et al., 1994). Thus, this system can be considered reliable, coherent and relatively easy to learn.

3.2.3 Beckman et al. (1994, 2005)

Beckman and other colleagues at the MIT started to promote the use of ToBI within the linguistic community by organizing seminars and workshops. They also devised training courses and materials, in which sets of transcribed utterances were provided as examples. In particular, they developed a manual for transcription of large speech corpora called 'The Guidelines for ToBI Labelling' (Beckman and Ayers, 1997), available at (http://www.ling.ohio-state.edu//~tobi/ame_tobi/labelling_guide_v3.pdf).

The ToBI system, originally created for American English (see § 3.2 for Mainsteam ToBI - MaE ToBI), was applied to a number of languages, by reflecting the specific phonetic and phonological characteristics within the intonational and prosodic grammar. Before engaging into a ToBI analysis, some conditions must be realized. First of all, the investigator should record an audio file and visually/auditorily inspect it with the aid of a computer program for speech analysis (such as Praat). Once that F0 have been examined, a transcription of the utterance is given by creating a tone tier for the tonal analysis; a word tier for the orthographic transcription of words; a break index tier to account for the perceived degree of juncture among intonation phrases; a miscellaneous tier to list disfluences and additional comments (such as, breathy or creaky voice).

The creation of a new ToBI version is subordinated to the development of a rigorous analysis of the main intonation phonology phenomena of the targeted language (Beckman et al., 1994, 2005). The set of conventions necessary to create a coherent system of intonational phonology transcription were codified, by the researchers working at the Ohio State University, as a set of principles to be applied to a ToBI version.

The principles of the ToBI systems have been outlined as follows:

Principles of the ToBI system:

1. 'The conventions are as accurate as possible, given the current state of knowledge. Ideally, they will be based on a large and long-established body of research in intonational phonology, dialectology, pragmatics and discourse analysis for the language variety, but at the very least, they are based on a rigorous analysis of the intonational phonology.
2. The conventions do not replace a permanent record of the speech signal with a symbolic record. An electronic recording of the transcribed utterance is an essential component of a complete ToBI framework transcription.
3. The conventions are efficient. They do not waste transcriber time by requiring the transcriber to symbolically mark non-distinctive pitch rises and falls that can be extracted from the signal automatically, or anything else that could be extracted from resources such as online pronunciation dictionaries.
4. The conventions are easy enough to teach that their use is not limited to a few experts to do the transcription. Therefore, there must be a freely available manual for teaching the system to new transcribers, with many recorded examples of transcribed utterances graded from easy to difficult.
5. The conventions are used and maintained consistently across transcription sites. Therefore, in the course of developing a ToBI framework system, there must be rigorous tests of inter-transcriber consistency, and there should be an agreed-upon center for maintaining the standard with periodic rechecks and evaluation of any proposed revisions'.

Table 10. Principles of the ToBI system (from http://www.ling.ohio-state.edu/~tobi/)

In addition to formulating the general rules under which a ToBI system should be devised, Beckman collaborated with Pitrelli and Hirschberg (see Pitrelli et al., 1994) to create an integrated model of description of tonal targets associated to specific meanings and sentence modes (i.e. questions vs. statements). Differences in pitch accents patterns were found to account for the variation of meanings provided by intonation (Pierrehumbert and Hirschberg, 1990). In fact, contours and pitch accents 'convey relationships between the propositional content of utterances and the speaker and hearer's mutual beliefs' (Hirschberg and Ward, 1995: 409). Intonation serves different linguistic and paralinguistic functions that may range from the expression of sentence modality to the marking of emotional and attitudinal nuances. In spoken language, intonation not only helps to express a speaker's attitudes such as involvement, concern, surprise, boredom and so on, but also indicates the distinction between sentence types (Hirschberg, 2002, 2004; Wells, 2006; Grice and Baumann, 2007).

Tone targets realize intonation prominence and convey different messages by modulating F0 height (Terken, 1997; Campbell and Beckman, 1997; Hirschberg, 2002). Pitch patterns have a role in providing syntactic, semantic, and pragmatic meaning (Hirschberg and Avesani, 1997; Hirschberg, 2004; Grice and Baumann, 2007). In fact, every pitch target may be associated to a specific meaning and sentence mode. Pitrelli et al. (1994) proposed the system reported in tab. 11.

Label	**Description**	**Meaning/mode**
H*	Simple high	Canonical declarative
L*	Simple low	Yes-no question
L+H*	Rising to high from low	Contrastive focus
L*+H	'Scooped' late rise	Pragmatic uncertainty
H+!H*	Fall onto stress	Pragmatic inference

Table 11. Pitch accents and their typical meaning/mode. (From Pitrelli et al., 1994)

According to Pitrelli et al. (1994), the tone target H*, described as simple high, is generally used to utter canonical declaratives. H* typically stresses an item that is considered to add new information to the discourse (Pierrehumbert and Hirschberg, 1990). By contrast, yes-no questions are usually introduced by the simple low, L*. This pitch accent characterizes prominent items, phonetically modulated as valleys[6]. As for the fall onto stress, H+!H* indicates pragmatic inference and implied familiarity (Pitrelli et al., 1994).

The notations L+H* and L*+H identify early and late rise. While early rise (i.e. L+H*) is used to realize contrastive focus utterances, late rise (i.e. L+H*) signals uncertainty or incredulity. To explain how these two different pitch accents are produced, I quote an example given in Hirschberg (2004: 5).

[6] For a study on the analysis of L target in English intonation see Erickson et al. (1995).

(1) The Smiths aren't inviting anybody important.

a. They invited **(L+H*)** Loraine.

b. They invited **(L*+H)** Loraine.

While in (a) pitch peak is realized early in the accented syllable, in (b) pitch peak is reached late. This difference has an effect on the general meaning of the two sentences. A L+H* pattern, in (a), presupposes a contrastive effect, as to deny the affirmation previously stated. Thus, this pitch accent is used 'to emphatically contradict the initial claim that Loraine is unimportant' (Hirschberg, 2004: 5). By contrast, the L+H* pattern in (b), 'where the low tone is aligned with the stressed syllable, can convey uncertainty about whether or not Loraine is an important person' (Hirschberg, 2004: 5).

3.2.4 Veilleux at el. (2006)

Intonation contours are described by dividing them into smaller intonation groups (Cruttenden, 1997; Pierrehumbert and Beckman, 1986; Venditti, 2006). Pitch accents, phrase accents and boundary tones are combined together into larger prosodic units called intonation phrases. Within the prosodic hierarchical order, IP is the main prosodic constituent of an utterance (Pierrehumbert and Beckman, 1986; Ladefoged, 2006; Wells, 2006). Every IP is divided into two or more intermediate phrases (iP). Each iP contains one or more pitch accents (L* or H*), followed by a phrase accent (L- or H-). The boundary tone (L% or H%) is the last part of the IP and contains no stressed syllables. The relation of pitch accents, phase accents, and boundary tones with intermediate phrases and intonation phrases has been visually represented in fig. 11.

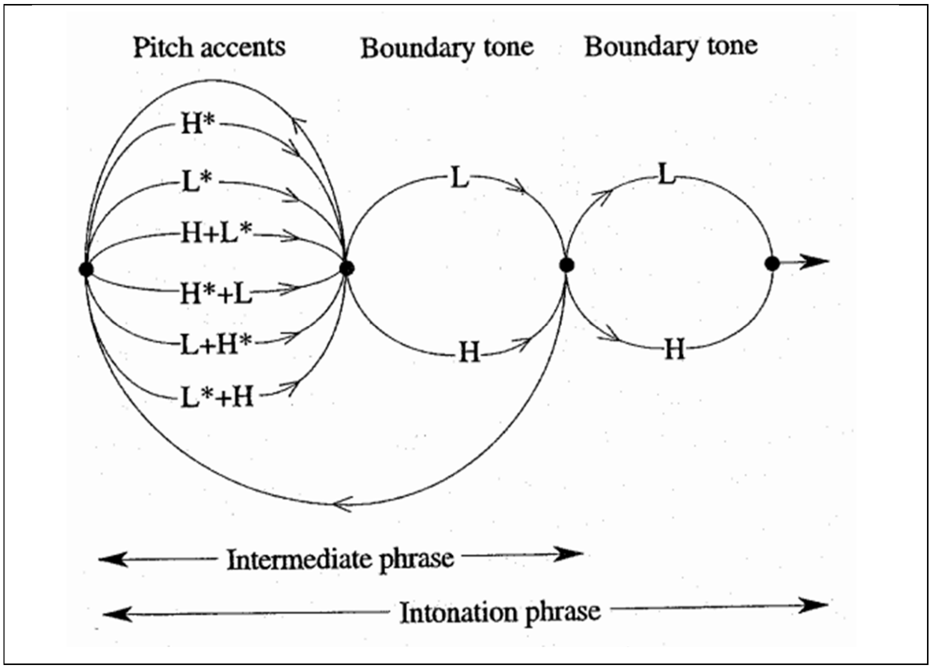


Figure 11. Relation between tonal target (pitch accents, phrase accents, and boundary tones) and phrase constituents divided into intermediate and intonation phrases. (From Liu, 2009: 2).

After about two decades of ToBI being in use, a number of contributions have implemented this system. Tonal descriptions are determined by rules governing the prosodic structure of utterances. One of the more recent tutorials available online is the 'Transcribing Prosodic Structure of Spoken Utterances with ToBI' by Veilleux et al. (2006). This open courseware developed at the Massachusetts Institute of Technology contains audios and transcript examples. To give an idea of the variety of meanings and functions that can be provided by using different intonation pattern, I recall the sample sentence 'Marianna made the marmalade', examined in several studies (among others, Beckman and Ayers, 1997; Harrington, 2008; Veilleux et al., 2006).

Since the prominence and stress allocated in an utterance cannot be predicted just on the basis of grammar (Bolinger, 1972), the same sentence, 'Marianna made the marmalade', uttered in different contexts can provide distinct meanings (Beckman and Elam, 1997; Veilleux et al., 2006; Harrington, 2008). The following examples (1-6) describe some contextual meanings that may be transferred by different tonal productions:

(1) *Marianna* made the *marmalade*.
H* H* L-L%

(2) *Marianna* made the *marmalade*?
L* L* H-H%

In (1) and (2), there are two syllables that are relatively more prominent than any other: the accented syllables in the words ‘Marianna’ and ‘marmalade’. These pitch accents are realized as high tones in (1) and as low tones in (2). This is due to the fact that (1) represents the typical tonal pattern for plain statements, while (2) is a yes-no question. Apart from pitch accents, the distinction between a statement and a yes-no question is given especially by boundary tones. Generally speaking, most of the utterances ending with a rise (H-H%) are perceived as yes-no questions while statements typically end in a fall (L-L%).

In fig. 12, the top panel shows a contour that can be transcribed as H*H* L-L%, see (1), while the bottom panel visualizes the pattern L*L* H-H%, see (2).

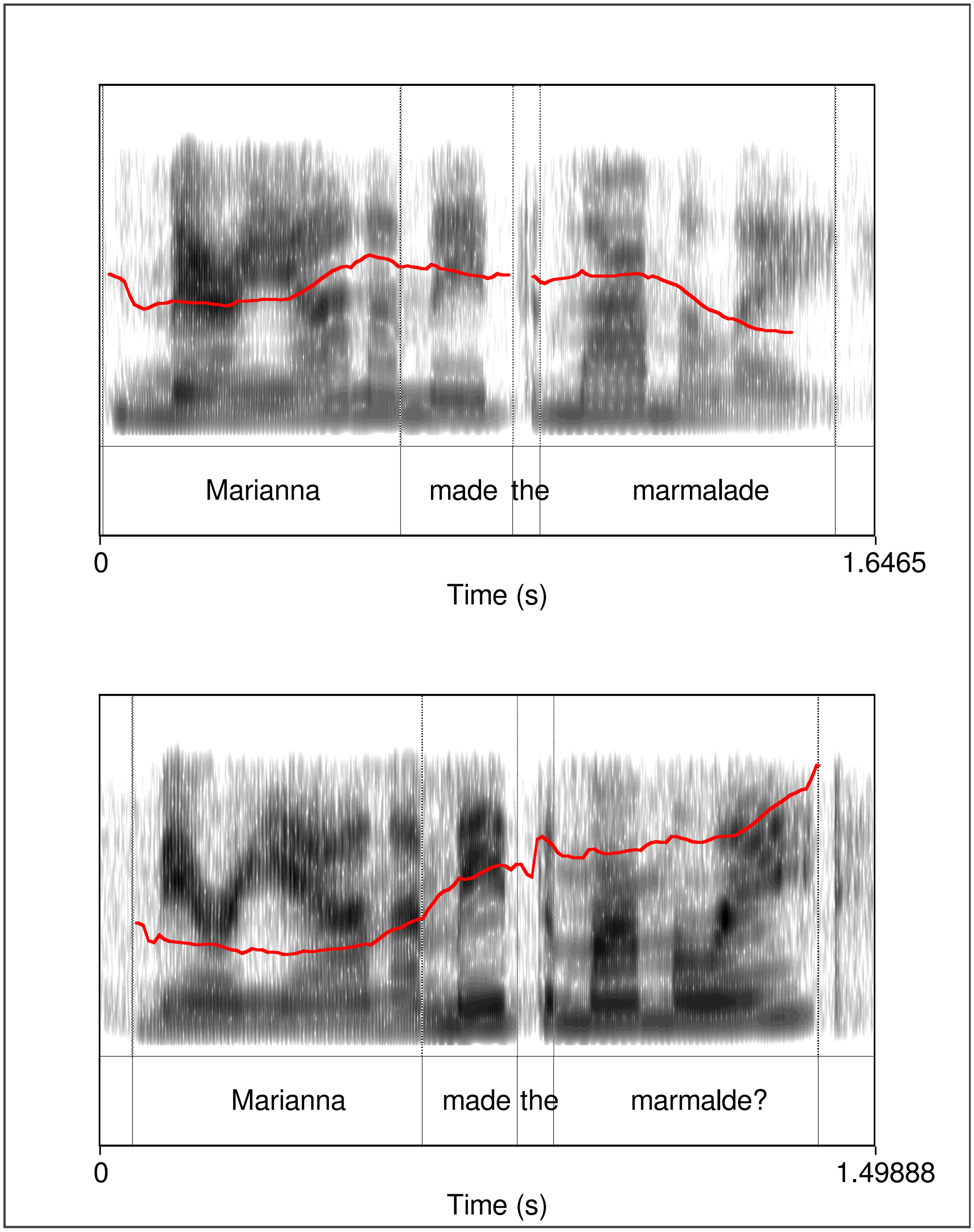


Figure 12. F0 contours for 'Marianna made the marmalade'. The same sentence is uttered in the top panel as a statement (H*H* L-L%,) and in the bottom panel as a yes-no question (L*L* H-H%). The recordings are from Beckman and Ayers (1993), http://www.ling.ohio-state.edu/~tobi/ame_tobi/, and the labels from Veilleux et al. (2006). The pitch visualization is taken from the Ohio State University Archive, http://www.ling.ohio-state.edu/~pwong/ling500/Module2-F0/pitch_animation.ppt.

(3) a. What Marianna made this time?

b. Marianna made the marmalade.

c. Marianna made the *marmalade*? I thought she made a cake.

L* H-H%

(4) a. Who did the marmalade?

b. *Marianna* made the marmalade?

L+H* L-L%

Different transcriptions of an utterance should be annotated depending on the position of the prominent elements. The yes-no question in (2) is transcribed as L*L* H-H% because there are two prominent elements: 'Marianna' (L*) and 'marmalade' (L*). Unlike (2), (3c) and (4b) have only one stressed item. In (3c) the more prominent syllable is in 'marmalade' and 'Marianna' is deaccented. Vice versa, in (4b) the more prominent syllable is in 'Marianna', and 'marmalade' is deaccented. Thus, in the transcription of (2) there are two pitch accents (one marking 'Marianna' and the other marking 'marmalade'), whereas in (3c) and (4b) there is only one pitch accent. This is due to focus reasons. In (3c) the speaker wants to ask his or her interlocutor whether Marianna made the marmalade and not something else, thus stress is given to the word 'marmalade'. In (4b) the speaker is questioning whether it was Marianna or someone else who made the marmalade. So, emphasis is given to 'Marianna' to show that its position is in contrastive focus.

In fig. 13, the top panel shows a contour that can be transcribed as L*L* H-H%, see (2), while the bottom panel visualizes the pattern L* H-H%, see (3c).

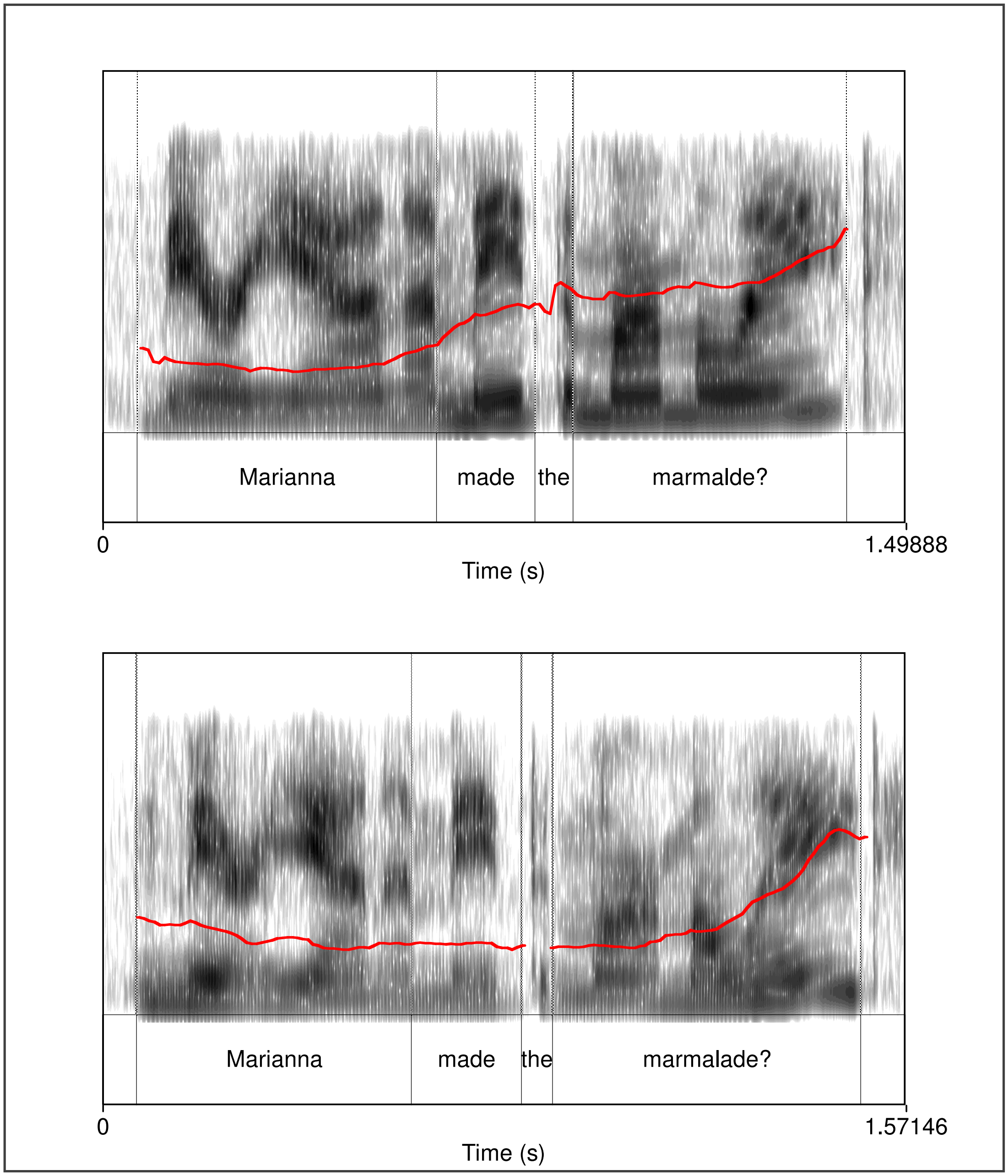


Figure 13. Marianna made the marmalade. The same sentence is uttered in the top panel as a default yes-no question (see example 2) and in the bottom panel with focus on the word marmalade (see example 3c). The recordings are from Beckman and Ayers (1993), http://www.ling.ohio-state.edu/~tobi/ame_tobi/, and the labels from Veilleux et al. (2006). The pitch visualization is taken from the Ohio State University Archive, http://www.ling.ohio-state.edu/~pwong/ling500/Module2-F0/pitch_animation.ppt.

In the following examples (5-7), the word examined is ‘Marianna’. Different tonal patterns can express emphasis, incredulity, surprise and other emotions.

(5) *Marianna*, are you ready to leave?
H* H-L%

(6) a. Marianna made some great marmalade.
b. *Marianna* made the marmalade? I thought she was allergic to
L+H* L-H% strawberries.]

(7) *Marianna*! What are you waiting for?
L+H* L-L%

In (5) ‘Marianna’ is pronounced as a H* H-L% tonal string. This pattern is typically used when speakers address someone and they want to raise the attention of this person and they call his or her name aloud. The intonation contour used in this case has been referred to as the ‘sustained’ calling contour (Ladd, 1996). In (6b) ‘Marianna’ is placed in a focused position. By using this tune, a speaker may convey an effect of surprise and incredulity. In fact, there is a contrast emerging from the meaning of (6a) and (6b). In (6a), a speaker asserts that ‘Marianna made some great marmalade’. The speaker in (6b) shows astonishment about the content of the message conveyed by the other interlocutor (in 6a) by expressing the idea that Marianna is believed to be allergic. In order to convey this idea of surprise and incredulity, the rise-fall on the stressed syllable is followed by a final rise. In (7), ‘Marianna’ is uttered as an exclamation. The tonal pattern consists of a sharp rise on the stressed syllable followed by a fall. This pattern is typically used when the speaker wants to show a strong reaction and calls for someone’s attention. In some contexts, it can conveys an irritated tone and a reprimanding intent.

In addition to the tonal patterns described in (5-7), some of the most recurrent patterns are exemplified by a drawing of the pitch contour and their tonal transcription in tab. 12 (see Harrington, 2008). Following the literature on tonal strings (Ladefoged, 2006; Veilleux et al., 2006; Halliday and Greaves, 2008; O’Connor, 2008; Truckenbrodt, 2012) I included in the right part of the table a brief description of the tonal meanings generally associated with these patterns:

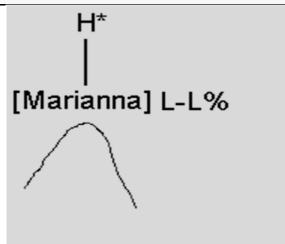
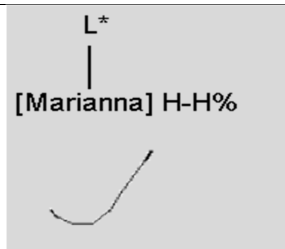
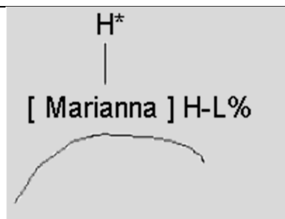
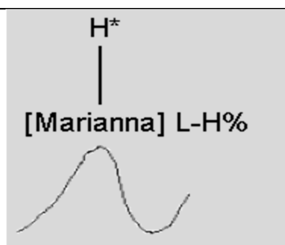
Pitch contour	Description
	Simple statement in response to a plain question. This tune is used by default in statements and is somehow a neutral contour to provide information and explanations about something.
	Plain question with a tune that does not convey an attitude. It is used mainly in yes-no questions to ask for information or permission.
	Tune indicating that the speaker is addressing Marianna. It conveys the idea that it is her turn to speak. This tune is typically used also to give suggestions or enumerate lists.
	Tune expressed by a rise fall. It usually conveys an idea of surprise or incredulity. It expresses uncertainty. When the pitch accent expresses focus position (L+H*) it conveys an idea of emphasis and contrast.

Table 12. 'Marianna'. The same word is uttered with four different tunes to convey different meanings in context. The drawings are from Harrington (2008), and the interpretations are based on Ladefoged (2006), Veilleux et al. (2006), Halliday and Greaves (2008), O'Connor (2008), and Truckenbrodt (2012).

Any pattern described in the previous examples and many others have been carefully analyzed and decoded within the ToBI annotation system. Fig. 14 shows all legal tonal strings in American English (see Veilleux et al., 2006). This schema is a straight-line approximation of possible pitch contours consisting of a combination of pitch accents, phrase accents and boundary tones. It is a useful resource to limit doubtful situations when transcribing utterances with ToBI and to provide and overall idea of how many alternative patterns can be used in an utterance.

Figure 14. Straight-line approximation of possible pitch contours consisting of a combination of pitch accents, phrase accents and boundary tones. From Veilleux et al. (2006), freely downloadable at: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/lecture-notes/tobi_tails.pdf.

3.3 Review of previous pitch range analyses across different languages

A fairly big number of studies have been carried out about intonation and pitch variation across languages. However, there is a reported lack of studies that provide systematically comparable and large-scale data. In fact, it may happen that findings across different experiments are not consistent with other studies carried out for the same languages. This is probably due to methodological differences across investigators and not only to language differences of pitch variation (Guimarães, 2007). Pitch variation basically depends on the language spoken, gender and physical characteristics of the speakers, and socio-cultural contexts.

Pitch range is expected to be a robust phenomenon that plays a major role in the description of prosody. In fact, pitch contours are fundamental to discriminate among distinct melodic patterns across languages. For example, qualitative and quantitative pitch differences relative to the patterns of yes-no questions, wh-questions and statements have been observed in a number of studies and languages. In particular, intonation movements clearly capture variation in pitch range across sentence-types.

> ‘Among the parameters than seem more susceptible to dialectal variation are the recurrent patterns typical of unemphatic declaratives, the presence or absence of a final high pitch on yes-no questions as well as the presence or absence of a final high pitch on unemphatic declaratives’ (Hirst and Di Cristo, 1998: 40).

In the following paragraphs, I will briefly review previous studies on pitch range and emphasize similarities and differences across language types. This survey, based on Romance and Germanic languages, does not propose a generally-agreed-upon description of pitch range — as it is not possible to draw universal conclusions based on studies carried out with different methods and data. However, the most important tendencies will be outlined.

3.3.1 Italian and other Romance languages

- **Italian** (Bertinetto, 1981; Grice, 1995; Rossi, 1998; Farnetani and Zmarich, 1997; Schmidt, 1999; D'Imperio, 2002; Sorianello, 2006; Savino and Grice, 2007; Mennen et al., 2010; Busà and Urbani, 2011; Niebuhr et al., 2011).

Italian makes use of intonation patterns to signal discourse information structure and prominence (Busà, 2008). It is a free-stress language that uses duration, F0 and intensity as the main stress cue (Bertinetto, 1980). Dialectal differences in rhythmic structure have been reported for Italian. 'Italian has been classified as syllable-timed, like French, but southern varieties tend towards stress-timing' (Grice, et al., 2004). Most of Italian varieties are syllable-timed, along with other Romance languages (Bertinetto, 1981; Vayra et al., 1984; Rossi, 1998; Farnetani and Busà, 1999; Ramus et al., 1999) and stressed syllables are generally the penultimate syllables (Rossi, 1998; D'Imperio and Rosenthal, 1999; Lepschy and Lepschy, 1977). The main syntactic tendency is the ordering of SVO- subject, verb and object positions (Rossano, 2010), though different word orders (e.g. dislocations and topicalizations) are possible.

The use of intonation to distinguish among sentence types is crucial, given the fact that Italian does not make an obligatory use of any morphological or syntactic mean to distinguish yes-no questions from statements (Avesani, 1990; D'Imperio, 2002; Bertinetto e Loporcaro, 2005). Generally speaking, the two main contours realized in the so-called Standard Italian are: falling movements for statements and rising movements for questions (Lepschy, 1978). Despite this clear opposition between patterns for statements and questions, the uses of intonation patterns are far more complex and articulated across regional varieties.

Italian has not been studied systematically. More research has been conducted on Central and Southern Italian varieties than on the Northern ones. Intonation research on the Northern Italian varieties is often still at a preliminary stage. Some of the Italian regional varieties that have been analyzed so far are:

1. *Northern*:

   Milan (Marotta and Sardelli, 2006); Parma (Felloni, 2010, 2011); Bologna (De Dominicis, 2005); Torino (Inerlandi et al., 2007); Veneto (Busà and Urbani, 2011;

Busà and Stella, 2012); Treviso (Payne, 2005); Padua (Magno Caldognetto et al., 1978; Busà and Stella, 2012).

2. *Central*:

   Florence (Magno Caldognetto et al., 1978; Avesani, 1995; Firenzuoli, 2000); Pisa (Sardelli, 1998; Gili Fivela, 2002; Savino et al., 2006); Siena (Sorianello, 2006); Lucca (Marotta, 2000); Rome (De Dominicis et al., 2006).

3. *Southern*:

   Palermo (Grice, 1995); Catanzaro (Marotta and Sardelli, 2006); Cosenza (Sorianello, 2001); Naples (D'Imperio and House, 1997; D'Imperio, 2000, 2002, 2004; Crocco, 2006); Bari (Grice and Savino, 1995; Savino et al., 2006); Lecce (Romano, 1997, 2001; Stella and Fivela, 2009) .

In the study of these Italian varieties, it has been shown that it is possible to find common traits among pitch patterns across sentence types. Thus, despite the lack of a widely recognized intonation system of Standard Italian, the vast majority of the Italian varieties share common features.

In most of Italian yes-no questions, the typical intonation contour consists of a final rise on the last syllable (Magno Caldogneto et al., 1978; Kori and Farnetani, 1983; Canepari, 1992; Avesani; 1990, 2005; D'Imperio, 2002; Giordano, 2006). However, it has been noted that, in many Italian varieties, the movement is a rise-fall rather than a rise (Canepari, 1992)[7]. This interesting difference across Italian regional varieties has been emphasized also by D'Imperio (2002: 38) who observed that 'while most Northern and Central varieties seem to be characterized by a terminal rise (i.e., a low nuclear accent followed by a rising phrase accent), Southern varieties exhibit a local rise on the nuclear accented syllable followed by a later fall'. The use of final rise on the last syllable is attested in the Milano, Padova, Liguria, Emilia, Sardegna and Northern Salento varieties (Endo and Bertinetto, 1997; Contini, 1971, 1983; Schirru, 1982; Rossi, 1998). By contrast, the movement of rise-fall has been observed in Southern Italian varieties, such as Bari, Napoli, Palermo, Cosenza and Southern Salento (Maturi, 1988; Caputo, 1994; Grice

[7] Canepari (1992: 204) described the typical intonation pattern of yes-no questions by affirming that in many widespread areas in Italy, the typical movement is not rising but rise-falling. The original version in Italian 'in parecchie zone sparse d'Italia, il movimento tipico non è ascendente, ma ascendente-discendente'.

and Savino, 1995, Grice, 1995; Sorianello, 2001, Romano, 2001). Indeed, the description of question contours depends on the regional variety considered.

	Question tune	
	Nuclear accent	Phrasal tones
Neapolitan	L*+H	HL- L%
Bari	L+H*	L-L% or L-H%
Palermo	L*+H	L-L% or L-H%
Florentine	H*	L-H%

Table 13. Typical intonation patterns in yes-no questions across Italian varieties. (From Grice, D'Imperio, Savino and Avesani, 2005: 367).

As shown in tab. 13, nuclear accents in yes-no questions are usually produced by sequences of rises (e.g. L*+H, L+H*, H*). Also phrasal and boundary tones are produced with rises (e.g. L-H% and H-H%) with the exception of Southern varieties (see Neapolitan, Bari and Palermo). For example, Grice (1995) analyzed the rise-fall accentual pattern of yes-no questions in Palermo Italian and she found that the final fall is, by far, the most common pattern in yes-no questions. In Bari and Palermo Italian, both statements and questions may end in a fall from a mid or high pitch. According to Grice (1995), the discriminant between statement and question status is given by the pitch level immediately before the starting point of the falling contour. When it is relatively high, the utterance is a statement; when it is relatively low, the utterance is a question. In sum, in Bari and Palermo Italian, the final fall is used 'by default' in yes-no questions while the final rise is considered as an 'optional stylistic variant' (Grice et al., 2004: 367).

Most of Italian wh-questions are not identified by a typical contour which works as a universally agreed intonation pattern describing the pitch contours of this kind of sentence. Usually the wh-word is characterized by a local rise-fall. The intonation contours of wh-questions depend on the prominence and the syntactic role of the wh-elements (Poletto and Pollock, 2000). However, little is known about the nature of the final boundary tone in wh-questions. In agreement with previous studies (see Magno Caldognetto, Ferrero, Lavagoli and Vagges, 1978; Busà and Urbani, 2011) speakers of Northern varieties of Italian have been found to pronounce both yes-no questions and wh-questions with terminal rises. A possible explanation of this phenomenon could be that, given the lack of any morphological and syntactic means to identify questions in Italian, it is the rising contour that has the crucial function of signaling questions (D'Imperio, 2007). This could be the case of Northern varieties of Italian whose yes-no questions and wh-questions are both characterized by rising contours. Thus, even though a standard contour for wh-questions has not been defined yet, the most common pattern is identified with an initial rise-fall on the wh-element, followed by a gradual fall along the sentence.

As far as statements are concerned, the patterns which most Italian speakers use by default is the high/low fall. Plain statements are usually uttered with falling contours. However, some problems arise when comparing broad vs. narrow focus. Broad focus sentences have been compared among four Italian varieties (i.e. Neapolitan, Bari, Palermo and Florentine Italian). Each variety uses a H+L* pattern for sentences in broad-focus condition (Grice et al., 2004). By contrast, the narrow-focus condition is realized with a variety of patterns. In Neapolitan and Florentine, one specific pitch pattern is used to signal broad-focus and another one to signal narrow-focus, while in Bari and Palermo, the same pattern is produced with the same pitch pattern to produce both broad and narrow focus (Grice et al., 2004). This has caused a number of interpretations about the ways in which such a phenomenon could be interpreted. Despite these differences between broad vs. narrow focus, every Italian variety is reported to have final falls signaled by L% in declarative sentences (Grice et al, 2004; Sorianello, 2006; Busà and Urbani, 2011; Busà and Stella, 2012). Thus, the final fall in the intonation contour is a mark for statements.

- **Spanish** (Llisterri et al., 2002, 2003; Prieto and Shih, 1996; Prieto, 2006, 2010; Ramirez Verdugo, 2005)

Spanish intonation patterns have been investigated by focusing the attention mainly on three aspects: (1) the interaction between narrow/broad focus and pitch range, (2) the relation between pitch range and expressions of politeness, (3) the role of pitch in the phonetic realization of lexical stress.

Within the domain of focus analysis, pitch range has been found to be a strong correlate for narrow and broad foci. In particular, narrow-focused sentences are characterized by a substantial increase in pitch range, as compared to the broad-focused sentences. For example, Prieto asserted that the speakers in her study were likely to conform to a similar tendency: they produced 'pitch accents with larger pitch ranges in narrow focus than in broad focus' (2006: 7).

In several studies on raising and falling Spanish contours (Prieto, 2011; Prieto and Shih, 1996; Ramirez Verdugo, 2012), it has been found that an increase in the pitch range of the final part of the utterance tone corresponds to a decrease of perceived politeness. By contrast, a decreasing of the pitch range was shown to have no effect in the perception of politeness. By analyzing prosodic cues, such as pitch range, in interaction with facial gestures, Prieto (2011: 841) has come to the conclusion that ‘there is nothing intrinsically polite about using an increased pitch range, unless it is accompanied by consistent contextual information [that is gestural information]’. Thus, when assessing the relevance of pitch in the perception of an utterance, attention has to be paid to various aspects of gestural information. Despite the fact that Spanish intonation patterns are fairly similar to the English ones, it has been shown that, in the Spanish subjects, there is a preference for falling contours over falling-rising contours to express uncertainty (Ramirez-Verdugo, 2005). What is more, phonetic analysis reveals that the vast majority of falling contours produced by Spanish native speakers are characterized by a narrower pitch range, as compared to the pitch range used by English non-native speakers (Ramirez-Verdugo, 2005).

Together with amplitude and duration, pitch is considered to be one of the main correlates of lexical stress (Llisterri et al., 2002; Schwab and Llisterri, 2012). Specifically, ‘F0 is the main parameter systematically related to the identification of the stressed syllable of a word, while duration is a secondary cue, also conditioned by the stress pattern’ (Llisterri et al., 2002). Thus, F0 range and duration are highly relevant parameters for the identification of lexical stress. The role of pitch has been assessed in the realization of stress placement, by analyzing words in isolation and carrier sentences. The results obtained have shown no systematic effects of the context (isolated words vs. words in the carrier sentences) on the perception of pitch range variation (Llisterri et al., 2002).

In addition to standard Castilian Spanish (Estebas-Vilaplana and Prieto, 2010), a number of regional varieties have been examined in different production and perception experiments. Some of the Spanish varieties examined from a intonation perspective are: Argentinian Spanish (Gabriel et al., 2010), Canarian Spanish (Cabrera Abreu and Vizcaìno Ortegaire, 2010), Cantabrian Spanish (López-Bobo and Cuevas-Alonso, 2010), Chilean Spanish (Ortiz et al., 2010), Dominican Spanish (Willis, 2010), Ecuadorian Andean Spanish (O’Rourke, 2010), Mexican Spanish (Prieto et al., 2007; De la Mota et al. 2010), Puerto Rican Spanish (Armstrong, 2010a, 2010b) and Venezuelan Andean

Spanish (Astruc et al., 2010). These studies on different varieties of Spanish aim at comparing data collected with similar methods and analyzed with Sp_ToBI (the ToBI version for Spanish). Several research projects on intonation and pitch range are available on the webpage of the Group of Prosodic Studies, also called GrEP (see online materials at http://prosodia.upf.edu/home/en/index.php).

Within Spanish varieties, especially Mexican Spanish has been the object of a deep study on the nature of intonation contours across different sentences types: yes-no questions, wh-questions and statements. Yes-no questions are reported to have a typical high-rise contour (Avila, 2003; De la Mota et al., 2010). As for wh-questions, different pitch contours, such as rising, falling and rising-falling patterns, have all been attested in different studies (Prieto et al., 1996, 2009-2012; De la Mota, 2010). In Mexican Spanish, the most common pitch pattern for wh- questions is the so-called circumflex contour L+H* H-L%. The circumflex contour configuration consists of 'a F0 rise associated with the last stressed syllable which continues during the onset of the following syllable. After the peak, the pitch falls to a low that is realized a bit higher than the initial low' (De la Mota, 2010: 335). Contrary to questions, statements are generally produced with low falls; this pattern is the standard configuration for broad focus statements across most varieties of Spanish.

- **Catalan** (Prieto, 1995, 2008, 2011, 2013; Estebas-Vilaplana, 2000; Vanrell, 2006; Borràs-Comes, 2012)

Catalan displays pitch patterns that are significantly different from other Romance languages such as Spanish, Italian, French and Portuguese. In particular, yes-no questions have a peculiar falling contour which is not found in any other Romance language (Prieto et al., 2005; Borràs-Comes et al. 2010).

Prieto described this phenomenon as being very uncommon: 'contrary to the dominant cross-linguistic pattern, most Catalan dialects have a falling yes-no question with a low boundary tone' (Prieto, 2013: 18). Even though both yes-no questions and statements are commonly uttered with falling contours, these contours are realized in two different ways. Yes-no questions are uttered with a high plateau in the pre-nuclear part of the contour while statements do not have a plateau, because pitch falls immediately after the rise (Estebas-Vilaplana, 2000; Prieto, 1995, 2008).

Vanrell (2006) analyzed the falling contours in yes-no questions and wh-questions in Catalan. It was found that the F0max of the rise gradually lowers as the sentence length increases. In particular, a high tone H* is placed on the accented wh-word and the falling contour continues with a descending pattern until the final boundary tone in the utterance (Pilar, 2013). In a study on Majorcan Catalan, Vanrell (2008) tested the hypothesis that the pitch accent choice in wh-questions may be affected by different functions of focus. With a rising pitch accent, Catalan speakers probably express the intention of bringing the wh-word into focus.

The role of pitch range in establishing intonational contrasts in Catalan sentences was extensively analyzed in a study by Borràs-Comes et al. (2010). It was shown that a gradual scaling in the height of the rising pitch accent L+H* may signal different sentence types: statements and echo questions (i.e. questions where speakers challenge what has been said). Results demonstrate that the degree of more or less narrow/wide pitch span corresponds to different sentence types. Thus, under certain conditions, different sentences are believed to occupy specific pitch range areas. In Catalan,

> 'there is a categorical difference between the statement interpretation and the echo question interpretation. In this case, the pitch range variation is the main cue that Catalan listeners use to decide between one and the other. This fact represents further evidence that pitch range can be used to make phonological distinctions between a variety of pragmatic meanings, and that this needs to be represented descriptively at the phonological level' (Borràs-Comes et al., 2010: 4).

Borràs-Comes et al. (2010) assume that pitch range should no longer be considered as a merely phonetic factor but also as a phonological marker that can be represented at the phonological level. However, to date, a method on the systematic investigation of phonological aspects of pitch range has not been proposed yet. In the following studies, I will try to consider how pitch range can be revisited and theorized under a phonological perspective.

- **French** (Mertens, 1993; Di Cristo, 1998; Jun and Fougeron, 2000, 2002; Welby, 2003, 2006; Vion and Colas, 2006; German and D'Imperio, 2010; Le Gac et al., 2012, Schwab and Llisterri, 2012)

French intonation is described as having a sequence of rising pitch movements that demarcate phrase boundaries (Jun and Fougeron, 2000). Unlike other Romance languages, French is characterized by sequences of intonation units (called intonation phrases) marked by rising intonation patterns. These intonation phrases are 'obligatorily marked by a prominent F0 rise near the end of the phrase, and optionally include an additional F0 rise near the beginning of the phrase' (German and D'Imperio, 2010: 1). This is probably due to the fact that, in French, stress does not have a distinctive function and it is primarily signaled by duration rather than F0 (Jun and Fougeron, 2000; Schwab and Llisterri, 2012). In particular, 'the location of stress is fixed at the word level, but its realization depends upon the position of a word within a phrase' (Jun and Fougeron, 2002: 147). By contrast, in Romance languages such as Spanish and Italian, stress is the result of changes in F0 and duration.

Several scholars have attempted to shed light on F0 variation in French, especially across question types. For example, Di Cristo (1998) distinguishes among the patterns used in yes-no questions and wh-questions. Yes-no questions typically have pitch contours corresponding to two functions: confirmation and information. Yes-no questions uttered to confirm something are characterized by a final fall while yes-no questions providing some kind of information are usually uttered with an overall rising pitch with a final rise on the last stressed syllable of the question. However, it has been claimed by Di Cristo (1998) that French yes-no questions are characterized by a final rise when the speaker seeks information and not confirmation (in such a case utterances are marked by final fall). As far as wh-questions are concerned, they are marked by a final fall. Thus, wh-questions exhibit final falling contours which are similar to those of statements.

Welby (2003) examined the structure of French intonational rises in her Ph.D. thesis. The major focus of her investigation is the so-called 'early rise', a typical pattern in French. Intonation contours are marked in French by 'a late rise, an obligatory fundamental frequency rise at the end of a phrase that is not the last phrase of the utterance, and an early rise, an optional rise near the beginning' (Welby, 2003: 225). In fact, rises can assume two different positions, depending on the alignment of the rise with the prominent syllable: early or late rise. What is more, unlike other Romance languages, French rises may be placed on stressed or even unstressed syllables. Thus, French intonation is characterized by a different tonal alignment in rising tunes that systematically operates as boundary tones (late rises) or speech segmentation (early rises). In fact, for both rises 'only one end point is anchored to a segmental landmark (the L beginning of the early rise; the H end of the late rise)' (Welby, 2006). In a way, these structurally different rising patterns are combined together to convey an effect of segmentation of speech into chunks, separated by boundaries.

Recently, Le Gac et al. (2012) reported the results of an experiment testing F0 variation across questions manipulated at the end of intonation phrases. The aim of this study was to establish whether F0 variation can be categorical or gradient at the end of final question contours in French. After manipulating the final F0 boundary tone of several questions (with a variation in F0 span of 1 ST per sentence), the principle of categorical vs. continuous perception was used to test the data. So, listeners were asked to do an identification task in order to test the categorical distinction of F0 variation in French sentences.

In their study, Le Gac et al. (2012) implemented the results of an experiment proposed by Post (2000) by providing evidence on the existence of at least three distinct pitch categories. Data suggest that pitch range variation is generally gradient rather than categorical. Yet, the authors argue that data show this effect due to the imitation task attempted in their experiment. The conclusion that can be drawn from this investigation is that a model of pitch variation in French should be based on the categorical distinction of at least three-way opposition. Thus, according to Grabe et al. (2012), the Auto-segmental Model and other models based on a bi-directional opposition between H and L tones fails to capture pitch variation in French.

- **European Portuguese** (Cruz-Ferreira, 1983; Cruz-Ferreira, 1987; Vigario and Frota, 2003; Guimarãs and Gouveia, 2007) and **Brazilian Portuguese** (Morais, 1998; Frota and Vigário, 2001).
-

Northern varieties of European Portuguese are believed to be more conservative than the Standard (Lisbon) Portuguese, with some similarities with Spanish (Vigario and Frota, 2003). Several distinctions across intonation patterns have been made on the basis of sentence types: yes-no questions vs. wh-questions vs. statements. The intonation contour of yes-no questions is characterized by fall-rise melody where a fall through the nuclear syllable is followed by a steep rise in the final syllable (Cruz-Ferreira, 1998; Frota, 2002). Patterns of wh-questions typically follow the standard trend of declaratives, by realizing sharp final fall through the last stressed syllable of a sentence (Cruz-Ferreira, 1998; Frota, 2002). The standard intonation contour for statements consists of an initial rise, a plateau and a sharp final fall (Cruz-Ferreira, 1998; Frota, 2002, 2003; Vigario and Frota, 2003).

In a study on European Portuguese by Cruz- Ferreira (1983), it was shown that Portuguese learners of English and English learners of Portuguese successfully managed to disambiguate meaning among sentences, depending on specific intonation patterns. In particular, speakers were likely to properly match the sentences with their meaning when the intonation patterns in L1 and L2 were similar. It has been claimed that sentences are correctly interpreted by developing 'general intuitions about the more likely meanings associated with lower and higher pitch' (Cruz-Ferreira, 1987: 116). This strategy has to do with the speakers' ability to identify pitch height.

Pitch range in European Portuguese has been analyzed in a study of two regional varieties of Portuguese (Lisbon and Funchal) by Guimarães and Abberton (2005). Portuguese females aged between 19 and 40 years old had a F0mean of 196.9 Hz, while older females, who were 41-67 years old, had a F0mean of 177.5 Hz. In addition, it was found that the F0mean for females from Lisbon was 199.4 Hz and for females from Funchal was 198.4 Hz. Thus, results highlighted an almost perfect match for F0mean of the Lisbon and Funchal Portuguese varieties.

Studies on Brazilian Portuguese have pinpointed pitch patterns that are different from the standard patterns found in European Portuguese. Morais (1998), and Frota and Vigário (2000, 2001) analyzed pitch variation in a comparative study of European and Brazilian Portuguese. Data collected in their studies show that, unlike Standard European Portuguese, Northern European Portuguese and Brazilian Portuguese share common phonological and rhythmic structures: 1) realization of prominence on lexically unstressed syllables, 2) presence of rhythmic stress to the left of the lexically stressed syllables, 3) higher density of pitch accents than in Standard European Portuguese.

Standard European Portuguese seems to be characterized by less tonal events than the Northern regional varieties. This peculiarity has been described by Frota (2002: 33) who asserted that the 'sparseness of tonal events within the pre-nuclear stretch is a distinctive feature of European Portuguese intonation that sets this language apart from other Romance languages, such as Italian, Spanish, and Brazilian Portuguese.'

3.3.2 English and other Germanic languages

- **American English** (Pierrehumbert, 1979; Ohala, 1983; Clark, 1999; Aoyama and Guion, 2007; Dilley and Brown, 2007; Dilley, 2010; Hedberg and Sosa, 2011)

Historically, the focus of most research on intonation and pitch analysis has been based on American English. Since the AM phonological and prosodic theory (Lieberman, 1975; Bruce, 1977; Pierrehumbert, 1980) was devised and developed on the basis of the Standard American English, this implies that Standard American is by far the most documented language. The prosodic research remains largely focused on tonal transcription and interpretation (see ToBI). In the United States, many researchers working on intonation in the framework of the AM approach[8] collaborated to large ambitious projects, such as the conventions set for the Mainstream American English, (Silverman et al., 1992; Pitrelli et al., 1994) and smaller case-specific studies.

American English varieties have been traditionally diversified and separately treated on the basis of their vowel inventory and system (Labov, 1998; Labov et al., 2006). Unlike standard American English, Southern varieties of American English are traditionally stereotyped as been characterized by slower speech rate and larger pitch excursions (Feagin, 1997). Arvaniti and Garding (2007) investigated two American English varieties: Minnesota and Southern California English. They found that the two varieties used different inventories of pitch accents: H* vs. L+H* vs. L*+H. In Southern California English, these pitch accents are perceived as distinct while, in Minnesota English, speakers do not distinguish between H* and L+H*. What is more, in bitonal pitch accents (i.e. L+H* and L*+H), F0 peak alignment is displayed later in Southern California English as compared to Minnesota English.

[8] Collaborative projects on the prosodic annotation of standard American English have been developed at the Ohio State University (http://www.ling.ohio-state.edu/~tobi/) and at the 14th International Congress of Phonetic Science, in the Models and Transcription Workshop organized in San Francisco in 1999. Some of the largest speech corpora annotated with the MAE_ToBI include the BU FM Radio news database created by teams at the Boston University and MIT (Ross and Ostendorf, 1994) and the MAGIC corpus developed at the Columbia University (Pan and McKewown, 1999).

As for F0 analysis, the most influential studies conducted on American English varieties have been outlined in Henton (1989: 303) and are shown in Appendix B. F0 values were calculated for groups of subjects having different size, age, gender and speaking different American regional varieties. In addition to these data, tonal patterns have been studied across sentence types: yes-no questions vs. wh-questions vs. statements. Bolinger (1972, 1986, 1989) distinguished among the typical pitch patterns used in different sentence modes. While yes-no questions are usually uttered with a final rise, wh-questions and statements are produced with a final fall. In comparison to wh-questions, yes-no questions are considered to have 'a more demanding nature' (Bolinger, 1989). This could be the reason why yes-no questions are typically uttered with rising patterns.

The rising nature of yes-no questions has been tested in number of studies (Brown, 1980; Thompson, 1995; Levis, 1999; Hedberg et al., 2008). In agreement with previous literature, Hedberg and Sosa (2011) reported that American English yes-no questions tend to be low rising and they are characterized by a low pitch accent followed by a high boundary tone. In their study on 410 American English yes-no questions, they calculated occurrences of tonal patterns, as shown in tab. 14.

Nuclear contour	**ToBI category**	**Number**
Low rise	L*H-H% L*L-H%	325 2
High rise	H*H-H% !H*H-H%	39 5
High fall	H*L-L% !H*L-L% L+H*L-L%	9 1 7
Low fall	L*L-L%	6
Fall rise	H*L-H% L+H*L-H%	1 1
Level	H*H-L% !H*H-L% L*+HH-L%	9 3 2
Total		**410**

Table 14. Occurrences of American English yes-no questions tonal patterns calculated over 410 yes-no questions. (From Hedberg and Sosa, 2011: 847).

Tab. 14 shows that most yes-no questions were utterances with rising contours: 327 yes-no questions were produced with low rises, 44 with a high rise, and 2 with a fall rise. In total, 373 over 410 yes-no questions were uttered with a final rise. Other tonal patterns, such as the fall, low fall, and level contour, were registered for yes-no questions. Indeed, yes-no questions are largely associated to final rising tonal patterns, as opposed to wh-questions and statements.

Statements and wh-questions are uttered by default with falling patterns. However, two of the most distinguishing features of wh-questions, as opposed to declaratives, are the presence of a prominent rise-fall on the wh element and a steeper declination line (Waibel, 1979). Wh-questions have been extensively examined by Herdberg et al. (2010) in an experiment conducted over 200 wh-questions. In tab. 15, results on the number of occurrences of different tonal patterns are reported:

Nucleus	ToBI Category	Number
High Fall	H*LL% !H*LL%	64 34
Rise Fall	L+H*LL% L+!H*LL% L+¡H*LL% ¡L+H*LL%	42 6 1 1
Low Fall	L*LL%	14
Low Rise	L*HH% L*H¡H% L*LH%	25 1 1
High Rise	H*HH% !H*HH%	3 1
Fall Rise	H*LH%	2
Rise-Fall-Rise	L+H*LH% L+!H*LH%	2 1
Level	H*HL% H*!HL%	1 1
Total		200

Table 15. Occurrences of American English wh-questions tonal patterns calculated over 200 wh-questions. (From Hedberg et al., 2010: 2).

Tab. 15 shows that most of the analyzed wh-questions were utterances with falling contours: 98 wh-questions were produced with a high fall, 50 with a rise-fall, and 14 with a low fall. In total, 182 over 200 wh-questions were uttered with a final fall. Other tonal patterns, such as the low rise, high rise, fall rise, rise-fall-rise, and level contour, were registered for wh- questions. Data show that the falling contour occurred 81% of the time and the rising contour only 18% of the time. Thus, wh-questions in standard American English are largely associated with final falling tonal patters.

According to Hedberg et al., wh-questions are most often produced with final falls because a falling contour is 'used to get more detailed information about an ongoing topic, to open up a new subtopic or to influence the development of an ongoing topic. Rising questions are most often used to ask for background information, and also to clarify information that is not audible' (2010: 4). Thus, the fact that wh-questions are occasionally uttered with rising contours is probably related to their discourse function.

- **British English** (Chen, and Gussenhoven, 2001; Nolan, 2003; Chen et al., 2004; Fletcher et al., 2005; Ladd, 2006; Grabe et al., 2010; Mennen et al., 2007; 2008, 2012).

Unlike other Germanic languages such as German and Dutch, British English is characterized by an overall high F0 level and a wide span (Chen, and Gussenhoven, 2001; Chen et al., 2004; Mennen et al., 2012). Pitch range variation has been examined for many varieties of British English (Grabe et al., 2000; Grabe and Post, 2002; Van Leyden, 2004).

The Southern Standard British (SSB) English has been the object of most studies (Chen et al., 2004; Fletcher et al., 2005; Ladd, 2006; Grabe et al., 2010; Mennen et al., 2007; 2008, 2012). In addition to SSB English, a number of investigations on other English varieties has been carried out, thus creating a general framework on the intonation systems across the UK. In particular, within the IViE - Intonational Variation in English project, cross-varietal and stylistic variation in English intonation has been examined by a group of scholars in the Phonetics Laboratory at the University of Oxford and the Department of Linguistics at the University of Cambridge. To the best of my knowledge, studies on intonation systems were realized on Belfast English (Jarman and Cruttenden, 1976; Rahilly, 1991; Wells and Peppé , 1996; Lowry, 1997); Tyneside English (Pellowe and Jones, 1978; Local et al., 1986); Liverpool English (Knowles, 1978); Welsh English (Wells, 1982; Tench, 1990; Walters, 1999); Manchester English (Cruttenden, 2001); Glasgow English (Mayo et al., 1996; Vizcaino-Ortega, 2002).

The tonal differences across the English varieties mentioned above are considerable. The intonation of a certain variety can be perceived as ‘sing-songy’ or ‘flat’, depending on the pitch level and span. What is more, pitch patterns may be so different across English varieties to characterize sentence modes (yes-no questions vs. wh-questions vs. statements) in opposite ways (rises vs. falls). For example, Nolan observed that ‘one of the most distinctive dialects of English from the intonational point of view is Northern Irish English (NIE), which always goes up at the end’ (2006: 447). In agreement with previous studies (Rahilly 1991, Cruttenden 1995), Grabe asserted that ‘Southern British English speakers produce declaratives with falling intonation and questions without morpho-syntactic markers with final rises. Belfast speakers do not appear to make this distinction; both sentence types are produced with rising intonation’ (2004: 3).

Differences among declaratives and inversion questions in Cambridge and Belfast English have been examined by Nolan (2006); results on the tonal pattern occurrences are shown in fig. 15.

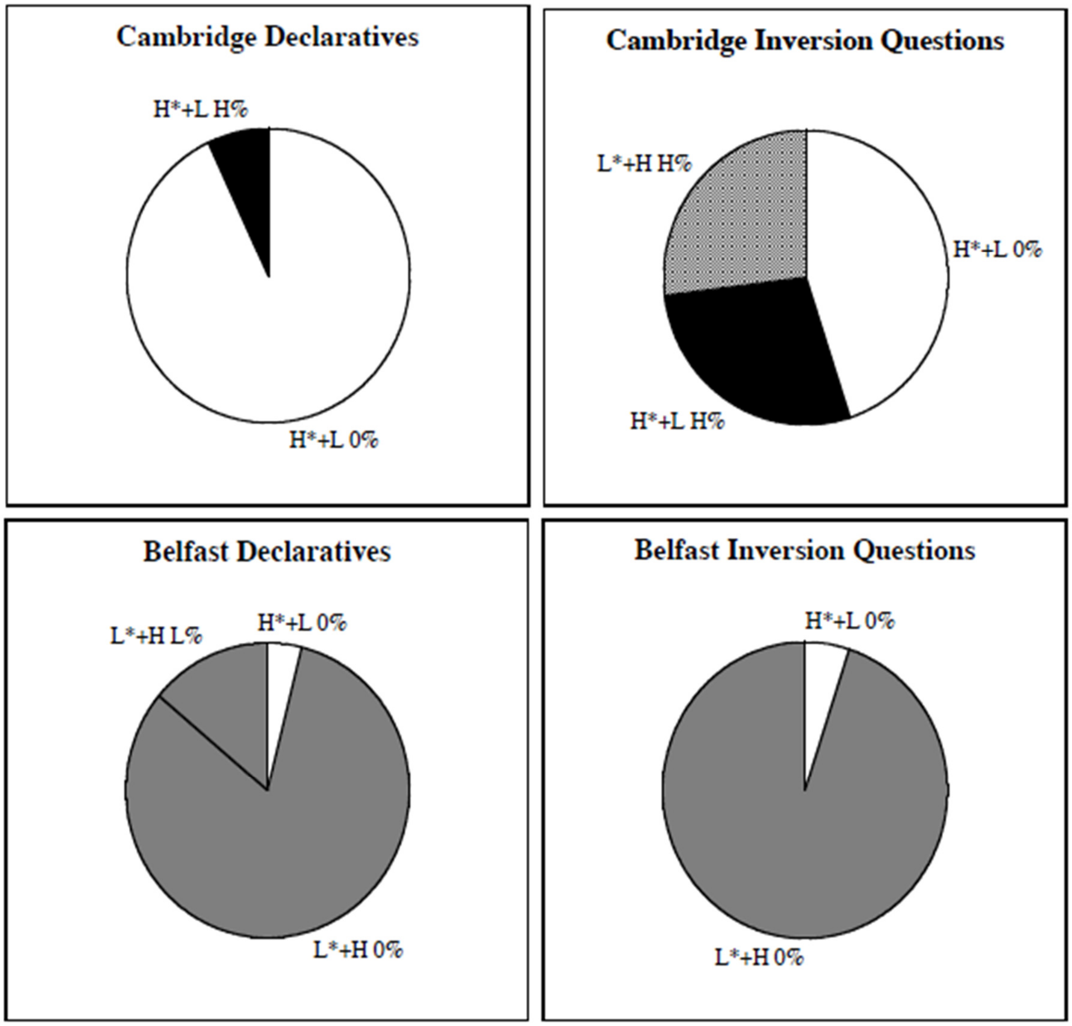


Figure 15. Distribution of English tonal patterns across declaratives and inversion questions in Cambridge and Belfast English. (From Nolan, 2006: 449).

In fig. 15, it is shown that in Belfast English statements and yes-no questions are mostly produced with rise-plateau patterns (L*+H L-L%), thus ‘revealing that these utterance types are generally not phonologically distinct’ (Nolan, 2006: 449). By contrast, in Cambridge English, statements and yes-no questions have distinct patterns. In fact, while statements are predominantly uttered with falling contours (over 90% of tonal patterns are falls, H* L-L%), more than 50% of yes-no questions are uttered with rising contours (i.e. either low rises, L*+H H%, or fall-rises, H*+L H%).

Even though differences are reported within British English varieties, a specific tonal trend has been found across sentence types. This gives a global and unitary picture of the entire tonal system, considered as a whole:

> ‘We found a consistent correlation between average F0 and the question/statement distinction. Average F0 in questions was higher than in statements and the height of the average was affected by the lexical and/or syntactic characteristics of the question. In all dialects, average F0 was highest in declarative questions, lower in other questions and lowest in declaratives. Comparable observations have been made in many other languages’. (Grabe et al., 2010: 11)

Ladd (1996) pointed out the importance of differences among tonal patterns across English varieties. For example, he claimed that a yes-no question such as ‘Can I have the bill, please?’, when produced with the fall-rise pattern (H* L-H%), may be perceived as perfectly polite by a British English speaker and condescending or peremptory by an American English speaker. Thus, intonation patters are an important cue to distinguishing among English varieties and to conveying different paralinguistic meanings.

- **Australian and New Zealand English** (Mitchell and Delbridge, 1965; Daly and Warren, 2001; Fletcher et al., 1999; Warren, 2005)

The intonational phonology of Australian English is not significantly different from Standard Southern British (SSB) English (Mitchell and Delbridge, 1965; Fletcher et al., 1999). As already noted for other varieties of English such as British English and American English, a fairly similar inventory of patterns is used within different varieties of English. Intonation differences have been reported within pitch patterns of Australian and New Zealand English, as compared to British English.

Australian English has been described as ‘a rising variety of English’ (Fletcher, 2010: 1), due to the large amount of rising contours in its intonation pattern inventory. In fact, the rising contour is often realized not just in questions but also in statements (that, cross-linguistically, are typically realized with falling contours). Pitch level and span have been found to correlate to some extent with the perception of question tunes vs. statement tunes. Across rising contours, high pitch level and wide span are typically associated with question modes while low level and narrow span are associated with statements (Fletcher, 2010).

In their review of these English varieties, Fletcher and colleagues (1999) traced the peculiarities of an interesting phenomenon observed in Australian and New Zealand English: the so-called high rising terminal (henceforth referred to as the HRT). The HRT is a pattern typically used in yes-no questions. Several studies have documented the use of HRT also in declarative sentences produced in Australian English (Horvath, 1985; Guy and Vonwiller, 1989) and New Zealand English (Britain, 1992, Cruttenden, 1994). In fact, it was claimed that

> ‘phonetically identical high-rising tunes can be used to signify these two different utterance types [yes-no questions and statements], which makes Australian English and New Zealand English intonation different from SSB English, for example where the high-rising nucleus is used primarily with yes-no questions, and never with declarative utterances’ (Fletcher et al., 1999: 2)

High rising on the nucleus is a condition for yes-no questions that is not realized in statements. What is more, early rises are typically associated with questions while late rises with statements (Warren and Daly, 2005).

This phenomenon has implications also across sentences produced by males vs. females. Generally speaking, it has been noted that women voices are characterized by more pitch variability than men voices (Daly and Warren, 2001). Within HRT, it was found that females start their rises later than males. In particular, females realize their late rise by producing L* L-H% or L* H-H% contours while males predominantly use the L+H* H-H% and exhibit sharp early rises on the accented syllables (Warren and Daly, 2005). This suggests that intonation contours are differently realized depending on sentence type and gender differences.

- **German** (Gibbon, 1998; Atterer and Ladd, 2003; Grice and Baumann, 2007; Fèry and Kügler, 2008; Reckling and Kügler, 2008; Niebuhr and Wolf, 2001; Mennen et al., 2006, 2007, 2008, 2012)

Prominence in German is realized by means of loudness, duration and pitch movements (Gibbon, 1998; Baumann and Grice, 2007). In languages such as English and German, prominence is realized on a designated syllable but the timing of the peak occurrence can change considerably depending on a number of factors. In fact, the same phonological category, either H or L, can be realized earlier or later in different language varieties (Mennen, 2006). This phenomenon is called alignment and it has been extensively studied in cross-linguistic studies.

As far as German is concerned, alignment has been considered across Germanic languages and German regional varieties (Atterer and Ladd, 2004, Mennen at al., 2012). These cross-linguistic differences in alignment have been exemplified by Atterer and Ladd as follows:

1. ‘Alignment in German is later than in English and Dutch (i.e., in German, both the beginning and the end of the rise are aligned later with respect to the segmental structure of the stressed syllable).

2. There is a difference in alignment between Northern German speakers and Southern German speakers, with the alignment in Southern German speech being later than that in Northern’ (Atterer and Ladd, 2004).

Thus, data on alignment in German suggest that peaks are placed later within the segmental strings than in English and Dutch. Within German regional varieties, Northern varieties align peaks earlier than Southern varieties. Differences of alignments across languages correlate with differences in pitch range. In fact, pitch range in German is found to be lower (low F0 level) and narrower (narrow F0 span) than in English (Mennen, 2006). Impressionistic observations on German reveal that German sounds quite level and flat, probably due to scarse pitch variation within sentences.

Gibbon (1998) asserted that German speakers are often perceived as ‘bored’ or ‘unfriendly’ due to their monotonic pitch contours and low-pitched voices. This assumption was tested by Reckling and Kügler (2008) in a production and perception study. Different materials, consisting of sentences positively or negatively connoted, were created in order to examine the correlation between pitch range and expression of positive/negative attitude. The findings confirmed the idea that positively connoted statements are characterized by a large pitch span while negatively connoted statements are marked by a small pitch span.

The data compared in the production study were further examined in a perception study, where the correlation between pitch range and expressions of politeness was tested. The results ‘indicate that the pitch range on negative connoted items is pragmatically relevant in a discourse. Speakers make use of the reduction of prominence via pitch range, and listeners are able to perceive it’ (Reckling and Kügler, 2008: 1673). Thus, it was shown that positively-perceived statements are prosodically marked by wide pitch span while negatively-perceived statements are prosodically marked by narrow pitch range. It is still an open question, though, whether or not this correlation can be considered as German specific or universal (see Gussenhoven, 2004 for the role of pitch range in the expression on positive/negative emotions and the relevance of the biological codes).

Intonational characteristics of F0 contours in different sentences types have been examined in a study by Brinckmann and Benzmüller (1999). Yes-no questions, wh-questions and statements were compared by calculating boundary tones, nuclear pitch accents, F0 offset, F0 onset, F0 range. The results show that the speakers in their study used F0 variation to distinguish between the three utterance types. In particular, yes-no questions were marked by rising contours ending in a high boundary tone. Both the F0 onset and offset were high; thus, F0 span was rather narrow. By contrast, patterns in wh-questions had a large F0 span, with a high F0 onset and a low F0 offset. Unlike yes-no questions, wh-questions were produced with falling contours ending in a low boundary tone. Patterns in German statements were characterized by high F0 onset, low F0 offset and a falling contour ending with a low boundary tone. As for F0 range, statements exhibited a narrow span.

A direct correlation between pitch span and dialect variation is found because 'pitch excursion functions as a dialect-specific cue, where the excursion of falling accents appears to be somehow more reliable and less variant than that of rising accents' (Kügler, 2009: 408). According to Gilles (2005), F0 span and its shape within the accented syllable are distinctive features across German dialect varieties. For example, the F0 span values range 'from 6.43 semitones in Dresden to 9.56 semitones in Duisburg' (Gilles, 2005: 165). Thus, dialect differences can be predicted on the basis of F0 span measurements, with minor excursions reported for Eastern German dialects and larger excursions found in Western German dialects (Gilles, 2005).

- **Dutch** (van Bezooijen, 1995; Shriberg et al., 1996; Ladd et al. 2000; Gussenhoven and Rietveld, 2000; Chen, and Gussenhoven 2001; Chen et al. 2004; Mennen, 1998, 2004; Reinish at al., 2011)

As for most of other Germanic languages, pitch accents work in Dutch as pointers to new information or as markers of a contrast-relation (Krahmer and Swerts, 1998, 2001). In Dutch, lexical stress is marked suprasegmentally by changes in duration, pitch, amplitude, and spectral tilt (Cutler, Wales, Cooper, & Janssen, 2007). Depending on the sequences of pitch accents, different sentences types are realized with rising or falling pitch movements. Typical patterns for yes-no questions, wh-questions and statements are associated to specific pitch movements that have been analyzed in different studies (Mennen, 1998; Haan, 2002; Swerts et al. 2002; Chen 2011).

Yes-no questions are characterized by rising contours (Gussenhoven, 2002; Haan, 2002; Chen, 2005). According to Lindsay (1985), the interrogative mode is expressed by several tonal mechanisms: 1) use of high tones, 2) prominence of raised peaks, 3) presence of high final boundary tones and 4) globally raised pitch. These mechanisms are exploited in Dutch yes-no questions, characterized by specific local and global acoustic properties: 'i) a high beginning; ii) a relatively high nuclear accent peaks; iii) a final rise; iv) a globally raised register level; v) absence of declination' (Haan, 2002: 52). Thus, it has been confirmed that Dutch pitch patterns are realized by means of a number of intonational properties distinguishing questions and statements. These properties include a combination of pitch patterns that contribute to mark yes-no questions with a specific feature: globally high pitch level.

Wh-questions have been differently interpreted among scholars. In some studies (see Haan, 2002 for a discussion of the studies on Dutch interrogativity), it has been argued that yes-no questions and wh-questions share common patterns; in other studies, it has been claimed that patters of these two question types are significantly different. In particular,

> while some authors claim that wh-questions typically lack a final rise (Guittart 1925:41; Stutterheim 1953:131; Droste 1972:124), others contend that presence or absence of a final rise in a wh-question reflect a speaker's attitude (Zwaardemaker and Van Eijk 1928:289).
> (Haan, 2002: 14)

When considering pitch span across sentence types, yes-no questions and wh-questions have a similar pitch span but a different pitch level, with yes-no questions having higher F0 level than wh-questions (Haan, 2000). In particular, 'wh-questions frequently combine a raised register level with a strongly narrowed register span' (Haan, 2002: 90). Generally speaking, the incidence of final rise is much more relevant in yes-no questions than wh-questions. Most of wh-questions are interpreted as questions, independently from the presence/absence of a final rise.

Statements typically have a falling contour, with low pitch level and narrow pitch span. F0 of statements considerably lowers along the contours until it reaches the baseline (presumably at the final part of the utterance). This kind of pattern is used 'by default' in a number of languages, including Dutch (Gussenhoven, 2002; Haan, 2002).

3.4 L2 acquisition models

The acquisition of an L2 language involves some kind of formal instruction, even when it takes place in a natural life context. Even though the L2 acquisition process is similar to the process that children undergo when learning their native language (Krashen, 1987), the L1 transfer on L2 acquisition plays a major role in making the acquisition process more or less problematic. Despite a general agreement on the crucial role of L1 transfer in the acquisition of interdependent production and perception skills, many unsolved issues have been raised within the linguistic community.

Bilinguals tend to use the 'grid' of their L1 phonology to interpret sounds of their L2 (Trubetzkoy, 1939; Weinreich, 1953, 1957; Wode, 1978; Lehiste, 1988; Flege, 1995). At a first sight, learning a new language (L2) that is similar to one's (L1) native language seems quite an advantage, because it enables language learners to parallel phonological categories of the L2 to the original (already acquired) inventories of the L1. However, this is not always the case. In several studies (Flege, 1995; Best, 1995; Kuhl and Iverson, 1995), it has been claimed that 'a relatively high degree of perceived dissimilarity [between L1 and L2] will eventually result in accurate segmental production and perception because it will promote the formation of a new category' (Alfano et al., 2010: 457).

Generally speaking, the L2 acquisition process is based upon the linguistic experience and competence of listeners, who perceive similarities and contrasts with their L1, depending also on the nature of their contact with L2. As previous research has shown,

> 'prior contact with the stimulus language [that is L2], and position along the trajectory of native or first language (L1) development, converge in some crucial way to shape one's perception of phonetic details and phonological structure in speech. Perception differs in important ways between naïve listeners and those who have experience with the stimulus contrasts as elements of a second language (L2). Experimental influences vary as well with age at onset of L2 acquisition, and/or with other crucial aspects of fluency and usage in both L1 and L2'
> (Best and Tyler, 2007: 14).

Thus, a series of factors (including language proficiency, exposure to L2 and age at onset of L2 acquisition) contribute to making the L2 acquisition process more or less attainable and productive. Undoubtedly, adults learning an L2 encounter many more difficulties than children. In this perspective, it is important to account for adults learning process in order to shed light on the physical, psychological, and social implications related to L1 pitch range transfer in L2.

Even though several studies have reported the transfer of L1 pitch contours to the L2 (Jenner, 1976; Adams and Munro, 1978; Backman, 1979; Willems, 1982; De Bot, 1986; Grover et al., 1987; Buysschaet, 1990), only in recent years pitch range has been the center of a discrete interest within the scientific community. To date, no model on L1 and L2 pitch range acquisition has been proposed, due to difficulties in mapping pitch range variation across languages. Nonetheless, it has been shown that a model such as the Speech Learning Model (see § 3.2.1), originally conceived to account for segmental data, is valid for making predictions about aspects of L2 prosody (Mennen, 1999). Thus, the following paragraphs will outline models proposed for segmental acquisition that, everything considered, may be applied also to non-segmental acquisition.

3.4.1 Speech Learning Model

The Speech Learning Model (SLM) has been developed by Flege and his collaborators (Flege, 1995; Flege, Munro and Fox, 1994; Flege, Munro and McKay, 1995a, 1995b) in order to account for 'age-related limits on the ability to produce L2 vowels and consonants in a native-like fashion' (Flege, 1995). The predictions made by the SLM are based on the L1 influence on adult's perception of the second language they speak. In particular, this model is based on data by different groups of L2 speakers, controlled for different age of arrival in a foreign country. Studies were carried out on the production and perception of L2 vowels (for English and Spanish see Flege and Bohn, 1989 and Flege, 1991; for English and Dutch see Flege, 1992; for English and Italian see Flege et al., 1995a) and L2 consonants (Flege and Hillenbrand, 1986; Flege and Eefting, 1987a, 1987b; Flege et al., 1995a, 1995b).

According to the SLM, speech sounds are stored in a common phonological space in which L1 and L2 phonetic categories are related. In particular, the SLM predicts that an L2 category may not be formed for a sound phonetically similar to an L1 sound, because of the phonetic similarities between L1 and L2 sounds (Flege, 1995). Regardless of how different L1 and L2 sounds are, ‘it seems that non-natives often do not perceive L2 sounds in exactly the same way monolingual native speakers of the L2 target do’ (Flege, 1995). For this reason, a distinction is made between L2 sounds that are relatively similar to L1 sounds and L2 sounds that are dissimilar from L1 sounds or completely new (Henning, 1996; Delattre, 1964, 1969; Flege, 1981; Alfano et al., 2009).

Some hypotheses and postulates were formulated to establish the grounds of a second language sound acquisition model. Namely, Flege (1995: 239) conceived the theoretical structure of the SLM in four postulates derived from empirical analyses presented in his studies on segmental aspects of L2 acquisition. The SLM postulates are defined as follows:

SLM Postulates

1. ‘The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.
2. Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.
3. Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.
4. Bilinguals strive to maintain contrasts between L1 and L2 phonetic categories, which exist in a common phonological space’.
   (from Flege, 1995: 239).

To my mind, the crucial point of this model is the creation of new phonetic categories. The process of gradual differentiation between old (belonging to the L1) and new (belonging to the L2) features in a language is relevant in order to establish new phonetic categories for the L2. These category representations may be established more easily when phonological contrasts across L1 and L2 are more evident. Why is the creation of a new category so decisive for the correct implementation of the L2 phonological system? According to Mennen (1999: 33) ‘when a category cannot be established for an L2 sound, the production of this sound will be phonetically inaccurate, resulting in accented production’. For this reason, when L1 and L2 sounds merge in a common phonetic space, this results in an accented and mispronounced realization of similar but not identical sounds. By contrast, when a learner's cognitive system classifies an L2 segment with a new category, the L2 segment is likely to be accent-free, due to the attainment of the goal (i.e. creation of a new category).

The SLM primarily focuses on the idea that a correct pronunciation is conceivable by means of the creation of new L2 categories. This achievement is possible only trough the distinction between L1 and L2 categories in that common phonological space within and across languages. In sum, SLM classifies L2 sounds as new or similar to L1 sounds depending on the phonetic differences between L1 and L2 sounds (Mennen, 1999).

Despite the general consensus achieved by Flege's SLM, in recent years, more clarifications and distinctions have contributed to improving this model. For example, one of the shortcomings of SLM is that it is based only on data obtained from adults. No indication is given about children L1/L2 acquisition. It has been claimed that,

> 'there may be a difference between child bilinguals (bilinguals acquiring both languages as infants or young children) and adult bilinguals (who started learning the L2 in adulthood), that is child bilinguals may have two separate phonological systems, whereas adult bilinguals may have a common phonological system for both L1 and L2'
> (Mennen, 1999: 173).

In order to test the validity of this thesis, the productions of child and adult bilinguals should be analyzed cross-linguistically (with a control group of monolingual native speakers). To date, such a study has not been attempted yet but several hypotheses have been formulated about the possible findings. If the aforementioned hypothesis is valid, the results obtained from the experiment on adult and child bilinguals should be different. While children, having separate phonological spaces for their L1 and L2, should be able to attain productions similar to those of monolinguals, adults would fail because, contrary to monolinguals, they share a common phonological space for their L1 and L2 (Mennen, 1999).

3.4.2 Perceptual Assimilation Model

The perceptual assimilation model, also called PAM, (Best, 1993 and 1994) expresses to which extent foreign language contrasts are perceived and perceptually assimilated to the native phonological system. Based on the idea that L1 plays a fundamental role in the perception of L2 segments, this model predicts three assimilation patterns by which non-native segments may be assimilated (Best, 1995; Mennen; 1999; Alfano et al., 2009).

Language *environment* and *experience* are two key-words of the PAM. The role of language environment in L2 acquisition is clear, as it develops along three separate dimensions: a) length of residence (LOR); b) usage of L2 and quantity/quality of input from L2 speakers; c) age of learning (Flege, 1999, 2002; Jia and Aaronson, 2003; Jia et al., 2006; Best and Tyler, 2007). As for language experience, on the basis of fluency (proficiency and competence) and balance (language dominance), bilinguals are distinguished between naïve and experienced learners (Best and Tyler, 2007).

In particular, PAM addresses the problem of explaining L2 speech perception by naïve listeners. It does so by elaborating the concept of assimilation. Three main situations may occur: successful assimilation, unsuccessful assimilation, non-assimilation. The degrees of assimilation may be explained as follow:

Assimilation Types

1. *Two Categories (TC) assimilation*

   If two L2 sounds are perceived as different, they will be mapped as perceptually distinct and assimilated into two different L1 categories (this being an example of good assimilation).

2. *Single Category (SC) assimilation*

   If two L2 sounds are perceived as similar to a native phoneme, both L2 sounds will be assimilated into a single L1 category (this being an example of poor assimilation).

3. *Non-assimilation*

   If two L2 sounds are too different from any native phoneme, they will fail to be assimilated within the native phonological space. Thus, they are ‘non-assimilable’.

Assimilation types consider the possible outcomes of interactions between L1 and L2 phonological systems. The predictions made by PAM are based on data from a group of adults' perception of a language unknown to them. This model predicts that 'non-native segments [...] tend to be perceived according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in the native phonological space' (Best, 1995: 193).

The assimilation of new sounds is more or less successful (see assimilation types) depending on differences in category goodness. It is assumed that 'learners are able to perceive variation in the goodness of fit of an L2 sound to an L1 category' (Mennen, 1999: 35). So, when two or more L2 sounds are perceived as tokens of an L1 phoneme, they differ in their level of goodness of fit to that phoneme. In fact, poor or good assimilation depends on the goodness of fit of L2 sounds to L1 sounds. In the case of SC assimilation, that is the assimilation of two L2 sounds into one L1 sound, one of the two L2 sounds is better assimilated than the other, depending on the proximity of the two sounds to the partially-similar L1 phoneme (Best and Tyler, 2007).

Unlike SLM, PAM defines similarities between L1 and L2 in terms of articulatory- phonetic contrasts (Mennen, 1999). In fact, PAM describes the nature of L1 influence on L2 'as being based on perceptual learning of phonetic-articulatory pattering at both the abstract contrastive level and, importantly, at the level of non-contrastive gradient phonetic detail' (Best and Tyler, 2007: 22). Since naïve listeners are believed to be unaware of phonological distinctions in their L2, they can rely only on their L1 to find out differences between unfamiliar L2 sounds and familiar L1 sounds.

3.4.3 Native language Magnet Model

The Native language Magnet Model, also called NLM (Kuhl, 1991; 2000), assumes that prototypic native sounds work as a magnet for non-native sounds. The magnet, in a way, attracts those L2 segments that are perceptually similar to L1 segments. More specifically, this model is based on the idea that 'language experience alters the mechanisms underlying speech perception, and thus the mind of the listeners' (Kuhl and Iverson, 1995: 121). Indeed, it is a matter of common evidence that

> 'listening to a foreign speaker shows how extensively language experience alters perception. The sounds emitted are a jumble to us; however, they make perfect sense to a native speaker of that language. […] The native listener hears familiar sounds and words; the other, a stream of unrecognizable noises. In other words, what differs is the mind of the beholder' (Kuhl and Iverson, 1995: 121).

It has been claimed that 'exposure to language early in life produces a change in perceived distances in the acoustic space underlying phonetic distinctions, and this subsequently alters both the perception of spoken language and its production' (Kuhl and Iverson, 1995: 122). In order to shed light on developmental and cross-language mechanisms affecting perception, *perceptual magnets*, which work as sort of prototypes of abstract categories, are examined.

Perceptual magnets are identified as the 'best instances' (Grieser and Kuhl, 1989; Kuhl, 1991) of phonetic categories in L1 that have an attractor effect on L2 sounds. Thus, certain perceptual distinctions are based on the distributional properties of phonetic categories that are analyzed as language inputs. Depending on the level of similarity of the L1 prototype and the L2 sound, the perceptual process becomes more problematic. Not only the magnet effect does not facilitate the acquisition of L2 categories but also it contributes to make it even more problematic. For instance, perceptual magnets (L1 sounds) may attract similar L2 sounds, thus distorting the perceptual space and making L2 contrasts difficult to acquire. This is due to the fact that, 'the native-language categories of the listeners somehow interfere with the ability to perceive the phonetic distinctions in the new language' (Kuhl and Iverson, 1995: 143). In line with Flege (SLM model) and Best (PAM model), Kuhl assumes that the closer an L2 sound is to an L1 sound, the more it will be assimilated into an L1 category. This may become problematic when there is not a one-to-one correspondence between the L1 and L2 phonological space.

Since it has been demonstrated that adults are perceptually affected by their L1 (Flege, 1995; Best and Tyler, 2007), the predictions made by the NLM must be limited to data on infant language acquisition. Based on the data gathered in NLM studies (Kuhl and Iverson, 1995), infants show an effect of L1 influence by 6 months of age. In fact, English, Swedish, and Japanese infants have the ability to recognize differences among phonetic categories and to exhibit language-specific magnet effects. For example, they were able to discriminate between vowel categories to locate them into the vowel space, on the basis of their exposure to their L1 (English vs. Swedish vs. Japanese) (Kuhl and Iverson, 1995). This means that infants successfully managed to acquire new categories in an amazingly short time (as compared to adults) and they kept these categories distinct within the perceptual space.

3.4.4 Feature Competition Model

The Feature Competition Model (FCM) is based on the assumption that some categories are more prominent than others; therefore, those categories that are more prominent in L1 will have a greater influence on the perception of L2 (Hancin-Bhatt, 1994). In particular, categories that are more or less prominent within a phonemic inventory are called 'features'. On the basis of the degree of prominence of a feature in L1 sounds, L2 sounds will be perceived accordingly. In fact,

> 'those features that are more prominent in the L1 system will tend to have a greater influence on learner's perception of new L2 sounds; that is, the feature prominences on the L1 will guide how L2 sounds are mapped onto existing L1 categories' (Brown, 2000: 10).

It is undeniable that, to some extent, FCM is very much based on previous models (see SLM; PAM; NLM). Moreover, the concept of *feature* is similar to the notion of *magnet*. On the one hand, a feature is defined as the most prominent category; on the other hand, a perceptual magnet identifies the best instance of phonetic categories. The difference between NLM and FCM lies on the fact that magnets are conceived as abstract prototypes while features are perceived as more dynamic components. Thus, 'unlike previous analyses, the FCM adopts a dynamic approach to phonology, one which assumes that features do not have discrete values, rather ones which are continuous, of greater or lesser prominence in an inventory' (Hancin-Bhatt, 1994: 241). In a way, the phonetic transfer from L1 to L2 occurs depending on the feature prominence. Hancin-Bhatt herself explains that, in the L2 acquisition process, what transfers from L1 to L2 is not

> 'general phonemic categories or patterns of features, rather feature prominences which may differentially affect the mapping of L2 feature patterns onto L1 patterns. Universally, then, all features are available; the patterns of activation and their likelihood of occurrence are language-specific' (Hancin-Bhatt and Govindjee, 1999: 157).

What is more, three different types of transfer from L1 and L2 were identified as based on L1-L2 sound mappings:

Transfer type:

1. Positive transfer: L1 and L2 sounds are the same
2. Transfer by feature prominence: L1 and L2 sounds are similar
3. Negative transfer: L1 and L2 sounds are completely different

The FCM model was tested for language-specific differences in interdental consonants mappings based on L1 input (Hancin-Bhatt and Govindjee, 1999). Differences in feature prominences across L1 were shown to be relevant in the perception of L2 interdentals. In particular, results demonstrated that L2 interdentals were perceived at different levels, depending on the feature prominence in L1.

3.5 Summary

Chapter 3 presents the outcome of cross-linguistic research on pitch variation across different languages. A summary and critical overview of previous studies conducted on first and second languages are provided. This is done by taking into consideration the inventories of intonation patterns reported for Italian varieties and Romance languages (i.e. Spanish, Catalan, French, European and Brazilian Portuguese) compared to English varieties (i.e. American English, British English, Australian and New Zealand English) and other Germanic languages (i.e. German and Dutch). The examination of previous studies on pitch range variation is motivated by the necessity to critically investigate the pitch differences envisaged in Italian and English.

Qualitative and quantitative pitch differences relative to the patterns of yes-no questions, wh-questions and statements have been observed in a number of studies and languages. The theoretical framework used for the phonological analysis of pitch range is based on the Auto-segmental Metrical approach. The system proposed for the annotation of prosodic information is called ToBI (Tones and Break Indices) and it represents a standard system for prosodic transcription, developed within and adapted to different scientific disciplines. In the ToBI notation system, pitch movements are described and analyzed in terms of categorically distinct elements that provide strings of H and L markers.

In order to better evaluate the impact of pitch range in the second language acquisition domain, four L2 acquisition models are critically analyzed: the speech learning model (SLM); the perceptual assimilation model (PAM); the native language magnet model (NLM); the feature competition model (FCM). To date, no model on L1 and L2 pitch range acquisition has been proposed, due to difficulties in mapping pitch range variation across languages. Nonetheless, it has been shown that a model such as the SLM, originally conceived to account for segmental data, can be used to make predictions about L2 pitch range. Based on the theories proposed in the SLM, PAM, NLM, FCM, considerations on the L2 acquisition of pitch range can be made.

Chapter 4:
Production study

4.1 Introduction

Although it is well documented that both segmental and prosodic aspects contribute to a perceived foreign accent, most research has focused on segmental acquisition and as a consequence current bilingual speech models (such as the Speech Learning Model or the Perceptual Assimilation Model) have not yet attempted to account for non-segmental aspects of bilingual speech learning, despite models being in existence for almost two decades.

Yet, intonation plays a key role in communication. In particular, it is used in different contexts to disambiguate the meaning of sentences, to emphasize specific parts of discourse or to convey emotional message, such as a sense of participation, interest, detachment etc. (Ladefoged, 2006) and it plays a major role in intelligibility (Maassen & Povel, 1984; Munro and Derwing, 2001; Isaacs and Trofimovic, 2011). Prosody is one of those fundamental factors which make a discourse intelligible and increase communication proficiency: the more a speaker pronounces an utterance with the correct pronunciation, accent and intonation; the more this utterance is likely to be easily understood by the listeners.

The intonation patterns we use when we speak have an immediate effect on the pragmatics of our communication. Intonation serves different linguistic and paralinguistic functions that may range from the expression of sentence modality to the marking of emotional and attitudinal nuances. In spoken language, intonation not only indicates the distinction between sentence types, but also conveys a speaker’s attitudes such as involvement, concern, surprise, boredom and so on (Wells, 2006). The general meaning of a sentence can be predicted by considering its tone.

4.2 Rationale behind the experiments

It is often suggested that Italian-accented English sounds like a sing-song and is more rhythmic than English (Einsenchlas and Tsurutani, 2011). Not just intonation, rhythm and stress patterns are responsible for such an effect, pitch range is also likely to play a role in the perception of the Italian lilt. The aim of this study is to compare pitch range in selected utterances produced by American English native speakers and Italian learners of English, in order to analyze and better understand differences in pitch level and span.

How can pitch range affect the way they are perceived? Preliminary studies on Italian prosody in English (Busà and Urbani, 2011; Busà and Stella, 2012) suggest that there may be substantial differences in the intonation patterns used by the Italians' non-native speakers and the English native speakers. Little is known about the Italian and English differences in pitch range.

Hirschberg and Avesani (2000) compared the mechanisms speakers employ to disambiguate syntactically and semantically ambiguous utterances in English and Italian. Their study shows that the strategies used to disambiguate sentences can differ among English, Italian and Spanish and that, in their use of intonation, Italian and Spanish speakers pattern together more often than either pattern with English. Mennen (2007) identified certain dimensions of intonation where non-native speakers (NNS) may differ from native speaker (NS) productions. These are the inventory of boundary tones and pitch accents (systemic dimension); the phonetic implementation of these structural elements (realisational dimension); the distribution of boundary tones and pitch accents (phonotactic dimension); and functionality (semantic/pragmatic dimension). These studies suggest that it is possible that NNSs may transfer their L1 intonation patterns into their L2 utterances while they are speaking a second language. They are probably unaware that this transfer process may lead them to unsuccessful communication. As a result, the difficulty in communication in interactions between NSs and NNSs is that L2 learners often use intonation patterns which do not convey the intended meaning (Busà, 2008).

Extensive research on contrastive English L1 and L2 prosody is needed in order to better understand and evaluate the mechanisms of English interactions between English NSs and Italian learners of English. To this end a database will be gathered of a group of native speakers of English and two groups of Italian second language learners of English (with and without specific intonation instruction). The groups will be matched for all factors known to play a role in intonation (e.g. sex, region, age, age of acquisition, etc.)

and form the basis of a cross-language comparison to identify the differences between native and non-native intonation characteristics in each of the above mentioned dimensions. After identification of these differences, a series of perception experiments will be devised where the various intonation parameters are manipulated separately and presented to listeners to identify their role in intelligibility as well as in foreign accent ratings.

This chapter discusses the methodology used for the comparative analysis of pitch level and span in English and Italian, and illustrates the results of two production studies.

4.3 First Experiment

The experiments described in this chapter were aimed to investigate the differences in the realization of pitch level and span in English and Italian. In the first experiment, pitch range variation was measured across sentences selected from a passage and uttered by several English native speakers (henceforth NSs) and non-native speakers (henceforth NNSs). Three kind of speech materials were collected and examined: American English speech produced by NSs, English speech produced by NNSs (Italian learners of English), and Italian speech produced by NSs. Pitch range variation was compared across speakers with English as L1/L2 and Italian as L1, in order to find out how pitch varies across languages.

Oral reading of some passages was performed by several speakers. They were asked to read the materials in a natural conversational way. The procedures which led to data collection were comparable across speakers: they read the same speech materials under the same conditions (e.g. the recording sessions took place in a sound-attenuated booth with sophisticated laboratory equipment). Moreover, personal information of all subjects was gathered in order to obtain a homogeneous corpus. The population size had roughly the same characteristics (matched by age, gender and level of education) and had a reasonable size.

In sum, the characteristics of the first experiment can be outlined as follow:

Characteristics of the first experiment:

- Subjects: both male and female subjects
- Size of the corpus: 8 subjects per language groups and gender groups
- Languages: English (L1 and L2) and Italian (L1)
- Materials: conversational paragraph containing lively speech.

4.3.1 Research Questions

In line with previous research, one might think that no matter the language, L2 speech is characterised by a narrower pitch range than L1 speech. Thus, it is possible that English sentences produced by Italians have an overall lower and narrower pitch range than those produced by Americans. However, to date, this has not been consistently confirmed by a large-scale study on cross-linguistic data. Impressionistically, American English is thought to have a larger pitch span than Italian. The present study measures pitch span in utterances produced in English by NSs (American English) and NNSs (Italians) of English, to test the first hypothesis that Italians make use of a narrower pitch range than the American subjects.

In addition, the English sentences produced by the Italians are expected to have a narrower pitch span than those produced by the Americans. Hence, they might have a slightly monotonous tone and rather flat pitch contour. The second hypothesis being tested is that Italian learners of English transfer the L1 pitch range variation into their L2, due to the influence of their L1. Thus, they will show a similar pitch range in their L1 and L2 speech.

Finally, this study is aimed at investigating whether or not pitch range varies in F0 level and span across genders. It has been shown that F0 values differ in males and females because of physiological (see § 2.2.1 and § 2.2.3) and socio-cultural factors (see § 2.2.4, the biological codes in § 2.3). However, it is not known whether both males and females modify their pitch values depending on the language they are speaking. It is expected that males and females have similar pitch patterns, due to the influence of their L1.

In sum, this experiment was built up in order to shed light on pitch variation in English and Italian, in English as a L1 and an L2, across the male and female population. The following three directional hypotheses were formulated to be tested in the present study:

(1) L2 speakers have narrower F0 span than L1 speakers;

(2) There is no difference in F0 level and span values between productions in (L1 and L2) English and Italian;

(3) Regardless of gender, utterances in English and Italian are produced with similar pitch variation by males and females.

4.3.2 Subjects

Eight adult native speakers of American English (4 males and 4 females) and 8 Italian (4 males and 4 females) subjects participated in the experiment. Table 16 shows the personal information about the American speakers. All American participants were speakers of American English, they all came from California and were students at the University of California – Los Angeles. Half of the American subjects were also proficient in Italian and 3 American subjects out of 8 had lived in Italy for at least one year (at the time of the experiment), by taking part into exchange programs held in Bologna, Florence and Milan.

Sp	NS	Gender	Age	Region	Major
1	American En	M	24	La Crescenta, CA	Education
2	American En	M	22	San Diego, CA	Mathematics
3	American En	M	22	Riverside, CA	Linguistics
4	American En	M	20	Cupertino, CA	Linguistics & Spanish
5	American En	F	21	Riverside, CA	Italian Film
6	American En	F	24	Benicia, CA	Italian Studies
7	American En	F	26	Riverside, CA	Italian Literature
8	American En	F	23	Los Angeles, CA	Linguistics

Table 16. Personal data of the eight American participants (Sp) in the experiment. Information is about their native language, gender, age, birthplace, and major.

Table 17 shows the personal information of the Italian speakers. All Italian speakers were either spending a period abroad in Los Angeles or they were graduate students at the University of California – Los Angeles. They came from northern Italy and only four of them reported speaking dialect in their daily life. Three out of the 8 Italian subjects had been living in Los Angeles for more than one year at the time of the experiment.

Sp	NS	Gender	Age	Region	Major
1	Italian	M	27	Belluno	Mathematics
2	Italian	M	25	Padova	Political Science
3	Italian	M	28	Vicenza	Electrical Engineering
4	Italian	M	28	Padova	Political Science
5	Italian	F	28	Verona	Italian Film
6	Italian	F	28	Treviso	Civil engineering
7	Italian	F	27	Vicenza	Linguistics
8	Italian	F	22	Belluno	Italian Studies

Table 17. Personal data of the eight Italian participants (Sp) in the experiment. Information is about their native language, gender, age, birthplace, and major.

All participants were students in different departments at UCLA: Italian (5 students); Linguistics (4 students); Engineering (2 students); Mathematics (2 students); Political Science (2 students); Education (1 student). Two students were undergraduate; six students were doing a Master degree; eight students were pursuing a Ph.D. The age of the American participants ranged from 20 to 26 years (mean age: 22,7 years). The age of the Italian participants ranged from 22 to 29 (mean age: 27). Even though the American subjects were slightly younger than Italians, a difference of up to 5 years is considered to have no impact on the homogeneity of the two groups. Other than age, the homogeneity of the subjects was controlled by the fact that most of the participants in the experiment were speakers of American English, as it is spoken in Southern California (Los Angeles area) and of Italian, as it is spoken in the North-East of Italy (Veneto area). None of the speakers reported any speech, hearing or communication disorder and all of them were non-smokers. There was no screening for formal training in music or singing. The experimenter personally knows all the participants she recruited for the experiment. They gave their consent for the treatment of their personal data (a sample of the consent form is shown in Appendix B) and volunteered for the experiment without receiving any monetary compensation.

4.3.3 Materials

This study compares native and non-native productions of 5 sentences selected from a short passage from 'The Little Prince' by Antoine de Saint-Exupéry[9]. In the data set, three different passages were recorded (see Appendix B for the complete list of materials). However, only the second passage (read twice in English and Italian) was selected for the experiment. In tab. 18, the sentences selected from the second passage for the analysis of pitch range are shown:

[9] See, de Saint-Exupéry, Antoine (2000). *The Little Prince,* (translated by Richard Howard). San Diego/New York: Harcourt.

	Second Passage:
English version	A: "I am very fond of sunsets. Come, let us go look at a sunset now." B: "But we must wait," I said. A: "Wait? For what?" B: "For the sunset. We must wait until it is time."
Italian Version	A: "Mi piacciono tanto i tramonti. Andiamo a vedere un tramonto adesso." B: "Ma bisogna aspettare..." dissi. A: "Aspettare che?" B: "Il tramonto. Dobbiamo aspettare fino a quando è ora."

Table 18. English and Italian version of the sentences selected from the second passage for the analysis of pitch range.

The second passage was chosen mainly because it was short, dialogic, and it did not contain difficult words to be pronounced. In particular, it was found to have a livelier prose than the other two passages due to the fact that it contained dialogs from two different characters (namely the narrator and the little prince). Only the first five sentences of the chosen passage were retained for the analysis, in order to have a corpus of a feasible and suitable size. The corpus created consisted of 120 utterances (5 sentences x 8 speakers x 3 language groups).

4.3.4 Procedure

During the recording session, subjects were instructed to read three short passages with a natural conversational intonation. The set of three passages were read both in English and Italian, every passage was recorded twice. Therefore, a set of two repetitions in English and two repetitions in Italian of all material was recorded. The order of material in the recording session was the same for every speaker, viz., first passage read in English and then in Italian; second passage read in English and then in Italian; third passage read in English and then in Italian.

The texts were read aloud by the 8 American English (male vs. female) speakers from California and the 8 Italian (male vs. female) speakers from the North East of Italy. The American English subjects read the materials in English; the Italian subjects read them in English and Italian. As a result, data were extracted from three different groups: (1) Americans speaking English, (2) Italians speaking English, and (3) Italians speaking Italian.

The subjects were constantly monitored by the experimenter while they were reading the sentences and they were required to repeat any sentence when they misread it. In some cases, when the speakers did not feel comfortable with the utterance pronounced, they asked to do the recording again. Before starting with the recording process, subjects were permitted to read silently the texts in order to familiarize themselves with it before reading it aloud for recording. Each recording session lasted about 20 minutes. At the end of the session, every subject was asked to fill in a questionnaire, containing questions about their personal information (see Appendix A for a sample of the questionnaire in English and Italian). Speakers were requested to indicate their first and last name; age; birth place; sex; native language; second languages and proficiency levels; university status; periods abroad; ways and daily use of learning a second language.

The materials were collected by the experimenter (within a period of three weeks) at the Linguistic Department of the University of California – Los Angeles. All the audio files were recorded and digitally acquired in a sound-attenuated booth in the UCLA Phonetic Laboratory. Recordings were collected using a Shure SM10 head-mounted microphone, recorded direct-to-disk on another computer located outside the sound booth, and digitized at a sample frequency of 44.1 KHz and a 32 bit quantization rate, using an AudioBox. By using a sampling rate of 44.1 KHz (i.e. CD quality), it was possible to collect data with excellent quality. After recording the short texts, the experimenter saved the data and labeled them as separate WAV audio files with Praat (Boersma and Weenick, 2010).

4.3.5 Method

The data were analyzed by following the method proposed by Mennen et al. (2012). Thereby pitch values were analyzed and compared across groups by calculating long-term distributional (LTD) and linguistic measures.

LTD measures are based on the analysis of F0 distribution. The analysis was carried out as follows. Values of F0 maximum (F0 max), F0 minimum (F0 min), F0 mean and F0 median were calculated over the entire sentences to measure pitch level. Measures analyzed for pitch span were: F0 maximum minus F0 minimum (max-min F0) in Hz and ST, standard deviation (SD), the difference between the 95^{th} and 5^{th} percentile (90% span), the difference between the 90^{th} and 10^{th} percentile (80% span), skewness and kurtosis. As F0 may not be normally distributed around the mean, skewness and kurtosis were measured. Skewness signals the asymmetry of values while kurtosis measures the distribution of values.

The selection of LTD measures for the analysis of pitch level and span was based on Mennen et al.'s approach (2012), see (see § 1.3.4). Values for LTD measures were obtained automatically by inquiring pitch information in Praat such as minimum, maximum, range, average, standard deviation, etc. The same protocol was used to calculate all measures using the same standards and procedure.

Pitch tracking was performed with a standard algorithm based on the autocorrelation method. This algorithm is the standard option to process speech and detect pitch locations in Praat. The autocorrelation computation is a very reliable and fairly straightforward method to process a high number of data directly on the waveform in order to detect and track a pitch line (Rabiner, 1997; Boersma and Weenink, 2010). After uploading sound files in .wav format into Praat, each sound file was transformed into a pitch file by selecting the function 'sound: to pitch'.

Just as the spectrograms are visualized according to parameters of view range (Hz), window length (s) and dynamic range (dB), pitch settings influence the display of pitch variations. In the pitch setting window, one can adjust the pitch range, the pitch unit and the drawing method. In the advanced pitch settings, it is possible to modify the maximum number of candidates, the silence threshold, the voice threshold, the voiced/unvoiced cost etc. In some cases, these parameters needed to be adjusted in the pitch-tracker to get a better-looking F0 contour.

As shown in fig. 16, the time step was automatically set at 0.0 sec, while the pitch floor and ceiling were set, by default, at 75 Hz and 600 Hz, respectively.

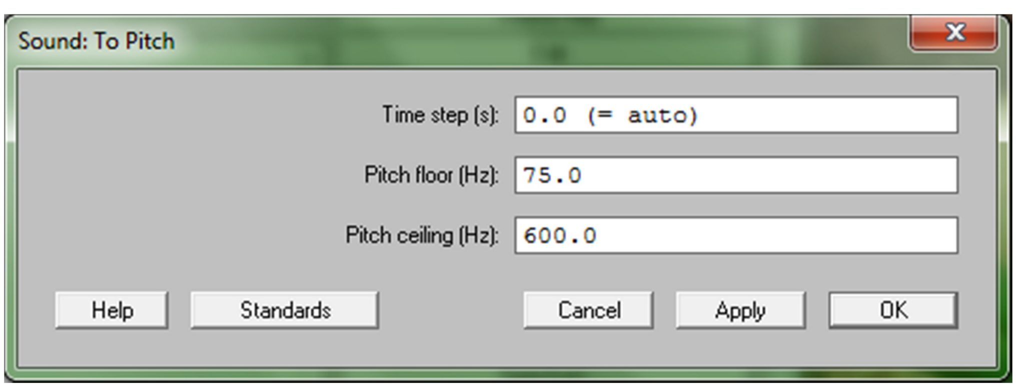


Figure 16. Standard settings for the calculation of LTD measures in Praat.

The reason why I chose to set the same standard values for pitch floor and ceiling in every sentence and for every speakers is that comparable data can be obtained only by setting standard values and keeping them constant in every analysis. My choice of values was influenced by a careful inspection of data. At a first sight, it would have been natural to use the default values in Praat (75 Hz - 600 Hz) or the values recommended in Praat manuals and tutorials (see van Lieshout, 2003; Wood, 2005; Boersma and Weenink, 2010; Styler; 2012), that is, 75 Hz for pitch floor and 300 Hz for pitch ceiling for male speakers; 100 Hz for pitch floor and 600 Hz for pitch ceiling for female speakers.

However, most of the speakers in my study reached values from about 80 Hz to about 550 Hz. In order not to miss some of these values, my choice was to set the pitch floor at 75 Hz for both males and females, and the pitch ceiling at 400 Hz for males and 600 Hz for females. This allowed me to capture any slight variation in the pitch line. What is more, a pitch floor down to 75 Hz permits to capture data about a particular phonation type called creaky voice[10]. A pitch floor set at 100 Hz would probably exclude any case of creaky voice (Mennen et al., 2012).

As suggested in Styler (2012: 17), it is recommended to ‘check any measures which seem unreasonable against single-cycle F0 measurements or against harmonic frequencies’ in order to avoid pitch-tracking mistakes made by the program. Thus, manual correction was used to adjust and edit pitch points that were shifted upwards or downwards, due to octave errors in voiceless parts of the signals, noise in the background, and pitch tracking errors. This was done in files called pitch objects that consent to manipulate pitch points by adding, shifting and deleting them.

Fig. 17 shows the original visualization in the sound file of the sentence ‘Do you need any money?’, uttered by the female American speaker 1, while fig. 18 shows the visualization in a pitch object of the same sentence. It is clear that the pitch track in the pitch object cannot be completely superimposed to the original sound file, as in the end, some additional pitch points that are not present in the sound file are visible .

[10] Creaky voice (also called vocal fry) is produced with vibrating vocal folds but at a very low frequency. In particular, creaky voice has lower F0 than modal or breathy phonation (Johnson, 2003: 136-139).

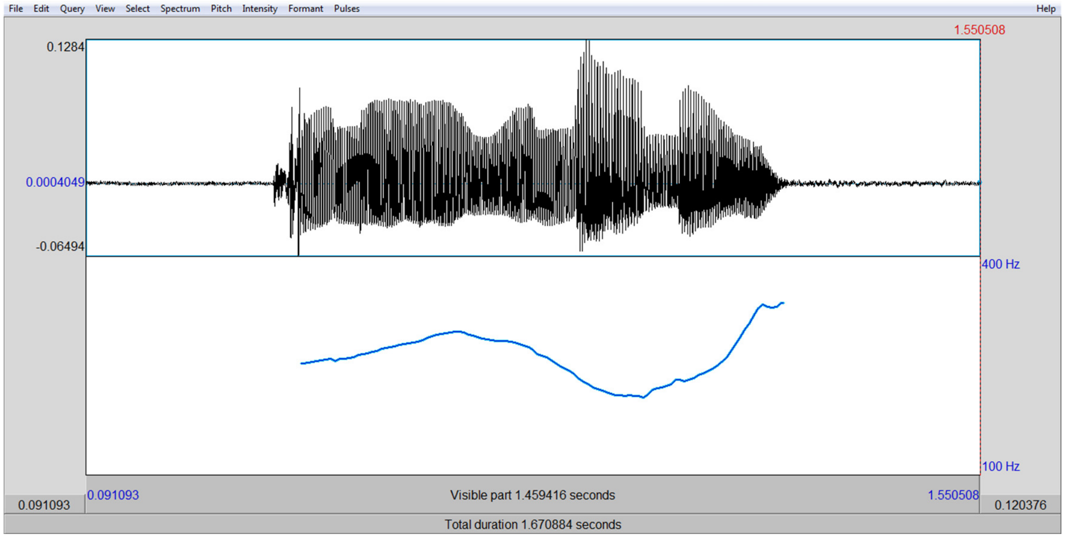


Figure 17. Screen view of the pitch line in the sound file of the sentence ‘Do you need any money?’, uttered by the female American speaker 1 .

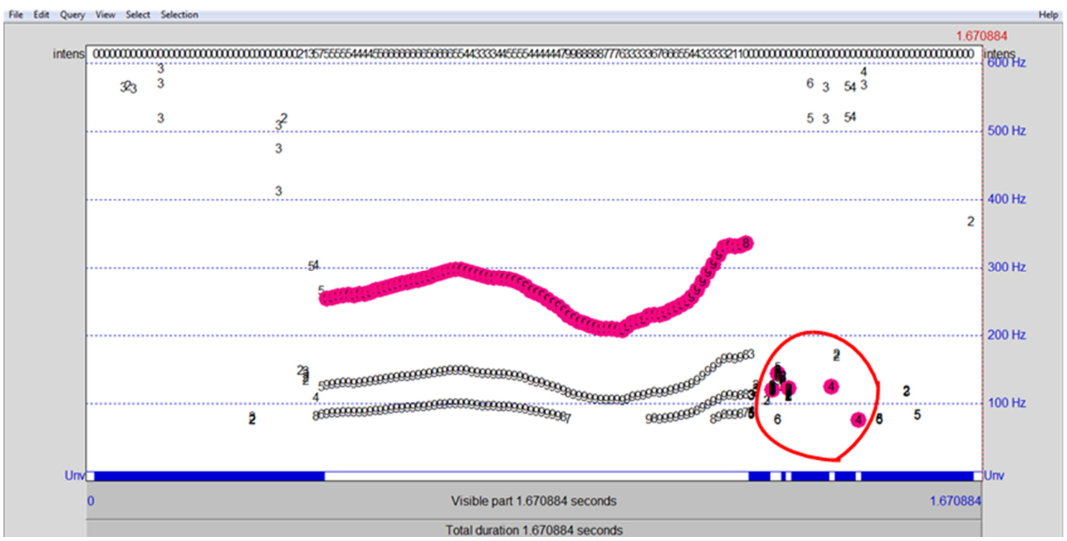


Figure 18. Screen view of the pitch line in the pitch object of the sentence ‘Do you need any money?’, uttered by the female American speaker 1.

Within the pitch line, some pitch points placed at the end of the utterance look ambiguously placed at a very low frequency. Thus, after listening to this sentence a few times and visually inspecting the audio file in fig. 18, I realized that those last pitch points (highlighted by a red circle) were the result of a pitch tracking mistake, probably due to background noise. Thus, I proceeded to erase them and I retained the corrected version of the sentence for further analysis. Fig. 19 shows the corrected version of the sentence, where erroneous pitch points have been erased.

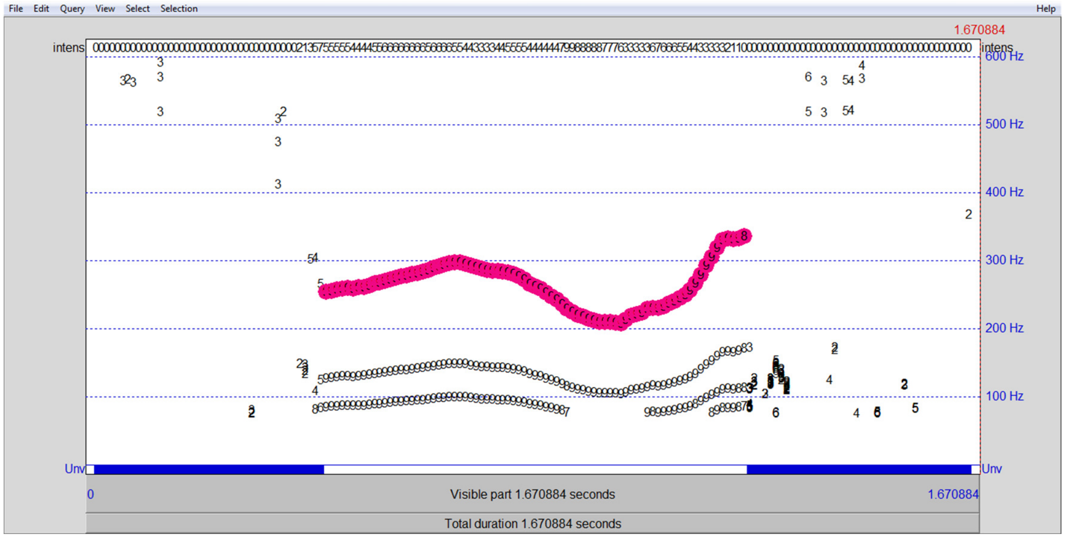


Figure 19. Screen view of the pitch line in the pitch object of the sentence 'Do you need any money?', uttered by the female American speaker 1.

In order to avoid pitch-tracking mistakes, spurious values on the pitch object visualizations were manually inspected, adjusted and, in some cases, erased. After manually correcting errors within the utterances, pitch objects were saved as binary files. Then, a script elaborated by Mennen et al. (2012) was used to automatically calculate lists of values for different measures.

To calculate linguistic measures, F0 range stylization was performed with the function 'to manipulation' in Praat. Fig. 20 shows a example of the F0 stylization process where every local F0 maximum and minimum is signaled by a pitch point. Every pitch point is the results of a auditory and visually inspection.

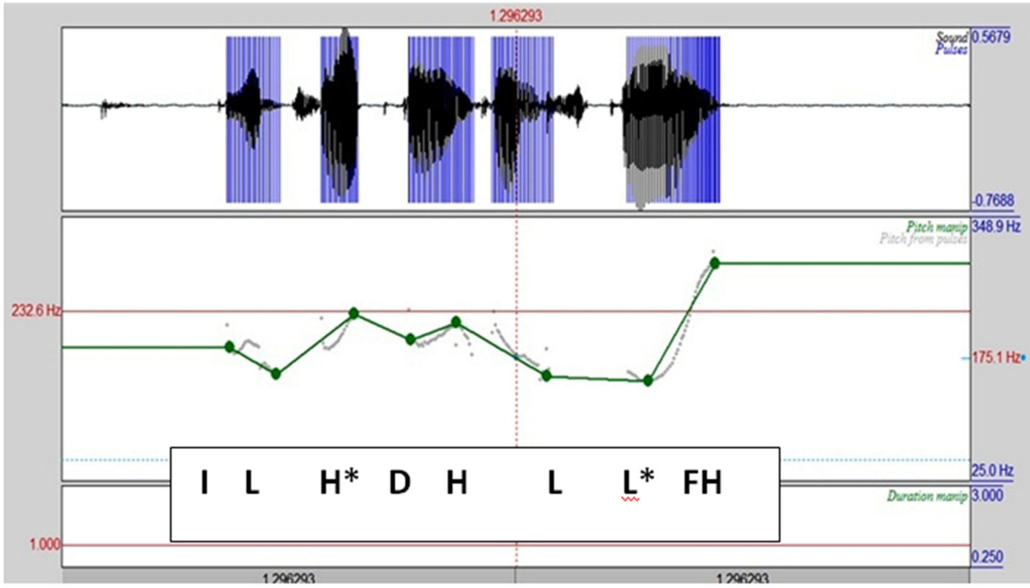


Figure 20. Screen view of the F0 stylization process where every local F0 maximum and minimum is signaled by a pitch point. Every pitch point receives a label. The manipulated sentence, 'You dropped down from the sky?', was extracted from the first passage of my corpus of sentences.

After manually inserting pitch points at local peaks or valleys, the whole corpus was manually labeled with Praat by adding F0 landmarks within the pitch track. Three simple steps were followed. First, pitch points were inserted at the beginning and the end of the intonation phrase; they were labeled respectively as 'I' and 'FH'/ 'FL' (depending on the final rise or fall within the pitch line). Second, local peaks or valley on prominent syllables were identified acoustically and visually, and they were labeled respectively as H* and L*. Third, any peak or valley on non-prominent syllables was labeled as H and L.

Label	***Description***
I	Phrase initial value
H*i	Local peak at phrase starting point
H*	Local peak, prominent syllable
H	Local peak, non-prominent syllable
L*	Local valley, prominent syllable
L	Local valley, non-prominent syllable
FH	Final local maximum
FL	Final local minimum

Table 19. Description of labels used to annotate the corpus. The method and guidelines proposed by Mennen et al. (2012) contain additional labels (!H*, D, $L*, U) that were not used for the analysis of the present corpus.

4.3.6 Results

Comparative analyses based on LTD and linguistic measures were drawn on the source languages (L1 American English vs. L1 Italian) and the target language (L2 American English) by calculating F0 range across speakers, sentence types, and gender (males vs. females).

4.3.6.1 Linguistic measures

After placing linguistic landmarks to peaks and valleys, a script in Praat was used to calculate the F0 of each pitch point. Then, values were averaged across speakers (males vs. females) and language groups (Americans speaking English vs. Italians speaking English vs. Italians speaking Italian). Values of linguistic measures were obtained for level (tab. 20) and span (tab. 21). In tab. 20, measures calculated for level were grouped depending on native language, language spoken and sex of the subjects.

Linguistic level	*Am M*	*It-En M*	*It-It M*	*Am F*	*It-En F*	*It-It F*
I	120	139	155	213	209	216
H*i	204	198	229	306	308	328
H*	139	179	163	292	254	352
Hi	158	158	188	279	288	320
H	150	175	210	263	231	326
L*	111	122	116	162	170	170
L	125	131	123	193	192	177
FL	92	111	104	150	148	164

Table 20. Overview of linguistic measures for level. Mean values for each landmark were calculated in Hz for male (M) and female (F) subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It).

For F0 level, L*, L and FL were counted as the measures of valleys, corresponding to the bottom line of the pitch contour. H*i, H*, Hi and H identified peaks within the intonation contour, and thus the top line. The sentence initial and final target points, I and FL, were included because they stand for reference points for the F0 movements across the contours.

Linguistic span	*Am M*	*It-En M*	*It-It M*	*Am F*	*It-En F*	*It-It F*
I – L*	10	17	38	50	39	46
I – FL	28	27	50	62	60	52
H*I – L*	93	76	113	144	138	158
H*I – FL	112	87	125	156	159	164
H* – L*	28	57	46	129	84	182
H* – FL	47	68	58	141	105	188

Table 21. Overview of linguistic measures for span. Mean values for each landmark were calculated in Hz for male (M) and female (F) subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It).

For F0 span, selected measures were calculated to describe the pitch movements along the contours: I-L*, I-FL, H*i-L*, H*i-FL, H*-L*, H*-FL. As shown in tab. 21 landmarks such as Hi, H and L were not included in the measures for span because their values were less extreme than those of H*I, H* and L*. Results for span show that the widest pitch excursions were reached by the H*i-FL measure, while the narrowest span values were obtained by the I-FL measure.

The values obtained from the linguistic measures for males (fig. 21) and females (fig. 22) were averaged and plotted on a graph containing three lines: the Americans speaking English (the black line of diamonds), the Italians speaking English (the light grey line of squares), and the Italians speaking Italian (the dark grey line of triangles). The synthesized pitch lines show that the Italian males used similar patterns when speaking in their L1/L2. By contrast, the Italian females tried to approach the model of American English speakers.

In fig. 21, it is shown that the Italian males used similar F0 level in L1/L2 with a wider pitch span in their L1, as compared to their L2. The non-initial peaks (H* and H) have considerably lower values in the pitch pattern used by the American males.

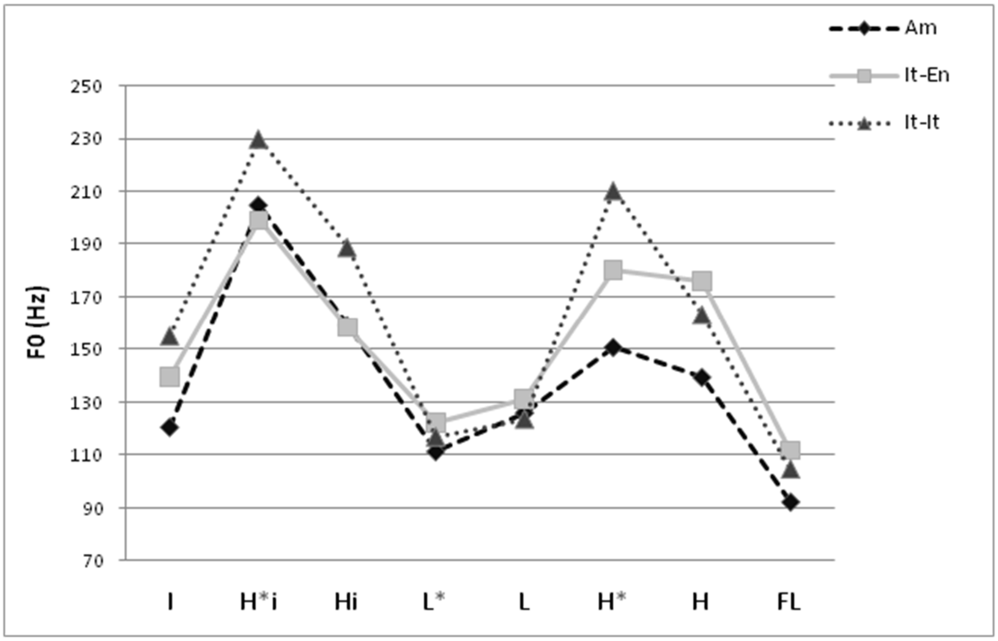


Figure 21. Linguistic measures by male speakers divided into three groups: American males speaking English (Am), Italian males speaking English (It-En), Italian males speaking Italian (It-It).

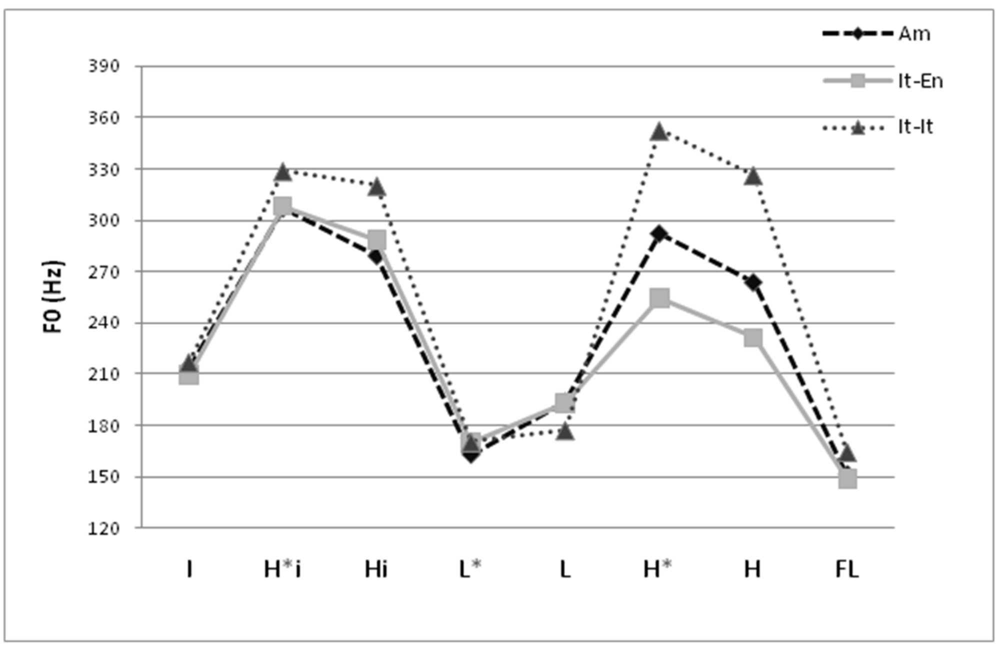


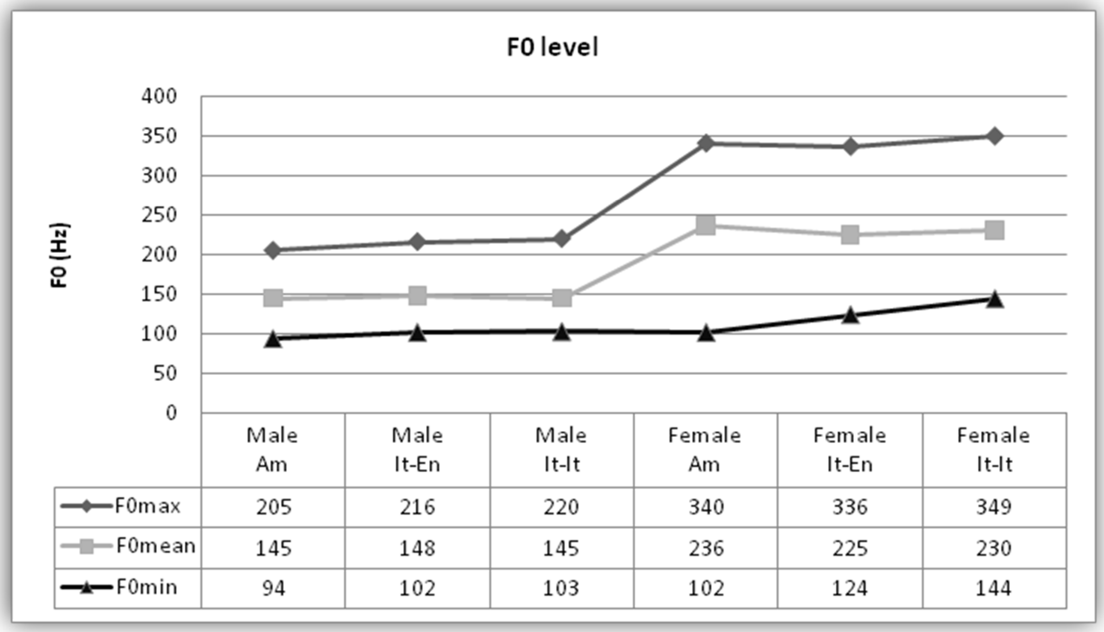
Figure 22. Linguistic measures by female speakers divided into three groups: American females speaking English (Am), Italian females speaking English (It-En), Italian females speaking Italian (It-It).

Contrary to males, the female speakers of this experiment tried to adapt their pitch patterns in L2 to the native speakers model. This can be inferred by the fairly similar values obtained for initial peaks (I, H*i, Hi) and valleys (L*, L, FL).

However, considerably different values were obtained for non-initial peaks (H* and H). As shown in fig. 22, the Italian females reached very high values in Italian (L1) and low values in English (L2). When speaking English, Italian subjects considerably lowered their non-initial peaks that obtained values even inferior to those of American females.

4.3.6.2 LTD measures

The graph in tab. 22 shows the distribution of LTD measures. F0 max, mean and min were calculated in Hz for the three language groups. This evidenced clear differences in pitch level patterns across genders and L1/L2.



	Male Am	Male It-En	Male It-It	Female Am	Female It-En	Female It-It
F0max	205	216	220	340	336	349
F0mean	145	148	145	236	225	230
F0min	94	102	103	102	124	144

Table 22. Graph containing F0 max, mean and min values in Hz by male and female subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It).

As it is clear from tab. 22, the differences across groups were much more extreme for the female than the male subjects. F0 level across males seems fairly similar, no matter the language. However, American males had lower values for F0 min (94 Hz) and F0 max (205

Hz), as compared to the Italians speaking in English (102 Hz for F0 min; 216 Hz for F0 max) and speaking Italian (103 Hz for F0 min; 220 Hz for F0 max). On a linear scale, the American men had a lower F0 range than the Italians with a difference of about 10 Hz. While the American women had very low F0 min (102 Hz, which incidentally is the same value for Italian males speaking English), the Italian women reached a F0 mean of 144 Hz when speaking Italian and 124 Hz when speaking English. Less dramatic F0 excursions occurred within F0 mean and F0 max values across the female groups.

The data for estimated marginal means of F0 max, mean, and min were plotted for males vs. females using the statistics program SPSS[11]. Values were distributed on the x-axis across three groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3).

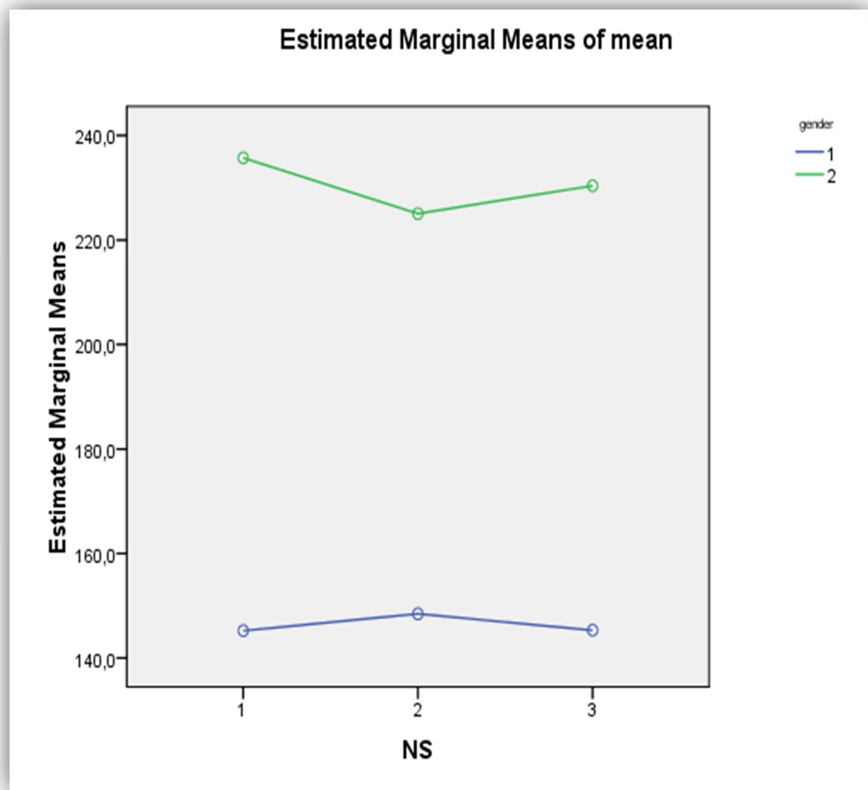


Figure 23. Estimated marginal Means of F0 mean distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects.

[11] Estimated marginal means are the unweighted means calculated in SPSS . In a study comparing the means of unequal sample sizes (as in ANOVA), it is necessary to take into consideration each mean in proportion to its sample size.

Estimated marginal F0 mean values were measured in SPSS and plotted in fig. 23. The green line (upper line) describes the estimated mean F0 mean values for females while the blue line (lower line) traces the mean F0 mean values for males. While F0 mean is about 10 Hz higher for the American females than the Italian females, the F0 mean values for the males are almost the same, with the Italian males speaking English (group 2) having slightly higher values.

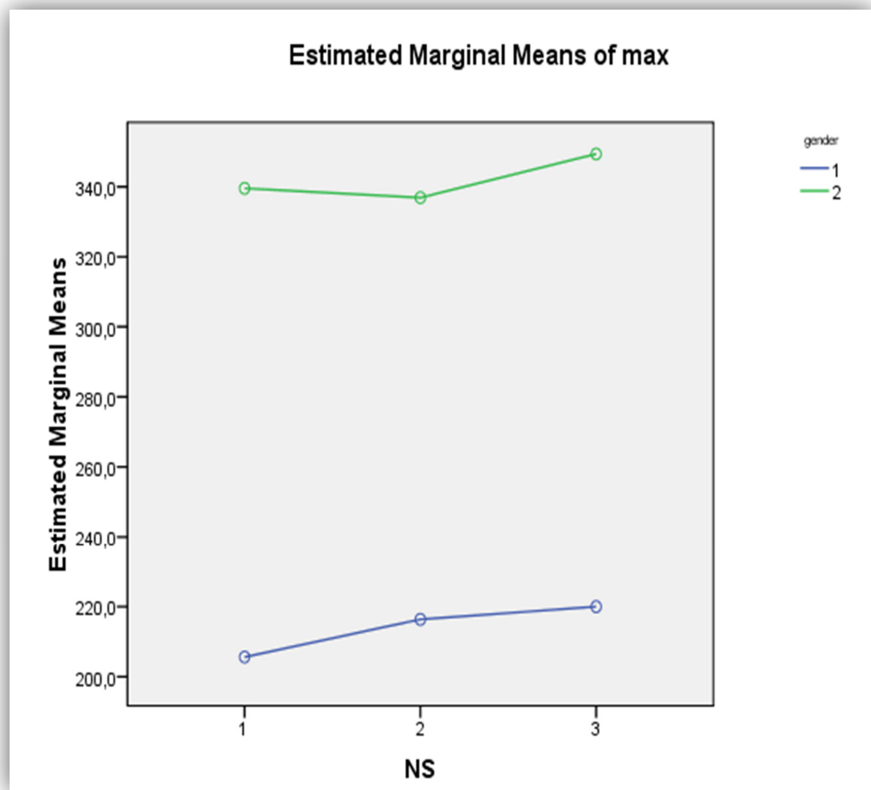


Figure 24. Estimated marginal Means of F0 max distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects.

The estimated marginal F0 max values, shown in fig. 24, are fairly similar across language groups. However, a completely different trend is noticeable in males vs. females. For the males (see blue line), the F0 max values are similar in Italians speaking English and Italian (i.e. groups 2 and 3) and are about 10 Hz higher than the F0 max values of the Americans (group 1). By contrast, for the females (green line), the F0 max values are almost the same for the Americans speaking English (group 1) and the Italians speaking English (group 2), with the values of Italian females speaking Italian (group 3) being higher than both Americans speaking English and Italians speaking English.

This shows that the Italian males, no matter which language they are speaking (English or Italian), have a higher F0 max than Americans, both in English and Italian. In particular, it looks like that the Italian males do not try to approach the American model by lowering their F0 max. On the contrary, Italian females considerably lower their F0 max when they are speaking English. This claim is supported by the fact that F0 max values for the Italian females speaking English (group 2) are significantly lower than that those of the Italian females speaking Italian (group 3). These contrastive trends across males and females are even more evident in the analysis of F0 min, in fig. 25.

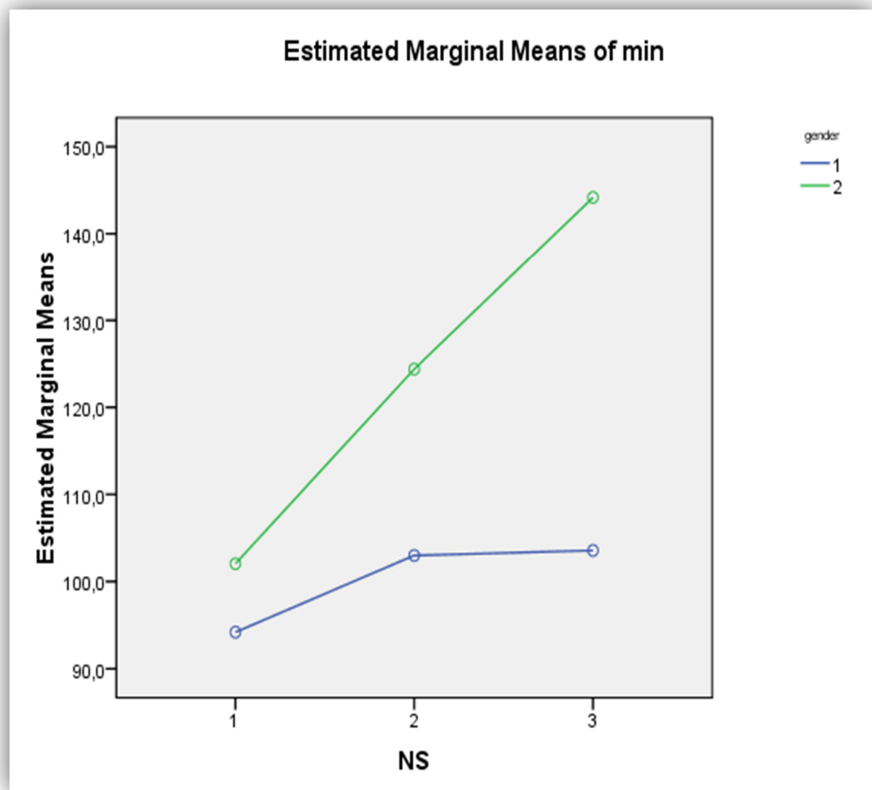


Figure 25. Estimated marginal Means of F0 max distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects.

The estimated marginal F0 min values are dramatically different across the females' language groups and fairly similar across the males' language groups. Thus, the trend shown in males vs. females is definitely opposite. The Italian males have the same F0 min in English (group 2) and Italian (group 3), and their F0 min is significantly higher than that of the American males (see blue line). By contrast, the Italian females reach higher F0 min in their L1 (group 3) than in their L2 (group 2). The F0 min value of the Italian females speaking English is about 20 Hz higher than that of the American females and 20 Hz lower than that of the Italian females speaking Italian (see green line). Once again, the Italian males show a strong influence of the L1 and use the same pitch patterns for English and Italian. On the contrary, females seem able to minimize the influence of the L1 transfer, by adopting pitch patterns in English that come close to those used by the American native speakers.

Data from the LTD measures were tested with repeated measures and one-way ANOVAs, in order to see whether or not differences across language groups were statistically significant. For level (mean F0), the between-subject factor 'gender' was highly significant for all measures. A one-way ANOVA showed that the factor 'gender' was significant across F0 mean ($F(119)=237.814$, $p<0.001$), F0 median ($F(119)=209.510$, $p<0.001$), F0 max ($F(119)=129.089$, $p<0.001$), and F0 min ($F(59)=17.283$, $p<0.001$). This result was expected, since there is anecdotic evidence about F0 differences between men and women. The between-subject factors 'native speaker' and 'language' did not reach significance for the F0 mean, the F0 median, and the F0 max; neither for the males nor for the females. Thus, no matter the language spoken and the native language of the speakers, the F0 level values calculated for F0 mean, F0 median, and F0 max were fairly similar for all groups. On the contrary, a one-way ANOVA showed that F0 min was significantly different across languages both for males and females. For the between-subjects factor 'native speaker', F0 min reached significance for males ($F(59)=4.531$, $p=0.038$) and for females ($F(59)=7.018$, $p<0.004$). Also for the between-subjects factor 'language', F0 min reached significance for males ($F(59)=20.435$, $p<0.001$) and for females ($F(59)=8.761$, $p<0.004$). Repeated measures showed that the within-subject contrast of F0 level*native speaker ($F(1)=8.17$, $p<0.005$) and F0 level*gender ($F(1)=116.76$, $p<0.000$) were significant. By contrast, the within-subject contrast of F0 level*language did not reach significance ($F(1)=1.81$, $p=0.181$). This suggests that the native language of the subjects (L1) plays a relevant role in pitch variation.

When comparing F0 span across sentences, it is fundamental to select the correct unit to measure these values. Previous studies claimed that logarithmic scales (e.g. ST) better than linear scales (e.g. Hz) manage to capture F0 span, by giving accurate evaluations of F0 intervals as they are perceived by the human auditory system (Daly and

Warren, 2001; Nolan, 2003; Mennen et al. 2012). Other studies show that the ST unit is a better predictor of F0 excursion than the Hz unit (for further discussion on differences between logarithmic and linear scales, see § 2.2.3). Thus, results for F0 span were calculated in Hz and ST and then compared.

Fig. 26 and 27 show the estimated marginal means calculated in Hz and ST, respectively.

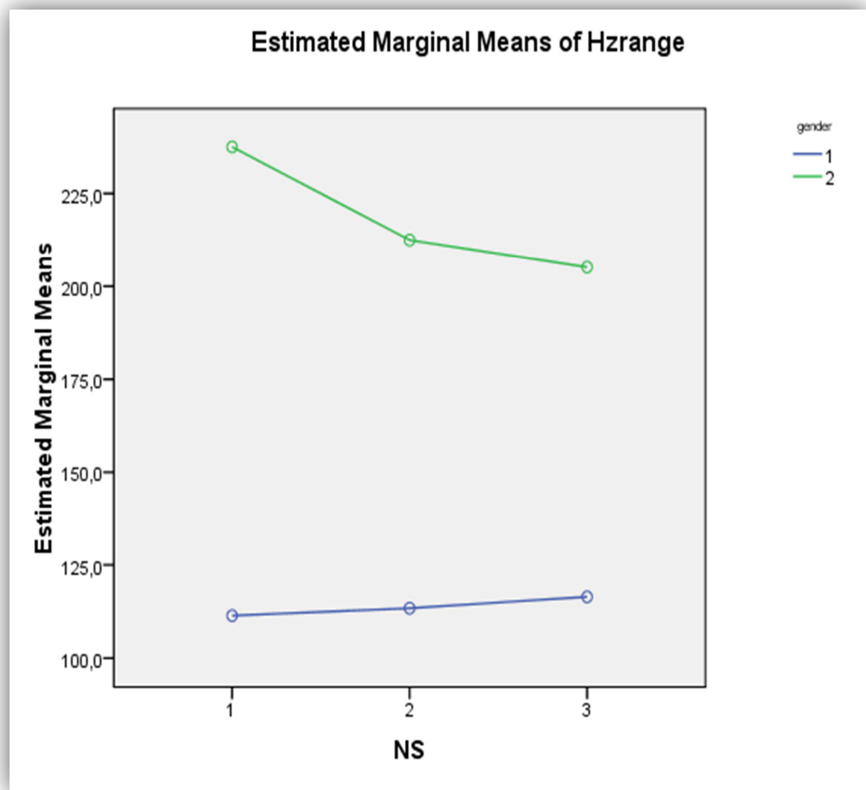


Figure 26. Estimated marginal Means of HZrange, that is F0 span calculated in Hz. Values are distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects.

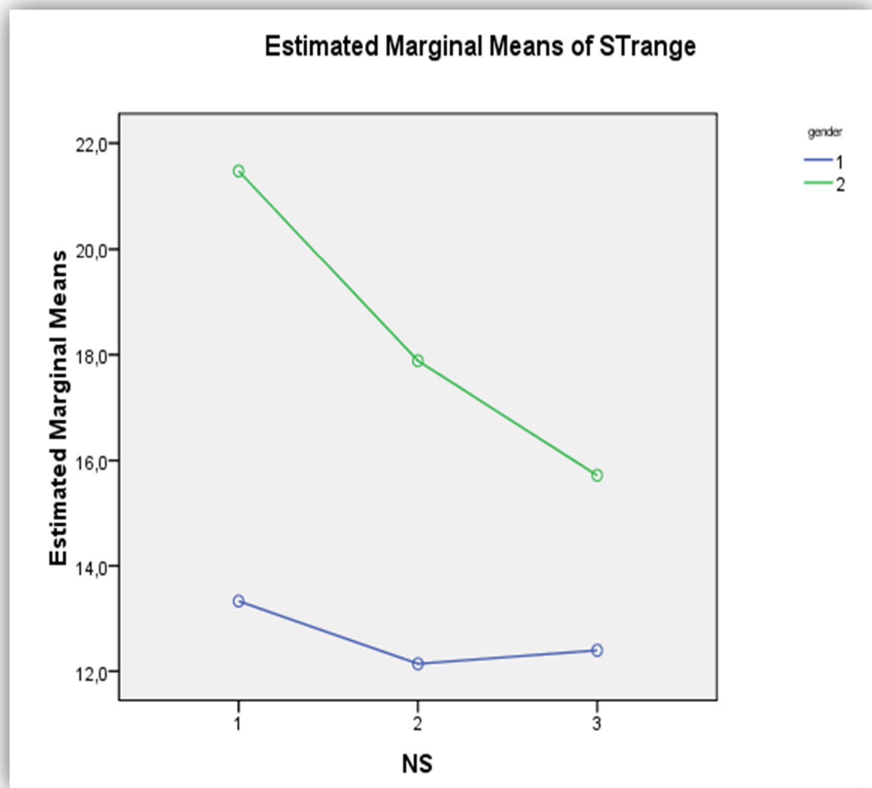


Figure 27. Estimated marginal Means of STrange, that is F0 span calculated in ST. Values are distributed on the x-axis across language groups: Americans speaking English (group 1), Italians speaking English (group 2), and Italians speaking Italian (group 3). The lower blue line (1) identifies values obtained from the male subjects while the upper green line (2) identifies values obtained from the female subjects.

By measuring F0 span in ST, it is possible to evidence a clear trend in span variation across language groups: the Americans (both males and females) produced sentences with higher F0 span than the Italians. In particular, the Italian females more than the males reached dramatically different values for their English speech as compared to their Italian speech (see values obtained for the groups 2 and 3 in fig. 27).

As shown in fig. 26, the graph measuring F0 span in Hz fails to capture these contrasts across groups. In fact, the variation of values measured in Hz across language groups is less weighty than the variation in ST: the F0 excursions across the data in ST are much more extreme than those calculated in Hz. Most importantly, the F0 span calculated in Hz data does not capture significant differences across measures obtained for the males. If one examines the F0 span of males (see blue line), the data in ST register a remarkable difference between the Americans and the Italians, with the Americans having higher values than the Italians. By contrast, the data in Hz do not provide any meaningful difference between the Americans and the Italians.

The results for F0 span in Hz appear to be quite contradictory when compared to the results measured in ST. This is due to the fact that the Hz unit is less suitable than the ST unit to measure F0 span (Daly and Warren, 2001; Johnson, 2003). ST is considered the most suitable and appropriate measure to compare F0 span (Nolan, 2003; Mennen et al., 2012) .

For these reasons, F0 span was calculated and compared in ST. The bars in tab. 23 show the distribution of the F0 values in ST, for the three language groups, evidencing clear differences in pitch span patterns across genders.

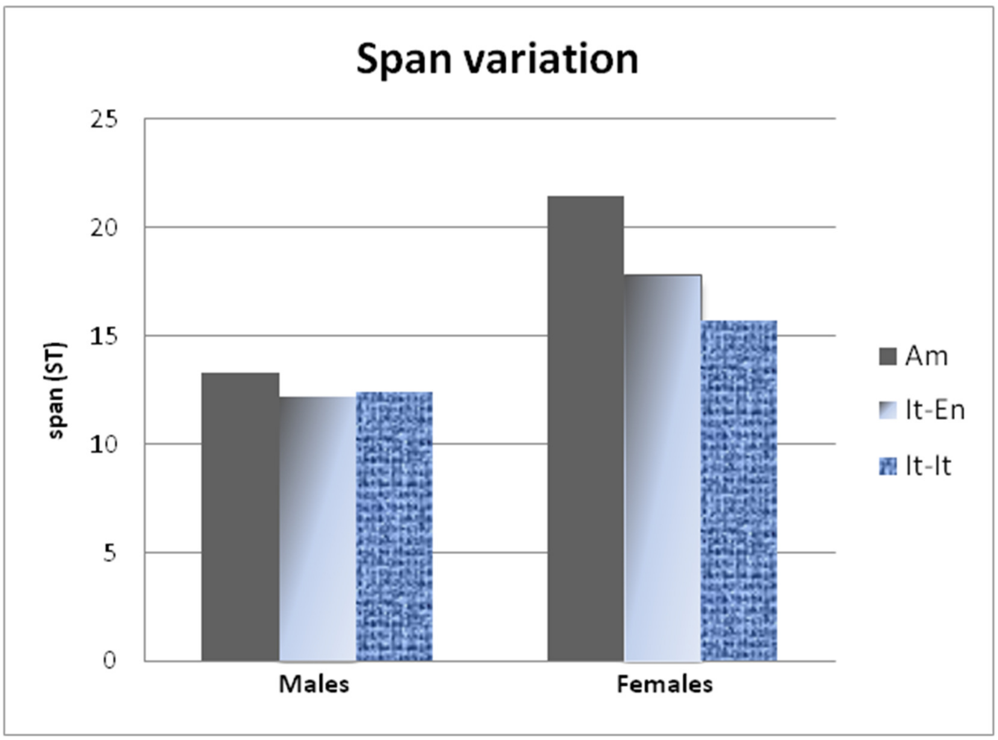


Table 23. Span values in male and female subjects divided into three groups: American speaking English (Am), Italian speaking English (It-En) and Italian speaking Italian (It-It).

The Italian males have fairly similar pitch span when speaking their L1 and L2, with span in Italian (12.39 ST) slightly larger than in English (12.14 ST). The American males' span (13,3 ST) is larger than that of the Italian males (compare the values above with 13.3 ST for the Americans). This difference across language groups is sensibly bigger among the females. The span values for the females are: 21.47 ST for the Americans; 17.88 ST for the Italians speaking English; 15.72 ST for the Italian speaking Italian. In sum, when speaking

English, the Italian females tried to approach the model of the American English native speakers by increasing their pitch span; the Italian males did not use this strategy (they even used a narrower span in English than in Italian).

Repeated measures and one-way ANOVAs were calculated for F0 span, to determine whether or not the differences across language groups were statistically significant. The between-subject factor ‘gender’ reached significance for F0 span calculated in ST ($F(119)=40.059$, $p<0.0001$). The between-subject factor ‘native speaker’, calculated for F0 span, was significant for the females ($F(59)=6.105$, $p<0.004$) but not for males ($F(59)=0,457$, $p=0.635$). Also the between-subject factor ‘language’, calculated for F0 span, reached significance for the females ($F(59)=4.125$, $p<0.025$) but not for the males ($F(59)=0.057$, $p<0.812$). This show the existence of a different trend among sexes. While the Italian women significantly varied F0 span depending on the language spoken (L1 vs. L2), the Italian males used a similar span in the two languages.

For skewness, the between-subject factor ‘gender’ was highly significant ($F(119)= 9.576$, $p<0.002$). Also for kurtosis, the between-subject factor ‘gender’ was statistically significant ($F(119)= 5.481$, $p<0.021$). The between-subject factor ‘language’ was statistically significant for skewness ($F(119)= 10.921$, $p<0.001$) but not for kurtosis ($F(119)= 0,533$, $p<0.467$). Also the ‘native speaker’ factor was statistically significant for skewness but not for kurtosis.

4.3.7 Discussion

The results of this first experiment show that Italians use higher pitch levels when speaking Italian and lower levels when speaking English. As for span, Italian females’ span is wider in English and narrower in Italian. By contrast, Italian males’ span is almost the same in L1 and L2.

The first hypothesis stating that L2 speakers have narrower F0 span than L1 speakers is confirmed for both Italian males and females, because the Italians showed narrower span than the Americans. The second hypothesis testing the impact of the L1 (Italian) transfer on L2 (English) is neither confirmed nor refused, due to different results across genders. While data from the male subjects support the idea that L1 has an influence on the pitch range used in L2, the results obtained by the female participants do not. The third hypothesis stating that pitch varies to the same extent in males and females was not confirmed. Contrary to males, the female speakers of this experiment tried to adapt their pitch patterns in L2 to the native speakers’ model. This is shown by the fairly similar values obtained from linguistic and LTD measures. Thus, the third hypothesis must be rejected.

The primary goal achieved in the first experiment was that of comparing language groups in order to support, with empirical data, the idea that pitch variation differs across languages and, more specifically, across L1 and L2. It was observed that pitch dynamism varied in English as an L1, English as an L2, and Italian as an L1. Pitch level showed fairly similar values across groups, with the Italian male speakers displaying slightly higher mean values for F0 min, max and mean than the American males. On the contrary, the female speakers got different results for pitch level. In line with the data obtained for the males, for the female speakers F0 min was higher in Italian L1 (144 Hz) and English L2 (124 Hz), than in English L1 (102 Hz). However, the data on F0 mean controvert this trend showing lower values in Italians L1 (230 Hz) and English L2 (225 Hz), than in English L1 (236 Hz).

This shows that pitch variation is much less predictable in females than in males. The results obtained for pitch span further support this idea by showing that the Italian males used similar pitch spans in English (12,14 ST) and Italian (12,39 ST) while the Italian females clearly diversified their pitch span when speaking English (17,88 ST) and Italian (15,72 ST). Generally speaking, the Americans (both males and females) used wider pitch span than the Italians. What is more, the Italian females better managed to reproduce the Americans pattern by widening their pitch span when speaking English.

The speakers in this study tried, at different levels, to reproduce the pitch range of native speakers of English. The results obtained in the first experiment suggest that the female better than the male participants succeeded in replicating the native speakers' pitch pattern. This generalization, however, needs to be confirmed by data from a larger sample of speakers, because I assume that F0 range in L2 may, to some extent, be influenced also by the speakers' competence and motivation in mastering an L2.

In line with previous research, the results of the present study show that 'F0 range is influenced by the phonological and/or phonetic conventions of the language being spoken and is not solely an artifact of physiological factors or cultural differences, as often assumed' (Mennen et al., 2012: 259). What is more, the discrepancy between the results obtained from the male and female subjects leads to the conclusion that gender plays a significant role in F0 variation. While utterances produced by the Italian male subjects had similar pitch range in Italian and English, the pitch range in utterances produced by the Italian female subjects varied considerably across languages. These differences may reflect the fact that 'cultural influences may be stronger for women's voice pitch than for that of men's' (Graddol and Swan, 1983). In addition, sociolinguistic factors may have an impact on the motivation and the intent to replicate the native speakers' model. This may lead some speakers (in this case, the female subjects) to perform better than others (in this case, the male the subjects) in the L2.

4.4 Second experiment

In the first experiment, the data obtained from male and female speakers shed light on the differences of pitch range variation across genders. Since females appeared to be more successful than males in replicating the native speakers model, this naturally leaded to the necessity to further examine data from female subjects in order to find greater empirical evidence on pitch variation in L1 and L2. For this reason, in the second experiment, I decided to focus attention on specific differences in pitch range variation within one population (i.e. females) and one language (i.e. English), spoken as L1 and L2. This choice has two advantages: 1) it eliminates the differences in pitch intervals due to gender (male vs. female) and 2) it provides a set of completely identical spoken materials, uttered in English. Working with the same language, analyzed as L1 and L2, has the advantage of obtaining highly comparable results.

The comparison of speech material produced in English and Italian may present difficulties because of the inherent differences in the two languages. These differences regard both segmental features (such as prevalence of voiced sounds over voiceless sounds in Italian, as compared to English) or prosodic factors (such as different number of syllables, placement of stress etc.). To overcome problems due to the segmental and suprasegmental differences between the two languages, in the second experiment I chose to examine only data in English. Thus, the productions of the same English sentences as produced by American native speakers and Italian learners of English were compared.

Even though the results obtained in the first experiment are reliable and robust, they were based on data from a limited number of participants (4 x language and 4 x gender). In order to increase the statistical significance of the data, in the second experiment the number of participants was increased to 20: 10 Italians and 10 Americans took part in the experiment, all females. Working with 20 subjects belonging to two homogenous groups (where subjects are females in their 20s-30s, university students with similar L2 proficiency) allows to compare data more effectively.

Based on the idea that pitch range varies to different extent depending on the pitch contours of sentences, I was interested in further examining pitch variation. In the second experiment, an analysis of distinct sentence types such as yes-no questions (henceforth YNQ), wh-questions (henceforth WHQ) and statements (henceforth STM) was carried out. The first experiment investigated the pitch level and span dynamics vary in a conversational passage, regardless of the types of sentences in that passage. The aim was to shed light on the differences of pitch range in Italian and English and to examine variation across genders. The results showed that the female speakers were more successful than the males in replicating the native speakers' model. To examine pitch variation further, the second experiment explores how pitch range varies locally, at the single sentence level, rather than globally, at the passage level (for the concepts of *local* and *global* pitch see Ullakonoja, 2007). Thus, the second experiment was designed to determine how pitch varies depending on sentence type.

The main features of the second experiment can be briefly summarized as follows:

Characteristics of the second experiment:

- Subjects: Only females
- Size of the corpus: increased number of subjects
- Languages: English as an L1 and L2
- Materials: three groups of different sentence type (YNQ vs. WHQ vs. EXM).

In the next paragraphs, some examples of YNQ, WHQ and STM are analyzed and interpreted in order to identify a specific model of pitch pattering for each sentence type produced in English as L1 and L2.

4.4.1 Yes-no questions (YNQ)

Polar questions (also called yes-no questions — YNQ) are usually pronounced with a rising tone that can be a high rise H* H-H% or a low rise L* H-H% (Levis, 2002). Since they presuppose a request for information or a demand for goods and services, the typical answers to YNQ are *yes* or *no*. Generally speaking, YNQ have a typical interrogative mode, expressed by a rising intonation contour. The incidence of rising patterns associated to YNQ was measured by Bolinger (1978) who asserted that 70% of questions were uttered with a final rise and 30% with a high pitch level. This study was carried out on samples of nearly 250 languages. In agreement with the data collected by Bolinger (1978), Hirst and Di Cristo (1998: 24) claimed that

> ‘a distinction between interrogative and declarative modes is one of the most universal characteristics of intonation systems […]. Even in the case of interrogative versus declarative, however, the nature of the distinction is far from uncontroversial since a question can be said to differ from a corresponding statement in its syntax, its semantics and its pragmatics, as well as any combination of the three’.

The focal point of pitch variation in YNQ is placed on the last part of the utterance, which, depending on the intonation contour used, usually have a high F0 max and F0 span. In fact, the biggest pitch excursion in a question is generally registered in the final part of the YNQ.

In American English, the typical pattern of YNQ is generally believed to be a H* H-H%, that is, a high rise, while the standard intonation pattern of British English YNQ is L* H-H%, that is, a low rise (Brown et al., 1980; Levis, 1996; Ladd, 1997; Cruttenden, 1997; Levis, 1999; Wells, 2006). However, depending on their pitch contours, YNQ can be uttered with different pitch levels and span. What it more, it has been claimed that differences in the contours of questions are a function of the pragmatic intent rather than of grammar or language variety (Pierrehumbert: 1980, Hedberg et al., 2008). In particular, speakers pronouncing YNQ with a falling contour appear to signal a greater certainty and self-confidence (Thompson, 1995; Hirschberg; 2002; Ramirez Verdugo, 2006). Regardless of the pragmatic context and the intent of the speaker, the focus of this study is to highlight how native speakers and learners of English produce pitch variation in YNQ. Thus, I am not directly concerned with the different syntactic, semantic and pragmatic implication in the intonation of YNQ.

This experiment challenges the idea that L1 and L2 speakers of English may use similar pitch variation. Even though it is generally agreed upon that most languages use similar patterns to utter YNQ (i.e. high and low rises), the impact of pitch variation plays a great role in the perception of differences in rising tones and may contribute to creating a mismatch among utterances pronounced by the speakers of L1 and L2 English. In fig. 28, three YNQ produced by American native speakers are compared to those produced by Italian learners of English. Six samples of the sentence 'Do you need any money?' were selected in order to give account of the basic differences across speakers' productions that are impressionistically evident in these samples.

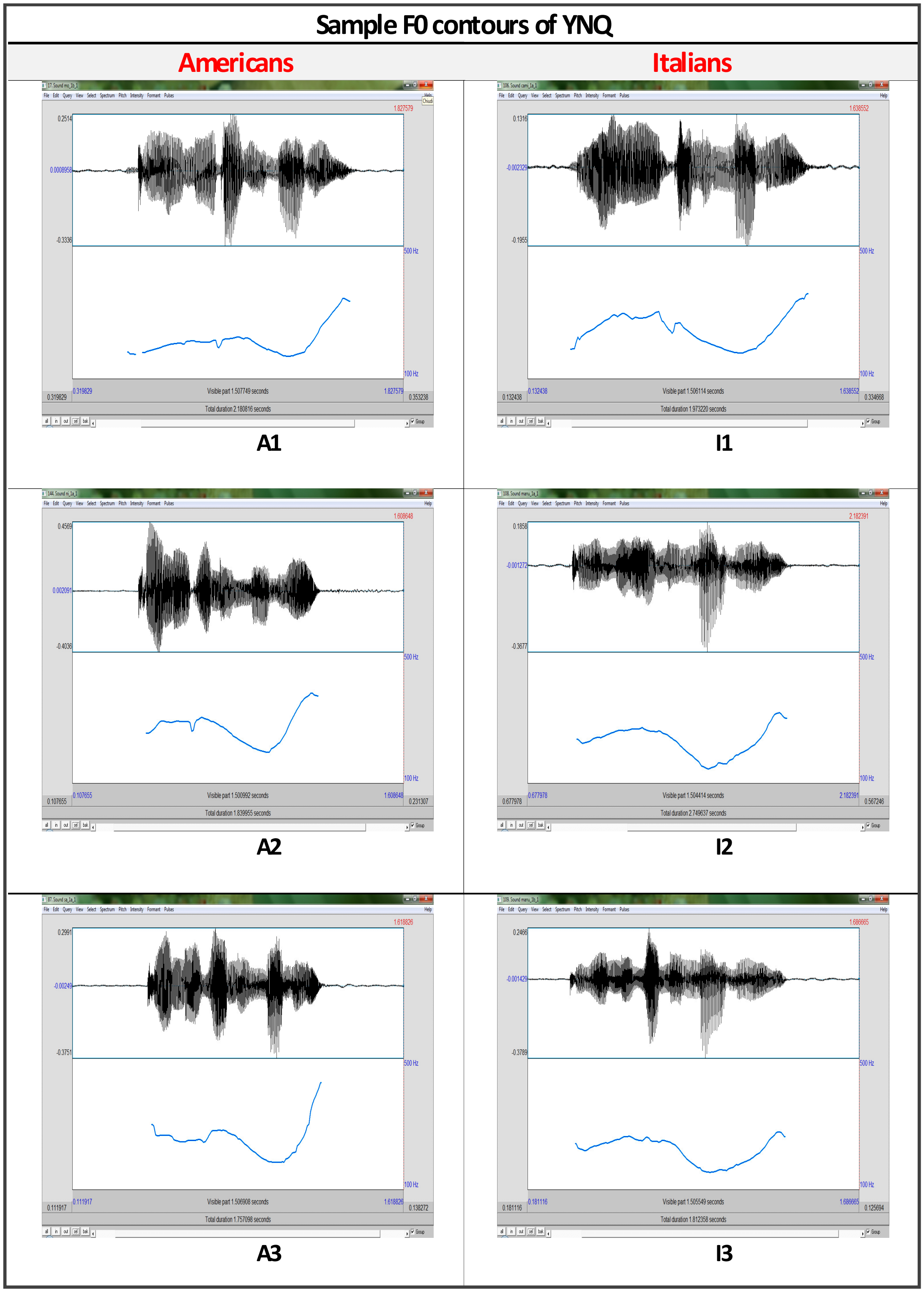


Figure 28. Patterns of yes-no questions (YNQ). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Do you need any money?' are compared.

Pitch contours of YNQ with waveforms lasting about 1,5 ms were compared. The pitch floor was set at 100 Hz and the pitch ceiling at 500 Hz. From a phonological point of view, the pitch pattern of these sentences is fairly similar, all of them being a low rise (L* H-H%). From a phonetic point of view, YNQ produced by the American speakers (e.g. A1, A2, A3) have a distinctive sharp rise at the end of the utterances and are relatively shorter in duration than YNQ uttered by the Italians (e.g. I1, I2, I3). When comparing A1 and I1, the F0max located in the final rise is almost equal: 341,6 Hz for A1 and 352,2 Hz for I1. By contrast, A2 and I2 reach quite different values in the final peak: 374,7 Hz for A2 and 314,3 Hz for I2. This contrast is even more evident in the last example, A3 vs. I3. The peak in A3 (427,5 Hz) is notably higher than that in I3 (275,4 Hz), with a difference of 152.1 Hz.

These impressionistic considerations need to be further examined in an experimental study (see results reported in § 4.4.7) in order to obtain reliable and robust data supported by statistical significance.

4.4.2 Wh-questions (WHQ)

Wh-questions (WHQ) are normally introduced by interrogative words such as *who*, *whom*, *which*, *what*, *when*, *where*, *whose*, *how* and require a more complex answer than just 'yes' or 'no'. Unlike YNQ, whose dominant pattern is the rising contour, WHQ are usually uttered with falling contours (Wells, 2006). According to Cruttenden (1997: 159), 'falls are the dominant pattern [for WHQ] in contrast to the rises associated with 'yes-no' questions'. Indeed, in languages such as English, Spanish, Romanian, Russian and Greek, the typical pattern for WHQ is considered more similar to that of STM than that of YNQ (Hirst and Di Cristo, 1998: 26). The standard intonation contours used in different sentence types have been briefly outlined as follows:

> 'A statement ends with a falling pitch. A question may end with a rising or a falling pitch. The two most common types of questions in English are: questions that ask for information with a question word (pitch falls). Questions that can be answered "Yes" or "No" (pitch rises)'.
> (Gilbert, 1993: 103)

Even though WHQ have a falling contour as the dominant pattern; the rising contour is an available alternative in cases where the speaker wants to show interest, sympathy and kindness or just wants to ask himself/herself a question (Pierrehumbert and Hirschberg, 1990; Wells, 2006; Steedman, 2007). What is more, reclamatory WHQ and echo WHQ (when speakers reiterate what has just been said because they either want to underline a concept or simply fail to understand something) tend to be pronounced with a rising contour (Bolinger, 1989; Halliday, 1994; Hedberg et al. 2010).

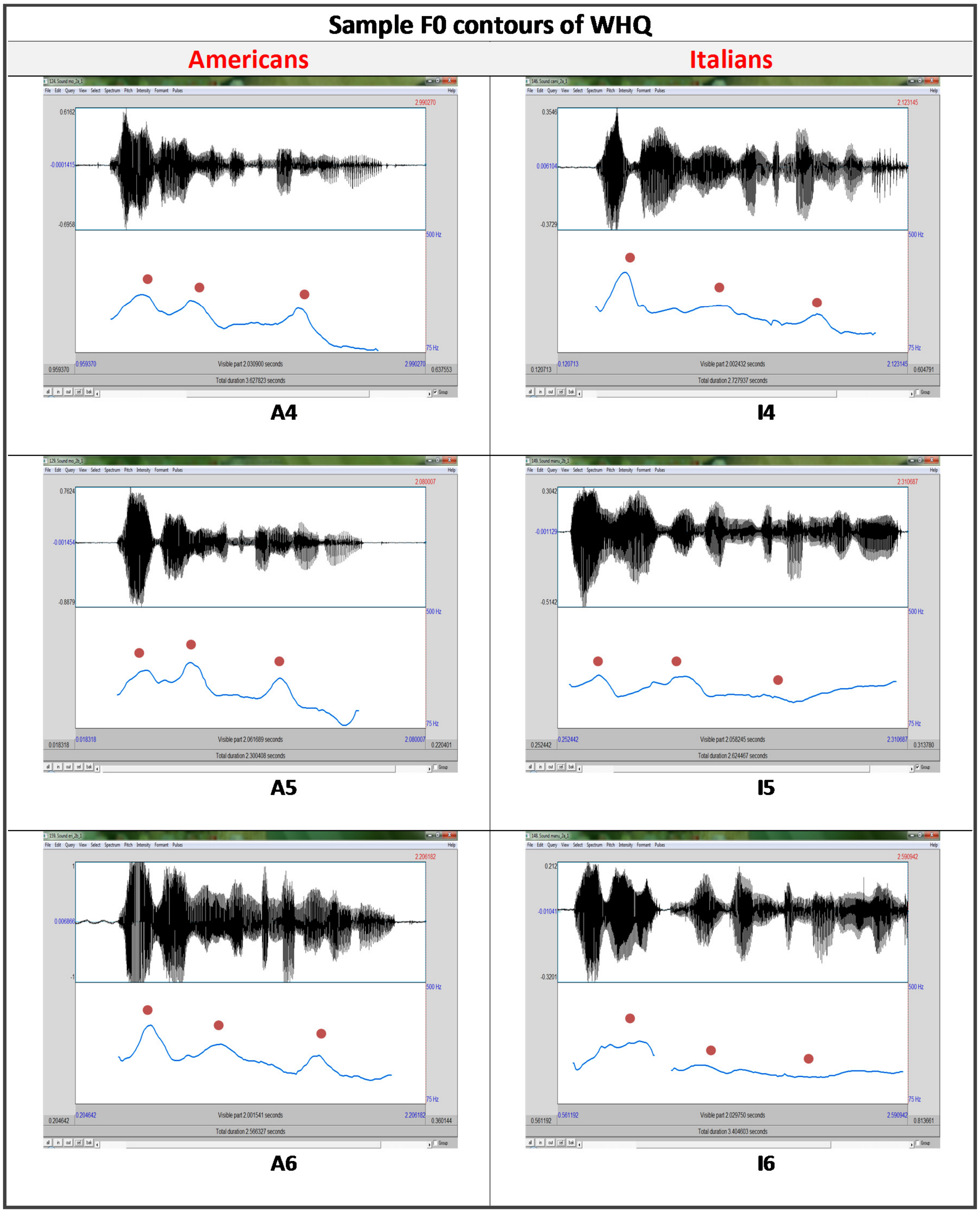


Figure 29. Patterns of wh-questions (WHQ). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Where were you when the money ran out?' are compared.

In fig. 29, three WHQ produced by American native speakers are compared to those produced by Italian learners of English. Six samples of the sentence ‘Where were you when the money ran out?’ were selected in order to give account of the basic differences across speakers’ productions.

Pitch contours of WHQ with waveforms lasting about 2,0 ms were compared. The pitch floor was set at 75 Hz and the pitch ceiling at 400 Hz. WHQ produced by the American speakers (e.g. A4, A5, A6) have a distinctive three-peaks pattern, where peaks are located at the beginning, in the centre and at the end of the utterances. These peaks are much sharper and more defined in the WHQ uttered by the Americans than the WHQ uttered by the Italians (e.g. I4, I5, I6). At first sight, the differences are evident by visually comparing pitch lines. Each peak is signaled by a red dot. The three pitch contours of the American subjects, A4, A5, A6, have three peaks each. By contrast, the Italian subjects pronounced the WHQ with fewer peaks: I4 and I5 have two distinct peaks; I6 has only one peak.

Not just the number of peaks but also their height contrasts between the two subjects groups. The peaks produced by the Americans appear to be characterized by higher and steeper slopes. This could be an effect of variation in pitch range (especially pitch span). In the second experiment, empirical evidence for this phenomenon is provided (see results reported in § 4.4.7).

4.4.3 Statements (STM)

STM can be uttered with a variety of pitch contours but they are frequently pronounced with a falling contour. According to Wells (2006: 25), ‘we say statements with a fall unless there is a particular reason to use some other tone. […] By using a fall we indicate that what we say is potentially complete, and that we express it with confidence, definitely and unreservedly’.

My data are in line with the claim that STM are commonly pronounced with a falling contour (Dauer, 1993; Gilbert, 1993; Grant, 1993; Hagen and Grogan, 1992; Orion, 1996). All the American and Italian subjects in this study pronounced the STM with a fall. Despite this similarity of patterns, STM are phonetically different, as far as pitch range, duration, speech rate are concerned.

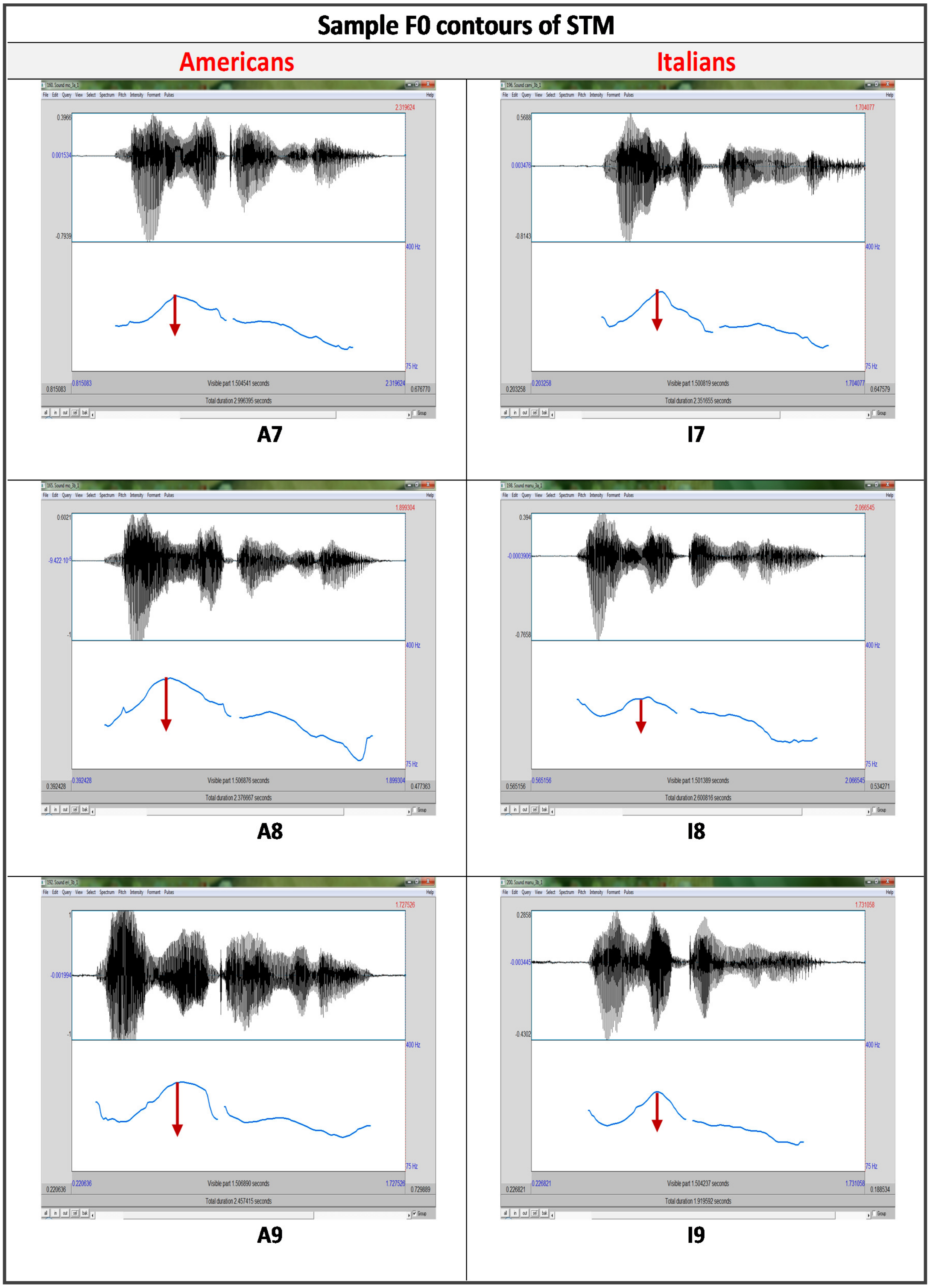


Figure 30. Patterns of statements (STM). Six versions (3 for the Americans and 3 for the Italians) of the sample sentence 'Now you are going away' are compared.

In fig. 30, three STM produced by American native speakers are compared to those produced by Italian learners of English. Six samples of the sentence 'Now you are going away' were selected in order to give account of the basic differences across speakers that are impressionistically evident in these samples.

Pitch contours of STM with waveforms lasting about 1,5 ms were compared. The pitch floor was set at 75 Hz and the pitch ceiling at 400 Hz. From a phonological point of view, the pitch pattern of these sentences is fairly similar, all of them being a falling contour (L+H* L-L%). From a phonetic point of view, the STM produced by the American speakers (e.g. A7, A8, A9) have a distinctive sharp rise at the beginning of the utterances and are relatively longer in duration than the STM uttered by the Italians (e.g. I7, I8, I9). In fact, the initial peaks not only look steeper in the utterances by the Americans but also longer. Thus, they seem more extended in time and width. Tab. 24 shows the values obtained for F0max, F0min, span, and duration in the speakers' productions.

	A7	**I7**	**A8**	**I8**	**A9**	**I9**
F0max	264.1 Hz	274.3 Hz	305.6 Hz	257.3 Hz	296.4 Hz	271.9 Hz
F0min	187.2 Hz	181.9 Hz	183.7 Hz	209.1 Hz	197.6 Hz	188.2 Hz
span	76.9 Hz	92.4 Hz	121.9 Hz	48.2 Hz	98.8 Hz	83.7 Hz
duration	0.500sec	0.498 sec	0.583 sec	0.452 sec	0.579 sec	0.450 sec

Table 24. Values for F0max, F0min, span, and duration of the STM 'Now you are going away', compared across language groups: American speakers (A7, A8, A9) and Italian speakers (I7, I8, I9).

When comparing A7 and I7, the pitch span (calculated as F0max- F0min and visually represented by a red arrow) of the initial peak is 76,9 Hz for A7 and 92,4 Hz for I7. The duration of the rise-fall almost identical: 0,500 sec for A7 and 0,498 for I7. By contrast, A8 and I8 reach quite different values in the span of the initial peak: 121,9 Hz for A8 and 48,2 Hz for I8. This contrast is even more evident in the duration of the two sample utterances. The rise-fall in A8 lasts 0,583 sec while the one in I8 lasts 0,452 sec. The span of the peak of A9 (98,8 Hz) is higher than that of I9 (83,7 Hz). Also the duration of the rise-fall is different: 0,579 sec for A9 and 0,450 for I9.

More analyses are needed to obtain reliable and robust data supported by statistical significance.

4.4.4 Research Questions

The aim of the present experiment is to find out whether or not pitch range considerably varies depending on sentence type (e.g. YNQ vs. WHQ vs. STM) and, most importantly, whether or not the English sentences produced by Americans have a pitch range similar to that of sentences produced by Italians.

The first hypothesis being tested has to do with the assumption that different intonation contours correspond to different pitch range. Pitch range (level vs. span) was measured for each utterance separately. Then, values were grouped depending on sentence type. Since it is generally agreed upon that YNQ, WHQ, and STM are uttered with typical intonation contours (see 4.4.1, 4.4.2, and 4.4.3), I expect that the dominant patterns for each sentence type may have an influence on pitch variation. In particular, due to the fact that YNQ are usually uttered with rising contours, they are expected to display more pitch variation that STM. For WHQ, to make previsions is an hazard because, unlike YNQ, WHQ are commonly uttered with falling contours. It is not known to what extent Italians are aware of this difference between YNQ and WHQ. Thus, it could be the case that Italians wrongly pronounce WHQ with rising patterns. This could have an effect on pitch range variation. Indeed, it has been shown that, YNQ, WHQ, and STM are likely to have a different pitch range variation (Busà and Urbani, 2011).

The second hypothesis being tested is that Italian learners of English are influenced by their L1, thus transferring the L1 pitch range variation into their L2. On the basis of the results obtained in the first experiment (see § 4.3.6), the English sentences uttered by the Italians are expected to have a narrower pitch span and higher pitch level than those produced by the Americans. Thus, L2 English sentences might have a slightly monotonous tone and rather flat pitch contour, due to the prosodic transfer from L1. If this hypothesis is confirmed, the pitch range shown by the Italians in their L2 speech will be much less dynamic and varied than the pitch range shown by the Americans in their L1 speech.

Thus, the null hypotheses being tested are:

(1) The F0 level and span values are the same across sentence types (YNQ vs. WHQ vs. STM),
(2) The F0 measures for F0max, F0mean, F0min, and STrange are the same in the two native language groups (Americans vs. Italians).

4.4.5 Subjects

Ten female adult native speakers of American English and 10 female adult native speakers of Italian volunteered for the second experiment. In table 25, personal information about the Americans is shown.

All the American participants were speakers of American English, they all came from California and were students at the University of California – Los Angeles. All the American subjects were also proficient in Italian at different levels and they had been living in Italy for several months at the time of the recording, by taking part into exchange programs held in Padova, Bologna, Florence and Milan.

Sp	**NS**	**Sex**	**Age**	**Smoke**	**Impair**	**Major**
1	American	F	20	no	no	English and Italian
2	American	F	20	no	no	History
3	American	F	20	no	no	Linguistics and Spanish
4	American	F	21	no	no	Italian Film
5	American	F	21	no	no	Psychology
6	American	F	22	no	no	Literature and Italian
7	American	F	24	no	no	Italian
8	American	F	26	no	no	Italian Literature
9	American	F	20	no	no	Literature and Italian
10	American	F	21	no	no	Marine Science

Table 25. Personal data of the ten American participants (Sp) in the experiment. Information is on their native language, their sex, their age, status of smoker/non smoker, presence of any reported language disorder or impairments, and their major.

In table 26, the Italian speakers' personal information is shown. All the Italian speakers were either graduate students at the University of California – Los Angeles or graduate students at the Università degli Studi di Padova. They all came from northern Italy (Veneto area) and only three of them spoke dialect in their daily life. Their dialectal inflection was not auditorily detectable. So, no distinction of any dialectal information was retained. Two out of the 10 Italian subjects have been living in Los Angeles for more than one year.

Sp	NS	Sex	Age	Smoke	Impair	Major
1	Italian	F	28	no	no	French
2	Italian	F	31	no	no	English Literature
3	Italian	F	29	no	no	English Literature
4	Italian	F	28	no	no	English Linguistics
5	Italian	F	28	no	no	English Literature
6	Italian	F	27	no	no	Linguistics
7	Italian	F	29	no	no	English
8	Italian	F	22	no	no	Italian
9	Italian	F	27	no	no	Linguistics
10	Italian	F	29	no	no	English Linguistics

Table 26. Personal data of the ten Italian participants (Sp) in the experiment. Information is on their native language, their sex, their age, status of smoker/non smoker, presence of any reported language disorder or impairments, and their major.

All the American participants were university students in different departments at UCLA: Italian, as a major or minor (6 students); Linguistics (1 student); History (1 student); Psychology (1 student); Marine Science (1 student). Eight students were undergraduate while two students were doing a Master degree. The age of the participants ranged from 20 to 26 years (mean age: 21,5 years). None of the speakers reported any speech, hearing or communication disorder and they were all non-smokers. There was no screening for formal training in music or singing.

All the Italian participants were university students in different departments at the University of Padova and UCLA: English literature and/or linguistics (6 students); Linguistics (2 students); Italian (1 student); French (1 student). All the Italian subjects but one were graduate students; two students were doing a Master degree while seven students were pursuing a Ph.D. The age of the participants ranged from 22 to 31 years (mean age: 27,8 years). Other than age, homogeneity of the subjects was controlled for the Italian variety they spoke: all the participants in the experiment were speakers of the Northern Italy Italian variety (Veneto area).

The experimenter personally knows all the participants she recruited for the experiment. They gave their consent for the treatment of their personal data (see consent form in Appendix B) and volunteered for the experiment without receiving any monetary compensation.

In tables 27 and 28, some additional information about exposure of the subjects to L2 is provided. In particular, I collected data about the number of years of L2 learning; the level of proficiency in L2 (based on the subject's own self-assessment); competence in other foreign languages; length of residence (expressed in months) in the United States, for the Italian subjects, in Italy, for the American subjects; percentages of use of Italian and English on a daily basis. Subjects were asked to rate how often they watched movies or TV programs broadcasted in their L2: 75% of Italians answered that they watch English programs at least once a week, only 20% of Americans regularly watch Italian programs.

Sp	***NS***	***Years of Italian***	***Spoken languages Level of proficiency in L2***	***Time Abroad***
1	Am	1	Italian (beginner), Spanish (intermediate)	<1
2	Am	3	Italian (intermediate), Spanish (beginner)	<1
3	Am	4	Italian (intermediate), French (intermediate), Spanish (beginner)	<1
4	Am	3	Italian (intermediate), Spanish (intermediate)	<1
5	Am	4	Italian (intermediate), Spanish (beginner), German (advanced)	<1
6	Am	4	Italian (intermediate), French (beginner)	<1
7	Am	3	Italian (intermediate), French (advanced), German (beginner)	<1
8	Am	3	Italian (intermediate), French (proficient)	<1
9	Am	3	Italian (intermediate), Spanish (advanced)	<1
10	Am	3	Italian (intermediate), Spanish (advanced)	<1

Table 27. Personal data of the ten American speakers (Sp) in the experiment. Information is on their native language, years of L2 learning (Italian as an L2 for the Americans), languages learned and spoken at different levels, length of residency in a L2 country (Italy for the Americans).

When inspecting the information about competence in L2 (years of learning, proficiency level, and time spent abroad), a great homogeneity in the group of Americans was found (table 27). In particular, all the American subjects were proficient in Italian. They were students of Italian (mostly at the intermediate level) and spent several months in Italy to practice their language skills (always periods inferior to one year, that is <1). One of the aim of this study is to establish to what extent there is an influence of L1 and L2 in pitch variation, and this is done by comparing bilinguals. If American English monolinguals would be mixed with American/Italian bilinguals this could lead to the production of erroneous, if not misleading and unreliable, results. Thus, it is important to focus on data produced by either bilinguals or monolinguals.

Sp	*NS*	*Years of English*	*Other spoken languages Level of proficiency in L2*	*Time Abroad*
1	It	15	English (advanced), French (near-native), Spanish (proficient)	>1
2	It	14	English (advanced), Spanish (advanced), French (upper-intermediate)	<1
3	It	12	English (upper-intermediate), Spanish (advanced), French (beginner)	<1
4	It	13	English (advanced), French (near-native), Spanish (intermediate)	<1
5	It	16	English (near native), Spanish (advanced), German (advanced)	<1
6	It	15	English (advanced), French (advanced)	<1
7	It	16	English (advanced), German (near-native)	>1
8	It	15	English (advanced), French (intermediate), German (beginner)	<1
9	It	15	English (advanced), German (near-native)	<1
10	It	15	English (advanced), German (near-native), French (intermediate), Albanian (beginner)	<1

Table 28. Personal data of the ten Italian speakers (Sp) in the experiment. Information is on their native language, years of L2 learning (English as an L2 for the Italians), languages learned and spoken at different levels, length of residency in a L2 country (the United States for the Italians).

When inspecting the information about competence in L2 (years of learning, proficiency level, and time spent abroad), a great homogeneity was found in the group of Italians (table 28). In particular, all the Italian subjects were proficient in English. They were students of Italian (mostly at the advanced level) and spent several months in the United States to practice their language skills. Eight students had spent less than one year (that is, <1) in the United States while two students had been living for more than one year (that is, >1) in the United States, at the time of the recording.

In sum, the total number of subjects selected for the experiment was 20: 10 American subjects and 10 Italian subjects. They were all females, in the same age and with similar competence in English and Italian.

4.4.6 Materials

This study compares native and non-native productions of 15 English sentences produced by 10 American and 10 Italian subjects. In the data set, all the sentences were divided into three groups depending on sentence type: 5 YNQ, 5 WHQ, and 5 STM. Every sentence was read by each participant at least twice (when the subjects were misreading a sentence, more repetitions were necessary). Only two repetitions were retained for every sentence.

In table 29, the 15 sentences created for the analysis of pitch range are shown:

	English sentences
Yes/no questions	Do you need any money? Have we met before? Are you still there? Can you open the door? Do you wanna come for dinner?
Wh-questions	Where were you when the money ran out? Why are you selling meat? What was her name again? What are you doing there? What's wrong with you?
Statements	Now you are going away. I hope I can see you on Monday. We should go and visit your uncle. I know you are leaving today. You should go to Hawaii.

Table 29. English sentences for the analysis of pitch range grouped according to sentence type: yes/no questions, wh-questions, and statements.

The materials created for the present experiment had to conform to specific standards. Sentences had to be short and they had to contain easy to pronounce words. Voiced sounds were prevalent over voiceless sounds because Praat fails to capture the pitch track of voiceless sounds. In particular, the three groups of sentences had the prosodic characteristics of different sentence types: YNQ vs. WHQ vs. STM.

The corpus created consisted of 600 utterances (5 sentences x 20 speakers x 3 language groups x 2 repetitions).

4.4.7 Procedure

In the experiment, the subjects were asked to read aloud short sentences in a natural way. The text was read aloud by 10 American English female speakers from California and 10 Italian female speakers from the North East of Italy (Veneto area). Both the American English and Italian subjects read the materials in English. Data were extracted from two different groups: (1) Americans speaking English, (2) Italians speaking English.

The materials were collected by the experimenter in two separate sessions. Part of the recordings were gathered together in a three weeks period at the Linguistic Department of the University of California – Los Angeles. The audio files were recorded and digitally acquired in a sound-attenuated booth in the UCLA Phonetic Laboratory. Another part of the recordings was acquired in Padova, during a recording session (lasting one day). In Padova, a sound-attenuated booth was not available. However, almost perfect conditions were created for the recording session that took place in a quiet and isolated room.

Recordings were saved into WAV audio files. They were collected using a Shure SM10 head-mounted microphone, recorded direct-to-disk on another computer located outside the sound booth, and digitized at a sample frequency of 44.1 KHz and a 32 bit quantization rate, using an AudioBox. By using a sampling rate of 44.1 KHz (i.e. CD quality), it was possible to collect data with excellent quality. After recording the short sentences, the experimenter saved the data and labeled them as separate WAV audio files with Praat (Boersma and Weenick, 2010).

During the recording session, subjects were instructed to read sentences with a natural conversational intonation. No indication was given about the intonation they had to use in the different types of sentence. The three sets of sentences (YNQ vs. WHQ vs. STM) were read in English, every sentence was recorded twice. Therefore, a set of two repetitions of all materials was recorded. The order of materials in the recording session was the same for every speaker. They read all the sentences in a sequence: first, list of YNQ; second, list of WHQ; third, list of STM.

The subjects were constantly monitored by the experimenter while they were reading the sentences and they were required to repeat any sentence when they misread it. In some cases, when the speakers did not feel comfortable with the utterance pronounced, they asked to do the recording again. Before starting with the recording process, subjects were permitted to read silently the sentences in order to familiarize themselves with them before reading them aloud for recording. Each recording session lasted about 20 minutes. At the end of the session, every subject was asked to fill in a questionnaire, containing questions about their personal information (see Appendix B for a sample of the questionnaire). Speakers were requested to indicate their first and last name; age; birth place; sex; native language; second languages and proficiency levels; university status; periods abroad; ways and daily use of learning a second language.

4.4.8 Method

The data were analyzed by following the method proposed by Mennen et al. (2012). Thereby pitch values were analyzed and compared across groups by calculating long-term distributional (LTD) and linguistic measures. The methodology used to measure pitch range in the second experiment was the same I adopted in the first experiment (see § 4.3.5).

Linguistic measures were calculated by manually annotating every sentence. The labels used in the second experiment were identical to those used in the first experiment. The beginning and the end of every sentence were marked with a I for the initial pitch and a FL for a final low pitch or a FH for a final high pitch. Local peaks on prominent syllables were marked as H* while local peaks on non-prominent syllables were marked as H. When peaks were placed at the phrase starting point they were labeled with an additional 'i' to signal their initial position, e.g. H*i and Hi. Valleys never appeared in initial position and they were labeled according to their prominence status. Local valleys on prominent syllables were identified by the label L* while local valleys on non-prominent syllables were identified by the label L.

LTD measures were based on the analysis of F0 distribution. Values of F0 maximum (F0max), F0 minimum (F0min), F0 mean (F0mean) and F0 median (F0median) were calculated over the entire sentences to measure pitch level. Measures analyzed for pitch span were: F0 maximum minus F0 minimum (max-min F0) in Hz and ST. The selection of LTD measures for the analysis of pitch level and span was based on Mennen et al.'s approach (2012), see (see § 1.3.4). Values for LTD measures were obtained automatically by inquiring pitch information in Praat such as minimum, maximum, range, average, standard deviation, etc. The same protocol was used to calculate all measures using the same standards and procedure. Pitch tracking was performed with a standard algorithm based on the autocorrelation method. This algorithm is the standard option to process speech and detect pitch locations in Praat.

In addition to linguistic and LTD measure, also the speaking rate and the duration of YNQ vs. WHQ vs. STM were measured. In fact, while examining data for pitch range I noticed that sentences produced by the Italians were generally longer than those produced by the Americans. For this reason, the duration and speaking rate of sentences may be a correlate of pitch range variation. To date, there is not a study reporting the existence of a correlation between duration and pitch range of sentences. However, some perception studies (for a review, see ‘t Hart et al., 1990) showed a great influence of duration in the perception of foreign accent. In order to shed light on the effect of measures of duration on pitch range variation, several measures are tested in the present experiment: number of syllables (nsyll); average syllable duration (ASD); duration (Dur); phonation time (PhT); speech rate (SR); articulation rate (AR).

4.4.9 Results

The following sections present an analysis of 600 sentences that are divided into three sentence types: yes-no questions, wh-questions, and statements (200 YNQ, 200 WHQ, and 200 STM). Thus, this study provides an overview of how pitch range variation is associated to specific sentence types in English as an L1 and L2. The results were obtained from the elaboration of data on linguistic measures, LTD measures, speaking rate of utterances.

By systematically comparing pitch range of YNQ, WHQ and STM, a three-way contrast in their F0 realizations is described. Questions and statements exhibit contrasting pitch range values depending on the L1 of the speakers (American English vs. Italian).

4.4.9.1 Linguistic measures

After placing linguistic landmarks to peaks and valleys, a script in Praat was used to calculate the F0 of every pitch point. The values were averaged across speakers and language groups (the Americans speaking English vs. the Italians speaking English). The values of linguistic measures were calculated for pitch level (tab. 30) and pitch span (tab. 31). In table 30, measures calculated for level were grouped depending on the native language of the speakers and the sentence types.

	YNQ		WHQ		STM	
Linguistic level	*AmE*	*It*	*AmE*	*It*	*AmE*	*It*
I	271	189	241	267	207	227
Hi	318	301	332	318	299	272
H*i	278	266	358	346	410	350
H*	265	269	294	315	378	277
H	159	193	265	332	271	248
L*	133	177	146	248	155	192
L	182	229	132	233	139	233
FH	416	353	_	_	_	_
FL	_	_	114	253	126	187

Table 30. Overview of linguistic measures for level. Mean values for each landmark were calculated in Hz for the American English (AmE) and the Italian (It) female subjects. Data were divided for each label into three groups depending on the sentence types: YNQ, WHQ and STM.

For F0 level, L*, L and FL were counted as the measures of valleys, that is, the bottom line of the pitch contour. H*i, H*, Hi and H identified peaks within the intonation contour, that is, the top line of the pitch contour. The sentence initial target point, I, and final target points, FH and FL, were included because they stand for reference points for the F0 movements across the contours. As shown in tab. 30, the landmark FL was not included in the measures of the YNQ because YNQ are characterized by intonation patterns ending in final rises (FH) and not in final falls (FL). By contrast, the WHQ and the STM did not show measures for the landmark FH because WHQ and STM are characterized by intonation patterns ending in final falls (FL) and not in final rises (FH).

In the YNQ, the highest F0 values were reached at the FH target point (i.e. 416 Hz for the Americans and 353 Hz for the Italians). The lowest F0 values were reached at the L* target point (i.e. 133 Hz for the Americans and 177 Hz for the Italians). In the WHQ, the highest F0 values were reached at the H*i (i.e. 358 Hz for the Americans and 346 Hz for the Italians). The lowest F0 values were reached at the FL target point for the Americans (i.e. 114 Hz) and at the L target point for the Italians (i.e. 233 Hz). In the STM, the highest F0 values were reached at the H*i target point (i.e. 410 Hz for the Americans and 350 Hz for the Italians). The lowest F0 values were reached at the FL target point (i.e. 126 Hz for the Americans and 187 Hz for the Italians).

	YNQ		WHQ		STM	
Linguistic span	*AmE*	*It*	*AmE*	*It*	*AmE*	*It*
I – L*	138	12	95	19	52	35
H*i – L*	145	89	212	98	255	158
H* – L*	132	92	148	67	223	85
FH – L*	283	176	_	_	_	_
I – FL	_	_	127	14	81	40
H* – FL	_	_	180	62	252	90
H*i – FL	_	_	244	93	284	163

Table 31. Overview of linguistic measures for span. Mean values for each landmark were calculated in Hz for the American English (AmE) and the Italian (It) female subjects. Data were divided for each label into three groups depending on the sentence types: YNQ, WHQ and STM.

For F0 span, targets points were calculated to describe the pitch movements along the measures: I-L*, H*i-L*, H*-L*, FH-L*, I-FL, H*-FL, H*i-FL. As shown in tab. 31, landmarks such as Hi, H and L were not included in the measures for span because their values were less extreme than those of H*I, H* and L*.

In the YNQ, results for span show that the widest pitch excursions were reached by the FH-L* measure (i.e. 283 Hz for the Americans and 176 Hz for the Italians), while the narrowest span values were obtained by the H*-L* measure for the Americans (i.e. 132 Hz) and by the I-L* measure for the Italians (i.e. 12 Hz). In the WHQ, results for span show that the widest pitch excursions were reached by the H*i-FL measure for the Americans (i.e. 244 Hz) and by the H*i-L* measure for the Italians (i.e. 98 Hz). The narrowest span values were obtained by the I-L* measure for the Americans (i.e. 95 Hz) and by the I-FL measure for the Italians (i.e. 14 Hz). In the STM, results for span show that the widest pitch excursions were reached by the H*i-FL measure (i.e. 284 Hz for the Americans and 163 Hz for the Italians), while the narrowest span values were obtained by the I-L* measure (i.e. 52 Hz for the Americans and 35 Hz for the Italians). The figures 31-33 show the linguistic measures calculated for the pitch patterns of YNQ, WHQ, and STM. The values obtained for the Americans are plotted against the values obtained for the Italians.

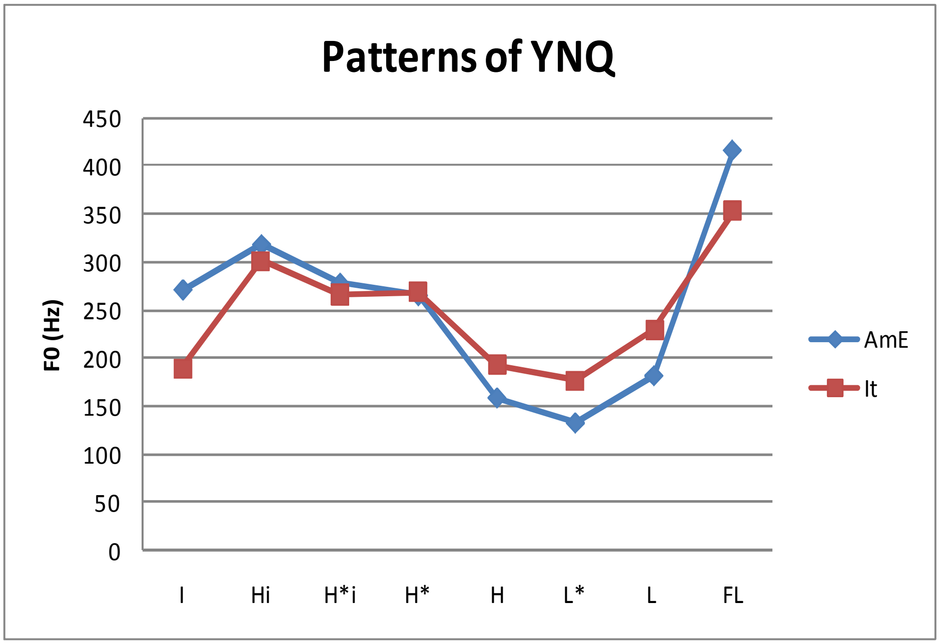


Figure 31. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along YNQ patterns individuated by pitch points corresponding to linguistic landmarks.

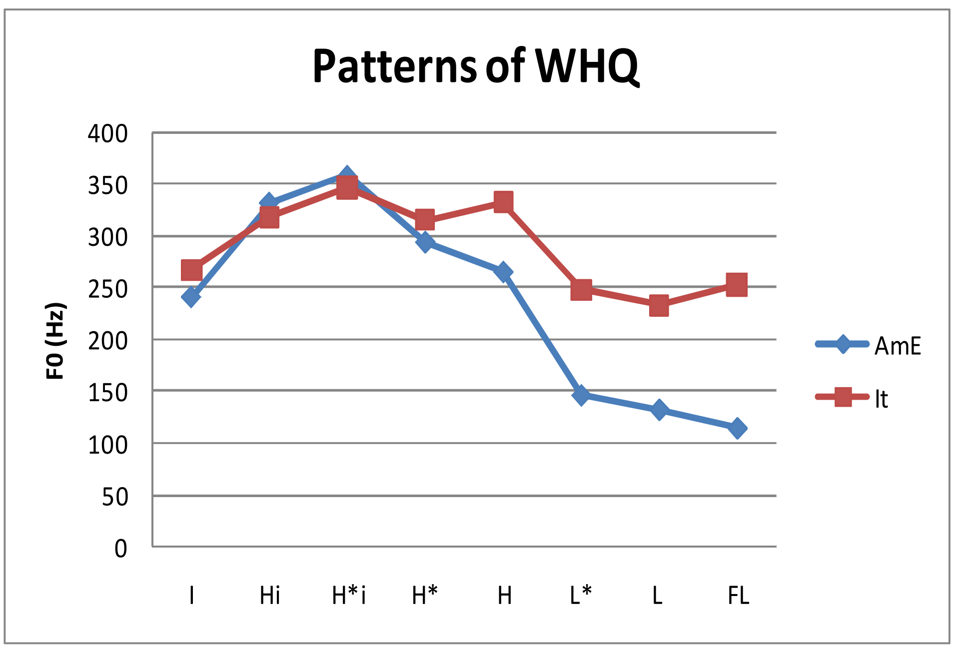


Figure 32. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along WHQ patterns individuated by pitch points corresponding to linguistic landmarks.

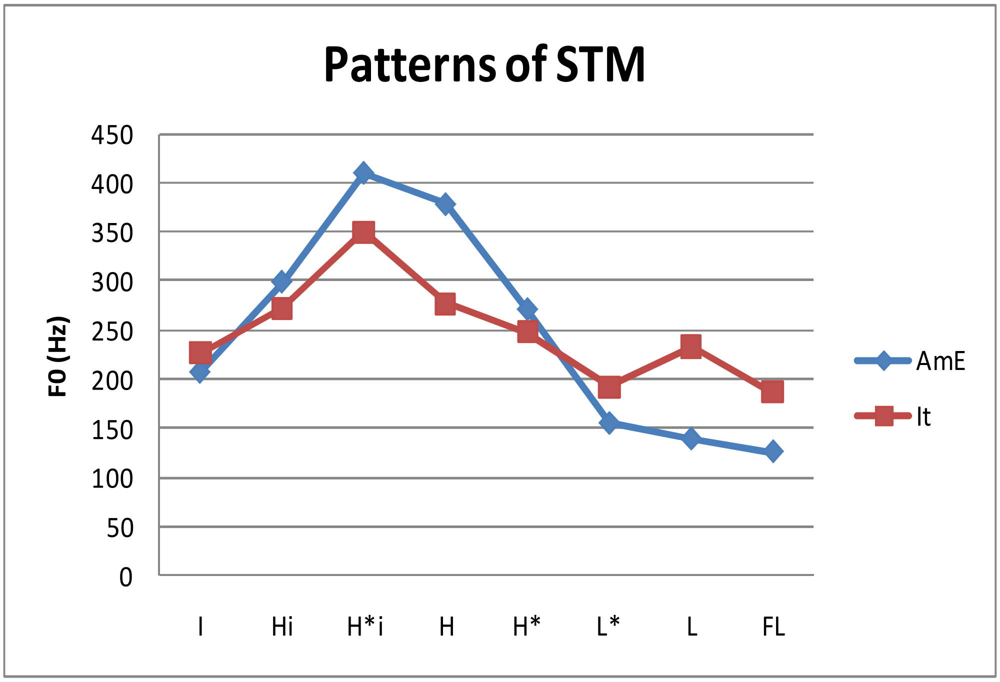


Figure 33. Linguistic measures calculated for the American (red squares) and Italian (blue diamonds) female speakers. Measures are described along STM patterns individuated by pitch points corresponding to linguistic landmarks.

The values obtained from the linguistic measures were averaged and plotted on a graph containing two patterns: the Americans speaking English (the blue line of diamonds), the Italians speaking English (the red line of squares).

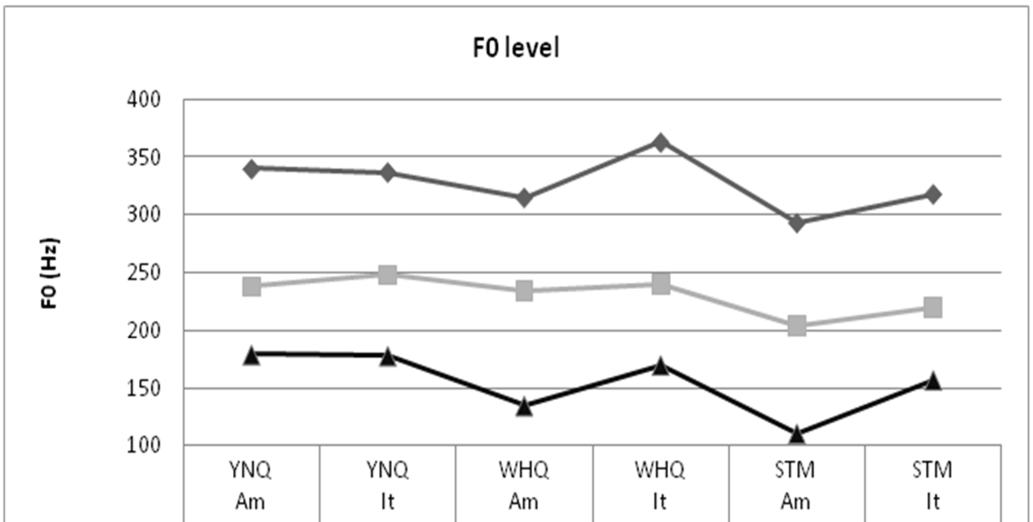
The **patterns for YNQ** (fig. 31) evidenced a clear similarity of the pitch contours of the Americans and the Italians, despite the fact that the Americans had higher values than the Italians at the beginning and at the end of the patterns: compare the values obtained for the initial pitch point I (i.e. 271 Hz for the Americans and 189 Hz for the Italians) and for the final rise FH (i.e. 416 Hz for the Americans and 353 Hz for the Italians). In addition, the mean difference across the H*i, Hi, H*, H, L*, L measures calculated for the Americans and the Italians was 26,33 Hz. By contrast, the mean difference between the Americans' and the Italians' values calculated for the I and the FH measures was 82 Hz for I and 63 Hz for FH. This means that, unlike the I and the FH measures, most of the pitch points measured in the YNQ obtained similar values for the Americans and the Italians.

The **patterns for WHQ** (fig. 32) showed fairly similar values obtained for initial and high peaks (I, H*i, Hi, H*) and different values obtained for final falls and valleys (L*, L, FL). The non-initial and non-prominent peak H had a considerably lower value in the pitch pattern used by the Americans than the Italians (i.e. the value for H was 265 Hz for the Americans and 332 Hz for the Italians). In addition, the mean difference between the Americans' and that Italians' values calculated for initial and high peaks (I, H*i, Hi, H*) was 18,25 Hz. By contrast, the mean difference in final falls and valleys (L*, L, FL) measures calculated for the Americans and the Italians was 114 Hz. This means that, the WHQ patterns were characterized by similar values for initial and high peaks and different values for final falls and valleys. Unlike the Italians, the Americans seem to realize a sharp slope at the end of the WHQ by reaching the lowest values of their pitch patterns.

The **patterns for STM** (fig. 33) showed fairly similar values obtained for some initial and non-prominent high peaks (I, Hi, H) and different values obtained for accented peaks (H*I and H*) and final lows (L, FL). The Americans had higher values than the Italians at the peaks and at the valleys of the utterances. Extremely different values for the Americans and the Italians were obtained for the initial peak H* (i.e. 378 Hz for the Americans and 277 Hz for the Italians) and for the final fall FL (i.e. 126 Hz for the Americans and 187 Hz for the Italians). In addition, the mean difference between the Americans' and the Italians' values calculated in the I, Hi, H L* measures was 26,75 Hz. The mean difference in the H*i and the FL measures calculated for the Americans and the Italians was 60 Hz for H*i and 61 Hz for FL. The highest differences between the pitch values obtained for the Americans and the Italians were reached for the H* measure (i.e. 101 Hz) and the L measure (94 Hz). This means that, the STM pitch line produced by the Italians was realized with a narrower pitch span than that produced by the Americans. Compared to the Americans' STM, the Italians' STM resulted as more flat and compressed. Unlike the Italians, the Americans realized much sharper rises and falls, by reaching the highest and lowest values in the STM patterns.

4.4.9.2 LTD measures

The processing of LTD measures required the analysis of a total of 600 utterances (5 sentences x 3 sentence types x 10 subjects x 2 native languages x 2 repetitions). Namely, every participants in the experiment, 10 Americans and 10 Italians, produced 30 sentences: 10 YNQ, 10 WHQ, and 10 STM. The graph in table 32 shows the mean values obtained for different measures of F0 level (i.e. F0max, F0mean, and F0min) for the two language groups (i.e. American native speakers and Italian learners of English), evidencing clear differences in pitch level patterns across sentence types (i.e. YNQ, WHQ, and STM).



	YNQ Am	YNQ It	WHQ Am	WHQ It	STM Am	STM It
F0max	340	337	315	363	293	318
F0mean	238	248	234	240	204	220
F0min	179	178	135	170	111	157

Table 32. F0 max, mean and min values in Hz by American and Italian subjects obtained in three sentence types: YNQ, WHQ, and STM.

F0 level considerably varies across sentence types and language groups, with WHQ and STM showing more significantly differences between the productions by the Americans and the Italians. It is clear, from the graph in table 32, that F0 level for WHQ and STM is shifted downwards in the utterances by the Americans, as compared to that in the Italian utterances. Surprisingly, YNQ showed very similar results for the Americans and the Italians. In the sentences produced by the Americans, every measure, F0max, F0mean, and F0 min, reached the highest values in YNQ and the lowest in STM, with values for WHQ in between. Also the sentences produced by the Italians were in line with this trend, with the exception of WHQ. In fact, F0max is higher for WHQ (363 Hz) than YNQ (337 Hz). Since, WHQ are commonly uttered with falling contours (see. 4.4.2), one would expect that they have rather low F0max values. Nevertheless, this does not happen.

F0 span was calculated in ST. The bars in table 33 show the distribution of F0 values in ST, for the three sentence groups (YNQ vs. WHQ vs. STM), evidencing clear differences in pitch span patterns across the language groups (Americans vs. Italians).

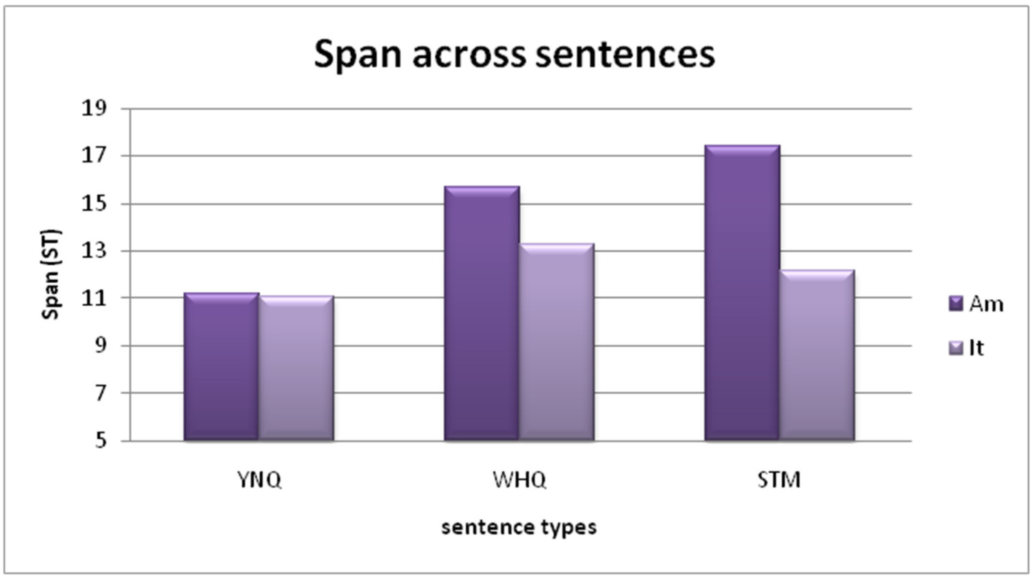


Table 33. Span values across sentences (YNQ, WHQ, and STM) by the American and the Italian subjects.

Span was measured in ST because logarithmic scales (e.g. ST) better than linear scales (e.g. Hz) manage to capture the excursions between F0 values (Daly and Warren, 2001; Nolan, 2003; Mennen et al. 2012). YNQ reached fairly similar pitch span values, with span in American YNQ (11.19 ST) slightly larger than in Italian YNQ (11.03 ST). Both WHQ and STM showed great differences in the span values of the American vs. Italian utterances. Span values for WHQ were 15,65 ST for the Americans and 13,29 ST for the Italians; span values for STM were 17,42 ST for the Americans and 12,19 ST for the Italians. In sum, the Americans' span was larger than that of the Italians in every sentence type.

The data from LTD measures were tested with t-tests, in order to see whether or not the differences across language groups were statistically significant. A t-test compares exactly two datasets (of some dependent variable such as F0max, F0min, F0mean, and F0 span) that differ with respect to one manipulation (of an independent variable such as 'sentence type' and 'native language'). With little or no overlap between the samples, one can be confident that the two datasets are statistically different. Whether or not there is overlap depends on the difference between the means of the samples (bigger difference means less overlap) and the variability in the samples (less variability means less overlap).

A paired t-test assuming equal variances between groups was selected as the preferred method to analyze data in this experiment, mostly for two reasons: 1) it is one of the most commonly used tests in phonetics studies (Rasinger, 2008; Lane, 2012); 2) it suits the kind of variables investigated in this experiment. The F0 values were loaded into excel to obtain inferential statistics.

For the **factor *'sentence type'*** (YNQ vs. WHQ vs. STM), two-sample paired t-tests were run for the dependent variables F0 level and span, calculated for the sentences produced by the Americans and the Italians. Statistical significance for the 'sentence type' factor was measured in a total of 12 t-tests. The dependent variables were two measures for pitch range (F0 level and span), the independent variables were the native languages of the speakers (American vs. Italian). For a more detailed description of data obtained from the 24 t-tests, the tables of the results of t-tests are shown in Appendix A.

Within the American subjects, the data for F0 level of YNQ were not significantly different from the data for WHQ, ($t(99)=1.15$ at the .05 level, $p=0.250$). The F0mean of YNQ was 238 Hz and that of WHQ was 234 Hz, with only 4 Hz difference. By contrast, the difference between YNQ and STM was very significant ($t(99)=10.85$ at the .05 level, $p<0.001$). The F0mean of YNQ was consistently higher than that of STM, with a difference of 34 Hz. The results of the t-test on F0 level in WHQ and STM indicate that these sentence types are significantly different, ($t(99)=8.76$ at the .05 level, $p<0.001$). The F0mean of WHQ was consistently higher than that of STM, with a difference of 30 Hz. As far as F0 span is concerned, data for the Americans subjects evidenced neat differences across sentence types. The data for F0span in YNQ were significanlty different from the data for F0 span in WHQ, ($t(99)= -6.13$ at the .05 level, $p<0.001$). Statistically significance was exhibited also by differences of F0 span in YNQ vs. STM ($t(99)= -9.65$ at the .05 level, $p<0.001$) and WHQ vs. STM ($t(99)= -2.15$ at the .05 level, $p=0.033$).

Within the Italian subjects, the data for F0 level of YNQ were significantly different from the data for WHQ ($t(99)=2.56$ at the .05 level, $p=0.011$). The F0mean of YNQ was 248 Hz and that of WHQ was 240 Hz. Also the difference between YNQ and

STM was significant (t(99)=9.68 at the .05 level, p<0.001). The F0mean of YNQ was consistently higher than that of STM, with a difference of 28 Hz. The results of the t-test on F0 level in WHQ and STM indicate that these sentence types are significantly different (t(99)=6.53 at the .05 level, p<0.001). The F0mean of WHQ was higher than that of STM, with a difference of 20 Hz. As far as F0 span is concerned, data for Italian subjects evidenced neat differences across sentence types. The data for F0span in YNQ were significanlty different from the data for F0 span in WHQ (t(99)= -4.71 at the .05 level, p<0.001). Statistically significance was displayed also by differences of F0 span in YNQ vs. STM (t(99)= -2.72 at the .05 level, p=0.007) and WHQ vs. STM (t(99)= 2.20 at the .05 level, p=0.029).

In sums, the statistically significant differences across sentence types (YNQ vs. WHQ vs. STM) were largely proved by several t-tests showing that both Italians and Americans modify F0 level and span, depending on sentence type.

For the **factor *'native language'***, two-tail paired t-tests were run for the dependent variables (F0max, F0mean, F0min, and STrange), separately calculated for every sentence type. Statistical significance for the 'native language' factor was measured in a total of 12 t-tests with F0 measures (F0max, F0mean, F0min, and STrange) as dependent variables and sentence type (YNQ vs. WHQ vs. STM) as independent variables (4 F0 measures x 3 sentence types). As shown in fig. 34 - 37, the two-tail paired t-tests show that there is a significant effect for the 'native language' factor across sentences produced by American native speakers and Italians native speakers. Separate analyzes were carried by creating a series of box plots for each dependent variable: F0max, F0mean, F0min, and STrange.

As far as **F0mean** is concerned, the difference between American and Italian YNQ was statistically different (t(99)= -2.16 at the .05 level, p=0.032). The F0mean of American YNQ was 238 Hz and that Italian YNQ was 248 Hz, with 10 Hz difference. By contrast, the difference between American and Italian WHQ was not statistically significant (t(99)= - 1.68 at the .05 level, p<0.094). The F0mean of American WHQ was only slightly lower than that of Italian WHQ, with a difference of 6 Hz. This difference was not significant. The results of the t-test on F0mean in American and Italian STM indicate that F0 mean is significantly different across the two language speakers (t(99)= -5 at the .05 level, p<0.001). The F0mean of American STM was consistently lower than that of Italians, with a difference of 16 Hz.

As far as **F0max** is concerned, data for WHQ and STM evidenced neat differences across language groups (native speakers of American vs. native speakers of Italian). By contrast, no significant difference was found in YNQ produced by the Americans and the Italians ($t(99)= 0.47$ at the .05 level, $p=0.635$). The F0max of American YNQ was 340 Hz and that Italian YNQ was 337 Hz, with only 3 Hz difference. Statistical significance was shown by F0max in WHQ uttered by the Americans and Italians ($t(99)= -6.21$ at the .05 level, $p<0.001$). The F0max of American WHQ (315 Hz) was dramatically lower than and that Italian WHQ (363 Hz), with a 48 Hz difference. Also the differences between STM uttered by the Americans vs. the Italians were statistically significant ($t(99)= -3.55$ at the .05 level, $p<0.001$). The F0max of American STM was 293 Hz and that Italian STM was 318 Hz, with a 25 Hz difference.

As far as **F0min** is concerned, the difference between American and Italian YNQ was not statistically significant ($t(99)= 0.08$ at the .05 level, $p=0.933$). The F0min of American YNQ was 179 Hz and that Italian YNQ was 178 Hz, with only 1 Hz difference. By contrast, the F0min difference between American and Italian WHQ was statistically significant ($t(99)= -6.26$ at the .05 level, $p<0.001$). The F0min of American WHQ (135 Hz) was considerably lower than that Italian WHQ (170 Hz), with a difference of 35 Hz. The results of the t-test on F0min in American and Italian STM indicate that F0min is significantly different across the two language speakers ($t(99)= -10.26$ at the .05 level, $p<0.001$). The F0min of American STM (111 Hz) was dramatically lower than that of Italians (157 Hz), with a difference of 46 Hz.

As far as **STrange** is concerned, data for WHQ and STM evidenced neat differences across language groups (native speakers of American vs. native speakers of Italian). By contrast, no significant difference was found for STrange in YNQ produced by the Americans and the Italians ($t(99)= 0.36$ at the .05 level, $p=0.714$). The STrange of American YNQ was 11.19 ST and that Italian YNQ was 11.03 ST. Statistically significance was exhibited by STrange in WHQ uttered by Americans and Italians ($t(99)= 3.26$ at the .05 level, $p=0.001$). The STrange of American WHQ (15.65 ST) was lower than that of Italian WHQ (13.26 ST), with a 2.35 ST difference. Also the differences between STM uttered by the Americans vs. the Italians were statistically significant ($t(99)= 8.02$ at the .05 level, $p<0.001$). The STrange of American STM was 17.42 ST and that Italian STM was 12.19 ST, with a 5.23 ST difference.

The box plots in fig. 34 - 37 describe F0mean, F0max, F0min and STrange in sentences uttered by the Americans and the Italians. In descriptive statistics, the height and length of boxes or whiskers depend on several measures. The bottom and top of the box are the first and second quartiles (also called Q1 and Q3, respectively), and the band near the middle of the box is the second quartile (Q2) that is also called 50th percentile or median. The bottom and top of the whiskers represent the smallest observation (sample minimum) and the largest observation (sample maximum) within data. Any data not included between the top and bottom whiskers is plotted into the graph as an outlier with a dot, small circle, or star (for further descriptions of box plots, see McGill et al., 1978; Benjamini, 1988; Rasinger, 2008; Lane, 2012). The box plots in fig. 34 describe F0mean in sentences uttered by the Americans (see blue box plots) and the Italians (green box plots).

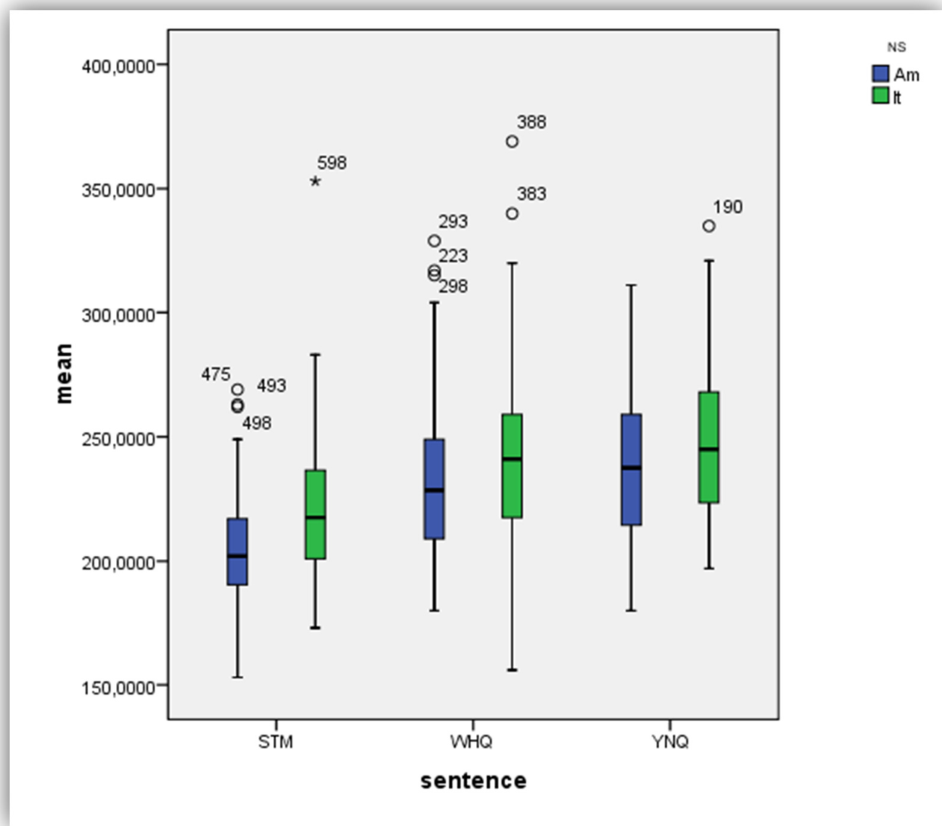


Figure 34. F0mean values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots).

Across sentences, F0mean was significantly higher for the Italians than for the Americans (compare the blue box plots to the green ones). Median values for F0mean were grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). As reported in tab. 34, the F0mean values were higher in Italian than American sentences with a difference of 8.25 Hz in YNQ, 12.85 Hz in WHQ, and 15.80 Hz in STM.

	Median of F0mean	
	AM	**IT**
YHQ	237.15 Hz	245.40 Hz
WHQ	228.75 Hz	241.60 Hz
STM	202.45 Hz	218.25 Hz

Table. 34. Median values for F0mean, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM).

The box plots in fig. 35 describe F0max in sentences uttered by the Americans (see blue box plots) and the Italians (green box plots).

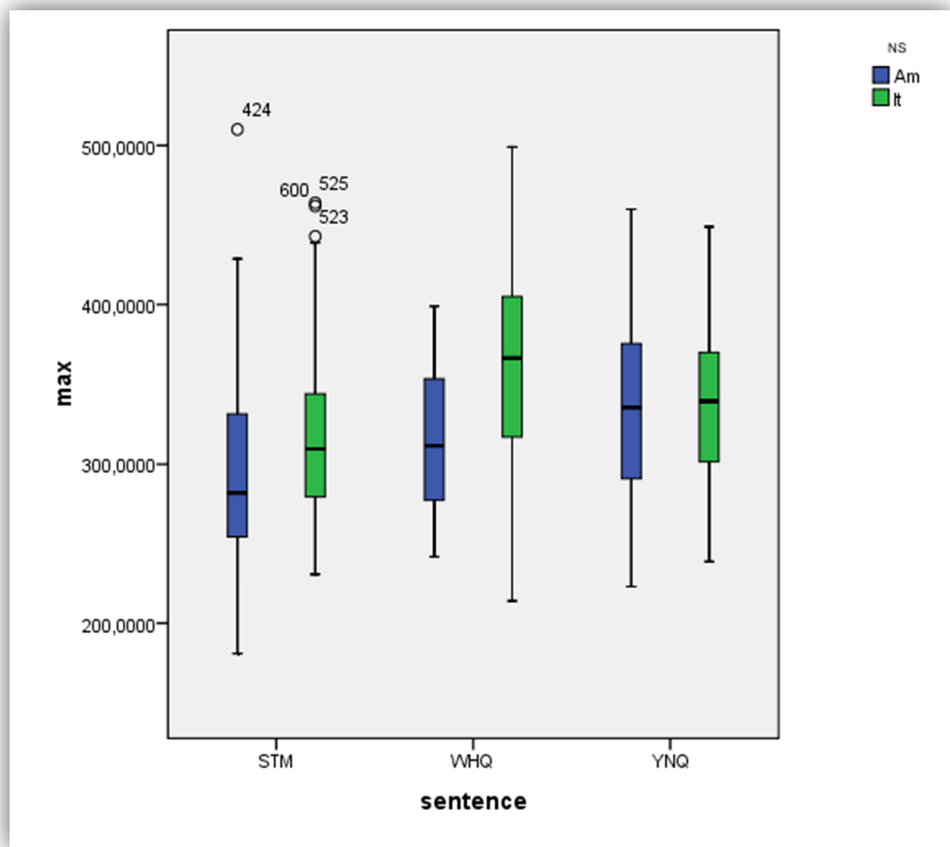


Figure 35. F0max values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots).

Across sentences, F0max was significantly higher for the Italians than for the Americans (compare the blue box plots to the green ones). Median values for F0max were grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). As reported in tab. 34, the F0 max values were higher in Italian than American sentences with a difference of 10.95 Hz in YNQ, 55.05 Hz in WHQ, and 27.60 Hz in STM.

	Median of F0max	
	AM	**IT**
YHQ	328.55	339.50
WHQ	311.90	366.95
STM	282.30	309.90

Table 35. Median values for F0max, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM).

The box plots in fig. 36 describe F0min in sentences uttered by the Americans (blue box plots) and the Italians (green box plots).

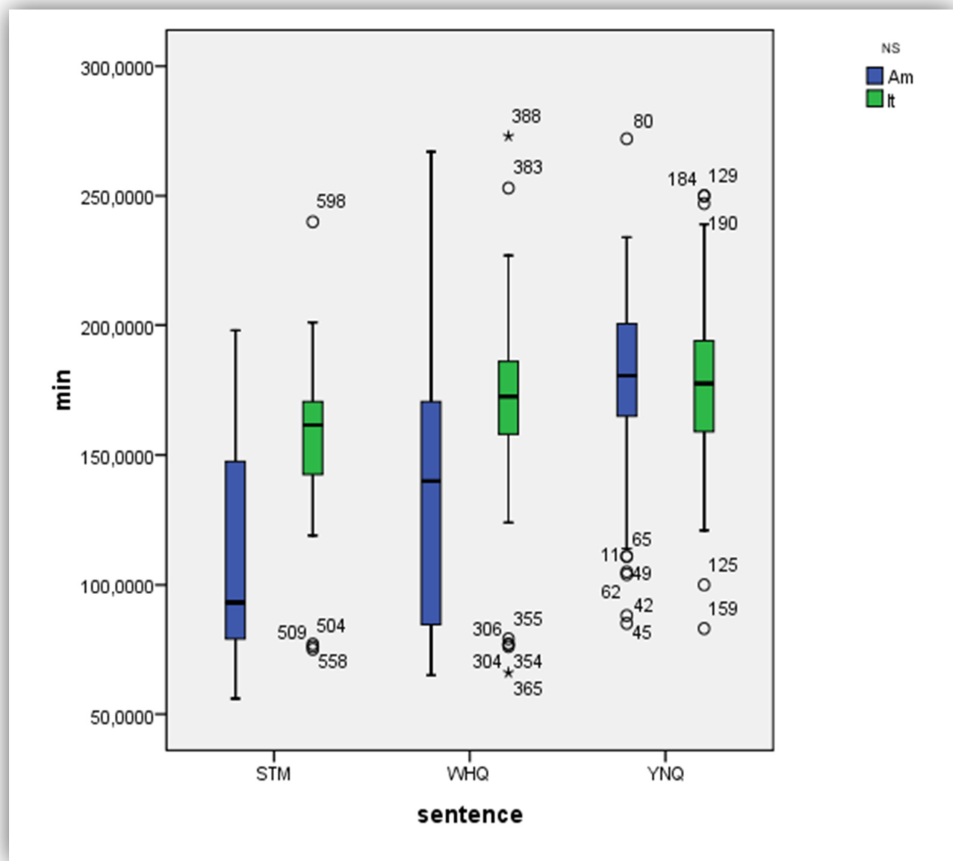


Figure 36. F0min values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots).

Across sentences, F0min was significantly higher for the Italians than for the Americans, with the exception of YNQ which display similar median and quartiles for the Americans and the Italians (compare the blue box plots to the green ones). Median values for F0min were grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). As reported in tab. 36, the F0min values in American YNQ were slighlty higher than in Italian YNQ with a difference of 4.5 Hz. By contrast, American WHQ and STM had lower F0min values than Italian WHQ and STM, with a difference of 32.80 Hz in WHQ and 67.40 Hz in STM.

	Median of F0min	
	AM	**IT**
YHQ	182.25	177.75
WHQ	140.30	173.10
STM	93.70	161.10

Table 36. Median values for F0min, grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM).

The box plots in fig. 37 describe STrange in sentences uttered by the Americans (blue box plots) and the Italians (green box plots).

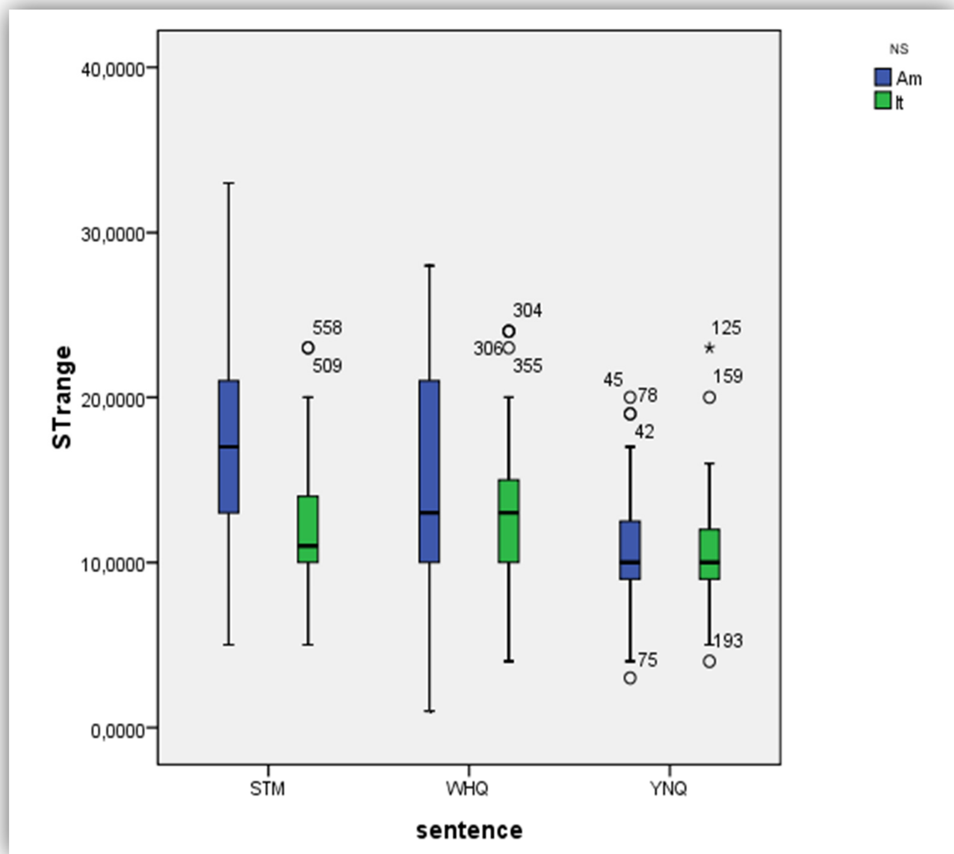


Figure 37. STrange values across sentences (STM vs. WHQ vs. YNQ) produced by Americans (blue box plots) and Italians (green box plots).

Across sentences, STrange was significantly higher for the Italians than for the Americans, with the exception of YNQ which display almost identical median and quartiles for the Americans and the Italians (compare the blue box plots to the green ones). Median values for STrange were grouped depending on native language (American vs. Italian) and sentence type (YNQ vs. WHQ vs. STM). As reported in tab. 37, the STrange values in American YNQ were slighlty higher than in Italian YNQ with a difference of 0.3 ST. By contrast, American WHQ and STM had lower STrange values than Italian WHQ and STM, with a difference of 0.75 ST in WHQ and 6.55 ST in STM.

	Median of STrange	
	AM	**IT**
YHQ	10.55	10.85
WHQ	13.85	13.10
STM	17.90	11.35

Table 37. Median values for STrange, grouped depending on native language (Americans vs. Italian) and sentence type (YNQ vs. WHQ vs. STM).

Repeated measures showed that the between-subject factors 'native language' and 'sentence type' reached significance for most dependent variables, characterizing F0 level and span. This suggests that the native language of the subjects (L1) plays a relevant role in pitch variation.

Within **YNQ**, F0mean was significantly different across language groups (English L1 vs. English L2), as indicated by the one-way ANOVA ($F(199)=7.769$, $p=0.030$). Also F0median was different in YNQ produced by the Americans and the Italians, as evidenced by the one-way ANOVA ($F(199)=8.128$, $p<0.005$). By contrast, no statistical difference was shown in results for F0min ($F(199)=6.845$, $p=0.931$) and F0 max ($F(199)=0.212$, $p=0.646$). Also F0 span did not reached significance, as indicated by the one-way ANOVA ($F(199)=1.125$, $p=0.732$).

Within **WHQ**, F0mean and F0median were not significantly different across language groups (English L1 vs. English L2), as indicated by the one-way ANOVA for F0mean ($F(199)=1.840$, $p=0.177$) and the one-way ANOVA for F0median ($F(199)=1.796$, $p=0.182$). By contrast, F0max and F0min were statistically different in WHQ produced by the Americans and the Italians, as evidenced by the one-way ANOVA for F0max (F(199)=44.302, p<0.001) and the one-way ANOVA for F0min (F(199)=35.156, p<0.001). Also F0 span reached significance in WHQ, as indicated by the one-way ANOVA (F(199)=8.843, p=0.003).

Within **STM**, F0mean was significantly different across language groups (English L1 vs. English L2), as indicated by the one-way ANOVA (F(199)=19.729, p<0.001). Also F0median was different in STM produced by the Americans and the Italians, as evidenced by the one-way ANOVA (F(199)=7.380, p<0.007). The differences in F0max (F(199)=10.999, p<0.001) and F0 min (F(199)=109.894, p<0.001) were also significant. Also F0 span reached significance, as indicated by the one-way ANOVA (F(199)=56.882, p<0.001). In sum, every dependent variable (F0mean, F0median, F0max, F0min, and span) was significantly different in English STM produced by the Americans, as compared to those produced by the Italians. This shows that pitch range is sensibly different in English sentences uttered by the English native speakers and non-native speakers .

4.4.9.3 Speaking rate and duration

Temporal features, such as speaking rate and phoneme-, word-, sentence-, duration, vary significantly across speakers. Duration and speaking rate have been reported to have an influence on foreign accent perception; in particular, number of syllables, duration and speaking rate are some of the most common measures calculated by voice specialists to measure and assess speech variation (see § 2.2.4).

Even though different models have been proposed to pinpoint the parameters that best represent temporal variation, to combine the analysis of time (duration) and frequency (pitch) is not an easy task. Arslan and Hansen (1996, 1997) tried to match this time-frequency variation by analyzing how native speakers of different languages produce slopes (frequency/time) in their intonation patterns. The results produced in their studies show that pitch variation across slopes is so large that this parameter 'is not expected to provide significant accent discrimination ability' (1996: 13). By contrast, De Jong and Wempe (2007, 2009) proposed a method to measure speech rate automatically. Their analysis is based on the assumption that the number of syllables per time unit is one of the best measures to capture speech rate differences across speakers and languages. Based on the use of a script for syllable nuclei detection, the following measured can be calculated: number of syllables, total duration, phonation time, speech rate, articulation rate, and speaking time over number of syllables (for a definition of these measure, see § 2.1.4).

Together with the analysis of F0 movements (described by linguistic and LTD measures), the differences of duration and speaking rate were analyzed between two language groups: the Americans speaking English and the Italians speaking English. The Italians speaking Italian are not included in the analysis of data because of the unavoidable differences in spoken materials. In fact, the duration, the rhythm and the number of syllables of the English sentences are different from those of the Italian sentences. For this reason, to compare English and Italian sentences could be misleading and entice false interpretations.

The utterances produced by the subjects of the second experiment were grouped depending on sentence types. In line with the protocol for the analysis of data followed so far, the results were grouped depending on the L1 of the speakers (American vs. Italian) and the sentences types (YNQ vs. WHQ vs. STM). The measures calculated were: number of syllables (nsyll); average syllable duration (ASD); duration (Dur); phonation time (PhT); speech rate (SR); articulation rate (AR). These measures were calculated in 15 sentences (5 YNQ vs. 5 WHQ vs. 5 STM) produced by the American and Italian speakers who repeated these sentences twice. (See figures 55-69 in Appendix C for the graphs showing mean values for ASD, D, PhT, SR, AR).

The **number of syllables** (nsyll) was tested in two one-way ANOVAs with ‘L1 of the speakers’ and ‘sentence type’ as between-subjects factors. The data showed that the Italians produced significantly more syllables than the Americans ($F(718)=19{,}368$ $p<0.001$). In addition, the number of syllables varied across sentences. In the YNQ, the mean nsyll was 3.95 for the Americans and 4.07 for the Italians. In the WHQ, the mean nsyll was 4.56 for the Americans and 5.61 for the Italians. In the STM, the mean nsyll was 5.61 for the Americans and 6.22 for the Italians. While YNQ had similar nsyll values for the Americans and the Italians, WHQ and STM were found to have different nsyll values across speakers. This shows that mean nsyll significantly varied across sentence types ($F(717)=79{,}968$ $p<0.001$).

The **average syllable duration** (ASD) was considerably higher for the Italians than the Americans. In the YNQ, the mean ASD was 0.27 s for the Americans and 0.30 s for the Italians. In the WHQ, the mean ASD was almost identical in the Americans and the Italians (i.e. for both the American and the Italian speakers, ASD was 0.26 s). However, as shown in Appendix C (fig. 68), ASD was higher for the Italians than the Americans in all the WHQ, but one. In the STM, the mean ADS was 0.23 s for the

Americans and 0.26 s for the Italians. As evidenced by two one-way ANOVAs with 'L1 of the speakers' and 'sentence type' as between-subjects factors, the mean ASD of sentences was found to be significantly higher in the Italians than the Americans ($F(718)=9,462$ $p<0.002$) and to significantly differ across sentence types ($F(717)=13,249$ $p<0.001$).

The mean **duration** (Dur) of sentences was tested in one-way ANOVAs with 'L1 of the speakers' and 'sentence type' as between-subjects factors. Dur was calculated in seconds (s) and its value varied across sentence types. In the YNQ, the mean Dur was 2.13 s for the Americans and 2.67 s for the Italians. In the WHQ, the mean Dur was 2.42 s for the Americans and 2.86 s for the Italians. In the STM, the mean Dur was 2.58 s for the Americans and 2.89 s for the Italians. In sum, the mean Dur of sentences was found to be significantly higher in the Italians than the Americans ($F(718)=81,716$ $p<0.001$) and to significantly differ across sentence types ($F(717)=16,669$ $p<0.001$).

The **phonation time** (PhT), that is defined as total time spent speaking divided by total time to produce speech sample (Wang, 2008: 21), was tested in one-way ANOVAs with 'L1 of the speakers' and 'sentence type' as between-subjects factors. Since the nsyll, the ASD, and the Dur of the sentences were higher in the Italians than the Americans, consequently, also the PhT of the Italians' sentences was significantly higher than that of the Americans' sentences ($F(718)=96,808$ $p<0.001$). In the YNQ, the mean PhT was 1.025 s for the Americans and 1.114 s for the Italians. In the WHQ, the mean PhT was 1.116 s for the Americans and 1.424 s for the Italians. In the STM, the mean PhT was 1.257 s for the Americans and 1.554 s for the Italians. In sum, the mean PhT of sentences was found to be significantly different across sentence types ($F(717)=58,284$ $p<0.001$).

The **speech rate** (SR), defined as syllables per minute, was tested in two one-way ANOVAs with 'L1 of the speakers' and 'sentence type' as between-subjects factors. In the YNQ, the mean SR was 1.97 for the Americans and 1.60 for the Italians, thus SR was higher in the Americans than the Italians. By contrast, in the WHQ, SR was higher in the Italians than the Americans, with a mean SR of 1.94 for the Americans and 2.00 for the Italians. In the STM, the mean SR was fairly similar across the subjects: 2.23 for the Americans and 2.21 for the Italians. In sum, the mean SR of sentences was found to differ significantly across the Italian and the American subjects ($F(718)=4,368$ $p<0.037$) and across sentence types ($F(717)=24,469$ $p<0.001$).

The **articulation rate** (AR), defined as syllables per minute without pauses, was tested in two one-way ANOVAs with 'L1 of the speakers' and 'sentence type' as between-subjects factors. Unlike SR, AR was found to be significantly higher in the utterances produced by the Americans than the Italians, for every sentence. In the YNQ, the mean AR was 3.93 for the Americans and 3.56 for the Italians. In the WHQ, the mean AR was 4.18 for the Americans and 3.90 for the Italians. In the STM, the mean AR was 4.50 for the Americans and 4.00 for the Italians. In sum, the mean AR of sentences was found to be significantly higher in the Americans than the Italians ($F(718)=25{,}496$ $p<0.001$) and to significantly differ across sentence types ($F(717)=14{,}775$ $p<0.001$).

In conclusion, the data show cross-linguistic differences across sentence types in the number of syllables (nsyll); average syllable duration (ASD); duration (Dur); phonation time (PhT); speech rate (SR); articulation rate (AR). In fact, the American speakers produced YNQ vs. WHQ vs. STM with lower nsyll, ASD, Dur, and PhT than the Italian speakers. By contrast, the American speakers produced YNQ vs. WHQ vs. STM with higher SR and AR than the Italian speakers. Differences between the two groups of speakers were found to be significant.

4.4.10 Discussion

The second experiment examined pitch range variation in different sentence types: yes-no questions (YNQ), wh-questions (WHQ), and statements (STM). This study was designed to determine how pitch level and span vary in English sentences produced by American and Italian females. In particular, pitch variation was tested across YNQ, WHQ and STM in order to identify a specific model of pitch patterns for each sentence type produced in English as L1 and L2. The aim of the present experiment was to find out whether or not pitch range considerably varies depending on the sentence types (e.g. YNQ vs. WHQ vs. STM) and, most importantly, whether or not English sentences produced by the Americans have a pitch range similar to that of sentences produced by the Italians.

Five dependent variables (F0mean, F0median, F0max, F0min, and span) were tested. The results showed that Americans' F0 span was larger than that of the Italians while the Americans' F0 level was lower than that of the Italians. This gives indication about the fact that pitch range is sensibly different in English sentences uttered by native speakers (in this case, Americans) and non-native speakers (in this case, Italians).

As far as F0 level is concerned, the results showed that F0 level considerably varied across sentence types and different L1 speakers (the Americans and the Italians), with

WHQ and STM showing more significant differences between the productions of the American and the Italians. The F0 level for WHQ and STM was shifted downwards in the utterances by the Americans, as compared to the Italian utterances. By contrast, YNQ obtained very similar results for the Americans and the Italians. F0mean, F0max, and F0 min were significantly higher for the Italians than for the American.

As far as F0 span is concerned, data for WHQ and STM evidenced neat differences across language groups (native speakers of American vs. native speakers of Italian). By contrast, no significant difference was found for F0 span in YNQ produced by the Americans and the Italians. Across sentences, F0 span was significantly higher for the Italians than for the Americans, with the exception of YNQ which display almost identical F0 span values for the Americans and the Italians. Generally, the Americans speakers produced sentences with lower numbers of syllable, average syllable duration, and phonation time than the Italians. In particular, the durations of the Americans' sentences were considerably lower than those of the Italians' sentences. Significant differences in speech rate and articulation rate were found between the two groups of speakers. The Americans produced sentences with higher SR and AR than the Italians.

4.5 Summary

Chapter 4 describes the goals and the achievements of the two production tests devised for this study. The experimental design underlying the production studies is aimed at identifying how pitch range is realized in English and Italian. A challenging issue in spoken prosody is the difficulty in the identification of the acoustic measures of pitch range. The main problem deals with the importance of quantifying a pitch movement within a determined F0 space. In fact, pitch range cannot be described in absolute terms but only in relation to the vocal characteristics of the speakers (the maximum and minimum F0 they can reach) and to the preceding or following targets in the continuum. For instance, 'the F0 level associated with a high pitch accent for one speaker, while higher than a low pitch accent for the same speaker in the same context, might correspond to the F0 level of a low accent for another speaker' (Shue, 2010: 130). Therefore, F0 values are related to target high or low pitch values, modeled on the speakers' characteristics. The experiments discussed in chapter 4 are based on an integrated model for the analysis of F0, based on the combination of LTD (long-term distributional) and linguistic measures.

The first comparative study explores the pitch range of men and women in American English (L1), English (L2), and Italian (L1). The pitch range values produced by the Italians speaking English are shown to be significantly different from those produced by the American native speakers and the Italian native speakers. Generally, the English sentences produced by the Italians are found to have a narrower pitch span than those produced by the Americans. This is due to two possible reasons. First, regardless of the nature of the languages investigated, L2 speech is always characterised by a narrower pitch range than L1 speech. Second, Italian learners of English are influenced by their L1 and transfer their L1 pitch range variation into their L2. In addition to L1, gender is also found to play an important role in distinguishing between pitch variation in L1 and L2. F0 values differ in males and females because of physiological and socio-cultural factors. Moreover, males and females have been found to modify their pitch values to different extents, depending on the language they speak. Unlike males, female speakers try to adapt their pitch patterns in L2 to the native speakers' model.

The second comparative study presents and discusses different realizations of pitch patterns across three sentence types: yes-no questions, wh-questions and statements. The main goal of this investigation is to identify whether or not American and Italian females adopt a similar pitch range (that represents phonetic details) across different phonological patterns. Hence, the development of pitch analysis is considered across a phonetic and a phonological interface. In order to grasp the whole aggregate of aspects influencing pitch variation, data on the duration, phonation time, speech and articulation rate, and number of syllables in the utterances are provided, too. Despite the similarities in phonological patterns produced by the American and the Italian subjects, meaningful phonetic differences are found across sentence types. All the sentences types produced by the Italians are characterized by higher F0 level and lower F0 span than the sentences produced by the Americans. In particular, the F0 level for wh-questions and statements is shifted downwards in the utterances produced by the Americans, as compared to the utterances produced by the Italians. Unlike the other sentence types, yes-no questions present similar F0 level and span values across speakers.

Chapter 5:
Perception Study

5.1 Introduction

The results obtained from the production studies, conducted on interrogative and affirmative sentences uttered by American native natives speakers and Italian learners of English, showed that pitch contours differ across sentence types and language groups. It was found that 'English declarative sentences typically end in a low, falling pitch and questions typically end in a high, rising pitch' (Vicenik, 2011: 40). The data presented in my analyses confirmed that the STM produced by the Italians and the Americans were generally uttered with falling F0 contours while the YNQ were uttered with rising contours. Similarly to the STM, the WHQ ended with final low contours. However, the tonal patterns of the WHQ were significantly different from the patterns of the STM, especially because pitch accents reached much higher F0 levels in the WHQ than the STM.

In addition, the Italian speakers' intonation in English was characterized by rather level and unvaried contours in different sentence types (e.g. YNQ, WHQ, STM). By contrast, the American native speakers used a variety of intonation contours and modulated pitch range depending on the sentence types. Since in English a level contour is typically associated with boredom or lack of interest, the production of unmodulated intonation by Italian speakers of English is likely to be perceived as a sign of scarce engagement in the conversation or lack of interest. In addition, the extensive use of this 'default' level contour may contribute to creating a distorted image of the Italian speakers' attitude or emotional state in specific communicative contexts.

Monotony (i.e. flat pitch range) is claimed to have a negative impact on listeners comprehension (Holub, 2010). In fact, several experiments have shown that inaccurate but lively speech is more positively judged by listeners than correct but monotonous speech (Collados Aís, 1998). This phenomenon has been observed especially in simultaneous interpreting (Ahrens, 2005; Holub, 2010). Interpreters have been found to 'be perceived as being less professional when speaking in a monotonous voice, regardless of whether they correctly convey the content of the original speech' (Holub, 2010: 125).

In a study examining intelligibility rate in speech by healthy female talkers, F0 was indicated as a factor of drop in intelligibility (Watson and Schlauch, 2008). Listeners were asked to rate the intelligibility of sentences presented under two conditions: 1) unmodified original sentences and 2) re-synthesized sentences with F0 reflecting the average values of F0 min, max, and mean for each subject's productions. Results showed that unmodified original sentences were more intelligible than sentences with flattened F0 contours. Across re-synthesized sentences, it was found that the level of intelligibility decreases as the index of flattened F0 increases (e.g. sentences were rated on a scale from poor to high intelligibility, with F0 flattened at average F0 max, F0 median and F0 min). Thus, the sentences flattened at the average F0 min were more intelligible than the sentences flattened at the average F0 max. After examining their data, Watson and Schlauch came to the conclusions that 'F0 height accounted for only a small amount of the drop in speech understanding in speech with a flattened F0' (2008: 348).

Despite the validity of the results achieved by Watson and Schlauch (2008), I believe that their conclusions are partially biased. I believe it is not advisable to compare natural speech to flattened speech. One of the problems related to the comparison of natural vs. flattened speech is that listeners who have to judge sentences rely more heavily on cues indicating naturalness rather than focusing on F0 variation. In addition to differences in F0 variation, natural and flattened sentences are characterized by differences of rhythmic and timing cues. In fact, the pitch contours of flattened sentences have a distinctive flat and monotone pitch that makes them sound robotic and unlively. When listeners have to rate these kinds of sentences they probably rely more on the monotone pitch than on the F0 level. Thus it is not correct to assume that 'F0 height is a secondary factor in the drop in intelligibility seen in monotone speech for female talkers' (Watson and Schlauch 2008: 348), on the basis of results obtained from a study where natural and flattened sentences are mixed together.

In order to test the influence of intonation, rhythmic, durational and timing information, as separate factors, stimuli must be diversified accordingly. For this purpose, re-synthesized speech permits to consider different cues separately, by eliminating some factors and preserving others. In a study on the ability to identify languages depending on suprasegmental cues (Ramus and Mehler, 1999), sentences were re-synthesized by preserving different aspects: 1) phonotactics, rhythm and intonation; 2) rhythm and intonation; 3) intonation only; 4) rhythm only.

In line with the results achieved with this model, a study on language identification by American infants (Vicenik, 2011) proved that intonation is one of the most robust cues to discriminate among languages. Vicenik's investigation (2011) was carried out with American (L1) and German (L2) stimuli in order to test whether or not infants rely on intonation to distinguish between the two languages. Stimuli were presented under three different conditions: 1) full-cue stimuli, that is natural speech in unmodified sentences; 2) low-pass filtered stimuli, that are sentences in which segmental information is removed while suprasegmental information (intonation, stress, prominence, etc.) is preserved; 3) re-synthesized intonation stimuli, that are sentences in which the original F0 contour is replaced with another contour modulated on the investigator's specific needs. The stimuli were modified and re-synthesized with different pitch contours, by adopting the following procedure:

> 'A simple undulating pitch contour was added to the American English and German stimuli so that all sentences in both languages had the same unfamiliar pitch pattern. This was done instead of simply removing the pitch contour, leaving a monotone pitch, because it was thought monotone intonation would be too boring and increase attrition. If infants are relying on intonation information, then they should fail to discriminate between the re-synthesized English and German passages. However, if infants are relying on rhythmic durational and timing information to discriminate English and German, then discrimination should still be observed'.
> (Vicenik, 2011: 104)

The results of Vicenik's investigation (2011) showed that intonation was a fundamental factor and an essential condition to distinguish American and German sentences. In fact, in the experiments on full-cue stimuli and low-pass filtered stimuli (where prosodic information was preserved), infants were able to successfully discriminate between American and German. In particular, the two languages were easily identified across low-pass filtered stimuli, where segmental information is absent. This shows that segmental cues were not necessary for discrimination.

Contrary to the previous experiments, the experiment on re-synthesized intonation stimuli (where sentences had been resynthesized to eliminate intonational information) did not work properly, as infants failed to capture differences between American and German. Since infants lacked access to intonational cues, they were no longer able to distinguish the two languages. This confirms that idea that 'infants can rely on prosodic information alone when discriminating between languages, even for prosodically similar languages like English and German' (Vicenik, 2011: 99).

5.2 The effects of pitch variation

In my production studies (see § 4.3 and § 4.4) the pitch span of the Italians speaking English was considerably narrower than that of the American English native speakers. Munro (1995) claimed that L2 speakers of English often use a narrower span, compared to the native norm. It is still an open question whether narrow span is a language-specific feature (a typical characteristic of every language) or an indicator of non-nativeness.

According to most scholars (Johns-Lewis, 1986; Mennen, 1998; Hincks, 2009), L1 speech is normally characterized by more pitch variation than L2 speech. In her study on speakers speaking German (as an L1) and English (as an L2), Mennen asserted that

> 'given that wider pitch ranges are generally perceived more positively, speakers of languages with a habitually narrower pitch range may be perceived as more negative by speakers of languages with a wider pitch range, and vice versa. It is likely that the negative perceptions towards German speakers [as compared to the English native speakers' model] could be partly due to such differences in pitch range'.
> (Mennen, 2006: 13).

According to Hincks, 'speech that is delivered without pitch variation affects a listener's ability to recall information and is not favored by listeners' (2009: 34). In addition, it has been reported that wide F0 span variation correlates with perceptions of speaker charisma (Rosenberg and Hirschberg, 2005; Strangert and Gustafson, 2008) and liveliness (Traunmüller and Eriksson, 1995; Hincks, 2005).

The use of high levels of pitch variation and the correct placement of prosodic cues in speech (such as stress and focus) conveys more liveliness and contributes to rising the interest of the listeners. English native speakers' reactions to English spoken by an international teaching assistant (a non-native speaker of English) were measured in a study by Hahn (2004). Results showed that monotonous speech, as well as the incidence of prosodic mistakes, reduced listeners' ability to process and understand information. What is more, it was demonstrated that 'when listening to speech with correct primary stress, the participants recalled significantly more content and evaluated the speaker significantly more favorably than when primary stress was aberrant or missing' (Hahn, 2004: 201). The important role of pitch variation to signal meaning, emotion and prominence in discourse has been emphasized by Wennerstrom (1994) who studied how non-native speakers use pitch and intensity contrastively, in oral presentations. She came to the conclusion that

> 'neither in oral-reading nor in free-speech tasks did the L2 (second language) groups approach the degree of pitch increase on new or contrastive information produced by native speakers. Similarly, there was less reduction of pitch and volume on redundant words in the oral reading on the part of L2 subjects relative to native speakers'
> (Wennerstrom, 1994: 415-416)

In order to test the contribution of intonation to the perception of foreign accent, Jilka (2002) carried out an experiment in which stimuli were tested under three different conditions: 1) original sentences, in which prosodic components were preserved; 2) low-passed filtered stimuli, in which segmental information was removed while intonation was preserved; 3) monotonized low-pass filtered stimuli, in which both segmental and prosodic information was removed. The three types of stimuli were tested for language identification, native speaker identification and rating of foreign accents. Results showed that, in some cases, rhythmic features of a language and speaking rate were sufficient elements to distinguish between Americans and Germans (Jilka, 2000). However, since most results did not reached statistical significance, no large effect of one prosodic cue over the others was registered. The general tendency observed in Jilka's study was that the stimuli retaining intonational information (i.e. low-passed filtered stimuli) were better rated than monotonized low-pass filtered stimuli (i.e. those lacking intonation information).

Even though many scholars (Wennerstrom, 1994; Traunmüller and Eriksson, 1995; Mennen, 1998, 2006; Jilka, 2000; Hahn, 2004; Hincks, 2005, 2009; Rosenberg and Hirschberg, 2005; Strangert and Gustafson, 2008) agree on the idea that speech with higher F0 range correlates with more positive judgments, little is known about the interaction between pitch level and span.

In her Ph.D. thesis, Chen (2005) investigated the role of pitch level and span in the perception of British English and Dutch speech. Stimuli were rated on four scales of rating parameters: 1) confident vs. not confident; 2) friendly vs. not friendly; 3) emphatic vs. not emphatic; 4) surprised vs. not surprised. The rating parameters were found to be quite good predictors of language membership. In addition, it was found that, pitch level (called by Chen 'pitch register') was less effective than pitch span in correlating with higher degree of friendliness, as shown in fig. 38:

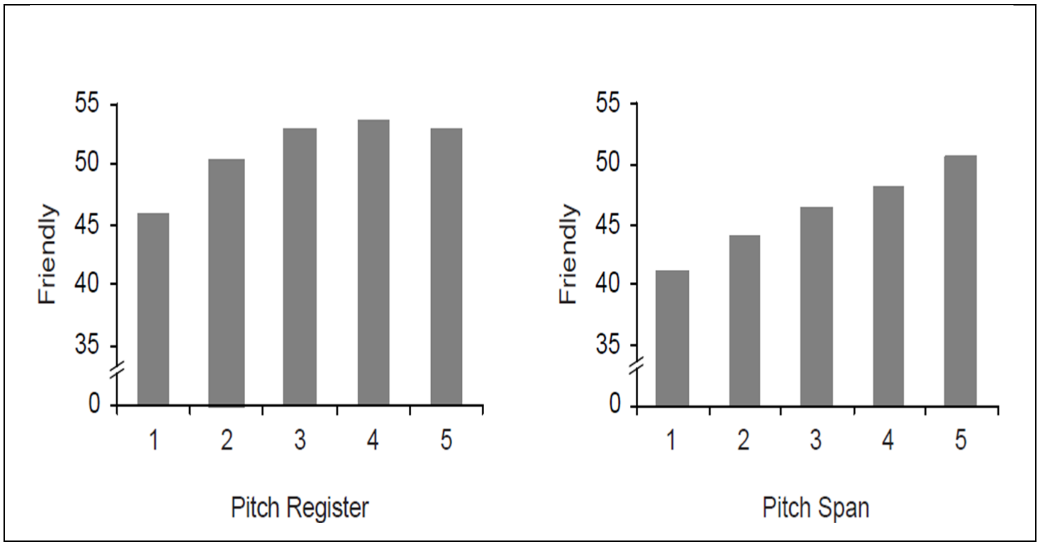


Figure 38. Ratings for 'friendly' scores calculated for 64 listeners in 5 pitch level (i.e. register) conditions and 5 pitch span conditions. (From Chen, 2005: 83).

Fig. 38 shows that 'a larger pitch span was perceived as signaling a higher degree of friendliness' (Chen, 2005: 84). The more pitch span increased from level 1 to level 5, the more the 'friendly' ratings increased. Thus, pitch span and 'friendly' ratings were directly proportional. By contrast, even though a correlation between pitch register and 'friendly' ratings was found, scores for friendliness reached their peak at the level 3 and 4 of pitch register, with a slight decrease in the degree of friendliness after level 4.

In line with the predictions of the Frequency Code (see § 2.3.1), Chen claimed that 'the wider/higher the pitch span/pitch register, the higher the perceived degree of friendliness and the lower the perceived degree of confidence are' (2005: 84).

5.3 Third experiment

The perception experiments described in this chapter were aimed at investigating how differences in the realization of pitch range in English (L1 and L2) are perceived by

Americans and Italians. In the third experiment, 30 sentences (selected from the production studies, see § 4.3 and § 4.4) were used to test whether or not the variation in pitch level and span correlates with positive or negative judgments.

Differences of pitch range were tested across two different dimensions: 1) English NSs vs. NNSs (that is Americans vs. Italians); 2) sentence types, yes-no questions vs. wh-questions vs. exclamations (henceforth YNQ vs. STM vs. EXM). As shown in previous studies (Traunmüller and Eriksson, 1995; Mennen, 1998, 2006; Jilka, 2000; Chen, 2005; Rosenberg and Hirschberg, 2005; Strangert and Gustafson, 2008), pitch span more than pitch level is considered to be a cue for non-nativeness. What is more, the experimental data obtained from my productions studies (in chapter 4) indicate that the mode of sentences is better captured by F0 span than F0 level. Given that differences of F0 level across language groups are not significant in discriminating among sentences, F0 level could be conceived as an inherent property of the voice quality and speaking style of every speaker. Unlike F0 span, F0 level is not a suitable measure to capture differences across sentence types and language group.

In the present experiment, F0 span was calculated across different stimuli (YNQ vs. STM vs. EXM) and speakers (American vs. Italian females). Then, both American and Italian listeners were asked to listen to 30 stimuli and to rate them. The listeners recruited for the perception study did not participate in the production study, so they had no previous knowledge of the materials used in the experiment. They were asked to judge the stimuli on the basis of a 5-point rating scale, assessing three different parameters: interesting vs. not interesting, exciting vs. not exciting; credible vs. not credible.

Since wide F0 span is considered to be positively perceived while narrow F0 span may attract negative scores, my prediction was that Americans are perceived more positively than Italians. Accordingly, I formulated the following hypothesis: speakers using wide pitch span are perceived as more interesting, exciting, and credible than speakers delivering a monotone speech. No matter the L1, L2 is usually characterized by a narrower pitch span (Jilka, 2000; Mennen, 2006; Hincks, 2009). Thus, narrow F0 span may be a cue to detect L2 speakers.

The goal of this experiment is to examine how F0 variation is perceived by Americans and Italians and test whether or not stimuli receive the same ratings by both the American and the Italian listeners. In addition, it is crucial to establish the direction of variation: F0 span and positive perception are directly proportional (i.e. [+ span, +

positive] and [- span, - positive]) or inversely proportional (i.e. [+ span, - positive] and [-span, + positive]). Finally, it is interesting to compare whether or not YNQ, STM, and EXM receive similar or different scores.

In sum, the characteristics of the third experiment can be outlined as follow:

Characteristics of the third experiment:

- Listeners: Americans and Italians (both male and female subjects)
- Size of the corpus: sentences produced by 5 American and 5 Italian females
- Languages: English as L1 and L2, Italian as L1
- Materials: three groups of different sentence types (YNQ vs. STM vs. EXM)

5.3.1 Research questions

In line with previous research, one might think that no matter the language, a wide pitch span variation is more positively perceived by listeners than a narrow pitch span. Since it has been found that Americans natives speakers use a wider F0 span than Italian learners of English, it is natural to infer that Italians may be negatively stereotyped also due to their small pitch range variation. The English produced by Italians is characterized by a number of segmental and prosodic features that differ from those of the native speakers. These concern patterns of prominence placement, deaccentuation, vowel reduction, the incorrect placement of stress on unstressed syllables, the replacement of final rising with final falling contours, and the distribution of pitch accents (Busà, 1995; Boula de Mareüil et al., 2004; Vieru-Dimulescu and Boula de Mareüil, 2005).

These prosodic mistakes made by Italians speaking English are mainly due to the differences in the inventories of intonation contours in the two languages. Italian and English present language-specific characteristics of rhythm and syllable length. For example, a strong Italian foreign accent is often recognized when Italian speakers 'add schwa to closed syllables, creating additional open syllables. This particular phenomenon concerning CV-structure can thus be classified as affecting phonotactic and rhythmic aspects of speech' (Jilka, 2000: 2).

As far as the placement of prominence is concerned, while in English accent is allocated according to the level of prominence of an item, Italian pitch accents 'tend to go on the final lexical item' (Grice and Baumann, 2007: 9). Thus, Italians give emphasis to the last word of an utterance, regardless of whether this word is prominent or non-prominent. Italian tends to rely on word order more than English (Ladd, 1996). The phenomenon of vowel reduction typically occurs in English as unstressed vowels are constantly reduced. By contrast, vowel reduction is seldom realized in Italian. Even though Italian is one of the languages that are strongly non-deaccenting, 'Italian also fairly readily allows deaccenting of large constituents, especially where the resulting accent is on an auxiliary, and especially in negative sentences' (Ladd, 1996: 177). According to Farnetani and Busà (1999), in English vowel reduction is a phonological process while in Italian it is not a phonological process and it occurs very sporadically at a phonetic level.

In addition to the patterns of prominence listed above, the small pitch variation and, most importantly, the considerably narrow pitch span used by Italians can be a real obstacle to successful communication in L2. The use of a narrow pitch range may be one of those factors that contribute to creating a distorted and negative image of Italian guises that are associated by English native-speakers with 'incompetence and lack of confidence, somewhat low attractiveness, but high sociability' (Eisenchlas and Tsurutani, 2011: 220). To my knowledge, the influence of narrow pitch span on the negative perception of a speaker has not been consistently confirmed by a large-scale study on cross-linguistic data. The present study is designed to measure the correlation between pitch span and positive/negative ratings of utterances produced in English by NSs (American English) and NNSs (Italians) of English.

The first hypothesis examines whether or not the fact that Italians make use of a narrower pitch range than Americans may have an influence on the way they are perceived. For this purpose, a group of American and Italian NSs are required to rate English stimuli (produced by Italians and Americans), presented in a randomized order. Stimuli are rated depending on three parameters: ± interesting, ± exciting, and ± credible. My hypothesis is that wide pitch span correlates with positive judgments and with higher levels of interest, excitement, and credibility [+interesting, +exciting, +credible]. By contrast, negative judgments [-interesting, -exciting, -credible] are expected to correlate with narrow pitch span.

The second hypothesis being tested is that the correlation between wide span and positive judgments is a universal and not a language-specific factor. Namely, I expect that both the American and the Italian listeners prefer wide span over narrow span. This assumption can be formulated on the basis of a universal preference of large pitch excursions, perceived as more lively and attractive, and thus more positively connoted. To my knowledge, there is no study showing that speakers whose L1 is characterized by narrow span perceive narrow span more positively than wide span. For example, L1 Dutch is reported to be spoken with a narrower span than L1 English (Willems, 1982; Gussenhoven, 2002; Chen, 2005). According to Chen (2005), pitch span is conceived within a relative scale in which speakers project their standard pitch span. Thus, 'speakers of the narrow-range language Dutch associate a larger meaning difference with a given interval of pitch variation than speakers of the wide-range language British English' (2005:103). My hypothesis is that, no matter the pitch span in their L1, both the Americans and the Italians perceive wide span more positively than narrow span.

Finally, the correlation between pitch span and rating scores is realized in different sentence types (YNQ, STM, and EXM). The third hypothesis tests whether or not this correlation systematically occurs in every type of sentence. It could be the case that wide pitch span is positively connoted in YNQ and negatively connoted in STM, or vice versa. In order to verify the universal meaning of F0 span variation, I assume that pitch variation is systematically correlated to specific judgements in L1 and L2 speech, no matter the differences across the sentence types investigated.

In sum, this experiment is built up in order to shed light on the effects of pitch span variation in English and Italian. In particular, the two groups of listeners (the Americans and the Italians) are expected to rate stimuli depending on the perceived pitch span. The following three directional hypotheses were formulated to be tested in the present study:

(1) F0 span is directly proportional to positive judgments, identified as [+interesting, +exciting, +credible];

(2) Assuming that (1) is confirmed by the data, this direct proportionality relation is found in scores obtained from both American and Italian raters;

(3) F0 span is directly proportional to positive judgments in every sentence type tested (YNQ, STM, and EXM).

5.3.2 Subjects

Ten adult native speakers of American English (1 male and 9 females) and 26 Italian (3 males and 23 females) subjects participated in the experiment. Tab. 38 shows the personal information about the American subjects while tab. 39 shows the personal information about the Italian subjects.

Subject	NS	Age	Town	State	Major	Gender
1	am	20	Fairfield	Connecticut	English and Italian	F
2	am	20	Boston	Massachusetts	History	F
3	am	20	Boston	Massachusetts	Art History	M
4	am	21	Cambridge	Massachusetts	Italian Film	F
5	am	21	Smithtown	New York	Psychology	F
6	am	22	Congport	New Jersey	Literature and Italian	F
7	am	24	Topsfield	Massachusetts	Linguistics	F
8	am	26	Boston	Massachusetts	Italian Literature	F
9	am	20	Bala Cynwyd	Pennsylvania	Modern Languages	F
10	am	21	New Haven	Connecticut	Marine Science	F

Table 38. Personal data of the 10 American participants in the third experiment. Information is about their native language, age, town, state, major, and gender.

All American participants were native speakers of American English and they were university students in different majors. They came from different states, in the East Coast. All Italian speakers were native speakers of Italian and they were university students in Communication Studies at the University of Padova. They took part in the experiment while attending an English class.

All students were in their second year of BA and came from the Veneto area. The age of the American participants ranged from 20 to 26 years (mean age: 22 years). The age of the Italian participants ranged from 19 to 24 (mean age: 21). None of the speakers reported any speech, hearing or communication disorder and all of them were non-smokers. There was no screening for formal training in music or singing. The students gave their consent for the treatment of their personal data (see consent form in Appendix B) and volunteered for the experiment without receiving any monetary compensation.

Subject	NS	Age	Town	Area	Class	Gender
1	it	20	Monselice	Padova	English	F
2	it	20	Lonigo	Vicenza	English	F
3	it	21	Padova	Padova	English	F
4	it	21	Padova	Padova	English	F
5	it	20	Padova	Padova	English	F
6	it	20	Camposampiero	Padova	English	F
7	it	20	Camposampiero	Padova	English	F
8	it	21	Padova	Padova	English	F
9	it	20	Este	Padova	English	F
10	it	20	Legnago	Verona	English	F
11	it	22	Asolo	Treviso	English	F
12	it	20	Arzignano	Vicenza	English	F
13	it	21	Padova	Padova	English	M
14	it	21	Monselice	Padova	English	M
15	it	21	Padova	Padova	English	M
16	it	21	Cittadella	Padova	English	F
17	it	19	Piove di Sacco	Padova	English	F
18	it	20	Padova	Padova	English	F
19	it	20	Piove di Sacco	Padova	English	F
20	it	22	Mestre	Venezia	English	F
21	it	20	Mirano	Venezia	English	F
22	it	20	Vittorio Veneto	Treviso	English	F
23	it	20	Padova	Padova	English	F
24	it	20	Venezia	Venezia	English	F
25	it	20	Padova	Padova	English	F
26	it	24	Padova	Padova	English	F

Table 39. Personal data of the 26 Italian participants in the third experiment. Information is about their native language, age, town, area, class and gender.

5.3.3 Materials

This study compared native and non-native productions of 3 sentences selected from the materials analyzed in chapter 4. In the data set, one YNQ, one STM and one EXM were produced by 10 speakers (5 American English NSs and 5 Italian NSs). A total of 30 stimuli (3 different sentences x 2 native languages x 5 subjects) were rated by 10 American English NSs and 26 Italian NSs.

The sentences were not filtered, re-synthesized or manipulated, they were selected on the basis of their mode (YNQ, STM, and EXM) and their gradual F0 variation. Since F0 span was the dependent variable to be tested, the experiment was designed to compare linear excursions of F0 span across sentences uttered by the Americans and the Italians, with wider F0 span variation reported for the Americans than for the Italians. The F0 span variation of the 15 stimuli produced by the Americans and the 15 stimuli produced by the Italians is shown in fig. 39.

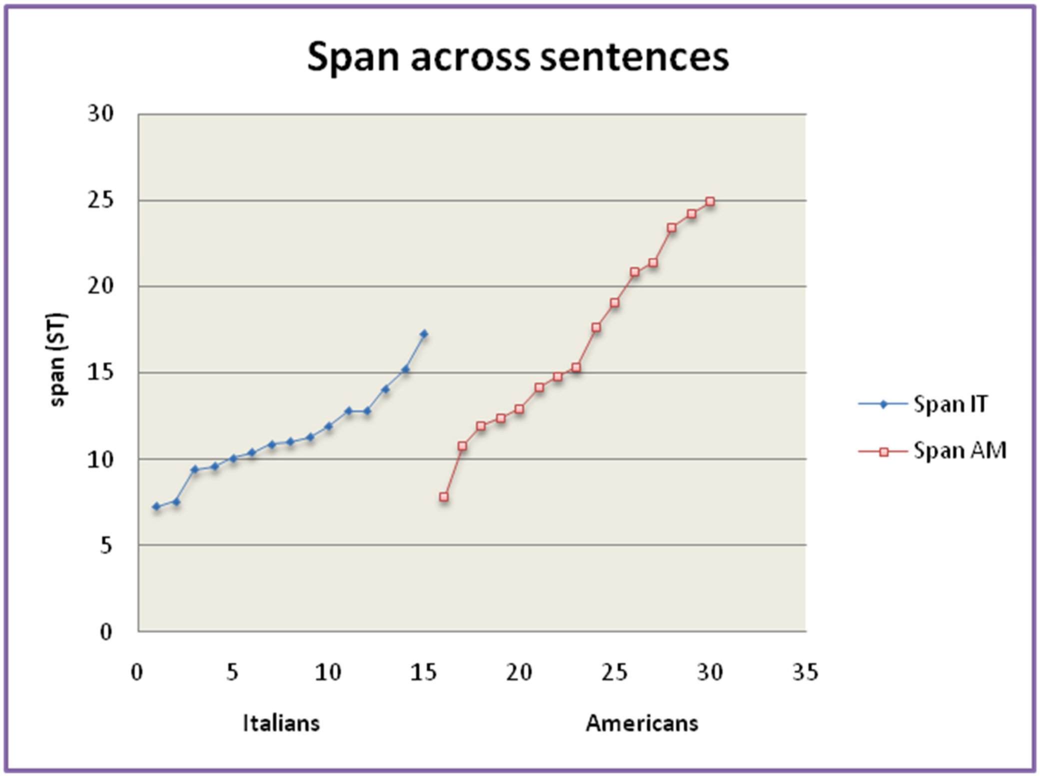


Figure 39. Values for F0 span across 30 English sentences (15 of them were produced by the Italians and the other 15 by the Americans).

Fig. 39 shows that the F0 span values calculated for the tested English stimuli ranged from 7.3 ST to 17.3 ST for the Italians and from 7.8 ST to 24.9 ST for the Americans. The mean F0 span was 11.446 ST in the utterances by the Italian NSs and 16.766 ST in the utterances by the American NSs. A difference of 5.32 ST between the mean F0 span values for the Italians and the Americans was functional to test the hypothesized differences of positive/negative judgments perceived by the English native and non-native listeners (that is, the American and the Italian raters).

As for the differences in F0 span reported across language types, tab. 40 shows the F0 span values in YNQ, STM, and EXM uttered by the Italians (I1, I2, I3, I4, and I5) and the Americans (A1, A2, A3, A4, and A5).

	F0 span calculated in different sentences		
stimuli	YNQ	STM	EXM
I1	7.3	10.4	11
I2	7.6	10.9	11.9
I3	9.4	11.3	12.8
I4	9.6	12.8	15.2
I5	10.1	14.1	17.3
SPAN Italians	**8.8**	**11.9**	**13.64**
A1	10.8	7.8	11.9
A2	12.9	12.4	20.8
A3	14.8	14.2	23.4
A4	17.6	15.3	24.2
A5	21.4	19.1	24.9
SPAN Americans	**15.5**	**13.76**	**21.04**

Table 40. Values for F0 span across YNQ, STM, and EXM. Mean F0 span values are calculated for utterances produced by Italians vs. Americans.

Tab. 40 shows that the F0 span values calculated for the tested 30 English stimuli differed across the Italians and the Americans. The F0 span in the YNQ stimuli ranged from 7.3 ST to 10.1 ST for the Italians and from 10.8 ST to 21.4 ST for the Americans. The F0 span in the STM stimuli ranged from 10.4 ST to 14.1 ST for the Italians and from 7.8 ST to 19.1 ST for the Americans. The F0 span in the EXM stimuli ranged from 11 ST to 17.3 ST for the Italians and from 11.9 ST to 24.9 ST for the Americans.

In the YNQ, the mean F0 span was 8.8 ST for the Italians and 15.5 ST for the Americans. In the STM, the mean F0 span was 11.9 ST for the Italians and 13.76 ST for the Americans. In the EXM, the mean F0 span was 13.64 ST for the Italians and 21.04 ST for the Americans. The difference between the mean F0 span values for the Italians and the Americans was: 6.7 ST in the YNQ, 1.9 ST in the STM, and 7.4 ST in the EXM. These F0 span differences were functional to test the hypotheses (formulated in § 5.3.1) and to make predictions on the nature of positive/negative judgments perceived by the English native and non-native listeners.

5.3.4 Procedure

The experiment was initially set up by means of the software for online surveys called survey monkey (see http://it.surveymonkey.com/). My first idea was to deliver my test online in order to collect data from as many listeners as possible. However, after a few trials, I realized that it was not advisable to exploit an online survey to collect data for a number of reasons. First of all, I could not control for the experimental setting and check the equipment for the test. A perception test is to be taken in a quiet room and subjects need to focus on their task without been distracted from external noises. For these reasons, listeners should be wearing headphones while listening to sentences. Reaction time is an important factor, thus listeners should not be allowed too much time to answer the questions and they should not listen to the same stimulus as many times as they wish.

Since I wanted to control for all these aspects, I decided to administer the test to a selected group of subjects (university students at the University of Padova) who were constantly monitored for all the duration of the test. Ten native speakers of American English doing their study abroad period in Padova and twenty-six native speakers of Italian taking an English class participated in the experiment. The listening session took place in a laboratory equipped with personal computers for every student. Stimuli were presented to subjects over headphones. Subjects were instructed to pay attention to the intonation of the stimuli and decide which interpretation was more appropriate for each stimulus by rating the stimuli on a 5-point scale. The 30 stimuli were divided into three blocks (with 10 stimuli for each block) and presented in a questionnaire. A training session was conducted prior to each block, to get the subjects used to the task. Every stimulus in the continuum was presented to the subjects in a randomized order, with a 3 second silent pause in between to let listeners write down their score.

Subjects accessed the test online by logging into their First Class account (http://fc.cla.unipd.it/Login). I used the First Class platform to administer the test because this software allowed me to control for and gather data quite efficiently. In fact, students were not allowed to open and access the questionnaires submitted by their colleagues, thus they could not copy from each other. They could not stop the audio files or listen to the same stimulus more than once. I supervised the delivering of the test by administering the audio files and making sure students handed in their answers in time. Subjects had 20 minutes to complete their questionnaires. At the end of the last block, they were not allowed any extra time and they had to send their questionnaire to the main conference right away.

During the training session, they were instructed to ignore the quality of the voice of the speakers and to pay attention to how the sentences were said. They had to imagine themselves as the addressees and indicate what kind of impression the different speakers made on them. As already said, the experiment consisted of three blocks: session 1, session 2, and session 3. Subjects were asked to indicate their impression on a interesting vs. not interesting scale (session 1), exciting vs. not exciting scale (session 2), and a credible vs. not credible scale (session 3). They had to put a cross on one of the five boxes allotted for each stimulus. Stimuli that sounded 'not interesting' were assigned to the left end of the scale; stimuli that sounded more 'interesting' were assigned to the right end of the scale. Students did the same for the exciting and credible scale.

After collecting all the questionnaires[12], I assigned 1 point to the stimuli that had obtained the worst ratings (not interesting, not exciting, and not credible) and 5 points to the stimuli that had obtained the best scores. In the end, I transferred all the data onto an excel files where I compared stimuli uttered by the Americans and the Italians. For each stimulus, I analyzed the scores given by the two groups of listeners and calculated the F0 span.

5.3.5 Results

The results showed that the listeners in the present experiment (both the Italians and the Americans) rated the sentences depending on the different F0 span values of YNQ, STM and EXM. A correlation between wide span and positive judgments was found in the scores given by both the American and the Italian listeners. In fact, regardless of their native language, all listeners perceived large pitch excursions (that is wide span) as being more lively and attractive than narrow span. By contrast, the scores given to narrow span stimuli were much lower than those given to wide span stimuli. This means that both the Americans and the Italians perceived wide span more positively than narrow span.

Fig. 40 shows the correlation between the rating scores (on the y-axis) and the values of F0 span calculated for the stimuli (on the x-axis).

[12] A sample spreadsheet of the questionnaire is in Appendix B.

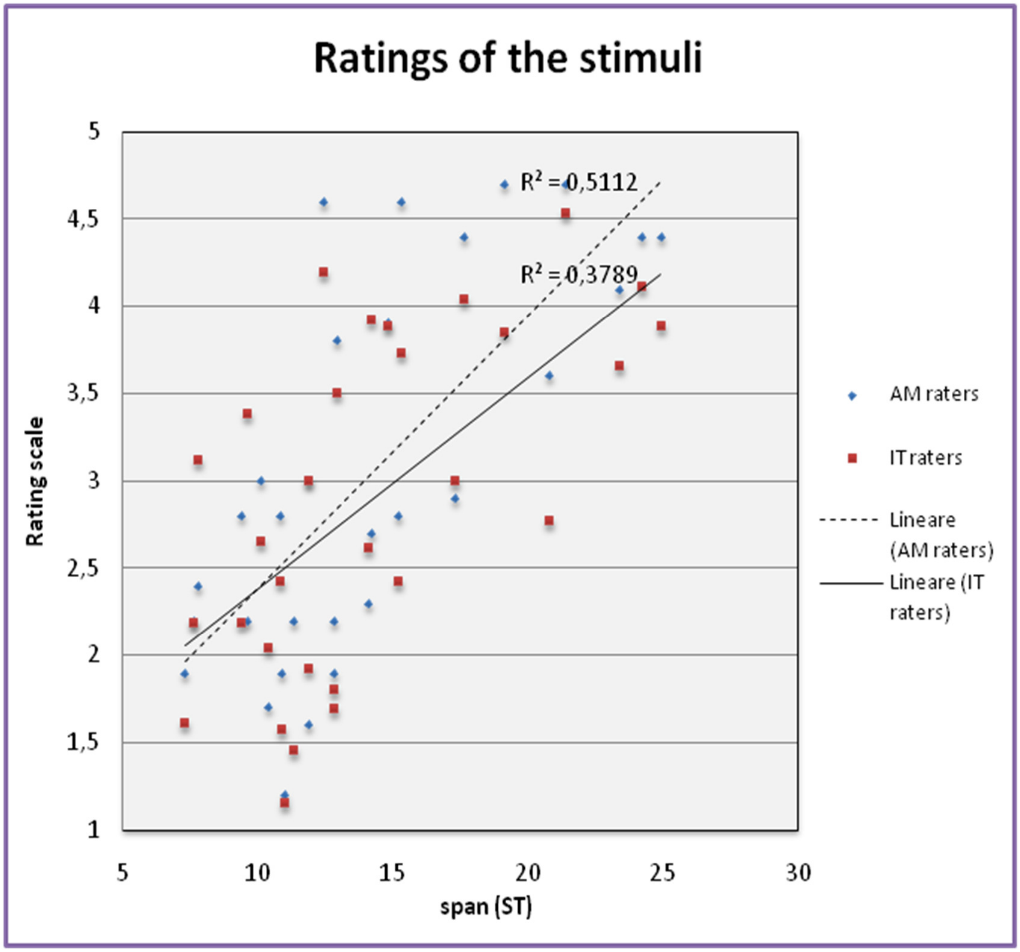


Figure 40. Ratings of the stimuli obtained from the American raters (blue diamond) and the Italian raters (red square). The rating scores obtained for the stimuli (on the y-axis) are plotted against the values of F0 span (on the x-axis).

The scatterplot in fig. 40 shows how the values for one group (e.g. the American raters) compare to the other (e.g. the Italian raters), and is useful for showing correlations. The mean scores given to the stimuli by the American and the Italian raters are plotted on the x-axis and the mean F0 span values are plotted on the y-axis. The data show that the ratings produced by the American and the Italian raters increase linearly as the F0 span of the stimuli increases. This means that the ratings given to the stimuli are strongly correlated with the values of F0 span across the sentences.

The trend lines (R^2 values) calculated for each group of raters (Americans vs. Italians) are measured by fitting a straight line to the spread of each group. The R^2 values calculate how much variance in the actual data is accounted for by the fitted lines.

The R^2 values displayed on the chart show that the scores by the American raters are more correlated to F0 span than the scores by the Italian raters (with R^2 of the American raters equal to 0.5112 and R^2 of the Italian raters equal to 0.3789). Overall, a strong correlation between scores and F0 span was found between both groups of raters.

On average, the stimuli produced by the Italians (i.e. I1, I2, I3, I4, and I5) and the Americans (i.e. A1, A2, A3, A4, and A5) received similar scores from the two groups of raters. Tab. 41 shows the rating scores given by the American and the Italian listeners to the English stimuli produced by a group of American and Italian speakers.

	YNQ		STM		EXM	
stimuli	AM raters	IT raters	AM raters	IT raters	AM raters	IT raters
I1	1.7	2.0	1.9	1.6	1.2	1.1
I2	1.9	1.5	2.2	2.1	1.6	1.9
I3	1.9	1.8	2.2	3.3	2.2	1.6
I4	2.2	1.4	2.8	2.1	2.8	2.4
I5	2.3	2.6	2.8	2.4	2.9	3.0
A1	2.4	3.1	3.0	2.6	3.0	3.0
A2	2.7	3.9	3.8	3.5	3.6	2.7
A3	4.6	3.7	3.9	3.8	4.1	3.6
A4	4.6	4.1	4.4	4.0	4.4	4.1
A5	4.7	3.8	4.7	4.5	4.4	3.8
MEAN SCORE	2.90	2.79	3.17	2.99	3.02	2.72
IT STIMULI	2.00	1.86	2.38	2.30	2.14	2.00
AM STIMULI	3.80	3.72	3.96	3.68	3.90	3.44

Table 41. Rating scores of English stimuli produced by some Americans and Italians. The stimuli were rated by the Americans and the Italians across sentence types.

The stimuli received similar scores from the group of the American and the Italian raters. On average, YNQ received a mean score of 2.90 points from the American raters and 2.79 points from the Italian raters. STM received a mean score of 3.17 points from the American raters and 2.00 points from the Italian raters. EXM received a mean score of 3.02 points from the American raters and 2.72 points from the Italian raters. In sum, the ratings expressed by the two groups of listeners were fairly similar, with the ratings given by the

American listeners slightly higher than those given by the Italians. Thus, the Italian raters were somehow stricter than Americans.

The scores obtained for the stimuli produced by the Americans compared to those produced by the Italians were strikingly different, with the American stimuli scoring higher than the Italian ones across every sentence type. The American raters gave higher scores to utterances produced by the American NSs as compared to the utterances produced by the Italian NSs.

Within the scores given by the **American raters**, the YNQ produced by the Italians received a mean score of 2.00 points while those produced by the Americans were rated 3.80 points. As for the STM, the stimuli produced by the Italians were rated 2.38 points while those produced by the Americans were rated 3.96 points. EXM received a mean score of 2.14 points for the stimuli produced by the Americans and 3.90 points for the stimuli produced by the Italians. In sum, ratings expressed for the two groups of stimuli (Italians' stimuli vs. Americans' stimuli) were considerably different. The American listeners gave higher evaluations to the Americans' stimuli than the Italians' stimuli.

Within the scores given by the **Italian raters**, the YNQ produced by the Italians received a mean score of 1.86 points while those produced by the Americans were rated 3.72 points. As for the STM, the stimuli produced by the Italians were rated 2.30 points while those produced by the Americans were rated 3.68 points. EXM received a mean score of 2.00 points for the stimuli produced by the Americans and 3.44 points for the stimuli produced by the Italians. In sum, also the Italian raters found strong differences between the Italians' and Americans' stimuli. The Italian listeners gave higher evaluations to the Americans' stimuli than the Italians' stimuli.

The stimuli were rated along three different scales: ± interesting, ± exciting, and ± credible. It was found that wide pitch span correlates with positive judgments and with higher levels of interest, excitement, and credibility [+interesting, +exciting, +credible]. By contrast, narrow pitch span correlates with negative judgments [-interesting, -exciting, -credible]. Since the Americans' stimuli were characterized by a wider pitch span than the Italians' stimuli, the American NSs were perceived more positively than the Italians NSs.

The linear correlation between pitch span and raters' judgments is shown by the three rating scales, measured in the present experiment: ± interesting (fig. 41), ± exciting (fig. 42), and ± credible (fig. 43).

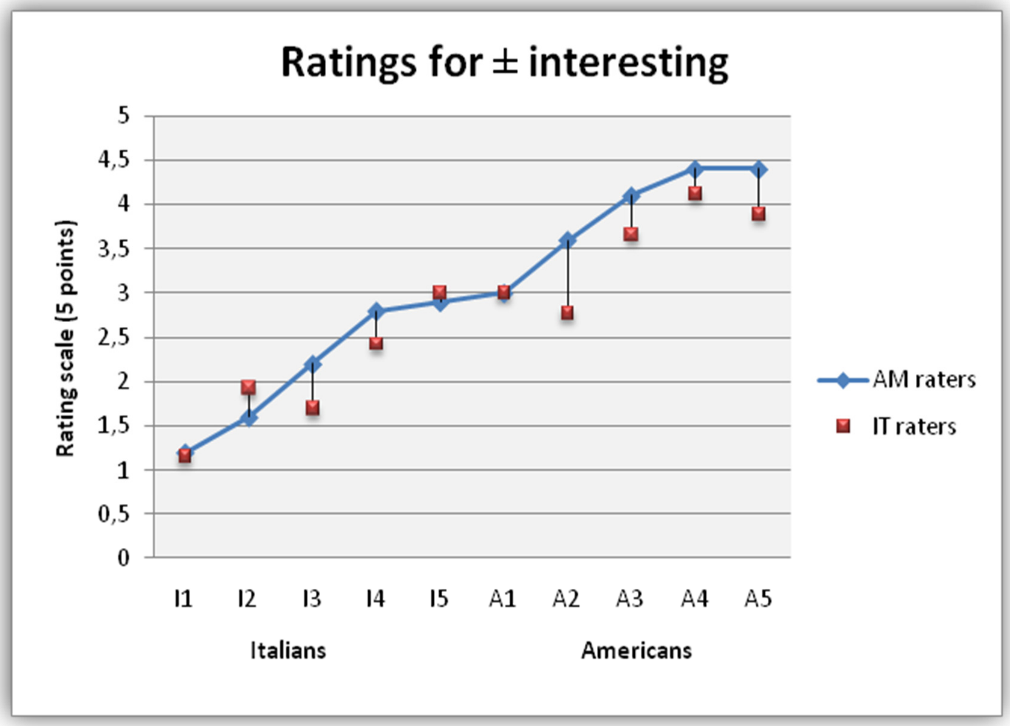


Figure 41. Ratings along the '± interesting' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners.

The scores describing the perception of **interest** generated in the raters by the different stimuli are shown in fig. 41. The correlation between the judgments given by the American and the Italian raters is very high. The blue line represents the general trend for the scores given by the American raters. Being native speakers of American English, the American raters are expected to be able to distinguish among the interesting and not interesting stimuli by relying on a number of prosodic cues, including F0 span. The average rating scores given by the Americans are compared to those given by the Italians (red squares). From the graph (in fig. 41), it is evident that the scores assigned to the stimuli from the two groups of raters are very similar.

In fact, the scores for the stimuli produced by some Americans (e.g. A1 and A4) and Italians (e.g. I1 and I5) are almost equal. Most of the stimuli were judged to be slightly less interesting by the Italian raters, as compared to the American raters. In particular, the stimulus produced by the American subject 2 (i.e. A2) received a mean score of 3.6 points from the American raters and 2.7 points from the Italian raters. One-way ANOVAs with 'native language of the raters' as between-subject factor were run to calculate the statistical difference among the scores of the two groups of raters (Americans vs. Italians). No significant difference was found across the ratings produced by the two groups of listeners (Americans vs. Italians).

The data collected from the listeners, whose task was to rate how excited the speakers of the stimuli sounded, seemed to be quite similar. In fact, for both the American and Italian raters, the higher the F0 span of the stimuli, the more exciting the stimuli were rated. The scores describing the perception of **excitement** generated in the raters by the different stimuli are shown in fig. 43.

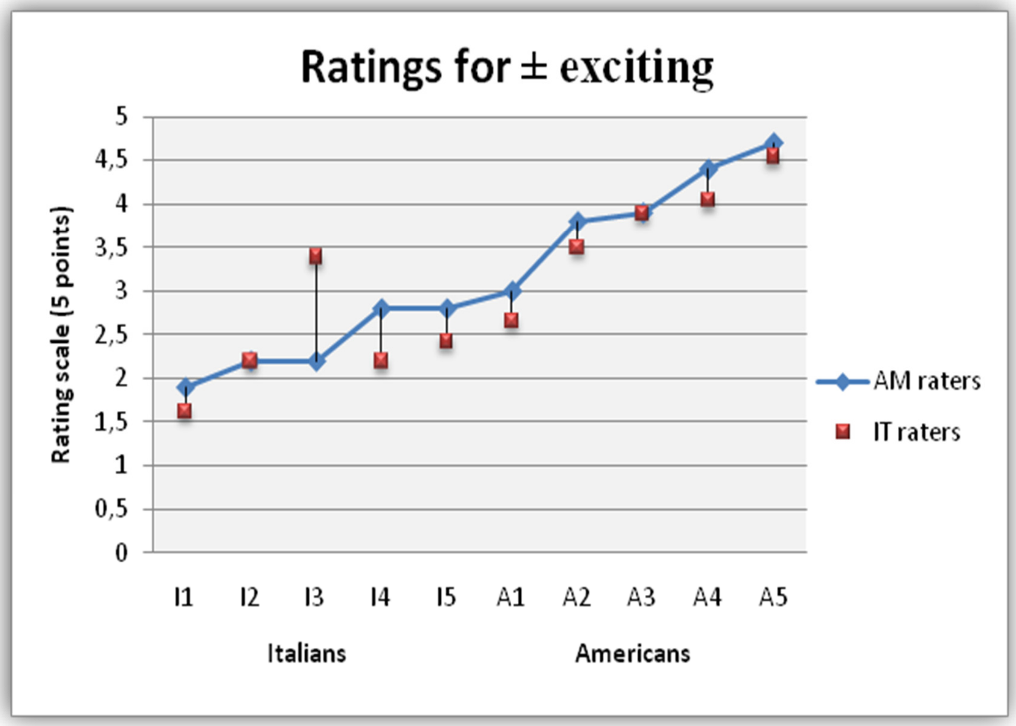


Figure 42. Ratings along the '± exciting' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners.

Fig. 42 shows that the correlation between the judgments given by the Americans and the Italian raters was very high, with scores for the stimuli produced by most of the American and Italian speakers being very similar. Only the stimulus produced by the Italian subject 3 (i.e. I3) received a mean score of 2.2 points from the American raters and 3.3 points from the Italian raters, showing a difference of 1.1 points. All the other stimuli were judged as slightly less exciting by the Italian raters, as compared to the American raters. The one-way

ANOVAs with 'native language of the raters' as between-subject factor showed no significant difference across scores by the American and the Italian raters.

The scores describing the perception of **credibility** generated in the raters by the different stimuli are shown in fig. 43.

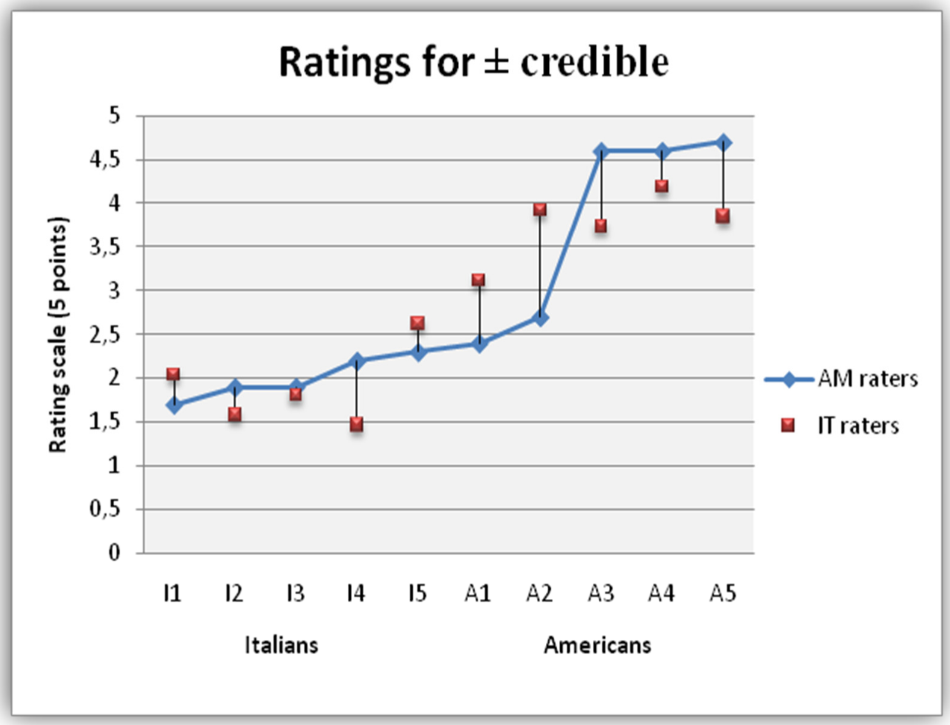


Figure 43. Ratings along the '± credible' evaluation. The stimuli produced by the Italian vs. the American native-speakers were rated on a 5 points-scale. The stimuli were rated by the American (blue diamond) vs. the Italian (red squares) listeners.

As shown in fig. 43, the correlation between the judgments given by the Americans and the Italian raters was high, with scores for the stimuli produced by some Americans less similar than those produced by the Italians (compare A 1-5 to I 1-5). Compared to the American raters, the Italian raters judged some stimuli as less credible (i.e. I2, I3, I4, A3, A4, A5) and some other stimuli as more credible (i.e. I1, I5, A1, A2). The difference among the scores allotted by the American raters and those allotted by the Italian raters were not large. Only the stimulus produced by the American subject 2 (i.e. A2) registered a mean score of 3.9 points from the American raters and 2.7 points from the Italian raters, with a difference of 1.2 points. The statistical difference of the scores given by the two groups of raters

(Americans vs. Italians) was tested in one-way ANOVAs with 'native language of the raters' as between subject factor. No significant difference was found across the ratings.

Every rating scale ('± interesting', '± exciting', and '± credible') was found to be a good predictor of F0 span variation, as it correlated with more or less positive judgments. Unlike the 'credible' parameter, the parameters 'exciting' and 'interesting' seemed to be equally distributed across the stimuli, for both the Italian and the American raters. The ratings concerning the credibility of the stimuli (fig. 43) appeared to be less strongly correlated than the 'interesting' and 'exciting' ratings. This may be due to the fact that Americans raters perceived as more credible some stimuli that appeared to be less credible for the Italian raters. Thus, minor differences across data for the American and Italian raters were found: a change in one direction (e.g.+credible) for the American raters did not yield a patterned change for the Italian raters. In spite of this difference, the credibility ratings registered between the American and the Italian groups of raters were found to be not significantly different, with a difference of only 0.06 points between the mean 'credibility' scores allotted by the Americans and the Italians.

The correlation between pitch span variation and rating scores was found in different sentence types (YNQ, STM, and EXM). Thus, F0 span was directly proportional to positive judgments and this linear correlation systematically occurred in every sentence type tested in the present experiment. In order to verify the effect of F0 span variation on different sentences, pitch span variation was correlated to the judgements given by both the American and Italians listeners.

The ratings of YNQ are shown in fig. 44. The scores given by the American and the Italian listeners (on the x-axis) are plotted against the F0 span values of the English stimuli (on the y-axis).

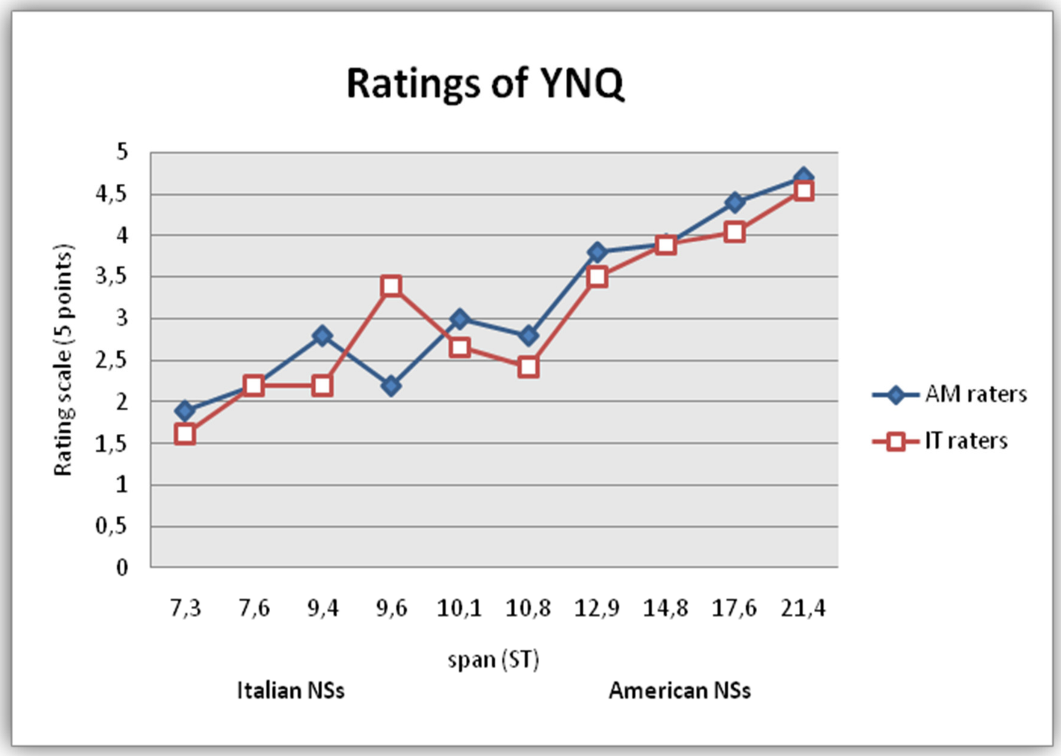


Figure 44. Ratings of YNQ. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel).

The scores given to the Italians' YNQ, whose pitch span ranged from 7.3 ST to 10.1 ST, were much lower than the scores given to the Americans' YNQ (that ranged from 10.8 ST to 21.4 ST). In fact, the Italians' stimuli scored about 1.5-3.0 points while the American's stimuli scored about 2.5-4.7 points. Thus, the rating range of the two groups of stimuli was significantly different, as indicated by a t-test based on the ratings given by the American listeners ($t(4)=6.51$ at the .05 level, $p<0.002$) and a t-test based on the ratings given by the Italian listeners ($t(4)=5.29$ at the .05 level, $p<0.006$).

The ratings of STM are shown in fig. 45. The scores given by the American and the Italian listeners (on the x-axis) are plotted against the F0 span values of the English stimuli (on the y-axis).

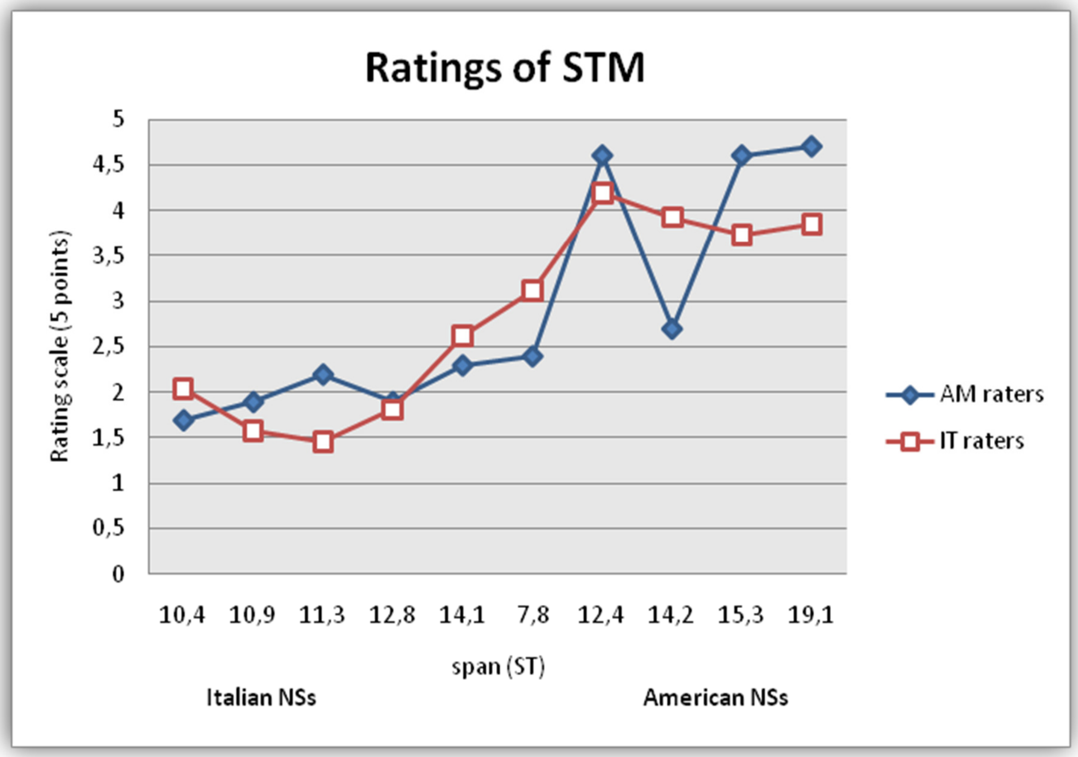


Figure 45. Ratings of STM. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel).

The scores given to the Italians' STM, whose pitch span ranged from 10.4 ST to 14.1 ST, were much lower than those of the Americans' STM (that ranged from 7.8 ST to 19.1 ST). In fact, the Italians' stimuli scored about 1.7-2.6 points while the American's stimuli scored about 2.4-3.8 points. Thus, the rating range of the two groups of stimuli was significantly different, as indicated by a t-test based on the ratings given by the American listeners ($t(4)=3.64$ at the .05 level, $p=0.02$) and a t-test based on the ratings given by the Italian listeners ($t(4)=5.96$ at the .05 level, $p<0.003$).

The ratings of EXM are shown in fig. 46. The scores given by the American and the Italian listeners (on the x-axis) are plotted against the F0 span values of the English stimuli (on the y-axis).

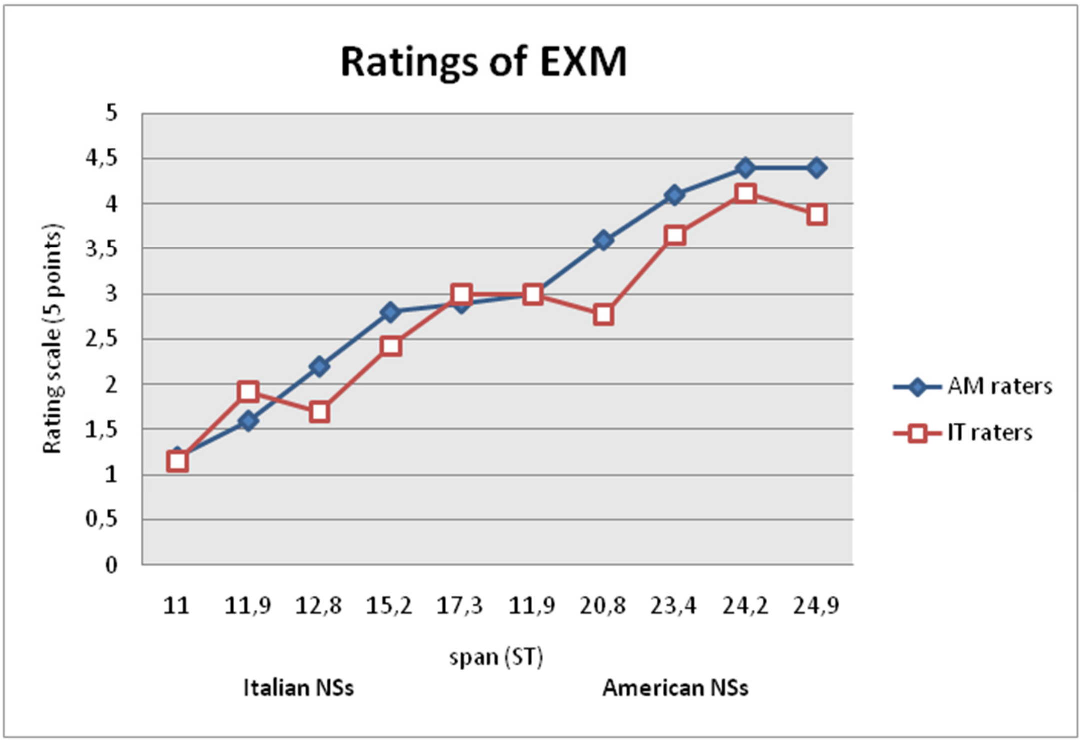


Figure 46. Ratings of EXM. The scores given by the Americans (blue diamond) vs. the Italians (red squares) are plotted against the F0 span values calculated for the English stimuli produced by the Italians (left side of the panel) and the Americans (right side of the panel).

The scores given to the Italians' EXM, whose pitch span ranged from 11 ST to 17.3 ST, were much lower than those of the Americans' EXM (which ranged from 11.9 ST to 24.9 ST). In fact, the Italians' stimuli scored about 1.2-3.0 points while the American's stimuli scored about 3.0-4.4 points. Thus, the rating range of the two groups of stimuli was significantly different, as indicated by a t-test based on the ratings given by the American listeners ($t(5)=18.97$ at the .05 level, $p<0.001$) and a t-test based on the ratings given by the Italian listeners ($t(5)=6$ at the .05 level, $p<0.003$).

5.3.6 Discussion

The third experiment was designed to shed light on the effects of pitch span variation on the perception of English stimuli produced by the American and the Italian NSs. Two groups of listeners (Americans vs. Italians) rated the stimuli with larger pitch span as more interesting, exciting and credible than the stimuli with narrower pitch span. Thus, the listeners relied on the perceived pitch span to differentiate among the stimuli. This means that pitch span is an important cue that both the American English and the Italian NSs associate to more or less positive judgements.

The first hypothesis, stating that F0 span is directly proportional to positive judgments identified as [+interesting, +exciting, +credible], was confirmed. Both the American and the Italian listeners rated the stimuli with large pitch span excursions with higher scores than the stimuli with narrow span variation. In addition, the results showed that the more the pitch span increases the more positive judgments are given to the stimuli. The relation between pitch span and ratings of the stimuli was found to be directly proportional, across the three rating scales.

The second hypothesis, investigating whether or not this direct proportionality relation is found in the scores given by both the American and Italian listeners, was confirmed. In fact, no matter their native language, both groups of (English L1 and L2) listeners perceived the stimuli with wide span variation more positively than the stimuli with narrow pitch variation. This is in line with the idea that, regardless of their language background, 'speakers are capable of interpreting phonetic implementation in the speech of the others (in known or unknown languages) in accordance with the universal paralinguistic meanings of pitch variation' (Chen, 2005: 43). Therefore, pitch span variation is not only a phonetic cue to L2 speech perception but also an indicator of universal paralinguistic meaning.

Finally, the third hypothesis formulated in the present experiment was also confirmed, as the data showed that F0 span is directly proportional to positive judgments in every tested sentence types (YNQ, STM, and EXM). The correlation between pitch span and the scores of the tested stimuli was realized across all three types of sentence tested. The patterns of YNQ, STM, and EXM were characterized by different pitch span values that were rated by the listeners depending on the size of the pitch excursions: the larger the span, the higher the scores. In addition, the results showed that the differences between the stimuli produced by the Americans and the Italians were statistically significant for each sentence type.

In conclusion, the effect of pitch span variation can be described as both universal and language-specific. As reported in the literature, in any language, the paralinguistic meaning of pitch variation is universal while the linguistic meaning is language-specific (Gussenhoven, 2002). The results of my study provide support to this claim by showing that there is a correlation between pitch span variation and the positive/negative perception of stimuli for both language groups under investigation (Americans vs. Italians). Also the language-specificity of pitch variation was assessed in the results. In fact, depending on the tested native language (Italian vs. American), the stimuli were perceived as more or less interesting, exciting and credible. This is an effect of pitch span variation that was shown to be narrower in the stimuli produced by the Italians and wider in the stimuli produced by the Americans.

5.4 Fourth experiment

The fourth experiment was designed to explore the contribution of pitch range to the perception of YNQ in English manipulated stimuli, based on sentences pronounced by Americans and Italians. To this end, a perception test was created with stimuli that contained a gradual increase and decrease of the pitch span at the end of the intonation contours to test whether there is a correlation between an increase in pitch span and evaluations of intonation contours. Manipulated YNQ pronounced by an English NS and an English NNS were mixed together in three pitch span conditions (narrow vs. original vs. wide pitch span).

Listeners had to rate sentences depending on the pitch span differences they perceived. In particular, the aim of the experiment was to test the impact of pitch span, conceived as a determinant factor for distinguishing friendly from unfriendly speakers. In order to manipulate pitch span across sentences, sentences were manually re-synthesized to modify (step by step) the pitch span in one selected part of the utterances. The resulting stimuli were perceived by listeners as natural sentences and every suprasegmental cues was preserved.

The characteristics of the fourth experiment can be outlined as follow:

Characteristics of the fourth experiment:

- Listeners: Italian listeners (both male and female subjects)
- Size of the corpus: 3 yes-no questions produced by an American and an Italian female subject
- Languages: English as an L1 and an L2
- Materials: re-synthesized sentences under three manipulation conditions

5.4.1 Manipulation techniques

Several manipulation strategies were took into consideration in order to generate stimuli in which pitch span variation was the main factor of difference across the sentences. To create the perfect conditions for the experiment (including the limitation of the effects of segmental cues), the flat intonation re-synthesis and the low-pass filtered manipulation were evaluated. It has to be remarked that these methodologies have been largely adopted in recent studies (Munro 1995; Ramus and Mehler, 1999; Jilka, 2000; Boula de Mareüil and Vieru-Dimulescu, 2006; Trofimovic and Baker, 2006; Munro et al., 2010) to mask either segmental and suprasegmental features. As Rognoni (2012) asserted

> ‘The main problem when testing the impact of single prosodic aspects lies in the fact that, in natural speech, prosody cannot be disentangled from the segmental dimension. One way to deal with these problems is to manipulate the speech signal and degrade or remove some parts of the information while preserving others, in order to separate the different streams of information and to evaluate their relative contribution in speech perception.’
> (Rognoni, 2012: 89)

Tab. 42 shows three versions of the sentence ‘Do you wanna come for dinner?’, uttered by an American native speaker. The top panel shows the original version (the non manipulated sentence), the mid panel shows the low-pass filtered version, and the bottom panel shows a flat intonation re-synthesis.

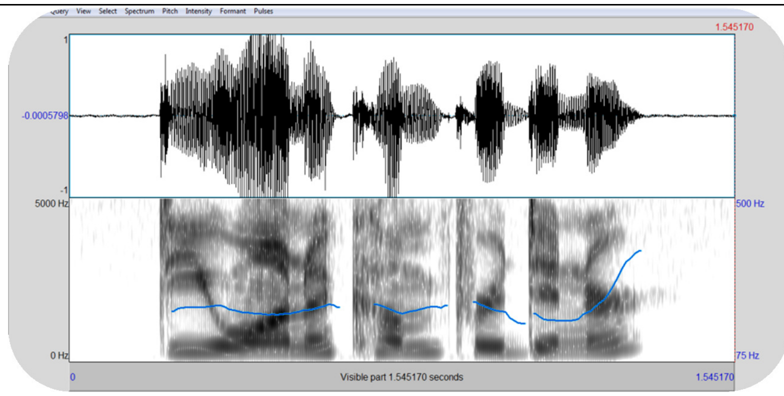
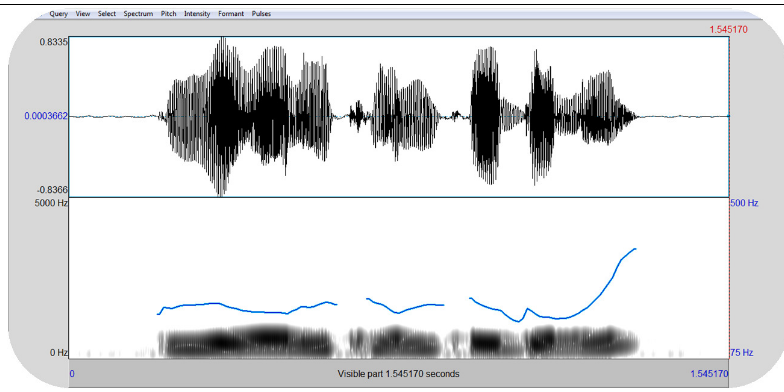
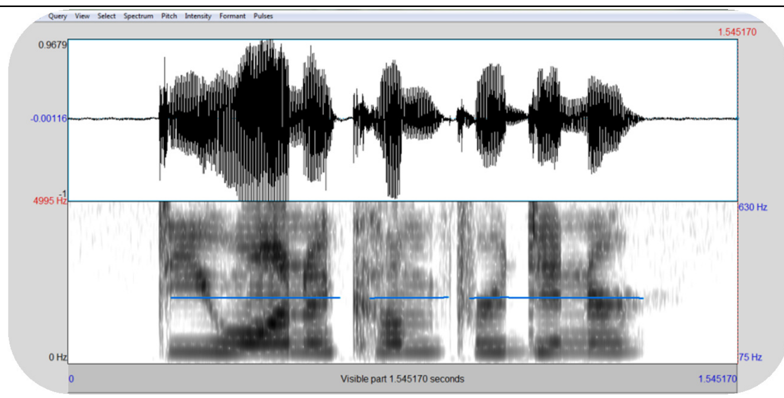

Original utterance: segmental + suprasegmental information is available

Low-pass filtered: only suprasegmental information is avaliable

Flat intonation re-synthesis: only segmental information is available

Table 42 . Three versions of the same sentence: original version (non manipulated sentence), low-pass filtered version, and re-synthesized version with flat intonation. The manipulated sentence is 'Do you wanna come for dinner?' uttered by an American native speaker.

A flat intonation re-synthesis can be realized by using the ‘Flat Intonation Re-synthesizer’ script, created by Chad Vicenik (it is available on the website of the UCLA phonetic lab (http://www.linguistics.ucla.edu/faciliti/facilities/acoustic/praat.html). This script permits to re-synthesize all the sound files in a specified directory and to create stimuli with a flat pitch of the specified frequency. The flat intonation re-synthesis technique consists in the monotonization of F0. The F0 contour is leveled to a constant value that may be selected and modified by the experimenter so that, the intonation contour is transformed into a flat line. Through the flat intonation re-synthesizer, the segmental information is preserved while the suprasegmental information is erased. As the bottom panel in tab. 42 shows, the rises and the falls in the pitch line are not tracked. Thus the resulting stimulus is totally flat.

The Low-pass filtering is a technique that permits to erase ‘segmental influences on listeners’ judgments and to evaluate the contribution of prosody’ (Jilka, 2010a: 115). The manipulation of stimuli with the low-pass filter consists of removing high F0 values and preserving low F0 values. Thus, the segmental information is removed while the suprasegmental information is preserved.

> ‘In the low-pass filtered signal only prosodic features such as fundamental frequency, stress, rhythm and speaking rate remain available to further analysis, although with a diminished information content as there is no direct relation to the segmental basis. The low-pass filtered signal approximately recreates the impression of indistinct murmuring as one would perceive speech from an adjacent room’.
> (Jilka, 2010a: 115)

As the mid panel in tab. 42 shows, the rises and the falls in the pitch line are tracked. Thus the intonation pattern of a sentence is preserved. The erasing of F0 values over about 150 Hz makes the sentence sound unintelligible. In fact, low-pass filtering masks the content of speech so that one cannot understand what is being said.

This effect can be appropriate to erase segmental information; nonetheless it can generate biased results in the experiments. According to Chen and Gussenhoven (2004: 317) ‘using unintelligible speech (i.e. low-pass filtered stimuli) appears to lead to fuzzier results with smaller effect sizes’. What is more, low-pass filtering may have a negative impact on the results of a study because subjects fail to ‘discriminate between the re-synthesized stimuli, not because the missing intonational cues eliminated important distinctions between the languages, but simply because the re-synthesis made the stimuli sound unnatural or disturbing' (Vicenik, 2011: 108). Indeed,

> 'Low-pass filtering is not an ideal way to degrade utterances with the aim of deleting segmental information and preserving prosody. Basically, filtering does not allow one to know which properties of the signal are eliminated and which are preserved. As a first approximation, segmental information should be eliminated because it is mainly contained in the higher formants of speech, and pitch should be preserved because it rarely rises higher than 400 Hz. But this is only an approximation'.
> (Ramus and Mehler, 1999)

The advantage of working with the flat intonation re-synthesis and the low-pass filtering is that the experimenter can modify large amount of data in a few-seconds time. In fact, the process is completely automatic and there is no need to do any manual correction. As discussed above, the flat intonation re-synthesis and the low-pass filtering techniques may not be the best techniques for testing the perception of intonation contours. For this reason, I decided to discard those filtering techniques in favor of another kind of manipulation that allows the experimenter to obtain natural sounding stimuli. In order to manipulate pitch span across sentences, I decided to manually re-synthesize sentences and to modify (step by step) the pitch span in one selected part of the utterances.

5.4.2 Research questions

Since in the third experiment, it was found that both the American and the Italian listeners based their ratings of the stimuli depending on F0 span variation, the goal of the present study is to test whether or not the same linear correlation can be found in stimuli manipulated with changes in F0 span variation. This study is designed to measure the correlation between pitch span and the listeners' evaluations of English utterances produced by NSs (Americans) and NNSs (Italians) of English.

Voice quality may contribute, even indirectly, to judgments of pitch variation (Bishop and Keating, 2012) and may also have an effect on the perception of attractiveness or friendliness (Feinberg et al., 2008). Undoubtedly, some husky and guttural voices may sound more or less attractive than squeaky and high-pitched voices. In particular, a recent study (Feinberg et al., 2008) examined the existence of a positive linear relationship between voice pitch and attractiveness ratings. Manipulated pitch in women's voices with low (lower than average), average, and high (higher than average) starting pitches were rated by female and male listeners. It was found that men preferred high-pitched voices while women preferred average-pitched voices.

One of the criticism that could be drawn to an experiment comparing stimuli produced by different speakers is that voice quality might play a role in distinguishing among the relative attractiveness of the stimuli. Thus, since I intend to measure the effect of pitch span (and not voice quality) on the listeners, to compare stimuli produced by different speakers might not be a good idea. In order to test the independent role of pitch span, for the present experiment, I selected only two speakers (an American NS and an Italian NS) with similar pitch characteristics (similar F0 level and span). I did so because I needed to test a single factor (pitch span) out of all the differences in the stimuli produced by the American NSs and the Italian NSs. For this reason, sentences produced by two speakers having a pitch range as similar as possible were used for the manipulation and re-synthesis process.

The first hypothesis being tested examines the influence of pitch span on the perception of friendliness. In order to test the impact of different ranges of pitch span, original sentences produced by an American and an Italian NS were manipulated and re-synthesized. Pitch span was manipulated with Praat to create three types of stimuli: narrow pitch span, original pitch span, and wide pitch span stimuli. In the narrow pitch span condition, the pitch span value of each stimulus was lowered in order to obtain flatter sentences than the original ones. In the original pitch span condition, the pitch span value was kept identical to the one of the original unmodified sentence. In the wide pitch span condition, the pitch span value of each stimulus was raised in order to obtain higher final slopes than those in the original sentences.

In line with the results obtained in the third experiment, the speakers uttering original span stimuli are expected to be perceived on average as moderately friendly, with Americans rated as more friendly than the Italians. Assuming that the role of pitch span is determinant in the productions of positive judgments, the wide span stimuli are expected to be perceived as much more friendly than the narrow span stimuli. Thus, my hypothesis is that wide pitch span correlates with friendly judgments and that this correlation is found across the three types of manipulated stimuli (narrow vs. original vs. wide span).

The second hypothesis being tested is that the correlation between pitch span and judgments of friendliness is a universal and not a language-specific factor. Namely, I expect that listeners prefer wide span over narrow span stimuli, regardless of the native language of the speakers. This assumption can be formulated on the basis of the universal preference exhibited by listeners for large pitch excursions, perceived as more lively and attractive (see Gussenhoven, 2002; Chen, 2005; Hincks, 2005; Rosenberg and Hirschberg, 2005; Feinberg et al., 2008; Hincks and Edlund, 2009; Gravano et al., 2011).

When comparing English stimuli produced by American and Italian NSs, it is natural that foreign accent plays a role in the perception of friendliness. In fact, accented speech may be disfavored because of the perceived foreign accent. However, due to the lack of context, the task of detecting foreign accent in a perception experiment (where sentences are perceived in isolation) is more difficult than it is in a natural conversation (Jilka, 2000). In a spontaneous conversation, listeners have more clues at their disposal to help them distinguish among L1 and L2 speakers.

In order to limit the influence of foreign accent on the judgments of the stimuli, only Italian listeners were selected for the present experiment. Since L2 speakers are expected to be less capable than L1 speakers to detect foreign accent, it is likely that L2 listeners fully rely on prosodic cues (in particular, pitch span) rather than foreign accent, when rating the stimuli. Thus, the Italian listeners are expected to rate the stimuli produced by the American and the Italian NSs, depending on the span values variation. If the second hypothesis is confirmed, this shows that the contribution of pitch span is relevant in distinguishing narrow span from wide span conditions, regardless of the native language of the speakers. The wide span stimuli produced by L1 and L2 English speakers are expected to be rated as more friendly than the narrow span stimuli.

The third hypothesis being tested focuses on the different judgments formulated by male vs. female listeners. Depending on socio-cultural and emotional factors, male and female subjects may display different rankings for pitch range variation (van Beezoijen, 1995; Feinberg et al., 2008; Yuasa, 2008). Intonation is considered as one of the main means to convey affective information (Ladd, 1980) and pitch modulation is universally related to the expression of different emotions (such as surprise, astonishment, anger, excitement etc.). Typically, different kinds of emotions are transmitted by changes in pitch range. The major tendency is that ‘with excitement there are greater extremes of pitch; with depression the range is narrowed’ (Bolinger, 1989: 13).

It has been argued that ‘individuals may shape the pitch range of their speech to fit the prevailing pitch range in a given linguistic community’ (Dolson, 1994: 323). Deutsch and colleagues (1992, 2009) found that ‘the pitch level of an individual’s speaking voice is strongly influenced by the pitch levels of speech in his or her linguistic community’ (2009:

208). In fact, some individuals (e.g. females) may perceive wide pitch span as more suitable for their pitch range target while other individuals (e.g. males) may consider the same pitch span as not appropriate. Indeed, not only females, but also males are sensitive to socio-cultural effects that influence their perception of pitch range. This is due to the fact that, 'through long-term exposure to the speech of others, the individual acquires a mental representation of the expected pitch range and pitch level of speech (for male and female speech taken separately)' (Deutsch et al., 2009: 208). Everything considered, one would expect that males and females create in their mind different representations of pitch range and, thus they formulate different expectations and quality judgments of pitch range. The aim of the third hypothesis is to test whether or not male and female listeners have similar expectations about pitch range.

In sum, this experiment was built up to shed light on the effects of pitch span variation across manipulated sentences with narrow, original and wide pitch span. The group of Italian (male vs. female) listeners are expected to rate the stimuli depending on their mental representation of the perceived pitch span. The following three directional hypotheses were formulated to be tested in the present study:

(1) Judgments of friendliness correlate with different manipulated stimuli, whose F0 span have been re-synthesized under three conditions (narrow span, original span, and wide span);

(2) Assuming that (1) is confirmed by the data, this direct proportionality relation is found in scores obtained from the stimuli produced by both the American and the Italian NS;

(3) The evaluations given by male and female listeners are different across the three pitch span conditions, because males and females are expected to differ in their mental representations of pitch range.

5.4.3 Subjects

Twenty-two adult native speakers of Italian (11 males and 11 females) participated in the experiment. Tab. 43 shows the personal information about the subjects: their native language, age, city, area, class, major, year at the university and gender.

Subject	NS	Age	City	Area	Class	Major	Year	Gender
1	it	20	Padova	Padova	English	Communication	2	M
2	it	20	Belluno	Belluno	English	Communication	2	F
3	it	21	Padova	Padova	English	Communication	2	M
4	it	21	Cittadella	Padova	English	Communication	2	F
5	it	20	Soave	Verona	English	Communication	2	F
6	it	20	Negrar	Verona	English	Communication	2	F
7	it	20	Padova	Padova	English	Communication	2	M
8	it	20	Dolo	Venezia	English	Communication	2	M
9	it	20	Mestre	Venezia	English	Communication	2	M
10	it	20	Cittadella	Padova	English	Communication	2	M
11	it	20	Cittadella	Padova	English	Communication	2	F
12	it	20	Rovigo	Rovigo	English	Communication	2	F
13	it	21	Marostica	Vicenza	English	Communication	2	M
14	it	23	Monselice	Padova	English	Communication	2	F
15	it	20	Mira	Venezia	English	Communication	2	F
16	it	21	Asolo	Treviso	English	Communication	2	M
17	it	25	Padova	Padova	English	Communication	2	F
18	it	19	Soave	Verona	English	Communication	2	F
19	it	20	Mirano	Venezia	English	Communication	2	M
20	it	20	Padova	Padova	English	Communication	2	F
21	it	20	Padova	Padova	English	Communication	2	M
22	it	23	Padova	Padova	English	Communication	2	M

Table 43. Personal data of the 22 Italian participants in the fourth experiment. Information is about their native language, age, city, area, class, major, year, and gender.

All the subjects were native speakers of Italian and they were students in Communication Studies at the University of Padova. All of them were in their second year and came from the Veneto area. Their age ranged from 19 to 25 (mean age: 20.6). None of the speakers reported any speech, hearing or communication disorder and all of them were non-smokers. There was no screening for formal training in music or singing.

All the participants were undergraduate students attending an English class. They were contacted on line and were asked to participate into a perception experiment. Subjects accessed the test online by logging into their First Class account (http://fc.cla.unipd.it/Login). They gave their consent for the treatment of their personal data and volunteered for the experiment without receiving any monetary compensation.

5.4.4 Materials

This study compared native and non-native productions of 3 sentences selected from the materials analyzed in chapter 4 (see § 4.4). In the data set, the stimuli presented to the listeners were three English YNQ with similar F0 level and span, produced by an American and an Italian NS. The thee tested sentences were: ‘Do you need any money?’, ‘Do you wanna come for dinner?’, and ‘Can you open the door?’. A total of 36 stimuli (3 different sentences x 3 pitch span conditions x 2 native languages x 2 repetitions) were rated by 22 Italian listeners.

When selecting the stimuli for the perception test, I decided to consider only YNQ for the analysis of pitch span under three different conditions (narrow vs. original vs. wide span). This choice is functional to the theory-controlled manipulation of specific intonational characteristic of YNQ. In fact, when manipulating F0 range, it is advisable to modify only a limited tonal portion of a sentence. If I had to artificially modify all the pitch span along the intonation contour, the resulting sentence would dramatically change its lexical and grammatical meaning: a questions could be transformed into a statement and vice versa. In order not to distort the intonation pattern of a sentence, I deliberately decided to manipulate only the last part of the YNQ: the final rise.

Typically, YNQ ends in a final rise. The focal point of pitch variation in YNQ is placed on the last part of the utterance, which, depending on the intonation contour used, usually has a high F0 max and F0 span. In fact, the biggest pitch excursion in a question is generally registered in the final part of the utterance. That is why the last final rise is particularly suitable to be manipulated. Moreover, the YNQ produced by the Italians and the Americans, analyzed in the second experiment (see § 4.4), appeared to have fairly similar mean F0 level and span, with the utterances produced by the Americans having a wider F0 span.

Fig. 47 and 48 show the F0 level and span calculated across YNQ produced by the American and the Italian NSs and analyzed in the second experiment.

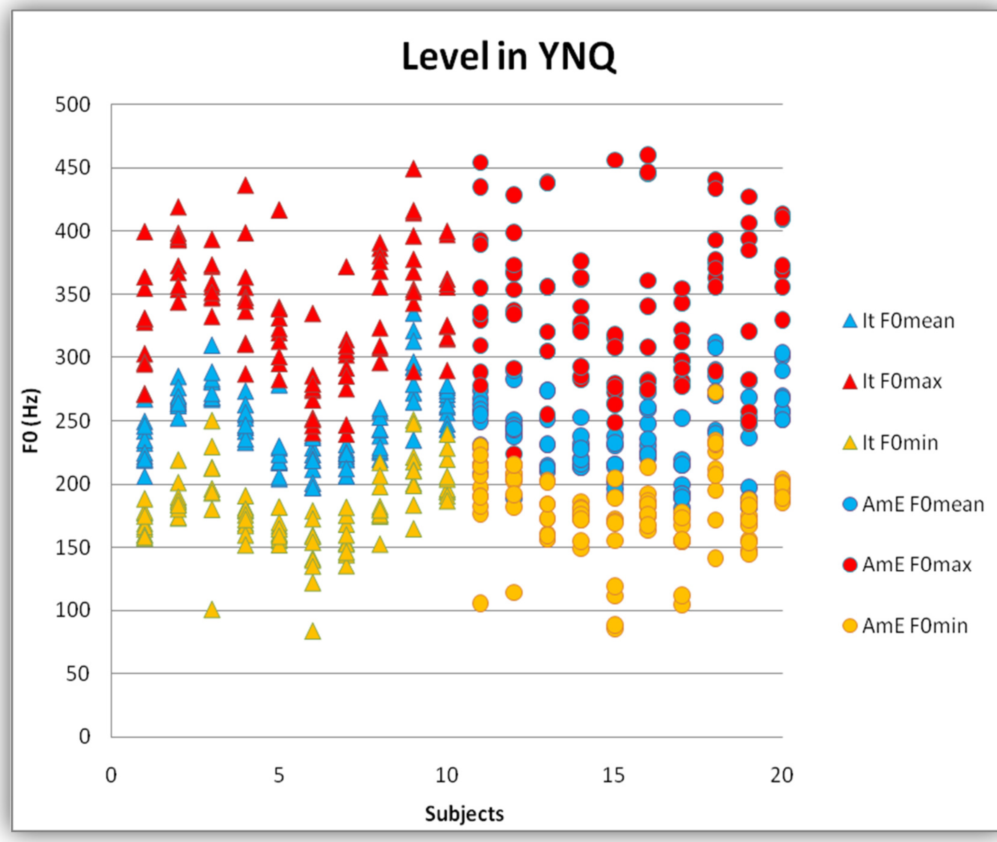


Figure 47. F0 levels (F0 max, F0 mean, and F0 min) across the YNQ produced by American and Italian native speaker.

Fig. 47 shows the dispersion of F0 mean (in blue), F0 max (in red), and F0 min (in yellow) across YNQ produced by the Americans (indicated by a circle) and the Italians (indicated by a triangle).

The mean F0 mean values calculated for the tested English stimuli were: 248 Hz for the Italians and 238 Hz for the Americans, with a difference of about 10 Hz. The mean F0 max was 337 Hz in the utterances produced by the Italians and 340 Hz in the utterances by the Americans. The mean F0 min was 178 Hz for the Italians and 179 Hz for the Americans. Thus, the YNQ produced by the Americans and the Italians showed very similar mean F0 level values, that were reported to be not significantly different from a statistical point of view (see § 4.4 for the statistical data).

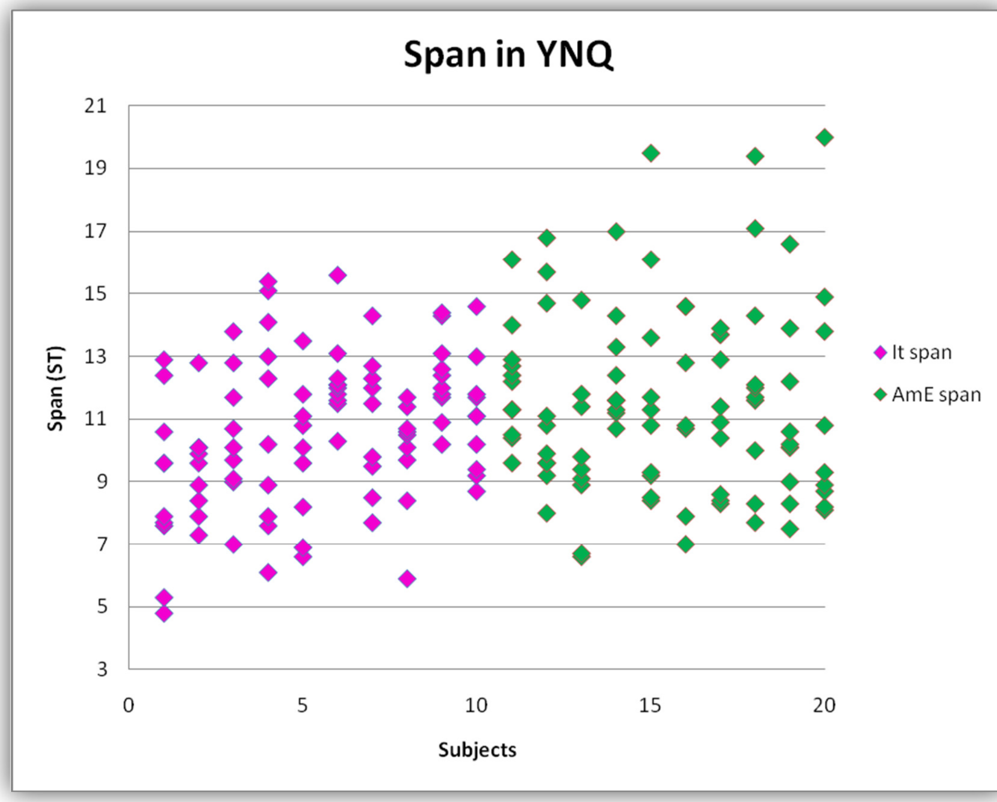


Figure 48. F0 span across YNQ produced by American and Italian native speakers.

Fig. 48 shows the dispersion of F0 span across YNQ produced by the Americans (in green) and the Italians (in purple). YNQ reached fairly similar mean pitch span values, with span in the American YNQ (11.19 ST) slightly larger than in the Italian YNQ (11.03 ST). The F0 span in the YNQ stimuli ranged from 7.3 ST to 10.1 ST for the Italians and from 10.8 ST to 21.4 ST for the Americans. No significant difference was found for pitch span in the YNQ produced by the Americans and the Italians ($t(99)= 0.36$ at the .05 level, $p=0.714$).

These similarities of F0 level and span across YNQ produced by the American and the Italian NS were functional to test the hypotheses (formulated in § 5.4.1). The original stimuli produced with a natural span variation were manipulated and re-synthesized in order to obtain sentences with a narrow, a mid and a wide span condition.

5.4.5 Procedure

Using a F0 generation re-synthesis method based on PSOLA (pitch-synchronous overlap-and-add algorithm) in Praat, manipulated versions of the original unmodified utterances (original span stimulus) were created. Even though the PSOLA algorithm alters F0, other aspects of the voice, such as speech rate and formant frequencies, remain unaffected (Feinberg et al., 2005). The manipulated sentences differed among each other only in one particular intonational aspect: their final pitch span. The re-synthesis values were automatically calculated in Praat by setting the minimum (for the narrow pitch span condition) and the maximum (for the wide pitch span condition) pitch span for the stimuli.

The three YNQ selected for the experiment were:

- *Do you need any money?*
- *Do you wanna come for dinner?*
- *Can you open the door?*

Since (narrow vs. original vs. wide) pitch span was the dependent variable to be tested, the stimuli were manipulated in their gradual F0 variation. In order to compare linear excursions of F0 span across sentences uttered by the Americans and the Italians, manipulation versions of several sentences were created.

The following figures (fig. 49 and 50) show the waveform and pitch lines of stimuli re-synthesized with a original, wide and narrow pitch span in the last movement of the YNQ (final rise). Fig. 49 compares stimuli produced by two American NSs (speaker 1 and 2) while fig. 50 compares stimuli produced by two Italian NS (speaker 3 and 4). Every sentence is presented under three pitch span conditions: narrow, original and wide pitch span.

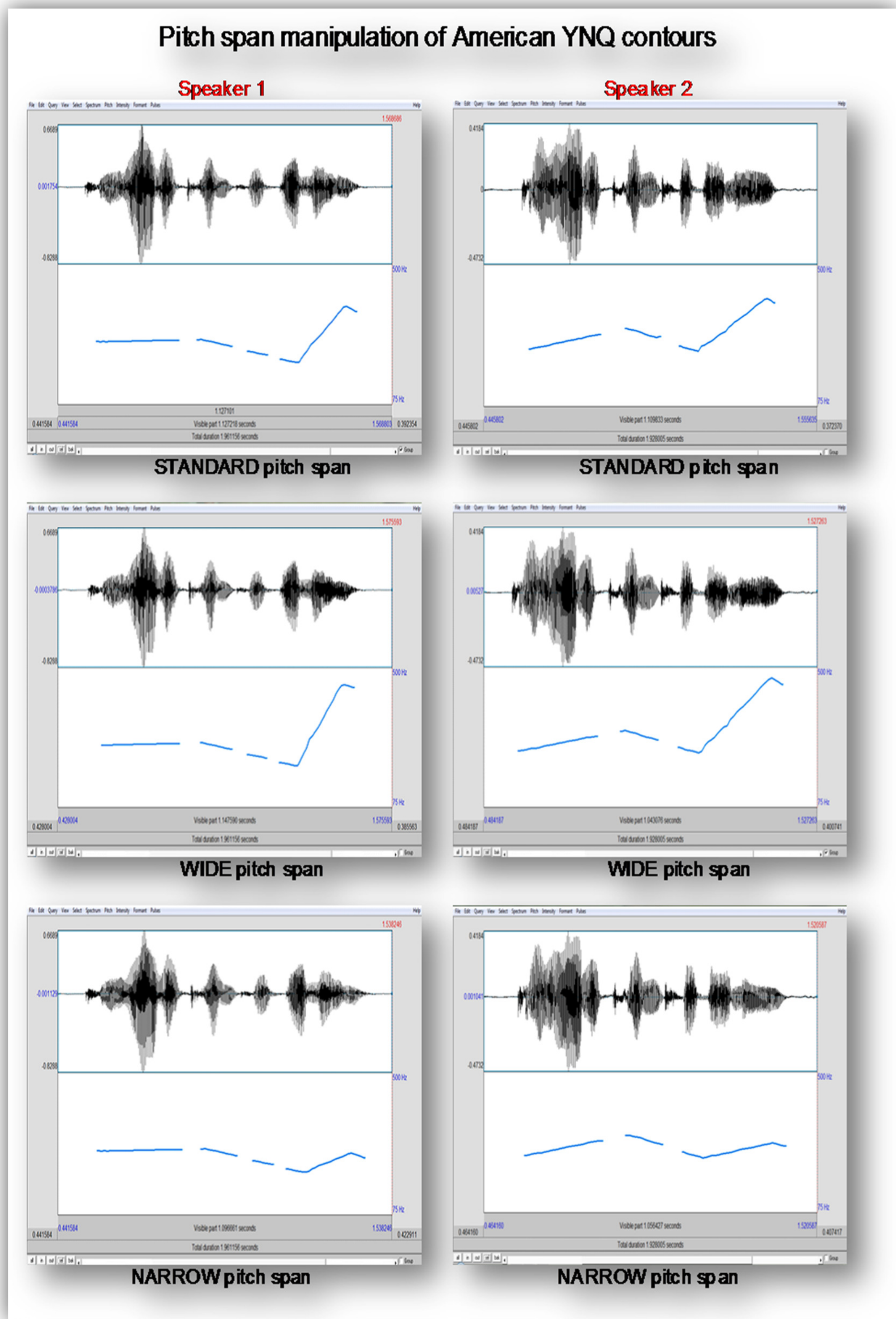


Figure 49. Manipulated versions: original (standard), wide and narrow pitch span of the YNQ 'Do you need money?' produced by two American NSs (speaker 1 and 2).

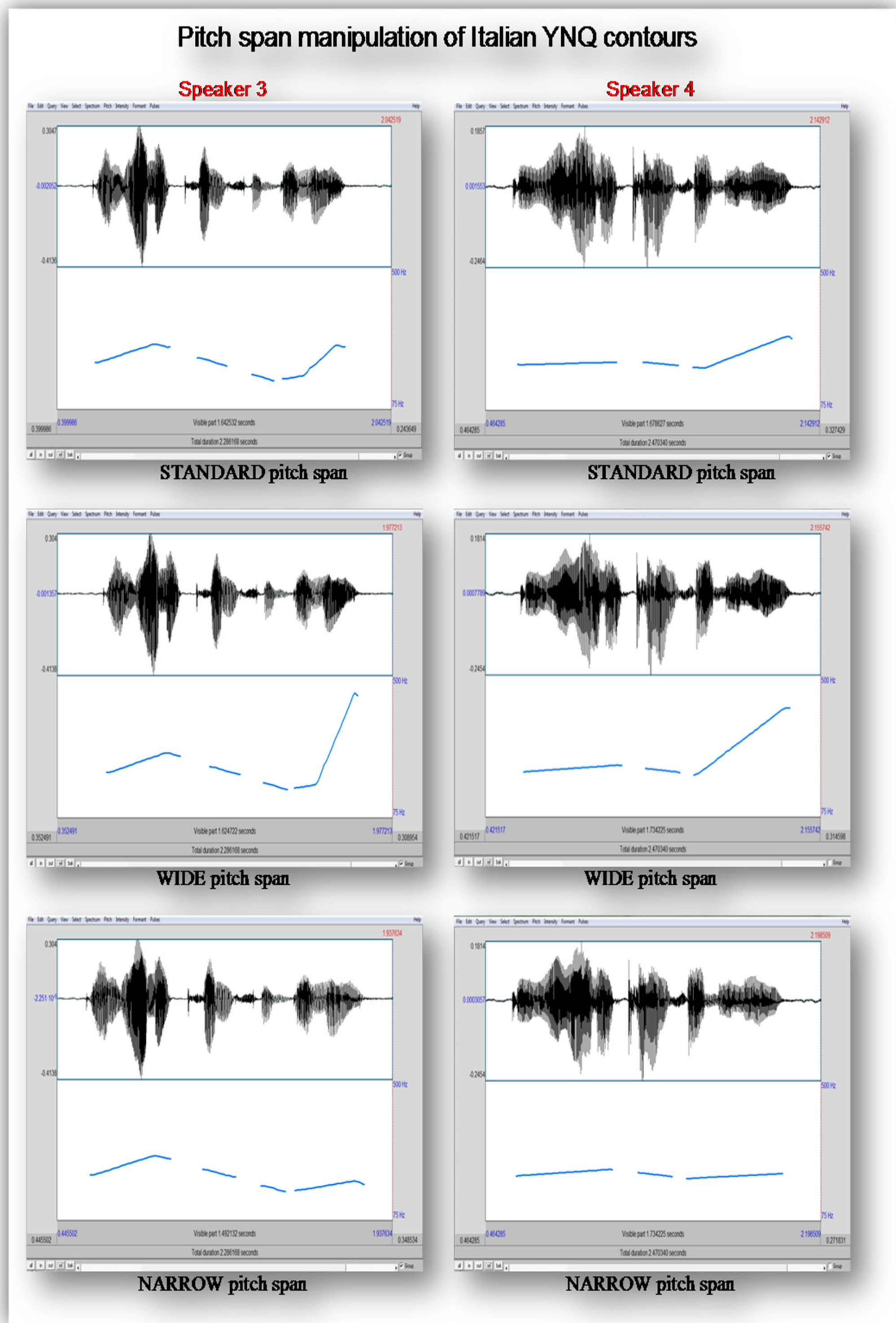


Figure 50. Manipulated versions: original (standard), wide and narrow pitch span of the YNQ ‘Do you need money?’ produced by two Italian NSs (speaker 3 and 4).

As shown in fig. 49 and 50, original pitch span stimuli (that are the original sentences produced by the American and Italian NSs) were manipulated to gradually raise the final pitch rise of about 150 Hz, in the wide span condition, and to gradually lower the final pitch rise of about 150 Hz, in the narrow span condition. Exact values of descrease/increase of pitch span were automatically generated by the PSOLA-based re-synthesis in Praat.

Once the manipulated versions of YNQ were created, 18 stimuli were selected for the experiment: 3 YNQ in three pitch span conditions produced by an American and an Italian NS. Fig. 51 shows how F0 level was manipulated in YNQ in order to create a continuum across the stimuli. For the American subject, mean F0 ranged from 173.6 Hz (lowest F0 mean in the narrow condition) to 226.6 Hz (highest F0 mean in the wide span condition). For the Italian subject, mean F0 ranged from 196.6 Hz (lowest F0 mean in the narrow condition) to 247.8 Hz (highest F0 mean in the wide span condition).

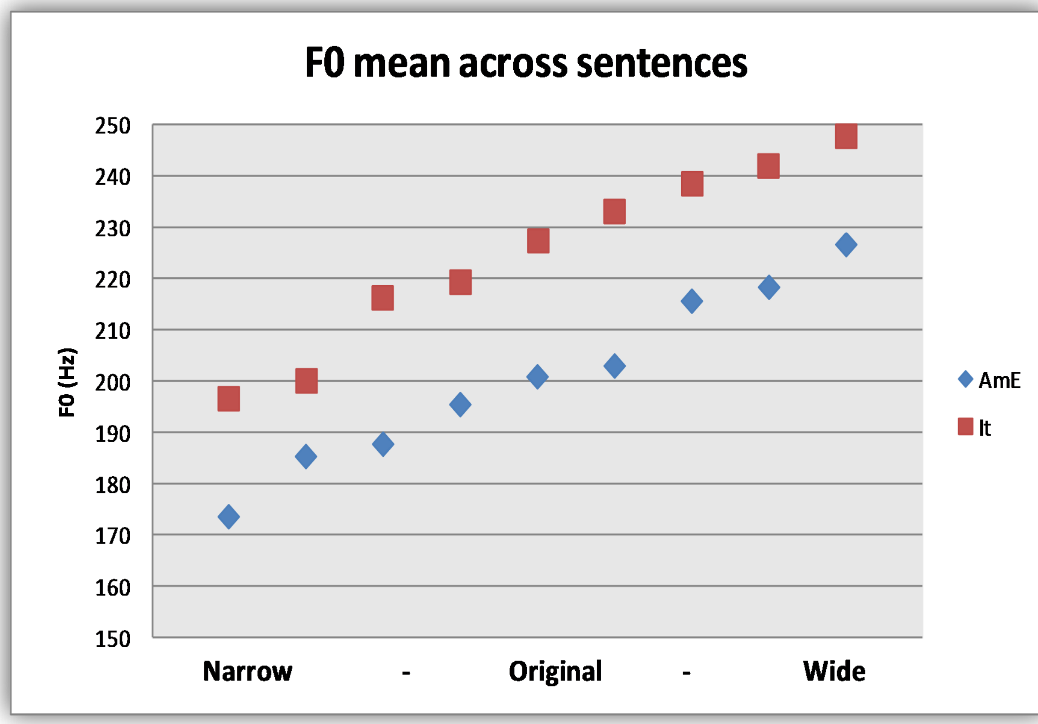


Figure 51. F0 level across YNQ produced by an American (blue diamond) and an Italian (red square) native-speaker, in three pitch span conditions (narrow vs. original vs. wide pitch span).

The F0 level calculated for the narrow span condition was 182.2 Hz for the American NS and 204.2 Hz for the Italian NS. The F0 level calculated for the original span condition was 199.8 Hz for the American NS and 226.5 Hz for the Italian NS. Finally, the F0 level for the wide span condition was 220.1 Hz for the American NS and 242.7 Hz for the Italian NS.

In sum, the F0 mean values of the American and Italian NS gradually increased from the narrow span condition to the wide span condition. In line with the results obtained in the second experiment (see § 4.4), the overall F0 level of the stimuli produced by the Italian NS was always higher than that of the American NS. If one interpolates a line between each pitch target shown in fig. 51, it is evident that the mean F0 values obtained from the American NS (see the blue diamonds in the graph) are parallel to those obtained from the Italian NS (see the red squares in the graph).

Fig. 52 shows how F0 span was manipulated in YNQ in order to create a continuum across the stimuli.

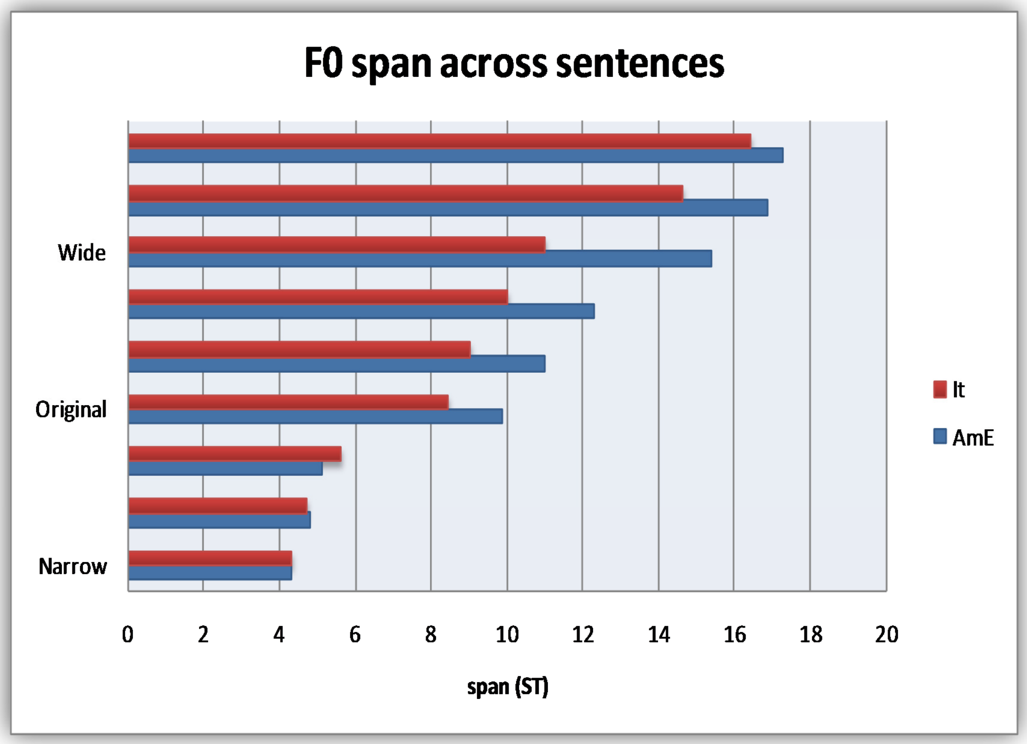


Figure 52. F0 span across YNQ produced by an American and an Italian native-speaker, in three manipulated versions (narrow vs. original vs. wide pitch span).

For the American subject, F0 span ranged from 4.3 ST (lowest F0 span in the narrow condition) to 17.3 ST (highest F0 span in the wide span condition). For the Italian subject, F0 span ranged from 4.3 ST (lowest F0 span in the narrow condition) to 16.4 ST (highest F0 span in the wide span condition). The F0 span calculated for the narrow span condition was 4.7 ST for the American NS and 4.8 ST for the Italian NS. The F0 span calculated for the original span condition was 11 ST for the American NS and 9 ST for the Italian NS. Finally, the F0 span calculated for the wide span condition was 16.5 ST for the American NS and 14 ST for the Italian NS. Both the F0 span values of the American and Italian NS gradually increased from the narrow span condition to the wide span condition. In line with the results obtained in the second experiment (see § 4.4), the overall F0 span of the stimuli produced by the Italian NS (see red bars in the graph) was narrower than that of the American NS (see the blue bars in the graph).

The stimuli were submitted to a selected group of subjects (university students at the University of Padova) who were asked to rate the stimuli on the basis of how friendly they sounded. Twenty-two native speakers of Italian taking an English class participated in the experiment. The perception test took place in a laboratory equipped with personal computers for all students. The stimuli were presented to the subjects over headphones. The subjects were instructed to pay attention to the intonation of the stimuli and decide which interpretation was more likely for each stimulus by rating the friendliness of the stimuli on a 5-point scale. They were asked to ignore the quality of the voice and to pay attention to how the sentences were said. Listeners had to imagine themselves as the addressees and indicate what kind of impression (friendly vs. not friendly) the speaker made on them. They had to put a cross on one of the five boxes allotted for each stimulus. The more to the left they placed the cross, the more they interpreted the sentence as ‘not friendly’; the more to the right they placed the cross, the more they interpreted the sentence as ‘friendly’.

The 36 stimuli were presented in an answer sheet and divided into two blocks, with 18 stimuli in each block. A training session was conducted prior to the experiment, to get subjects used to the stimuli and the task. Every stimulus in the continuum was presented to the subjects in a randomized order, with a 3 second silent pause in between to let subjects write down their score. Subjects accessed the test online by logging into their First Class account. Students were not allowed to open and access the answer sheets submitted by their colleagues, thus they could not look at each other’s sheets. They could not stop the audio files or listen to the same stimulus more than once. I supervised the administration of the test, making sure students handed in their answers on time. The subjects were allowed 30 minutes to complete their answer sheets[13]. At the end of the last block, the students were not allowed any extra time and they had to send their answer sheets to the main conference right away.

After collecting all the answer sheets, I assigned 1 point to stimuli judged as 'not friendly' and 5 points to stimuli judged as 'friendly', with other scores distributed along the scale. In the end I transferred all the data onto an excel files in which I compared the answers to the manipulated stimuli by grouping scores together depending on the three pitch span conditions: narrow, original, and wide pitch span.

5.4.6 Results

The results show a correlation between wide span in the sentences and judgments of friendliness: the stimuli re-synthesized with a wide pitch span were perceived as more friendly than the stimuli with a narrow pitch span. Not only did the pitch span in the manipulated sentences but also the native language of the speakers play a role in the perception of level of friendliness. In fact, in every pitch span condition (narrow vs. original vs. wide pitch span), the stimuli produced by the Italian NS were rated more poorly than those produced by the American NS. This means that both the 'pitch span' and the 'native language' factors contribute to make utterances produced by the American NS sound more friendly than the utterances produced by the Italian NS.

Fig. 53 shows the scores given by the listeners to the stimuli with manipulated pitch span. The stimuli were compared across the native language of the speakers (American vs. Italian) and across pitch span variation (narrow vs. original vs. wide pitch span). The total score of friendliness reached in the narrow span condition was 313 points for the YNQ produced by the Italian NS and 326 points for the YNQ produced by the American NS. The total score reached in the original span condition was 359 points for the YNQ produced by the Italian NS and 405 points for the YNQ produced by the American NS. Finally, the total score reached in the wide span condition was 461 points for the YNQ produced by the Italian NS and 478 points for the YNQ produced by the American NS.

[13] A sample spreadsheet of the questionnaire is in Appendix B.

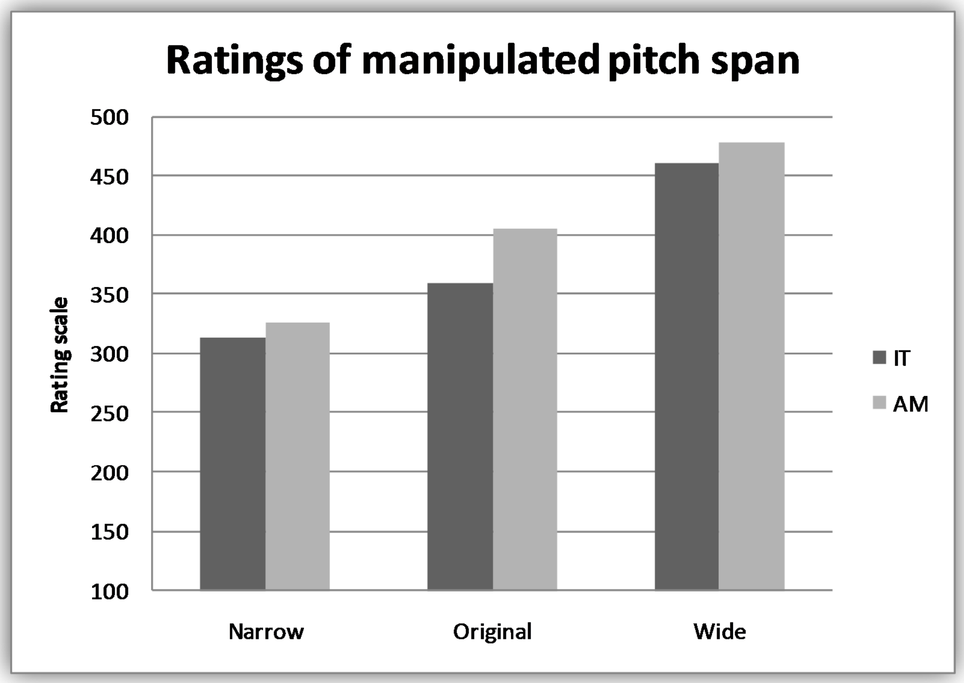


Figure 53. Ratings of stimuli with manipulated pitch span. The stimuli are compared across native the language of the speakers (American vs. Italian) and pitch span variations (narrow vs. original vs. wide pitch span).

Tab. 53 shows the rating scores given to English YNQ produced by an American and an Italian NS. The total values calculated for three different YNQ are the sum of scores produced by the listeners during the two repetitions of the test (rep. 1 and rep. 2). In the two repetitions, the listeners judged the same 18 stimuli that were presented in different randomized orders. On average, the stimuli produced by the Italian NS and the American NS received comparable scores: the scores increased as the pitch span increased. In addition, the judgments given by the listeners during the first and the second part of the experiment were fairly similar. This means that a different randomized order did not have an impact on the consistency of the final results.

NS		SENT	SPAN	Rep. 1	Rep. 2	TOT
		1	N	43	53	96
		2	N	61	57	118
		3	N	47	52	99
Italian	*Narrow span*					**313**
		1	S	54	47	101
		2	S	70	66	136
		3	S	63	59	122
Italian	*Original span*					**359**
		1	W	71	68	139
		2	W	89	77	166
		3	W	74	82	156
Italian	*Wide span*					**461**
		1	N	51	45	96
		2	N	59	58	117
		3	N	56	57	113
American	*Narrow span*					**326**
		1	S	64	54	118
		2	S	65	66	131
		3	S	77	79	156
American	*Original span*					**405**
		1	W	71	76	147
		2	W	85	84	169
		3	W	79	83	162
American	*Wide span*					**478**

Table 44. Ratings of YNQ with manipulated pitch span. The stimuli are compared across the native language of the speakers (American vs. Italian), pitch span variation (narrow vs. original vs. wide pitch span), the listening session (first and second repetition).

As the data in fig. 53 and in tab. 44 show, the stimuli were perceived as more or less interesting depending on two factors: 1) pitch span condition (narrow vs. original vs. wide pitch span) and 2) native language of the speaker. To calculate whether or not the different scores given to the stimuli were statistically significant, several t-tests were run for two factors: ‘native language of the speaker’ and ‘pitch span condition’.

A t-test compares exactly two datasets (of some dependent variable such as F0 span) that differ with respect to one factor (of an independent variable such as ‘pitch span condition’ and ‘native language of the speaker’). With little or no overlap between the samples, one can be confident that the two datasets are statistically different. Whether or not there is overlap depends on the difference between the means of the samples (bigger difference means less overlap) and the variability in the samples (less variability means less overlap). A paired t-test assuming equal variances between groups was selected as the preferred method to analyze the data in this experiment, mostly for two reasons: 1) it is one of the most commonly used tests in phonetics studies (Rasinger, 2008; Lane, 2012); 2) it suits the kind of variables investigated in this experiment. The F0 values were loaded into excel to obtain inferential statistics. For a more detailed description of data obtained from the 9 t-tests, the tables of the results of the t-tests are shown in Appendix A.

For the **factor *'native language of the speaker'*** (American English vs. Italian), two-sample paired t-tests were run across stimuli with different pitch span, calculated for the sentences produced by an American NS and an Italian NS. The statistical significance for the factor 'native language of the speaker' was measured in three t-tests. The dependent variable consisted of the scores given by the listeners (i.e. ratings) and the independent variables were the three conditions of pitch span (i.e. narrow vs. original vs. wide).

The difference in ratings between the narrow pitch span stimuli produced by an American and an Italian was not statistically significant ($t(65)= 1.11$ at the .05 level, $p=0.270$). By contrast, the difference in ratings between the original pitch span stimuli produced by an American and an Italian was statistically significant ($t(65)= 2.82$ at the .05 level, $p= 0.006$). The results of the t-test on the difference in ratings betweenthe wide pitch span stimuli produced by an American and an Italian indicate that the scores for friendliness did not differ significantly for the two speakers of the two languages ($t(65)= 0.86$ at the .05 level, $p= 0.392$).

For the **factor *'pitch span condition'*** (narrow vs. original vs. wide pitch span) two-tail paired t-tests were run across YNQ as produced by an American and an Italian NS. The statistical significance for the 'pitch span condition' factor was measured in a total of 6 t-tests with the ratings of the listeners as dependent variables and the native languages of the speakers (American vs. Italian) as independent variables.

Within the stimuli produced by the Italian speaker, the results of a t-test on the difference of ratings between narrow vs. original pitch span stimuli indicate that scores for friendliness significantly differed between the two span conditions ($t(65)= 2.47$ at the .05 level, $p= 0.015$). Also the difference in ratings between the narrow and wide pitch span stimuli was statistically significant ($t(65)= 2.39$ at the .05 level, $p= 0.019$). The difference in ratings between the original and wide pitch span stimuli was statistically significant, too (t(65)= 4.33 at the .05 level, p< 0.001).

Within the stimuli produced by the American speaker, the results of a t-test on the difference in ratings between narrow vs. original pitch span stimuli show that scores for friendliness significantly differed between the two span conditions (t(65)= 2.19 at the .05 level, p= 0.031). Also the difference in ratings between the narrow and wide pitch span stimuli was statistically significant (t(65)= 3.28 at the .05 level, p< 0.001). The difference in ratings between the original and wide pitch span stimuli was statistically significant, too (t(65)= 2.63 at the .05 level, p= 0.01).

In order to examine whether or not the male and the female listeners displayed similar trends in rating the YNQ produced in different pitch span conditions, the ratings produced by the males and the females were compared and tested separetely. Fig. 54 shows the scores given by the males vs. female listeners to the stimuli with manipulated pitch span. The ratings are compared across the native language of the speakers (American vs. Italian) and across pitch span variation (narrow vs. original vs. wide pitch span).

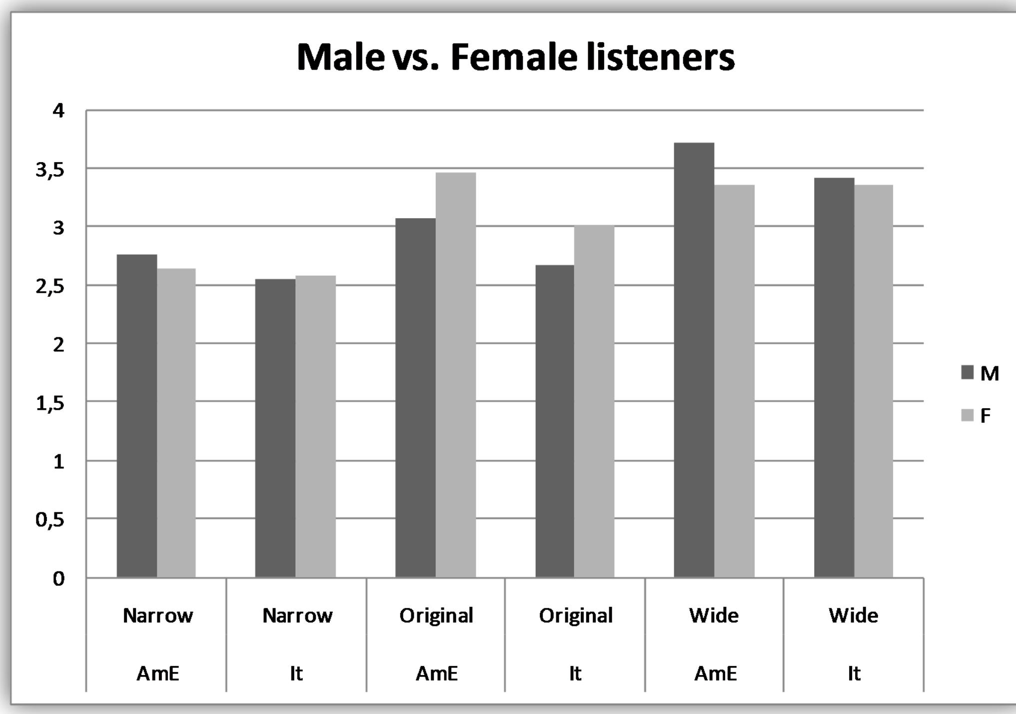


Figure 54. Ratings of males and females. The scores given by the male and the female listeners are compared across the native languages of the speakers (American vs. Italian) and the pitch span variation (narrow vs. original vs. wide pitch span).

Fig. 54 shows that that the YNQ produced by the American NS scored higher than the YNQ produced by the Italian NS in every pitch span condition. The ratings given to the narrow span stimuli, by both male and female listeners, were fairly similar (average 3.5 points). By in the original pitch span condition, the female listeners gave higher ratings than the males while, in the wide pitch span condition, the males gave higher ratings than the females.

These data seem to support the idea that the males preferred the wide span stimuli over the original ones, vice versa the females were more inclined to give relatively higher scores to the stimuli with an original pitch span rather than the ones with a wide span. However, this assumption is not supported the statistical analysis. A one- way ANOVA run for the factor 'gender' did not show any significant difference across ratings given by the male and the female listeners.

5.4.7 Discussion

The results of the experiment show that the American and the Italian speakers were judged as more friendly when the pitch span of their sentences was widened (wide span condition) and less friendly when the pitch span was narrowed (narrow span condition). This happened for all the stimuli, regardless of the native language of the speakers (American vs. Italian). This show that listeners are sensitive to pitch span variation and used pitch span as a determinant factor to distinguish between more or less friendly speakers.

This is in line with the idea that, even though L2 accent is a crucial factor in the formulations of judments on L1 and L2 speech, the variation of pitch span seems to be one of the most important cues in distinguishing among friendly and unfriendly questions. Thus, 'in spite of the differences between speakers with different language backgrounds, an important implication […] is that the perception of questions seems to be activated on an identical set of prosodic parameters across languages; these parameters all contribute to the perception of high pitch' (Chen, 2005: 47).

The present experiment showed that listeners with different language backgrounds accessed the intonational cue of pitch span variation when making judgments on stimuli manipulated for different pitch span conditions. Stimuli were re-synthesized from original samples produced by an American and an Italian NS, selected for their pitch range similarities. This was done to maximally activate listeners' skills to discriminate among different pitch span ranges and minimize the voice quality influence.

The first hypothesis tested whether or not judgments of friendliness correlated with different manipulated stimuli, whose F0 span had been re-synthesized under three conditions (narrow span, original span, and wide span). The results showed that the ratings of the stimuli increased as the pitch span increased. Thus, the first hypothesis was confirmed. In fact, wide-span manipulated stimuli were perceived as more friendly than narrow-span manipulated stimuli, with the Americans rated as more friendly than the Italians. Thus, the role of pitch span was determinant in the productions of ± friendly judgments. In sum, the pitch span variation linearly correlated with friendly judgments across all three types of manipulated stimuli (narrow vs. original vs. wide span). The relation between the pitch span and the ratings of the stimuli was found to be directly proportional.

The second hypothesis, investigating whether or not the correlation between pitch span and judgments of friendliness is a universal factor, was also confirmed. In fact, no matter their native language, both the L1 and the L2 speakers were perceived as more friendly in the wide pitch span condition and less friendly in the narrow pitch span condition. The listeners expressed judgments by fully relying on pitch span variation rather than on the native language of the speakers. Thus, the results show that the relative friendliness of the speakers was rated on the basis of pitch range variation rather than the L1/L2 influence. In fact, listeners preferred wide span over narrow span stimuli, regardless of the native language of the speakers. This is confirmed by the fact that differences reported for the factor 'pitch span variation' across the stimuli reached statistical significance while the differences reported for the factor 'native language of the speakers' did not (see the results of the t-tests).

Finally, the third hypothesis, testing the evaluations given by male and female listeners across the three pitch span conditions, was not validated by the results. Contrary to my expectations, this hypothesis was not confirmed, as the males and the females listeners did not significantly differ in their judgments of pitch range. Since males and females create in their mind different representations of pitch range (Dolson, 1994; Deutsch 1992, 2009; van Beezoijen, 1995; Feinberg et al., 2008; Yuasa, 2008) they were expected to formulate different quality judgments of pitch range. In a study by Feinberg et al. (2008), it was found that both male and female listeners perceive wide span variation in a speaker to be more attractive than narrow span variation. However, ‘men prefer high voice pitch to average voice pitch in women’s voices’ (Feinberg et al., 2008: 615) while women seem to place a limit to how high a voice can go and still be considered attractive.

In line with this finding, the results of my study show that the stimuli with a wide pitch span condition were considered more friendly by the male listeners and less friendly by the female listeners. This effect could be determined by the fact that, according to Feinberg et al. (2008), female listeners may perceive female voices with an extremely raised pitch as not attractive. ‘Overall, however, it seems that as long as a voice is within the normal range of pitches, the pattern is for higher-pitched women’s voices to be perceived as more attractive’ (Munger, 2008). The differences in ratings given by the male and the female listeners replicated the finding by Feinberg and colleagues. In fact, the stimuli with an original pitch span condition received higher scores by the female than the male listeners, vice versa the stimuli with a wide pitch span condition received higher scores by the male than the female listeners. However, in the present experiment, the differences in ratings across the male and the female listeners were not above chance level and thus, they were not statistically significant.

5.5 Summary

Chapter 5 discusses the incidence of the results obtained from the production studies from a perceptual perspective. The acoustic differences of data examined with LTD and linguistic measures are further tested by the judgments of a group of American and Italians listeners. Two perception studies are designed to measure and interpret the perception of cross-linguistic differences in pitch range. The general hypothesis at the basis of these experiments is that both the American and the Italian listeners may display a similar understanding and evaluation of pitch range variation. The experimental data in the production studies indicate that the mode of sentences is better captured by F0 span than level. In particular, pitch span more than level is found to be a cue for non-nativeness.

The third comparative study examines the correlation between pitch span variation and the perception of different sentence types, yes-no questions, statements, and exclamations by American and Italian listeners. The results show that the stimuli with larger pitch span are perceived as more interesting, exciting, credible and friendly than the stimuli with narrower pitch span. As previously mentioned, the effect of pitch span variation can be described as both universal and language-specific. The universal component of pitch perception is given by the correlation between pitch span variation and the perception of stimuli rated by the listeners with different language backgrounds (American vs. Italian listeners). The language-specific component of pitch variation is determined by the different evaluations given to the tested native languages (Italian vs. American speakers).

The fourth comparative study verifies whether or not a linear correlation between positive/negative judgements and wide/narrow span is activated in the speech produced by L1 and L2 speakers. To accomplish this, original stimuli uttered by native and non-native speakers of English were manipulated and re-synthesised to create three pitch span conditions: narrow span, original span, and wide span. Both the American and the Italian speakers were judged to be more friendly when the pitch span of their sentences was widened (wide span condition) and less friendly when the pitch span was narrowed (narrow span condition). Thus, pitch span variation was found to be a rubust cue for the perception of friendly/unfriendly questions. Regardless of their native languages, speakers uttering sentences with wide pitch span were perceived as more friendly than speakers uttering sentences with narrow pitch span. This further clarifies the crucial role of pitch range in conveying linguistic and paralinguistic meaning by raising positive or negative judgments from the listeners.

Chapter 6:
Conclusion

6.1 Summary of the findings

The experimental work carried out in this investigation can be divided into two main areas: one that is concerned with the acoustic differences of pitch range (level and span) across American and Italian speakers (chapter 4), and another with the perception of the pitch span differences reported for American and Italian productions (chapter 5). Pitch span is found to systematically differ across speakers with different native languages (Americans and Italians), thus it can be considered as a critical factor to distinguish between speech produced by native and non-native speakers.

Pitch range is at the same time a universal characteristic of any language and a language-specific factor that differs across speakers who have different language backgrounds. As a consequence, a language may possess recurring pitch range patterns that do not apply to other languages. What is more, pitch range has a socio-cultural impact because it may be variously interpreted and perceived, based on the different expectations of a community of speakers. In this work, I investigated not only the differences in pitch range across speakers (Americans vs. Italians) but also speakers' expectations about positively or negatively connoted pitch patterns.

6.2 Production studies

Two experiments were carried out on speech material produced by some Californian American and Veneto Italian native speakers to analyze the patterns of pitch range in English as L1 and L2. The investigation of the language-specific use of F0 prompted a number of speculations and hypotheses related to the characteristics of pitch range across languages, speakers, genders, and sentence types.

6.2.1 First experiment

In the first experiment, it was observed that the pitch level and span displayed a considerable dynamism across English as L1 and L2, and Italian as L1. Pitch level appeared to have fairly similar values across groups, with the Italian male speakers displaying slightly higher mean values for pitch level than the American males. On the contrary, the female speakers showed different results for pitch level. The pitch realizations of the American females were spread all along the acoustic space, reaching a mean F0max level equal to 340 Hz, while the pitch realizations of the Italian females were shifted upwards, with a mean F0max level equal to 349 Hz. In addition, the mean F0min level of the female subjects was about 42 Hz higher in the Italians (with a mean F0min level of 144 Hz), as compared to the Americans (i.e. with a mean F0min level of 102 Hz). As for span, the Americans (both the males and the females) used wider pitch span than the Italians. What is more, the Italian females better than the males managed to reproduce the Americans' pattern by widening their pitch span when speaking English.

These results show that pitch range varies considerably across languages. In fact, 'languages are not equally inclined to exploit the phonetic space of prosodic variables that only indirectly affect variations in pitch range' (Chen, 2001: 88). American English and Italian significantly differ in the pitch range adopted by their L1 and L2 speakers. The Italian subjects of my experiment appeared to project their standard L1 pitch range onto their L2. What is more, the different phonological and phonetic conventions displayed in languages such as American English and Italian have an influence on the modulation of pitch range trends, that are perceived and interpreted depending on the socio-cultural expectations of the linguistic community. For example, the American speakers seem speak with lower F0min levels than the Italians, thus their pitch space looks as shifted downwards. As far as pitch span is concerned, the Americans' speech is characterized by largely wider F0 span than the Italians' speech.

Gender also plays a significant role in F0 variation. While utterances produced by the Italian male subjects have similar pitch range in Italian and English, the pitch range of utterances produced by the Italian female subjects varies considerably across languages. These differences may reflect the fact that women are more sensitive than men to perceive and modulate their pitch range. As a result, Italian females seem to better approach the model of the American native speakers, by adapting the pitch range of their English L2 pitch range to the English L1 trends. Sociolinguistic factors may have an impact on the motivation and the intent to replicate the native speakers' model. This may lead some speakers (in this case, the female subjects) to perform better than others (in this case, the male subjects) in the L2.

6.2.2 Second experiment

In the second experiment, F0 variation was tested in the analysis of the pitch contours of different sentence types: yes-no questions (YNQ), wh-questions (WHQ) and statements (STM). The main aim of the second experiment was to determine how pitch level and span differed in the intonation patterns of English sentences produced by American and Italian females. The results show that the Italian subjects produce all sentence types with a narrower pitch span and a higher pitch level than the Americans. This means that the English as L2 is more high-pitched than English as L1. This finding is in line with data from other languages, such as Swedish, standard Chinese, Japanese, and Hungarian (Bolinger, 1978; Ohala, 1983; van Beezoijen, 1995; Gussenhoven, 2002; Yuasa, 2008). A considerable drop in pitch span was observed for the sentences produced by the Italians speaking English. This is an effect of two main factors: 1) the average Italian pitch span is narrower than the average American English pitch span and 2) the English L2 pitch span is narrower than the English L1 pitch span.

Important differences were registered across sentence types. YNQ were uttered with similar pitch level and span by both the Americans and Italians. High pitch level and wide pitch span are common traits in YNQ across languages. Consequently, the Italian learners of English may either be familiar with the English final rising patterns or may simply transfer patterns from their L1 to their L2. The Italian subjects of this study, whether subconsciously or consciously, successfully managed to approach the American model for YNQ pitch contours. By contrast, the American and Italian subjects differed in their production of WHQ and STM pitch contours. The data showed that the F0 level for WHQ and STM was shifted upwards in the acoustic space of the utterances produced by the Italians, as compared to those produced by the Americans. Also the data calculated for pitch span evidenced neat differences across WHQ and STM produced by the Americans and the Italians. Generally, the dependent variables measured for pitch level (F0mean, F0median, F0max, F0min) were significantly higher in the utterances produced by the

Italians. On the other hand, the dependent variable of pitch span (STrange) was significantly wider in the utterances produced by the Americans.

The results of the first and the second experiments lead to the conclusion that F0 span has a significant role in distinguishing pitch patterns displayed in English as L1 and L2, and Italian as L1. The pitch span of the Italians speaking English is considerably narrower than that of the American English native speakers. What is more, the experimental data gathered across sentence types indicates that the mode of sentences is better captured by the measures of F0 span than level. In conclusion, the Italian speakers of English used a narrower span, compared to the native norm. This is in line with previous findings on data from other languages. In several studies (Backman, 1979; Jenner, 1976; Willems 1982; Ladd, 1996; Munro, 1995; Mennen, 2006; Mennen et al., 2012), it was found that L2 English is characterized by a narrower pitch span than L1 English. Hence, a narrow pitch span can be considered as an indicator of non-nativeness.

6.3 Perception Studies

Two experiments were carried out on speech materials (produced by American and Italian native speakers) that were rated by two groups of listeners (American and Italian listeners). These studies were designed to measure the correlation between pitch span variation and positive/negative perception of utterances. Since the production studies showed that the American native speakers use a wider F0 span than the Italian learners of English, I was interested in determining the effect that a small or large pitch range variation has on the listeners.

6.3.1 Third experiment

In the third experiment, a group of American and Italian native speakers rated English stimuli (produced by Italians and Americans), presented in a randomized order. The stimuli were rated on three separate parameters: ± interesting, ± exciting, and ± credible. The correlation between pitch span and rating scores was measured in different sentence types (yes-no questions, statements, and exclamations). The two groups of listeners (the Americans and the Italians) were found to base their judgments of the stimuli on the perceived pitch span.

The results show that both the Italian and the American listeners perceive the wide-span stimuli as being more interesting, exciting, and credible than the narrow-span stimuli. In fact, a systematic correlation was found between pitch span and rating score:. the results show that the wider the pitch span the more positive judgments are given to the stimuli. The relation between pitch span and the ratings of the stimuli was found to be directly proportional, across the tested parameters (± interesting, ± exciting, and ± credible). Again, this confirms the idea that wide pitch span is perceived as more lively and attractive than narrow pitch span. Since the Italian subjects in this study have narrower pitch span than the Americans, they are perceived as less interesting, exciting, and credible than the latter.

Pitch span can be considered as a language-specific and speaker-specific factor of prosodic variation. The language-specific component is given by standard pitch range values measured across different languages. Generally, languages such as German and Dutch are considered narrow-span languages as compared to English. For example, 'speakers of the narrow-range language Dutch associate a larger meaning difference with a given interval of pitch variation than speakers of the wide-range language British English' (Chen, 2005: 103). The speaker-specific component is given by the individuals' personal interpretation of pitch range. Regardless of their native languages, different speakers can display similar pitch span and be more or less sensitive to pitch variation, depending on the way they project a given pitch span within their relative habitual pitch space.

Regardless of the habitual pitch span observed in their L1, both the Americans and the Italians perceived wide span more positively than narrow span. This finding has important implications in the domain of L1 and L2 acquisition and is in contrast with the assumption (Chen, 2005) that a given pitch span may be positively connoted in a wide-span language (such as English) and negatively connoted in a narrow-span language (such as Italian) or vice versa. It seems that speakers using a wide span range are universally perceived more positively than those who adopt a narrow span range, regardless of whether they are speaking a language characterized by a narrow or a wide span.

Whatever their language background, 'speakers are capable of interpreting phonetic implementation in the speech of the others (in known or unknown languages) in accordance with the universal paralinguistic meanings of pitch variation' (Chen, 2005: 43). In my study, the universal paralinguistic meaning of pitch variation was interpreted by pitch span which is not only a phonetic cue to L2 speech perception but also an indicator of different levels of credibility and friendliness.

6.3.2 Fourth experiment

The fourth experiment was designed to test whether the correlation between pitch span and the evaluations of intonation contours given by the listeners conform to the following rule: the larger the span, the higher the scores. In order to isolate the single component of pitch span variation from other factors of differences across sentences (i.e. general rhythm, duration, sentence mode, voice quality, etc.), yes-no questions (YNQ) with similar pitch characteristics (similar F0 level and span) were selected. Stimuli were re-synthesized from original samples produced by an American and an Italian native speaker, selected for their pitch range similarities, to minimize inter-speaker differences. Pitch span was gradually increased or decreased at the end of every sentence in order to generate a set of stimuli in which pitch span variation was the main factor of difference across the different sentences. As a result, three pitch span conditions were created: narrow span, original span, and wide span. Male and female Italian listeners were asked to rate how friendly/unfriendly the sentences sounded to them.

In line with the results obtained in the third experiment, the results of the fourth experiment confirmed the rule by which the rating scores given by the listeners are directly proportional to F0 span. The stimuli re-synthesized with an original pitch span condition were perceived on average as moderately friendly, with the Americans rated as more friendly than the Italians. Friendliness ratings gradually increased from the sentences with a narrow span to those with a wide span. This suggests that the data followed the general principle: the larger the span, the higher the scores, the more friendly the stimulus was interpreted.

Both the American and the Italian speakers were judged more friendly when the pitch span of their sentences was enlarged (wide span condition) and less friendly when the pitch span was narrowed (narrow span condition). Consequently, listeners preferred wide span over narrow span stimuli, regardless of the native language of the speakers. The listeners did not significantly differ in their judgments of pitch range: both the male and the female listeners perceived wide span as more friendly than narrow span.

6.4 General Discussion

This investigation analyzed the characteristics of pitch range production and perception in English sentences uttered by Americans and Italians. The results of this study support the following generalizations.

First, pitch span more than level is a cue for non-nativeness because generally L2 English is characterized by a narrower pitch span than L1 English. What is more, the experimental data in the production studies indicate that the mode of sentences (YNQ vs. WHQ vs. STM) is better captured by F0 span than level.

Second, Italian learners of English transfer their L1 pitch range variation into their L2. The Italians use overall higher pitch levels when speaking Italian and lower levels when speaking English. The English sentences produced by the Italians have overall higher pitch levels and narrower pitch span than those produced by the Americans. Conversely, the Italians' pitch spans are overall higher in English and lower in Italian. When comparing productions in English, Italian females use higher F0 levels than the American females; vice versa, Italian males show slightly lower F0 levels than the American males.

Third, wide pitch span correlates systematically with listeners' positive interpretations, as wide span is associated with being interesting, exciting and credible. Listeners with different language backgrounds seem to access the intonational cue of pitch span variation when making judgements on stimuli manipulated for different pitch span conditions. Thus, both the American and the Italian speakers are likely to be judged as being more friendly when they use a wide pitch span than when they use a narrow pitch span, regardless of the native language of the speakers (American vs. Italian).

6.5 Future Directions

My investigation examines the pitch range differences in the production and perception of sentences in American English and Italian. Throughout my work I tried to understand the complex mechanisms by means of which pitch range varies across speakers, sentences, and languages. In carrying out this research, one of the most daunting tasks was to create an integrated system in which different LTD and linguistic measures were crossed and tested. By adopting a method based on the analysis of LTD and linguistic measures (Mennen et al., 2012), I managed to explain and describe the complexity of pitch variation across the dimensions of pitch level and span. In my opinion, the kind of analysis I pursued in my thesis can be proposed as a model for future research on pitch range because: 1) it considers the double dimension of pitch range: pitch level and span, 2) it analyzes the types of measures that better capture cross-linguistic differences in pitch range, 3) it focuses on the comparison of production and perception data.

As a follow-up to the present study, it would be interesting to collect data obtained from speakers of L1 and L2 French, German, Spanish, etc. This would contribute to providing an accurate account of how pitch range differs across languages. In addition, the gathering of information from different languages is useful to confirm the universal and language-specific components of pitch variation I analyzed in the present study.

Since the investigation of pitch range involves various disciplines such as linguistics, engineering, psychology, speech therapy, and physiology, there is a great variety of possible ways to implement a study on pitch range. For example, the study of speech parameters such as voice quality and intensity could provide an additional overview on the inter-speaker variability of pitch range. One option for extending the present study would consist in considering the pitch range properties associated with the expression of emotion in speech. In this respect, several attempts to explain the correlation between pitch range and emotion have already been done (Mozziconacci and Hermes, 1999; Gussenhoven, 2002; Mozziconacci 2002; Liscombe et al., 2003; Chen, 2005; Hincks, 2005; Rosenberg and Hirschberg, 2005; Feinberg et al., 2008; Yuasa, 2008; Hincks and Edlund, 2009; Gravano et al., 2011).

In conclusion, the increasing interest in research on pitch range opens new opportunities for studies on the socio-phonetic use of pitch range across languages. The dynamic effect of L1 and L2 influence on pitch variation has an impact also on the creation of a model that accounts for prosodic and intonational aspects of L2 speech learning.

References

Abe, I. (1998). Intonation in Japanese. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 360-375.

Adams, C. and Munro, R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and nonnative speakers of English. *Phonetica*, 35, 125-156.

Ahrens, B. (2005). Prosodic phenomena in simultaneous interpreting. *Interpreting*, 7 (1), 51-76.

Alcoba, S. and Murillo J. (1998). Intonation in Spanish. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 152-166.

Alfano, I., Schwab, S., Savy, R., Llisterri, J. (2009). Cross-language speech perception: Lexical stress in Spanish with Italian and Francophone subjects. *Proc. Associazione Italiana di Scienze della Voce (AISV)*, Zürich, 455- 474.

Armstrong, M.E. (2010). Intonational encoding of pragmatic meaning: The case of absolute interrogatives in Puerto Rican Spanish. *Proc. Speech Prosody*, Chicago, IL.

Armstrong, M.E. (2010). Puerto Rican Spanish intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 155-189.

Apple, W., Streeter, L.A., Krauss, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37 (5), 715-727.

Aoyama, K. and Guion, S.G. (2007). Prosody in second language acquisition: An acoustic analysis on duration and F0 range. In: Bohn, O.-S. and Munro, M.J. (eds.), *The role of language experience in second-language speech learning: In honor of James Emil Flege.* Amsterdam: John Benjamins, 281-297.

Arslan, L.H. and Hansen, J.H.L. (1996). Language accent classification in American English. *Speech Communication*, 18 (4), 353-369.

Arslan, L.H. and Hansen, J.H.L. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *Journal of the Acoustical Society of America*, 102 (1), 28-40.

Arvaniti, A., Ladd D.R., Mennen I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics,* 26, 3-25.

Arvaniti, A. and Baltazani, M. (2000). Greek ToBI: A system for the annotation of Greek speech corpora. *Proc. 2nd International Conference on Language Resources and Evaluation (LREC),* Athens, 2, 555-562.

Arvaniti, A. and Baltazani M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In: Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 84-117.

Arvaniti, A., Ladd, D.R., Mennen, I. (2006). Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech*, 49 (4), 421-450.

Arvaniti, A. and Ladd D.R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26: 43-74.

Arvaniti, A. (2009). Greek intonation and the phonology of prosody: polar questions revisited. *Proc. 8th International Conference on Greek Linguistics*, 14-29.

Astruc, L., Mora, E., Rew, S. (2010). Venezuelan Andean Spanish Intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 191-226.

Atterer, M., and Ladd, D. R. (2004). On the phonetics and phonology of ''segmental anchoring'' of F0: evidence from German. *Journal of Phonetics*, 32, 177-197.

Avanzi, M., Obin, N., Lacheret-Dujour, A., Victorri, B. (2011). Toward a continuous modeling of French prosodic structure: using acoustic features to predict prominence location and prominence degree. *Proc. Interspeech*, Florence, 2033-2036.

Avesani, C. (1990). A contribution to the synthesis of Italian intonation. *Proc. of the International Conference on Spoken Language Processing*, 833–836.

Avesani, C. and Vayra, M. (2005). Quale informazione codificare per la sintesi dell'intonazione?. In: Biffi, M., Salibra, L., Calabrese, O. (eds.), *Italia linguistica. Discorsi di scritto e di parlato*. Siena: Protagon, 235-248.

Avesani, C. (2011). *Tono ed Intonazione*. Materials distributed in class for the lessons on Intonation, (not available online).

Backman, N.E. (1979). Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin,* 4 (2), 239-266.

Baken, R.J., and Orlikoff, R.F. (2000). *Clinical Measurement of Speech and Voice*. San Diego: Singular Publishing Group.

Barnes, J., Brugos, A., Veilleux, N., Shattuck-Hufnagel, S. (2011). Voiceless intervals and perceptual completion in F0 contours: evidence from scaling perception in American English. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 108-111.

Beckman, M. and Pierrehumbert, J.(1986). Intonational Structure in Japanese and English, *Phonology Yearbook*, 3, 15-70.

Beckman, M.E. and Hirschberg, J. (1993). The ToBI annotation conventions. Unpublished manuscript, (last assessed, December, 17th 2012). Available at: http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html

Beckman, M.E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17-67.

Beckman, M.E. and Ayers, G.M. (1997). Guidelines for ToBI labeling. Online MS and accompying files, (last assessed, December, 17th 2012). Available at http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/

Beckman, M., Díaz-Campos, M., McGory, J.T., Morgan, T.A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, 14, 9-36.

Beckman, M.E., Hirschberg, J., Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. OUP, 9-54.

Bent, T., Bradlow, A.R., Wright, B.A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32 (1), 97-103.

Benjamini, Y. (1988). Opening the box of a boxplot. *The American Statistician*, 42 (4), 257–262.

Bertinetto, P.M. (1980). The perception of stress by Italian speakers. *Journal of Phonetics,* 8, 385-395.

Bertinetto, P.M. (1981). Strutture prosodiche dell'italiano. *Quaderni degli studi di grammatica italiana*. Firenze: Accademia della Crusca.

Best, C., McRoberts, G., Sithole, N. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 45-60.

Best, C. (1995). A direct realist view of cross-language speech perception. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research.* Baltimore: York, 171-204.

Best, C., McRoberts, G., Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America,* 109, 775-794.

Best, C.T. and Tyler, M.D. (2007) Nonnative and second-language speech perception. Commonalities and complementarities. In: Bohn, O-S and Munro, M.J. (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam & Philadelphia: John Benjamins, 13–34.

Bird, S. and Harrington, J. (2001). Speech annotation and corpus tools. *Speech Communication*, 33, 1-4.

Bishop, J. and Keating, P. (2010). Perception of pitch location within a speaker's own range: fundamental frequency, voice quality and speaker sex. *UCLA Working Papers in Phonetics,* 108, 113-140.

Bishop, J. and Keating P. (2012). Perception of pitch location. *Journal of the Acoustical Society of America,* 132, 1100-1112.

Black, A. and Hunt, A. (1996). Generating F0 contours from ToBI labels using linear regression. *Proc. ICSLP*, 3, Philadelphia, 1385-1388.

Bocci, G. and Avesani, C. (2011). Phrasal prominences do not need pitch movements: postfocal phrasal heads in Italian. *Proc. Interspeech*, Florence, 1357-1360.

Boersma, P. and Weenink, D. (2010). *Praat: doing phonetics by computer* [computer software, version 5.3.32]. Amsterdam: University of Amsterdam. Available at http://www.praat.org

Bolinger, D. (1972). *Intonation: selected readings*. Harmondsworth: Penguin.

Bolinger, D. (1978). Intonation across languages. In: Greeberg, J.H., Ferguson, C.A., Moravcik, E.A. (eds.), *Universals of Human Language,* 2. Stanford, CA: Stanford University Press, 471-524.

Bolinger, D. (1986). *Intonation and its parts. Melody in spoken English.* Palo Alto, CA: Stanford University Press.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse.* Palo Alto, CA: Stanford University Press.

Bolinger, D. (1998). Intonation in American English. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 45-55.

Borràs-Comes, J., Vanrell, M., Prieto, P. (2010). The role of pitch range in establishing intonational contrasts in Catalan. *Proc. Speech Prosody*, Chicago, IL.

Borràs-Comes, J. (2012). *The role of intonation and facial gestures in conveying interrogativity.* Ph.D. thesis. University Pompeu Fabra.

Botinis, A. (1998). Intonation in Greek. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 288-310.

Boula de Mareüil, P., Marotta, G., Adda-Decker, M. (2004). Contribution of prosody to the perception of Spanish/Italian accents. *Proc. 2nd International Conference on Speech Prosody,* Nara, Japan, 681-684.

Boula de Mareüil, P., Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63* (4), 247-67.

Brazil, D. (1984). The intonation of sentences read aloud. In: Gibbon, D. and Richter, H. (eds.), *Intonation, accent and rhythm: Studies in discourse phonology.* Berlin, Germany: Walter de Gruyter, 46-66.

Brazil, D. (1985). *The communicative value of intonation in English.* Birmingham, England: English Language Research.

Brinckmann, C. and Benzmüller, R. (1999). The relationship between utterance type and F0 contour in German. *Proc. Eurospeech*, 1, 21-24.

Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, 4, 77-104.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In: Archibald, J. (ed.), *Second Language Acquisition and Linguistic Theory*. Blackwell, 4-63.

Brown, G., Currie, K., Kenworthy, J. (1980). *Questions of intonation.* Baltimore: University Park Press.

Bruce, G., Elert, C., Engstrand, O., Wretling, P. (1999). Phonetics and phonology of the Swedish dialects – a project presentation and a database presentor. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 321-324.

Bruce, G. (2003). Aspects of North Swedish intonational phonology. *Phonum, 9*, 17-20.

Busà, M.G. (1995). *L'Inglese degli Italiani. L'acquisizione delle Vocali.* Padova: Unipress.

Busà, M.G. (2008a). New perspectives in teaching pronunciation. In: Baldry, A., Pavesi, M., Taylor Torsello, C., Taylor, C. (eds.), *From DIDACTAS to eCoLingua: an ongoing research project on translation and corpus linguistics*. Trieste: EUT, 171-188.

Busà, M.G. (2008b). Teaching prosody to Italian learners of English. Working towards a new approach. *Ecolingua. The Role of E-corpora in Translation and Language Learning*, 113-126.

Busà, M.G. (2010). Effects of L1 on L2 pronunciation. Italian prosody in English. In: Gagliardi, C. and Maley, A. (eds.), *EIL, ELF, Global English: Teaching and learning processes*. Bern: Peter Lang, 207-228.

Busà, M.G. and Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 380-383.

Busà, M.G. and Stella, A. (2012). Intonational variations in focus marking in the English spoken by North-East Italian speakers. In: Busà, M.G. and Stella, A. (eds.), *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*. Padova: CLEUP, 31-35.

Cabrera Abreu, M. and Vizcaíno Ortegarie, F. (2010). Canarian Spanish intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa, 87-121.

Campbell, N. and Beckman, M. (1997). Stress, prominence, and spectral tilt. In: Botinis, A., Kouroupetroglou, G., Carayiannis, G. (eds.), *Intonation: Theory, Models and Applications*. Athens: ESCA, 67-70.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning,* 59, 2, 448-451.

Chen, A. (2011). What's in a rise. Evidence for an off-ramp analysis of Dutch intonation. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 380-383.

Chen, A., Gussenhoven, C., Rietveld, T. (2004). Language specificity in perception of paralinguistic intonational meaning. *Language and Speech*, 47 (4), 311-349.

Chen, A., Rietveld, T., Gussenhoven, C. (2001). Language-specific effects of pitch range on the perception of universal intonational meaning. *Proc. Eurospeech*, Aalborg, 1403-1406.

Cho, H. and Flemming, E. (2011). The phonetic specification of contour tones: the rising tone in Mandarin. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 112-115.

Chun, D.M. (2002). *Discourse intonation in L2. From theory and research to practice*. Amsterdam: John Benjamins.

Clark, R. A. J. (1999). Using prosodic structure to improve pitch range variation in text to speech synthesis. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 69-72.

Clark, R. A. J. (2003). *Generating Synthetic Pitch Contours Using Prosodic Structure*. PhD thesis, The University of Edinburgh.

Collados Aís, Á. (1998). *La evaluación de la calidad en interpretación simultánea: La importancia de la comunicación no verbal*. Granada: Comares.

Conture, E. G. and Curlee, R. F. (2007). Stuttering and related disorders of fluency. (3rd ed.). New York, NY: Thieme Medical.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9 (6), 864–881.

Crocco, C. (2006). Prosodia e funzioni delle domande polari in italiano napoletano. In: Savy, R. and Crocco, C. (eds.), *Analisi Prosodica: Teorie, Modelli e Sistemi di Annotazione*, Padova: ISTC/EDK. 164-181.

Cruttenden, A. (1992). Review of 't Hart, Collier and Cohen 1990. *Journal of Linguistics* 28*,* 522-526.

Cruttenden, A. (1994). Rises in English. In: Windsor-Lewis, J. (ed.), *Studies in general and English Phonetics: Essays in Honour of Professor J.D. O'Connor*. London: Routledge, 155-173.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: CUP.
Cruz-Ferreira, M. (1998). Intonation in European Portuguese. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation System. A survey of 20 languages.* Cambridge: CUP, 167-178.
Cucchiarini, C., Strik, H., Boves, L. (2002). Quantitative assessment of second language learners' fluency. Comparisons between read and spontaneous Speech. *Journal of the Acoustical Society of America*, 111 (6), 2862-2873.
Curtin, S. and Werker, J.F. (2007). The perceptual foundations of phonological development. In: Gaskell M.G. (ed.). *The Oxford Handbook of Psycholinguistics*. New York: OUP, 579-599.
Cutler, A. (2002). Native listeners. *European Review*, 10 (1), 27-41.
Cutler, A., Wales, R., Cooper, N., Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. *Proc. 16th International Conference of Phonetic Sciences (ICPhS)*, Saarbrücken, 1913-1916.
Daly, N. and Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5 (1), 85-96.
Dankovičová, J. (1999). Articulation rate variation within the intonation phrase in Czech and English. *Proc. 14th International Conference of Phonetic Sciences (ICPhS),* San Francisco, 269-272.
Dauer, R. M. (1993). *Accurate English: A complete course in pronunciation*. Englewood Cliffs, NJ: Regents/Prentice Hall.
De Dominicis, A. (2003). Assertive e interrogative a Bologna e Roma. In: *La fonetica acustica come strumento di analisi della variazione linguistica in Italia*. Atti delle XII e Giornate di Studio del G.F.S.. Roma: Il Calamo, 129-136.
De Jong, N.H. and Wempe, T. (2007). Automatic measurement of speech rate in spoken Dutch. *ACLC Working papers,* 2 (2), 49-58.
De Jong, N.H. and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods,* 41, 385-390.
De la Mota, C., Butragueño, P. M., Prieto, P. (2010). Mexican Spanish intonation. In: Prieto, P., and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*. München: Lincom Europa, 319-350.
Derwing, T.M., and Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition*, 19 (1), 1-16.
Deutsch, D. (1992). Some new pitch paradoxes and their implications, *Philos. Trans. R. Soc. London*, Ser. B, 336, 391–397.
Deutsch, D., Le, J., Shen, J., Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *Journal of the Acoustical Society of America*, 125 (5), 208- 2013.
Dilley, L. and Brown, M. (2007). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics*, 35, 523-551.
Dilley, L. (2010). Pitch range variation in English tonal contrasts. Continuous or categorical? *Phonetica*, 67, 63-81.
D'Imperio, M.P. and House, D. (1997). Perception of questions and statements in Neapolitan Italian. *Proc. Eurospeech,* Rhodes, 251-254.

D'Imperio, M.P. (2002). Italian intonation. An overview and some questions. *Probus*, 14, 37-69.
Di Cristo, A. and Hirst, D. (1993). Prosodic regularities in the surface structure of French questions. *Prosody*, 268-271.
Di Cristo, A. (1998). Intonation in French. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: Cambridge University Press, 195-218.
Docherty, G.J. and Ladd, D.R. (eds.) (1992). *Papers in laboratory phonology II. Gestures, segments, prosody.* Cambridge: Cambridge University Press.
Drugman, T. And Alwan, A. (2011). Joint robust voicing detection and pitch estimation based on residual harmonics. *Proc. Interspeech*, Florence, 1973-1976.
Eisenchlas, S. and Tsurutani, C. (2011). You sound attractive! Perceptions of accented English in multi-lingual environment. *Australian Review of Applied Linguistics*, 34 (2), 216-236.
Epstein, M. (2002). *Voice quality and prosody in English.* Ph.D. thesis. The University of California – Los Angeles.
Erickson, D., Honda, K., Hirai, H., Beckman, M. (1995). The production of low tones in English intonation. *Journal of Phonetics*, 23, 179-188.
Escudero, D., Aguilar, L., Vanrell, M.M., Prieto, P. (2011). Analysis of inter-transcriber consistency in the Cat ToBI prosodic labelling system. *Speech Communication,* 54, 566-582.
Estebas-Vilaplana, E. and Prieto, P. (2010). Castilian Spanish intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 17-48.
Face, T. and Prieto, P. (2007). Rising accents in Castilian Spanish: a revision of Sp_ToBI. *Journal of Portuguese Linguistics*, 6 (1), 117-146.
Faharani, F., Georgiou, P.G., Narayanan, S.S. (2004). Speaker identification using supra-segmental pitch pattern dynamics. *Proc. ICASSP*, 1, 89-92.
Farnetani, E. and Kori, S. (1990). Rhythmic structure in Italian noun phrases. *Phonetica,* 47, 50-65.
Farnetani, E. and Zmarich, C. (1997). Prominence patterns in Italian. An analysis of F0 and duration. *Proc. INT-1997*, Athens, 115-118.
Farnetani, E. and Busà, M.G. (1999). Quantifying the range of vowel reduction in Italian. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Franciso, 491-494.
Feagin, C. (1997). The African contribution to southern states English. In: Bernstein, C., Nunally, T., Sabino, R. (eds.). *Language variety in the south revisited*. Tuscaloosa, AL: University of Alabama, 123-139.
Feinberg, D.R., Jones, B.C., Little, A.C., Burt, D.M ., Perrett, D.I. (2005). Manipulation of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour,* 69 (3), 561-568.
Feinberg, D.R., De Bruine, L.M., Jones, B.C., Perrett, D.I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, 37 (4), 615-623.
Feldhausen, I., Peskova, A., Kireva, E., Gabriel, C. (2011). Categorical perception of Porteño nuclear accents. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 116-119.

Felloni, M.C. (2010). *Uno studio sociofonetico sulla intonazione. Produzione e percezione della interrogativa globale nell'italiano regionale di Parma*. Ph.D. thesis. Università degli Studi di Pavia.

Féry, C., Kaiser, E., Hörnig, R., Weskott, T., Kliegl, R. (2009). Perception of intonation contours on given and new referents: a completion study and an eye-movement experiment. In: Boersma, P. and Hamann, S. (2009). *Phonology in Perception*. Berlin: Mouton de Gruyter.

Flege, J., Hillenbrand, J. (1986). Differential use of temporal cues to the / s/-/z/ contrast by native and non native speakers of English. *Journal of the Acoustical Society of America*, 79, 508- 517.

Flege, J. (1995). Second language speech learning. Theory, findings, and problems. In Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Baltimore: York Press, 233- 273.

Flege, J. E. (1999). Age of learning and second-language speech. In: Birdsong, D.P (ed.). *New Perspectives on the Critical Period Hypothesis for Second Language Acquisition*. Hillsdale, NJ: Erlbaum.

Fougeron, C. and Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics*, 26, 45-69.

Frota, S. (2002). Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation. *Probus*, 14-1, 113-146.

Frota, S. (2003). The phonological status of initial peaks in European Portuguese. *Catalan Journal of Linguistics*, 2, 133-152.

Frota, S. and Vigario, M. (2001). On the correlated of rhythmic distinctions: the European Brazilian Portuguese case. *Probus*, 13, 247-273.

Gabriel, C., Feldhausen, I., Pešková, A., Colantoni, L., Lee, S.-A., Arana, V., Labastía, L. (2010). Argentinian Spanish intonation. In: Prieto, P. and Roseano, P. (eds.). *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 285-317.

Gac, D. and Yoo, H. (2011). Intonation of left dislocated topics in modern Greek. *Proc. Interspeech*, Florence, 1361-1364.

Garding, E. (1998). Intonation in Swedish. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 112-130.

Garellek, M. and Keating, P. (2010). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *UCLA Working Papers in Phonetics*, 108.

German, J. And D'Imperio, M. (2010). Focus, phrase length, and the distribution of phrase-intial rises in French. *Proc. Speech Prosody*, Chicago IL, 1-4.

Giannini, A. and Pettorino, M. (1992). *La fonetica sperimentale*. Roma: Edizioni Scientifiche Italiane.

Gibbon, D. and Richter, H. (1984). *Intonation, Accent and Rhythm. Studies in Discourse Phonology.* Berlin: Walter de Gruyter.

Gibbon, D. (1998). Intonation in German. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 78-95.

Gilbert, J. B. (1993). *Clear speech: Pronunciation and listening comprehension in North American English.* (2nd ed.). New York: Cambridge University Press.

Gili Fivela, B. (2002). Tonal alignment in two Pisa Italian peak accents. *Proc. Speech Prosody*, Aix-en Provence, 339-342.

Gili Fivela, B. (2008). *Intonation in Production and Perception. The case of Pisa Italian*. Alessandria: Edizioni dell'Orso.

Gili Fivela, B., Stella, A., D'Apolito, S., Sigona, F. (2011). Coarticulation across prosodic domains in Italian: an ultrasound investigation. *Proc. Interspeech*, Florence, 393-396.

Gilles, P. (2005). Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung. Berlin: Mouton de Gruyter.

Giordano, R. (2006). The intonation of polar questions in two central varieties of Italian. *Proc. Speech Prosody*, Dresden, Germany, 1-4.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29, 383-406.

Grabe, E. (2002a). Variation adds to prosodic typology. *Speech Prosody*, 127-132.

Grabe, E. and Post, B. (2002). Intonational variation in the British isles. *Speech Prosody*, 343-346.

Grabe, E., Kochanski, G., Coleman, J. (2003). Quantitative modeling of intonational variation. *Proc. Speech Analysis and Recognition in Technology, Linguistics and Medicine*.

Grant, L. (1993). *Well said: Advanced English pronunciation*. Boston: Heinle & Heinle.

Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., Nenkova, A. (2011). Acoustic and prosodic correlates of social behavior, *Proc. Interspeech*, Florence, 97-100.

Grieser, D. and Kuhl, P. (1989). Categorization of speech by infants. Support for speech-sound prototypes. *Developmental Psychology,* 25, 577-588.

Gronnum, N. (1998). Intonation in Danish. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages*. Cambridge: CUP, 131-151.

Grice, M. (1991). The intonation of interrogation in two varieties of Sicilian Italian. *Proc. 12th International Conference of Phonetic Sciences (ICPhS)*, Aix-en-Provence, 210-213.

Grice, M. (1995). *The intonation of interrogation in Palermo Italian: implications for intonation theory*. Niemeyer: L.A. series.

Grice, M. and Savino, M. (1995). Low tone versus 'sag' in Bari Italian Intonation: a perceptual experiment. *Proc. 13th International Conference of Phonetic Sciences (ICPhS)*, Stockholm, 658-61.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., Batliner, A. (1996). Consistency in Transcription and Labelling of German Intonation with GToBI. *Proc. ICSLP*, Philadelphia.

Grice, M., D'Imperio, M., Savino, M., Avesani. C. (2004). Strategies for intonation labelling across varieties of Italian. In: Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 362- 389.

Grice, M., Baumann, S., Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In: Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 55-83.

Grice, M. and Baumann, S. (2007). An introduction to intonation – Functions and Models. In: Trouvain, J. and Gut U. (eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin/New York: De Gruyter, 25-51.

Guimarães, I. and Abberton, E. (2005). Fundamental frequency in speakers of Portuguese for different voice samples. *Journal of voice*, 19 (4), 529-605.

Guimarães, I. and Gouveia, S. (2007). Do Portuguese women have a lower pitch than other female subjects? *Proc. 27th World Congress of the International Association of Logopedics and Phoniatrics*, Copenhagen.

Gussenhoven, C. and Rietveld, T., (2000). The behavior of H*and L* under variations in pitch range in Dutch rising contours. *Language and Speech*, 43, 183-203.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. *Proc. Speech Prosody*, Aix-en-Provence, 47-57.

Gussenhoven, C. (2002). Phonology of intonation. *GLOT International 6*, 9/10, 1-32.

Gussenhoven, C., Rietviled, T., Kekhoff, J., Terken, J. (2003). ToDI second edition. Online courseware available at http://todi.let.kun.nl/ToDI/home.htm (last accessed, December, 17th 2012).

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In: Sun, A.-J. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP.

Gut, U. (2000). *Bilingual acquisition of intonation. A study of children speaking German and English.* Tübingen, Niemeyer: Mouton de Gruyter.

Guy, G. and Vonwiller, J. (1989). The high rise tones in Australian English. In: Collins, P. and Blair, D. (eds), *Australian English*. St Lucia: UQP.

Haan, J., and Van Heuven, V.J. (1999). Male vs. female pitch range in Dutch questions. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 1581–1584.

Hancin-Bhatt, B. (1994) Segment transfer. A consequence of a dynamic system. *Second Language Research,* 10 (3), 241-269.

Hancin-Bhatt, B. and Govindjee, A. (1999). A computational model of feature competition in L2 transfer. In: Broeder, P. and Murre, J. (eds.) *Language and thought in development: Cross-linguistic studies*. Tübingen: Gunter Narr, 145-161.

Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M.A.K. and Greaves, W.S. (2008). *Intonation in the grammar of English*. London: Equinox.

Hadding-Koch, K. and Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica*, 11, 175-185. Reprinted in Bolinger, D. (1972). *Intonation: selected readings*. Harmondsworth: Penguin, 348-358.

Hagen, S.A., and Grogan, E. (1992). *Sound advantage. A pronunciation book.* Englewood Cliffs, NJ: Prentice Hall.

Hahn, L. D. (2004). Primary stress and intelligibility. Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38* (2), 201-223.

Hanson, H. and Chuang, E. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, 106, 1064-77.

Harrington, J. (2008). Intonation: Transcribing with ToBI. Sidney: Macquarie University. (Last assessed 10th January 2013). Available at: http:// clas.mq.edu.au/speech/phonetics/phonology/intonation/tobi_transcribe.html

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39 (3), 379-440.

Hedberg, N., Sosa, J.M., Görgülü, E. (2008). Early and late nuclei in yes-no questions: tails or high rises? *Proc. Speech Prosody*, Campinas, Brazil.

Hedberg, N., Sosa, J.M., Görgülü, E., Mameni, M. (2010). The prosody and intonation of wh-questions in American English. *Proc. Speech Prosody*, Chicago, Illinois.

Hedberg, N. and Sosa, J.M. (2011). The phonetics of final pitch accents in American English Polar questions. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 847-350.

Hermes, D.J. and van Gestel, J.C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, 90, 97-102.

Hillenbrand, J.M. (2003). Some effects of intonation contour on speech intelligibility. *Journal of the Acoustical Society of America*, 114 (4), 2338.

Hinkel, E. (2005). *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech. A proposal for an automatic feedback mechanism. *System, 33* (4), 575-591.

Hincks, R. and Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning and Technology*, 13 (3), 32-50.

Hirschberg, J. and Avesani, C. (2000). Prosodic disambiguation in English and Italian. In: Botinis, A. (ed.), *Intonation. Analysis, Modelling, and Technology*. Dordrecht: Kluwer Academic Publishers, 87-95.

Hirschberg, J. (2002). Communication and prosody. Functional aspects of prosody. *Speech Communication*, 36, 31-43.

Hirschberg, J. (2004). Pragmatics and intonation. In: Horn, L.R. and Ward, G. (eds.), *The Handbook of Pragmatics*. Oxford: Basil Blackwell, 515-537.

Hirschberg, J., Benus, S., Gravano, A. (2007). Prosody, emotions, and...'whatever', *Proc. Interspeech*, Antwerp.

Hirst, D. (1998). Intonation in British English. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 56-77.

Ho, C.-H. (2001). *Speaker Modelling for Voice Conversion*. Ph.D. thesis. Brunel University.

Hollien, H. (1974). On vocal registers. *Journal of Phonetics*, 2, 125-143.

Holub, E. (2010). Does intonation matter? The impact of monotony on listener comprehension. *The Interpreters' Newsletter*, 15, 117-126.

Horvath, B. (1986). *Variation in Australian English. The sociolects of Sydney.* Cambridge: Cambridge University Press.

Hyman, L.M. (2006). Between stress and tone. *Phonology*, 23 (2), 225-257.

Hyman, L.M. (2011). Tone: Is it different?. In: Goldsmith, J., Riggle, J., Yu, A. (eds.) *The Handbook of Phonological Theory*, (2nd ed.), Oxford: Wiley-Blackwell.

Iivonen, A. (1998). Intonation in Finnish. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 56-77.

Iivonen, A. (1999). F0 Contours of utterances superimposed on the temporal voice range profile of the speaker. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 953-956.

Interlandi, G.M., Romano, A., de Castro L.M., Coimbra, R.L. (2007). Cross-linguistic prosodic variation in two Romance languages: Portuguese and Italian. *Proc. 3rd Congreso Internacional de Fonética Experimental*, Santiago de Compostela, 337-387.

Iseli, M., Shue, Y.-L., Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America, 121*, 2283-2295.

Iverson, P. and Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America,* 97 (1), 553-562.

Iverson, P. and Kuhl, P. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America,* 99 (2), 1130-1140.

Jassem, W. (1971). Pitch and compass of the speaking voice. *Journal of the International Phonetics Association*, 1, 59-68.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenner, B. (1976). Interlanguage and foreign accent. *Interlanguage Studies Bulletin*, 1 (2-3), 166-195.

Jiang, H. (2011). Gender difference in English intonation. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 974-977.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent.* Ph.D. thesis, University of Stuttgart.

Jilka, M. (2000). Testing the contribution of prosody to the perception of foreign accent. *Proc. New Sounds*, Amsterdam, 199 - 207.

Johnson, K. (2003). *Acoustic and auditory phonetics*. (2nd ed.). Oxford: Blackwell.

Johnson, K. (2008). *Quantitative methods in linguistics.* Oxford: Blackwell.

Jun, S.-A. (2000). KToBI (Korean ToBI) Labelling Conventions (version 3.1). Online courseware. (Last accessed December, 17th 2012). Available at: http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html

Jun, S.-A. (ed.) (2005). *Prosodic typology. The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, S.-A and Oh, M. (2000). Acquisition of second language intonation. *Proc. International Conference on Spoken Language Processing,* Beijing, China, (4), 76-79.

Jun, S.-A. and Fougeron, C. (2000). A phonological model of French intonation. In: Botinis, A. (ed.), *Intonation: Analysis, Modeling and Technology.* Dordrecht: Kluwer Academic Publishers, 209-242.

Jun, S.-A. and Fougeron, C. (2002). The realizations of the accentual phrase in French intonation. *Probus*, 14: 147-172.

Karlsson, A., House, D., Svantesson, J.O., Tayanin, D. (2011). Comparison of F0 range in spontaneous speech in Kammu tonal and non-tonal dialects. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 1026-1029.

Keating, P. (2003). Linguistic phonetics in the UCLA phonetics lab. *UCLA Working Papers in Phonetics,* 103, 12-29.

Keating, P. and Kuo, G. (2010). Comparison of speaking fundamental frequency in English and Mandarin, *UCLA Working Papers in Phonetics*, 108, 164-187.

Keating, P. and Kuo, G. (2012). Fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America,* 132, 1050-1060.

Keating, P., Esposito, C., Garellek, M., Khan, S., Kuang, J. (2011). Phonation contrasts across languages. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 1046-1049.

Kennedy, S. and Tofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech. The role of listener experience and semantic context. *The Canadian Modern Language Review*, 64 (3), 459-489.

Klatt, D. H. and Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-57.

Kormos, J. and Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32 (2), 145-164.

Krahmer, E. and Swerts, M. (1998). Reconciling two competing views on contrastiveness. *Proc. ICSLP*, Sydney, 1299-1302.

Krahmer, E. and Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, 34 (4), 391-405.

Krashen, S. D. (1987). *Principles and practice in second language acquisition*. London: Prentice Hall International.

Kuang, J. (2012). Registers in tonal contrasts. *Proc. 3rd TAL-Tonal Aspects of Languages,* Nanjing, China, 1-4.

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93- 107.

Kuhl, P.K., and Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-language Research.* Baltimore: York, 121-154.

Kuhl, P. K. (2000). A new view of language acquisition. *Proc. National Academy of Sciences of the United States of America*, 97 (22), 11850-11857.

Kuwabara, H. and Sagisaka, Y. (1995). Acoustic characteristics of speaker individuality: Control and conversion. *Speech Communication*, 16 (2), 165-173.

Labov, W. (1998). The three dialects of English. In: Linn, M.D. (ed.), *Handbook of dialects and language variation*. San Diego: Academic Press, 39-81.

Labov, W., Ash, S., Boberg, C. (2006). *The Atlas of North American English*. Berlin: Mouton de Gruyter.

Ladd, D.R., Silverman, K.E.A., Tolkmitt, F., Bergmann, G., Scherer, K.R. (1985). Evidence for the independent function of intonation contour type, voice quality, and f0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, 78, 435-444.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: CUP

Ladd, D.R., Mennen, I., Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, 107, 2685-2696.

Ladd, D.R. (2007). Alignment of tonal targets: 30 years on. Text of a talk presented in Bruce's 60th birthday symposium in Lund. Available at: http://www.lel.ed.ac.uk/~bob/bruce60.pdf (Last accessed 19th Nov 2012).

Ladefoged, P. (1996) *Elements of Acoustic Phonetics*. (2nd ed.). Chicago: University of Chicago Press.

Ladefoged, P. (2003). *Phonetic data analysis: an introduction to instrumental phonetic fieldwork*. Oxford: Blackwells.

Ladefoged, P. (2006). *A course in Phonetics*. Boston: Thomson.

Lane, D.M., (eds.) (2012). Online Statistics Education: A Multimedia Course of Study. At: http://onlinestatbook.com/Online_Statistics_Education.pdf (Last accessed 14th Dic 2012).

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.

Lehiste, I. (1996). Suprasegmental features of speech. In: Lass, N.J. (ed.). *Principle of Experimental Phonetics*. Saint-Louis: Mosby, 226-244.

Le Gac, D., Yoo, H., Bartkova, K. (2012). Pitch level variation in French: categorical or gradient?, *Speech Prosody*, Shanghai.

Leino, T. (1998). In search of 'optimal pitch': the lowest possible tone as the reference in the evaluation of speaking pitch in Finnish speakers. *Proc. 24th IALP Congress*, 56-59.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-412.

Lepetit, D. (1989). Cross-linguistic influence on intonation: French/Japanese and French/English. *Language Learning* 39: 397-413.

Lepschy, A. L. (1978). Appunti sull'intonazione. In: Lepschy, A. L. (ed.), *Saggi di linguistica italiana*. Bologna: Il Mulino, 111-126.

Lepschy, A. L. and Lepschy, G. (1977). *The Italian Language Today*. London: Hutchinson.

Levis, J. M. (1996). *An experimental study of the status of two low-rising intonation contours in American English.* PhD thesis, University of Michigan, Ann Arbor.

Levis, J.M. (1999). The intonation and meaning of normal yes/no questions. *World Englishes*, 18, 373-380.

Levis, J.M. (2002). Reconsidering low-rising intonation in American English. *Applied Linguistics*, 23 (1), 56-82.

Li, A. and Zu, Y. (2008). Speaking rate effects on discourse prosody in Standard Chinese. *Proc. Speech Prosody*, Campinas, Brazil, 449-452.

Li, K., Zhang, S., Li, M., Lo, W.-K., Meng, H. (2011). Prominence model for prosodic features in automatic lexical stress and pitch accent detection. *Proc. Interspeech*, Florence, 2009-2012.

Lieberman, P. (1967). *Intonation, Perception and Language*. Cambridge: MIT Press.

Liberman, M. (1975). *The intonational system of English*. Ph.D thesis. Cambridge: MIT.

Liberman, M. and Pierrehumbert J. (1984). Intonational invariance under changes in pitch range and length'. In: M. Aronoff and R.T. Oherle, *Language Sound Structure*. Cambridge, MA: MIT Press, 157-233.

Lin, H.-Y., and Fon, J. (2011). The role of pitch reset in perception at discourse boundaries. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 1242-1245.

Lin, H.-Y., and Fon, J. (2012). Prosodic and acoustic features of emotional speech in Taiwan Mandarin. *Proc. Speech Prosody*, Shanghai, China, 450-453.

Liscombe, J., Venditti, J., Hirschberg, J. (2003). Classifying subjective ratings of emotional speech. *Proc. Eurospeech*, Geneva.

Llisterri, J., Machuca, M., de la , Mota, C., , Riera, M., Ros, A. (2003). The perception of lexical stress in Spanish. *Proc. 15th International Conference of Phonetic Sciences (ICPhS)*, Barcelona, 2023- 2026.

López-Bobo, M.J. and Cuevas-Alonso, M. (2010). Cantabrian Spanish Intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 49-85.

Magno Caldognetto, E., Ferrero, F.E., Lavagnoli, C., Vagges, K. (1978). F0 contours of statements, yes-no questions and wh-questions of two regional varieties of Italian. *Journal of Italian Linguistics*, 3, 57-68.

Mayewski, W., Hollien, H., Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica*, 25 (2), 119-25.

McGill, R, Tukey, J.W., Larsen, W.A. (1978). Variations of box plots. *The American Statistician*, 32 (1), 12–16.

Mennen, I. (1998). Can language learners ever acquire the intonation of a second language? *Proc. STILL,* Marholmen, Sweden, 17-20.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543-563.

Mennen, I. (2006). Phonological and phonetic influences in non-native intonation. In: Trouvain, J. and Gut, U. (eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice.* Berlin: Mouton De Gruyter, 1-18.

Mennen, I., Schaeffler, F., Docherty, G. (2007). Pitching it differently: a comparison of the pitch ranges of German and English speakers. *Proc. 16th International Conference of Phonetic Sciences (ICPhS)*, Saarbrücken, 1769-1772.

Mennen, I., Schaeffler, F., Docherty, G. (2008a). A methodological study into the linguistic dimensions of pitch range differences between German and English. *Proc. IV Conference on Speech Prosody*, University of Campinas, 527-530.

Mennen, I., Schaeffler, F., Docherty, G. (2008b). Cross-language differences in pitch range: Full Research Report ESRC End of Award Report, RES-000-22-1858. Swindon: ESRC.

Mennen, I., Chen, A., Karlsson F. (2010). Characterizing the internal structure of learner intonation and its development over time. *Proc. New Sounds*, Poznan.

Mennen, I., Schaeffler, F., Docherty, G. (2012). Cross-language difference in f0 range: a comparative study of English and German. *Journal of the Acoustical Society of America*, 131 (3), 2249-2260.

Mitchell, A.G. and Delbridge, A. (1965). *The pronunciation of English in Australia*, Sidney: Angus and Robertson.

Mok, P.P.K., and Wong, P.W.Y. (2010). Perception of the merging tones in Hong Kong Cantonese. Preliminary data on monosyllables. *Proc. of Speech Prosody*, Chicago, 2010.

Moraes, J. A. (1998). Intonation in Brazilian Portuguese. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 179-194.

Mozziconacci, S. and Hermes, D.J. (1999). Role of intonation patterns in conveying emotion in speech. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 2001-2004.

Mozziconacci, S. (2002). Prosody and emotions. *Proc. Speech Prosody*, Aix-en-Provence, 1-9.

Munger, D. (2008). High-pitched voices are most attractive - with a few exceptions. *ScienceBlogs LLC.* Posted on 3rd July 2008. Retrieved at: http://scienceblogs.com/cognitivedaily/2008/07/03/highpitched-voices-are-most-at/

Munro, M.J. (1995). Nonsegmental factors in foreign accent. Ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17-34.

Munro, M.J., Derwing, T.M., Burgess, C.S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52* (7-8), 626- 637.

Nadeu, M. and Prieto, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics*, 43, 841-854.

Nelson, C.L. (1992). Intelligibility and non-native varieties of English. In: Kachru, B.B. (ed.), *The other tongue: English across cultures*, (2nd ed.). Urbana: University of Illinois Press, 58-73.

Niebuhr, O., D'Imperio, M., Gili Fivela, B., Cangemi, F. (2011). Are there 'shapers' and 'aligners'? Individual differences in signalling pitch accent category. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 120-123.

Niebuhr, O. and Wolf, A. (2011). Low and high, short and long by crook or by hook?, *Proc. Interspeech*, Florence, 1869-1872.

Nolan, F. (2003) Intonational equivalence: an experimental evaluation of pitch scales. *Proc. 15th International Conference of Phonetic Sciences (ICPhS),* Barcelona, 771-774.

Nolan, F. (2006). Intonation. In: Aarts, B. and McMahon, A. (eds.), *Handbook of English Linguistics*. Oxford: Blackwell.

Nygaard, L.C., Herold, D., Namy, L. (2009). The semantics of prosody. Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33, 127–146.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica,* 40, 1-18.

Ohala, J. J. (1984). An ethological perspective on common cross - language utilization of F0 of voice. *Phonetica*, 41, 1-16.

Ohala, J.J. (1994). The frequency code underlines the sound symbolic use of voice of pitch. In: Hinton, L. (ed.). *Sound symbolism.* Cambridge: CUP. 325-247.

Orion, G. (1996). *Pronouncing American English.* (2nd ed.). New York: Newbury House.

Ortiz, H., Fuentes, M., Astruc, L. (2010). Chilean Spanish Intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 255-283.

O'Connor, R. (2008). A prosodic projection for role and reference grammar. In : Van Valin, R.D. (ed.), *Investigations of the Syntax-Semantics-Pragmatics Interface*. Amsterdam/Philadelphia: John Benjamins, 227-244.

O'Rourke, E. (2010). Ecuadorian Andean Spanish intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 227-253.

Pakosz, M. (1982). Intonation and attitude. *Lingua*, 56 (2), 153–178.

Pan, H.-H. (2007). Focus and Taiwanese unchecked tones. In: C. Lee, M. Gordon, D. Büring. Topic and Focus. Cross-linguistic Perspectives on Meaning and Intonation. Dordrecht: Springer, 195-213.

Pan, S. and Mckeown, K. (1999). Word informativeness and automatic pitch accent modeling. *Proc. SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*. University of Maryland.

Patterson, D. and Ladd, D. R. (1999). Pitch range modelling: linguistic dimensions of variation. *Proc. 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, 1169-1172.

Patterson, D. (2000). *A Linguistic Approach to Pitch Range Modelling.* PhD dissertation, University of Edinburgh.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes, 23*, 19-43.

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 219-233.

Pierrehumbert, J. (1979) The Perception of Fundamental Frequency Declination, *Journal of the Acoustical Society of America,* 66, 363-369.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*, Ph.D. Thesis, MIT.

Pierrehumbert, J. and Beckman, M.E. (1988). *Japanese Tone Structure*, Cambridge (MA): MIT Press.

Pierrehumbert, J., and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In: Cohen, P.R., Morgan, J., Pollack, M.E. (eds.), *Intentions in Communication*. Cambridge MA: The MIT Press, 271–311.

Pierrehumbert, J. (2006). Prosody, intonation, and speech technology. In Bates, M. and Weischedel, R.M. (eds.), *Challenges in natural language processing*. Cambridge: CUP, 257-280.

Pisoni, D.B. and Remez, R.E. (eds.) (2005). *The Handbook of Speech Perception*. Malden (MA)/Oxford: Blackwell.

Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Post, B., D'Imperio, M., Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning. *Proc. 16th ICPhS*, Saarbrücken, 191-196.

Prieto, P., Van Santen, J., Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429-451.

Prieto, P., Shih, C., Nibert, H. (1996). Pitch Downtrend in Spanish, *Journal of Phonetics*, 24, 445-473.

Prieto, P., D'Imperio, M., Gili Fivela, B. (2005). Pitch accent alignment in Romance: primary and secondary association with metrical structure. In:

Warren, P. (ed.) *Language and Speech, Special Issue: Intonation in Language varieties*. London: Kingston Press, 48 (4), 359-396.

Prieto, P. (2008). Tonal alignment patterns in Catalan nuclear falls. *Lingua*, special issue on *The phonology and phonetics of tone from a typological perspective*.

Prieto, P., and Roseano, P. (coords.) (2009-2012). *Atlas interactivo de la entonación del español*. (Last accessed 12th Dec. 2012). Available at: http://prosodia.upf.edu/atlasentonacion/

Prieto, P. (2013). The Intonational Phonology of Catalan. In: S.-A. (ed.), *Prosodic Typology 2. The Phonology of Intonation and Phrasing*. Oxford: OUP.

Ramirez, Verdugo M.D. (2005). The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: pragmatic effects. *Journal of Pragmatics*, 37 (12), 2086- 2115.

Ramus, F., and Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech re-synthesis. *Journal of the Acoustical Society of America*, 105 (1), 512-521.

Ramus, F., Nespor, M., Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

Rasinger, Sebastian M. (2008). *Quantitative Research in Linguistics*. London: Continuum.

Reckling, F. and Kügler, F. (2011). Pitch range in positive and negative connoted statements of German. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 1670-1673.

Reinisch, E., Jesse, A., Mc Queen J.M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech*, 54 (2), 147-165.

Rognoni, L. (2012). The impact of prosody in foreign accent detection. A perception study of Italian accent in English. In: Busà, M.G. and Stella, A. (eds.), *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*. Padova: CLEUP, 89-93.

Romano, A. (1997). Definizione di sei sotto-varietà intonative del Salento: prime valutazioni dei risultati di un test di riconoscimento. In: Cutugno F. (ed.), *Fonetica e fonologia degli stili dell'italiano parlato. Proc. VII Giornate di Studio del G. F. S.,* Napoli, 27-49.

Rosenberg, A. and Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. *Proc. of Interspeech 2005*, Lisbon.

Ross, K. and Ostendorf, M. (1994). A Dynamic system model for generating F0 for synthesis. *Proc. ESCA/IEEE Workshop on Speech Synthesis*, 131-134.

Rossano, F. (2010). Questioning and responding in Italian. *Journal of Pragmatics*, 42, 2756-2771.

Rossi, M. (1998). Intonation in Italian. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 219-238.

Sas, Library. Repeated measures ANOVA using SAS PROC GLM. *UCLA: Statistical Consulting Group*. (Last accessed 30th Jan 2013). Available at http://www.ats.ucla.edu/stat/sas/library/repeated_ut.htm

Savino E., Grice M., Gili Fivela B., Marotta G. (2006). Intonational cues to discourse structure in Bari and Pisa Italian: perceptual evidence. *Proc. Speech Prosody,* Dresden,114-117.

Savino, M. and Grice, M. (2007). The role of pitch range in realizing pragmatic contrasts. The case of two questions types in Italian. *Proc. 16th International Conference of Phonetic Sciences (ICPhS)*, Saarbrücken, 1037-1040.

Scherer, K.R. (1979). Personality markers in speech. In: Scherer, K.R. and Giles, H. (eds.), Social Markers in Speech, Cambridge University Press, 147-209.

Sebastián, N. (2005). Cross-language speech perception. In: Pisoni, D.B. and Remez, R.E. (eds.). *Handbook of Speech Perception*. New York: Blackwell.

Sebastian-Galles, N. and Diaz B. (2012). First and second language speech perception: Graded learning. *Language Learning*, 62 (2), 131-147.

Shriberg, E.E., Ladd, D.R., Terken, J M.B., and Stolcke, A. (1996). Modeling pitch range variation within and across speakers: Predicting F0 targets when 'speaking up'. *Proc. ICSLP*, Stockholm, 2, 386–389.

Shue, Y.-L., Keating, P. Vicenik, C. (2009). VoiceSauce: A program for voice analysis. *Journal of the Acoustical Society of America, 126*, 2221.

Shue, Y.-L., Keating, P., Vicenik, C., Yu, K. (2011) VoiceSauce: A program for voice analysis, *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 1846-1849.

Silverman, D. (2003). Pitch discrimination during breathy versus modal phonation, In: Local, J., Ogden, R., Temple, R. (eds.), *Laboratory Phonology VI*. Cambridge: CUP, 293-304.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Ph.D thesis. Cambridge University.

Silverman K., Beckman, M.E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992). ToBI: A Standard for Labeling English Prosody, *Proc. 2th ICSLP*, Banf , Canada, 2, 867-870.

Sorianello, P. (2001). Modelli intonativi dell'interrogazione in una varietà di italiano meridionale (Cosenza). *Rivista Italiana di Dialettologia* XXV, 85-108.

Sorianello P. (2006). *Prosodia. Modelli e ricerca empirica*, Roma, Carocci.

Steedman, M. (2007). Information-structural semantics for English intonation. In: Lee, C., Gordon, M., Büring, D. *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*. Dordrecht: Springer, 245-264.

Stella, A. and Gili Fivela, B. (2009). L'intonazione nell'italiano dell'area leccese: prime osservazioni dal punto di vista autosegmentale-metrico. In: Romito, L., Galatà, V., Lio, R. (eds.), *La fonetica sperimentale: metodo e applicazioni. Proc. Associazione Italiana di Scienze della Voce (AISV)*, Cosenza, 259-292.

Stockwell, R. and Bowen, D.J. (1965). *The sounds of English and Spanish*. Chicago: Chicago Press.

Strangert, E., and Gustafson, J. (2008). *Subject ratings, acoustic measurements and synthesis of good-speaker characteristics. Proc. of Interspeech*, Brisbane, Australia.

Styler, W. (2012). Using Praat for Linguistic Research. University of Colorado at Boulder Phonetics Lab. [Document version, 1.3]. http://savethevowels.org/praat/UsingPraatforLinguisticResearch.pdf.

Svetozarova, N. (1998). Intonation in Russian. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 261-274.

Swerts, M., Krahmer, E., Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics*, 30, 629-654.

't Hart, J., Collier, R., Cohen, A. (1990). *A perceptual study of intonation.* Cambridge: Cambridge University Press.

't Hart, J. (1998). Intonation in Dutch. In: Hirst, D. and Di Cristo, A. (eds.), *Intonation Systems. A survey of 20 languages.* Cambridge: CUP, 96-111.

Tench, Paul (1996). *The Intonation Systems of English*. London: Cassell.

Terken, J. (1997). Variation of accent prominence within the phrase: Models and spontaneous speech data. In: Yoshinori, S., Campbell, N., Higuchi, N. (eds.). *Computing Prosody: Computational Models for Processing Spontaneous Speech*. New York: Springer, 95-116.

Thompson, S. (1995). Teaching intonation on questions. *ELT Journal,* 49, 235–243.

Thorsen N. (1983). Standard Danish sentence intonation: phonetic data and their representation. *Folia Linguistica*, 17, 187–220.

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America*, 83, 1536–1552.

Toivanen, J. (2003). Tone Choice in the English Intonation of Proficient Non-native Speakers. *PHONUM, 9*, 1-4.

Toivanen, J., and Waaramaa, T. (2005). Tone choice and voice quality of dispreferred turns in the English of Finns. *Logopedics Phoniatrics Vocology,* 30, 181–184.

Tokuma, S. and Xu, Y. (2011). L1/L2 perception of lexical stress with F0 peak-delay: effect of an extra syllable added. *Proc. Interspeech*, Florence, 1881-1884.

Tøndering, J. (2011). Preplanning of intonation in spontaneous versus read aloud speech. Evidence from Danish. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 2010-2013.

Towell, R., Hawkins, R., Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84-119.

Traunmüller, H., and Eriksson, A. (1995). The perceptual evaluation of *F0* excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America, 97* (3), 1905-1915.

Trofimovich, P. and Baker, W. (2006). Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Studies in Second Language Acquisition*, *28* (1), 1-30.

Ueyama, M. (1997). The phonology and phonetics of second language intonation: the case of "Japanese English". *Proc. 5th European Conference on Speech Communication and Technology,* 2411-2414.

Ueyama, M. and Jun, S.-A. (1997). Focus realization in Japanese English and Korean English intonation. *UCLA Working Papers in Phonetics 94.*

Ueyama, M. and Jun, S.-A. (1998). Focus realization in Japanese English and Korean English intonation. In: Hajime, H. (ed.), *Japanese and Korean Linguistics*. CSLI, Stanford University Press, 629-645.

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). In: Trouvain, J. and Barry, J.W.(eds.). *Proc. 16th International Conference of Phonetic Sciences (ICPhS),* Saarbrücken,1701-1704.

Ullakonoja, R. (2008). Speech Rate as an Indicator of Fluency in the Russian of Finnish Learners. In: O'Dell, M. and Nieminen, T. (eds.) *Fonetiikan päivät 2008 – The Phonetics Symposium 2008*. Tampere: Tampere University Press, 97-109.

Ullakonoja, R. and Dufva, H. (2008). Perception of L2 fluency in study abroad context. *Academic Exchange Quarterly*, (12) 3, 62-66.

Ullakonoja, R. (2010a). Pitch contours in Russian yes/no questions by Finns. In: Hasegawa-Johnson, M., Bradlow, A.,Cole, J., Livescu, K., Pierrehumbert, J., Shih, C. (eds.) *Proc. Speech Prosody*, Chicago.

Ullakonoja, R. (2010b). How do native Russian speakers evaluate yes/no questions produced by Finnish L2 learners?. *Rice Working Papers in Linguistics,* 2, 92-105.

Vaissière, J. (2005). Perception of Intonation. In: Pisoni, D.B. and Remez, R.E. (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell, 236- 263.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech,* 38, 253-265.

van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures - oral performances of students aged 11-12 years. *Language Testing*, 11, 291-319.

Vanrell, M.M. (2006). *The phonological role of tonal scaling in Majorcan Catalan interrogatives*. MA Thesis, Universitat Autònoma de Barcelona.

Vanrell, M.M. (2008). Function of focus and intonation in Majorcan Catalan WH-questions, *Workshop on Information Structure and Prosody*, Studiecentrum Soeterbeeck, Netherlands.

Veilleux, N., Shattuck-Hufnagel, S., Brugos, A. (2006). 6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI, January IAP 2006**.** (MIT OpenCourseWare), http://ocw.mit.edu (Last accessed 17th Dec, 2012). License: Creative Commons BY-NC-SA.

Venditti, J. (2006). Prosody in sentence processing. In: Mazuka, R., Nakayama, M., Shirai, Y. (eds.). *The Handbook of East Asian Psycholinguistics, Volume II: Japanese*, Cambridge: CUP, 208-217.

Vicenik, C. (2011). *The role of intonation in language discrimination by infants and adults*. Ph.D thesis, University of California, Los Angeles.

Vieru-Dimulescu, B. and Boula de Mareüil, P. (2005). Contribution of prosody to the perception of a foreign accent: a study based on Spanish/Italian modified speech. *Proc. ISCA Workshop on Plasticity in Speech Perception*, London, 66-68.

Vigario, M. and Frota, S. (2003). The intonation of Standard and Northern European Portuguese. *Journal of Portuguese Linguistics*, 2-2, 115- 137.

Vion, M. and Colas, A. (2006). Pitch cues for the recognition of yes – no questions in French. *Journal of psycholinguistic research,* 35, 427 – 445.

Waibel, A. (1979). The pitch contour in wh-questions. *Journal of the Acoustical Society of America*, 65 (1), 33-37.

Wang, B. (2008). What can we tall from these temporal measures? Temporal measures as indices of oral proficiency. *English Language Teaching*, 1 (2), 21-31.

Warren, P. (2005). Patterns of late rising in New Zealand English: Intonation variation or intonational change? *Language Variation and Change*, 17, 209-230.

Watson, P.J. and Schlauch, R.S. (2008). The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *American Journal of Speech-Language Pathology,* 17, 348-55.

Welby, P. (2003). *The slaying of lady mondergree, being a study of French tonal association and alignment and their role in speech segmentation.* Ph.D. Thesis. The Ohio State University.

Welby, P. (2006). French intonational structure. Evidence from tonal alignment. *Journal of Phonetics*, 34 (3), 343-371.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics, 15* (4), 399-421.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition,* 20, 1-25.

Wennerstorm, A. (2000). The role of intonation in second language fluency. In: Riggenbach, H. (ed.), *Perspectives on fluency*. Michigan: The University of Michigan Press, 102-127.

Wichmann, A. (2000). *Intonation in Text and Discourse. Beginnings, Middles and Ends*. Harlow: Longman.

Willems, N.J. (1982). *English intonation from a Dutch point of view*. Dordrecht: Foris Publications.

Willis, E.W. (2010). Dominican Spanish Intonation. In: Prieto, P. and Roseano, P. (eds.), *Transcription of Intonation of the Spanish Language*, Munich: Lincom Europa, 123-153.

Wood, S. (2005). *Praat for beginners. The beginners guide to Praat.* Lund University Phonetic Lab. (Last accessed, 10th Nov. 2012) http://person2.sol.lu.se/SidneyWood/praate/frames.html

Yang, L.-C. (2011). Tonal variations in Mandarin: new evidence from spontaneous and read speech. *Proc. Interspeech*, Florence, 1397-1400.

Yuan, J. (2010). Linguistic rhythm in foreign accent. *Proc. Interspeech*, Makuhari, 1848-1849.

Yuasa, P.I. (2008). *Culture and Gender of Voice Pitch. A Sociophonetic Comparison of the Japanese and Americans*. London: Equinox.

Zadeh, V.A., Gussenhoven, C., Bijankhan, M. (2011). A pitch accent position contrast in Persian. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 188-191.

Zipp, L. and Dellwo, V. (2011) Reading-Speech-Normalization: A method to study prosodic variability in spontaneous speech. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 2328-2331.

Appendix A:

Appendix A contains all the t-tests and ANOVAs calculated to carry out statistical analysis. Before comparing the results of t-tests and ANOVAs, the characteristics and settings of these two kinds of test analyzing the distribution of data are briefly outlined. The explanations given below are based on the materials (personal notes and slides presented in class) provided during lessons in the Experimental Phonetics (Keating and Garellek) and my readings on statistics (Rasinger, 2008; Li, 2010; Lane, 2012).

1. T-test

A t-test compares exactly two datasets (of some dependent variable) that differ with respect to one manipulation (of an independent variable). With little or no overlap between two samples, it is clear that they are different. Whether there is overlap depends on both the difference between the means of the samples (bigger difference means less overlap), and the variability in the samples (less variability means less overlap).

Selecting a t-test in Excel:

In my production and perception studies, I chose to use paired two-tails t-tests (also called paired two samples for means). However, other t-test types may be used depending on the nature and distribution of data across samples. There are 3 different t-tests in excel that one can select for an experimental design: Paired Two Sample for Means, Two-Sample Assuming Equal Variances, and Two-Sample Assuming Unequal Variances.

1. *Paired Two Sample for Means*

A paired two sample for means is a paired test that has to be used when the data in the two samples are matched, item by item – e.g. each pair from one speaker, or the speakers have been matched in pairs in some way. Paired tests are common in phonetics.

2. *Two-Sample Assuming Equal Variances* (this is unpaired)

An unpaired test, such as the two-sample assuming equal variances, is used when the data can be regarded as two unordered sets. In fact, variances in the two samples are defined as equal when variability is roughly the same in the two samples.

3. *Two-Sample Assuming Unequal Variances* (this is unpaired)

An unpaired test, such as the two-sample assuming unequal variances, is used when the data can be regarded as two unordered sets. This test is selected for data where variability is expected to be unequal in the two samples.

Setting of parameters in a t-test

The first step to do in a t-test is to identify the data samples for the two variables to be compared (variable 1 vs. variable 2). In order to tell the program which data samples to compare, it is necessary to select a range of cells in the excel file for each sample. As a second step, after having selected the data to be compared, one has to consider two fundamental parameters such as the hypothesized mean difference and the alpha value (The Italian translation of terms is given in brackets to facilitate the examination of t-tests shown in the following pages).

- *Hypothesized Mean Difference (Differenza Ipotizzata per le Medie)*: it describes the overall difference one is testing for. Its value is usually 0 (i.e. no difference) because that is the null hypothesis.
- *Alpha Value (Valore Alpha)*: it defines the probability of error that can be considered acceptable for stating that there is a difference between two samples. The default value is 0.05, that means a 5% chance of error. If a test requires stricter parameters, the investigator can decide to lower the percentage of acceptability (e.g. the alpha value can be set at 0.01, that means a 1% chance of error).

Interpretations of values and outputs of a t-test

- *Mean (Media)*: The average of the data in each group.
- *Variance (Varianza):* The variability around that mean.

- *Observations (Osservazioni):* The number of cases considered in the analysis.
- *Pearson Correlation (Correlazione di Pearson)*:
  Measure of the correlation between variables 1 and 2.
- *Df (Gdl):*
  Degrees of freedom in sample given the mean. In other words, it is the number of values that are free to vary. The df is calculated as the number of pairs – 1.
- *t Stat (Stat t):*
  Ratio of difference between means of samples. It is a measure of variability (the standard error of the difference). The larger t stat is, the more likely the samples are to be different (and the sign tells the direction of any difference between the two samples).
- *P one-tail (P una coda); P two-tail (P due code):*
  Probability that difference is by chance; the smaller this number, the less likely the difference is by chance. P-value is compared to the alpha value, which is the stated criterion for p-value. A p-value equal to or less than alpha means that there is a small enough chance of error (5% or less for alpha of .05) that there is a reliable difference between the samples. A p-value greater than alpha means that there is a high enough chance of error (greater than 5% for alpha of .05) that there is not a reliable difference between the samples. The definition one tail vs. two-tails indicates in how many directions (1 or 2) the investigator is testing for a difference between the data samples.
- *t Critical one-tail (t critico una coda); t Critical two-tail (t critico due code):* Required minimum value of t for significance at the chosen alpha. The definition one tail vs. two-tails indicates in how many directions (1 or 2) the investigator is testing for a difference between the data samples. In linguistics, the directions tested are almost always two.

2. ANOVA

An analysis of variance (ANOVA) compares the significance level of difference among groups of data that differ with respect to one or more factors. In the testing, the variables examined are either independent or dependent. The means of the samples are calculated in order to test whether the mean values obtained across the dependent variables depend one or more independent variable.

Selecting a ANOVA in SPSS:

In my production and perception studies, I chose to use one-way ANOVAs. However, other kinds of ANOVAs may be used depending on the nature and distribution of data across samples. There are three different types of ANOVAs one can select for an experimental design: one-way ANOVA, multi-factor ANOVA, repeated measures ANOVA.

- *One-way ANOVA*

  One-way ANOVA tests the equality of means and is used when data are tested for one independent variable and one or more dependent variables. Typically, the one-way ANOVA is used to test for the difference among at least three groups. 'In the case that only two groups are compared, one-way ANOVA is equivalent with t-test' (Li, 2010: 139).

- *Multi-factor ANOVA*

  Multi-factor ANOVA tests the effects of more than one independent variables. For example, an ANOVA measuring the interaction between two independent variables is called two-way ANOVA. When the tested independent variables are three, the ANOVA is called three-way ANOVA and so forth. 'As the number of independent variables increases, the number of interactions increases and the interpretation of the model becomes more difficult' (Li, 2010: 140).

- *Repeated measures ANOVA*

  Repeated measures ANOVA is used especially in longitudinal studies 'when all members of a random sample are measured under a number of different conditions. As the sample is exposed to each condition in turn, the measurement of the dependent variable is repeated' (Sas, 2013). 'Repeated measures design is suitable for the studies in which the same measures are collected multiple times for each subject but under different conditions' (Li, 2010: 140).

Setting of parameters in an ANOVA

An ANOVA calculates '*between-group* variation by measuring the differences between each group mean and the overall mean, and then averaging the differences' (Li, 2010: 138). It also calculates '*within-group* variation by calculating the difference between each data value and the mean of its group, and then averaging the differences' (Li, 2010: 138).

Interpretations of values and outputs of an ANOVA

As already observed in the explanation of p-value in the t-test, the difference among groups of data is significant depending on the p-value. In statistical significance testing, The p-value is defined as 'the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The lower the p-value, the less likely the result is if the null hypothesis is true, and consequently the *more* "significant" the result is, in the sense of statistical significance' (Li, 2010: 139). P-value is compared to the alpha value, which is the stated criterion for p-value. A p-value equal to or less than alpha means that there is a small enough chance of error (5% or less for alpha of .05) that there is a reliable difference between the samples. A p-value greater than alpha means that there is a high enough chance of error (greater than 5% for alpha of .05) that there is not a reliable difference between the samples.

Production Studies: first experiment

One-way ANOVA testing the results obtained for F0 min, max, mean, range, median, SD, slope, slope without octave jumps in the male population. The factor being tested is the language spoken in the utterances (English vs. Italian). Post-hoc tests are measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
min	Between Groups	712,826	1	712,826	4,531	,038
	Within Groups	9125,050	58	157,328		
	Total	9837,875	59			
max	Between Groups	77,044	1	77,044	,017	,896
	Within Groups	259015,485	58	4465,784		
	Total	259092,528	59			
mean	Between Groups	1271,338	1	1271,338	,349	,557
	Within Groups	211481,228	58	3646,228		
	Total	212752,566	59			
range	Between Groups	80,294	1	80,294	,057	,812
	Within Groups	81250,900	58	1400,878		
	Total	81331,194	59			
median	Between Groups	100,935	1	100,935	,170	,681
	Within Groups	34350,135	58	592,244		
	Total	34451,070	59			
SD	Between Groups	19,955	1	19,955	,056	,814
	Within Groups	20649,126	58	356,019		
	Total	20669,081	59			
slope	Between Groups	2338,184	1	2338,184	,055	,815
	Within Groups	2454559,198	58	42319,986		
	Total	2456897,382	59			
slope w/oct	Between Groups	51,929	1	51,929	,539	,466
	Within Groups	5586,081	58	96,312		
	Total	5638,010	59			

One-way ANOVA testing the results obtained for F0 min, max, mean, range, median, SD, slope, slope without octave jumps in the male population. The factor being tested is the native language of the speakers (American English vs. Italian). Post-hoc tests are measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
min	Between Groups	2563,069	1	2563,069	20,435	,000
	Within Groups	7274,806	58	125,428		
	Total	9837,875	59			
max	Between Groups	7,268	1	7,268	,002	,968
	Within Groups	259085,261	58	4466,987		
	Total	259092,528	59			
mean	Between Groups	2888,064	1	2888,064	,798	,375
	Within Groups	209864,502	58	3618,353		
	Total	212752,566	59			
range	Between Groups	574,343	1	574,343	,412	,523
	Within Groups	80756,850	58	1392,359		
	Total	81331,194	59			
median	Between Groups	,563	1	,563	,001	,976
	Within Groups	34450,507	58	593,974		
	Total	34451,070	59			
SD	Between Groups	125,966	1	125,966	,356	,553
	Within Groups	20543,115	58	354,192		
	Total	20669,081	59			
slope	Between Groups	16628,656	1	16628,656	,395	,532
	Within Groups	2440268,726	58	42073,599		
	Total	2456897,382	59			
slope w/oct	Between Groups	84,790	1	84,790	,886	,351
	Within Groups	5553,221	58	95,745		
	Total	5638,010	59			

One-way ANOVA testing the results obtained for F0 min, max, mean, range, median, SD, slope, slope without octave jumps in the female population. The factor being tested is the language spoken in the utterances (English vs. Italian). Post-hoc tests are measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
min	Between Groups	15792,693	1	15792,693	8,761	,004
	Within Groups	104556,037	58	1802,690		
	Total	120348,730	59			
max	Between Groups	27754,877	1	27754,877	2,932	,092
	Within Groups	548972,260	58	9465,039		
	Total	576727,137	59			
mean	Between Groups	85412,819	1	85412,819	7,977	,006
	Within Groups	621036,106	58	10707,519		
	Total	706448,925	59			
range	Between Groups	418,795	1	418,795	,142	,708
	Within Groups	171498,875	58	2956,877		
	Total	171917,670	59			
median	Between Groups	165,599	1	165,599	,101	,752
	Within Groups	95537,633	58	1647,201		
	Total	95703,231	59			
SD	Between Groups	626,410	1	626,410	1,079	,303
	Within Groups	33679,510	58	580,681		
	Total	34305,920	59			
slope	Between Groups	645626,700	1	645626,700	7,805	,007
	Within Groups	4797551,236	58	82716,401		
	Total	5443177,936	59			
slope w/oct	Between Groups	13,581	1	13,581	,114	,737
	Within Groups	6937,667	58	119,615		
	Total	6951,248	59			

One-way ANOVA testing the results obtained for F0 min, max, mean, range, median, SD, slope, slope without octave jumps in the female population. The factor being tested is the. native language of the speakers (American English vs. Italian). Post-hoc tests are measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
min	Between Groups	3406,409	1	3406,409	1,689	,199
	Within Groups	116942,322	58	2016,247		
	Total	120348,730	59			
max	Between Groups	9367,224	1	9367,224	,958	,332
	Within Groups	567359,913	58	9782,067		
	Total	576727,137	59			
mean	Between Groups	24063,787	1	24063,787	2,045	,158
	Within Groups	682385,137	58	11765,261		
	Total	706448,925	59			
range	Between Groups	369,024	1	369,024	4,125	,025
	Within Groups	171548,647	58	2957,735		
	Total	171917,670	59			
median	Between Groups	741,834	1	741,834	,453	,504
	Within Groups	94961,397	58	1637,265		
	Total	95703,231	59			
SD	Between Groups	352,330	1	352,330	,602	,441
	Within Groups	33953,590	58	585,407		
	Total	34305,920	59			
slope	Between Groups	123341,232	1	123341,232	1,345	,251
	Within Groups	5319836,704	58	91721,322		
	Total	5443177,936	59			
slope w/oct	Between Groups	69,754	1	69,754	,588	,446
	Within Groups	6881,495	58	118,646		
	Total	6951,248	59			

Production Studies: second experiment

Test t: due campioni accoppiati per medie, valori di F0 mean in YNQ

	mean AM	*mean IT*
Media	238.585	248.002
Varianza	1002.357652	843.1327232
Osservazioni	100	100
Correlazione di Pearson	-0.0263229	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-2.16389128	
P(T<=t) una coda	0.016439139	
t critico una coda	1.660391157	
P(T<=t) due code	0.032878278	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 max in YNQ

	max AM	*max IT*
Media	340.514	337.047
Varianza	3324.442024	2251.188577
Osservazioni	100	100
Correlazione di Pearson	0.045719999	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	0.475088439	
P(T<=t) una coda	0.317885202	
t critico una coda	1.660391157	
P(T<=t) due code	0.635770403	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 min in YNQ

	min AM	*min IT*
Media	179.032	178.694
Varianza	1011.057956	822.4173374
Osservazioni	100	100
Correlazione di Pearson	0.102696756	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	0.083306272	
P(T<=t) una coda	0.466888113	
t critico una coda	1.660391157	
P(T<=t) due code	0.933776227	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di range (in ST) in YNQ

	Strange AM	*Strange IT*
Media	11.191	11.038
Varianza	10.23416061	8.700763636
Osservazioni	100	100
Correlazione di Pearson	0.084054995	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	0.367332627	
P(T<=t) una coda	0.357077496	
t critico una coda	1.660391157	
P(T<=t) due code	0.714154993	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 mean in WHQ

	mean AM	*mean IT*
Media	234.398	240.776
Varianza	1037.663834	1118.812145
Osservazioni	100	100
Correlazione di Pearson	0.336868469	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-1.686297705	
P(T<=t) una coda	0.047443782	
t critico una coda	1.660391157	
P(T<=t) due code	0.094887564	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 max in WHQ

	max AM	*max IT*
Media	315.254	363.136
Varianza	1850.913216	3307.885358
Osservazioni	100	100
Correlazione di Pearson	-0.155115694	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-6.219788574	
P(T<=t) una coda	5.97318E-09	
t critico una coda	1.660391157	
P(T<=t) due code	1.19464E-08	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 min in WHQ

	min AM	*min IT*
Media	135.249	170.084
Varianza	2384.419292	1071.08762
Osservazioni	100	100
Correlazione di Pearson	0.112412682	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-6.260375125	
P(T<=t) una coda	4.95865E-09	
t critico una coda	1.660391157	
P(T<=t) due code	9.91731E-09	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di range (in ST) in WHQ

	Strange AM	*Strange IT*
Media	15.652	13.297
Varianza	45.37565253	15.19281919
Osservazioni	100	100
Correlazione di Pearson	0.153720978	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	3.25032526	
P(T<=t) una coda	0.00078785	
t critico una coda	1.660391157	
P(T<=t) due code	0.0015757	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 mean in STM

	mean AM	*mean IT*
Media	204.911	220.99
Varianza	484.4135141	819.2912121
Osservazioni	100	100
Correlazione di Pearson	0.216858337	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-5.008884611	
P(T<=t) una coda	1.19589E-06	
t critico una coda	1.660391157	
P(T<=t) due code	2.39178E-06	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 max in STM

	max AM	*max IT*
Media	293.916	318.982
Varianza	3041.224186	2692.305733
Osservazioni	100	100
Correlazione di Pearson	0.131790596	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-3.552226637	
P(T<=t) una coda	0.000293409	
t critico una coda	1.660391157	
P(T<=t) due code	0.000586819	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di F0 min in STM

	min AM	*min IT*
Media	111.392	157.814
Varianza	1398.400339	561.9850545
Osservazioni	100	100
Correlazione di Pearson	-0.048459264	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-10.26216087	
P(T<=t) una coda	1.46673E-17	
t critico una coda	1.660391157	
P(T<=t) due code	2.93346E-17	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, valori di range (in ST) in STM

	Strange AM	*Strange IT*
Media	17.424	12.194
Varianza	36.43133737	10.52541818
Osservazioni	100	100
Correlazione di Pearson	0.114532261	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	8.025172217	
P(T<=t) una coda	1.06001E-12	
t critico una coda	1.660391157	
P(T<=t) due code	2.12003E-12	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Americans

	mean YNQ	*mean WHQ*
Media	238.585	234.398
Varianza	1002.357652	1037.663834
Osservazioni	100	100
Correlazione di Pearson	0.357496768	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	1.156457925	
P(T<=t) una coda	0.125138462	
t critico una coda	1.660391157	
P(T<=t) due code	0.250276925	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Americans

	mean YNQ	*mean STM*
Media	238.585	204.911
Varianza	1002.357652	484.4135141
Osservazioni	100	100
Correlazione di Pearson	0.376589347	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	10.85725755	
P(T<=t) una coda	7.4459E-19	
t critico una coda	1.660391157	
P(T<=t) due code	1.48918E-18	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Americans

	mean WHQ	*mean STM*
Media	234.398	204.911
Varianza	1037.663834	484.4135141
Osservazioni	100	100
Correlazione di Pearson	0.276000598	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	8.769070679	
P(T<=t) una coda	2.63025E-14	
t critico una coda	1.660391157	
P(T<=t) due code	5.26049E-14	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Americans

	Strange YNQ	*Strange WHQ*
Media	11.191	15.652
Varianza	10.23416061	45.37565253
Osservazioni	100	100
Correlazione di Pearson	0.065424511	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-6.139826006	
P(T<=t) una coda	8.60733E-09	
t critico una coda	1.660391157	
P(T<=t) due code	1.72147E-08	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Americans

	Strange YNQ	*Strange STM*
Media	11.191	17.424
Varianza	10.23416061	36.43133737
Osservazioni	100	100
Correlazione di Pearson	0.12929512	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-9.655465412	
P(T<=t) una coda	3.08811E-16	
t critico una coda	1.660391157	
P(T<=t) due code	6.17623E-16	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Americans

	Strange WHQ	*Strange STM*
Media	15.652	17.424
Varianza	45.37565253	36.43133737
Osservazioni	100	100
Correlazione di Pearson	0.174210097	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-2.15456337	
P(T<=t) una coda	0.016810956	
t critico una coda	1.660391157	
P(T<=t) due code	0.033621911	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Italians

	mean YNQ	*mean WHQ*
Media	248.002	240.776
Varianza	843.1327232	1118.812145
Osservazioni	100	100
Correlazione di Pearson	0.602178086	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	2.567281055	
P(T<=t) una coda	0.005872902	
t critico una coda	1.660391157	
P(T<=t) due code	0.011745804	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Italians

	mean YNQ	*mean STM*
Media	248.002	220.99
Varianza	843.1327232	819.2912121
Osservazioni	100	100
Correlazione di Pearson	0.53261256	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	9.689957755	
P(T<=t) una coda	2.59677E-16	
t critico una coda	1.660391157	
P(T<=t) due code	5.19355E-16	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 level* across sentences by Italians

	mean WHQ	*mean STM*
Media	240.776	220.99
Varianza	1118.812145	819.2912121
Osservazioni	100	100
Correlazione di Pearson	0.533412716	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	6.534930041	
P(T<=t) una coda	1.39096E-09	
t critico una coda	1.660391157	
P(T<=t) due code	2.78192E-09	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Italians

	Strange YNQ	*Strange WHQ*
Media	11.038	13.297
Varianza	8.700763636	15.19281919
Osservazioni	100	100
Correlazione di Pearson	0.042145317	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-4.718094544	
P(T<=t) una coda	3.91323E-06	
t critico una coda	1.660391157	
P(T<=t) due code	7.82647E-06	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Italians

	Strange YNQ	*Strange STM*
Media	11.038	12.194
Varianza	8.700763636	10.52541818
Osservazioni	100	100
Correlazione di Pearson	0.065962396	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	-2.72746964	
P(T<=t) una coda	0.003775603	
t critico una coda	1.660391157	
P(T<=t) due code	0.007551205	
t critico due code	1.9842169	

Test t: due campioni accoppiati per medie, *F0 span* across sentences by Italians

	Strange WHQ	*Strange STM*
Media	13.297	12.194
Varianza	15.19281919	10.52541818
Osservazioni	100	100
Correlazione di Pearson	0.025152072	
Differenza ipotizzata per le medie	0	
gdl	99	
Stat t	2.202386113	
P(T<=t) una coda	0.014979327	
t critico una coda	1.660391157	
P(T<=t) due code	0.029958654	
t critico due code	1.9842169	

One-way ANOVA testing the results obtained for: number of syllables (nsyll); number of pauses (npause); duration (Dur); phonation time (PhT); speech rate (SR); articulation rate (AR); average syllable duration (ASD). The factor being tested is the L1 of the speakers (English vs. Italian). Post-hoc tests were measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
nsyll	Between Groups	63,013	1	63,013	19,368	,000
	Within Groups	2335,919	718	3,253		
	Total	2398,932	719			
npause	Between Groups	,168	1	,168	5,935	,015
	Within Groups	20,331	718	,028		
	Total	20,499	719			
dur (s)	Between Groups	336571,513	1	336571,513	81,761	,000
	Within Groups	2955662,675	718	4116,522		
	Total	3292234,187	719			
phonationtime (s)	Between Groups	105584,668	1	105584,668	96,808	,000
	Within Groups	783091,097	718	1090,656		
	Total	888675,765	719			
speechrate (nsyll/dur)	Between Groups	21967,401	1	21967,401	4,368	,037
	Within Groups	3611057,986	718	5029,329		
	Total	3633025,388	719			
articulation rate (nsyll / phonationtime)	Between Groups	265843,368	1	265843,368	25,496	,000
	Within Groups	7486395,431	718	10426,735		
	Total	7752238,799	719			
ASD (speakingtime/ nsyll)	Between Groups	728,626	1	728,626	9,462	,002
	Within Groups	55288,413	718	77,003		
	Total	56017,038	719			

One-way ANOVA testing the results obtained for: number of syllables (nsyll); number of pauses (npause); duration (Dur); phonation time (PhT); speech rate (SR); articulation rate (AR); average syllable duration (ASD). The factor being tested is the sentence type (YNQ vs. WHQ vs. STM). Post-hoc tests were measured with the Bonferroni Alpha Value (0.05).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
nsyll	Between Groups	437,519	2	218,760	79,968	,000
	Within Groups	1961,413	717	2,736		
	Total	2398,932	719			
npause	Between Groups	,144	2	,072	2,544	,079
	Within Groups	20,354	717	,028		
	Total	20,499	719			
dur (s)	Between Groups	146276,758	2	73138,379	16,669	,000
	Within Groups	3145957,429	717	4387,667		
	Total	3292234,187	719			
phonationtime (s)	Between Groups	124274,803	2	62137,401	58,284	,000
	Within Groups	764400,962	717	1066,110		
	Total	888675,765	719			
speechrate (nsyll/dur)	Between Groups	232126,658	2	116063,329	24,469	,000
	Within Groups	3400898,729	717	4743,234		
	Total	3633025,387	719			
articulation rate (nsyll / phonationtime)	Between Groups	306842,453	2	153421,226	14,775	,000
	Within Groups	7445396,346	717	10384,095		
	Total	7752238,799	719			
ASD (speakingtime/nsyll)	Between Groups	1996,404	2	998,202	13,249	,000
	Within Groups	54020,635	717	75,343		
	Total	56017,038	719			

Perception Studies: third experiment

Test t: due campioni accoppiati per medie nelle EXM valutate dagli americani

	Italian stimuli	*American stimuli*
Media	2.14	3.9
Varianza	0.548	0.36
Osservazioni	5	5
Correlazione di Pearson	0.973743524	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-18.97856808	
P(T<=t) una coda	2.27024E-05	
t critico una coda	2.131846782	
P(T<=t) due code	4.54048E-05	
t critico due code	2.776445105	

Test t: due campioni accoppiati per medie nelle EXM valutate dagli italiani

	Italian stimuli	*American stimuli*
Media	2.038461538	3.484615385
Varianza	0.49704142	0.333284024
Osservazioni	5	5
Correlazione di Pearson	0.663304027	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-6.000848981	
P(T<=t) una coda	0.001940262	
t critico una coda	2.131846782	
P(T<=t) due code	0.003880524	
t critico due code	2.776445105	

Test t: due campioni accoppiati per medie nelle YNQ valutate dagli americani

	Italian stimuli	*American stimuli*
Media	2.42	3.92
Varianza	0.212	0.527
Osservazioni	5	5
Correlazione di Pearson	0.709047137	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-6.515583641	
P(T<=t) una coda	0.001432219	
t critico una coda	2.131846782	
P(T<=t) due code	0.002864437	
t critico due code	2.776445105	

Test t: due campioni accoppiati per medie nelle YNQ valutate dagli italiani

	Italian stimuli	*American stimuli*
Media	2.407692308	3.676923077
Varianza	0.43387574	0.629881657
Osservazioni	5	5
Correlazione di Pearson	0.742231584	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-5.291016784	
P(T<=t) una coda	0.003062135	
t critico una coda	2.131846782	
P(T<=t) due code	0.00612427	
t critico due code	2.776445105	

Test t: due campioni accoppiati per medie nelle STM valutate dagli americani

	Italian stimuli	*American stimuli*
Media	2	3.8
Varianza	0.06	1.315
Osservazioni	5	5
Correlazione di Pearson	0.275907306	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-3.643993486	
P(T<=t) una coda	0.010943631	
t critico una coda	2.131846782	
P(T<=t) due code	0.021887262	
t critico due code	2.776445105	

Test t: due campioni accoppiati per medie nelle STM valutate dagli italiani

	Italian stimuli	*American stimuli*
Media	1.9	3.761538462
Varianza	0.209023669	0.159319527
Osservazioni	5	5
Correlazione di Pearson	-0.323440175	
Differenza ipotizzata per le medie	0	
gdl	4	
Stat t	-5.968489846	
P(T<=t) una coda	0.001979081	
t critico una coda	2.131846782	
P(T<=t) due code	0.003958161	
t critico due code	2.776445105	

Perception Studies: fourth experiment

Test t: due campioni accoppiati per medie per la variabile 'madrelingua del soggetto'

	narrow IT	*narrow AM*
Media	3.015151515	3.242424
Varianza	1.030536131	1.171096
Osservazioni	66	66
Correlazione di Pearson	-0.255470886	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	-1.11079469	
P(T<=t) una coda	0.135374444	
t critico una coda	1.668635976	
P(T<=t) due code	0.270748888	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'madrelingua del soggetto'

	standard IT	*standard AM*
Media	3.484848485	2.863636
Varianza	1.576689977	1.534965
Osservazioni	66	66
Correlazione di Pearson	-0.026071713	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	2.824410434	
P(T<=t) una coda	0.003141359	
t critico una coda	1.668635976	
P(T<=t) due code	0.006282718	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'madrelingua del soggetto'

	wide IT	*wide AM*
Media	2.560606061	2.69697
Varianza	1.234731935	0.860606
Osservazioni	66	66
Correlazione di Pearson	0.212107979	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	-0.860345286	
P(T<=t) una coda	0.196380528	
t critico una coda	1.668635976	
P(T<=t) due code	0.392761055	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	narrow IT	*standard IT*
Media	3.015151515	3.484848485
Varianza	1.030536131	1.576689977
Osservazioni	66	66
Correlazione di Pearson	0.090702576	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	-2.475525248	
P(T<=t) una coda	0.007959645	
t critico una coda	1.668635976	
P(T<=t) due code	0.015919289	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	narrow IT	*wide IT*
Media	3.015151515	2.560606061
Varianza	1.030536131	1.234731935
Osservazioni	66	66
Correlazione di Pearson	-0.048561559	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	2.396257388	
P(T<=t) una coda	0.00972513	
t critico una coda	1.668635976	
P(T<=t) due code	0.019450259	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	standard IT	*wide IT*
Media	3.484848485	2.560606061
Varianza	1.576689977	1.234731935
Osservazioni	66	66
Correlazione di Pearson	-0.065489139	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	4.339297828	
P(T<=t) una coda	2.54829E-05	
t critico una coda	1.668635976	
P(T<=t) due code	5.09658E-05	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	narrow AM	*standard AM*
Media	3.242424242	2.863636364
Varianza	1.171095571	1.534965035
Osservazioni	66	66
Correlazione di Pearson	0.277479039	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	2.196939126	
P(T<=t) una coda	0.015797959	
t critico una coda	1.668635976	
P(T<=t) due code	0.031595917	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	narrow AM	*wide AM*
Media	3.242424242	2.696969697
Varianza	1.171095571	0.860606061
Osservazioni	66	66
Correlazione di Pearson	0.10495008	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	3.283810308	
P(T<=t) una coda	0.000825711	
t critico una coda	1.668635976	
P(T<=t) due code	0.001651422	
t critico due code	1.997137887	

Test t: due campioni accoppiati per medie per la variabile 'condizione di pitch span'

	standard AM	*wide AM*
Media	2.863636364	4.093969697
Varianza	1.534965035	0.860606061
Osservazioni	66	66
Correlazione di Pearson	0.017036114	
Differenza ipotizzata per le medie	0	
gdl	65	
Stat t	2.636666652	
P(T<=t) una coda	0.190498903	
t critico una coda	1.668635976	
P(T<=t) due code	0.010997806	
t critico due code	1.997137887	

Appendix B:

Sample of the consent form signed by the subjects participating in my experiments:

MODULO DI CONSENSO ALLA PARTECIPAZIONE A STUDIO LINGUISTICO E AL TRATTAMENTO DEI DATI PERSONALI

Con la presente io sottoscritto/a ____

FIRST NAME LAST NAME

Acconsento che la mia voce sia audio registrata nell'ambito dello studio linguistico intrapreso dalla dottoranda Urbani Martina.

Acconsento inoltre al trattamento dei miei dati personali ai sensi della Legge 196/03, nella consapevolezza che i risultati del test verranno pubblicati anonimamente e che i dati non verranno in nessun caso divulgati per scopi diversi da quelli della ricerca scientifica.

In fede,

____________________________ (firma del partecipante) Padova, ________________

METADATA

Personal Information:	**Answers:**
Where are you from? (city and state)	
How old are you?	
Are you an American English native speaker?	□ yes □ no
Do you usually smoke?	□ yes □ no
Do you have any reported speech disorder?	□ yes □ no
Academic Information	
What is your major?	
How long have you been studying Italian? What is your level of Italian?	
Do you speak other foreign languages? Wh ch ones? What is your level?	
Language Proficiency	
How often do you speak Italian?	□ 80% □ 50% □ 20%
Have you spent some time in an Italian speaking country? How long?	
Do you watch movies in Italian or watch TV progr ms in Italian? How often?	

Thank you for your precious collaboration!

Speech material recorded for the production experiments. The passages, taken from 'The Little Prince' by Antoine de Saint Exupéry (online version).

ENGLISH MATERIALS

Yes-no questions

1. Do you need any money?
2. Have we met before?
3. Are you still there?
4. Can you open the door?
5. Do you wanna come for dinner?

Wh-questions

6. Where were you when the money ran out?
7. Why are you selling meat?
8. What was her name again?
9. What are you doing there?
10. What's wrong with you?

Statements

11. Now you are going away.
12. I hope I can see you on Monday.
13. We should go and visit your uncle.
14. I know you are leaving today.
15. You should go to Hawaii.

Extract from *The Little Prince*, chapter 2:

"What is that object?"
"That is not an object. It flies. It is an airplane. It is my airplane."
And I was proud to have him learn that I could fly.
He cried out, then:
"What! You dropped down from the sky?"
"Yes," I answered, modestly.
"Oh! That is funny!"

Extract from *The Little Prince*, chapter 6:

"I am very fond of sunsets. Come, let us go look at a sunset now."
"But we must wait," I said.
"Wait? For what?"
"For the sunset. We must wait until it is time."
At first you seemed to be very much surprised. And then you laughed to yourself.
You said to me:
"I am always thinking that I am at home!"

Extract from *The Little Prince*, chapter 20:

But it happened that after walking for a long time through sand, and rocks, and snow, the little prince at last came upon a road. And all roads lead towards men.

"Good morning," he said.

He was standing before a garden, all a-bloom with roses.

"Good morning," said the roses.

The little prince gazed at them. They all looked like his flower.

"Who are you?" he demanded, thunderstruck.

"We are roses," the roses said.

"Ah!" he said.

And he was overcome with sadness. His flower had told him that she was the only one of her kind in all the universe. And here were five thousand of them, all alike, in one single garden!

ITALIAN MATERIALS

Yes-no questions:

1. Hai bisogno di soldi?
2. Ci siamo già incontrati?
3. Sei ancora lì?
4. Puoi aprire la porta?
5. Vuoi venire a cena?

Wh-questions:

6. Dov’eri quando sono finiti i soldi?
7. Perché stai vendendo carne?
8. Qual era il suo nome?
9. Cosa stai facendo lì?
10. Cosa c’è che non va?

Statements:

11. Ora stai andando via.
12. Spero di vederti lunedì.
13. Dovremmo andare a trovare tuo zio.
14. So che partirai oggi.
15. Dovresti andare alle Hawaii.

Extract from *The Little Prince*, chapter 2:
"Che cos'e' quest'oggetto?"
"Non e' un oggetto. Vola. E' un aeroplano. E' il mio aeroplano".
Ero molto fiero di fargli sapere che volavo.
Allora grido':
"Come? Sei caduto dal cielo?"
"Si", risposi modestamente.
"Ah! Questa e' buffa..."

Extract from *The Little Prince*, chapter 6:
"Mi piacciono tanto i tramonti. Andiamo a vedere un tramonto adesso."
"Ma bisogna aspettare..." dissi.
"Aspettare che?"
"Il tramonto. Dobbiamo aspettare fino a quando è ora."
Dapprima mi sei sembrato avere un'aria molto sorpresa, e poi hai riso di te stesso e mi hai detto:
"Credo sempre di essere a casa mia!"

Extract from *The Little Prince*, chapter 20:
Ma capito' che dopo aver camminato a lungo attraverso le sabbie, le rocce e le nevi, il piccolo principe alla fine arrivò su una strada. E tutte le strade portavano verso gli uomini.
"Buon giorno", disse.
Era in piedi davanti a un giardino fiorito di rose.
"Buon giorno", dissero le rose.
Il piccolo principe le guardo'. Assomigliavano tutte al suo fiore.
"Chi siete?" domando' loro stupefatto.
"Siamo delle rose", dissero le rose.
"Ah!" fece il piccolo principe.
E si senti' molto infelice. Il suo fiore gli aveva raccontato che era il solo della sua specie in tutto l'universo. Ed ecco che ce n'erano cinquemila, tutte simili, in un solo giardino.

Answer sheets created for the third and the fourth experiments (perception studies).

1. Listen to the following sentences. You will hear several people repeating the same sentence: 'Come, let us go look at the sunset now'. How inviting does this proposal sound to you? Discriminate among invitations that sound more or less appealing.
Example of NOT INTERESTING proposal
Example of INTERESTING proposal

	Not interesting				Interesting
Speaker 1	○	○	○	○	○
Speaker 2	○	○	○	○	○
Speaker 3	○	○	○	○	○
Speaker 4	○	○	○	○	○
Speaker 5	○	○	○	○	○
Speaker 6	○	○	○	○	○
Speaker 7	○	○	○	○	○
Speaker 8	○	○	○	○	○
Speaker 9	○	○	○	○	○
Speaker 10	○	○	○	○	○

2. Well done! Now listen to another question: 'Do you wanna come for dinner?'. Some speakers are truly inviting you for dinner, other speakers just want to be polite. You should find out how true their invitations are by deciding whether the speakers are excited or annoyed about having dinner with you.
Example of NOT EXCITED invitation
Example of EXCITED invitation

	Not excited				Excited
Speaker 1	○	○	○	○	○
Speaker 2	○	○	○	○	○
Speaker 3	○	○	○	○	○
Speaker 4	○	○	○	○	○
Speaker 5	○	○	○	○	○
Speaker 6	○	○	○	○	○
Speaker 7	○	○	○	○	○
Speaker 8	○	○	○	○	○
Speaker 9	○	○	○	○	○
Speaker 10	○	○	○	○	○

3. Listen to this sentence: 'I hope I can see you on Monday'. Is the speaker really looking forward to seeing you on Monday? How credible is this sentence? Rate the sentences on a credibility-based scale.

Example of NOT CREDIBLE sentence

Example of CREDIBLE sentence

	Not credible				Credible
Speaker 1	○	○	○	○	○
Speaker 2	○	○	○	○	○
Speaker 3	○	○	○	○	○
Speaker 4	○	○	○	○	○
Speaker 5	○	○	○	○	○
Speaker 6	○	○	○	○	○
Speaker 7	○	○	○	○	○
Speaker 8	○	○	○	○	○
Speaker 9	○	○	○	○	○
Speaker 10	○	○	○	○	○

The perception test is available online at:
http://www.surveymonkey.com/s/5VH38T

Please write your age, birthplace and mother tongue in the yellow box and read the following instructions:

Age:
Birthplace:
Mother tongue:

Instructions:

1. You will hear several repetitions of the same questions, pronounced in different ways. The questions are:
   - *Do you need any money?*
   - *Do you wanna come for dinner?*
   - *Can you open the door?*
2. You should find out how **FRIENDLY** these questions are by considering how the speaker pronounced them. Pay attention to the tone of voice of the speaker. Listen to the following examples.

   FRIENDLY example

   UNFRIENDLY example
3. Put a cross, (**X**), to rate every sentence on a scale from 1 (unfriendly) to 5 (friendly).

Part 1

		Unfriendly				**Friendly**
		1	**2**	**3**	**4**	**5**
Sent.	1					
Sent.	2					
Sent.	3					
Sent.	4					
Sent.	5					
Sent.	6					
Sent.	7					
Sent.	8					
Sent.	9					
Sent.	10					
Sent.	11					
Sent.	12					
Sent.	13					
Sent.	14					
Sent.	15					
Sent.	16					
Sent.	17					
Sent.	18					

Now save this file as a .doc document on your desktop/folder. Take a short break if you like. When you are ready, continue with the second part.

Part 2

		Unfriendly				Friendly
		1	2	3	4	5
Sent.	1					
Sent.	2					
Sent.	3					
Sent.	4					
Sent.	5					
Sent.	6					
Sent.	7					
Sent.	8					
Sent.	9					
Sent.	10					
Sent.	11					
Sent.	12					
Sent.	13					
Sent.	14					
Sent.	15					
Sent.	16					
Sent.	17					
Sent.	18					

Remember to save this document after you are done with the second part!

Thank you for your precious collaboration

Appendix C:

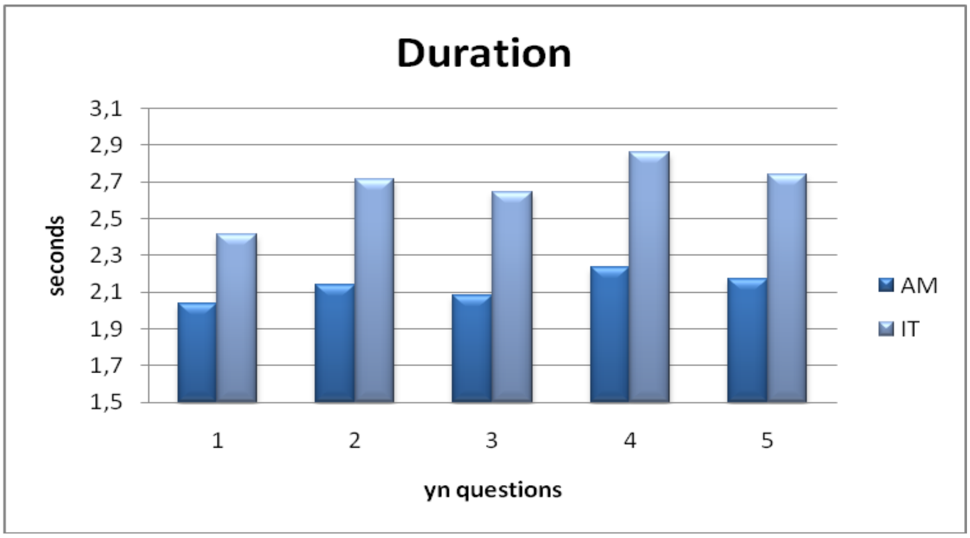


Fig. 55. Duration in yes-no questions.

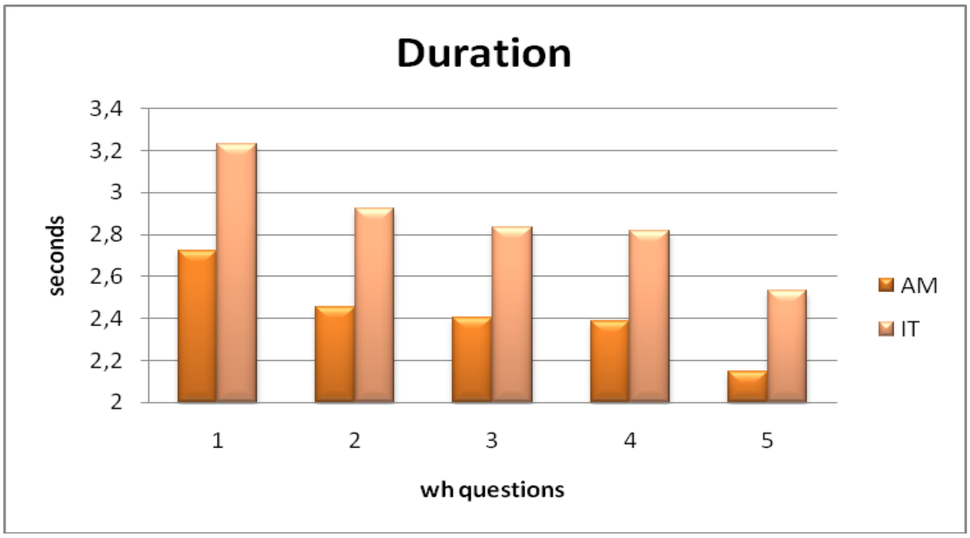


Fig. 56. Duration in wh-questions.

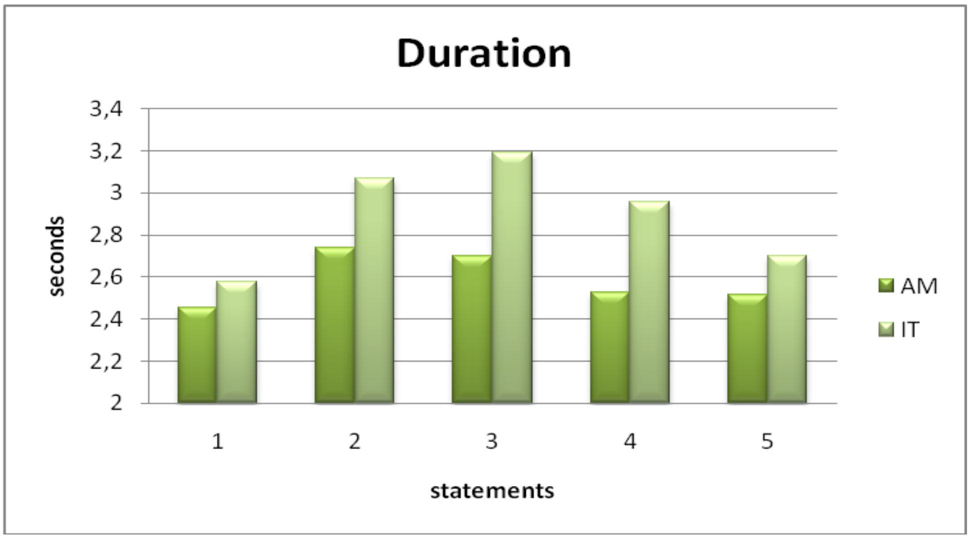


Fig. 57. Duration in statements.

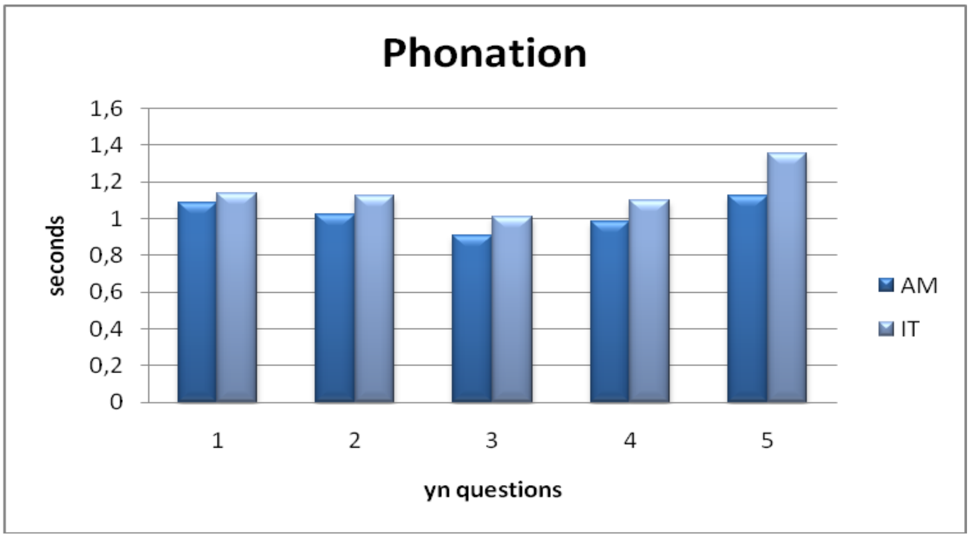


Fig. 58. Phonation in yes-no questions.

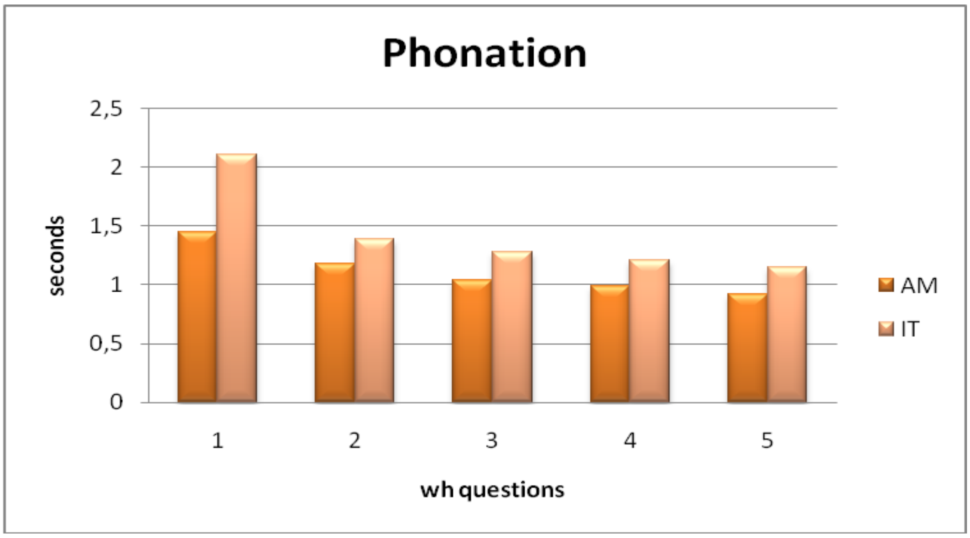


Fig. 59. Phonation in wh-questions.

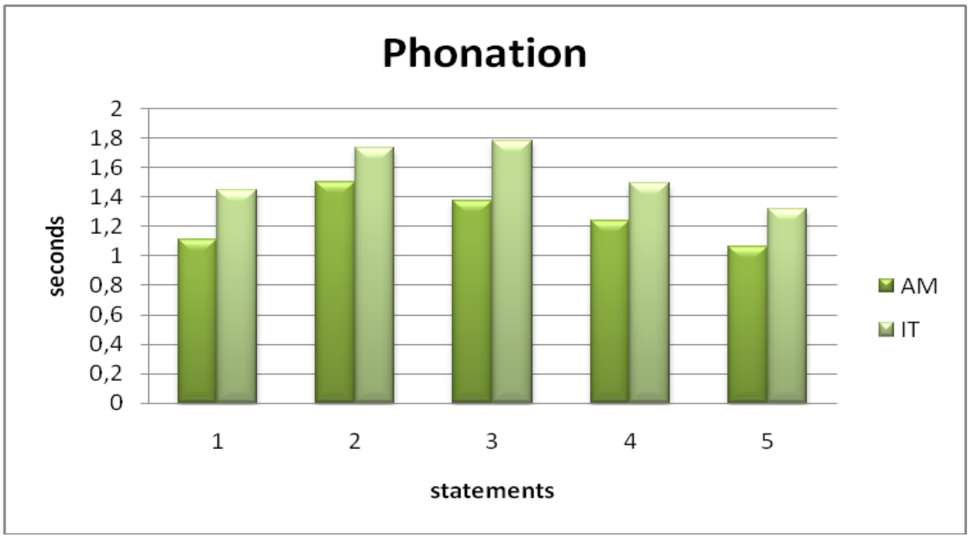


Fig. 60. Phonation in statements.

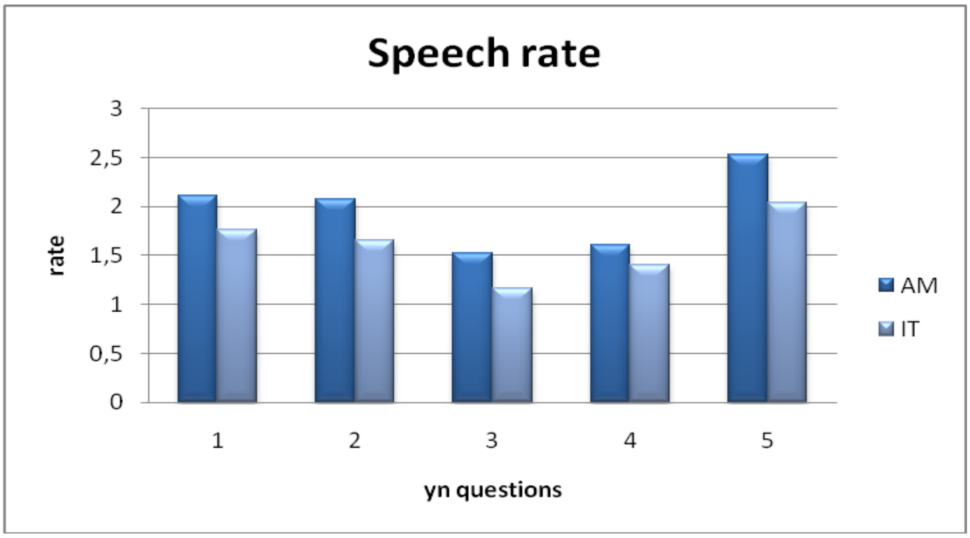


Fig. 61. Speech rate in yes-no questions.

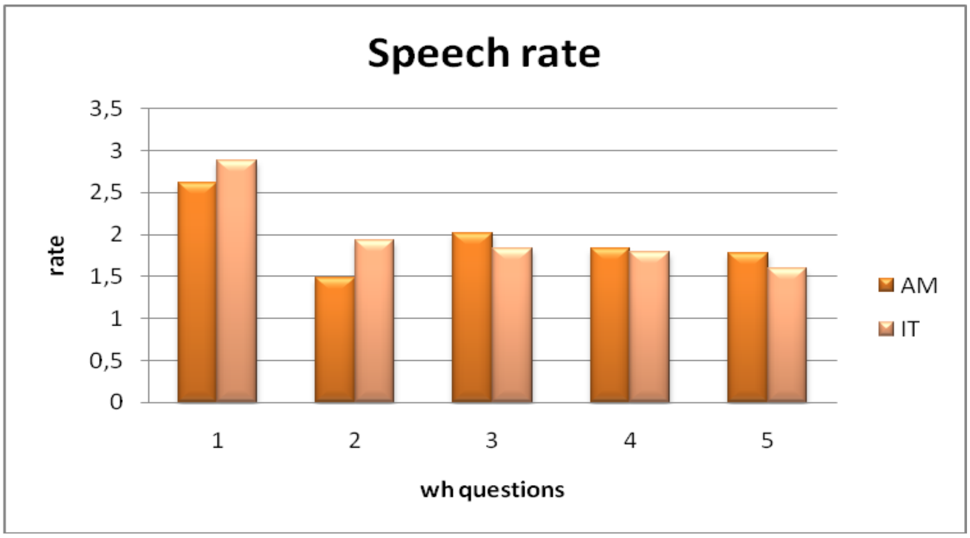


Fig. 62. Speech rate in wh-questions.

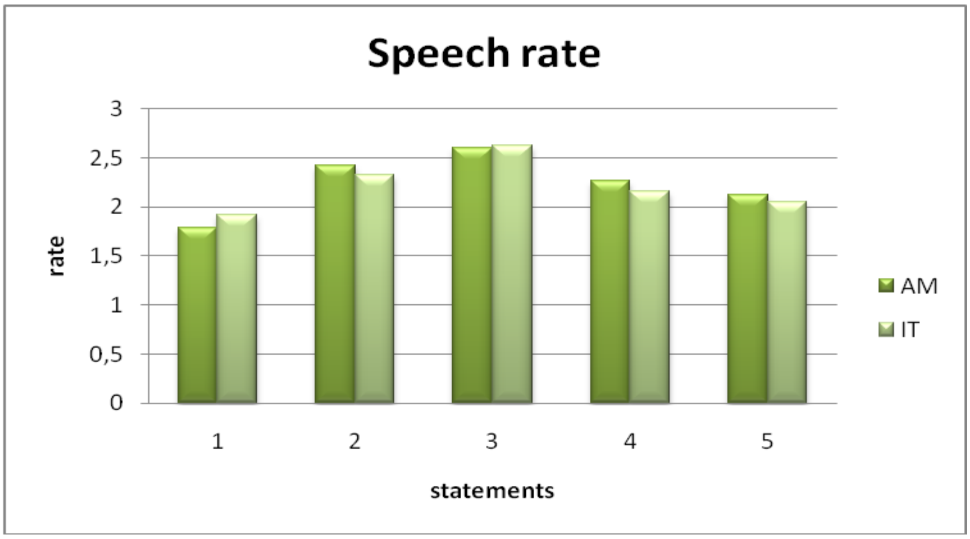


Fig. 63. Speech.rate in statements.

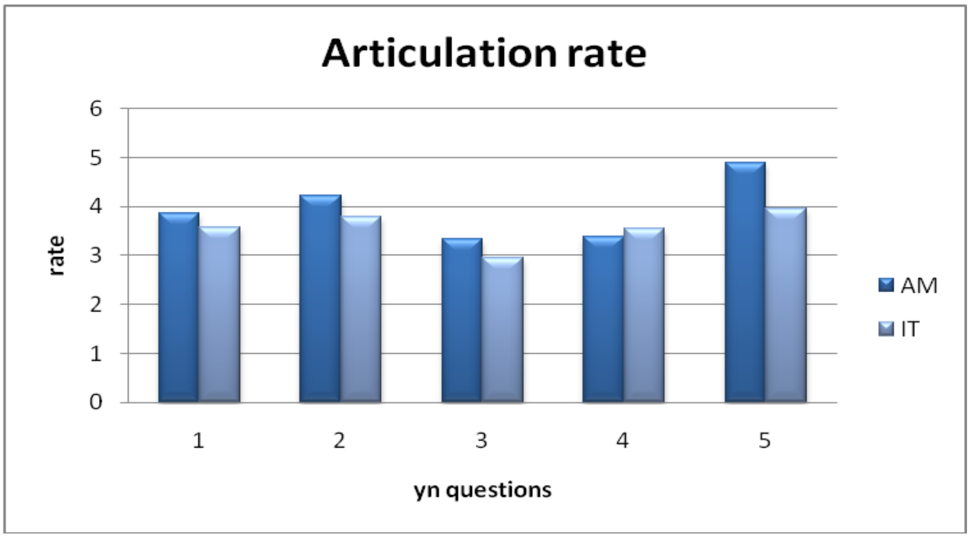


Fig. 64. Articulation rate in yes-no questions.

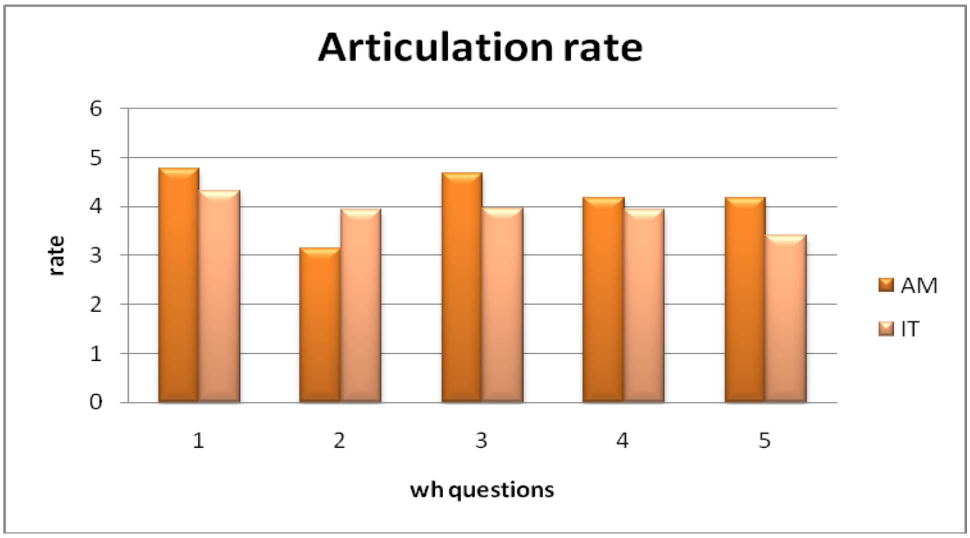


Fig. 65. Articulation rate in wh-questions.

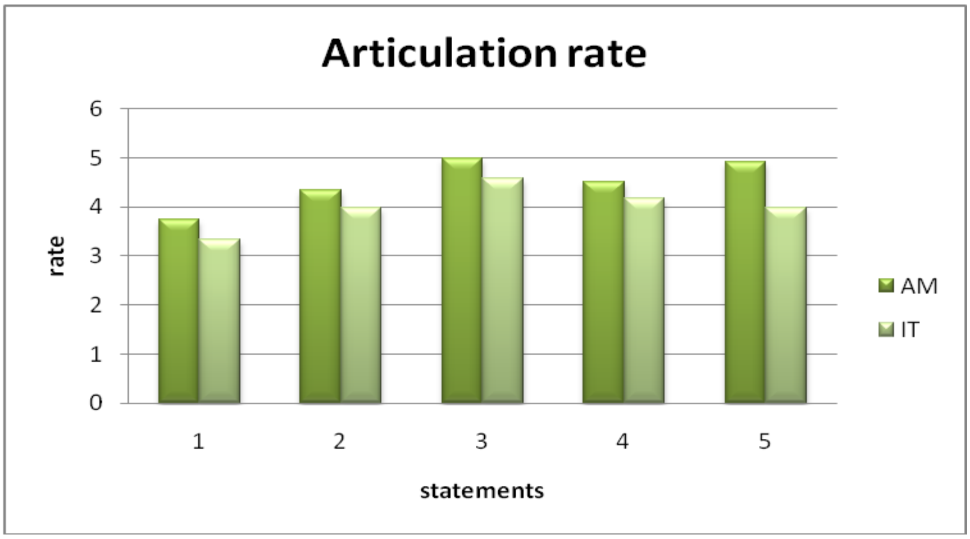


Fig. 66. Articulation rate in statements.

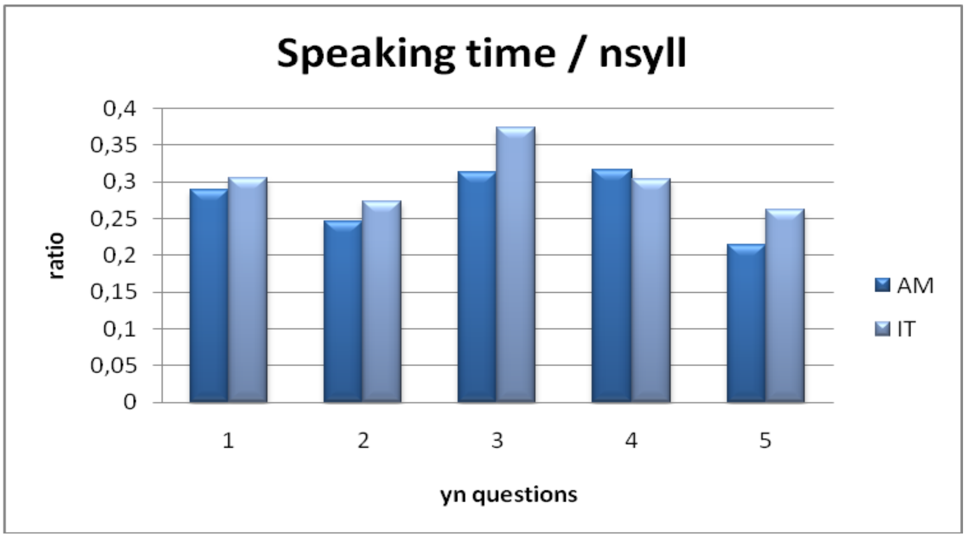


Fig. 67. Speaking time divided for number of syllables in yes-no questions.

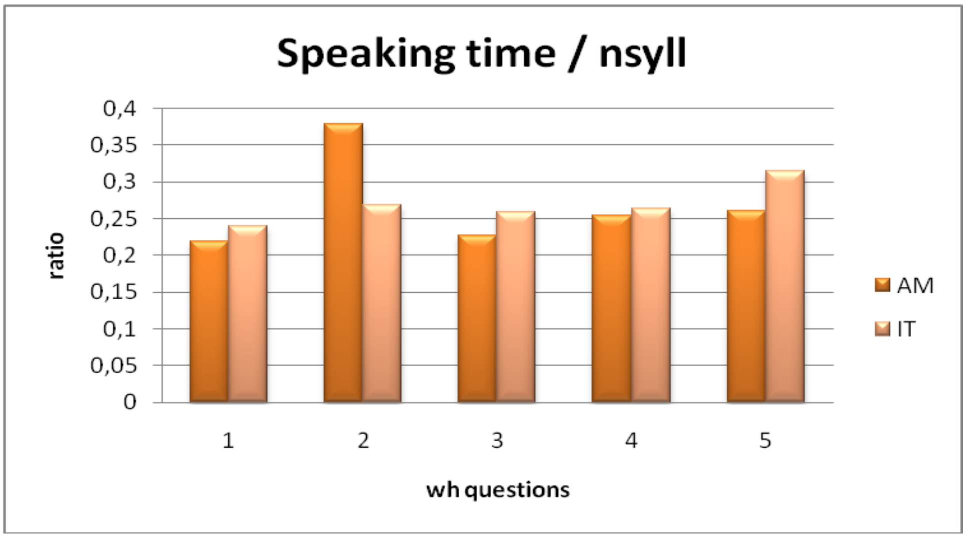


Fig. 68. Speaking time divided for number of syllables in wh-questions.

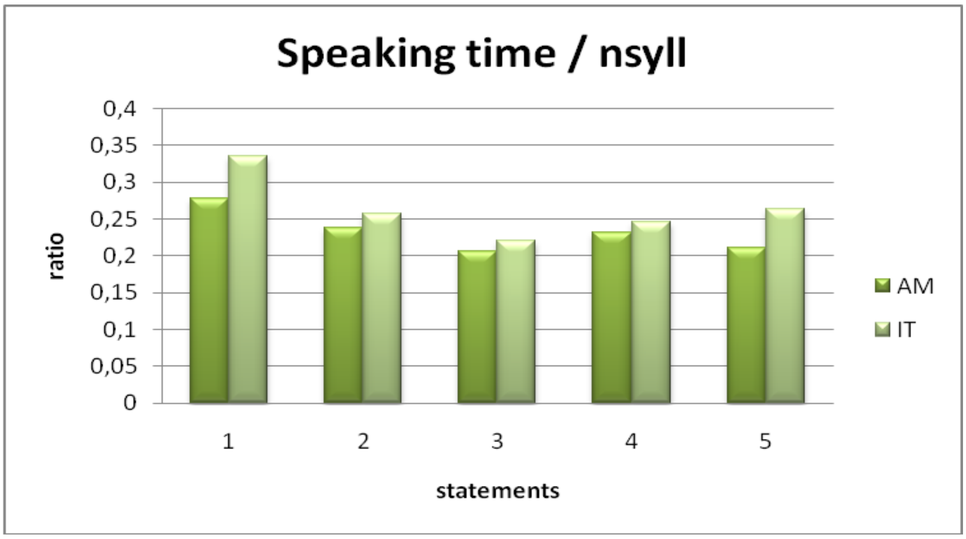


Fig. 69. Speaking time divided for number of syllables in statements.