La comunicazione parlata

I

La comunicazione parlata

Il GSCP (Gruppo di Studio sulla Comunicazione Parlata) è uno dei tre gruppi di interesse della SLI (Società di Linguistica Italiana), nato nel 2003 dalla constatazione del crescente interesse per il parlato, divenuto punto di intersezione di aree di studio diverse. Esempi di questo interesse e della centralità del tema della comunicazione parlata sono l'attenzione per il soggetto parlante e per le basi biologiche della comunicazione, lo studio delle strategie messe in atto dai parlanti nelle interazioni dialogiche, gli studi sui disturbi del linguaggio, quelli sulla fenomenologia fonica di emozioni e sentimenti e sugli atti linguistici, le applicazioni relative al trattamento automatico del parlato e le attività per la conseguente predisposizione di strumenti (banche dati, sistemi di analisi, di sintesi e di codifica automatiche), l'emergere di nuove professioni (periti fonici giudiziari, esperti in ingegneria linguistica, linguisti computazionali) e l'espandersi, nell'accademia e nella società, di professioni classiche, come foniatri e logopedisti.

Il Gruppo promuove e coordina gli studi sulla comunicazione parlata, favorendo la collaborazione e lo scambio tra studiosi di ogni orientamento e specializzazione. La collana La comunicazione parlata ospita i risultati delle attività scientifiche promosse in seno al Gruppo di Studio.

SLI – GSCP International Conference

Naples 13–15 June 2016. Atti del convegno

a cura di

Anna De Meo
Francesca Maria Dovetto

Contributi di

Albano Leoni Federico, Alfano Iolanda, Anastasio Simona
Bianchi Valentina, Bongelli Ramona, Riccioni Ilaria
Zuczkowski Andrzei, Bonvino Elisabetta, Cortes Velasquez Diego
Fiorenza Elisa, Bossaglia Giulia, Mello Heliana
Bullio Paula Cristina, Scarpa Ester, Busà Maria Grazia
Bugnerotto Sara, Castagneto Marina, Patrucco Anna Maria
Sidraschi Diego, Cervoni Valerio, Pietrandrea Paola
Chiusaroli Francesca, Cresti Emanuela, D'Avanzo Stefania
De Gregoris Gregorio, Di Salvo Margherita, Diadori Pierangela
Giuliano Patrizia, Lima e Silva Luis Filipe, Mello Heliana
Negri Lìgia, Pagliaro Anna Chiara, Rasulo Margaret
Russo Rosa, Sorianello Patrizia, Gurrado Glenda
Tommaso Raso, Frederico A. Cavalcante, Maryualè M. Mittmann
Vernillo Paola, Panunzi Alessandro

Aracne editrice

www.aracneeditrice.it
info@aracneeditrice.it

www.gioacchinoonoratieditore.it
info@gioacchinoonoratieditore.it

via Vittorio Veneto, 20
00020 Canterano (RM)
(06) 45551463

ISBN 978-88-255-xxxx-x

I edizione: luglio 2018

Indice

SLI – GSCP International Conference
ISBN 978-88-255-xxxx-x
DOI 10.4399/97888255xxxxx9
pag. 153–171 (luglio 2018)

Being communicative

Does it mean using wider pitch range and more gestures?

Maria Grazia Busà, Sara Brugnerotto*

Introduction: Gestures and Prosody

Research has shown that gestures and speech are interconnected. Gestures appear to develop together with speech, and be synchronous and co–expressive with speech at the semantic, pragmatic, and phonological levels (Kendon, 2004; McNeill, 2005). In fact, it is suggested that gestures share the same system that generate both thought and speech (Kita *et al.* 2017).

There is general consensus that gestures are temporally aligned with prominent parts of speech (e.g., Birdwhistell, 1970; Kendon, 1980, 2004; Bull & Connelly, 1985; Loehr, 2004; Esteve–Giberta & Prieto, 2013). However, the precise nature of this temporal relationship is far from clear. Studies have provided contradictory evidence as to whether gestures and prominence are coordinated at the level of the focused word (Butterworth & Beattie, 1978; Roustan & Dohen, 2010), the lexically stressed syllable (Loehr, 2004, 2007; Rochet–Capellan *et al.* 2008), or the syllable with an intonation peak (De Ruiter, 1998; Nobe, 1996).

Different types of gestures seem to have different effects on the production and perception of prosodic prominence. For example, beats (that is, sudden, baton–like, up and down, or back and forth movements that tend to reflect discourse structure by marking important words and phrases) have been shown to increase the acoustic duration, energy, and values of F_0 and F_1 of the pitch accent of the words co–produced with them; beat production also increases the perceptual prominence of the co–produced words, so that beats act as facilitators of the listeners' task in communication (Krahmer & Swerts, 2007).

Little attention has been paid to the relationship between speakers' global pitch range and use of gestures. Anecdotal evidence suggests that the speak-

* Dept. of Linguistic and Literary Studies, University of Padova, Italy; mariagrazia.busa@unipd.it, sara.brugnerotto@studenti.unipd.it.

ers' pitch variation in speech and their use of gestures might be related. For example, monotonic speakers generally do not gesticulate much, while animated speech is often characterized by a considerable amount of gesturing. It is likely in fact that speakers convey paralinguistic meanings both through their voice and their gestures.

In public speaking classes teachers tell students to use a 'lively' voice and to accompany 'purposeful' gestures when delivering a speech, emphasizing the idea that congruence in language, voice and gestures is at the basis of successful communication; it is suggested that while variations in the speaker's pitch range (and intonation, rhythm and volume) help listeners follow the information flow, gestures can contribute to maintaining the listeners' attention by providing them with a visual channel, in addition to the audio channel, that helps them follow what the speaker is saying. However, there is still little empirical evidence to support the methods used in teaching public speaking.

Assessing whether speakers' global pitch range and use of gestures are related is important also in L2 studies. In L2 speech, the speakers' combined use of non–native pitch and gestures (together with the use of non–native language structures) may affect communication. Mennen *et al.* (2008, 2012) found that Southern Standard British English speakers have a higher and more varied pitch range than Northern Standard German speakers. As a result, Germans may sound "bored" or "unfriendly" to British listeners (Gibbon, 1998), and British voices (especially female) may sound "over–excited" (Eckert & Laver, 1994) or even "aggressive" (Gibbon, 1998) to German listeners. Similarly, the use of frequent, broad, full arm, animated gestures during speech may be common and/or accepted in some linguistic communities, but may be considered distracting and cause annoyance to the listener in some other linguistic communities. On the other hand, too little gesturing may put off interlocutors that are used to greater gesturing in conversations (Axtell, 1991; Efron, 1972; Ekman and Friesen, 1969; Graham and Argyle, 1975; Okada and Brosnahan, 1990).

The studies that have investigated the relationship between speakers' global pitch range and gestures have not found that this relationship is significant. Hoetjes *et al.* (2014) compared the speech characteristics of subjects that were asked to describe a tie knot with or without using gestures. The hypothesis was that not being able to gesture increases the cognitive complexity of the speech task, and this affects the speakers' speech by making it less fluent and more monotonic (i.e., characterised by more filled pauses, smaller pitch range, lower intensity, lower speech duration, and lower speech rate). The authors also looked at the number of speakers' attempts (or repetitions), under the assumption that the first attempt is cognitively more difficult than the following ones. The results showed no

effect of ability to gesture on the number of words produced, speech duration, speech rate, production of pauses, and pitch range. The number of attempts had an effect on the number of words produced, speech duration, and production of filled pauses, which all decreased after the first attempt, showing that, with repetitions, speakers become more fluent and use fewer words; but pitch and speech rate did not change significantly.

Instructing subjects to gesture, or alternatively not to gesture, is not an uncommon procedure in gesture studies. This procedure has often been used to investigate various aspects of the co–production of gestures and speech (see review in Parrill *et al.* 2016). However, it is not clear that instructions to gesture (or not gesture) will change speakers' gesture rate or type (Parrill *et al.* 2016). Thus, it is possible that it may not be the best procedure to elicit variations in speakers' global pitch range or other speech characteristics in relation to gestures. Because prosodic variation is used to express paralinguistic meanings, a better way to elicit speakers' variations in pitch might be to encourage speakers to be communicative in their speech, as is actually done in public speaking classes. This is tested in the present investigation.

This paper investigates global pitch range and use of gestures as produced by Italian speakers of English L2 before and after they are instructed to be communicative in their speech. Speech in Italian is also examined for comparison. The main hypothesis is that when speakers want to be communicative in the L2 they will use a more varied pitch range and more gestures. Other speech characteristics in the three datasets are also examined. It is expected that when the subjects repeat the task in English, this will be less cognitively complex, and be characterised by greater fluency, a higher articulation rate and a higher speech rate.

1. Background: the interrelation of language and gestures

Gestures produced with speech, or co–speech gestures, fulfill a number of functions (Galati & Brennan, 2014; Holler & Wilkin, 2009; Holler et al., 2011). Gestures have important cognitive functions, and facilitate thought conceptualization and speaking (Kita, 2000). Recent research suggests that gestures may have a self–oriented function since they are used to activate, manipulate, package and explore spatio–motoric information for speaking and thinking (Kita et. al 2017). The fact that speakers gesture also when they are communicating in non face–to–face situations (e.g., when they are talking over the phone) or when they are visually–impaired is evidence of the cognitive function of gestures (Cohen, 1977; Cohen & Harrison, 1973; de Ruiter, 2000; Iverson & Goldin–Meadow, 1998; Mol et al., 2011).

Gestures also have a communicative function, and be planned and produced with the addressee's needs in mind. Speakers seem to produce more and larger gestures when they see their interlocutor(s) than when they do not. Also, speakers may provide more accurate descriptions of objects or figures when they are allowed to gesture than when they are required to keep their arms folded (Graham & Argyle, 1975; Riseborough, 1981; Rogers, 1978). However, even when gestures have a communicative function, a self–oriented, cognitive function cannot be excluded (Kita *et al.* 2017).

Gestures are constrained also by individual and socio–contextual factors. These are still largely unexplored but are likely to have a considerable effect on speakears' behavior, as is true for language.

A number of factors seem to affect the use and extent of gestures in L2 speakers' language and speech. Since gestures play an important role in facilitating language access in L1 speech production, L2 speakers may gesture more than monolingual speakers, due to the cognitive complexity that speaking more than one language requires. For example, gestures would help L2 speakers in lexical access retrieval (Kita, 2000; Nicoladis, 2007; Pika et al., 2006). However, gestures are also used by L2 speakers to enhance spoken communication. L2 speakers' gestures seem to be both quantitatively and qualitatively different than in the L1. For example, L2 speakers produce gestures that are more elaborate, and larger when they are speaking to an interlocutor who is present and visible (Bavelas *et al.* 2008). Also, L2 speakers produce more iconic gestures (i.e., gestures that, by their form, depict some feature or event being described), especially when the interlocutor is visible (Morett *et al.* 2012).

Like other aspects of language and speech, gestures seem to be transferred from the L1 to the L2 in the process of L2 acquisition (Gullberg, 2006; Nicoladis, 2007; Pika et al., 2006), though in fact the relationship between the L1 and the L2 could be bidirectional, rather than unidirectional (Brown, 2008; Brown & Gullberg, 2008). It is less clear whether language–specific gesture rates are also transferred (Nicoladis, 2007). In the transfer of gestures from the L1 to the L2, variations can be accounted for individual factors, language proficiency level, task complexity, speakers' expressivity, and affective state (i.e., speakers' performance anxiety or nervousness to speak in a particular situation) (Nicoladis et al., 2007; Parrill *et al.* 2016).

In both L1 and L2, speakers' use and extent of co–speech gestures can be affected by the information that the addressees are assumed to know already, what is referred to as *common ground* (Clark, 1996). Common ground is known to affect verbal language; for example, when speakers make reference to information that is shared with the interlocutor they tend to use fewer words and less informative utterances (Clark & Wilkes–Gibbs, 1986; Isaacs & Clark, 1987; Fussell & Krauss, 1992). The effects of common

ground on co–speech gestures are less clear. Some research has shown that assuming common ground with the interlocutors makes speakers use more interactive gestures during speech, possibly as a way to mark the mutually shared knowledge (Holler & Wilkin, 2009; Holler et al., 2011). Other studies have shown that common ground leads speakers to use fewer, smaller, less precise, and less informative gestures than when no common ground can be assumed, producing the same effects on gestures that it does on speech, that is, a production of gestures that is reduced (Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Jacobs & Garnham, 2007).

2. Measures of global pitch range and coding of gestures

In most languages variations of the fundamental frequency (or F_0) of the human voice are used to create meaning and emphasis. The range over which these variations may occur is called pitch (or F_0) range. Typically, a voice that is heavily inflected, or has a wide pitch range, will sound animated; a voice that has a narrow pitch range will sound monotonic. Thus, pitch range has been used as a measure of speaker's perceived liveliness (Hinks, 2004, 2005; Mennen et. al. 2008) –though the use and interpretation of pitch range may vary depending on language (Graham, 2013; Mennen et. al. 2008, 2012) and sociocultural/ sociophonetic factors (van Bezooijen, 1995).

L2 speech may be characterised by limited pitch variation and a narrower pitch range than L1 speech (Aoyama & Guion, 2007; Hinks, 2004, 2005; Mennen et. al. 2008, 2012; Pickering, 2004; Traunmüller & Eriksson, 1995; Ullakonoja, 2007). In fact, it is possible that prosodic information is processed differently by native and non–native speakers because of their different levels of competence in the L1/L2. For example, as suggested by Jenkins (2002), non–native speakers may rely more on segmental, as opposed to prosodic, information to get their meanings across, given that they lack the amount of extra–linguistic knowledge that native speakers can rely on when communicating. Differences in pitch range in L1 and L2 may also be more conspicuous in certain speaking styles, such as formal presentations (Hinks, 2004, 2005; Johns–Lewis, 1986), during which non–native speakers may be particularly focussed on getting their meanings across, at the expense of prosody.

A framework for measuring global pitch range cross–linguistically was first established by Ladd (1996), then elaborated by Patterson (2000), and further by Mennen et al. (2008, 2012). Within this framework, a number of measures are used to quantify differences in pitch range, including F_0Max, F_0Min, F_0Mean. Hincks (2004, 2005) found that a value of pitch variation, which she called pitch variation quotient (PVQ) and is obtained by normal-

izing the speakers' pitch dividing the standard deviation of all F_0 values in an utterance by their mean value, strongly correlates with speakers' perceived liveliness, though only weakly with speakers' proficiency.

Perceived liveliness cannot be dissociated from perceived fluency. For this reason, it appears important to also investigate the characteristics of speakers' rhythm, such as pauses, durations, articulation rate and speech rate.

With regard to gestures, different taxonomies exist that can be used for coding co–speech gestures. These taxonomies reflect the different ways in which co–speech gestures have been described by various authors. A widely accepted taxonomy, based on McNeill (1992, 2005), distinguishes gestures belonging to *ideation processes* (propositional gestures, representing linguistic referents: iconics, metaphorics, and deictics) and gestures characterizing *discursive activity* (non–propositional gestures: cohesive and beats). To the first type belong gestures that enact or metaphorically represent action, motion or shape, or that indicate location or trajectory. These are also called *representational* gestures. To the second type, *discursive* gestures, belong all gestures that regulate speech, but have no clear meaning attached to. Finally *emblems* refer to conventional and cultural gestures that are meaningful without speech. According to Kita *et al.* (2017) representational gestures are mainly self–oriented. However, in speech, the cognitive and communicative functions of gestures are not clear–cut and not mutually exclusive: « indeed, the very same gestures that contribute to activating, packaging and exploring spatio–motoric information may communicate such information to others » (*ibidem*, 17).

3. Experiment

An experiment was devised to compare global pitch range and use of gestures in the productions, in Italian and in English (twice), of 8 Italian speakers of English as a second language.

Three hypothesis are tested: (1) When performing a task in the L2, the cognitive effort required by speaking the L2 will show up as more narrow pitch, less fluent speech and increased gesturing in the L2 as compared to the L1; (2) If the subjects repeat the task in English, this repetition will decrease the cognitive effort required by the speakers to perform the task; this will show up as more fluent speech and decreased gesturing with respect to the first repetition; (3) If, at the repetition of the task in English, speakers are instructed to be communicative, this will have an effect on their pitch and gestures, which will be greater and more varied than in the first repetition.

3.1. *Subjects and Materials*

8 native Italian speakers took part in the experiment. All were English L2 learners (B2 level of the CEFR), participating in a public–speaking class, master–degree level, taught by the first author. All subjects were female, mean age 23.55, speakers of Italian L1 and students at the University of Padova, Italy.

The subjects were instructed to read Aesop's fable "The Fox and The Crow" to themselves, and then tell the fable to the class first in Italian and then in English. A week later, they told the story in English again, but this time they were instructed to try to be as communicative as possible when delivering the story. All students were naive to the purpose of the experiment. The experiment took place at the beginning of the course so the students were not yet familiar with common techniques use in public speaking. Both times, the speakers were video–recorded by the teacher. Each recording lasted about 90–120 seconds.

The corpus consists of three data sets, one in Italian –produced at time1, which will be referred to as Italian t1; one in English –produced at time1, which will be referred to as English t1; and another one in English –produced at time2, which will be referred to as English t2.

3.2. *Methods and Analysis*

The analysis was carried out in three main phases: a first accurate visual examination of the audiovisual material focussing on the subjects' gestural behavior; a prosodic analysis; and a gesture analysis.

The audio signal was extracted from the videos using the AVC software (at http://www.any\T1\textendashvideo\T1\textendashconverter.com/) and then imported in Praat (www.praat.org). Pitch was measured setting the pitch floor to 75 Hz, and the ceiling to 500 Hz, since all the speakers were female. Praat textgrids were used to annotate the following:

a) *words*, i.e, the orthographic transcription of each pronounced word;
b) *respiratory pauses,* i.e., breaths;
c) *silent pauses,* i.e., all silences longer than 100ms in duration. [These are not used for the physical need to inhale (unlike respiratory pauses), but rather have a grammatical function. However, they may also occur when speakers are unsure of what they have to say, when they hesitate to say something, or for lexical retrieval, especially in the case of L2 speakers];

d) *filled pauses* are all those vocalizations or vocal phenomena that have no semantic meaning such as 'uhm', 'ehm', 'mmm', but signal the length of the delay of upcoming speech (Clark & Fox Tree, 2002);
e) *disfluencies,* i.e., any vocal interruption of speech that does not add any propositional or semantic content to the utterance.

Disfluencies were subcategorised in *repetitions* and *corrections*. *Repetitions* are reiterations of the word/s the speaker just said. In spontaneous speech, repetitions generally involve a first instance of the repeated word, a possible silent pause, a second instance of the repeated word, and the continuation of the utterance (Rangarajan, & Shrikanth 2006). *Corrections* are all verbal mistakes, that is, every time the speakers stop to correct a word they considered wrong or inappropriate. Corrections involve a wrong word, a possible silent pause and/or a repetition, followed by the correct word. For the purpose of the total word count repetitions and corrections were counted as words.

The following measures were taken. Total duration of speech; Number and duration of pauses (silent and respiratory); Duration of articulated speech (Tot.Dur.Speech–Tot.Dur.Pauses); Articulation rate (n. words/duration articulated speech); Speech rate (n.words/total duration speech) (Pettorino 2003); Total number of words. For pitch range: F_0Min, F_0Max, F_0Mean, St.Dev. (Ladd 1996; Patterson 2000; Mennen *et al.* 2008, 2012). To calculate the PVQ, following a procedure indicated in Hinks (2005), the pitch listings were extracted from each audio file, the outliers were removed, mean and standard deviation were calculated, and the data were normalized dividing the standard deviation of F_0 by the mean (PVQ = St.Dev.F_0/F_0Mean). The statistical significance of the results was tested with within–subject ANOVAs, unless specified otherwise.

The last phase of the data analysis was carried out with Elan (https://tla.mpi.nl/tools/tla-tools/elan/). Each co–speech gesture in the three data sets (Italian t1, English t1 and English t2) was annotated. The gestures were divided into three main categories: *depictive, discursive* and *emblems* (broadly following McNeill (2005)'s distinction (see 3 above)). In the first type, depictive, the authors included iconic, metaphorical and deictic gestures. In the second type, discursive, the authors included beats and any other gesture that has no clear iconic or metaphorial meaning.

The speakers' gestures were counted by category, and gesture rate was calculated with the formula Total N.gestures/Total N.words * 100 (Nicoladis et al., 2007). This was done to relate gesture rate with speech rate, as well as to allow a comparison of the number of gestures in the first and second repetition.

4. Results

4.1. *Pitch*

Table 1 shows the results of the pitch values for the three datasets, averaged across the subjects. In Italian t1 and English t1 the values are similar. In English t2 the subjects show a higher Mean pitch, higher Max pitch, and higher standard deviation, indicating that F_0 is much more varied here that in the other two datasets.

Tabella 1: Pitch values in the three datasets.

	Italian t1	English t1	English t2
Mean pitch	201.30	199.50	228.22
Min pitch	78.15	79.96	76.7
Max pitch	445.5	432.61	475.14
ST DEV	42.70	42.96	57.01

Figure 1 shows the PVQ data for the eight speakers in the tree datasets. The data confirms what is shown in Table 1, that is, all speakers varied their pitch more in English t2 than in English t1 or Italian t1. For all speakers the PVQ of Italian is comparable to the PVQ of English 1, showing that the speakers did not use a very varied pitch in Italian and in their their first repetition in English; the speakers' pitch is more varied when they repeat the story in English the second time, when they were asked to be more communicative.

At a between–subject ANOVA test, the difference in PVQ values in the three datasets was highly significant: $F(2, 779) = 49.96$, $p < .0001$. The results of a t–test showed that the dataset that is significantly different from the others is English t2.

4.2. *Speech data*

Table 2 shows the data relating to the speakers' speech characteristics in the three datasets.

The Italian t1 dataset has the shortest total duration, English t1 has the longest duration, and English t2 has values between the other two datasets, but these differences are not significant.

The speakers make fewer pauses in Italian t1 than in English t1 and English t2. There is a significant difference between the Italian t1 data and both the English t1 and English t2 data ($F(2, 21) = 49.96$, $p < .0001$). There is no significant difference between the English t1 and the English t2 datasets.

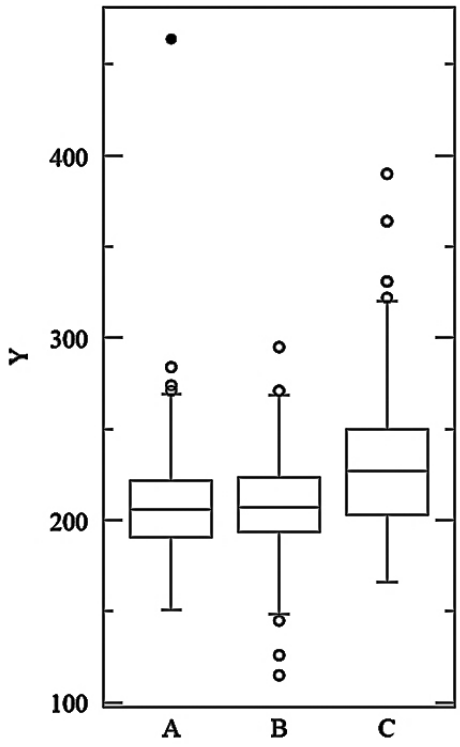

Figura 1: Pitch variation quotient (PVQ) in the three datasets.

The speakers also make significantly shorter pauses in Italian t1 than in English t1 ($F(1,7) = 14.189$, $p < 0.01$) and English t2 ($F(1,7) = 10.732$, $p < 0.05$), while the difference in the duration of pauses between English t1 and English t2 is not significant. These results show that, as expected, the subjects are more fluent in Italian than in either English t1 and English t2; the second repetition in English does not significantly change the subjects' level of speech fluency, suggesting that fluency is related to speakers' proficiency.

The English t1 dataset has the smallest number of words, and English t2 has the greatest number of words, Italian t1 having values in between. The difference in number of words is significant only in the case of the English t1 and English t2 datasets ($F(1,7) = 12.699$, $p < 0.01$).

Figure 2 shows the relative ratio of pauses and spoken words in the three datasets: In Italian t1 the speakers produce longer stretches of speech before they make a pause; in English t1 the stretches of speech between pauses are shortest; in English t2 the stretches of speech are slightly longer but still shorter than in Italian.

The speakers also differ in the duration of articulated speech (Tot. Dur. Speech–Tot. Dur. Pauses), with the Italian t1 dataset having lower values than the English t2 data and, more so, of the English t1 data. These differences are, however, not significant.

Tabella 2: Speech data in the three datasets.

	Italian t1	English t1	English t2
Total dur. speech (ms)	84.93	96.23	92.39
Total n. pauses (silent, filled and resp.)	23.25	35.75	37.5
Total dur. of pauses (silent, filled and resp.)	9.14	12.86	13.33
Total dur. articulated speech (ms)	75.78	83.36	79.05
N. of words	208.62	189.75	222.37
Articulation rate	2.75	2.27	2.81
Speech rate	2.45	1.97	2.40

For articulation rate, English t1 has lower values than both Italian t1 and English t2. There is a significant difference between Italian t1 and English t1 ($F(1,7) = 6.413$, $p < 0.05$), and between English t1 and English t2 ($F(1,7) = 15.48$, $p < 0.01$), while Italian t1 and English t2 have similar values (the difference in values is not significant).

Similarly, for speech rate, English t1 has lower values than both Italian t1 and English t2. There is a significant difference between Italian t1 and English t1 ($F(1,7) = 10.269$, $p < 0.05$), and between English t1 and English t2 ($F(1,7) = 12.436$, $p < 0.01$), while the difference between Italian t1 and English t2 is not significant.

These data show that in the first repetition in English the subjects speak at significantly slower articulation and speech rates than in their native language, but with the second repetition in English they reach values that are not too dissimilar from the native ones.

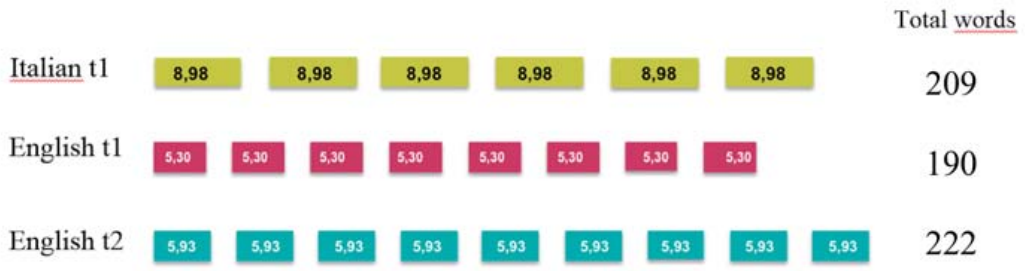

Figura 2: Comparison of the frequency of coccurrence of pauses in relation to spoken words in the three datasets.

4.3. *Occurrence of gesture types*

Figure 3 shows the occurrence of each gesture category in the three different datasets. The data shows that there is an overall higher number of gestures

in English t2 than in the other two datasets. Also, out of all three categories of gestures, depictive gestures increase more than the other two categories.

For the depictive gestures, the Italian t1 dataset does not differ significantly from the English t1, but it does differ significantly from English t2 ($F(1,7) = 8.607$, $p < 0.05$). Also, there is a significant difference between the English t1 and the English t2 datasets ($F(1,7) = 17.775$, $p < 0.01$).

As for the discursive gestures, the Italian t1 dataset does not differ significantly from either the English t1 or the English t2 dataset, nor is there a significant difference between the English t1 and the English t2 datasets.

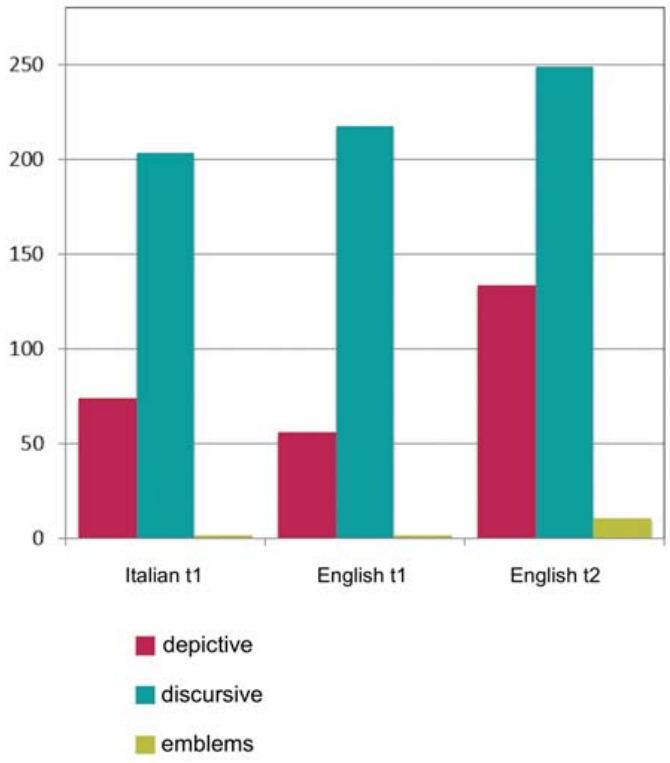

Figura 3: Occurrrence of Gesture Types.

4.4. *Relation of pitch and gestures*

Figure 4 compares the PVQ values and the Gesture Rate values for the three datasets. The PVQ data show that the PVQ value is similar for Italian t1 and English t1, and rises sharply from English t1 to English t2. The pattern is very similar for gesture rate. This shows a clear effect for the English t2 dataset.

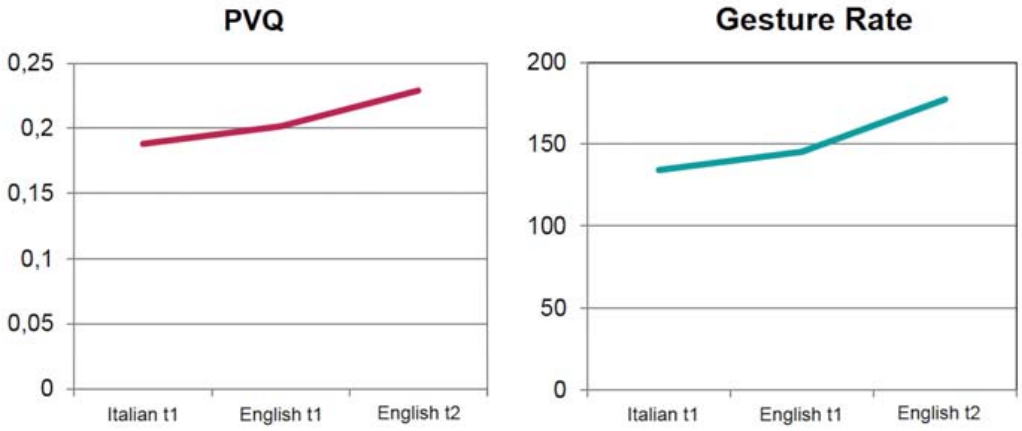

Figura 4: Comparison of PVQ and Gesture rate in the three datasets.

Discussion and Conclusions

The study examined the productions of Italian speakers of English L2 telling a story in Italian and then twice in English at two different times. Object of investigation were the global pitch range, speech characteristics and gestures of the speakers' productions and the effects of the instruction to the subjects to be communicative in the second English repetition.

The subjects show similar pitch range values in their first telling of the story in Italian and English. This result suggests that the speakers transfer their pitch values from Italian to English, and does not provide support to previous findings (reviewed in 3) showing that speakers' pitch in the L2 is narrower than in the L1. It is possible, however, that differences in pitch range between the L1 and the L2 may be more evident when non–native speakers are focussing on getting their meanings across (for example in formal academic presentations) than on the way in which they deliver the content –as reviewed in 3. In the present investigation, the instruction to be communicative (in the second repetition in English) causes speakers to use a significantly more varied, wider, pitch.

The subjects are more fluent in Italian than in English, as expected. Thus, even though the speakers in the second repetition use a more varied pitch, they continue to produce speech that has more and longer pauses than in Italian. This suggests that the speakers' fluency is related to the proficiency in the L2 rather than to the decreased cognitive effort.

The subjects produce a significantly higher number of words in the second repetition in English than in either of the other two datasets. This is somewhat unexpected. It is known that when speakers make reference to information that is shared with the interlocutor they tend to use fewer

words and less informative utterances (see studies on common ground reviewed in 2). Thus, because the students were repeating the task in front of the same audience (their classmates), we would have predicted to find a lower number of words in the second repetition in English than in the first, and not the opposite. A possible interpretation of this result is that, in their attempt to be more communicative, the subjects use more words to add details to the story they are telling.

As for articulation and speech rate, the subjects' articulate and speak more slowly in the first repetition in English than in Italian, possibly due to cognitive and linguistic problems with the task in the L2. However, when repeating the story in English they manage to reach articulation and speech rate similar to those of the native language, showing a significant improvement on their first repetition, probably due to the decreased cognitive effort required for the task.

The subjects gesture more in the second repetition in English than in the other two tasks. If the subjects in the present investigation had used gestures simply with a cognitive function, they would have used most gestures in English t1, when the task was most complex cognitively, and fewer gestures in English t2. Instead, the increase in gestures in the second repetition in English shows that the subjects are using gestures with a communicative function. Interestingly, in the second repetition in English the speakers also use the most varied pitch range, confirming our hypothesis that speakers' communicative intentions cause an increase in both gestures and pitch range.

In gesture studies, a common procedure used to assess variations in gesture rate across individuals is to instruct subjects to gesture or, alternatively, not to gesture. The present study shows that asking subjects to be more communicative in their speech is also effective in bringing about a change in number and type of gestures and in gesture rate. In this investigation, the increase in gestures concerns primarily depictive gestures, confirming previous findings (reviewed in 2) that L2 speakers use more iconic gestures with a communicative function, especially when the interlocutor is in sight. While representational gestures are generally considered to have a self–oriented function (i.e., help speakers in their cognitive processes), the present investigation shows an instance in which these gestures also have a communicative function.

This study has shown that speakers who intend to be more communicative will use a more varied pitch and more gestures to do so. Public speaking classes urge students that to be more communicative and to enhance listeners' comprehension of the public speeches they should use varied intonation and meaningful gestures. However, studies showing the actual effectiveness of this kind of instructions are lacking. The findings

of this study are relevant in this respect. Public speaking instruction also insists on the importance of repetition and rehearsal, which are claimed to be essential for bringing about changes in speakers' speaking and delivery styles. This investigation shows that, in fact, repetition of a task can bring about substantial changes in the speakers' verbal (i.e., words) and non–verbal patterns (i.e., speech and articulation rate, pitch range, gesture type and rate).

The results of this study are relevant also to L2 instruction. Even though students' fluency may be related to speakers' proficiency and thus be hard to improve in the short period of a week's time, repeating a task can significantly lead to a change in learners' habits both at the verbal and non–verbal levels in the L2. Providing students with instructions to be communicative can be particularly effective in bringing about the change.

References

Aoyama, K. & Guion, S. G. (2007). Prosody in second language acquisition: An acoustic analysis on duration and F_0 range. In O.–S. Bohn and M. J. Munro [Eds], The Role of Language Experience in Second–Language Speech Learning. In Honor of James Emil Flege. Amsterdam, John Benjamins, 281–297. [2028?]

Axtell, R.E. (1991). *Gestures – The Dos and Taboos of body language around the world*. New York: John Wiley & Sons.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. Journal of Memory and Language, 58 (2), 495–520.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*, Philadelphia, University of Pennsylvania Press.

Brown, A. (2008). Gesture viewpoint in Japanese and English: Cross–linguistic interactions between two languages in one speaker. *Gesture, 8*(2), 256–276. [2028?]

Brown, A. & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1–L2 encoding of manner in speech and gesture. *Studies in Second Language Acquisition*, *30*(2), 225–251.

Bull, P. & Connelly, G. (1985). Body movement and emphasis in speech", Journal of Nonverbal Behavior, 9: 169–187.

Butterworth, P. & Beatty, G. (1978). Gesture and silence as indicators of planning in speech", in R. Campbell and G.T. Smith [Eds.], Recent Advances in the Psychology of Language: Formal and Experimental Approaches. New York: Plenum Press.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H.H., Fox Tree, J. (2002). "Using uh and um in spontaneous speech", Cognition, 84, 73–1.

Clark, H.H., & Wilkes–Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Cohen, A.A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27*(4), 54–63.

Cohen, A.A., & Harison, R.P. (1973). Intentionality in the use of hand illustrators in face–to–face communication situations. *Journal of Personality and Social Psychology, 28*, 276–279.

de Ruiter, J.P. (1998). *Gesture and speech production*. Unpublished dissertation. University of Nijmegen.

de Ruiter, J.P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture*, 284–311, Cambridge University Press, Cambridge. [2028?]

Eckert, H., & Laver, J. (1994). *Menschen und ihre Stimmen: Aspekte der vokalen Kommunikation*. Weinheim: Psychologie Verlags Union.

Ekman, P., & Friesen, W.V. (1969) The repertoire of nonverbal behavior. *Semiotica* 1, 49–98.

Efron, D. (1972). *Gesture, race, and culture*. Hague: Mouton.

Esteve–Giberta, N. and Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56: 850–864. [2028?]

Fussell, S.R., & Krauss, R.M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*, 378–391.

Galati, A. & Brennan, S. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co–speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435–451.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157–195.

Gibbon, D. (1998). German Intonation. In D. J. Hirst & A. Di Christo (eds.), *Intonation systems: A survey of twenty languages* (pp. 78–95). Cambridge, MA: Cambridge University Press.

Goldin–Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Graham, C. (2013). Revisiting fo range production in Japanese–English simultaneous bilinguals. *UC Berkeley Phonology Lab Annual Report*, 110–125.

Graham, J.A., & Argyle, M. (1975). A cross–cultural study of the communication of extra–verbal meaning by gestures. *International Journal of Psychology, 10*, 57–67.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). International Review of Applied Linguistics, *44*(2), 103–124.[2028?]

Hincks, R. (2004). Processing the prosody of oral presentations. *Proc. InSTIL/ICALL 2004 – NLP and speech technologies in advanced language learning systems,* 63–69.

Hincks, R. (2005). Measuring liveliness in presentation speech. *Proc. INTERSPEECH 2005 – Eurospeech, 9th European Conference on Speech Communication and Technology,* 01.

Holler, J. & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267–289.

Holler, J., Tutton, M. & Wilkin, K. (2011). Co–speech gestures in the process of meaning coordination. *Proceedings of the 2nd Conference on Gesture and Speech in Interaction (GESPIN 2011)*, Bielefeld, Germany.

M Hoetjes, E Krahmer, M Swerts (2014). Speech Communication 57, 257–267.

Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General, 116*, 26–37.

Iverson, J.M. & Goldin–Meadow, S. (1998). Why people gesture when they speak, Nature, *396*, 228.

Jacobs, N. & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291–303.

Jenkins, J. (2002). "A sociolinguistically–based, empirically–researched pronunciation syllabus for English as an International Language", Applied Linguistics, 23(1): 83–103.

Johns–Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns – Lewis, (Ed.), *Intonation in discourse*, Breckenham, Kent: Croom Helm, 199–220.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication*, The Hague: Mouton, 207–227.

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (ed.), *Language and gesture*. Cambridge: Cambridge University Press, 162–185.

Kita, S., Alibali, M. W., & Chu, M. (2017, February 27). How Do Gestures Influence Thinking and Speaking? The Gesture–for–Conceptualization Hypothesis. Psychological Review. Advance online publication. http://dx.doi.org/10.1037/rev0000059.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396–414.

Ladd, D. R. (1996). Intonational phonology, Cambridge, Cambridge University Press.

Loehr, D. P. (2004). Gesture and intonation, Doctoral dissertation, Georgetown University, Dissertation Abstracts International, 65 06, 2180, UMI No. 3137056.

Loehr, D. P. (2007). Aspects of rhythm in gesture and speech, *Gesture*, 7, 179–214.

McNeill, D. (1992). *Hand and Mind*. Chicago: The University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

Mennen, I., Schaeffler, F. & Docherty, G. (2008). 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. (2012). Cross– language differences in fundamental frequency range: A comparison of English and German. *Journ. of the Acoustical Society of America*, *131*(3), 2249–2260. [2028?]

Mol, L., Krahmer, E., Maes, A. & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer–Mediated Communication, 17*(1), 77–100.

Morett, L. M., Gibbs, R. W., & MacWhinney, B. (2012). The role of gesture in second language learning: Communication, acquisition, & retention. Retrieved from http://mindmodeling.org/cogsci2012/papers/0143/paper0143.pdf

Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics, 28*, 441–454.

Nicoladis, E., Pika, S., Yin, H. & Marentette, P.(2007). Gesture use in story recall by Chinese–English bilinguals. *Applied Psycholinguistics 28*(4), 721–735.

Nobe, S. (1996). *Representational Gestures, Cognitive Rhythms, and Acoustic Aspects of Speech: A Network/Threshold Model of Gesture Production*, Doctoral dissertation, University of Chicago, Dissertation Abstracts International, 57 07, 4736. UMI No. 636827.

Okada, T., & Brosnahan, L. (1990). *Japanese and English gesture: Contrastive nonverbal communication*. Tokyo: Taishukan Press.

Parrill, F.,Cabot, J., Kent, H., Chen, K., and Payneau, A. (2016). Do people gesture more when instructed to? *Gesture*, 15, 3, 357–371.

Patterson, D. (2000). A linguistic approach to pitch range modeling, PhD dissertation, Univ. of Edinburgh.

Pettorino, M. (2003). La velocità di articolazione. Atti delle XIV Giornate di Studio del Gruppo di Fonetica Sperimentale–GFS.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional [2028?]

Discourse. English for Specific Purposes, *23*, 19–43.

Pika, S., Nicoladis, E. & Marentette, P. (2006). A cross–cultural study on the use of gestures: Evidence for cross–linguistic transfer? Bilingualism: Language and Cognition, *9*(3).

Rangarajan, V., & Narayanan, S. (2006). Analysis of disfluent repetitions in spontaneous speech recognition. Signal Processing Conference, 2006 14th European. IEEE.

Riseborough, M.G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*, 172–183.

Rochet–Capellan, A., Laboissierre, R., Galvan A. and Schwartz, J. (2008). The speech focus position effect on jaw– finger coordination in a pointing task, *Journal of Speech, Language, and Hearing Research*, 51: 1507– 1521. [2028?]

Rogers, W.T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behaviors within utterances. *Human Communication Research, 5*, 54–62.

Roustan, B. & Dohen, M. (2010). Co–production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus, *Proceedings of Speech Prosody 2010*, 100110, 1–4. Retrieved from www. speechprosody2010.illinois.edu/papers/100110.pdf.

Traunmüller, H. & Eriksson, A. (1995). The perceptual evaluation of *Fo* excursions in speech as evidenced in liveliness estimations, *Journ. of the Acoustical Society of America*, 97, 1905–1915.

Ullakonoja, R., (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). *Proc. 16th ICPhS,* 1701–1704.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women, *Language and Speech*, *38*(3): 253–256.