UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia Dello Sviluppo e Della Socializzazione

SCUOLA DI DOTTORATO DI RICERCA IN PSICOLOGIA
INDIRIZZO SCIENZE COGNITIVE
CICLO XXIII

Segmental and supra-segmental aspects of speech perception in familiar and unfamiliar languages: Influence of the listener's native language as revealed by Event-related potentials (ERPs)

Direttore della Scuola: Ch.ma Prof.ssa Clara Casco

Coordinatore d'indirizzo: Ch.ma Prof.ssa Francesca Peressotti

Supervisore: Ch.ma Prof.ssa Lucia Colombo

Dottoranda: Chizuru Deguchi

31 Gennaio 2011

Segmental and supra-segmental aspects of speech perception in familiar and unfamiliar languages: Influence of the listener's native language as revealed by Event-related potentials (ERPs)

RIASSUNTO

La struttura fonologica varia tra le lingue. I sistemi neurali nel cervello umano che sottostanno la percezione del linguaggio parlato vengono strutturati mediante l'esposizione precoce e la lunga esperienza con la lingua nativa, in modo che la sensibilità a specifici segnali acustici che sono pertinenti in quella lingua si sviluppi in modo ottimale ed estremamente efficiente. Tale processo adattivo di plasticità cerebrale facilita il riconoscimento dei fonemi e l'acquisizione degli indizi prosodici nella lingua nativa. D'altro canto, lo stesso processo aumenta la difficoltà di percepire alcuni segnali linguistici in lingue straniere. La presente tesi presenta degli studi effettuati per indagare l'elaborazione corticale delle componenti segmentali e soprasegmentali del linguaggio parlato utilizzando le tecniche di potenziali evento-relati (ERP).

Nello Studio I, sono stati condotti due esperimenti sulla percezione uditiva delle vocali che appartengono o non appartengono all'inventario fonemico della lingua nativa dei soggetti. La Mismatch Negativity (MMN) è una componente di ERP che riflette la rilevazione pre-attentiva della devianza acustica dello stimolo attuale rispetto alla rappresentazione integrata dello stimolo *standard* nella memoria sensoriale (Näätänen, 2007; Kujala et al., 2007) ed è stata utilizzata anche per sondare le rappresentazioni lingua-specifiche nella memoria a lungo termine (Näätänen, 1997; Cheour et al., 1998; Winkler et al., 1999a). Una sequenza uditiva è stata costruita con quattro vocali (una standard e tre devianti), ciascuna pronunciata da tre parlanti diverse, ed è stata presentata ai soggetti nelle condizioni di ascolto attivo e passivo. Quando tutte le vocali erano distintive nell'inventario della lingua nativa, la MMN è stata generata da ciascuna vocale deviante indipendentemente dalla variabilità delle parlanti, dimostrando la categorizzazione pre-attentiva basata sulle rappresentazioni nella memoria a lungo termine (Esperimento 1). Al contrario, quando una parte delle vocali non apparteneva all'inventario fonemico nativo, la MMN è stata generata da nessuna delle vocali devianti, che indica la mancanza di categorizzazione automatica (Esperimento 2). Inoltre, la variabilità dovuta alle parlanti ha interagito con la categoria vocale nella condizione di ascolto attivo, come risulta dai dati comportamentali e dalle componenti cognitive di ERP, quando il deviante per discriminare è fisicamente simile allo standard (Esperimento 1) oppure non è supportato dall'inventario nella lingua nativa (Esperimento 2), suggerendo che l'analisi acustica consapevole ed altre strategie cognitive sono impiegate per categorizzare i devianti acusticamente difficili o ambigui.

Nello Studio II (Esperimento 3), è stata indagata la percezione dei cambiamenti piccoli e grandi di altezza tonale (*pitch*) immessi nelle frasi parlate nella lingua nativa degli ascoltatori, jabberwocky o in una lingua straniera. Sono state analizzate le prestazioni in un compito di rilevazione delle incongruenze prosodiche e gli ERP generati dai cambiamenti di altezza tonale (una componente negativa precoce e una positività più tardiva). I risultati hanno dimostrato che i soggetti sono più efficienti ad elaborare l'altezza tonale nel contesto frasale della loro lingua nativa che della lingua straniera, mentre i risultati per le frasi in jabberwocky erano intermedi, suggerendo che sia la prosodia familiare che il contesto naturale e semanticamente interpretabile della lingua nativa aiutano gli ascoltatori a rilevare i cambiamenti sottili di altezza tonale.

I risultati globali di questi studi hanno confermato che la lunga esperienza con la lingua nativa ha un impatto sull'elaborazione uditiva nella corteccia fin dagli stadi precoci della percezione del linguaggio parlato, di conseguenza modulando la percezione cosciente degli stimoli parlati in lingue diverse.

ABSTRACT

The phonological structure differs between languages. Human brain structure subserving speech perception is shaped through early exposure to and long-term experience with one's native language, so that the sensitivity to specific acoustic cues that are relevant in that language develops. Such adaptive process of brain plasticity facilitates to recognize phonemes and capture prosodic signals correctly in the native language, whereas it augments difficulty to perceive some linguistic cues in foreign languages. Speech phonology consists of segmental and suprasegmental components. Segmental components are expressed by formants defining phonemes, and suprasegmental components are defined by pitch (fundamental frequency: F0), intensity and temporal parameters (duration of phonemes and of pauses). The present studies are aimed at investigating cortical auditory processing of segmental and suprasegmental components of speech using the techniques of event-related brain potentials (ERPs).

In Study I, two experiments were conducted to explore the auditory perception of vowels that belong or not to the phonemic inventory of listeners' native language. The mismatch negativity

(MMN) is an ERP component reflecting pre-attentive detection of auditory deviance compared with the integrated representation of standard features in the sensory memory (Näätänen et al., 2007; Kujala et al., 2007 for reviews). In the present study, this component was used to tap language-specific long-term memory traces for vowels (Näätänen et al., 1997; Cheour et al., 1998; Winkler et al., 1999a). An auditory sequence including four vowels (one standard and three deviants), each pronounced by three speakers, was presented in passive and active listening conditions. When all the vowels were distinctive in the native inventory (Experiment 1), the MMN was elicited by each deviant vowel category independently of speaker variability, demonstrating pre-attentive categorization based on long-term memory traces. By contrast, when a part of vowels did not belong to the native phonemic inventory (Experiment 2), deviant vowels did not elicit MMNs, indicating the lack of automatic categorization. Moreover, speaker's variability interacted with vowel category only in the active listening condition, as reflected by behavioral data and cognitive ERP components, suggesting that listeners depended more on conscious acoustic analysis of each stimulus than on pre-attentive categorization, when the deviant to discriminate is physically similar to the standard or is not supported by the inventory in the native language.

In Study II, another experiment (Experiment 3) was conducted to investigate the perception of small and large pitch changes embedded in sentences spoken in listeners' native language, jabberwocky or a foreign language. Behavioral performance in a pitch change detection task and the ERP components elicited by pitch changes (an early centro-temporal negativity and a later positivity) were analyzed. Results showed that subjects processed phrasal pitch better in their native language than in the foreign language, while the results for jabberwocky sentences were intermediate, suggesting that both familiar prosodic patterns and natural and meaningful context in the native language help listeners to detect subtle changes in phrasal pitch contour.

Overall results confirmed that long-term experience with one's native language has effects from early stages of cortical auditory processing involved in speech perception, modulating conscious perception of speech stimuli in different languages.

ACKNOWLEDGEMENTS

The present work was carried out in the laboratory of Dipartimento di Psicologia Generale, University of Padova (Padova, Italy), and at the Institut de Neurosciences Cognitives de la Méditerranée, CNRS-Marseille (Marseille, France).

I would like to express my sincerest gratitude to my supervisors Prof. Lucia Colombo and Dr. Mireille Besson for their constant support and guidance during all the research activities and training programs.

Special thanks are extended to my co-authors Dr. Magali Boureux, Dr. Michela Sarlo, Dr. Julie Chobert, Dr. Angèle, Brunellière and Prof. Noël Nguyen for their precious collaboration and scientific advices.

I wish to thank my colleagues in the PDG and DPSS in Padova, especially Elisa Puvia, Giulia Pavan, Veronica Montani, Michele Scaltritti and Mauro Marchetti, for their help in various scientific and non-scientific issues.

I am also very grateful to all my colleagues in INCM-CNRS in Marseille, Daniele Schön, Julie Chobert, Céline Marie, Clément François, Mitsuko Aramaki, Jean-luc Velay, Marieke Longcamp, Aurélie Lagarrigue, Jean Vion-Dury, Aline Frey, Dee Bolger, Vera Tsogli, and others, for helping me during the research work and for supporting me in the unforgettably friendly atmosphere inside and outside of the laboratory.

January, 2011 Chizuru Deguchi

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications.

[Study I, Experiment 1]

I. Deguchi, C., Chobert, J., Brunellière, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. Brain Research, 1366, 149–161.

[Study II, Experiment 3]

II. Colombo, L., Deguchi, C., Boureux, M., Sarlo, M., & Besson, M. (in press). Detection of pitch variation depends upon the familiarità of intonational contour of sentences. Doi: 10.1016/j.cortex.2010.03.008.

ABBREVIATIONS

ABR	auditory brainstem response
AST	asymmetric sampling time
EEG	electroencephalogram
ERP	event-related potential
F0	fundamental frequency
F1	the first sormant
F2	the second formant
IFG	inferior frontal gyrus
ITG	inferior temporal gyrus
L1	native language
L2	second language
LTM	long-term memory
MEG	magnetoencephalogram
MMN	mismatch negativity
MMNm	the magnetic counterpart of MMN
MLR	mid-latency response
MTG	middle temporal gyrus
NLM	the native language magnet theory
PAM	the perceptual assimilation model
PN	processing negativity
pITL	posterior inferior temporal lobe
SLM	the speech learning model
SOA	stimulus onset asynchrony
STM	short-term memory
STS	superior temporal sulcus
STG	superior temporal gyrus
TWI	transient window of integration
VOT	voice onset time

CONTENTS

EXPERIMENTS

STUDY I: Perception of native and non-native vowel contrasts (Experiments 1 & 2)

STUDY II: Perception of phrasal pitch contour in the native and unfamiliar languages (Experiment 3)

1. INTRODUCTION

Today, the acquisition of foreign languages has gained growing interest and utility. However, many of us encounter difficulties in understanding fluently spoken conversation in a foreign language, even when good grammatical knowledge has been acquired. This is because the brain system subserving auditory speech processing is tuned specifically to the phonological structure of our native language. Such adaptive brain plasticity is known to occur from the first year of age and enhances sensitivity to relevant acoustic cues of the native speech sounds to facilitate further language acquisition. Cross-linguistic studies of speech perception explore how the acoustic cues relevant for speech comprehension in a certain language are perceived depending on the listener's native language.

The present work aims to investigate the perception of segmental and suprasegmental aspects of speech in the native and in a foreign language, using the technique of event-related potentials (ERPs). Thanks to its excellent temporal resolution, the ERP method is a particularly suitable tool to study neural responses to speech stimuli that develop inherently in the temporal dimension. The languages considered in this work are Italian and French. These languages are both classified into 'syllable-timed' languages (Grabe & Low, 2002) and have similar grammatical structure. However, they also present notable differences at phonological level. For example, the inventory of vowels is larger in French than in Italian, and the stress position in multisyllabic words is relevant in lexical distinction in Italian but not in French. Such differences provide optimal conditions to compare cross-linguistic differences focused on phonological perception.Study I was designed to investigate the perception of segmental aspects, more specifically, the categorization of spoken vowels that do or do not belong to the phonemic inventory of listeners' native language. An auditory sequence of vowels selected from the French phonemic inventory was used in an odd-ball paradigm, with naturally spoken stimuli including multiple tokens from each category. The same set of stimuli was presented to two different native-language subject groups in two experiments – French and Italian native speakers, respectively. To assess the effect of attentional control on the auditory processing, ERPs were recorded in active and passive listening conditions. Based on Näätänen's model of auditory change detection, the influence of long-term linguistic experience on the early stages of auditory processing was examined by analyzing the N1 component and the mismatch negativity (MMN). Other cognitive ERP components, as well as behavioral measures, were also analyzed to investigate conscious processes involved in a deviant vowel discrimination task.

In Study II, the third experiment was conducted to study cross-linguistic effect on the perception of phrasal pitch contour, as one of the suprasegmental aspects. Based on previous studies (Schön et al., 2004; Marques et al., 2007), small and large pitch changes were applied on the last word of the sentence, and ERPs elicited by pitch changes were analyzed. Electroencephalogram (EEG) was recorded while subjects were engaged in a pitch incongruity detection task. In this experiment, subjects were Italian native speakers, and sentences in three different linguistic conditions were created: Italian, jabberwocky (strings of non-sense pseudowords preserving the phonological structure of Italian) and French. These conditions served to separate effects of familiar prosodic cues and of meaningful phrasal context on linguistic pitch processing.

The following chapters present the theoretical framework for the present studies. Chapter 2 outlines the characteristics of speech perception and its development, with particular interest on language-specific adaptation of the auditory system. In Chapter 3, essential cortical substrates in the human brain underlying speech perception are summarized. Chapter 4 first describes major ERP components reflecting different processes concerning spoken language perception. In the following and important part of this chapter, the mechanisms of auditory change detection and the related ERP components (Näätänen, 1990; Näätänen et al., 2011) will be explained. The three experiments are presented in Chapters 5–8.

2. PERCEPTUAL AND COGNITIVE ASPECTS OF SPEECH PERCEPTION AND ITS DEVELOPMENT

2.1. Main characteristics of speech perception

Speech sound perception shows some peculiar aspects compared to perception of non-linguistic sounds. Segmental phonology in a given language is formed by a limited number of discrete phonetic units (phonemes), and the inventory of relevant units varies between different languages. When we hear naturally spoken conversation, we can easily identify linguistic units and correctly reconstruct the sentences despite factors such as environmental noises and articulatory errors. It is as if a part of the auditory processing system in our brain is specialized to extract linguistic signals and decode significant information. Speech sounds can be described in terms of the articulatory movements engaged in pronunciation (articulatory phonetics) or in terms of their physical properties (acoustic phonetics).

Articulatory phonetics focuses on how speech sounds are produced by the articulatory organs. According to the distinctive feature theory (Jakobson et al., 1952) phonemes are classified on the basis of certain articulatory gestures that establish minimal acoustic contrasts between them (distinctive features). For example, /b/ contrasts with /p/ on a ‘voicing’ feature (voiced vs. voiceless), and with /d/ on a ‘place of articulation’ feature (labial vs. alveolar), etc. Jakobson et al. (1952) proposed that mental representations of linguistic sounds are stored as a configuration of such features used as the minimal units. While consonants can be classified according to voicing (voiced or voiceless), manner of articulation (nasal, oral, stop, fricative, etc.) and place of articulation (labial, alveolar, palatal, etc.), vowels, which are always voiced, are distinguished by the tongue height (high, mid and low), place of articulation (front, mid and back) and shape of the lips (e.g., round).

Acoustic phonetics aims at physically characterizing phonemes based on spectrographic analysis. The spectrogram visualizes frequency-amplitude-time analysis of an utterance computed by Fourier transformation. On the spectrogram phonemes are typically seen as bands at different acoustic frequencies (see Figure 2.1). These bands are called formants. In particular, the first three formants (F1, F2, F3) show good correspondence with articulatory movements such as place and manner of articulation and have a crucial role in phoneme recognition.

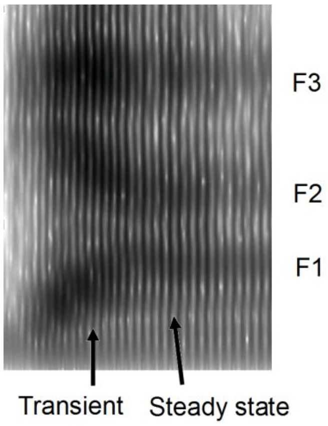


Figure 2.1. A spectrogram representing a CV syllable (/ga/ in this example)

The typical waveforms for a consonant-vowel (CV) syllable consist of rapid formant transitions in the initial segment (transient) and the following constant waveforms lasting from 100 to 300 ms (steady-state). The transient roughly characterizes the consonant, while the formant frequencies in the steady-state correspond to the vowel. Some phonetic features can be easily recognized on the spectrogram, such as stop consonants as a blank interval and fricatives or sibilants as a 'noise' including a wide range of frequencies. However, it is impossible to determine the formant values corresponding to individual consonants by dividing a CV syllable into the consonant and vowel components. This is partly due to a phenomenon called 'coarticulation', for which different phonetic features often overlap with one another in the course of speech production. For example, the transient corresponding to /d/ has different shapes when waveforms corresponding to /di/ and /du/ syllables are compared (Liberman et al., 1967), showing that the consonant is coarticulated with the vowel that follows it. In addition, a stop consonant separated from the vowel is not perceived as a consonant, but as a 'clic' noise, showing that perception of some consonants depends on both the transient and the initial segment of the steady-state. By contrast, the steady-state seems to have relatively stable formant values that determine the vowel category.

Moreover, spectrograms for voiced consonants and vowels specify fundamental frequency (F0) and a series of its harmonics. Pitch contour of spoken language, one of the major components of speech prosody, can be studied by tracking the waveform of the F0.

Categorical perception is a peculiar phenomenon of speech sound perception. Although acoustic parameters defining phonemes are continuous and greatly vary in natural speech sounds according to age and sex of the speaker, speech rate, co-articulation effects, dialectal accent, etc., phonemes can nevertheless be categorized as linguistically significant units. There is no doubt that such a mechanism has an adaptive value in facilitating speech comprehension. This phenomenon was demonstrated experimentally by researchers at Haskins Laboratories using a series of synthetic speech samples extracted from a formant transition continuum from /b/ to /d/ to /g/ covering the entire range of place of articulation changes. They found that listeners can discriminate easily between sounds that cross a phonemic boundary, whereas discrimination between sounds that fall within the same phonemic category tends to be poor, resulting in a steep slope around a perceptual phonemic boundary on a categorization curve (Liberman et al., 1957). Categorical perception has also been found for other dimensions of phonetic contrast characterizing consonants such as voice-onset-time (VOT)[1] (Liberman et al., 1961) and manner of articulation (Miyawaki et al., 1975).

By contrast, evidence for categorical perception in vowels is less clear, as some studies suggest that tokens within the same vowel category are discriminated quite well by adults (Pollack, 1952; 1953; Fry et al., 1962) and by infants (Swoboda et al., 1976), while other studies show the opposite (Iverson and Kuhl, 1995).

The phonological structure of a language does not depend only on the phonemes, but also on suprasegmental properties. Prosody refers to suprasegmental components of speech that consist of variations in pitch (F0), intensity, and rhythm (durational pattern of syllables and pauses). Although prosodic information does not convey any semantic meaning or concept, it has very important functions for smooth speech communication. For example, phrasal prosody serves for emphasizing words that mark the main topic in the sentence, indicating if the sentence is affirmative, or is a command or a question (pragmatic function), and transmitting the emotional state of the speaker. Moreover, prosodic cues are used to differentiate lexical items in some languages with consequent semantic implication, such as the lexical stress in Italian and Spanish and the pitch contour in tone languages (e.g., Chinese). Finally, prosodic structure is highly correlated with morphological and syntactic structures. Thus, prosodic information is used to disambiguate local and global syntactic structure (Christophe et al., 2003; Christophe et al., 2004; see Cutler et al., 1997 for a review), and inconsistency between the prosodic structure and the syntactic parsing of a spoken sentence results in processing costs (D'Imperio et al., 2007; Marslen-Wilson et al., 1992; Warren et al., 1995). Crowder et al. (1990) and Speer et al. (1993) suggested that the prosodic structure of a spoken

---

[1] Voice-onset-time (VOT) is one of the critical acoustic parameters that characterize phonetic features. The VOT distinguishes between voiced and unvoiced stop consonants and is measured as the time interval from the consonant burst to the initiation of the vocal cords' vibration. Stop consonants are perceived as 'voiced' when the VOT is short, and as 'voiceless' when the VOT exceeds a certain threshold. In the case of labial plosive consonants, if the vocal cord vibration begins within 25 ms after the release of the burst, the sound is heard as /b/, whereas if the VOT is longer than 25 ms, it is perceived as /p/.

sentence is partially integrated with segmental information in memory without mediation of semantics, and a rudimentary parsing can operate based on prosodic boundaries. This mechanism also explains how infants' sensitivity to prosodic cues can help them to discover basic morphosyntactic units (see Section 2.2).

As in the case of the phonemic inventory, the prosodic structure also differs from language to language. For example, languages can be classified based on rhythmic properties. According to traditional rhythm class theory (Abercrombie, 1967), in stress-timed languages (e.g., English, Dutch and German) the interval between stressed syllables is kept approximately constant by reducing the length of weak vowels, while in syllable-timed languages (e.g., Italian, French and Spanish) the syllable length is relatively stable. Grabe & Low (2002) validated this classification calculating durational variability in successive phonetic units (*Pairwise Variability Indices*, PVI). However, many languages remain unclassified, presenting intermediate or mixed rhythmic properties between stress-timed and syllable-timed classes. This further illustrates phonological dissimilarities between various languages.

2.2. Development of language-specific perception in infants

Previous studies demonstrate that newborns can differentiate speech-relevant acoustic features and discriminate phonetic contrasts of any language. In the course of early speech acquisition, infants learn to focus their attention selectively on acoustic cues that are most relevant in their language (Jusczyk et al., 1990). This section reviews how infants' perceptual sensitivity to speech sounds becomes progressively specialized to the phonemic inventory of their native language during the first year of life (see Kuhl, 2004 for a review).

Early studies on speech perception in infants investigated their capacity to discriminate minimal phonetic contrasts between two sounds. One- and 4-month-old English-learning infants are capable of discriminating consonant pairs differing in phonetic features such as VOT (Eimas et al., 1971), place of articulation (Eimas, 1974; Levitt et al., 1988) and manner of articulation (Eimas, 1975; Eimas & Miller, 1980; Hillenbrand et al., 1979), as well as vowel contrasts (Kuhl & Miller, 1982; Swoboda et al., 1976, 1978; Trehub, 1973). Furthermore, discrimination ability for some phonetic contrasts (consonants and vowels) has been shown in infants even 4–5 days after birth (Bertoncini & Bijeljac-Babic et al., 1987).

After 6 months of age, infants' sensitivity to non-native speech contrasts begins to decline (Werker & Tees, 1984; Werker & Lalonde, 1988; Polka & Werker, 1994), while their perception of native phoneme contrasts improves. During this period, infants seem to engage some learning strategies such as pattern recognition, statistical learning, and alteration of perceived acoustic space (Kuhl, 2004).

Pattern recognition can be investigated by testing the ability to categorize phonemes by ignoring acoustic and phonetic differences among tokens of each phonemic category produced by different speakers. Previous studies demonstrated 6-month-old infants' skill to categorize vowels (Kuhl, 1979, 1983; Kuhl et al., 1992) and consonants (Hillenbrand, 1983) in syllables spoken by both male and female speakers. Moreover, Jusczyk et al. (1992b) showed that 2-month-old infants are capable of sorting two items that differ by a consonant ('bug' and 'dug') spoken by 12 different speakers (six male and six female), although they did not discriminate the words when a 2-minute delay was inserted between exposure to the training stimuli and testing. Furthermore, more recent studies, using electrophysiological techniques (see Section 4.2.2.2 for MMN studies on phoneme categorization), revealed that the brain of newborn babies can distinguish phonetic contrast (/pa/ vs. /ta/) by extracting relevant features from acoustically variable speech sounds (Dehaene-Lambertz & Pena, 2001). However, some studies suggest that, when presented with words (as opposed to syllables) spoken by different speakers, infants aged about 10.5 months may store in memory representations of specific instances produced by different speakers rather than abstract phonological prototypes (Houston et al., 1998; Houston & Jusczyk, 2000).

Several studies show that language is learned by infants in an implicit manner, by computing the distributional probabilities of occurrence of specific units (i.e., phonemes, syllables, words) or sequences of units in the linguistic environment to which infants are exposed. Such distributional information provides salient cues for learning phonetic categories from the early stage of language acquisition, because relevant phonemes in the native language occur much more frequently than non-native speech sounds. Consequently, frequent sounds are stored in memory as prototypical exemplars of native-language categories and other less frequent sounds are categorized according to their physical similarity to prototypical sounds. For example, Maye et al. (2002) demonstrated sensitivity to short-term statistical learning in 6-month-old infants using VOT contrasts. Thus, prototypes of native phonemes are considered to be implicitly learned by exposure, based on the distributional properties of different speech inputs that infants have heard (Kuhl et al., 1997). The Native Language Magnet model (Kuhl et al., 1992, 2008) proposes that development of language-specific prototypes reorganizes the perceptual space onto which native and non-native speech sound inputs are mapped, so that it better reflects the phonemic inventory of one's native language. Such reorganization causes perceptual distortion by altering speech-relevant acoustic dimensions. In particular, prototypes act as perceptual magnets by attracting non-typical sounds, and emphasizing boundaries between the native language categories. At the same time, certain non-native boundaries disappear, which explains why infants' sensitivity to some non-native contrasts declines. The language-specific 'perceptual magnet effect' (Kuhl, 1991) has been shown in 6-month-old infants, and their speech perception reflects the statistical distribution of vowels in the native language (Kuhl et al., 1992).

Statistical learning contributes not only to acquisition of the phoneme inventory, but also to discovery of relevant cues for the segmentation of continuous speech flow into important units such as words and longer units of linguistic organization. Some previous studies emphasize the importance of prosodic cues in the initial acquisition of speech segmentation (Jusczyk et al., 1992a, 1999). As in the case of phonetic contrasts, typical prosodic patterns differ between languages. Nazzi et al. (1998) suggested that newborn babies are already sensitive to the rhythmic characteristics of utterances in their native language, and other studies provide evidence that infants seem to learn prosodic aspects before acquiring segmental contrasts of their native language (Best, 1994; Jusczyk et al., 1993; Mehler et al., 1988; Mehler & Christophe, 1995). The typical rhythmic pattern of one's native language provides salient cues in word segmentation. For example, words carrying stress on the initial syllable are most frequent in English (Cutler & Carter, 1987). In fact, English-learning infants at 7.5 months of age have been shown to segment words by identifying word onsets at stressed syllables, thereby mis-segmenting weak/strong words (Jusczyk et al, 1999), while children aged 10.5 months show some skill in segmenting weak/strong words (Jusczyk et al., 1999; Myers et al., 1996). Before completing the acquisition of lexical knowledge, segmentation of words having a prosodic pattern that is not predominant in the native language must depend on other types of acoustic information. For example, in every language there are phonotactic rules that decide if a given sequence of phonemes or syllables in words is permissible. When they are between 6 and 9 months of age, infants learn to use probabilistic phonotactic rules governing the word structure to segment speech flow (Jusczyk et al., 1994). Infants at 8 months of age have been shown to be capable of learning to segment words based only on the statistical likelihood of a sequence of syllables (i.e., transitional probabilities) after short-term (2 minutes) auditory exposure to experimental syllable strings (Aslin et al., 1998; Saffran et al., 1996). After 9 months of life, infants' use of cues deriving from the segmental structure increases compared with prosodic ones and so does the ability to segment words successfully with a non-predominant prosodic pattern (Johnson & Jusczyk, 2001; Mattys et al., 1999). Development of infants' sensitivity to segmental cues in this period has also been demonstrated in a cross-linguistic study by Jusczyk et al. (1993) using spoken word lists in two languages having similar prosodic structures (English and Dutch):

English and Dutch 9-month-old infants showed a preference to listen to the list in their native language, while 6-month-olds did not. Further development of children's strategies for word segmentation occurs with lexical acquisition. Adult-like strategies are viewed as a constraint satisfaction process, integrating various sources of information: phonetic, prosodic, lexical, and phonotactic cues characterizing the beginnings and endings of spoken words.

Moreover, statistical learning of prosodic cues also has a role in segmentation of larger units of linguistic structure. For example, pauses, falling pitch contour, and prolongation of final syllables frequently mark clause boundaries, and sensitivity to these cues has been found in 6-month-olds (Nazzi et al., 2000). Because of the close relationship between prosodic and syntactic boundaries, sensitivity to acoustic cues associated with prosodic boundaries should be helpful for children to learn rudimental syntactic parsing and the corresponding segmental markers such as grammatical morphemes (Jusczyk, 1999). Furthermore, the ability to use statistical cues to learn grammatical structure by short-term passive listening has been demonstrated in 12-month-olds (Gómez & Gerken, 1999), and such ability is considered as the foundation for further acquisition of the specific syntactic organization of the native language.

2.3. Effects of linguistic experience on the perception of foreign speech sounds

2.3.1. Theories on cross-linguistic perceptual interference

The selective sensitivity to the native phonemic inventory shaped during the first year of life (Werker & Tees, 1984; Werker & Lalonde, 1988) persists in adulthood, supporting automatic categorization and recognition of speech codes in the native language. On the other hand, the same mechanism also hinders late acquisition of foreign languages, distorting our perception of non-native speech sounds. The decline in sensitivity was found to affect some but not all non-native contrasts (Best et al., 1988, 1995).

Some models that have been proposed to explain how sensitivity declines for a part of non-native contrasts consider as an important factor the acoustic relationship of the non-native sounds in question with the native phonemic categories. For example, the Perceptual Assimilation Model (PAM) proposed by Best (1995) assumes that a non-native speech sound can be assimilated to a native category, or perceived as a non-categorizable speech sound, or perceived as a non-speech sound (noise). Accordingly, if a given non-native contrast is mapped onto two different native phoneme categories, it should remain easy to discriminate. By contrast, if two non-native sounds are assimilated to a single phonemic category, they are the most difficult to discriminate. This class of non-native contrasts should therefore be susceptible to the decline of sensitivity that occurs during native language acquisition. Finally, difficulty of discrimination for non-native sounds falling outside native language categories should be intermediate and depends on how they are mapped in relation to native categories. The Speech Learning Model (SLM; Flege, 1995) also postulates that the perceived similarity between second language (L2) and native language (L1) sounds determines the probability of categorization: the greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely a new phonetic category is established (Flege, 1995). Finally, the Native Language Magnet Theory (NLM) proposed by Kuhl (1993) has focused on the developmental process, during which infants acquire representations of prototypical instances of native phonemic categories by exposure to their native language environment (Kuhl, 1991; Kuhl et al., 1992, 1997). Such prototypical representations exert a perceptual magnet effect by attracting other less prototypical instances falling into the same category. Consequently, listeners are less likely to discriminate contrasts between a prototypical instance and another exemplar of the same category than to discriminate two peripheral instances of the category (Kuhl, 1991).

However, Lotto et al. (1998) pointed out that the perceptual magnet effect may be due to the fact that a pair of non-prototypical exemplars is more likely to cross a boundary between phonemic categories than a pair of prototypical ones. They examined discrimination ability using instances of the vowels between /i/ and /e/ and obtained data that could be explained by higher discriminability for cross-category contrasts than for within-category contrasts, as accounted for by categorical perception (Liberman et al., 1957), but without the additional effect of prototypicality.

2.3.2. Previous cross-linguistic studies

As described in the previous sections, the perception of foreign speech sounds is influenced by phonological properties of one’s native language. More specifically, some non-native phonemes that are acoustically similar to a phonetic category in one’s native language are difficult to discriminate. For example, it is well known that Japanese native speakers show persistent difficulty in distinguishing between the two consonants /l/ and /r/, which are both mapped onto the same category in Japanese (Goto, 1971; Bradlow & Pisoni et al., 1997). Similarly, the Hindi contrast between the retroflex /D/ and dental /d/ is easily distinguished by Hindi native speakers, but English native speakers find it hard to discriminate (Werker & Lalonde, 2005). Pallier et al. (1997, 2001) have shown the remarkably persistent effect of very early experience with the native language by comparing two groups of fluent Catalan-Spanish bilinguals having Catalan-speaking and Spanish-speaking parents, respectively. Catalan-dominant bilinguals distinguished significantly better than Spanish-dominant bilinguals the Catalan /e/-/ɛ/ vowel contrast, which is not distinctive in Spanish.

Cross-linguistic perceptual differences have also been found for suprasegmental features that are relevant for lexical distinction in some languages but not in other languages, such as stress (Spanish), lexical tones (Chinese, Thai), and vowel length (Finnish, Thai). Dupoux et al. (1997) reported a perceptual phenomenon defined '*stress deafness'*, for which French native speakers, compared to Spanish subjects, showed much more difficulty in discriminating non-words that differ only in the stress position (e.g., [vásuma] vs [vasúma] vs [vasumá]). In French, stress almost always falls on the last syllable of the word. Consequently, French speakers presumably do not focus their attention on the stress position in words, because it has no utility for correctly identifying lexical items in their native language. By contrast, Spanish speakers are considered to be more trained to process stress position to distinguish different lexical items (e.g., *bébe* ['drink'] vs. *bebé* ['baby']) through long-term experience with their native language.

The perception of lexical tones in tone languages has been studied extensively by comparing native speakers of a tonal with those of a non-tonal language. For example, American English native speakers have difficulty in distinguishing Mandarin Chinese tones (Wang et al., 1999). In addition to perceptual sensitivity, Lee et al. (1993) also suggested different strategies in processing lexical tones between the language groups: tone and segmental information are perceived in an integrated way by Chinese native speakers, while these dimensions are processed independently by British English speakers. Furthermore, Gandour et al. (2003) revealed using fMRI (functional magnetic resonance imaging) that Chinese lexical tones are processed differently in the brain by Chinese and American English native speakers. Lexical tones activated areas in the left hemisphere only in Chinese subjects, while American English subjects showed bilateral activation. By contrast, pitch contours of longer phrases (intonation) activated predominantly areas in the right hemisphere in both language groups. These results illustrate well how prosodic processing is linked with other linguistic functions.

These studies show how our speech perception system is specialized to receive signals that are relevant in our native language and accommodates sounds of unfamiliar languages according to the native phonetic structure. The impact of very early linguistic experience is due to the neural plasticity in children's brains that optimizes the neural substrates supporting speech perception in an adaptive way according to the linguistic environment with which children interact.

3. NEURAL BASES OF SPEECH PERCEPTION

Certain perceptual phenomena specific to speech sounds indicate that linguistic and non-linguistic sounds are processed by at least partially separate mechanisms. Linguistic and non-linguistic sound inputs are first treated in the common subcortical auditory pathways, from which signals are sent to the hierarchically structured cortical auditory systems. Speech processing proceeds from general acoustic analysis to more specialized higher-order linguistic functions (Hickok & Poeppel, 2000). Cortical sound processing begins in the primary auditory cortex (portions of Heschl's gyrus), which is activated even by unstructured auditory stimuli such as noise. The next step of processing appears to occur in the supratemporal plane, which responds not only to speech sounds but also to non-speech sounds that are spectro-temporally structured (e.g., pure or complex tones) better than to unstructured noise. Areas in the bilateral dorsal superior temporal gyrus (STG), adjacent to the primary auditory cortex, are considered to be involved in the analysis of physical features of sounds and respond similarly to speech and non-speech sounds (Liebenthal et al, 2005).

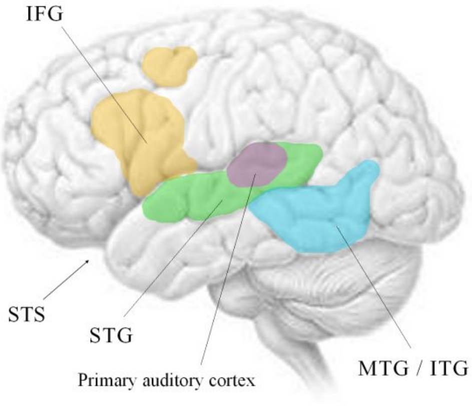


Figure 3.1. Functional anatomy of speech-sound processing
The initial acoustic/phonetic analysis occurs in the bilateral superior temporal gyri (STG). The outputs from the STG are processed in areas in the left hemisphere including the inferior frontal gyrus (IFG), middle temporal gyrus (MTG) and inferior temporal gyrus (ITG).

By contrast, the left middle and anterior superior temporal sulcus (STS) has been shown to respond better to familiar syllables than to comparably complex non-linguistic sounds (Liebenthal et al, 2005). Thus, it seems that while all sound inputs are processed bilaterally in the primary auditory cortices (STG), only the output for speech sounds is sent from the STG to the left non-primary auditory cortices (STS). Zatorre & Belin (2001) explained this asymmetrical activity in the STS by assuming that the two hemispheres have different and complimentary spectral and temporal sensitivities: the left STS is more suitable for processing rapid acoustic changes requiring high temporal resolution, while the right STS can better analyze fine spectral properties. Poeppel (2001) further proposed the Asymmetric Sampling in Time (AST) hypothesis, which defines more precisely different sampling rates (*Temporal integration windows*) at which the left (20–50 ms) and right (150–250 ms) auditory cortices extract spectro-temporal sound information. Among acoustic components of speech sounds, formant transition determining consonants typically occurs in a very short time window (< 50 ms), whereas spectral structure and pitch contour characterizing vowels and musical tones are defined in longer temporal windows. Consequently, consonants are prevalently processed in the left hemisphere, and fine-grained pitch and prosodic contour are analyzed in the right hemisphere (Joanisse et al., 2003; Boemio et al., 2005; Zatorre et al, 2002). Thus, the left STS is considered to represent an advanced stage in the hierarchy of auditory processing, crucial for phoneme perception and as an interface between the primary acoustic analysis and following higher-level linguistic processes.

According to the Dynamic Dual Pathway model of auditory language functions (Friederici & Alter, 2004), after acoustic analysis in the STG, syntactic/semantic information is primarily processed in a left temporo-frontal hemispheric pathway, whereas sentence-level prosody is processed in a right temporo-frontal hemispheric pathway. The lateralization of cortical activation depends on stimulus properties (Pannekamp et al., 2005) and task demands (Plante et al., 2002), reflecting dynamic interaction between syntactic and prosodic information during speech sentence comprehension, which is possibly mediated by the corpus callosum (Friederici et al., 2007). Furthermore, semantic processing was identified in the inferior frontal subcomponent, including the middle temporal gyrus (MTG) and the BA 45 in the inferior frontal gyrus (IFG), and the syntactic circuit in the temporo-frontal subcomponents including the frontal operculum and the inferior portion of Broca's area (BA 44) (Friederici, 2002). The pathway for sentence-level prosody includes areas in the STG and the right frontal operculum (Meyer et al., 2002, 2004).

The left pathway for semantic and morphosyntactic processing was also postulated by Hickok & Poeppel (2000, 2004) as the ventral stream, which projects from the STG involving the STS to the posterior inferior temporal lobe (pITL) that includes portions of the middle temporal gyrus (MTG) and inferior temporal gyrus (ITG). The ventral stream is taken to serve as an interface between acoustic-phonetic representations and lexical/conceptual representations and between lexical items and morpho-syntactic operations. The left pITL was proposed as the locus of sound-to-meaning (or concept) interface system (Hickok & Poeppel, 2000), whereas the left anterior temporal regions (e.g., anterior STS) has been suggested as the locus of sentence-level grammatical processing (Friederici et al., 2000a; Humphries et al., 2001; Mazoyer et al., 1993; Schlosser et al., 1998). Hickok & Poeppel (2000, 2004) also postulated the dorsal stream, another pathway involved in the auditory-motor integration of speech by mapping sounds onto articulatory representations.

4. EVENT-RELATED POTENTIALS (ERPs) REFLECTING SPEECH PERCEPTION

4.1. Auditory ERP components mirroring spoken language processing

In the present studies, auditory cortical Event-related brain potentials (ERPs) were used to explore the neural substrates of perceptual and cognitive processes associated with speech perception. The excellent temporal resolution of ERPs is a great advantage for exploring speech processing in the brain, which inherently develops in time following incoming auditory materials.

Cortical ERPs can be classified into *exogenous* (or obligatory) components elicited by any external events reflecting sensory/perceptual (e.g., visual and auditory) processes and *endogenous* components reflecting internal cognitive processes, such as expectation, conscious processing of stimulus and response, and decision making. Early ERP components, such as the P1, N1 and P2, are exogenous, while the Mismatch Negativity (MMN) and other ERPs with longer latencies are classified into endogenous components.

Previous ERP studies using various linguistic materials have contributed to elucidate the time course of speech processing in the brain. Whereas basic acoustic-feature extraction in the primary auditory cortex was detected using magneto-encephalography to responses as early as 20–50 ms after the stimulus onset (Lutkenhoner et al., 2003), phonological representation for speech sounds was reflected by responses with latencies of 100–200 ms (Poeppel et al., 1996; Obleser et al., 2004). Among later ERP components, the N2 and P3 family reflects general cognitive processes such as conscious analysis of stimuli, decision and response processing (see Donchin & Coles, 1988; Kok, 2001; Polich, 2007 for reviews). Some ERP components are known to mirror specific linguistic processes. One of the most classic language-related components is the ‘semantic’ N400 (Kutas & Hilliard, 1980). In the experiment of Kutas & Hilliard (1980), when subjects read sentences ending with a semantically incongruous word (e.g., ‘*He takes coffee with cream and dog*’), the N400 was observed as a negative component peaking at about 400 ms after the beginning of the last word (i.e., ‘*dog*’) over central sites. Moreover, the amplitude of the N400 correlated with the degree of semantic inconsistency of the target word. The N400 has also been elicited by auditorily presented stimuli (Bentin et al., 1993; Friederici et al., 1993), suggesting that it is a modality-nonspecific index of semantic processes. A slightly earlier subcomponent (N350) has also been identified as an index of lexical processing (Bentin et al., 1999; Embick et al., 2001; Pylkkanen et al., 2002; Stockall et al., 2004). For syntactic processing, the *early left anterior negativity* (ELAN: Neville et al., 1991; Friederici et al., 1993, 2000b; Friederici & Mecklinger, 1996) and the P600, a positive component peaking at around 600 ms at centro-parietal sites (see Osterhout et al., 1997; Osterhout & Hagoort, 1999 for reviews), have been reported by using violations of grammatical rules or sentences that include syntactic ambiguity requiring revision of sentence structure (i.e., a 'garden-path' effect). The ELAN is associated with the early automatic parsing processes (Hahne & Friederici, 1999) and reflects an increased load on working memory associated with the syntactic segmentation (King & Kutas, 1995). By contrast, the P600 is not lateralized and is supposed to reflect conscious re-analysis of structural relationships detected by the syntactic parser (Friederici et al., 1996).

Neurophysiological correlates for processing of phrasal prosody have been investigated in some studies. For example, a positive component (Closure Positive Shift, CPS) peaking at around 300 ms has been observed in correspondence with clause boundaries. This component was elicited by using both spoken sentences and degraded speech sentences that maintained only prosodic structures without discernable segmental elements, thereby suggesting that it reflects processes related with prosodic information that listeners exploit to determine the sentence structure (Pannekamp et al., 2005; Steinhauer et al., 1999; Steinhauer & Friederici, 2001; Steinhauer, 2003). Moreover, Astésano et al. (2004), by orthogonally manipulating semantic congruity and modal prosody (i.e., typical intonational contours indicating declarative and interrogative sentences) in sentences, demonstrated interaction between prosodic processing reflected by a delayed positivity (P800) and semantic processing, reflected by an N400 component.

4.2 The N1 and the MMN: early phases of speech sound perception

4.2.1 N1 reflects acoustic feature analysis of sounds

The N1 component, the exogenous ERP component that has been most extensively documented, is found as a negative deflection peaking at about 100 ms after physical event over the fronto-central scalp sites (Näätänen & Picton, 1987). The N1 is considered to be a transient neural response to sensory stimulation including stimulus onsets, offsets, and changes in sound features (Näätänen & Picton, 1987; Näätänen et al, 1988; Näätänen & Winkler, 1999). The N1 seems to reflect activation of neural populations in the cortical area where the stimulus is represented (Näätänen et al, 1992). Rapid repetition of stimuli sharing common features diminishes the N1 amplitude, probably due to the physiological refractoriness of the neural populations responding to those features (see Näätänen & Picton, 1987 for a review) or to latent inhibition mechanisms (Sable et al., 2004).

In previous studies on speech sound processing, the N1 has been found to be sensitive to the spectral and temporal cues critical for speech sound perception (Martin et al., 1997; Ostroff et al., 1998; Whiting et al., 1998). Tremblay et al. (1998, 2001) have also shown an increase in amplitude of the N1-P2 complex after perceptual training with syllables, reflecting plastic changes in the afferent neural system induced by stimulation. Moreover, by using the Magnetoencephalography (MEG) method, Obleser et al. (2003) have shown that the N100m component is sensitive to spectral dissimilarities, as mapped in a F1-F2 vowel space. Diesch & Luce (2000) have also reported that the source strengths of the N100m elicited by individual formants or by two-formant vowels synthesized by additive superposition of single formants reflects contribution of both formants in a superadditive way, suggesting interactive processes for feature analysis at one or more stages of an afferent auditory pathway.

The generators for the auditory N1 are located in the supratemporal plane and lateral areas of the auditory cortices and in frontal areas (Giard et al., 1994; Näätänen & Picton, 1987). Näätänen & Picton (1987) have identified three anatomically and functionally distinct subcomponents of the N1. The first component peaks at around 100 ms predominantly at fronto-central scalp areas (Vaughan & Ritter, 1970), and appears to be generated in the bilateral supratemporal plane. The amplitude of this subcomponent shows refractoriness to repetitive stimuli due to adaptation of feature-specific neurons (Näätänen et al, 1988). Therefore, this component is functionally attributed to sensory analysis of the input, decoding of the afferent signals, or the mnemonic retention of sensory features of the stimulus (Näätänen & Picton, 1987). The second subcomponent of the N1 is observed over the temporal areas, with a positive peak at around 100 ms followed by another negative peak at around 150 ms. It is generated by radially oriented generators in the lateral STG (Scherg & Von Cramon, 1986; Picton et al, 1999). The amplitude of its component is generally larger on the right hemisphere, but enhanced on the counterlateral side to the stimulated ear (Wolpaw & Penry, 1977). The third component is a vertex-negative wave peaking at around 100 ms. Näätänen & Picton (1987) suggested that this component may functionally reflect a diffused transient arousal enhancement of the neural networks due to modality-general sensory stimulation, which facilitates detection and analysis of the stimulus and task-related cognitive functions. The source of this component is located in the thalamo-reticular system that projects to the frontal premotor cortices (Velasco & Velasco, 1986; Velasco et al., 1985).

Woods (1995) and Picton et al. (1999) also proposed a componential structure of the N1 including at least 5 subcomponents. The N1a (75–95 ms) and the N1c (130–170 ms) components, together with the interpolated positivity (Ta; peaking at 100–150 ms; Wolpaw & Penry, 1975) are generated by radially oriented generators in the lateral STP and lateral STG. The frontocentral N1'/P90 (85–110 ms) and vertex N1b (110–140 ms) are considered to be generated by tangentially oriented sources in the supratemporal plane. The N1'/P90 generators are probably tono-topically organized in the auditory cortices (Woods, 1995; Picton et al, 1999), whereas the N1b appears to be modality-general.

4.2.2. The Mismatch Negativity (MMN)

4.2.2.1. MMN reflects pre-attentive auditory change detection

The MMN is an endogenous ERP component typically elicited using the odd-ball paradigm by infrequent deviant stimuli inserted in a sequence of repeated standard stimuli. The MMN is considered to reflect automatic, pre-attentive change detection (Naätänen et al., 1978) and is elicited by changes in the physical features of the acoustic stimuli, such as frequency (pitch), duration, intensity and timbre, as well as by changes in the abstract relationships between sounds, such as an ascending tone pair in a sequence of descending tone pairs (Tervaniemi et al., 1994b) or by violations of a combinatory rule between pitch and intensity (Paavilainen et al., 2001; see Näätänen, 2001; Näätänen et al., 2007, Kujala & Näätänen, 2010 for reviews). Moreover, for the MMN elicitation it is not necessary that the standard stimuli are physically identical, but if one or more component features are commonly shared by many of the stimuli in the sequence, they are used as standard features to detect deviants (Näätänen et al., 2004; Pakarinen et al., 2010).

One of the important characteristics of the MMN is that its generation reflects automatic, pre-attentive processes that are independent of conscious awareness. The MMN can be elicited even when the subject is not paying attention to the stimuli (Woldorff et al., 1991, 1998; Näätänen, 1991; Näätänen et al., 1993a). This striking feature of the MMN has made it a potent tool for investigating children and infants (Alho et al., 1990; Cheour et al., 1998), and even fetuses (Huotilainen et al., 2005).

Another important characterisctic of the MMN generation is involvement of short-term memory: the MMN cannot be elicited before the memory trace has been developed by extracting the regular aspects from the sensory information provided by the preceding sound sequence

(Bendixen et al., 2007; Bendixen & Schröger, 2008; Cowan et al., 1993; Sams et al., 1985). Based on regularities governing the preceding stimuli and stored in the memory trace, which usually lasts a few seconds, predictions upon the forthcoming auditory events are automatically formed by extrapolation (Näätänen et al., 2001). The MMN reflects detection of deviance resulting from comparison between the incoming sound and such predictions (Näätänen & Winkler, 1999; Kujala & Näätänen, 2010). Therefore, single auditory events without any preceding sounds elicit obligatory ERP responses reflecting sensory stimulation such as P1, N1 and P2, but no MMN (Näätänen & Picton, 1987; Korzyukov et al., 1999).

Generally, the MMN amplitude is known to increase as a function of magnitude of the physical dissimilarity between the deviant and standard stimuli (Tiitinen et al., 1994; Pakarinen et al., 2007). Moreover, the MMN amplitude for deviant sounds has often been found to be positively correlated with discrimination accuracy even for equal physical dissimilarities (see, for example, Amenedo & Escera, 2000; Kujala et al., 2001; Novitski et al., 2004; Kujala & Näätänen, 2010 for a review), and as such it has been used to probe plastic changes in cortical memory traces due to perceptual training (Näätänen et al., 1993b; Kraus et al., 1995; Menning et al., 2002; Tremblay et al., 1998). However, the MMN amplitude is not a direct measure of behavioral accuracy, because it only reflects the early stages of cortical sound processing, but not conscious cognitive processes that can contribute to discrimination accuracy. For example, Tremblay et al. (1998) trained subjects to discriminate VOT contrasts using syllables in four sessions and monitored the training effects using the MMN and identification tests. Interestingly, some subjects showed increases in MMN amplitude even before improvement in the behavioral measures were apparent, suggesting that changes in neural responses associated with early sound processing, as reflected by the MMN, may precede the improvement in discrimination performance at a conscious level. The correlation of the MMN with auditory discrimination sensitivity has also served to reveal deficient auditory functions in dyslexic children for some auditory contrasts, such as consonant changes (/ba/ vs. /da/; Schulte-Körne et al., 1998), pitch changes (Baldeweg et al., 1999; Schulte-Körne et al., 1999) and changes in rhythm pattern in complex sound sequences (Kujala et al., 2000).

Source localization of the MMN has detected two subcomponents of the MMN. The first component is elicited by generators that have been localized in the bilateral supra-temporal cortices using the equivalent current dipole (ECD) method (Giard et al., 1995; Rinne et al., 1999; Scherg et al., 1989; Jemel et al., 2002;) and MEG recordings (Levänen et al., 1996; Alho et al., 1998; Sams et al., 1991; Csépe et al., 1992). This supra-temporal MMN subcomponent presents a typical polarity inversion at electrodes placed on the mastoids (below the Sylvian fissure) in nose-referenced recording. Another subcomponent is generated by processes localized predominantly in the right frontal lobe (Näätänen et al., 1978; Giard et al., 1990; Rinne et al., 2000). Functionally, the supratemporal MMN component seems to be related with early sound change detection, while the frontal component is taken to be associated with involuntary attention call to auditory deviance (Näätänen et al., 1978; Näätänen & Michie, 1979; Giard et al., 1990; Rinne et al., 2000; Escera et al., 1998; Schröger, 1996a, 1996b; see Näätänen et al., 2007 for a review). However, it should be noted that the MMN does not index the effective attention switch itself, which is induced only when the deviant sound is intrusive enough. Such attention switch is reflected in the P3a component (Donchin & Coles, 1988; Escera et al., 2000) that appears following the MMN (see Chapter 4.2.1).

It has also been shown that the loci of the MMN generator depend on the dimension of sound changes, indicating that partly separate neural populations in the cortex are activated by different acoustic dimensions (Giard et al., 1995; Rosburg, 2003). Different localizations of MMN generators can also explain different polarity-reversal ratios of MMNs elicited by the frequency, duration, and intensity deviants (Paavilainen et al., 1991). Moreover, MMNs elicited by nonlinguistic sounds are usually larger in the right than in the left hemisphere (Giard et al., 1995; Paavilainen et al., 1991), whereas activation of MMN generators to language-specific stimuli has often been found to be stronger in the left than in the right fronto-temporal cortex (Näätänen, 2001; Shtyrov et al., 2000). In addition, different sources in the fronto-central cortex have been identified for MMNs to different types of linguistic information, probably due to involvement of different linguistic processes (Pulvermüller et al., 2005; Shtyrov et al., 2004).

Although both the N1 and MMN typically occur within 100–200 ms from change onset independently of conscious processing of auditory information, these components present different characteristics due to their different elicitation mechanisms (Näätänen, 1990, Schröger and Wolff, 1998; Rinne et al., 2006). While the presentation of any type of sounds typically elicits an auditory N1, the MMN is associated with the presentation of sounds violating regularities extracted from previous sound sequences based on integrated representations of auditory events (Näätänen & Winkler, 1999). In fact, while the first stimulus in a sequence after a long period of silence elicits a very large N1, no MMN can be elicited without preceding standard sounds (Näätänen & Picton, 1987; Korzyukov et al., 1999; Näätänen et al., 1989; Sams et al., 1985). N1 amplitude also shows refractoriness to repetitive sounds, suggesting adaptation of feature detector neurons, while the MMN to deviants grows during the session (Näätänen et al., 1993b). Moreover, some types of deviants elicit the MMN but not the N1, such as sound intensity decrements (Näätänen et al., 1987; Näätänen, et al., 1989), omission of a sound (Tervaniemi et al, 1994a), or a violation of abstract relational regularities between sounds (Saarinen et al., 1992; Tervaniemi et al., 1994b; Paavilainen et al., 1995, 1999). These examples of dissociation indicate that activation of the N1 and MMN generators depends on different types of sensory code, which were described as the pre-representational and representational systems by Näätänen & Winkler (1999). While the N1 generator appears to use the sensory codes of a single event, the elicitation of the MMN depends on the full analysis of the acoustic features and integration of information provided by the different feature detectors and temporal and higher-order properties detected in a sequence of sounds into a unitary representation (Näätänen & Winkler, 1999).

In addition, different generation loci have been detected for the MMN and the N1 using MEG. For example, the MMNm to frequency change is generated in the supratemporal cortex 3–10 mm anteriorly to the N1m source (Sams et al., 1991; Csépe et al., 1992; Huotilainen et al., 1993; Tiitinen et al., 1993; Levänen et al., 1993, 1996; Korzyukov et al., 1999; Rosburg, 2003; Rosburg et al., 2004).

4.2.2.2. MMN as a probe of central sound representation

As another remarkable feature of the MMN, this component has been shown to reflect long-term memory traces, and in particular, it is a useful tool to probe the permanent language-specific representations for speech-sounds.

The MMN amplitude to speech sounds is modulated not only by physical distance between the standard and deviant stimuli, but also by typicality and familiarity of the speech exemplar, which are important factors influencing the formation of sensory memory representations (Huotilainen et al., 2001). For example, phonemes in one's native language have been found to elicit larger MMN responses than non-native speech sounds (Näätänen et al., 1997; Dehaene-Lambertz, 1997), presumably due to the activation of native-language phonetic prototypes (Ylinen et al., 2006). Based on the assumption that instances within each phoneme category are organized around the prototype according to the degree of typicality (Iverson & Kuhl, 1996, 2000), Aaltonen et al. (1997) could demonstrate the perceptual magnet effect (Kuhl, 1991) at the pre-attentive level using the MMN. They used synthesized vowels from the continuum between /i/ and /y/ vowels created by changing only the F2, and compared discrimination accuracy and MMN amplitude between two equidistant pairs extracted from different F2 ranges of the continuum, one including two exemplars close to the prototypical /i/ and the other including instances judged as non-prototypical exemplars of /i/ (close to the perceptual boundary with /y/). In line with the prediction based on the perceptual magnet effect, Finnish subjects showed poorer discrimination performance and also a smaller MMN to the contrast between prototypical tokens than to the contrast between non-prototypical tokens. The difference can be explained only by typicality of tokens, as both contrasts showed the same acoustical distance. These results suggest that long-term memory representations can modulate even the pre-attentive processing of speech sounds.

Pattern recognition of the native vowels has also been demonstrated by recording MMNs to multiple tokens of both standard and deviant stimuli. One particular problem in the MMN studies using only one token from each category is that the MMN to deviants may be due either to phonemic categorization or detection of any acoustic change irrelevant for phonetic distinction, such as pitch and timbre of the voice. Shestakova et al. (2002) resolved such confounding effects by using acoustically variable exemplars spoken by 150 different speakers for each of the three vowel categories (a, u, i). The MMN was generated by each deviant vowel change, thus demonstrating the pre-attentive categorization based on the long-term memory traces for vowels and the generalization of vowel-identity codes across highly variable instances as commonly encountered in the natural speech environment. In addition, the generator source was localized in the left hemisphere using MEG (Shestakova et al., 2002, 2004).

Other MMN studies have explored the language-related permanent memory traces for higher-order units (see Shtyrov & Pulvermuller, 2007 for a review). For example, memory traces for native words were probed by comparing the MMN to meaningful words with that to phonologically legal pseudowords: MMNs elicited by deviant words were larger than MMNs elicited by acoustically matched pseudowords (Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2004; Shtyrov et al., 2005). Pulvermüller & Shtyrov (2006) further suggested existence of widely distributed and strongly connected neuronal populations in the brain that accommodate memory traces for words. Further ERP and MEG studies investigated processing of morpho-syntactic information using grammatical mismatch between the pronoun and the verb suffix using pronoun-verb phrases in Finnish (Shtyrov et al., 2003) and English (Pulvermüller & Shtyrov, 2003), or between the determinant article and the noun in German (Pulvermüller & Assadollahi, 2007). Results of these experiments showed larger MMNs to deviant strings including grammatical mismatch than those that are grammatically congruous, demonstrating that even morpho-syntactic processing occurs pre-attentively as early as 100–150 ms after the critical word. These studies also located the generator source of the syntactic MMN in the left superior-temporal and the left inferior frontal cortices, in line with the localization of specific language functions in the left perisylvian cortex (Friederici & Kotz, 2003). Different from ELAN and P600, which are elicited by syntactic incongruities when subjects attend to linguistic aspects of the word strings, elicitation of the syntactic MMN does not require attention of subjects, indicating that early syntactic processing occurs automatically.

4.2.3. Näätänen's model of attentive and pre-attentive auditory processing

Study I of the present thesis was designed based on the model of auditory change detection proposed by Näätänen (1990) and then updated by Näätänen et al. (2011). The model has been developed based on a large number of previous ERP studies using the MMN paradigm, such as odd-ball paradigm and its variants, in which a sequence of auditory events is constructed as stimuli including one repetitive standard sound or standard feature embedded in many sounds and infrequent deviant sounds/features. As is shown in Figure 4.1, the relationship between the automatic processes and attention control, as well as the mechanisms linking the unconscious sound analysis to the conscious auditory perception is illustrated in the model. In this model, the N1 and MMN components are assumed to reflect different stages of pre-attentive processing, associated to two parallel routes to attention switch.

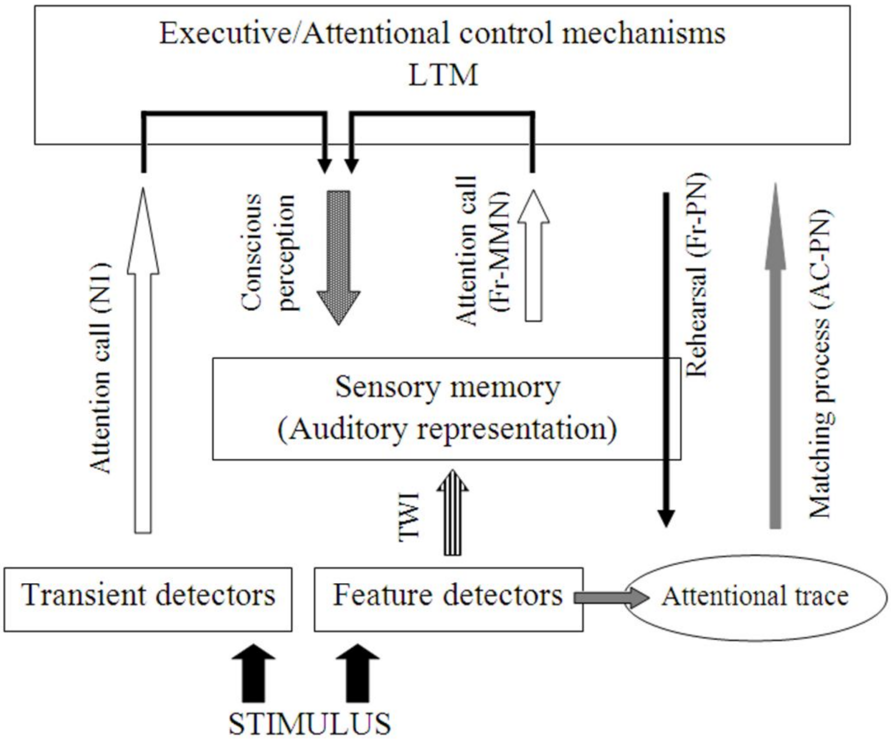


Figure 4.1. Näätänen's model of attentive and pre-attentive auditory change detection (adapted from Näätänen et al., 2011, Psychophysiology 48, 4–22, p.9)

The input sound /s/ first activates the feature-specific sensory neurons along the afferent pathway at subcortical level. These initial processes of sensory analysis are reflected by the auditory brainstem response (ABR; Picton et al., 1981; Starr & Don, 1988; Vaughan & Arezzo, 1988) and middle-latency ERP responses occurring between 8 and 50 ms after the stimulus (MLR; Picton et al., 1974).

The output signals from the subcortical mechanisms are received by the cortical Transient Detector system. The N1 component is considered to reflect transient detection of sound onset at this level and subsequent analysis of auditory features (Banai et al., 2005; Johnson et al., 2008; Kraus & Nicol, 2005). The N1 amplitude appears to reflect the magnitude of sensory stimulation received by the feature detector neurons (Näätänen, 1992).

Different auditory features of the input sound analyzed by single Feature Detectors are then integrated (Gomes et al., 1995, 1997; Takegata et al., 1999, 2001, 2005; Winkler et al., 2005) in the Temporal Window of Integration (TWI) of about 200 ms duration (Atienza et al., 2003; Näätänen & Winkler, 1999; Nousak et al., 1996; Yabe et al., 1997, 1998, 2001a, 2001b) to form a unitary auditory representation. The outputs from the TWI process are stored in the neural substrate of the Sensory Memory as a spectro-temporally integrated representation of regular features embedded in the previous sequence. The regularities represented in the sensory memory generate an expectation for the forthcoming stimulus, and the MMN is elicited by the matching process of the current stimulus with such expectation (see Section 4.2.2.1). Furthermore, and importantly, the long-term memory (LTM) and top-down processes can interact with processes at this representational level (Näätänen & Winkler, 1999). This may explain the effect of permanent language-related memory traces on the MMN for phonemes (see Naatanen, 1997, Cheour et al., 1998; Winkler et al., 1999a,b) and for higher-order units with associated semantic activation (see Pulvermüller & Shtyrov, 2006; Pulvermüller et al., 2009 for reviews).

The conscious perception of representations stored in sensory memory occurs either when the unattended deviant stimulus generates an attentional-call signal strong enough to exceed the threshold to trigger an involuntary attention switch (Näätänen, 1990, 1992), or when the stimulus features of the attended sounds are voluntarily maintained in the attentional trace. To explain the first case, two neural routes of involuntary attention switch are postulated in the model. The first route originates from the N1 transient-detector system, and the attention switch is triggered if the deviant stimulus is sensorily intrusive enough (e.g., large pitch change or a natural environmental sound versus standard tone) to exceed a certain threshold. The second route depends on the strength of the attention-call signal indexed by the frontal MMN process (Näätänen, 1990; Winkler, 2007), and the attention switch triggered by this route can bring the auditory representation in sensory memory into conscious perception.

Involuntary attention switch to unattended auditory deviant events is reflected by the P3a component, a fronto-central positivity peaking at around 250 ms after stimulus onset (Escera et al., 1998, 2000). Importantly, while the MMN indicates that the pre-attentive system has detected a deviance of the current stimulus from the predicted one, based on the regularities extracted from previous stimuli, the P3a indexes the actual attention shift towards the deviant event (Squires et al., 1975; Escera et al., 2000). The attention switch to task-irrelevant sounds as indexed by the P3a may occur without participants' awareness (van Zuijen et al., 2006; Paavilainen et al., 2007), but it can cause prolonged reaction times in the primary task (Berti et al., 2004; Escera et al., 1998).

There are still debates on the relationship between the N1/MMN generation processes and mechanisms triggering the P3a. Some previous studies have shown that the amplitudes of the N1, MMN and P3a are correlated with each other and with the magnitude of acoustic change of deviant sounds, suggesting a link between the processes subserving generation of these components (Berti et al., 2004; Escera et al., 2001; Yago et al., 2001). However, other studies have demonstrated that elicitation of the P3a is not fully governed by the N1 and MNN generation mechanisms. For example, Rinne et al. (2006) used infrequent intensity changes as task-irrelevant deviants and reported distinct effects of intensity increments and decrements: task-irrelevant intensity increments elicited an enhanced N1, the MMN and P3a, while the intensity decrements elicited only the MMN. The authors thereby suggested dissociation between the MMN and P3a processes, but rather a stronger link between the P3a and N1 generation processes. By contrast, Schröger et al. (2007) and Bendixen et al. (2007) reported dissociation between the N1 and P3a using tone sequences arranged with complex relational regularities, while their participants attended only to the tone duration. Violations of regularities in tonal relationship affected behavioral performance and elicited the MMN and P3a, but showed no effect on the N1 amplitude. These results suggest that P3a elicitation is not associated with physical sound changes underlying the N1 elicitation, but may be more linked to the MMN generator based on spectro-temporally integrated representation specifying abstract relational rules between sounds. Moreover, Ritter et al. (1999) and Sussman et al. (2003) showed that when the occurrence of deviants was made fully predictable by the use of visual cues, the P3a was not elicited, whereas the elicitation of the N1 and MMN was not affected, suggesting that the P3a is associated with a higher-order top-down control. Furthermore, the P3a to environmental sounds has been found to reflect familiarity (Cycowicz & Friedman, 1998) or personal significance (Roye et al., 2007) of the sound, thus supporting the involvement of a higher-order evaluative process.

Another case giving rise to conscious sound perception depends on the attentional selection of stimulus based on the voluntarily maintained memory representations of the to-be-attended stimulus (‘attentional traces’; Alho et al., 1987a; Näätänen, 1982), and this process is reflected by an endogenous ERP component called the Processing Negativity (PN). The PN was described by Näätänen et al. (1978) as a slow negative displacement of the ERP following the N1 peak elicited by the attended standards compared with the unattended standards. Two components of PN have been identified. The sensory-specific, auditory-cortex PN (Näätänen 1982), elicited over the temporal areas, is taken to reflect the conscious on-line comparison between the current stimulus and the attentional trace. By contrast, the frontal PN component (Näätänen & Michie, 1979) reflects the attentional control mechanisms on the attentional trace (Hansen & Hillyard, 1980, 1983, 1984; Näätänen, 1982; Okita et al., 1983). This matching process terminates rapidly for the stimulus that is very distant from the to-be-attended one, while it lasts for the lifetime of the attentional trace in the case of a perfect match (Alho et al., 1987a, b; Näätänen, 1982, 1990).

Other forms of attention control effect have also been described, such as the general arousal effect due to the subject’s vigilance level, which enhances the excitability of the Transient Detectors (Näätänen & Picton, 1987). Moreover, a channel-specific selective attention effect observed as an enhancement of the N1 elicited by a stimulus set presented to the attended ear compared to the N1 to the same stimuli presented to the unattended ear (Hillyard et al., 1973). In this case, attentional control is considered to modulate the neural excitability of the input-channel at very early stages of processing (McCallum et al., 1983; Michie et al., 1993; Rinne et al., 2008; Woldorff et al., 1991; Woldorff & Hillyard, 1991).

4.3. ERP studies revealing cross-linguistic interference in speech sound perception

As described in Section 4.2.2.2, the central phonetic representations shaped through long-term experience with one’s native language can be tapped by using the MMN. In this section, previous MMN studies that have demonstrated neural substrates of cross-linguistic differences in speech perception due to language-specific memory-traces are presented.

Näätänen et al. (1997) found that, in Finnish adult native speakers, an infrequent vowel deviant presented in a sequence of native-language vowel standards (i.e., /e/) elicited larger-amplitude MMN when it was a typical exemplar of a vowel category in the subject’s native language (Finnish vowels /o/ and /ö/) than when it was not (Estonian vowel /õ/). This finding was not linked to the acoustic distance, which was larger for Estonian contrast (/õ/-/e/) than for Finnish contrast (/ö/-/e/).

Using the same stimuli as in Näätänen et al. (1997) in infants, Cheour et al. (1998) demonstrated how language-specific memory traces develop and interact with acoustic sound processing. They recorded MMNs to Finnish and Estonian vowel contrasts in Finnish 6-month-old infants and in Finnish and Estonian 12-month-olds. Their results showed that in Finnish 6-month-olds MMN amplitude to deviants reflected only their acoustic distance from the standard stimuli, whereas MMNs obtained from Finnish and Estonian 12-month-olds better reflected language-specific phonetic representations, showing larger amplitude to their respective native vowel deviant. Thus, in line with behavioral results of previous developmental studies (Werker & Tees, 1984; Werker & Lalonde, 1988), the MMN demonstrated the emergence of language-dependent memory traces in the human brain before 12 months of life. Cheour et al. (2002) also found that MMNs to French vowel contrasts emerged in Finnish children at 4–6 years of age, soon after their entry to a French kindergarten.

Moreover, the cross-linguistic studies conducted by Winkler et al. (1999a, 1999b) have highlighted the influence of long-term experience with one's native language or a highly familiarized second language on pre-attentive processing of speech sounds. They recorded MMNs in adult native speakers of Finnish or Hungarian using vowel pairs that belonged to two different categories (across-category) in one language but to the same category (within-category) in the other language, and vice versa. The results showed that, in both language groups, across-category contrasts elicited larger MMNs than within-category contrasts in subjects' native language (Winkler et al., 1999a). They also found that MMNs elicited by the same deviants in Hungarian subjects fluent in the Finnish language were of similar amplitude as MMNs observed in Finnish subjects (Winkler et al., 1999b).

Taken together, these results clearly show that not only long-term experience with one's native language but also a high proficiency in a second language can have a significant impact on the perception of linguistic sounds and on the underlying neurophysiological mechanisms. Such effects of language-specific memory traces as reflected by the MMN have also been found for segmental contrasts in other language pairs, including French–Hindi (Dehaene-Lambertz, 1997), French–Japanese (Dehaene-Lambertz et al., 2000) and English–Hindi (Shafer et al., 2004).

Similar cross-linguistic perceptual effects have also been demonstrated for suprasegmental speech contrasts (Chandrasekaran et al., 2007; Nenonen et al., 2003, 2005; Tervaniemi et al., 2006; Ylinen et al., 2006; Kirmse et al., 2008; Hisagi et al., 2010). Frequency (pitch) and duration are common acoustic dimensions defining both speech prosody and non-linguistic sounds, and are exploited as phonetically distinctive cues in some languages but not in others. Previous behavioral results have evidenced enhanced sensitivity to syllabic pitch contours in tone language native speakers than in non-tonal language speakers (Lee et al., 1993; Wang et al., 1999). Further MMN study suggested that such enhancement is supported by neural substrates underlying pre-attentive sound processing: larger MMNs to difficult Mandarin-Chinese tone contrasts were elicited in native speakers of Chinese than in American-English native speakers (Chandrasekaran et al., 2007).

In quantity languages (e.g., Finnish, Japanese), vowel duration has an important role in lexical distinction. For example, in Finnish lexical items that differ only in vowel durations can have completely different meanings (e.g., /tuli/ [fire] and /tuuli/ [wind]). Similarly to the effect of tone-languages on syllabic pitch perception, native speakers of a quantity language (Finns) showed enhanced MMN amplitudes as well as higher discrimination performance for duration differences (but not for frequency changes) compared to non-quantity language speakers, such as Germans (Tervaniemi et al., 2006; Kirmse et al., 2008) and Russians with different grades of knowledge of the Finnish language (Nenonen et al., 2003, 2005; Ylinen et al., 2006). Furthermore, Tervaniemi et al. (2006) also demonstrated that such enhancement of sensitivity to duration in Finnish native speakers extends to non-linguistic sounds (see also Kirmse et al., 2008). It should be noted that duration processing for speech sounds differs at least partly from that for non-linguistic tones, because phonetic analysis appears to interact with acoustic analysis (Jaramillo et al., 2001; Nenonen et al., 2003; Kirmse et al., 2008). However, the previous results strongly suggest that long-term experience with the phonological system of the native language affects brain plasticity underlying basic auditory functions.

EXPERIMENTS

GENERAL SCOPE OF THE EXPERIMENTS

The following chapters present ERP studies conducted for the present thesis, which were designed to investigate cross-linguistic effects on different phonological components. Study I focused on the perception of vowels that do or do not belong to the phonemic inventory of listeners' native language. Speech stimuli belonging to four French vowel categories were presented to French native speakers (Experiment 1, Chapter 5) and Italian native speakers (Experiment 2, Chapter 6). In particular, two of the four vowel categories were also part of the phonemic inventory of standard Italian language, while the other two were not. The latter were selected so as to explore if they are processed differently according to their relationship with the phonemic inventory acquired and stored in long-term memory through experience with the native language of participants. The results from the two language groups were not compared directly, because the experiments were conducted in different laboratories, equipped with a different experimental apparatus. Nevertheless, qualitative analysis of the results obtained from the two experiments allows us to consider the perceptual and cognitive processes involved in vowel categorization and gives us a new perspective to further investigate processing mechanisms for native and foreign language speech sounds (see Chapter 7). In Study II, another experiment was conducted to investigate the perception of pitch contour of spoken sentences in different language contexts and linguistic conditions (Experiment 3, Chapter 8). Of main interest was if and how listeners exploit familiar prosodic patterns at perceptual and cognitive levels of pitch processing. Sentences in the native language of the participants (Italian), a foreign language (French) and jabberwocky (non-sense pseudolanguage with Italian-like phonological structure) were auditorily presented, and participants were required to evaluate pitch changes compared to the prosodic contour of the sentence context. Investigating processing of familiar and unfamiliar languages at different levels of speech representations should help us to explore and develop efficacious componential approaches for second-language acquisition.

STUDY I: Perception of native and non-native vowel contrasts

(Experiments 1 & 2)

5. EXPERIMENT 1: PERCEPTION OF VOWELS IN THE NATIVE LANGUAGE

5.1 INTRODUCTION

The present study aimed at investigating the neural bases of vowel perception in adult native speakers of French. Of main interest was to determine whether the effects of acoustic distance and of speaker variability on vowel perception differ in passive and active listening conditions. To this end, we compared two early ERP components, namely the N1 and the MMN, in the two attentive conditions, and also analyzed the cognitive ERPs in an active discrimination task.

The first aim of the present study was to investigate the effect of acoustic distance defined by major formant dimensions on the perception of French vowels in French native speakers. Acoustically, each vowel category is characterized by typical ranges of formant frequencies. Notably, one or two formants are sufficient for synthesizing identifiable vowels (Delattre et al., 1952). Jacobsen et al. (2004a) have found that synthetic sounds composed by F1 and F2 frequencies typical of vowels elicit MMNs similar to natural vowels. To further verify pre-attentive processing of formant information, we examined if MMN amplitudes to vowels reflect acoustic distance as defined by F1 and F2.

The second aim of the experiment was to further investigate the effects of speaker variability on the pre-attentive processing of natural French vowels. Jacobsen et al. (2004b) demonstrated that MMNs to vowel changes were elicited regardless of token variability by using 21 tokens of each vowel (/a/ or /i/) synthesized with constant F1 and F2 values but differing in fundamental frequency and intensity. Their results provide evidence that relevant formant information is separated from other variable acoustic dimensions and is used to detect vowel changes at pre-attentive level. However, natural speech sounds are more complex and more variable from one token to the other than synthetic vowels, and formant frequencies are continuous parameters that vary among speakers according to age, sex, dialectal accent, etc. How relevant parameters are extracted from variable speech sounds to produce categorical perception is therefore an important issue. Previous studies showed that MMNs were elicited by changes in vowel category even with natural rather than synthesized vowels pronounced by different speakers (Brunellière, et al., 2009, Shestakova et al., 2002). These studies clearly demonstrated that vowels were pre-attentively categorized despite large acoustic variability. However, from these studies it was not clear if the process of pre-attentive categorization was also influenced by within-category differences. In the present experiment, three speakers were employed to introduce a certain degree of within-category acoustic variability and also to assess if responses to different tokens within each category differ.

The paradigm used in the present study was the classical MMN paradigm, with a repetitive standard, and infrequent deviant sounds. The sequence of stimuli included four vowels phonemically contrastive in French and pronounced by three different speakers (9 deviants plus 3 standards). The standard vowel was /u/ and the deviant vowels were /o/, /y/ and /ø/. Importantly, the contrast /u/-/o/ is closer than both the contrasts /u/-/y/ and /u/-/ø/ in the acoustic space defined by the first two formants (i.e., F1 and F2; see Figure 5.1). These contrasts are further characterized by deviance in different acoustic dimensions: /u/-/o/ in F1, /u/-/y/ in F2 and /u/-/ø/ in both F1 and F2. We hypothesized that if listeners pre-attentively process vowel categories, MMNs should be elicited by each deviant vowel. Moreover, we also predicted that acoustically closer deviant-standard contrasts should elicit smaller MMNs than acoustically more distant contrasts (Tiitinen et al., 1994; Pakarinen et al., 2007). Of most interest was to determine whether speaker variability will also influence pre-attentive vowel processing as reflected by MMN amplitude and/or latency.

Moreover, it was also of interest in this study to investigate whether the effects of acoustic distance and of speaker variability on vowel perception differ in passive and active listening conditions. To this end, we computed the MMN in both the active and passive listening conditions. As one of few previous experiments designed to directly compare these two types of conditions, Szymanski et al. (1999) studied the effect of attention on the speech-evoked MMN, using synthesized disyllabic (VCV) speech sounds. The standard was /ibi/, and the middle consonant of deviants could differ in either of two deviant features, phoneme (/g/, /igi/) or intensity (/b/ and louder /b/). Oddball sequences were presented dichotically, and participants were asked to detect one of the deviant features, presented in one ear, ignoring the sounds presented in the other ear. The effect of attention was examined by comparing the MMNs elicited by each feature in the attended and ignored sequences when it was target and non-target. For both features, the MMN was larger to deviants presented in the attended ear than to deviants presented in the unattended ear, and this effect did not interact with task relevance (i.e., target or non-target feature). Based on these results, we expected the MMN to be larger in the active than in the passive listening condition.

In addition to the MMN, we investigated effects of interest on the N1 component, which is taken to reflect pre-representational stage of early auditory processing (Näätänen & Winkler, 1999; Näätänen et al., 2011) and also has been found to index perceptual sensitivity to some phonetic contrasts, such as /ba/-/da/ contrast (Martin et al., 1997; Whiting et al., 1998) and VOT contrast (Tremblay et al., 1998, 2001). Based on the previous results, we predicted that the amplitude of the N1 component should be larger for /y/ and for /ø/ than for /o/, because the former vowels present larger acoustic differences in formant characteristics from the standard /u/ than the latter. In addition, based on many results showing an attention effect on N1 amplitude (e.g., Näätänen & Picton, 1987; Hillyard et al, 1987) with larger N1 to attended than to unattended visual (e.g., Mangun, 1995) or auditory stimuli (e.g., Woldorff & Hillyard, 1991), we expected the N1 to be larger in the active than in the passive listening condition.

Finally, turning to the effect of speaker variability, Kaganovich et al. (2006) recorded ERPs while participants performed a Garner selective attention task (Garner, 1974). In this paradigm, four types of speech sounds, two vowels spoken by two different talkers, were repeatedly presented in a random order, and for each sound participants were requested to identify the vowel category (vowel classification task) or the talker identity (talker classification) by focusing attention upon one of the dimensions and ignoring the other. They found an interaction between vowel and talker dimensions that was reflected in an increased sustained negativity that developed after the N1-P2 components, around 300 ms post-stimulus onset. Of interest was to determine whether similar effects on the cognitive ERPs would be found by using an active discrimination task.

5.2. METHODS

5.2.1. Participants

Fifteen volunteer participants (mean age 21.7 years, range 18–29 years; 8 women) were paid to participate in an hour and a half experimental session. All were native French speakers with normal hearing and without neurological diseases. They signed an informed consent to participate in the experiment. Three participants were excluded from data analysis, because either no MMN was found or their error rate was too high (> 77%).

5.2.2. Materials

One standard and 3 deviant vowels were chosen from the French vowel inventory. The standard vowel was the back high vowel /u/. Deviant vowels included /o/ (back mid-high vowel), /y/ (front rounded high vowel) and /ø/ (front rounded mid-low vowel; see Figure 5.1). One Duration deviant, created from the standard vowel /u/ by cutting off 40-ms middle segment and keeping intact the initial and final slopes, was also included as a control deviant for further cross-linguistic experiments. Naturally spoken vowels were used as stimuli. Three female native French speakers produced four syllables (/bu/, /bo/, /by/ and /bø/) several times. The best token for each vowel and for each speaker was selected by experienced researchers to serve as experimental stimuli (see

Table 5.1 for the F1 and F2 values, and Figure 5.2 for waveforms and spectrograms of the stimuli). From each syllable, the initial 90 ms of the vowel was extracted and its intensity level was normalized (with a linear slope of 15 ms at the beginning and at the end of the sound). Figure 5.1 illustrates how the deviant vowels differ from /u/ in the two-formant (F1 and F2) acoustic space. The acoustic distance between each deviant vowel and the standard is represented by length of the segment linking the respective centers of gravity (expressed in Bark). As can be seen, acoustic distance is much smaller for /o/ (1.16) than for /y/ (6.10) and for /ø/ (4.42).

The 12 different deviants (four deviant categories by three speakers) with the same probability of occurrence (2% each) were combined with three different standards (the vowel /u/ produced by three speakers) that occurred with the same probability of occurrence (each standard: 25.3%). A total of 2250 stimuli, including 540 deviants (45 per deviant category) and 1710 standards (570 per speaker) were used. Within the sequence, stimuli were presented in a pseudorandom order with the constraints that no more than two consecutive stimuli were spoken by the same speaker and that two deviants from the same category were not presented consecutively. The sound onset asynchrony (SOA) was 600 ms. The Passive condition lasted for 23 minutes (3 blocks of 7 min 40 sec each) and the Active condition lasted for 24 minutes (6 blocks of 4 minutes each). More blocks were used in the Active than in the Passive condition because results of preliminary tests revealed that participants had difficulties maintaining their attention focused for more than 4 min.

Vowel	Speaker	F1	F2
/u/	1	2.67	8.00
	2	2.87	6.44
	3	2.51	6.71
/o/	1	3.79	7.58
	2	3.88	6.56
	3	3.88	7.12
/y/	1	2.41	13.27
	2	2.41	13.08
	3	2.19	13.08
/ø/	1	4.15	10.92
	2	3.79	11.70
	3	3.79	11.30

Table 5.1. The F1 and F2 values (in Bark) of the 4 French vowels (/u/, /o/, /y/ and /ø/), each pronounced by 3 speakers.

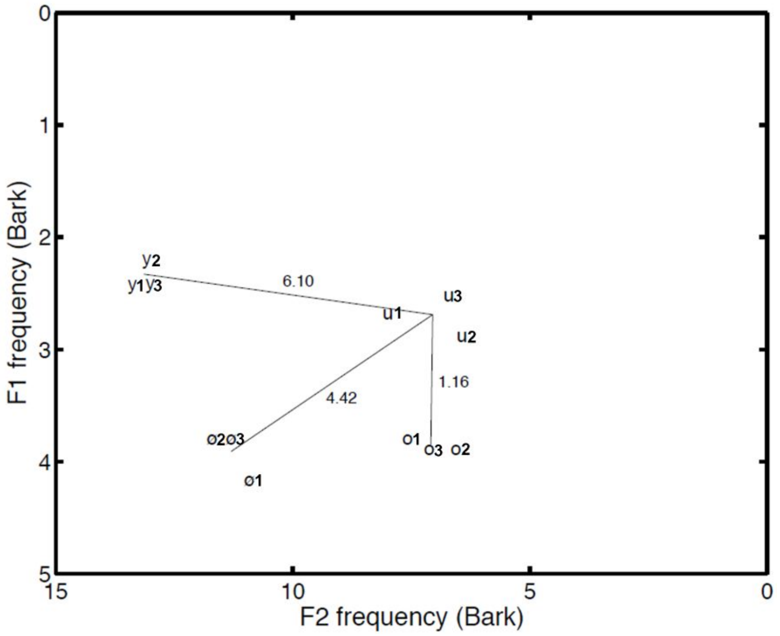


Figure 5.1. Acoustic space representing the F1 and F2 values of the 4 French vowels used in the experiment: /u/, /o/, /y/ and /ø/, each pronounced by 3 speakers. The center of gravity relative to each vowel was computed from the 3 samples. The 3 black lines link the center of gravity for /u/ with the centers of gravity for /y/, /ø/ and /o/. The length of each line is indicated in Bark.

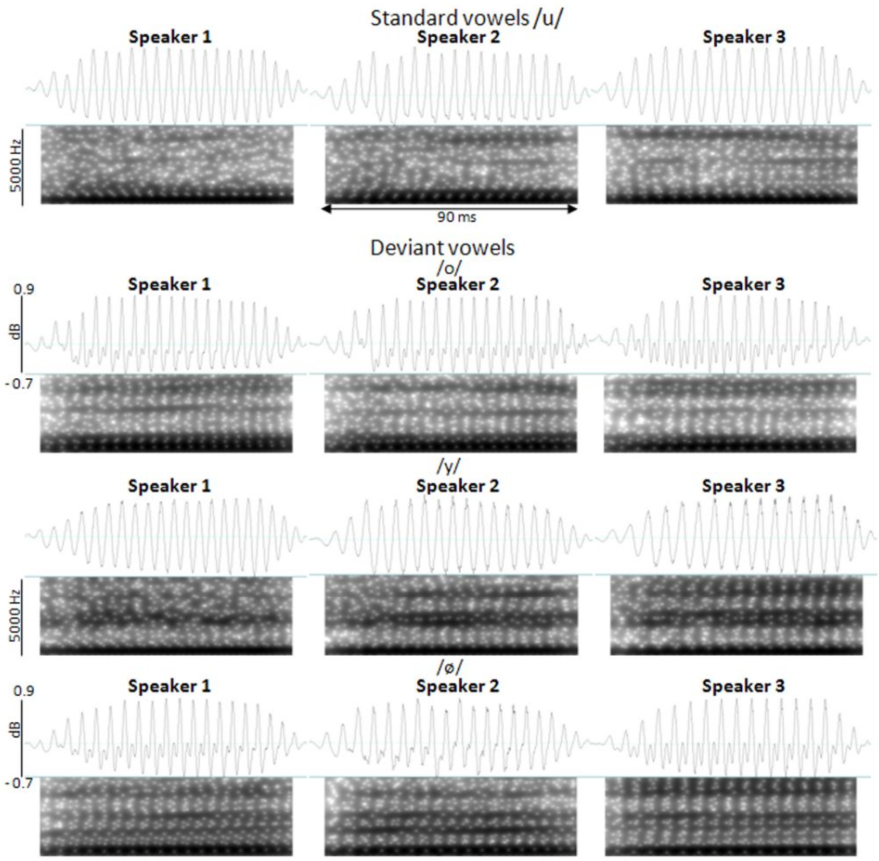


Figure 5.2. Waveforms and spectrograms of the vowels. Four French vowels (/u/, /o/, /y/ and /ø/) were produced by 3 different female speakers.

5.2.3. Procedures

Participants sat in a comfortable chair in a Faraday shielded room. The experiment comprised two conditions. In the Passive condition, participants watched a silent subtitled movie displayed on a computer screen and were asked not to pay attention to the sounds presented through headphones. In the Active condition, participants carried out a deviant detection task, pressing a response button each time they heard a vowel deviant. To avoid any carryover effects of attention (e.g., see Näätänen et al. 1993b), the Passive condition always preceded the Active condition.

5.2.4. EEG recordings and data analysis

In both the Passive and Active conditions, EEG data were recorded using an ActiveTwo system (Biosemi, Amsterdam, Netherlands) at a sampling rate of 512 Hz from 32 active electrodes attached to an Electrode cap (Biosemi Pintype) at standard positions of the International 10/20 System (Jasper, 1958): Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F7, F8, F3, F4, Fc5, Fc6, Fc1, Fc2, T7, T8, C3, C4, Cp5, Cp6, P3, P4, PO3, PO4, P7, P8, O1, O2. Moreover, to detect horizontal eye movements and blinks, the electrooculogram (EOG) was recorded from Flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Three additional electrodes were placed on the left and right mastoids and on the nose. The EEG was re-referenced off-line to the algebraic average of the left and right mastoids or to the nose recording, and filtered with a bandpass of 1–30 Hz (12 dB/oct). The electrode on the nose was used as a common reference to verify the mastoid polarity reversal of the MMN (see Näätänen et al, 2007 for a review). Recordings were segmented into 700 ms epochs (from -100 ms until 600 ms post-stimulus onset). Epochs with electric activity exceeding baseline activity by 75 µV were considered as artifacts and rejected from further processing.

Behavioral data analysis

In the Active condition, responses within 200 and 600 ms from deviant onset were considered as correct detections. Mean percentages of errors (% ER) and Reaction Times (RTs) for correct responses were computed for the 3 deviant vowels for each speaker. A two-way repeated-measures ANOVA including Vowel deviant (/o/, /y/ and /ø/) and Speaker (1, 2 and 3) as within-subject factors was carried out on both % ER and RTs.

ERP data analysis

To study the effect of attention on the early stages of vowel processing, the N1 and the

MMN IN the Passive and the Active conditions were compared. In order to be able to compare the two data sets, the electrophysiological data were averaged and analyzed for each participant using the mastoid off-line referenced data that typically show larger MMNs and better signal to noise ratio than nose-referenced data (see Schröger, 1998; Kujala et al., 2007 for reviews). However, we also computed the MNN by using the nose reference to verify the typical MMN inversion between Fz and Cz and the mastoids electrodes that differentiates the MMN from other negative components such as the N1 and N2 (see Näätänen et al, 2007 for a review).

The grand-average ERP waveform was computed in the two attention conditions for the standard /u/ and for each deviant category (vowels /o/, /y/ and /ø/) and separately for each speaker. N1 amplitude was computed as the mean amplitude between 100 and 190 ms after stimulus onset relative to 100 ms baseline.

For the MMN, difference waveforms were computed in the Active and Passive conditions for each participant by subtracting the ERPs elicited by the standards from those elicited by the deviants. Grand-average difference waveforms were computed for each deviant category (vowels /o/, /y/ and /ø/) and separately for each speaker. MMN amplitude was measured as the most negative peak between 150 and 250 ms post-deviant onset at Fz. Mean amplitudes at each electrode for each subject were computed in a 50 ms window centered at the peak as determined from the grand average. The latencies of MMN maximum amplitudes were determined for each subject at Fz between 110 and 210 ms for the deviant vowel /y/ and between 130 and 230 ms for the other deviant types to take the variability of latencies into account.

For N1 and MMN amplitudes, five-way repeated measures ANOVAs were computed including Attention (Passive and Active), Vowel (/o/, /y/ and /ø/)[2], Speaker (1, 2 and 3), Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; and right: F4, C4, P4) and Anterior-Posterior locus (frontal, central and parietal) as within-subject factors. For MMN latency at Fz, three-way repeated measures ANOVAs were conducted including Attention (Passive and Active), Vowel (/o/, /y/ and /ø/) and Speaker (1, 2 and 3) as within-subject factors.

To further analyze ERPs in the Active condition, the N2 and P3 components were measured in the 190–300 ms and 300–400 ms latency bands, respectively. Within each latency window, four-way repeated measures ANOVAs were computed including Vowel (/o/, /y/ and /ø/), Speaker (1, 2 and 3), Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; and right: F4, C4, P4) and

---

[2] ANOVAs were also conducted including the 4 deviant categories (3 deviant Vowels and Duration). Because similar results were obtained in these analyses and in the ANOVAs including only the 3 deviant vowels, we only report these results to simplify presentation.

Anterior-Posterior locus (frontal, central and parietal) as within-subject factors.

Greenhouse-Geisser corrections were applied to all p values and Tukey tests were used for all the post-hoc comparisons computed to determine the source of significant interactions.

5.3. RESULTS AND DISCUSSION

[Behavioral data]

Error rates (Figure 5.3, left)

The mean error rate was significantly higher for /o/ (10.3%) and /y/ (9.0%) than for /ø/ (4.8%; main effect of Vowel: F (2,22) = 7.57; p < .01). The Vowel by Speaker interaction was also significant (F (4,44) = 5.19; p < .01). Planned comparisons revealed a significantly higher error rate for /o/ produced by Speaker 2 than for /o/ produced by Speaker 1 (p < .05) and by Speaker 3 (p < .05).

Reaction Times (Figure 5.3, right)

Mean RTs were significantly longer for /o/ (405 ms) than for /y/ (388 ms) and for /ø/ (383 ms; main effect of Vowel: F (2,22) = 13.30; p < .001). The Vowel by Speaker interaction was also significant (F (4,44) = 16.07; p < .001). Planned comparisons showed that RTs were significantly longer for /o/ produced by Speaker 2 than for /o/ by Speaker 1 (p < .001), and for /y/ produced by Speaker 1 than for /y/ produced by Speaker 2 (p < .01) and by Speaker 3 (p < .001).

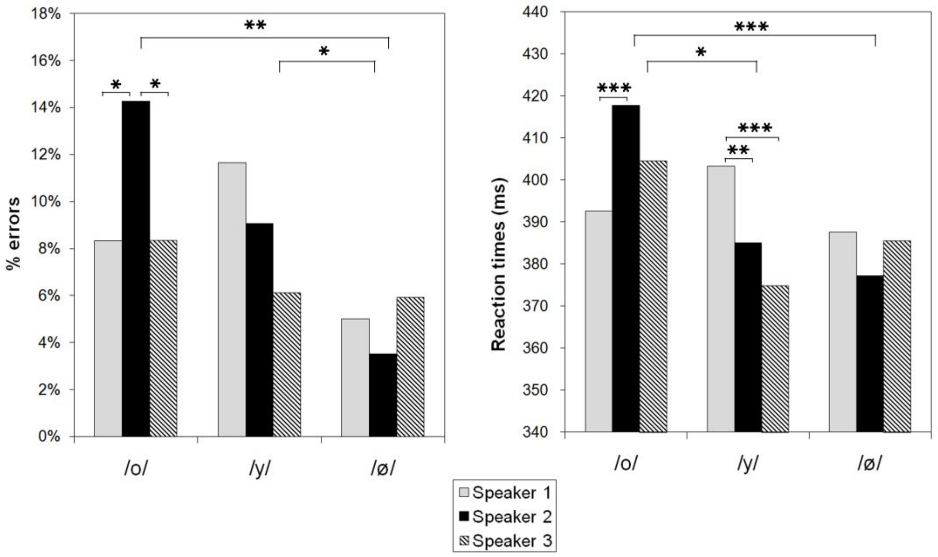


Figure 5.3. Error rates (left) and mean Reaction Times (RTs; right) in the vowel deviant discrimination task for the 3 deviant vowels (/o/, /y/ and /ø/) produced by Speaker 1 (grey), Speaker 2 (black) and Speaker 3 (hatched line). (Experiment 1)

Overall behavioral data in the active discrimination task showed acoustic distance effect on the discriminability of deviants. Although all vowels belonged to distinct phonemic categories in listeners' native language, their level of discrimination difficulty varied as a function of acoustic distance: the closest the deviant vowel from the standard /u/, the longest the RTs and the highest the error rate. Moreover, lower error rate and shorter RTs for /ø/ than the other vowels indicates that both F1 and F2 dimensions contribute to the perceptual discriminability. Finally, the Vowel by Speaker interaction revealed that /o/ produced by Speaker 2 (in error rate and RTs) and /y/ by Speaker 1 (in RTs) were more difficult to discriminate than the other tokens from the same categories, but no significant difference between speakers was found for /ø/. This may be due to uncertainty in determining a boundary between contrasts that are phonetically close and therefore difficult to discriminate. Considering that formants are relatively stable across tokens for each category (see Table 5.1 and Figure 5.1), it may be that discrimination of contrasts determined by more formants is more resistant to variations in speaker-specific properties than that of contrasts differing by one formant.

[**Electrophysiological data]**

N1 component

Figures 5.4 and 5.5 illustrate the grand-averaged ERP waveforms in the Passive and Active conditions at selected 9 electrodes. In both conditions, an early centro-parietally distributed negative component (N1), peaking between 100–190 ms after stimulus onset, was elicited by the standard and deviant vowels.

Results of the ANOVAs on N1 amplitude (see Table 5.2) revealed a significant main effect of Attention ($p < .01$), with larger N1 component in the Active (-1.63 μV) than in the Passive condition (-0.27 μV). The N1 was larger at central (-1.08 μV, $p < .05$) and at parietal (-1.35 μV, $p < .01$) sites than at frontal sites (-0.42 μV; main effect of Ant-Post locus, $p < .05$). The main effect of Vowel was also significant ($p < .001$), with smaller N1 for /o/ (-0.21 μV) than for /y/ (-1.50 μV, $p < .001$) and for /ø/ (-1.15 μV, $p < .001$). Interestingly, the Vowel by Speaker interaction was significant ($p < .05$). Planned post-hoc comparisons revealed that the N1 was significantly smaller for /o/ by Speaker 2 than for /o/ by Speaker 3. Importantly, further planned analyses revealed that the Vowel x Speaker interaction was significant in the Active listening condition ($p < .05$) but not in the passive condition ($F(4,44) = 1.75$, $p = .17$).

Table 5.2. Results of the ANOVAs on N1 amplitude in the 100–190 ms latency band after stimulus onset in the Passive and Active listening conditions.

Condition	Effects	F(df)	p
Passive vs. Active	Attention	F(1,11) = 12.03	**
	Vowel	F(2,22) = 18.73	***
	Ant-Post	F(2,22) = 8.44	*
	Vowel*Speaker	F(4,44) = 4.60	*
	Lateral*Ant-Post	F(4,44) = 4.01	*
Passive	Vowel	F(2,22) = 8.54	**
Active	Vowel	F(2,22) = 10.14	**
	Ant-Post	F(2,22) = 8.78	*
	Vowel*Speaker	F(4,44) = 3.50	*
	Lateral*Ant-Post	F(4,44) = 3.33	*

Attention (Passive and Active), Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: $p < .001$; **: $p < .01$; *: $p < .05$

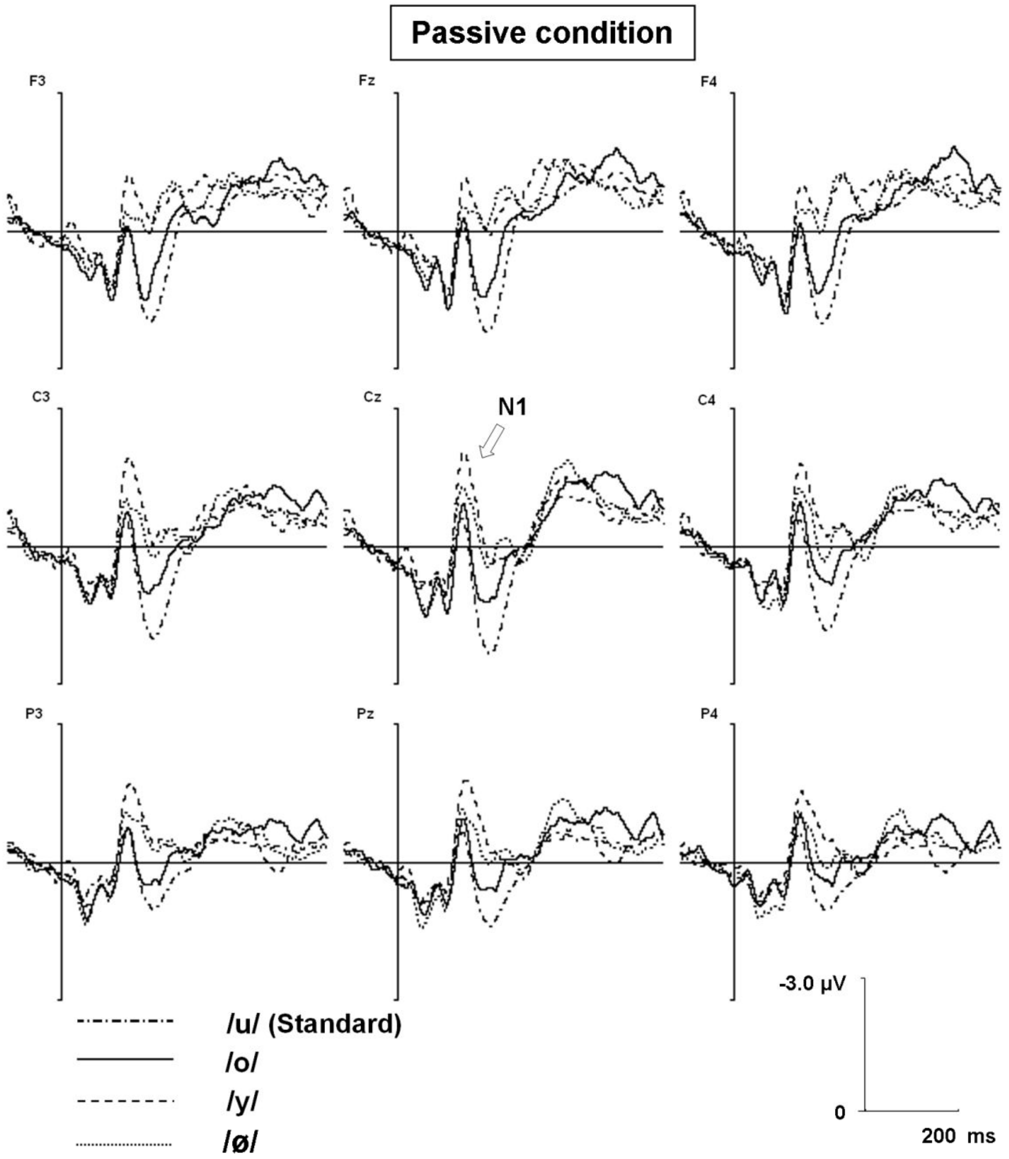


Figure 5.4. Grand-averaged ERP waveforms in the Passive condition in Experiment 1 for the standard /u/ (dot-dashed line) and for the 3 deviant vowels: /o/ (solid line), /y/ (dashed line) and /ø/ (dotted line). All speakers are included.

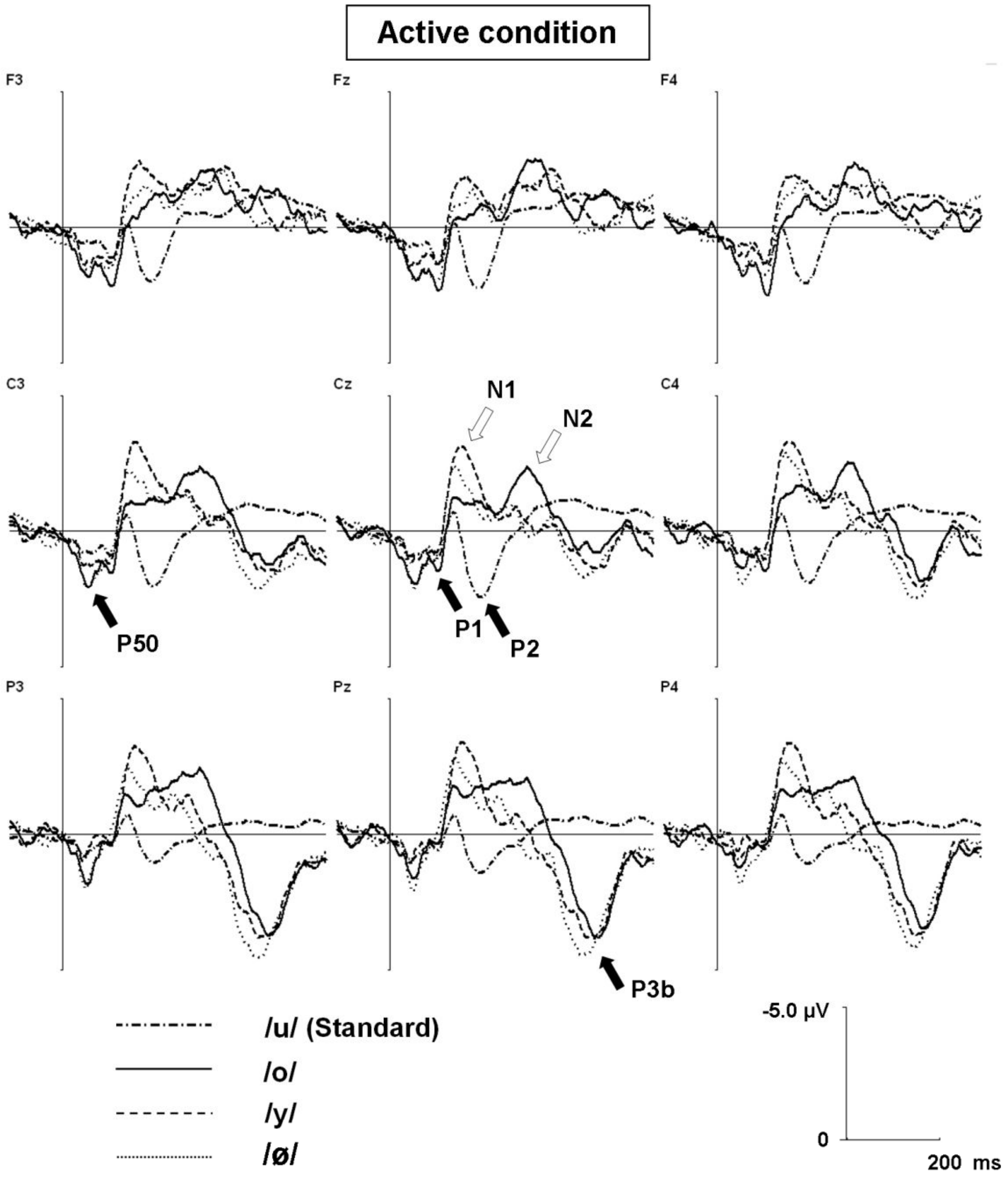


Figure 5.5. Grand-averaged ERP waveforms in the Active condition in Experiment 1 for the standard /u/ (dot-dashed line) and for the 3 deviant vowels: /o/ (solid line), /y/ (dashed line) and /ø/ (dotted line). All speakers are included.

The above results show that N1 amplitude varied between the different deviant vowels as a function of their acoustic distance from the standard in the F1-F2 space, with larger N1 to /y/ and /ø/ than to /o/. This is consistent with previous studies showing sensitivity of N1 amplitude to the magnitude of physical change in the auditory environment (Näätänen & Picton, 1987) and to spectral dissimilarities between vowels (e.g., Obleser et al., 2003; Diesch & Luce, 2000). Smaller N1 amplitude to deviants acoustically similar to the standard can be accounted for by the refractoriness of neuronal population of the N1 generator responding to repetitive features presented by the standard (Näätänen & Picton, 1987). In other words, if the vowel /o/ shares more acoustic features with the standard /u/ than the vowels /y/ and /ø/, physiological refractoriness involves more neural populations affecting N1 amplitude.

Moreover, N1 amplitude was also THE smallest to /o/ pronounced by Speaker 2, which was the most difficult token in the discrimination task as confirmed by behavioral data, and this effect was found only in the Active condition. This may indicate enhancement of sensitivity of feature-detector neurons due to attentional control (Näätänen et al., 2011). However, no similar effects were found on N1 amplitude for /y/ pronounced by Speaker 1, which was also found more difficult than the other tokens of the same category in behavioral measures.

Furthermore, the N1 showed an attention effect with larger amplitude in the Active than in the Passive condition (e.g., Näätänen & Picton, 1987; Hillyard et al, 1987). The enhancement of N1 amplitude in the Active condition may be partially due to higher level of arousal (Näätänen & Picton, 1987), but it may also be attributable to superimposition of another component reflecting selective attention, such as the processing negativity (Näätänen, 1982).

Mismatch negativity (MMN)

Figures 5.6 and 5.7 illustrate the grand-averaged difference waveforms (deviant-standard) in the Passive and Active conditions, respectively, at selected 9 electrodes. MMNs showed largest amplitude at fronto-central sites with smaller amplitude to /o/ than to both /y/ and /ø/.

Results of the ANOVA on MMN amplitude (see Table 5.3) revealed significant main effect of Attention ($p < .05$), with larger MMN in the Active (-2.70 μV) than in the Passive (-1.47 μV) condition. The main effect of Vowel was also significant ($p < .01$). Results of post-hoc Tukey tests showed that the MMN was smaller for /o/ (-1.44 μV) than for /y/ (-2.64 μV, $p < .001$) or for /ø/ (-2.19 μV, $p < .05$). The Vowel by Speaker interaction was not significant.

Table 5.3. Results of the ANOVAs on MMN amplitude in the Passive and Active listening conditions

Condition	Effects	F(df)	p
Passive vs. Active	Attention	F(1,11) = 8.58	*
	Vowel	F(2,22) = 11.37	**
	Attention*Ant-Post	F(2,22) = 4.49	*
	Lateral*Ant-Post	F(4,44) = 3.29	*
Passive	Vowel	F(2,22) = 5.66	*
	A-P	F(2,22) = 8.23	*
Active	Vowel	F(2,22) = 5.33	*
	Lateral*Ant-Post	F(4,44) = 3.50	*

Attention (Passive and Active), Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: p < .001; **: p < .01; *: p < .05

Results of the ANOVAs on MMN peak latency at Fz (see Table 5.4) showed a main effect of Vowel (p < .001). Post-hoc Tukey tests showed that mean MMN latency was longer for /o/ (176 ms, p < .001) and for /ø/ (170 ms, p < .01) than for /y/ (161 ms, p < .001). The Attention by Vowel interaction was also significant (p < .05) showing that MMN latency was significantly shorter in the Active than in the Passive condition only for /ø/ (163 ms vs. 176 ms, p < .01). The other main effects and interactions were not significant.

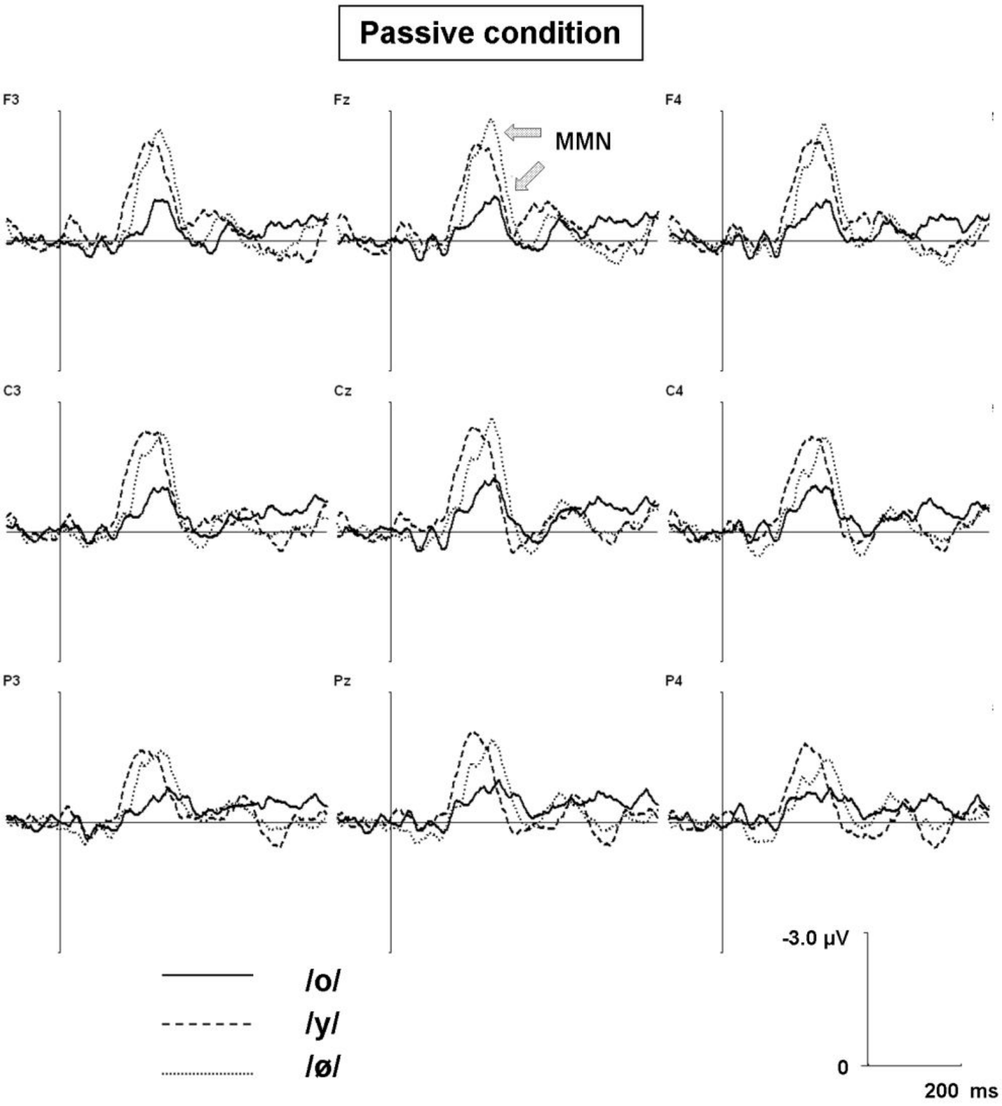
Table 5.4. Results of the ANOVAs on MMN latency at Fz in the Passive and Active listening conditions.

Condition	Effects	F(df)	p
Passive vs. Active	Vowel	F(2,22) = 17.82	***
	Attention*Vowel	F(2,22) = 5.25	*
Passive	Vowel	F(2,22) = 14.75	***
Active	Vowel	F(2,22) = 6.58	*

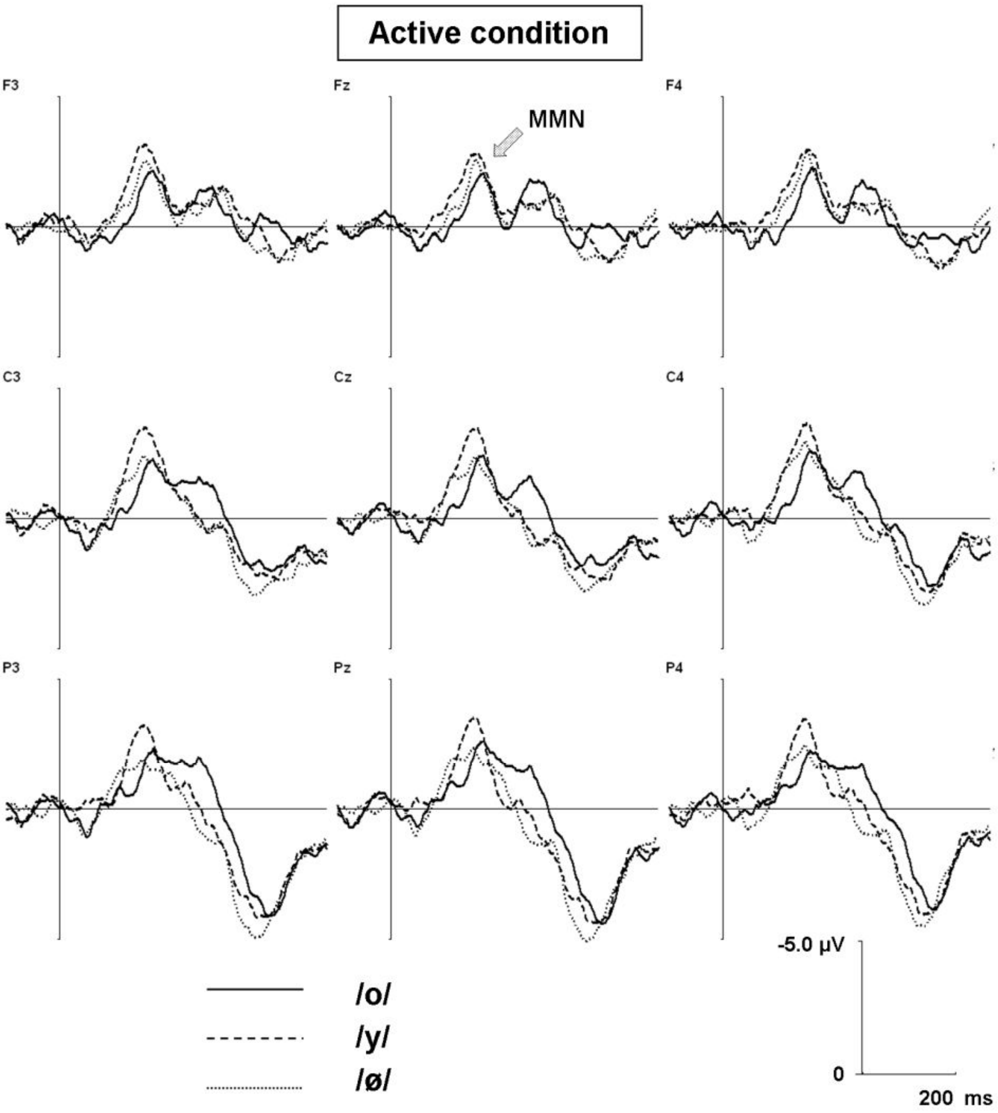
Attention (Passive and Active), Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: p < .001; **: p < .01; *: p < .05



Figures 5.6. Grand-averaged ERP difference waveforms (deviant minus standard) in the Passive condition in Experiment 1 for the deviant vowels /o/ (solid line), /y/ (dashed line) and /ø/ (dotted line). All speakers are included.



Figures 5.7. Grand-averaged ERP difference waveforms (deviant minus standard) in the Active condition in Experiment 1 for the deviant vowels /o/ (solid line), /y/ (dashed line) and /ø/ (dotted line). All speakers are included.

In line with previous studies (Jacobsen et al., 2004b; Shestakova et al., 2002), the MMN was elicited by each deviant vowel category independently of within-category acoustic variability due to speakers (no main effect of Speaker and no Vowel by Speaker interaction) in both attentional conditions, demonstrating pre-attentive categorization based on long-term phonetic representations of the native phonemic inventory. While the MMN to within-category contrasts is considered to reflect auditory sensory representations, the MMN to across-category contrasts is typically associated to categorical phonetic representations (e.g., Winkler et al, 1999a). Based on this account, our results further demonstrate the adaptive value of the language system that allows for basic and relevant linguistic units such as vowels to be perceived and categorized independently of the variability of their specific acoustic features. Moreover, and importantly, the Vowel by Speaker interaction was significant on N1 amplitude in the Active listening condition, but not in the Passive condition. Taken together, these results suggest that within-category acoustic changes are perceived through an attention-dependent detection process, but that they are automatically assimilated to vowel categories through a memory-based comparison process, as reflected by the MMN.

Moreover, MMN amplitude was modulated by the acoustic distance between the deviant vowel and the standard: MMNs were smaller for /o/ than for the other deviant vowels (i.e., /y/ and /ø/). Similar acoustic distance effect was also demonstrated by MMN latency, which was longer for /o/ than for /y/. However, MMN latency for /ø/ was as long as for /o/ in the Passive condition and as short as for /y/ in the Active condition. These results further suggest that F1 and F2 are processed differently by the MMN generator, and that MMN elicitation is influenced more strongly by F2 than by F1, which should be verified in further experiments by systematically varying F1 and F2 values.

Regarding the attention effect, the MMN showed larger amplitude in the Active than in the Passive listening condition, which is in line with previous results showing larger MMN when the stimuli were attended to, than when they were ignored (Woldorff et al., 1998). However, the increase in MMN amplitude in the Active listening condition may also reflect the overlap of N2b component on the MMN (Näätänen, 1995). In the present data, while the MMN in the passive listening condition was largest at frontal sites, the MMN recorded in the active listening condition was distributed more posteriorly, compatibly with the typical centro-parietal distribution of the N2b (Novak et al., 1992; Sussman et al., 2003).

Later ERPs in the Active listening

Further analyses were conducted in the active listening condition to investigate the N2 and P3 components (see Table 5.5 and Figure 5.5). Between 190 ms and 300 ms, the N2 component for

/o/ was larger than for /y/ and for /ø/ at centro-parietal sites (main effect of Vowel, $p < .05$ and Vowel by Ant-Post interaction, $p < .01$). The Vowel by Speaker interaction was almost significant ($p < .06$) with larger N2 for /o/ produced by Speaker 3 than by the other speakers.

Between 300 ms and 400 ms and at parietal sites, the P3b component was significantly smaller for /o/ (1.89 µV) than for /ø/ (3.18 µV, $p < .001$) and for /y/ (2.75 µV, $p < .001$) (Vowel by Ant-Post interaction, $p < .05$ and main effect of Ant-Post locus, $p < .001$). The main effect of Speaker was significant ($p < .05$) with overall larger P3b for Speaker 2 (1.19 µV) than for Speaker 1 (0.40 µV, $p < .05$). The P3b was of intermediate amplitude for Speaker 3 (1.00 µV) and not significantly different from the other two speakers.

Table 5.5. Results of the ANOVAs on N2 and P3 amplitude in the Active condition measured in the 190–300 ms and 300–400 ms latency windows, respectively.

Latency (ms)	Effects	F(df)	p
190–300	Vowel	F(2,22) = 3.90	*
	Vowel*Speaker	F(4,44) = 2.73	°
	Vowel*Ant-Post	F(4,44) = 7.03	**
300–400	Speaker	F(2,22) = 3.56	*
	Ant-Post	F(2,22) = 26.46	***
	Vowel*Ant-Post	F(4,44) = 4.83	*

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

According to previous studies, N2 amplitude is influenced by discrimination difficulty (Ritter et al., 1979, 1983; Simson et al., 1977), and the amplitude of the parietal P3b, which reflects categorization and decision processes, is negatively correlated with processing difficulty of the stimuli (Donchin & Coles, 1988; Kok, 2001 for a review). Based on these functional interpretations of the N2 and P3 components, the present results are in line with behavioral data (highest error rate and longest RTs for /o/): N2 amplitude was larger and P3 amplitude was smaller for /o/ than for /y/ and /ø/. Therefore, both behavioral and electrophysiological data demonstrate the effect of acoustic distance on attentive vowel processing.

Moreover, the /o/ pronounced by Speaker 3 was associated with larger N2 components than the /o/ pronounced by the other speakers, but this finding did not correspond to behavioral data. Although it is not possible to identify the specific feature responsible for the N2 enhancement, it would be of interest to specify the acoustic parameters that increase the intelligibility of the basic speech units.

In sum, overall results of Experiment 1 showed that acoustic distance influenced the amplitude and latency of the MMN elicited by familiar vowels, as well as the level of performance and the pattern of ERPs in the active discrimination task. Most importantly, while speaker variability did not influence the MMN in the passive listening condition, within-category acoustic differences were reflected by the results only in the active listening condition. These findings suggest that vowels are pre-attentively categorized and that within-category acoustic differences are processed at the attentive level.

6. EXPERIMENT 2: PERCEPTION OF NATIVE AND NON-NATIVE VOWEL CONTRASTS

6.1 INTRODUCTION

Experiment 2 was conducted to explore the impact of the long-term experience with one's native language on the brain plasticity subserving auditory vowel perception by using the same paradigm and speech materials as Experiment 1, but enrolling Italian adult native speakers as participants.

Among the vowels used in the present study, the standard vowel (/u/) and one of the deviants (/o/) are vowel categories of Italian. By contrast, the other two deviant vowels (/y/ and /ø/) are not distinctive categories in standard Italian language, but they are typical exemplars of distinctive categories only in some regional dialects (e.g., Lombard and Piemontese). We defined the deviant /o/ as 'native', and /y/ and /ø/ as 'non-native' vowels, based on the phonemic categories of standard Italian language. In acoustic terms, /y/ is intermediate between two Italian vowels, /u/ and /i/, and /ø/ is intermediate between /o/ and /e/, respectively differing in the F2 dimension. Importantly, acoustic distance from the standard vowel in the F1-F2 space is smaller for the native vowel /o/ than the non-native /y/ and /ø/ (see Figure 5.1). For this reason, if participants find the non-native deviants more difficult to discriminate than the native /o/, this cannot be attributable to acoustic distance. In terms of articulatory phonology, the non-native vowels belong to the class of 'front-rounded' vowels, which does not exist in the Italian vowel inventory. The native deviant /o/ and the non-native /y/ differ from the standard /u/ in one phonetic feature, 'tongue height' and 'place of articulation', respectively, while /ø/ differs from /u/ in both features. From this view, the /ø/ may be phonetically easier to discriminate than /o/ and /y/.

The first aim of the second experiment was to investigate how vowels that are not distinctive in listeners' native language are processed pre-attentively compared with native vowels. It is known that deviant vowels that are exemplars of distinctive categories in participants' native language elicit larger MMNs than deviants that are not (Näätänen, 1997; Winkler et al., 1999a), reflecting the influence of long-term phonological representations on the matching process between the input sound and the acoustic representation in the sensory memory buffer. Such effect of language-specific memory traces can even override the effect of acoustic distance between the standard and deviant sounds (Cheour et al., 1998). Thus, results from the previous cross-linguistic studies (Näätänen, 1997; Winkler et al., 1999a; Cheour et al., 1998) demonstrated very clearly the impact of long-term linguistic experience on the pre-attentive vowel categorization by using finely controlled synthetic vowels.

However, different from these studies, in the present study acoustically different tokens for the standard and deviant vowels were mixed in the same sound sequence. This manipulation should make automatic categorization more difficult, because categorical vowel changes are embedded in other randomly changing acoustic dimensions (e.g., pitch and timbre). In Experiment 1, native vowels could nevertheless be pre-attentively categorized presumably by extracting formant values, which were relatively stable across tokens within each category (see Figure 5.1). In the present study, two deviants, /y/ and /ø/, show intermediate values of F1 and F2 falling between the typical ranges of standard Italian vowel categories, and these vowels may hardly be assimilated to native categories (Best, 1995, see Section 2.3.1). Moreover, the presence of formants in intermediate ranges may disturb automatic categorization based on formant information, which can be perceptually masked by other acoustic features (e.g., pitch and timbre) that randomly change in the sequence.

Of most interest, therefore, was to investigate how native and non-native vowels are represented and processed, and whether and how these representations and processes depend on acoustic distance. If pre-attentive change detection occurs on the basis of formants even in absence of long-term memory traces, MMNs to native and non-native vowels should mainly reflect acoustic distance. Moreover, if being classifiable as a native vowel has an important effect, the retrieval of long-term memory traces to these vowels should occur automatically, producing well-identifiable MMNs. Alternatively, if formants fail to be pre-attentively extracted as salient constant features, MMNs may not even be elicited by non-native deviant vowel categories. In Experiment 1 the vowel /o/, a French native vowel, elicited in French participants smaller MMN compared to the other vowels due to its short acoustic distance from /u/ (the standard stimulus). Thus, we would not expect a large MMN to this vowel, although it belongs to the Italian inventory. Neither did we expect large and clear MMNs to the other deviants, because they do not belong to the native inventory. However, we expected a difference between native and non native vowels to be reflected in other early or late ERP components.

The second aim was to assess if the effect of speaker variability depend on the presence of long-term phonetic memory traces. In Experiment 1, the results differed between speakers in the Active condition, and the difference was more evident for the deviant vowels phonetically more difficult to discriminate (/o/ and /y/). If acoustic or phonetic distance is the most important factor in determining the ease of discrimination for Italian participants as well, /o/ and /y/ should be more difficult than /ø/, and the effect of speaker variability is likely to appear on the former vowels. By contrast, the native vowel (/o/) may be easier to discriminate than the non-native vowels (/y/ and /ø/) for Italian participants, because its phonetic representation is supported by the long-term memory and therefore well defined. If this is true, the speaker variability effect should be stronger on the non-native vowels than on the native one.

The second aim was to assess if the effect of speaker variability depends on the presence of long-term phonetic memory traces. Based on previous studies regarding the categorical perception (Liberman et al., 1957, 1961; Miyawaki et al., 1975), tokens of a native phonemic category may be perceived more similar one to the other than tokens that fall between two native categories. According to the perceptual magnet effect (Kuhl, 1991), a contrast between non-prototypical instances is better discriminated than an equidistant contrast between two prototypical instances of the same native category, as instances that are close to the prototype are 'captured' by it. Notably, this effect can operate at the pre-attentive stage of auditory processing (Aaltonen et al., 1997). Although these theories have been proposed using tasks different from the discrimination task used in the present study, they both suggest that tokens of non- native vowels that fall between two native categories, which are acoustically distant from native prototypes, may be perceived as more dissimilar one from the other than those of native vowels. Therefore, the effect of speaker variability is expected to be stronger on the non-native than on the native vowels.

Finally, the third aim was to investigate the effect of attention on vowel categorization. Results of Experiment 1 suggested that selective attention enhances analysis of within-category acoustic differences, particularly for deviant categories acoustically difficult to discriminate (i.e., /o/ and /y/). In Experiment 1, all the deviants were typical tokens of distinctive categories in listeners' native language, and therefore long-term phonological representations could support pre-attentive categorization. By contrast, when the incoming deviant is a non-native vowel, if the same pre-attentive mechanism does not operate because of the lack of corresponding long-term representation, other cognitive strategies may be required to categorize it. For example, the use of conscious cognitive strategies in the active discrimination task should be reflected by late ERPs in the Active condition, and such strategies are expected to differ between vowels that do or do not belong to the native phonemic inventory.

In the present experiment, as in Experiment 1, behavioral (error rates and reaction times) and electrophysiological measures (ERPs) were obtained in active and passive listening conditions. Whereas acoustic distance in the F1-F2 space predicts better discrimination for /y/ and /ø/ than for /o/, consideration of distance on articulatory phonology predicts that /ø/ is easier to discriminate than /o/ and /y/. Behavioral data should indicate which of these factors are more important to explain discrimination of native (/o/) and non-native (/y/ and /ø/) vowel sounds.

The N1 component should be elicited by all the standard and deviant vowels, with larger amplitude in the Active than in the Passive conditions. In Experiment 1, N1 amplitude also reflected acoustic distance, and speaker variability interacted with vowel category only in the Active condition. Moreover, difference in sensitivity to vowels that do not belong to the Italian inventory may be reflected in the N1 (Tremblay et al., 1998, 2001), resulting in smaller amplitude for /y/ and /ø/ in Italian participants. However, based on the results of Experiment 1 showing strong effects of acoustic distance on N1, the effect of this dimension might be counterbalanced by the other factor – that is, whether vowels belong to the Italian inventory, thus producing a null difference among vowels.

Finally, later ERP components in the Active condition were analyzed to explore how listeners consciously treat tokens of native and non-native vowels. The N2 and P3b components were clearly observed in Experiment 1, reflecting task difficulty and stimulus discriminability (Donchin & Coles, 1988; Kok, 2001 for a review). Of interest was to verify if these task-related components reflect behavioral performance. In particular, considering that the task required mainly to discriminate deviants from the standard, possibly the distinction between native and non native vowels was a non-relevant dimension and might not affect the auditory and cognitive processing. In this case, the ERPs should mainly reflect stimulus discriminability. By contrast, based on the assumption that vowels are obligatorily categorized, cognitive processes engaged in the discrimination task might differ not only depending on acoustic distance from the standard vowel /u/, but also on whether vowels belonged to the native or to the non-native categories. Indeed, we expected that the late ERP components in the Active condition should reflect cognitive strategies that participants employ to resolve potential dilemmas in classifying vowels and in response decision.

6.2. METHODS

6.2.1. Participants

Eighteen volunteer native Italian speakers (mean age 23.6 years, SD = 2.45, range 21–30 years; 11 women) participated in 2-hour experimental session. All were with normal hearing and without neurological diseases. They signed an informed consent to participate in the experiment. Six participants were excluded from data analysis, because of high error rate (> 77%) in one or more stimulus conditions or because of artifacts in EEG data, resulting in a total of 12 participants (mean age 22.8 years, SD = 1.71, range 21–26 years; 8 women) included in the data analyses.

6.2.2. Materials

The stimuli were the same sound sequences used in Experiment 1.

6.2.3. Procedures

Procedures for the experimental blocks were the same as in Experiment 2, with the only difference that the sounds were presented to the participant through laudspeakers instead of headphones. In addition, the experimental blocks in the Active condition were preceded by a practice block, in which participants received explanation about four French vowel categories and conducted a classification task with visual feedback on the response.

6.2.4. EEG recordings and data analysis

In both the Passive and Active conditions, Electroencephalogram (EEG) was recorded from 31 scalp electrodes, mounted on an elastic cap, and located at the following sites according to International 10/20 system (Jasper, 1958): Fpz, Fz, FCz, Cz, CPz, Pz, Oz, Fp1, Fp2, F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, P4, O1, O2, F7, F8, FT7, FT8, T3, T4, TP7, TP8, T5, T6. Electrode impedances never exceeded 9 kΩ. For the purpose of artifact scoring, vertical and horizontal electro-oculograms (EOGs) were recorded. Electrode pairs (bipolar) were placed at the supra- and sub-orbit of the right eye and at the external canthi of the eyes. Three additional electrodes were placed on the left and right mastoids and on the nose. These recording sites were referenced online to the nose electrode, which was used as a common reference to verify the mastoid polarity reversal of the MMN. The EEG and EOG signals were amplified by a Neuroscan Synamp amplifier (El Paso, TX, USA) with a bandpass filter of .05–100 Hz and were digitized at 500 Hz (16 bit AD converter, accuracy 0.034 μV/LSB) and stored on a Pentium II computer.

The EEG was re-referenced off-line to the algebraic average of the left and right mastoids, and filtered with a bandpass of 1–30 Hz (24 dB/oct). Continuous EEG data were segmented into 700 ms epochs (from -100 ms until 600 ms post-stimulus onset). EEG epochs were baseline-corrected against the mean voltage during the 100-ms pre-stimulus period. Epochs with electric activity exceeding baseline activity by ±75 μV were considered as artifacts and rejected from further processing.

Behavioral data analysis

In the Active condition, responses within 200 and 600 ms from deviant onset were considered as correct detections. Mean percentages of errors (% ER) and Reaction Times (RTs) for correct responses were computed for the 3 deviant vowels for each speaker. A 2-way repeated-measures ANOVA including Vowel deviant (/o/, /y/ and /ø/) and Speaker (1, 2 and 3) as within-subject factors was conducted on both % ER and Rts.

ERP data analysis

The grand-average ERP waveform was computed in the two attention conditions for the standard /u/ and for each deviant category (vowels /o/, /y/ and /ø/) and separately for each speaker using the mastoid off-line referenced data. The difference waveforms were computed in the Active and Passive conditions for each participant by subtracting the ERPs elicited by the standards from those elicited by the deviants.

Based on the procedures used in Experiment 1, the N1 and the MMN were compared between the Passive and the Active conditions to study the effect of attention on the early stages of vowel processing.

N1 amplitude was computed as the mean amplitude relative to 100 ms baseline between 105 and 155 ms after stimulus onset.

To verify the presence of an MMN, the difference between the deviant and the standard in ERP amplitude at the Fz, where the MMN usually shows largest amplitude, was tested using one-tailed t-tests at 95% IC over the MMN latency range (150–220 ms).

In the Passive condition, the mean amplitudes of the P3a and RON components were also computed from the difference waves using the following time windows: from 220 to 300 ms for the P3a and from 300 to 380 ms for the RON.

In the Active condition, later ERPs were analyzed using the mean amplitudes in the 170–250 ms, 250–350 ms and 350–500 ms latency bands.

For N1 amplitude, 5-way repeated measures ANOVAs were computed including Attention (Passive and Active), Vowel (/o/, /y/ and /ø/), Speaker (1, 2 and 3), Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; and right: F4, C4, P4) and Anterior-Posterior locus (frontal, central and parietal) as within-subject factors. For the amplitudes of the P3a and RON in the Passive condition and other ERP components in the Active condition, within each latency window, 4-way repeated measures ANOVAs were computed including Vowel (/o/, /y/ and /ø/), Speaker (1, 2 and 3), Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; and right: F4, C4, P4) and Anterior-Posterior locus (frontal, central and parietal) as within-subject factors.

Greenhouse-Geisser corrections were applied to all p values and Tukey tests were used for all the post-hoc comparisons computed to determine the source of significant interactions.

6.3. RESULTS AND DISCUSSION

[Behavioral data]

Results of ANOVAs on error rates (see Table 6.1 and Figure 6.1, left) showed significant main effects of Vowel and of Speaker: the mean error rate was significantly higher for /o/ (13.7%, $p < .01$) and /y/ (16.0%, $p < .001$) than for /ø/ (4.9%, $p < .001$) and higher for Speaker 1 (15.7%, $p < .001$) and Speaker 2 (11.9%, $p < .05$) than for Speaker 3 (7.0%). The Vowel by Speaker interaction was also significant. Planned comparisons revealed difference between speakers for the vowel /y/ and for /ø/, but no significant difference between speakers was found for /o/. For each of the non-native vowels, the error rate was significantly higher for Speaker 1 than the other speakers.

Results of ANOVAs on reaction times (see Table 6.1 and Figure 6.1, right) showed a significant main effect of Vowel: mean RTs were significantly longer for /o/ (415 ms, $p < .001$) and for /y/ (417 ms, $p < .001$) than for /ø/ (372 ms). The main effect of Speaker was also significant: RTs were longer for Speaker 1 (421 ms) than for Speaker 2 (401 ms, $p < .001$), and for Speaker 2 than for Speaker 3 (382 ms, $p < .001$). A significant Vowel by Speaker interaction was also found. Planned comparisons revealed difference between speakers for /y/ and /ø/, but not for /o/: RTs for /y/ were significantly longer for Speaker 1 than for Speaker 2 ($p < .001$) and longer for Speaker 2 than for Speaker 3 ($p < .001$), while RTs for /ø/ produced by Speaker 1 were longer than for /ø/ produced by Speaker 2 ($p < .001$) and by Speaker 3 ($p < .01$) without significant difference between the latter vowels.

Table 6.1. Results of the ANOVAs on error rate (ER, left) and reaction times (RTs, right) in the vowel discrimination task.

	Effects	F(df)	p
ER	Vowel	F(2,22) = 12.65	***
	Speaker	F(4,44) = 12.50	***
	Vowel*Speaker	F(4,44) = 6.88	**
RTs	Vowel	F(2,22) = 29.28	***
	Speaker	F(2,22) = 58.20	***
	Vowel*Speaker	F(4,44) = 33.06	***

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: $p < .001$: **: $p < .01$: *: $p < .05$

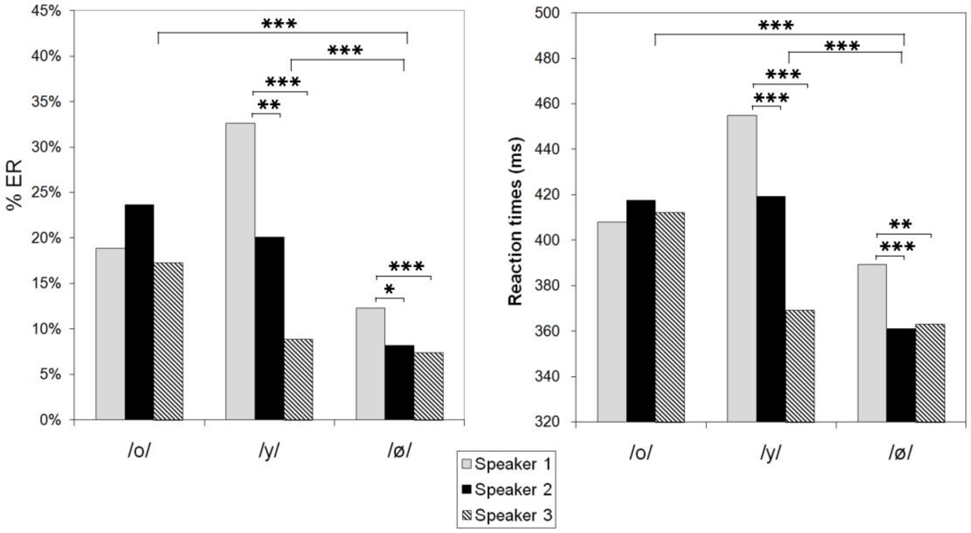


Figure 6.1. Error rates (left) and mean Reaction Times (RTs; right) in the vowel deviant discrimination task for the 3 deviant vowels (/o/, /y/ and /ø/) produced by Speaker 1 (grey), Speaker 2 (black) and Speaker 3 (hatched line). (Experiment 2)

Overall behavioral data confirmed that /ø/ was easier to discriminate than /o/ and /y/, suggesting that the difference in both F1 and F2 contributes to vowel distinction. This finding also supports the prediction based on articulatory phonology: a vowel contrast differing in two phonetic features (tongue height and place of articulation) is easier to discriminate than a contrast differing in only one feature. Moreover, significant differences between speakers emerged for the non-native vowels (/y/ and /ø/) but not for the native vowel (/o/). That is, the native /o/ was difficult to discriminate independently of the specific speaker's production, while the discriminability of /y/ depended on speaker-specific properties.

It is important to note how these results differ from the results from French native speakers. A comparison with Figure 5.3 shows that the level of performance of French participants was overall much higher, not only for the French vowels that do not belong to the Italian inventory (/y/ and /ø/), but also for the vowel that is native in both languages (/o/). This suggests that probably the instances of /o/ produced by French speakers were not really typical of the Italian homologous (Strange et al., 2007), thus adding to the difficulty to discriminate /o/ from /u/ due to the small acoustic distance. Moreover, although the level of performance of Italian participants for the non-native vowels (/y/ and /ø/) was lower than that of French participants seen in Experiment 1, as expected, discrimination was much easier for /ø/ than for /y/, suggesting that the difficulty in discriminating /y/ due to the fact that it is not a native Italian vowel added to the phonetic closeness (differing in only one phonetic feature) with /u/, further increasing discrimination difficulties. If the only factor responsible of discrimination difficulty was the fact that a vowel does or does not belong to the native inventory, we would have expected the same pattern to hold for /ø/ as well, which was not the case. Thus, /ø/ was the easiest of the three vowels, both for French and Italian participants, which suggests that phonetic closeness was the most reliable predictor of discrimination difficulty.

[Electrophysiological data]

N1 component (see Figures 6.2 and 6.3)

In both conditions, an early centro-parietally distributed negative component (N1), peaking at around 130 ms after stimulus onset, was elicited by the standard and deviant vowels.

Results of the ANOVAs on N1 amplitude (see Table 6.2) revealed a significant main effect of Attention ($p < .001$), with larger N1 component in the Active (-0.66 μV) than in the Passive condition (-0.26 μV), showing a robust attention effect, which can be attributed to a general arousal effect or to superimposition of the processing negativity on the N1 (Näätänen, 1982). The Attention by Ant-Post locus interaction was also significant ($p < .05$), showing that its amplitude was larger at central (-0.30 μV, $p < .05$) and at parietal sites (-0.38 μV, $p < .001$) than at frontal sites (-0.096 μV) in the Passive condition, while it did not differ between the three Ant-Post loci in the Active condition. The main effect of Vowel and its interaction with other factors were not significant, in line with what was suggested by behavioral data. ANOVAs conducted separately for each attentive condition confirmed the centro-parietal distribution of the N1 only in the Passive condition. Moreover, in the Passive condition the main effect of Speaker was significant, with larger N1 amplitude for Speaker 2 than for the other speakers.

Table 6.2. Results of the ANOVAs on the N1 amplitude measured in the 105–155 ms time window.

Condition	Effects	F(df)	p
Passive vs. Active	Attention	F(1,11) = 22.43	***
	Attention*A-P	F(4,44) = 5.99	*
Passive	Speaker	F(2,22) = 5.13	*
	A-P	F(2,22) = 9.95	**
	Speaker*Laterality	F(4,44) = 3.25	*
Active	ns		

Attention (Passive and Active), Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: $p < .001$; **: $p < .01$; *: $p < .05$

Mismatch negativity (MMN) (see Figures 6.4 and 6.5)

Results of t-tests revealed significantly larger negativity for the deviant compared to the standard only for Duration in the Passive and Active conditions, but not for deviant vowels. The absence of clear MMNs to deviant vowels indicates that vowels were not pre-attentively categorized.

An important finding in these results is that the N1 amplitude varied between speakers but not between vowels in the Passive condition. This may be because salience of speaker-specific auditory properties overrode formant information at this early stage of auditory processing. This interpretation also explains why the MMN was not elicited by vowel changes. It may be that pre-attentive detection system failed to automatically extract formant information to form an integrated representation based on this dimension.

Other components in ERP difference waves in the Passive condition (see Figures 6.4 and 6.5)

P3a component

In the 220–300 ms time window, a fronto-centrally distributed negativity was observed and identified as the P3a component (Donchin & Coles, 1988; Escera et al., 1998, 2000). Results of ANOVA on P3a amplitude revealed a significant main effect of Vowel ($F(2, 22) = 11.15$, $p < .001$), reflecting smaller positivity for /o/ (0.17 μV) than for /y/ (0.66 μV, $p < .01$) and for /ø/ (0.75 μV, $p < .001$). Other effects and interactions involving Vowel or Speaker were not significant.

RON (reorienting negativity)

In the 300–380 ms time window, a negative component largest at central scalp sites was elicited by /y/ and /ø/. This negativity elicited following the P3a was identified as the Reorienting Negativity (RON, Escera et al., 2000; Friedman et al., 2001 for a review). Results of the ANOVAs on RON amplitude (see Table 6.3) revealed a significant main effect of Vowel, showing that the RON was larger for /ø/ (-0.91 μV) than for /y/ (-0.53 μV, $p < .05$), while it was not elicited by /o/ (-0.005 μV). The main effect of Speaker and Vowel by Speaker interaction were also significant, reflecting that RONs elicited by /y/ and /ø/ were larger for Speaker 3 than for Speaker 1, and of intermediate amplitude for Speaker2. Moreover, the RON was largest at the Cz for /y/ and /ø/ (Laterality by A-P locus and Vowel by Laterality by A-P locus interactions) and for Speaker 2 and 3, but not for Speaker 1 (Speaker by Laterality by A-P locus interaction).

Table 6.3. Results of the ANOVAs on RON amplitude measured in the 300–380 ms time window using difference waves (deviant-standard) in the Passive condition.

Effects	F(df)	p
Vowel	F(2,22) = 21.48	***
Speaker	F(2,22) = 16.39	***
Lateral	F(2,22) = 7.48	**
A-P	F(2,22) = 6.21	*
Vowel*Speaker	F(4,44) = 3.99	*
Vowel*AP	F(4,44) = 4.06	*
Speaker*AP	F(4,44) = 5.97	**
Lateral*A-P	F(4,44) = 3.69	°
Vowel*Lateral*A-P	F(8,88) = 5.47	***
Speaker*Lateral*A-P	F(8,88) = 3.26	*

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: p < .001; **: p < .01; *: p < .05; ° p < .06

The P3a and RON components are taken to reflect an involuntary attention switch to unattended deviants that are intrusive or novel, and the subsequent redirection of attention back to the primary task, respectively (Escera et al., 2000; Friedman et al., 2001 for a review). The presence of the P3a and RON for /y/ and /ø/ suggests that non-native vowels captured attention and may have been classified as such, that is, as vowels not belonging to the native categories. In previous studies, the P3a has often been preceded by an MMN and/or an enhancement of the N1 (Berti et al., 2004; Yago et al., 2001), but these three components can be dissociated in certain experimental conditions (Rinne et al., 2006; Schröger et al., 2007; Bendixen et al., 2007; Ritter et al., 1999; Sussman et al., 2003; Horvàth et al., 2009). In the present results, the P3a and RON were elicited with no preceding MMN, suggesting that attention switch was not mediated by vowel categorization. Furthermore, elicitation of the P3a and RON to the non-native vowels was observed without enhancement of N1 amplitude, which did not differ between vowels, suggesting that attention switch was not triggered by sensory salience of auditory change, but depended rather on the evaluation of the sound at a later stage of information processing (Friedman et al., 2001 for a review). The evaluative processes underlying the P3a elicitation may regard attempts to classify vowels perceived as belonging to a continuum between two native vowel categories as closer to one or to the other. Behavioral data also supports that attention switch depended rather on classification of the vowels according to the native inventory: /ø/ was easiest to discriminate, and /o/ and was as difficult as /y/, but only /y/ and

/ø/ elicited the P3a and RON.

The P3a elicited by speech sounds was reported by Wang et al. (2005) using disyllabic pseudowords. In each experimental block, both the standard and deviants carried stress on the same position (1st or 2nd syllable), and phonetic changes (in VOT) occurred on the stressed or unstressed syllable of the infrequent deviants. Their results showed that while all deviants elicited MMNs, only the phonetic changes occurring on the stressed syllable elicited the P3a, suggesting that salient prosodic cues capture attention and probably enhance top-down control processes. From the present results, although P3a amplitude did not differ between /y/ and /ø/ nor between speakers, RONs to /y/ and /ø/ pronounced by Speaker 3, which were better discriminated, were larger than those elicited by the other speakers for each vowel category. This may support that involuntary attention switch indexed by the P3a and RON possibly mirrors allocation of cognitive resources for further top-down analysis of the stimulus (Escera et al., 2000; Friedman et al., 2001 for a review).

ERPs in the Active condition

Further analyses were conducted on the later ERP components in the active listening condition using the ANOVA.

Between 170 ms and 250 ms after stimulus onset (see Table 6.4), the P2 component for all deviant vowels and the following N2 component for /o/ were found in this latency window. The P2 was largest over parietal sites (main effect of Ant-Post locus: $p < .01$) and much smaller for /o/ than for /y/ and for /ø/, resulting in less positive mean amplitude for /o/ (0.25 μV) than for /y/ (1.18 μV, $p < .01$) and for /ø/ (1.16 μV, $p < .01$; main effect of Vowel: $p < .001$). Moreover, the main effect of Speaker was also significant ($p < .01$) with larger P2 amplitude for Speaker 3 (1.16 μV) than for Speaker 1 (0.77 μV, $p < .05$) and for Speaker 2 (0.66 μV, $p < .01$), which may correspond to the behavioral data showing better discriminability of vowels pronounced by Speaker 3 than those spoken by the other speakers. The Vowel by Speaker by Ant-Post locus interaction was also significant. Post-hoc comparison revealed that the effect of Speaker was mainly due to the vowel /o/: N2 amplitude for /o/ at frontal sites was significantly larger for Speaker 2 than for the other speakers, as was seen in Experiment 1.

Table 6.4. Results of the ANOVA computed on the mean amplitude in the 170–250 ms time window in the Active condition.

Effects	F(df)	p
Vowel	F(2,22) = 7.65	***
Speaker	F(2,22) = 8.35	**
Lateral	F(2,22) = 4.14	*
A-P	F(2,22) = 8.21	**
Vowel*Speaker*A-P	F(8,88) = 3.83	*

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: p < .001; **: p < .01; *: p < .05

Between 250 ms and 350 ms (see Table 6.5), the /o/ and /ø/ elicited a positive component prevalent at centro-parietal sites (P3), which was larger for /o/ (1.37 µV) than for /ø/ (0.76 µV, p < .01), whereas no similar positive component was identified in this latency band for /y/ (0.18 µV, [/y/ < /ø/] p < .05; main effect of Vowel: p < .001). From the latency and scalp distribution of this positivity, it is considered to be the task-related P3b. The P3 showed a shorter latency for /ø/ than for /o/, and this may be due to higher discriminability of /ø/ as reflected by the behavioral data. However, the reason for the apparent absence of the P3 for /y/ is unclear. It is possible that a small P3 component to /y/ merged with other cognitive ERP components.

The Vowel by Lateral, Lateral by Ant-Post and Vowel by Lateral by Ant-Post interactions were significant, reflecting that the pattern of ERP responses considerably varied among vowels. The P3 to /o/ was widely distributed over the scalp, whereas the P3 to /ø/ was more localized at centro-parietal sites. A negative component peaking at around 300 ms was elicited by /y/ and /ø/ with maximum amplitude at Cz and Pz. The latency of this negativity is compatible with the Phonological Mismatch Negativity (PMN; Connolly et al., 1992), which has been associated with phonological mismatch between the input and an expected target (Connolly & Phillips, 1994; Newman et al., 2003; Newman and Connolly, 2009). In the present results, a clear negative component was elicited by the non-native vowels (/y/ and /ø/) but not by the native vowel (/o/). This suggests that this component does not necessarily reflect magnitude of phonetic mismatch, but may be related with evaluation of the nature of the vowels, that is, their belonging or not to the native phonemic inventory.

In this latency window, /o/ elicited a recognizable N2/P3 complex similar to that elicited by all the deviant vowels in Experiment 1, whereas ERP waveforms for /y/ and /ø/ were quite different. For example, the P2 was much larger for /y/ and /ø/ than for /o/. The enhancement of P2 amplitude may be due to the lower familiarity of the non-native vowels compared to the native vowel, which probably required more selective attention to maintain phonetic representations for these vowels in the short-time memory (Wolach et al., 2001). Alternatively, this may reflect the perceptual discriminability due to acoustic distance from the standard, which was larger for /y/ and /ø/ than for /o/.

Table 6.5. Results of the ANOVA computed on the mean amplitude in the 250–350 ms time window in the Active condition.

Effects	F(df)	p
Vowel	F(2,22) = 20.88	***
Vowel*Lateral	F(4,44) = 3.90	*
Lateral*A-P	F(4,44) = 4.39	**
Vowel*Lateral*A-P	F(8,88) = 4.82	**

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: $p < .001$; **: $p < .01$; *: $p < .05$

Between 350 ms and 500 ms (see Table 6.6), the main effects of Vowel ($p < .01$) and of Ant-Post locus ($p < .001$) were significant. In this latency window, the N4 component was observed with largest amplitude at parietal sites (-1.28 μV) and progressively diminished towards anterior regions (central: -0.46 μV; frontal: 0.31 μV). N4 amplitude was larger for /ø/ (-0.91 μV) than for /y/ (-0.09 μV, $p < .01$) and was intermediate for /o/ (-0.44 μV), without significant difference of /o/ from the other two vowels. This component was relatively lateralized to the right (the main effect of Laterality: $p < .06$). The Vowel by Speaker and Vowel by Speaker by Lateral interactions were also significant. In particular, /y/ and /ø/ by Speaker 3 elicited larger N4 components than the same vowels pronounced by the other speakers.

The functional signification of this parietal N4 component is not clear. Connolly et al. (1992) and Connolly & Phillips (1994) functionally separated the PML and the N400 crossing phonological and semantic mismatches in target-words: the PMN reflected phonological processing at the pre-lexical stage, and the N400 was related with semantic violation of expectations developed by the preceding context. However, the N4 observed in the present results may be different from the "semantic" N400 component, which usually shows a central scalp distribution. Moreover, semantic processing is unlikely to occur with the isolated vowels used in the present experiment, although the present task may have involved categorization of each input vowel, which can be seen as a "semantic" task. Considering the correlation of its amplitude with discrimination performance (better discrimination and larger N4 to /ø/ than to the other vowels, and to /y/ and /ø/ by Speaker 3 than to those by the other speakers), the N4 may reflect a conscious analytic process evaluating the differences among vowels (in category, in acoustic distance, and as tokens reflecting speakers' differences) and between the deviant vowels and the standard.

Table 6.6. Results of the ANOVA computed on the mean amplitude in the 350–500 ms time window in the Active condition.

Effects	F(df)	p
Vowel	F(2,22) = 5.65	*
Lateral	F(2,22) = 3.28	°
A-P	F(2,22) = 53.78	***
Vowel*Speaker	F(4,44) = 5.19	**
Vowel*Speaker*Lateral	F(8,88) = 2.54	*

Vowel (/o/, /y/ and /ø/); Speaker (1, 2 and 3); Laterality (Right, Midline and Left); Anterior-posterior (Frontal, Central and Parietal)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

***: $p < .001$; **: $p < .01$; *: $p < .05$; °: $p < .06$

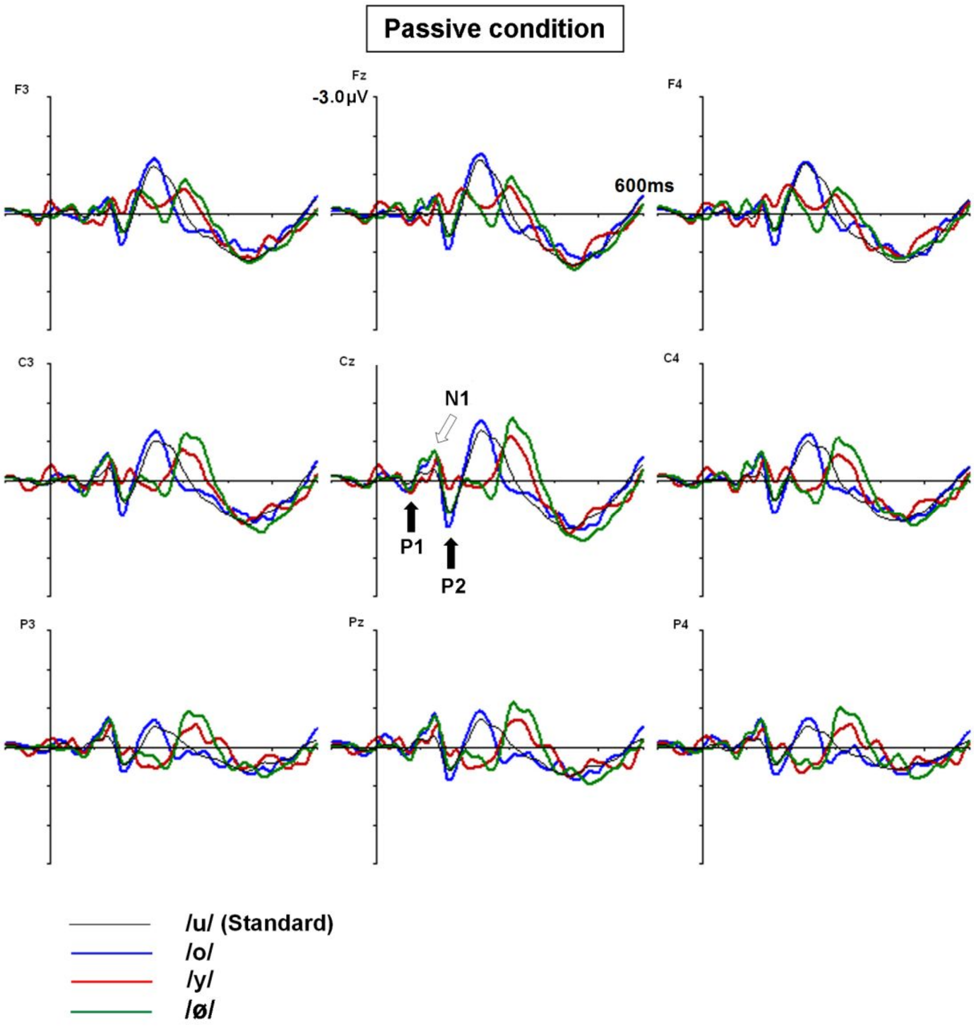


Figure 6.4. Grand-averaged ERP waveforms in the Passive condition in Experiment 2 for the standard /u/ (black line) and for the 3 deviant vowels: /o/ (blue line), /y/ (red line) and /ø/ (green line). All speakers are included.

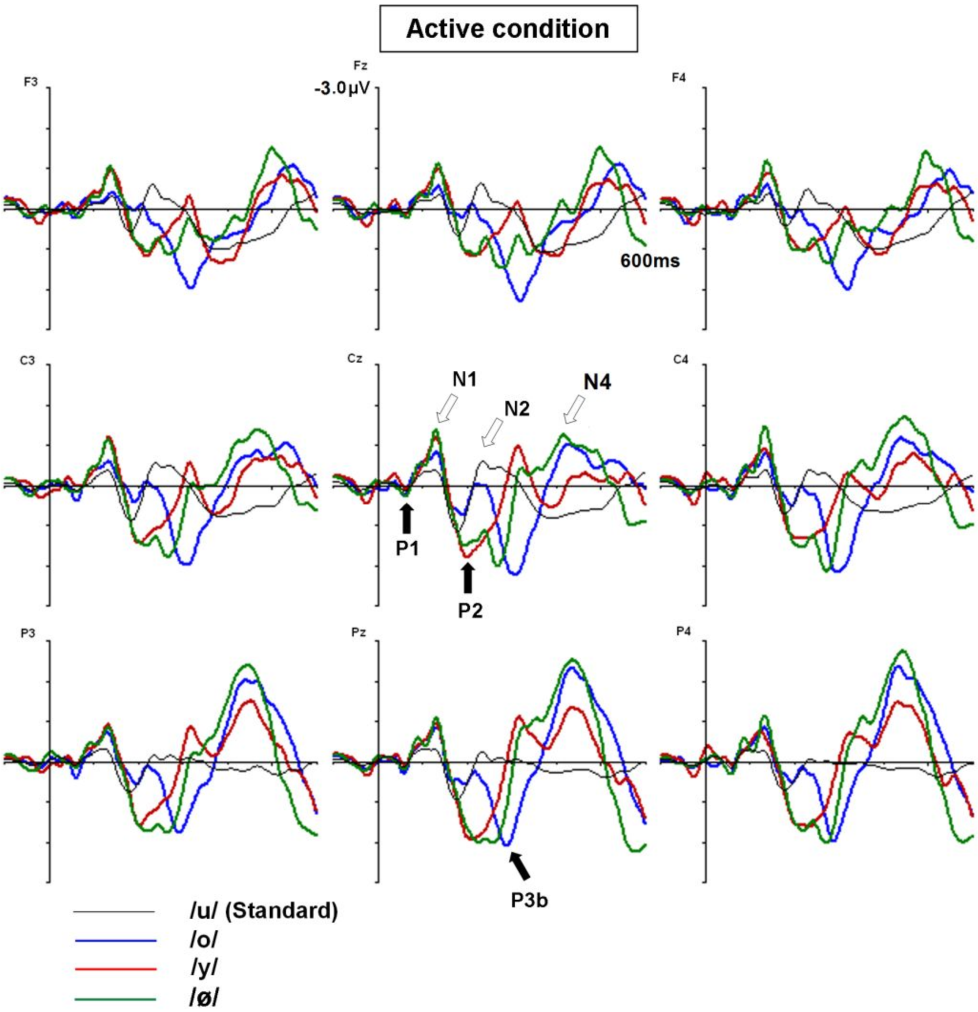
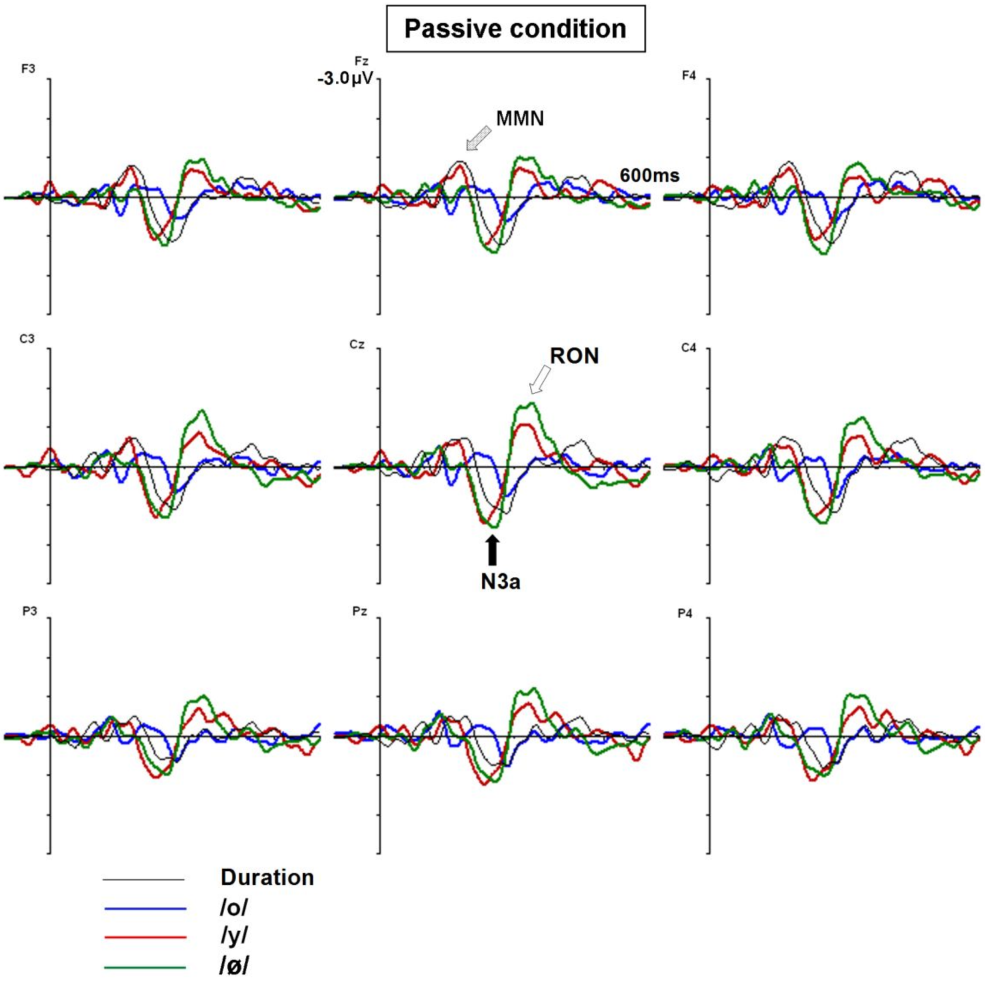
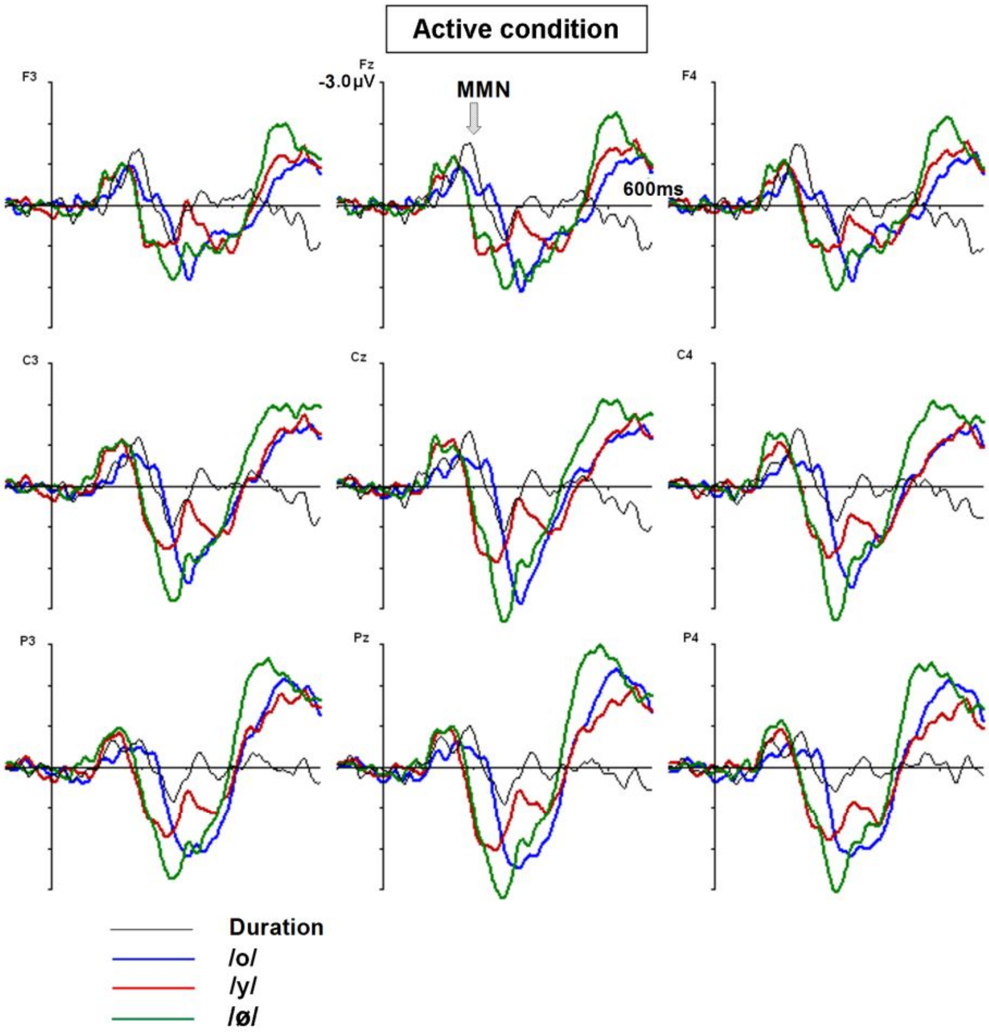


Figure 6.5. Grand-averaged ERP waveforms in the Active condition in Experiment 2 for the standard /u/ (black line) and for the 3 deviant vowels: /o/ (blue line), /y/ (red line) and /ø/ (green line). All speakers are included.



Figures 6.6. Grand-averaged ERP difference waveforms (deviant minus standard) in the Passive condition in Experiment 2 for the Duration deviant (black line) and the deviant vowels /o/ (blue line), /y/ (red line) and /ø/ (green line). All speakers are included.



Figures 6.7. Grand-averaged ERP difference waveforms (deviant minus standard) in the Active condition in Experiment 2 for the Duration deviant (black line) and the deviant vowels /o/ (blue line), /y/ (red line) and /ø/ (green line). All speakers are included.

7. GENERAL DISCUSSION (STUDY I)

In Study I, two experiments were conducted with French and Italian native speakers, respectively, using the same paradigm and stimuli in order to examine how perceptual and cognitive processes differ due to long-term linguistic experience. The results were not directly compared between these experiments because of different experimental equipments employed. However, the results obtained from two native-group language groups showed interesting differences, which provide us with information about cross-linguistic interference in speech sound perception. In this section, different aspects of between-group differences attributable to language-specific adaptation of the auditory system are highlighted, with their implications for future research perspectives.

Pre-attentive and attentive categorization of vowels

The results regarding the MMN were very different between the two experiments. In Experiment 1, MMNs were elicited by all the deviant vowels in the Passive and Active conditions, indexing pre-attentive categorization according to the vowel inventory of the native language of participants. By contrast, in Experiment 2 no clear MMN was elicited by any of the deviant vowels. Such difference between the native-language groups is important for understanding the mechanism of categorical perception. The MMN can be elicited by violation of regularity based on any acoustic dimension, while elicitation of the MMN by deviant vowels required detection of regularity, specifically in formant dimensions. An issue of interest was therefore if the MMN could be elicited according to formant changes even in the absence of long-term phonetic representations corresponding to deviant vowels. Importantly, formant values of naturally spoken vowels used in the present experiments showed a certain variability within each category (see Figure 5.1), and two deviant vowels (/y/ and /ø/) presented intermediate values falling between Italian native categories. For this reason, it may be that formants were perceived as more continuous than categorical parameters.

This tendency to perceive formants as continuous parameters may have been enhanced by the presentation of all the vowels in the same sequence, and in addition by the presence of the vowel (/o/), which is an Italian vowel, but it is also acoustically very close to /u/, and, as suggested by behavioral data, perhaps not prototypically pronounced by the three French speakers. This interpretation is even more likely considering that vowel /o/ did not elicit a clear MMN, although it belongs to the Italian inventory.

Moreover, speaker-specific acoustic properties that changed in a random order could have interfered strongly with formant perception. The obtained results suggest that when the sequence presents variations in formants covering frequency ranges not corresponding to long-term memory traces possessed by listeners, speaker-specific acoustic information may override formant information at the pre-attentive level.

Another striking finding was that the deviant vowels /y/ and /ø/ elicited the P3a and RON in the Passive condition in Italian native speakers, but not in French native speakers. This means that the non-native vowels captured involuntary attention, even though they were not consciously identified. The observed dissociation between the N1, MMN and P3/RON suggests that the involuntary attention switch did not depend on pre-attentive auditory change detection, as reflected by the N1 and MMN, but rather by a later evaluative process on the nature of the deviant sound. The dissociation between behavioral data and the P3a/RON elicitation (that is, although /y/ was as difficult to discriminate as /o/ and much more difficult than /ø/, /y/ and /ø/ both elicited the P3a/RON, but /o/ did not) also supports that attention switch was related with the classification of vowels rather than the perceptual salience of the stimulus. This interpretation can explain why these vowels triggered an attention switch in Italian participants but not in French participants.

The involuntary attention switch induced by non-native vowels in Italians has an important implication in learning foreign languages, in that it indicates that novel vowel categories that should be learned can be perceptually differentiated from the native vowels even without consciously focused attention. Moreover, probably not all the non-native speech sounds can generate the P3a/RON effect, but attention capture should depend on specific acoustic properties of the sound and familiarity of the non-native language considered. As far as the familiarity is concerned, French speech sounds are considered to be, to some extent, familiar to many Italian native speakers, because they are exposed to different European languages by the media, by listening to foreign tourists or by traveling abroad. The two non-native vowels used in the present study (i.e., /y/ and /ø/) are also similar to German /ü/ and /ö/, respectively, and can be recognized as phonetically familiar sounds by Italians without specific knowledge of French or German language, even though they cannot classify them into exact phonemic categories. In this sense, it might be the ‘familiarity’ rather than the ‘novelty’ of these vowels that induced attention switch, requiring further analysis to identify them. Therefore, the attention capture effect of non-native vowels should be further examined by including more categories that are acoustically classified as belonging or not to the native phonemic inventory and by comparing familiar foreign languages with unfamiliar ones.

In contrast with MMN results, behavioral data showed that participants in both groups could discriminate deviant vowels (although not always easily, as suggested by the 20–30% error rates for certain vowels and speakers) by focusing attention on the relevant formant dimensions and filtering speaker-specific information. ERP waveforms recorded in the Active condition were quite different between the two experiments, especially for /y/ and /ø/. Whereas ERPs for /o/ showed similar patterns in the two experiments with recognizable N2/P3 complex following the N1 and P2 components, the ERPs for /y/ and /ø/ presented more complex patterns. The larger N2 amplitude for /o/ than for /y/ and /ø/ was also observed in Experiment 1, and it may reflect the difficulty of discrimination due to acoustic similarity of /o/ to the standard.

The parietal N4 component was elicited by deviant vowels with larger amplitudes for /ø/ than for /y/ (and of intermediate amplitude for /o/) in the Active condition only in Experiment 2. In particular, only for the non-native vowels (/y/ and /ø/) the N4 was preceded by a PMN-like negativity. The functional interpretation of each component is only speculative, but they presumably reflect cognitive strategies employed to evaluate the mismatch between the incoming deviant and the representation of the standard consciously maintained in the memory buffer. . Such mismatch, as noted, reflects analysis of acoustic characteristics in interaction with phonetic representations of vowel categories. It should be of great interest to identify such strategies to develop efficient phonological training methods for learners of foreign languages. For example, it should be noted that the articulatory gestures employed to pronounce the novel vowels (/y/ and /ø/) were explained to participants prior to the discrimination task. It is possible that such instruction might have induced processing of articulatory movements associated with covert production.

Acoustic distance effect

Acoustic distance defined by F1 and F2 was consistently reflected by behavioral data and by the N1 and MMN amplitudes in the Passive and Active conditions in Experiment 1, but not in Experiment 2. This may have been enhanced by the above noted characteristic of the vowels to lie on a continuum, and by the speaker variability, making their discrimination less easy. In addition, speaker variability showed stronger effects on the N1 amplitude in the Passive condition in Experiment 2, but not in Experiment 1. By contrast, the P2 component, which has also been referred to as the N1/P2 complex that shows sensitivity to spectral and temporal cues (Martin et al., 1997; Whiting et al., 1998; Tremblay, 1998), was particularly prominent in Experiment 2, mainly for /ø/ and /y/.

These results suggest that, in the Passive condition, sensitivity of the N1 generator neurons was spontaneously tuned to receive formant information when the input vowels were all typical exemplars of native categories (Experiment 1), whereas it was better tuned to speaker-specific acoustic information than to formants when some vowels contained formant values in less easily categorizable ranges (Experiment 2). Such difference between the two experiments shows the persistent impact of long-term experience with the native language on speech sound perception due to the brain plasticity modulating neural sensitivity underlying early stages of auditory processing, which are reflected by the N1 and MMN. This is true even despite the lack of a clear effect for the native vowel /o/, considering that it may not have been perceived as a typical Italian /o/ because it was pronounced by French speakers (Strange et al., 2007). The observed error rate for /o/ as high as around 20% may reflect perceptual difficulties for vowel instances belonging to a native category but are phonetically realized differently in the context of another language (Gottfried, 1984).

The language-specific effect on pre-attentive auditory processing has previously been reported using the MMN (Näätänen et al., 1997; Winkler et al., 1999a; Cheour et al., 1998). The present results further demonstrated the selective attunement of sensitivity at the level of transient detectors, involving afferent neural populations responding to typical formant configurations of the native phonemes.

Speaker variability

Behavioral data obtained from Italian participants showed significantly better discrimination for deviant vowels pronounced by Speaker 3 than for vowels produced by the other speakers (significant main effect of Speaker), while French participants did not show differential sensitivity to this specific speaker. This suggests that acoustic cues specifying speaker identity had a stronger effect on Italian participants than on French participants. As additional information, Speakers 1 and 2 were from the north region of France (around Paris), while Speaker 3 was from the south region (Provence). Therefore, the ease of discrimination for Speaker 3 observed in Experiment 2 may be associated with the regional accent. However, this possibility should be verified by more detailed analysis of phonetic parameters collecting wider samples of natural spoken vowels, and such analysis may be useful to know what acoustic parameters are particularly relevant to improve the discriminability of vowels.

Moreover, discrimination performance differed depending on the speakers especially for the vowels difficult to discriminate (/o/ and /y/) in French participants, and for the non-native vowels (/y/ and /ø/) in Italian participants. The former case can be accounted for by uncertainty in response decision for most difficult tokens, while the latter case by the lack of prototypical representations for the non-native vowels. In other terms, these results may reflect difficulty of generalization across speakers for acoustically ambiguous or less familiar phonemes (Eisner and

McQueen, 2005).

Regarding the relationship between phonetic information and speaker specificity, Knösche et al. (2002) suggested that these dimensions interact at pre-attentive stages within the auditory cortex, but they are treated by parallel processes at a later stage of signal-evaluation. However, Kaganovich et al. (2006) also found earlier processing for speaker identity, as reflected by shorter P1 latency, than for vowel category. Our results demonstrated selective sensitivity to speakers or vowel categories at pre-attentive level modulated by listeners' native language, suggesting that processes for these dimensions are at least partially separate even in early auditory processing.

Attention effect

In both experiments, the N1 amplitude was enhanced in the Active condition compared with the Passive condition. Such attention effect appears to be independent of the language-specific factors, and attributable to a higher level of arousal in the active listening condition or to an overlapping component (i.e., the processing negativity) reflecting effects of selective attention to the task-relevant stimuli. More importantly, the N1 amplitude showed speaker variability effect only in the Active condition in French participants. By contrast, in Italian participants the N1 showed sensitivity to speaker variability only in the Passive condition, but not in the Active condition. Such difference between groups at the early stage of processing suggests that sensitivity of afferent neurons responding to specific dimensions can be tuned by selective attention according to the task-relevance.

Selective attention should have a crucial role in establishment of novel phonetic categories by phonological training in adults learning a foreign language. Dobel et al. (2009) reported non-efficacy of passive training by statistical learning for acquisition of a novel phonetic contrast. By contrast, efficacy of short-term training has often been reported using an active identification or discrimination task with stimulus presentation at short interstimulus interval and visual feedback (Kraus et al., 1995; Tremblay et al., 1998, 2001; Menning et al., 2002). Such a method is considered to induce listeners to focus their attention on the target phonetic contrasts and to evaluate the input sounds based on the echoic memory and acoustic analysis rather than on the phonetic representations in long-term memory (Kraus et al., 1995).

STUDY II: Perception of phrasal pitch contour in the native and unfamiliar languages

(Experiments 3)

8. EXPERIMENT 3

8.1. INTRODUCTION

Study II focused on the perception of pitch contour of spoken sentences to clarify the effect of long-term experience with listeners' native language on different linguistic components that interact with pitch processing.

Detection of incongruities in pitch contour (i.e., strong and weak pitch changes) was investigated in previous ERP studies in order to verify the hypothesis that musical training would result in higher sensitivity to pitch variations in spoken sentences by comparing between musicians and non-musicians. The task was to detect the sentence-final words carrying a weak or strong pitch incongruity, defined on the basis of the percentage increase in the fundamental frequency (F0) compared to the control condition (i.e., 200% or 135%, respectively). In two experiments, French adult native speakers were tested using sentences spoken in their native language (Schön et al., 2004) and sentences in a foreign language, Portuguese (Marques et al., 2007). In both experiments, faster reaction times (RTs) and lower error rates in a pitch detection task were observed for strong compared to weak pitch changes independently of the musical expertise. Moreover, electrophysiological data showed that the amplitude of an early negativity (peaking around 150 ms) and of a later positive component were larger to strong than to weak incongruities (Schön et al., 2004). These two components are considered to reflect, respectively, automatic perceptual processing of pitch change and following cognitive processes involved in analysis and evaluation of pitch congruity of the final word within the prosodic contour of the current sentence in order to provide a response. Importantly, significant differences between subject groups were observed only in the weak incongruity condition: lower error rates and larger positive components in musician than non-musician were commonly observed in these studies.

Overall, the pattern of results demonstrated a reliable effect of musical expertise on F0 change perception in French native speakers listening to French and Portuguese sentences. However, the results obtained from these experiments also showed some interesting differences between languages. First, RTs to strong pitch changes were slower, and error rates were higher in the foreign language compared to the native language. In addition, the onset of the positive component to the strong incongruity was delayed (by about 300 ms) in Portuguese compared to French. Finally, the negative component elicited by Portuguese sentences showed different morphology and time course compared that observed in Schön et al.'s study (2004). These differences suggest that the prosodic contour can be processed more easily when the sentence is spoken in the native language than when it is in an unfamiliar foreign language.

Marques et al. (2007) suggested that RTs and latencies of ERP components were longer in Portuguese than in French, because participants were not able to develop expectancies regarding when the critical sentence final word would occur based on the semantic context of the sentence in Portuguese. This interpretation implies that semantic and prosodic processes might interact between them in detection of prosodic incongruities in a sentence context.

However, an alternative interpretation is that detection of pitch changes is largely based on the prosodic contour of the sentence. Presumably, speakers implicitly acquire typical prosodic contours of sentences in their native language, and this implicit knowledge, together with syntactic and semantic information, allows a listener to predict the prosodic patterns of the forthcoming segment. In turn, prosodic cues can help segmenting the speech input into words and phrases, thanks to distributional regularities of word-level prosody (Cutler & Carter, 1987; Colombo, 1992) and close relationship between prosodic and syntactic boundaries (Jusczyk, 1999).

There is evidence that listeners exploit distal prosodic cues in spoken language processing (i.e., preceding cues that are far from the segmentation point in the sentence). For example, relative size of prosodic boundaries affects the way an ambiguous syntactic structure is parsed (Schafer et al., 2000). Dilley & McAuley (2008) investigated how, based on auditory perceptual organization principles, listeners exploited F0, duration or both as distal cues to define lexical boundaries and to segment speech. Pannekamp et al. (2005) showed that the CPS component was elicited by prosodic boundaries not only with normal sentences, but also with jabberwocky sentences (i.e., pseudowords replacing content words) or hummed sentences, in which only prosodic information was present. Thus, distal cues based on the prosodic structure of the native language are likely to be used to organize syntactic structures and lexical boundaries. Previous studies on language acquisition provide evidence that infants' speech perception is tuned to the linguistic environment with which they interact (Kuhl., 2004 for a review), and that adults tend to perceive speech sounds according to the phonological systems of their native language (Best & Strange, 1992; Flege, 1995).

In the present study, the familiarity of distal prosodic information was manipulated in order to determine whether discrimination of pitch changes depends on the possibility to carry out a syntactic and semantic interpretation of the sentence context or may solely rely on the familiarity and predictability of the intonational contour of the context. We used an experimental design similar to Schön et al. (2004) and Marques et al. (2007) by presenting to Italian participants sentences in which F0 of whole final words was slightly or strongly increased. Sentences were spoken in Italian (native language), French (unfamiliar foreign language) or in jabberwocky. The Italian-French contrast extends the French-Portuguese comparison investigated in separate studies by Schön et al. (2004), and by Marques et al. (2007) by using a within-subject design. According to the rhythm-based classification, both Italian and French are syllable-timed languages (Cutler and Norris, 1988). However, Italian is characterized by lexical stress, which is located in a variable and unpredictable position of the word, whereas French has a single rhythmic stress that is assigned to the final full syllable (i.e. not containing a schwa) of the last lexical item of a stress group (Di Cristo & Hirst, 1998; Magne et al., 2007). In French, the strong syllable at the end of the rhythmic group (i.e., a group of syntactically-related words, like prepositional or verbal phrases) is marked by pitch increase. By contrast, stress In Italian is marked by increased length of the tonic syllable, and not necessarily by pitch variation. Consequently, variations in pitch could only be interpreted to signal intonational changes independently of the stress pattern (Rossi, 1998).

While Italian sentences included all different types of information (prosodic, semantic and syntactic), jabberwocky sentences preserved the familiar prosodic pattern of Italian but had only limited syntactic information provided by function words and morphological affixes, and no lexical/semantic content. Familiar prosodic cues of their own language may help participants to detect unnatural pitch changes, and also to predict when the sentence final word occurs. Thus, if pitch incongruity discrimination is based on the intonational contour of the sentence, the level of performance should not differ between Italian and jabberwocky sentences (both contain familiar prosodic information), and should be higher than with French sentences. Moreover, the ERP pattern should be more similar between Italian and jabberwocky than between Italian and French. Specifically, based on previous results for the native (Schön et al, 2004) and a foreign language (Marques et al, 2007), we expected that both the early negativity to the strong incongruity and the late positivity (P3-like component) should be larger for Italian and jabberwocky sentences than for French sentences. By contrast, if semantic interpretation of the sentence is crucial in order to identify the word on which pitch change occurs, as proposed by Marques et al. (2007), and if participants are not able to exploit familiar prosodic information without semantic content, the level of detection performance for Italian sentences should be higher than for both jabberwocky and French sentences, that should not differ between them. Regarding ERPs, the amplitude of the P3-like component should be similar for both French and jabberwocky sentences and smaller than for Italian sentences. The contrast between Italian, jabberwocky and French should consequently allow us to verify if pitch incongruity detection is based on prosody, on semantics or on both.

8.2. METHODS

8.2.1. Participants

Thirty-nine Italian native speakers participated in the experiment. Because of the difficulty to discriminate the weak incongruity condition in French, the level of accuracy for 9 participants was too low, and their data were removed from further analyses. Moreover, one participant had too many artifacts in the ERP recordings and only behavioral data were included. Mean age of the 30 participants was 24.77 years (S.D. = 4.17). All participants were right-handed and had normal hearing according to self-report. None of them had learned or was able to understand French. They were paid for participation in the experiment that lasted 2 hours and a half.

8.2.2. Materials

Ninety-nine sentences were created for each language condition (Italian, jabberwocky and French). French and Italian sentences were declarative sentences with one or two main clauses, and/or a subordinate or a prepositional phrase. French sentences were partially taken from the materials used in Schön et al. (2004), but were mostly new sentences, with the constraint that they ended with a disyllabic word starting with a stop consonant, as Italian and jabberwocky stimuli did. Last words of Italian sentences were stressed on the antepenultimate syllable (as the majority of Italian words), while those of French sentences were stressed on the last syllable (as all French words are). Jabberwocky sentences were created by substituting the content words of the Italian sentences with pseudowords with the same number of syllables. All sentence types were matched for number of syllables, and as much as possible for syntactic structure. Examples of the sentences are the following:

Italian: *Marco non deve andare in bicicletta sulla corsia riservata ai taxi.*

Jabberwocky: *Forvo non dusna fadare in chirinetta sulla gornea tirubata ai tessi.*

French: *Le père de Sophie est accueilli à son arrivée par un délicieux gratin.*

The 297 sentences were spoken by a fluent Italian-French bilingual female speaker, recorded and digitized using the computer software “Audacity”. Italian and French sentences were read with a natural intonation appropriate to each language. Each jabberwocky sentence was read immediately after the corresponding Italian sentence, so that the speaker maintained a similar prosodic contour in both conditions. Mean sentence duration was 3909 ms (SD = 509) for Italian, 4393 ms (SD = 612) for jabberwocky and 4104 ms (SD = 537) for French. Mean duration of the final word was 507 ms (SD = 60) for Italian, 535 ms (SD = 61) for jabberwocky, and 511 ms (SD =

127) for French.

Three lists were built so that each sentence was presented in each of the three experimental conditions across participants. The final words of the sentences were prosodically congruous, weakly incongruous, or strongly incongruous with the preceding context. To create the incongruous conditions, Italian and French sentences were modified by using the software “Praat” (Boersma & Weenink, 1996). F0 of the whole last word was increased to 115% for the weak incongruity, and to 200% for the strong incongruity conditions, while maintaining the same pitch contour (see Figure 8.1). The increase in F0 for the weak incongruity was based on the results of pretests, so that accuracy levels in Italian would be comparable to that obtained by Schön et al. (2004) in French.

Thus, a total of 891 sentences were used in the experiment (99 sentences x 3 languages x 3 congruity conditions), with 33 sentences in each experimental condition. Each participant received one list of 297 sentences (33 trials x 3 languages x 3 congruity conditions). Each sentence was only presented in one condition.

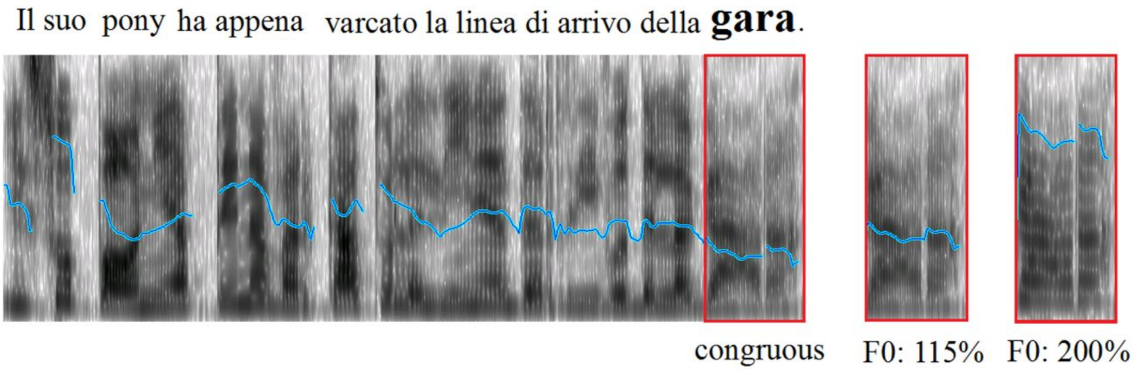


Figure 8.1. Spectrogram of an Italian sentence used in the experiment. The pitch contour is shown by the blue line tracing the F0. Pitch contours of the sentence-final word in the weak incongruity (F0: 115%) and in the strong incongruity (F0: 200%) are also shown.

8.2.3. Procedures

The experiment took place in an electrically shielded quiet room, and was programmed and presented using the software “e-prime”. Participants were seated in front of a PC monitor and were required to fixate the central asterisk while attentively listening to the sentences that were auditorily presented through a pair of speakers symmetrically positioned in front of them. They were instructed to decide as quickly and accurately as possible, whether the intonation of the final word of each sentence was congruous or incongruous with the former sentence context, pressing one of two buttons of a response box (one for “congruous” and the other for either weakly or strongly incongruous). Reaction times were measured from the onset of the final word (corresponding to the onset of the F0 modification in the incongruous conditions) until a response button was pressed.

For each participant, the three language conditions were presented separately in three experimental blocks. Each experimental block began with a practice block of 9 trials, during which feedback on the correctness of the response was given, to familiarize participants with the task and to train them to blink during the inter-trial interval. Feedback was only given on practice trials. In each block, sentences were presented in a random order. The hand of response and the order of presentation of the three language conditions were counterbalanced across participants.

8.2.4. EEG recordings and data analysis

Electroencephalogram (EEG) was recorded from 31 scalp electrodes, mounted on an elastic cap, and located at the following sites according to International 10/20 system: Fpz, Fz, FCz, Cz, CPz, Pz, Oz, Fp1, Fp2, F3, F4, FC3, FC4, C3, C4, CP3, CP4, P3, P4, O1, O2, F7, F8, FT7, FT8, T3, T4, TP7, TP8, T5, T6. These recording sites plus an electrode placed on the right mastoid were referenced online to the left mastoid electrode and digitally re-referenced off-line to the algebraic average of the left and right mastoids. Electrode impedances never exceeded 9 kΩ. For the purpose of artifact scoring, vertical and horizontal electro-oculograms (EOGs) were recorded. Electrode pairs (bipolar) were placed at the supra- and sub-orbit of the right eye and at the external canthi of the eyes. The EEG and EOG signals were amplified by a Neuroscan Synamp amplifier (El Paso, TX, USA) with a bandpass filter of 0.05–100 Hz and were digitized at 500 Hz (16 bit AD converter, accuracy 0.084 µV/LSB) and stored on a Pentium II computer. The notch filter was set at 50 Hz.

Continuous EEG data were corrected for eye blinks using a regression-based correction algorithm (Scan 4.1 software). EEG was then segmented off-line into single epochs from 150 ms before to 1600 ms after the onset of the sentence-final word. EEG epochs were baseline-corrected against the mean voltage during the 150-ms pre-stimulus period. All EEG epochs were visually scored for eye movement and other artifacts, and each portion of data containing artifacts greater than ±70 µV in any channel was rejected for all the recorded channels prior to further analysis. Artifact-free trials with correct behavioral responses were separately averaged for each subject in each experimental condition.

ERP data were analyzed by computing mean amplitudes, relative to a 150-ms baseline. Latency windows were selected based upon visual inspection of the traces and on preliminary 100 ms latency band analysis between 0 ms and 1200 ms after the onset of the sentence-final word.

When the effects were similar in successive 100-ms windows, latency bands were pooled together. For each language, ANOVAs were computed for midline and lateral electrodes separately and included Language (Italian, jabberwocky and French), Congruity (congruous, weakly incongruous, and strongly incongruous) and Electrodes (Fz, Cz and Pz) as within-subjects factors for midline analysis; and Language (3), Congruity (3), Hemisphere (2), ROI (3 regions of interest: fronto-central, temporal, and parietal) and Electrode (3) as within-subjects factors for lateral analysis. In each ROI three electrodes were grouped together (F3, F7, FC3 and F4, F8, FC4; C3, T3, TP7 and C4, T4, TP8; CP3, P3, T5 and CP4, P4, T6). All p values were adjusted with the Greenhouse-Geisser epsilon correction for non-sphericity. Tukey tests were used for all post-hoc comparisons.

8.3. RESULTS

[Behavioral data]

RTs data were analyzed for correct responses only. ANOVAs were carried out on both accuracy and RTs, with congruency (congruous, weak incongruity, strong incongruity) and language (Italian, jabberwocky and French) as within-subject factors.

Accuracy

Mean RTs and accuracy rates for the two groups of participants are displayed in Figure 1. As accuracy was at ceiling in the strong incongruity condition, statistical analyses were only carried out on the weakly incongruous and congruous conditions. Results showed a main effect of Language [F(2,58) = 26.36, MSE = .01, $p < .001$]. Post-hoc tests showed that the level of performance was similar for Italian (.77) and jabberwocky (.75) and lower for French sentences (.65). The main effect of Congruity was also significant, F(1,29) = 125.24, MSE = .06, $p < .001$] with higher accuracy in the congruous condition (.93) than in the weakly incongruous condition (.52). The Language by Congruity interaction was significant [F(2,58) = 7.12, MSE = .01, $p < .01$]. In the congruous condition, accuracy was higher for Italian than for French sentences (.95 vs .90; t(29) = 2.90, SE = .02, $p < .01$). Performance on jabberwocky (.93) did not differ from Italian, and only marginally from French [t(29) = 1.81, SE = .02, p = .08]. In the weak incongruity condition the difference between Italian (.60) and French sentences (.41) was also significant [t(29) = 5.15, SE = .04, $p < .001$] but the difference between jabberwocky (.57) and French sentences was significant [t(29) = 4.42, SE = .001]; again the difference between Italian and jabberwocky sentences was not significant .

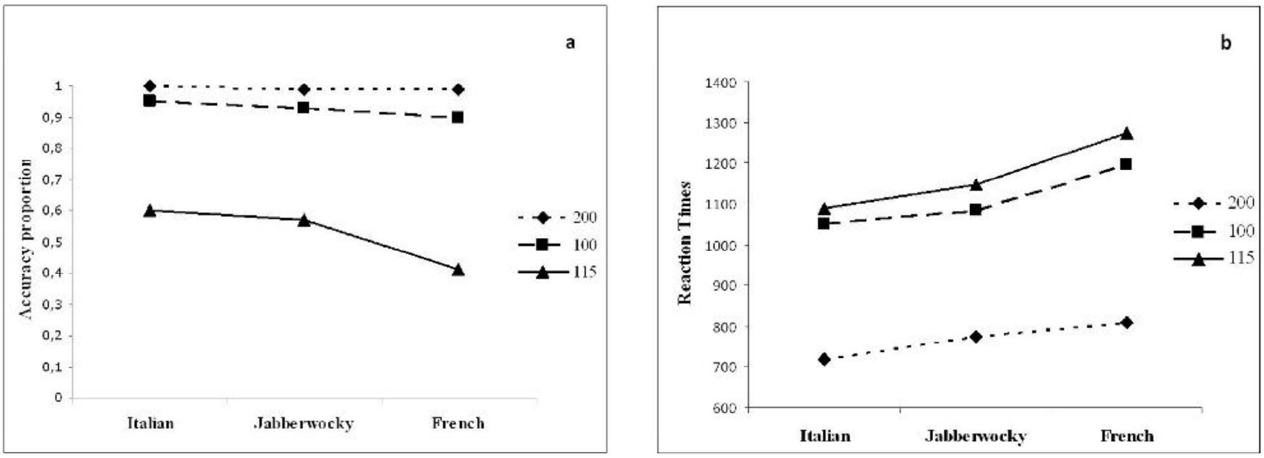


Figure 8.2. Mean accuracy proportion (left) and reaction times (right) in the weak incongruity, strong incongruity and congruous conditions for Italian, jabberwocky and French.

Reaction times

Results showed main effects of Language [F(2,58) = 22.31, MSE = 18505.29, p < .001] and Congruity [F(2,58) = 175.13, MSE = 22098.44, p < .001] and a significant interaction [F(4,116) = 3.48, MSE = 7054.91, p =.01]. Inspection of Figure 1b and Table 1 shows that the Italian-jabberwocky difference was not significant in the congruous and weak incongruous conditions, but it was reliable in the strong incongruous condition. In contrast, French sentences always significantly differed from both Italian and jabberwocky sentences except in the strong incongruous condition: French was not slower than jabberwocky.

Table 8.1. Results of t-tests (t) on RTs data and standard errors (SE) comparing Italian and jabberwocky sentences (It-Jab), Italian and French sentences (It-Fr) and French and jabberwocky sentences (Fr-Jab) in each congruity condition.

			Pitch variations			
	Congruous		*Weak*		*Strong*	
Contrasts	*t*	*SE*	*t*	*SE*	*T*	*SE*
It-Jab	< 1	22.5	< 1	29.4	2.3*	18.2
It-Fr	5.2**	26.8	5.1**	33.9	4.5**	18.5
Fr-Jab	5.2**	24.8	2.9*	38.8	< 1	22.2

*p < .05; ** p < .001

[Electrophysiological data]

Results of the general ANOVAs are reported in Table 8.2. In the 0–200 ms latency range and at lateral electrodes mean amplitudes were more negative for Italian (-1.10 µV, p < .001) and jabberwocky (-0.89 µV, p < .01) than for French (-0.44 µV; see Table 2). Moreover, in the 100–200 ms range, an early negative component was larger for strong incongruity (midline = -1.98 µV and lateral = -2.08µV) than for weak incongruity (midline = -0.86 µV and lateral = -0.88 µV, p < .05 and p < .001, respectively) and congruous endings (midline = -0.61 µV and lateral = -0.75 µV, p < .01 and p < .001, respectively). The Language by Congruity interaction was also significant at midline and lateral electrodes in both the 200–400 ms and the 500–700 ms latency bands.

Table 8.2. Results of overall ANOVAs on mean amplitudes of ERPs at midline and lateral electrodes.

		Latency Band	Factors	*df*	F	p
Overall data	Midline	100–200	C	2, 56	6.41	*
		200–400	L x C	4, 112	8.55	**
		500–700	L x C	4, 112	6.01	**
	Lateral	0–100	L	2, 56	11.43	**
		100–200	C	2, 56	12.98	**
		200–400	L x C	4, 112	9.21	**
		500–700	L x C	4, 112	7.35	**

L: Language (Italian vs. jabberwocky vs. French); C: Congruity (Congruous vs. Weak incongruity vs. Strong incongruity)

df: uncorrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

**: p < .001; *: p < .006

To further analyze these interactions, separate ANOVAs were conducted for each congruity condition, including as factors the three sentence conditions (Italian, jabberwocky and French) as well as Electrodes (Fz, Cz and Pz) for midline analysis and Hemispheres, ROIs and Electrodes (as defined above) for lateral analysis. Results are presented in Table 8.3. For strong incongruity and at both midline and lateral electrodes, mean amplitudes in the 0–100 ms range (N1 component) were smaller for French (midline = - 0.09 µV and lateral = - 0.02 µV; see Figure 2) than for Italian (midline = -0.61 µV and lateral = - 0.53 µV, p < .05 in both conditions) and jabberwocky (midline = - 0.72 μV and lateral = - 0.58 μV, p < .05 and p < .01, respectively). Moreover, in the 200–500 ms range and at parietal sites the positivity (P3b component was larger for Italian (midline = 8.34 μV and lateral = 5.90 μV), than for jabberwocky (midline = 6.61 and lateral = 4.43 μV, p < .001 in both conditions) and for jabberwocky compared to French (midline = 4.76 μV and lateral = 3.07 μV, p < .001 in both conditions). No differences were found between 500 and 700 ms. Finally, the Language by Electrode and Language by ROI interactions were also significant in the 700–900 ms range. A late positivity developed at parietal sites, that was larger for French (midline = 7.13 μV and lateral = 5.42 μV) than for jabberwocky (midline = 5.74 μV and lateral = 4.43 μV, p = .001 and p < .001, respectively) and intermediate for Italian (midline = 6.33 μV and lateral = 5.20 μV), which differed only from jabberwocky (p < .05). By contrast, at frontal sites, it was larger for Italian (midline = 2.37 μV and lateral = 1.03 μV) than for both jabberwocky (midline = 0.73 μV and lateral = -0.08 μV, p = .001 and p < .001, respectively) and French (midline = 0.71 μV, p < .001, and lateral = -0.26 μV, p < .001).

For the weak Incongruity and at lateral electrodes, mean amplitudes in the 200–300 ms range were more negative (N2/N3 component) for Italian (-2.10 μV, p < .001) and jabberwocky (-1.68 μV, p < .01) than for French (-0.45 μV; see Figure 3). Moreover, the Language by Electrode and Language by ROI interactions were also significant in the 300–700 ms range. The amplitude of the positivity (P3b component) at parietal sites was larger to Italian (midline = 4.74 μV and lateral = 3.07 μV), than to jabberwocky (midline = 2.52 μV, p < .001, and lateral = 1.17 μV, p < .001) and to jabberwocky than to French (midline = 1.27 μV, p < .05 and lateral = 0.30 μV, p < .05). Finally, the Language by Hemisphere interaction was significant in the 300–400 ms latency band: the amplitude was significantly more negative at right (-2.56 μV) than at left (-1.34 μV, p < .001) electrodes only for jabberwocky.

For congruous endings and at both midline and lateral electrodes in the 300–700 ms latency band, both the main effect of Language and the Language by Electrodes and Language by ROI interactions were significant. The amplitude of the positivity (P3b component) at parietal sites was larger to Italian (midline = 3.69 μV and lateral = 2.40 μV) than to jabberwocky (midline = 2.16 μV, p < .001, and lateral = 0.99 μV, p < .001) and to jabberwocky than to French (midline = 1.32 μV and lateral = 0.27 μV, p = .08 and p < .001, respectively; see Figure 4).

Table 8.3. Results of separate ANOVAs for each congruity condition on mean amplitudes of the ERPs at midline and at lateral electrodes.

		Latency (ms)	Band	Factors	*df*	F	p
Strong Incongruity	Midline	0–100		L	2, 56	3.75	*
		200–500		L x E	4, 112	27.18	***
		700–900		L x E	4, 112	5.64	**
	Lateral	0–100		L	2, 56	6.18	**
		200–500		L x R	4, 112	41.73	***
		700–900		L x R	4, 112	6.97	***
Weak Incongruity	Midline	300–700		L x E	4, 112	4.32	*
	Lateral	200–300		L	2, 56	10.52	***
		300–400		L x H	2, 56	4.78	*
		300–700		L x R	4, 112	6.08	**
Congruous Endings	Midline	300–700		L	2, 56	10.21	***
		300–700		L x E	4, 112	3.56	*
	Lateral	300–700		L	2, 56	12.01	***
		300–700		L x R	4, 112	4.99	**

L: Language (Italian vs. jabberwocky vs. French); C: Congruity (Congruous vs. Weak incongruity vs. Strong incongruity); R: ROI (region of interest); H: Hemisphere; E: Electrodes

df: un corrected degrees of freedom; p: probability level after Greenhouse-Geisser correction of F-values

*: p < .05; **: p < .01; ***: p < .001

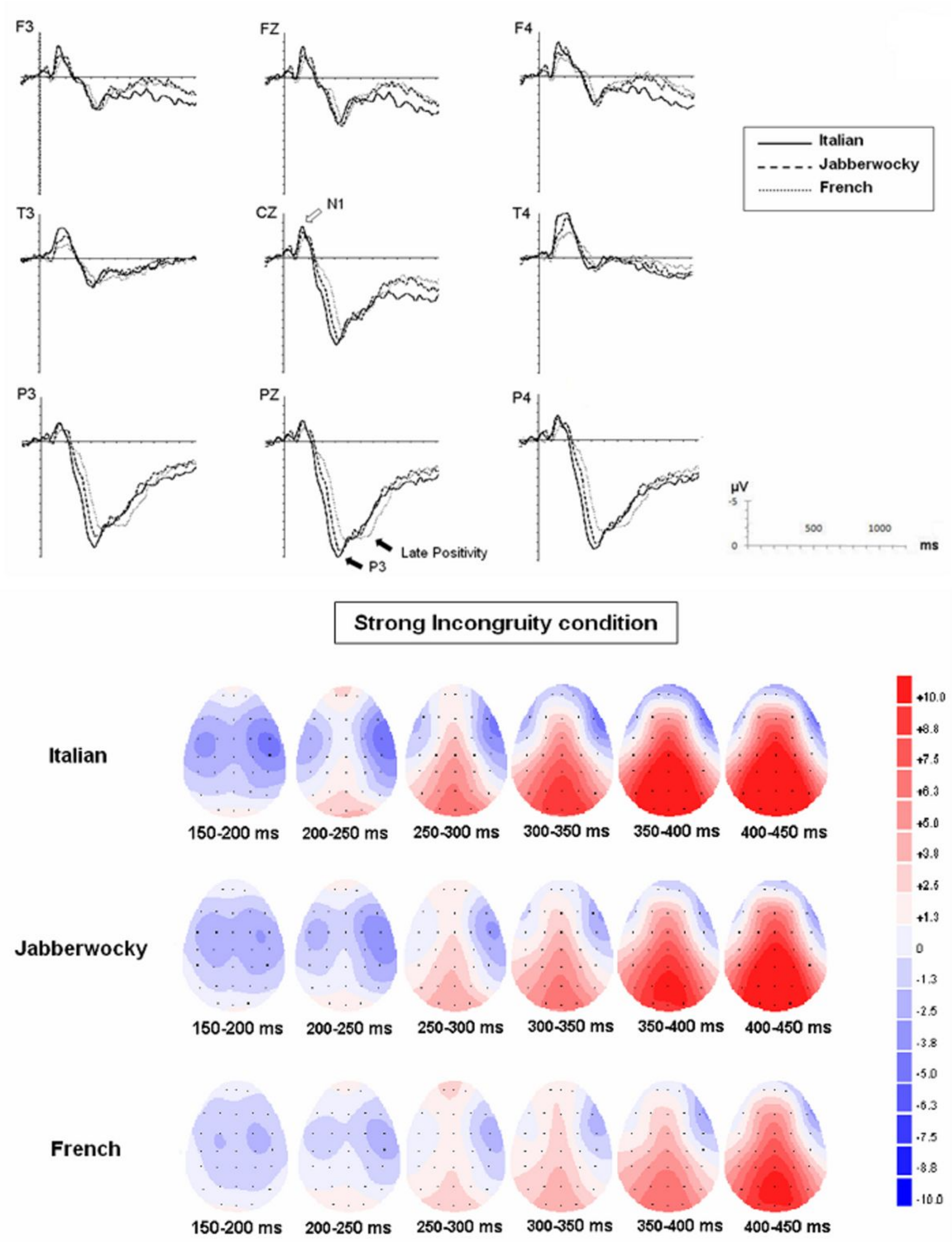


Figure 8.3. Upper panel: Grand-averaged event-related potentials elicited by the strongly incongruous condition in Italian, jabberwocky and French sentences. Selected traces from 9 electrodes are presented. The N1 component is marked by a white arrow. The P3 component and Late Positivity are marked by black arrows. **Lower Panel:** Topographic maps of scalp electrical activity computed every 50 ms from 150 to 450 ms from the onset of the last word in the strongly incongruous condition for Italian (top), jabberwocky (middle) and French (bottom) sentences.

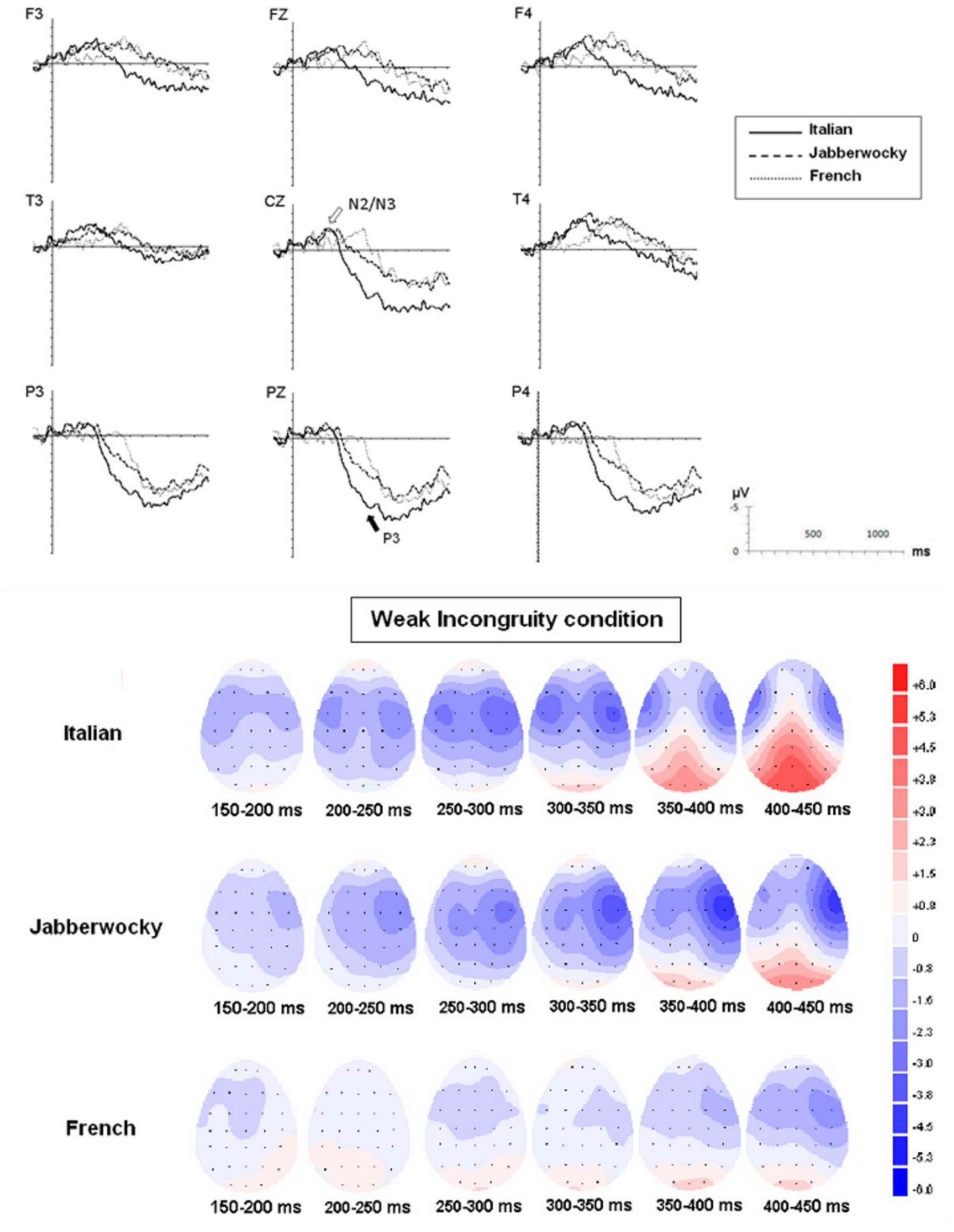


Figure 8.4. Upper panel: Grand-averaged event-related potentials elicited by the weakly incongruous condition in Italian, jabberwocky and French sentences. Selected traces from 9 electrodes are presented. The N2/N3 component is marked by a white arrow. The P3 component is marked by a black arrow. **Lower Panel:** Topographic maps of scalp electrical activity computed every 50 ms from 150 to 450 ms from the onset of the last word in the weakly incongruous condition for Italian (top), jabberwocky (middle) and French (bottom) sentences.

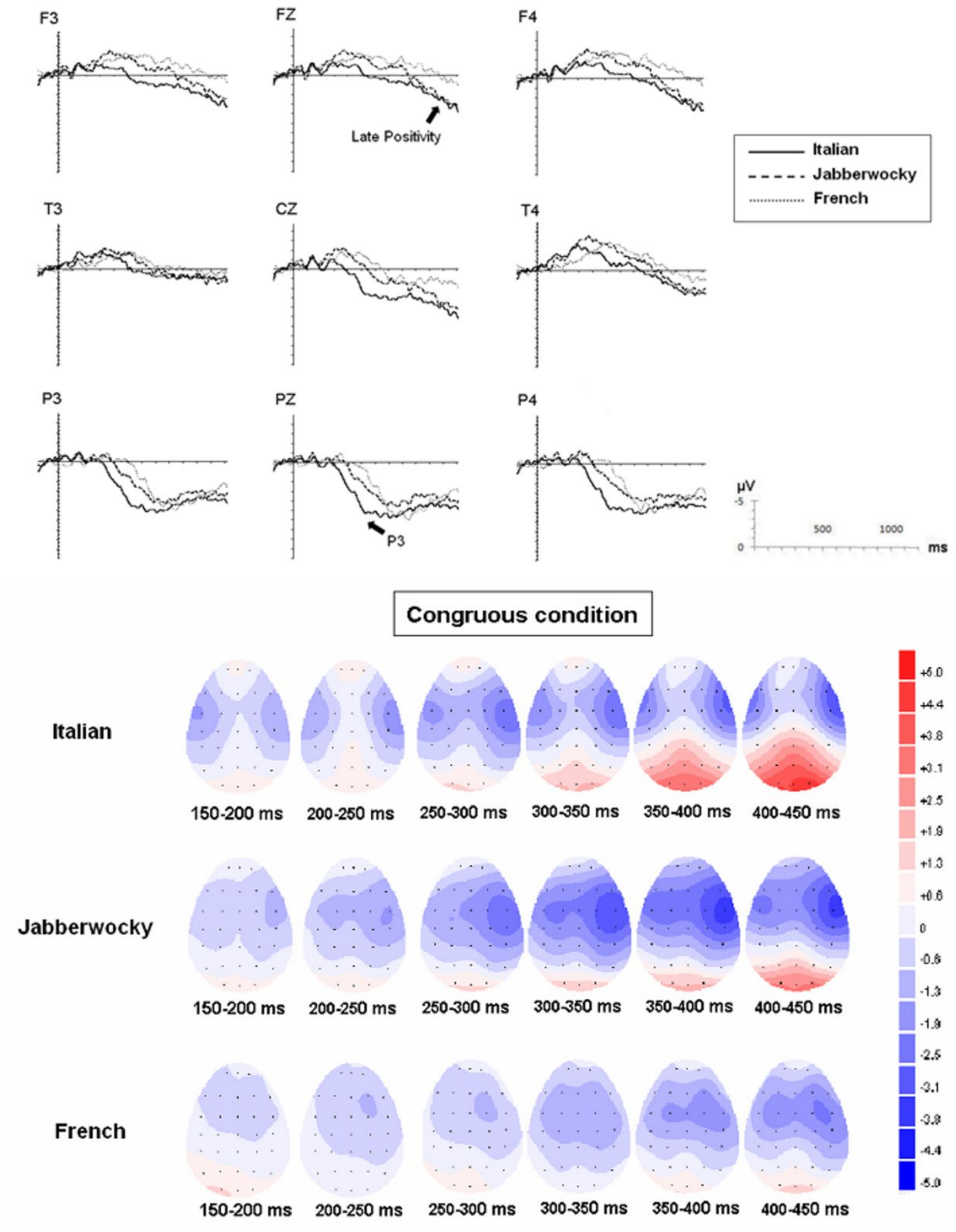


Figure 8.5. Upper panel: Grand-averaged event-related potentials elicited by the congruous condition in Italian, jabberwocky and French sentences. Selected traces from 9 electrodes are presented. The P3 component and Late Positivity are marked by black arrows. **Lower Panel:** Topographic maps of scalp electrical activity computed every 50 ms from 150 to 450 ms from the onset of the last word in the congruous condition for Italian (top), jabberwocky (middle) and French (bottom) sentences.

8.4. DISCUSSION

Behavioral data showed a strong and reliable effect of language familiarity. Moreover, the level of performance was higher for jabberwocky than for French sentences, showing that participants were able to use familiar prosodic contour of sentences to detect pitch changes.

As noted, the weakly incongruities were difficult to discriminate from the congruous endings, as was seen in previous studies (e.g., Schön et al, 2004). Although the congruous condition was almost as accurate as the strongly incongruous condition and both were at ceiling (Figure 8.2a), responses were faster to the strongly incongruous condition than to the congruous and weakly incongruous conditions (Figure 8.2b). These results suggests that the strongly incongruous endings were easy to discriminate from the congruous ones, whereas congruous and weakly incongruous endings were more difficult to discriminate from each other. The latter conditions presumably required a re-analysis of the auditory input in order to decide if a pitch change had occurred, by comparing the pitch contour of the last word with that of the sentence context. Thus, congruous and weakly incongruous conditions were most suitable to see effects of familiarity of the intonational patterns of the native language. In both conditions, accuracy was higher and RTs were faster for Italian and jabberwocky (that did not differ from one another) than in French. This pattern suggests that participants were able to efficiently exploit familiarity with the pitch contour of the sentences.

Moreover, in the strong incongruity condition, participants were almost equally accurate in all three languages, but they responded faster in Italian than in jabberwocky and French. This is in line with the explanation provided by Marques et al. (2007) that a full semantic interpretation allows listeners to anticipate the sentence final word. Thus, pitch detection was helped by complete linguistic and semantic information provided by Italian sentences, even when the pitch variation was acoustically very salient and easy to detect.

Thus, the present results add to the evidence brought forward by Marques et al. through a) a direct comparison of the native and foreign language in a within-subjects design, and b) the inclusion of a pseudo-language (jabberwocky) in which only familiar prosodic information provides important clues to the pitch change. The former confirmed that a full semantic interpretation to optimally perform the task, even when the pitch change to detect is perceptually very strong. The latter showed that the jabberwocky condition allowed listeners to perform the task as accuracy as in the native language based purely on familiarity of intonation.

The pattern of ERP data was in line with behavioral data. Overall, results showed that the amplitude of the P3 component, which was elicited by strongly incongruous endings over parietal sites peaking at 200/300 ms after final word onset, was largest for Italian sentences, intermediate for jabberwocky sentences and smallest for French sentences. This ERP component nicely complements behavioral data in the strong incongruity condition, in which accuracy was at ceiling. As noted above, the P3b component is taken to reflect post-perceptual processes involved in categorization and decision, and P3b amplitude positively reflects discriminability of the stimulus (e.g., Hillyard et al., 1971). Final words were easier to categorize in Italian because of converging evidence from complete linguistic cues required for the semantic and syntactic interpretation of the sentence. In line with this interpretation, a late positivity that developed for strong incongruities in French sentences in the 700–900 ms latency band over parietal sites is likely to reflect the greater difficulty of conducting such analysis in an unfamiliar language. The latency of this effect is similar to the onset latency of the late positivity (600–1000 ms) to strong incongruity in Portuguese sentences reported by Marques et al (2007).

While ERP responses to the strong incongruity condition might reflect automatic perception of a mismatch of the pitch variation in the sentence-final word with the preceding context (Okamoto et al., 2007; Sanders et al., 2009), the congruous and weakly incongruous conditions presumably required more selective attention to the pitch differences. This interpretation is supported by the occurrence of an N2/N3 in most difficult conditions (see Fujioka et al, 2006 and Moreno et al, 2009, for similar interpretations). This effect is clearly observed in the topographic maps of the three conditions (figures 8.3–8.5), which show a similar pattern for congruous and weakly incongruous conditions but a quite different pattern for the strongly incongruous condition.

In order to investigate the extent to which obtained effects were purely due to differences in F0 and not to linguistic components, acoustic analyses were carried out on the materials. In particular, the F0 of the target word and of the two words before the last were measured in each language, in the congruous and weakly incongruous conditions. Preliminary analyses showed that in the congruous condition, in all three languages, F0 did not significantly differ between the antepenultimate and penultimate word, but it decreased at the last word. In the materials used in the present study, the penultimate and the last words were not separated by a syntactic boundary. Pitch decrease in the sentence-final word can be considered as one of universal prosodic signals marking the ending of a sentence, together with pause and prolongation of the last syllables (Jusczyk et al., 1992a). For Italian, jabberwocky and French sentences, the average F0 of the antepenultimate and penultimate words combined were 236, 239, and 239 Hz, and the mean F0 of the final word in the congruous condition were 178, 175 and 194 Hz, respectively. The difference between preceding words and target was significant in all three languages, with a 58 Hz difference for Italian [$t(98) = 24.04$, SE =

2.42, p < .001]; a 64 Hz difference for jabberwocky [t(98) = 28.45, SE = 2.24, p < .001]; and a 45 Hz difference for French [t(98) = 19.84, SE = 2.28, p < .001].

In the weakly incongruous condition, where a 115% increase in F0 was made, F0 of the target word increased to 204, 201, 223 Hz, respectively for Italian, jabberwocky and French. The difference between F0 of the target and that of preceding words was larger in Italian and jabberwocky than in French (32, 38, 16 Hz, respectively), but it was again significant in all three languages [t(98) = 12.35, SE = 2.56 , p < .001; t(98) = 15.84, SE = 2.38, p < .001; t(98) = 6.94, SE = 2.52, p < .001].

These analyses showed that the trend of pitch decrease in the last word was similar in all three languages, suggesting that acoustic information provided important cues for detecting the end of the sentence. Moreover, the fact that the pitch decrease in the last word was larger in Italian and jabberwocky than in French, and that Italian and jabberwocky did not acoustically differ, may partly explain the differences in behavioral data; the level of performance was equivalent for Italian and jabberwocky but lower for French. Further. This effect due to acoustic change probably added to another factor related with the familiarity of the pattern of acoustic variations in the native language.

A further issue is whether the different performance in the three language conditions depended exclusively on acoustic information in the two words preceding the target or on processing of the whole sentence. If judgment on the pitch variations of the target was based only on the preceding words, no difference between Italian and jabberwocky would be expected, as acoustically the two conditions did not differ. By contrast, the ERP data showed significant differences between these languages; positive components in the congruous and weakly incongruous conditions were significantly less marked for jabberwocky than for Italian (see Results section and Figures 8.4 and 8.5). This suggests the relevance of the whole sentence context for evaluation of the pitch variation in the last word.

Moreover, in the strong incongruity condition (see Figure 8.3), the early N1 component was larger for Italian and jabberwocky sentences than for French sentences. The N1 component is typically taken to reflect sensory/perceptual analysis of the stimulus. Thus, perceptual processing of the strong incongruity seems to be influenced by the familiarity with the pattern of pitch contour.

Similar to previous results by Schön et al (2004), Mietz et al. (2008) also reported an early negative component, peaking at around 120 ms after critical word onset, and a large positive component, starting at around 500 ms (P600) elicited by prosodic mismatches. While the overall pattern of their results was similar to that obtained in the present study, the critical word was always preceded by a pause, associated to specific syntactic structures. That is, the final word was preceded by acoustic cues marking phrasal boundaries, such as pause and change in intonation, which could help anticipation of the occurrence of the critical word. By contrast, in the present study the pitch change occurred at the beginning of a lexical word, but never coincided with a syntactic/prosodic boundary. Thus, the early negativity found in the present study was not elicited by a syntactic/ prosodic boundary, where the prosodic variation in pitch contour signals that a phrase ends and another one begins. The increased N1 amplitude reported here for Italian and jabberwocky sentences may rather reflect the segmentation of the auditory input in lexical words and/or increased attention to words with strong pitch changes (see also Sanders et al., 2009).

Several interpretations can be proposed regarding the functional significance of the increased N1 amplitude to strong incongruities for Italian and jabberwocky sentences compared to French sentences: a) processing of the acoustic differences triggered by pitch changes; b) detection of a mismatch between prosody of the sentence context and the word onset, where pitch change occurs; c) segmentation of the last word from the context (Connolly, 1992; Connolly & Phillips, 1994; Sanders & Neville, 2002; Sanders, 2002; Sanders & Neville, 2003); d) selective attention processes associated to the pitch change and individuation of the critical word (Sanders et al., 2009). Obviously these hypotheses are not exclusive and might partially concur to the results.

In sum, the present study provided evidence that although a full (semantic, syntactic, prosodic) interpretation provides a better context for the detection of strong and weak F0 changes, familiarity with a prosodic contour provides a strong support to perform the task. The similarity of results for Italian and jabberwocky sentences shows that participants were able to predict the upcoming pitch change from the prosodic contour of the sentence context, suggesting that familiarity with the prosodic contour is an important element in speech processing (Isel et al., 2005).

9. CONCLUSIONS

The present work added evidence that the long-term experience with one's native language modulates perception of both the segmental and suprasegmental components of speech.

Study I demonstrated that acoustically variable tokens of vowels all belonging to the inventory of listeners' native language are pre-attentively categorized, as indexed by the MMN (Experiment 1). By contrast, such automatic categorization did not occur when a part of vowels did not belong to the native phonemic inventory (Experiment 2). The effect of attention to the stimuli seems to operate even at early stages of auditory processing, modulating sensitivity of afferent neural populations responding to specific acoustic features, as indexed by the N1 component, as well as at later cognitive processes to refine conscious analysis of acoustic dimensions and categorization. In particular, the non-native vowels induced involuntary attention switch in the passive listening condition, and task-related cognitive ERP components that seem to reflect conscious strategies to classify them into novel categories in the active discrimination task (Experiment 2). These results suggest an important role of selective attention in enhancing both early acoustic analysis and later cognitive processes that contribute to improve discrimination of non-native phonetic contrasts.

Study II demonstrated that processing of intonational contour of spoken sentences is facilitated by both familiar prosodic patterns and complete semantic context of the native language, and the whole sentence context serves to form distal prosodic cues to evaluate local pitch changes. The obtained results provide us with information to better understand how suprasegmental aspects and their interaction with other linguistic components (semantics/syntax) contribute to speech comprehension.

The present work dealt with two specific aspects of speech perception, which are functionally entangled with other acoustic properties and linguistic structures in naturally spoken language. I hope that our results contribute to further research on the cognitive and neural processes involved in speech sound perception, on the relationship of perception with production of speech sounds, and also on the interaction of speech sound perception with higher-order linguistic functions. Such perspectives should be usefully developed to improve not only learning methods of foreign languages but also phonological training procedures oriented for dyslexic children or aphasic patients.

REFERENCES

Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., & Lang, A.H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America*, 101, 1090–105.

Abercrombie, D.(1967). Elements of General Phonetics. Edinburgh: Edinburgh University Press.

Alho, K., Donauer, N., Paavilainen, P., Reinikainen, K., Sams, M., & Näätänen, R. (1987a). Stimulus selection during auditory spatial attention as expressed by event-related potentials. *Biological Psychology*, 24, 153–162.

Alho, K., Töttölä, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987b). Brain mechanism of selective listening reflected by eventrelated potentials. *Electroencephalography and Clinical Neurophysiology*, 68, 458–470.

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., Näätänen, R., (1990). Eventrelated brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, 77, 151–155.

Alho, K., Connolly, J.F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., & Ilmoniemi, R.J. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience Letters*, 258, 9–12.

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, 12, 2570–4.

Aslin, R.N., Saffran, J.R., & Newport, E.L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, 9, 321–324.

Astesano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research,* 18(2), 172–184.

Atienza, M., Cantero, J. L., Grau, C., Gomez, C., Dominguez-Marin, E., & Escera, C. (2003). Effects of temporal encoding on auditory object formation: A mismatch negativity study. *Cognitive Brain Research*, 16, 359–371.

Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology*, 45, 495–503.

Banai, K., Nicol, T., Zecker, S.G., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience*, 25, 9850–9857.

Bendixen, A., Roeber, U., & Schröger, E. (2007). Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience*, 19(10), 1664–1677.

Bendixen, A., & Schröger, E. (2008). Memory trace formation for abstract auditory features and its consequences in different attentional contexts. *Biological Psychology*, 78, 231–241.

Bentin, S., Kutas, M., & Hillyard, S.A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, 30(2), 161–9.

Bentin, S., Mouchetant-Rostaing, Y., Giard,M.H., Echallier, J.F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, 11, 235–260.

Berti, S., Roeber, U., & Schröger, E. (2004). Bottom–up influences on working memory: behavioral and electrophysiological distraction varies with distractor strength, *Experimental Psychology*, 51, 249–257.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CV's. *Journal of the Acoustical Society of America*, 82, 31–37.

Best, C.T. (1994). The emergence of language-specific phonemic influences in infant speech perception. (GoodmanJ., NusbaumH C., Ed.). *Development of speech perception: The transition from speech sounds to spoken words*, pp 167–224., Cambridge, MA.

Best, C.T. (1995). A Direct realist perspective on cross-language speech perception. In W. Strange & J.J. Jenkins (eds), *Cross-language speech perception*, Timotium, MD, York Press, pp 171-204.

Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988). Examination of the perceptual re-organization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360.

Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20(3), 305–330.

Best, C. T, Lafleur, R., & McRoberts, G. W. (1995). Divergent developmental patterns for infants' perception of two nonnative contrasts. *Infant Behavior and Development*, 18, 339–350.

Berti, S., Roeber, U., & Schröger, E. (2004). Bottom-up influences on working memory: behavioral and electrophysiological distraction varies with distractor strength, *Experimental Psychology*, 51, 249–257.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S.E., & Mehler, J. (1987). Discrimination in neonates of very short CV's. *Journal of the Acoustical Society of America*, 82, 31–37.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8, 389–395.

Boersma, P., & Weenink, D. (1996). Praat, a system for doing phonetics by computer, version 3.4.

Institute of Phonetic Sciences of the University of Amsterdam, Report 132.

Bradlow, A.R., & Pisoni, D.B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299–2310.

Brunellière, A., Dufour, S., Nguyen, N., Frauenfelder, U.H., (2009). Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition*, 111, 390–396.

Chandrasekaran, B., Krishnan, A., Gandour, J.T. (2007). mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148–156.

Cheour, M., Ceponiente, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351–353.

Cheour, M., Shestakova, A., Alku, P., Čeponienė, R., & Näätänen, R. (2002). Image eponienImage and R. Näätänen, Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters*, 325, 187–190.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics,* 31(3–4), 585–598.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language,* 51(4), 523–547.

Colombo, L. (1992). Lexical stress effect and its interaction with frequency in word pronunciation. Journal of Experimental Psychology: Human Perception and Performance, 18(4), 987–1003.

Colombo, L., Deguchi, C., Boureux, M., Sarlo, M., & Besson, M. (2010). Detection of pitch variation depends upon the familiarità of intonational contour of sentences. Doi: 10.1016/j.cortex.2010.03.008.

Connolly, J.F., Phillips, N.A., Stewart, S.H., & Brake, W.G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, 43(1), 1–18.

Connolly, J.F., & Phillips, N.A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256–266.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology:*

Human Perception and Performance, 19, 909–921.

Crowder, R.G., Serafine, M.L. & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory and Cognition*, 18, 468–476.

Csépe, V., Pantev, C., Hoke, M., Hampson, S., & Ross, B. (1992). Evoked magnetic responses of the human auditory cortex to minor pitch changes: localization of the mismatch field. *Electroencephalography and Clinical Neurophysiology*, 84, 538–48.

Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.

Cutler, A., & Norris, D.G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 113–121.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech,* 40 (2), 141–201.

Cycowicz, Y.M, & Friedman, D. (1998). Effect of sound familiarity on the eventrelated potentials elicited by novel environmental sounds. *Brain and Cognition*, 36, 30–51.

Deguchi, C., Chobert, J., Brunellière, A., Dufour, S., Nguyen, N., & Besson, M. (2010) Pre-attentive and attentive processing of French vowels, *Brain Research*, 1366, 149–161.

Delattre, P.C., Liberman, A.M., Cooper, F.S., & Gerstman, L.J., (1952). An experimental study of the acoustic determinants of vowel colour: Observation on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 8, 195–210.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8, 919–924.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological Correlates of Phonological processing : a cross-linguistic study. *Journal of Cognitive Neuroscience*, 12(4), 635–647.

Dehaene-Lambertz, G., & Pena, M. (2001) Electrophysiological evidence for automatic phonetic processing in neonates, *NeuroReport*, 12, 3155–3158.

Di Cristo, A., & Hirst, D. (1998). *Intonation systems*. Cambridge: Cambridge university press.

Diesch, E., & Luce, T., (2000). Topographic and temporal indices of vowel spectral envelope extraction in the human auditory cortex. *Journal of Cognitive Neuroscience*, 12, 878–893.

Dilley, L., & McAuley, J.D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language,* 59(3), 294–311.

D'Imperio, M., Petrone, C. & Nguyen, N. (2007). Effects of tonal alignment on lexical identification in Italian. In Gussenhoven C. and Riad T. (Eds), *Tones and tunes.* Berlin: Mouton de Gruyter, 79–106.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 3697–3709.

Donchin, E., & Coles, M., (1988). Is the P300 component a manifestation of context updating? *Behavioral Brain Science*, 11, 357–374.

Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, 36, 406–421.

Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306.

Eimas, P.D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 16, 513–521.

Eimas, P.D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, 18, 341–347.

Eimas, P.D., & Miller, J.L. (1980). Discrimination of the information for manner of articulation. *Infant Behavior & Development*, 3, 367–375.

Eisner, F., & McQueen, J.M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67 (2), 224–238.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research*, 10, 345–348.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention switching to novelty and change in the acoustic environment. *Journal of Cognitive Neuroscience*, 10, 590–604.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro-Otology*, 5, 151–166.

Flege, J.E. (1995). Second language speech learning theory, findings, and problems. In (ed. W. Strange), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.

Friederici, A.D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive*

Brain Research, 1(3), 183–92.

Friederici, A.D. & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, 25, 157–176.

Friederici, A.D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: L earning, Memory, and Cognition*, 22, 1219–1248.

Friederici, A.D., Meyer, M., & von Cramon, D.Y. (2000a). Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language*, 74, 289–300.

Friederici, A.D., Wang, Y.H., Herrmann, C.S., Maess, B., & Oertel, U. (2000b). Localization of early syntactic processes in frontal and temporal cortical areas: A magnetoencephalographic study. *Human Brain Mapping*, 11(1), 1–11.

Friederici, A.D., & Kotz, S.A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage*, 20, 8–17.

Friederici, A.D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model, *Brain and Language*, 89, 267–276.

Friederici, A.D., von Cramon, D.Y. & Kotz, S.A. (2007). Role of the corpus callosum in speech comprehension: Interfacing syntax and prosody. *Neuron*, 53, 135–145.

Friedman, D., Cycowicz, Y.M, & Gaeta, H. (2001).The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25, 355–373.

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171–189.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain, 129,* 2593–2608.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*, 84(3), 318–336.

Garner, W.R. (1974). *The processing of information and structure*. Lawrence Erlbaum Associates, Potomac, MD.

Giard, M.H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology*, 27, 627–40.

Giard, M.H., Perrin, F., Echallier, J.F., Thevenet, M., Froment J.C., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology*, 92, 238–252.

Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Thévenet, M., Pernier, J., & Naatanen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory. *Journal of Cognitive Neuroscience*, 7, 133–43.

Gomes, H., Ritter, W., & Vaughan Jr., H.G. (1995). The nature of preattentive storage in the auditory system. *Journal of Cognitive Neuroscience*, 7, 81–94.

Gomes, H., Bernstein, R., Ritter, W., Vaughan Jr., H.G., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, 34, 712–716.

Gomez, R.L., & Gerken, L.A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia*, 9, 317–323.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, 12, 91–114.

Grabe, E., & Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Laboratory Phonology*, 7 (Eds. Gussenhoven & Warner; Mouton [Berlin]), 515–546.

Hahne, A., & Friederici, A.D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience,* 11(2), 194–205.

Hansen, J.C., & Hillyard, S.A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, 49, 277–290.

Hansen, J.C., & Hillyard, S.A. (1983). Selective attention to multidimensional auditory stimuli in man. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 1–9.

Hansen, J.C., & Hillyard, S. A. (1984). Effects of stimulation rate and attribute cuing on event-related potentials during selective auditory attention. *Psychophysiology*, 21, 394–405.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.

Hillenbrand, J.M. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research*, 26, 268–282.

Hillenbrand, J.M., Minifie, F.D., & Edwards, T.J. (1979). Tempo of spectrum change as a cue in speech sound discrimination by infants. *Journal of Speech and Hearing Research*, 22, 147–165.

Hillyard, S.A., Squires, K.C., Bauer, J.W., & Lindsay, P.H. (1971). Evoked potential correlates of auditory signal detection. *Science, 172*, 1357–1360.

Hillyard, S.A., Hink, R.F., Schwent, V.L., & Picton, T.W. (1973). Electrical signs of selective attention in the human brain. Science, 182, 177–180.

Hillyard, S.A., Woldorff, M., Mangun, G..R., & Hansen, J.C. (1987). Mechanisms of early selective attention in auditory and visual modalities. *Electroencephalography and Clinical Neurophysiology*, *Suppl. 39*, 317–24.

Hisagi, M., Shafer, V.L., Strange, W., & Sussman, E.S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures. *Brain Research*, 1360, 89–105.

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, 79, 139–147.

Houston, D.M., & Jusczyk, P.W. (2000). The role of talkerspecific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception & Performance*, 26, 1570–1582.

Houston, D.M., Jusczyk, P.W., & Tager, J. (1998). Talker-Specificity and the Persistence of Infants' Word Representations. In A. Greenhill et al (Eds.) *Proceedings of the 22nd Annual Boston University Conference on Language Development*, (pp. 385–396). Somerville, MA : Cascadilla Press.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *NeuroReport*, 12, 1749–1752.

Huotilainen, M., Ilmoniemi, R.J., Lavikainen, J., Tiitinen, H., Alho, K., Sinkkonen, J., Knuutila, J. & Näätanen, R. (1993). Interaction between representations of different features of auditory sensory memory. *NeuroReport*, 4, 1279–81.

Huotilainen, M., Kujala, A., Alku, P., (2001). Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neuroscience Letters*, 310, 133–136.

Huotilainen, M., Kujala, A., Hotakainen, M., Parkkonen, L., Taulu, S., Simola, J., Nenonen, J., Karjalainen, M., Näätänen, R. (2005). Short-term memory functions of the human fetus recorded with magnetoencephalography. *NeuroReport*, 16, 81–84.

Isel, F., Alter, K., & Friederici, A.D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience,* 17(1):

154–167.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553–562.

Iverson, P. & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99, 1130–1140.

Iverson, P & Kuhl, P.K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*. 62(4), 874–886.

Jacobsen, T., Schröger, E., & Sussman, E. (2004a). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research*, 20, 473–479.

Jacobsen, T., Schröger, E., & Alter, K. (2004b). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, 41, 654–659.

Jakobson, R., Fant, C.M. & Halle M. (1952). *Preliminaries to Speech Analysis*. Cambridge MA, MIT Press.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P. Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research*, 12, 459-466.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.

Jemel, B., Achenbach, C., Müller, B.W., Röpcke, B., & Oades, R.D. (2002). Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topography*, 15, 13–27.

Joanisse, M.F., & Gati, J.S. (2003). Overlapping neural regions for the processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, 19, 64–79.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567.

Johnson, K.L., Nicol, T., Zecker, S.G., Bradlow, A.R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant-vowel stop syllables. *Clinical Neurophysiology*, 119, 2613–2635.

Jusczyk, P.W. (1999). Narrowing the distance to language: One step at a time. *Journal of Communication Disorders*, 32, 207–222.

Jusczyk, P.W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L.J., & Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development*, 5, 265–286.

Jusczyk, P.W., Hirsh-Pasek, K.; Nelson, D. J.K., Kennedy, L.J., Woodward, A., & Piwoz, J. (1992a). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.

Jusczyk, P.W., Pisoni, D.B., & Mullennix, J. (1992b). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition*, 43, 253–291.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M.I, Svenkrud, V.Y. & Jusczyk, A.M. (1993) Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402–420.

Jusczyk, P.W., Luce, P.A., & Charles Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630–645.

Jusczyk, P.W., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.

Kaganovich, N., Francis, A.L., & Melara, R.D., (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, 1114, 161–172.

King, J.W., & Kutas, M. (1995). Who did what and when? Using word- and clauselevel ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, J. (2008). Modulation of mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology*, 67, 131–143.

Knösche, T., Lattner, S., Maess, B., Schauer, M., & Friederici, A.D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage*, 17, 1493–1503.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38, 557–577.

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, R. J., Virtanen, J., Kropotov, J., & Näätänen, R. (1999). Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neuroscience Letters*, 276, 169–172.

Kraus, N., McGee, T., Carrell, T.D., King, C., Tremblay, K., & Nicol, N. (1995). Central auditory plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, 7, 25–32.

Kraus, N., & Nicol, T.G. (2005). Brainstem origins for cortical "what" and "where" pathways in the auditory system. *Trends in Neuroscience*, 28, 176–181.

Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1668–1679.

Kuhl, P.K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6, 263–285.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93–107.

Kuhl, P.K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 259–274). Dordrecht: Kluwer.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.

Kuhl, P.K., & Miller, J.D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics*, 31, 279–292.

Kuhl, P. K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.

Kuhl, P.K., Andruski, J.E., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units addressed to infants. *Science*, 277, 684–686.

Kuhl, P. K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 979–1000.

Kujala, T., Myllyviita, K., Tervaniemi, M., Alho, K., Kallio, J., & Näätänen, R. (2000). Basic auditory dysfunction in dyslexia as demonstrated by brain activity measurements. *Psychophysiology*, 37, 262–266.

Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R., (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, 112, 1712–1719.

Kujala, T., Tervaniemi, M., & Schröger, E., (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74, 1–19.

Kujala, T., & Näätänen, R., (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology*, 91, 55–67.

Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.

Lee, L., & Nusbaum, H.C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, 53(2), 157–165.

Levänen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one versus two stimulus features. *Experimental Brain Research*, 97, 177–83.

Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex*, 6, 288–96.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). The perception of place of articulation contrasts in voiced and voiceless fricatives by two-month-old infants. *Journal of Experimental Psychology: Human Perception & Performance*, 14, 361–368.

Liberman, A.M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117–123.

Liberman, A.M., Harris, K.S., Kinney, J.A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61(5), 379–388.

Liberman, A.M., Cooper F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–61.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., & Medler, D.A. (2005). Neural Substrates of phonemic perception. *Cerebral Cortex*, 15, 1621–1631.

Lotto, A.J., Kluender, K.R., & Holt, L.L. (1998). The Perceptual Magnet Effect depolarized. *Journal of the Acoustical Society of America*, 103, 3648–3655.

Lutkenhoner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinstrater, O., & Patterson, R.D. (2003). Localization of primary auditory cortex in humans by magnetoencephalography. *NeuroImage*, 18, 58–66.

Magne, C, Astésano, C, Aramaki, M, Ystad, S, Kronland-Martinet, R , & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioural and electrophysiological evidence, *Cerebral Cortex*, 17(11), 2659–2668.

Mangun, G.R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, 32, 4–18.

Marques, C., Moreno, S., Castro, S.L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence.

Journal of Cognitive Neuroscience, 19(9): 1453–1463.

Martin, B.A., Sigal, A., Kurtzberg, D., Stapells, D.R., (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America*, 101, 1585–1599.

Marslen-Wilson, W.S., Tyler, L.K., Warren, P., Genier, P., & Lee, C.S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 45A(1), 73–87.

Mattys, S.L., Jusczyk, P.W., Luce, P.A., & Morgan, J.L. (1999). Word segmentation in infants: How phonotactics and prosody combine. *Cognitive Psychology*, 38, 465–494.

Maye, J., Werker, J. F., & Gerken, L.A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5, 467–479.

McCallum, W. C., Curry, S. H., Cooper, R., Pocock, P. V., & Papakostopoulos, D. (1983). Brain event-related potentials as indicators of early selective processes in auditory target localization. *Psychophysiology*, 20, 1–17.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tilson, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 144–178.

Mehler, J., & Christophe, A. (1995). Maturation and learning of language in the first year of life. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences: A handbook for the field* (pp. 943–954). Cambridge, Mass.: MIT Press.

Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, Mora-timed contrasts of the Japanese language. *Learning and Memory*, 9, 253–67.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., & von Cramon, D.Y. (2002). Functional MRI reveals brain regions mediatine slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17, 73–88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., & von Cramon, D.Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89, 277–289.

Michie, P.T., Solowij, N., Crawford, J.M., & Glue, L.C. (1993). The effects of between-source discriminability on attended and unattended auditory ERPs. *Psychophysiology*, 30, 205–220.

Mietz, A., Toepel, U., Ischebeck, A., & Alter, K. (2008). Inadequate and infrequent are not alike:

ERPs to deviant prosodic patterns in spoken sentence comprehension. *Brain and Language,* 104(2), 159–169.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, 18, 331–340.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S.L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex, 19*, 712–723.

Myers, J., Jusczyk, P.W., Kemler Nelson, D.G., Charles-Luce, J., Woodward, A., & Hirsh-Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language*, 23, 1–30.

Näätänen, R., (1982). Processing negativity: An event-related potential reflection of selective attention. *Psychological Bulletin*, 92, 605–640.

Näätänen, R., (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral Brain Science*, 13, 201–288.

Näätänen, R., (1991). Mismatch negativity (MMN) outside strong attentional focus: a commentary on Woldorff et al. *Psychophysiology*, 28, 478–84.

Näätänen R. (1992). *Attention and Brain Function*. Hillsdale, N.J: Erlbaum.

Näätänen, R., (1995). The Mismatch Negativity: A Powerful Tool for Cognitive Neuroscience. *Ear & Hearing*, 16, 6–18.

Näätänen, R., (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S., (1978). Early selective attention effect on evoked potential reinterpretes. *Acta Psychologica*, 42, 312–329.

Näätänen, R., & Michie, P.T. (1979). Early selective attention effects on the evoked potential. A critical review and reinterpretation. *Biological Psychology*, 8, 81–136.

Näätänen, R., & Picton, T.W., (1987). The N1 wave of the human electric and magnetic response to sound: A review and analysis of the component structure. *Psychophysiology*, 24, 375–425.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1987). The mismatch negativity to intensity changes in an auditory stimulus sequence. *Electroencephalography Clinical Neurophysiology, Suppl. 40*, 125–31.

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E.N. (1988).

Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, 69, 523–531.

Näätänen, R., Paavilainen P, Alho K, Reinikainen K, & Sams M. (1989). Do eventrelated potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters*, 98, 217–21.

Näätänen, R., Teder, W., Alho, K., & Lavikainen, J. (1992). Auditory attention and selective input modulation: a topographical ERP study. *NeuroReport*, 3, 493–496.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993a). Attention and mismatch negativity. *Psychophysiology*, 30, 436–50.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993b). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, 4, 503–506.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheor, M., Huotilainen, M., Iivonen, R., Vainio, M., Ilmonoiemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434.

Näätänen, R.,& Winkler, I. (1999). The concept of auditory stimulus representation in neuroscience. *Psychological Bulletin*, 125, 826–859.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., &Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, 24, 283–288.

Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R. (2004). The mismatch negativity (MMN) - towards the optimal paradigm. *Clinical Neurophysiology*, 115, 140–4.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544–2590.

Näätänen R, Kujala T, & Winkler I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, 48, 4–22.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, 24(3), 756–766.

Nazzi, T., Kemler Nelson, D.G., Jusczyk, P.W., & Jusczyk, A.M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1, 123–147.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech–sound duration processing. *Cognitive Brain Research*, 16, 492–495.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R., (2005). Speech–sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, 92, 26–32.

Neville, H., Nicol, J.L., Barss, A., Forster, K.I., & Garrett, M.F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.

Newman, R.L., Connolly, J.F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, 40, 640–647.

Newman, R.L., & Connolly, J.F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, 80(1), 114-121.

Nousak, J.M.K., Deacon, D., Ritter, W., & Vaughan Jr., H.G. (1996). Storage of information in transient auditory memory. *Cognitive Brain Research*, 4, 305–317.

Novak, G., Ritter, W., & Vaughan, H.G. (1992). The chronometry of attention-modulated processing and automatic mismatch detection. *Psychophysiology*, 29, 412–430.

Novitski, N., Tervaniemi, M., Huotilainen, M., & Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research*, 20, 26–36.

Obleser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15, 207–213.

Obleser J., Lahiri, A., & Eulitz, C. (2004). Magnetic Brain Response Mirrors Extraction of Phonological Features from Spoken Vowels. *Journal of Cognitive Neuroscience*, 16(1), 31–39.

Okamoto, H., Stracke, H., Wolters, C.H., Schmael, F., & Pantev ,C. (2007).Attention improves population-level frequency tuning in human auditory cortex. *Journal of Neuroscience,* 27(39), 10383–90.

Okita, T., Konishi, K., & Inamori, R. (1983). Attention-related negative brain potential for speech words and pure tones. *Biological Psychology*, 16, 29–47.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, 1, 203–209.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, 14, 1–14.

Ostroff, J.M., Martin, B.A., & Boothroyd, A. (1998). Cortical responses to acoustic change within a syllable. *Ear & Hearing*, 19, 290–297.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology*, 78, 466–79.

Paavilainen, P., Saarinen, J., Tervaniemi, M., & Näätänen, R. (1995). Mismatch negativity to changes in abstract sound features during dichotic listening. *Journal of* Psychophysiology, 9:243–9.

Paavilainen, P., Jaramillo, M., Näätänen, R, & Winkler I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters*, 265, 179–82.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, 38, 359–65.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport*, 18, 159–163.

Pakarinen, S., Takegata, R., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*, 118, 177–185.

Pakarinen, S., Huotilainen, M., & Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology*, 121, 1043–1050.

Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in vowel acquisition. *Cognition*, 64, B9–B17.

Pallier, C., Colomé, A., & Sebastian-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445–449.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A.D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience,* 17(3), 407–421.

Picton, T.W., Hillyard, S.A., Krausz, H.I., & Galambos, R. (1974). Human auditory evoked potentials: I. Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, 36, 179–190.

Picton, T.W., Stapells, D.R., & Campbell, K.N. (1981). Auditory evoked potentials from the human cochlea and brainstem. *Journal of Otolaryngology*, 9 *(Supplement)*, 1–41.

Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., & Trujillo, N.J. (1999) Intracerebral sources of human auditory-evoked potentials. *Audiology & Neuro-Otology*, 4, 64–79.

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from entence processing: Activation interacts with taks demands. *NeuroImage*, 17, 401–410.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T.P., Rowley, H.A., Wexler, K. & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. Brain Research: *Cognitive Brain Research*, 4, 231–242.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code, *Cognitive Science*, 25, 679–693.

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148.

Polka,L., & Werker, J.F. (1994). Developmentalchangesinperception of nonnative vowel contrasts. Journal *of Experimental Psychology: Human Perception & Performance*, 20, 421–435.

Pollack, I. (1952). The information in elementary auditory displays. *Journal of the Acoustical Society of America*, 24, 745–749.

Pollack, I. (1953). The information in elementary auditory displays II. *Journal of the Acoustical Society of America*, 25, 765–769.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity (MMN). *NeuroImage*, 14, 607–616.

Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *NeuroImage*, 20, 159–172.

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, 41, 106–112.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R.J., (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, 17, 884–892.

Pulvermüller, F., & Shtyrov, Y., (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, 79, 49–71.

Pulvermüller, F., & Assadollahi, R. (2007). Grammar or serial order?: Discrete combinatorial brain mechanisms reflected by the syntactic mismatch negativity. *Journal of Cognitive*

Neuroscience, 19, 971–980.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110, 81–94.

Pylkkanen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, 81, 666–678.

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., & Näätänen, R. (1999). Analysis of speech sounds is left-hemisphere predominant at 100–150 ms after sound onset. *NeuroReport*, 10, 1113–7.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal mismatch negativity sources. *NeuroImage*, 12, 14–9.

Rinne, T., Sarkka, A., Degerman, A., Schröger, E., & Alho, K. (2006).. Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, 1077, 135–143.

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior collicus. *Journal of Neurophysiology*, 100, 3323–3327.

Ritter, W., Simson, R., Vaughan, H.G., & Friedman, D. (1979). A brain event related to the making of sensory discrimination. *Science*, 203, 1358–1361.

Ritter, W., Simson, R., & Vaughan, H.G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination task. *Psychophysiology*, 20, 168–179.

Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan Jr., H.G. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology*, 36, 835–838.

Rosburg, T. (2003). Left hemispheric dipole locations of the neuromagnetic mismatch negativity to frequency, intensity and duration deviants. *Cognitive Brain Research*, 16, 83–90.

Rosburg, T., Haueisen, J., & Kreitschman-Andermahr, I. (2004). The dipole location shift within the auditory evoked neuromagnetic field components N100m and mismatch negativity (MMNm). *Clinical Neurophysiology*, 115:906–13.

Rossi, M. (1998). Intonation in Italian. In Hirst and Di Cristo A. (Eds), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 219–238.

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: an event-related potential study with ringtones, *European Journal of Neuroscience*, 26 (3), 784–790.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract stimulus attributes in human brain. *NeuroReport*, 3, 1149–51.

Sable, J. J., Low, K. A., Maclin, E. L., Fabiani, M., & Gratton, G. (2004). Latent inhibition mediates N1 attenuation to repeating sounds. *Psychophysiology*, 41, 636–642.

Saffran, J., Aslin, N. & Newport, E.L. (1996) Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Sams, M., Hämäläinen, M., Antervo, A., Kaukoranta, E., Reinikainen, K., & Hari, R. (1985). Cerebral neuromagnetic responses evoked by short auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 61, 254–266.

Sams, M., Kaukoranta, E., Hämäläinen, M,. & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, 28, 21–9.

Sanders, L.D., & Neville, H.J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience,* 5(7), 700–703.

Sanders, L.D., Neville, H.J., & Woldorff, M. (2002). Speech segmentation of native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research,* 45(3), 519–530.

Sanders, L.D., & Neville, H.J. (2003). An ERP study of continuous speech processing. I. segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research,* 15(3), 228–240.

Sanders, L.D., Ameral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia,* 47(4), 1183–1186.

Schafer, A.J., Speer, S.R., Warren, P., & White, S.D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research,* 29(2), 169–182.

Shafer, L.V., Schwartz, R.G., Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, 18, 242–254.

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, 65, 344–360.

Scherg, M., Vajsar, J., Picton, T.W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, 1, 336–55.

Schlosser, M.J., Aoyagi, N., Fulbright, R.K., Gore, J.C., & McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Human Brain Mapping*, 6, 1–13.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology,* 41(3), 341–349.

Schröger, E. (1996a). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience*, 8, 527–39.

Schröger, E. (1996b). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology*, 100, 517–26.

Schröger, E., (1998). Application of ERP measures in psychological research: Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers*, 30, 131–145.

Schröger, E., & Wolff, C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: a new distraction paradigm. *Cognitive Brain Research*, 7, 71–87.

Schröger, E., Bendixen, A., Trujillo-Barreto, N.J., & Roeber, U. (2007). Processing of abstract rule violations in audition. *PLoS One*, 11, e1131.

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *NeuroReport*, 9, 337–40.

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1999). Pre-attentive processing of auditory patterns in dyslexic human subjects. *Neuroscience Letters*, 276, 41–4.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R.J., & Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *NeuroReport*, 13, 1813–1816.

Shestakova, A., , Brattico, E., Soloviev, A., Klucharev, V., & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. Cognitive *Brain Research*, 21, 342–350.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, 12, 657–663.

Shtyrov, Y., & Pulvermüller, F. (2002). Memory traces for in¯ectional affixes as shown by mismatch negativity. *European Journal of Neuroscience*, 15, 1085–1091.

Shtyrov, Y., Pulvermüller, F., Näätänen, R., & Ilmoniemi, R.J. (2003). Grammar processing outside the focus of attention: An MEG study. *Journal of Cognitive Neuroscience*, 15, 1195–1206.

Shtyrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European Journal of Neuroscience*, 19, 1083–92.

Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *NeuroImage*, 27, 37–47.

Shtyrov, Y., & Pulvermüller, F. (2007). Language in the Mismatch Negativity Design Motivations, Benefits, and Prospects. *Journal of Psychophysiology*, 21(3–4), 176–187.

Simson, R., Vaughan, H.G., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual discrimination tasks. *Electroencephalography and Clinical Neurophysiology*, 42, 528–535.

Speer, S.R., Crowder, R.G., & Thomas, L.M. (1993). Prosodic Structure and Sentence Recognition. *Journal of Memory and Language*, 32, 336–358.

Squires, N.K., Squires, K.C., & Hillyard, S.A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387–410.

Starr, A., & Don, M. (1988). Brain potentials evoked by acoustic stimuli. In T. W. Picton (Ed.), *Human event-related potentials* (Vol. 3, pp. 97–157). Amsterdam: Elsevier.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 8*6(1), 142–164.

Steinhauer, K., Alter, K., & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience,* 2(2), 191.

Steinhauer, K., & Friederici, A.D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research,* 30(3), 267–295.

Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language*, 90, 88–94.

Strange, W., Weber, A., Levy, E.S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America*, 122, 1111–1129.

Sussman, E., Winkler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, 40, 430–435.

Swoboda, P., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at-risk infants. *Child Development*, 47, 459–465.

Swoboda, P., Kass, J., Morse, P. A, & Leavitt, L. A. (1978). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Development*, 49, 332–339.

Szymanski, M.D., Yund, E.W., Woods, D.L. (1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *Journal of the Acoustical Society of America*, 106, 3492–3504.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity (MMN). *Neuroscience Letters*, 266, 109–112.

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *NeuroReport*, 12, 525–529.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for simultaneous, spatially distributed auditory objects. *Cognitive Brain Research*, 25, 169–179.

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, & Näätänen, R. (1994a). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, 38, 157–67.

Tervaniemi, M., Maury, S., & Näätänen, R. (1994b). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, 5, 844–846.

Tervaniemi, M., Jacobsen, T., Röttger, S., Kujala, T., Widmann, A., Vainio, M., Näätänen, R., & Schröger, E. (2006). Selective tuning of cortical sound feature processing by language experience, *European Journal of Neuroscience*, 23, 2538–2541.

Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R.J., Simola, J., & Näätänen, R. (1993). Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. *Psychophysiology*, 30, 537–40.

Tiitinen, H., May, P., Reinkainen, K., & Näätänen, R. (1994). Attentive novelty detected in humans is governed by pre-attentive sensory memory. *Nature*, 372, 90–92.

Trehub, S.E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, 9, 91–96.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiologic changes during speech-sound training. *NeuroReport*, 9, 3557–3560.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B., (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear & Hearing*, 22, 79–90.

van Zuijen, T.L., Simoens, V.L., Paavilainen, P., Näätänen, R., & Tervaniemi, M., (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *Journal of Cognitive Neuroscience*, 18, 1292–1303.

Vaughan, H.G., Jr., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, 28, 360–367.

Vaughan, H.G. Jr., & Arezzo, J.C. (1988). The neural basis of eventrelated potentials. In T. W. Picton (Ed.), *Human event-related potentials* (Vol. 3, pp. 45–96). Amsterdam: Elsevier.

Velasco, M., & Velasco, F. (1986). Subcortical correlates of the somatic, auditory and visual vertex activities. II. Referential EEG responses. *Electroencephalographpy and Clinical Neurophysiology*, 48, 377–386.

Velasco, M., Velasco, F., & Olvera, A. (1985). Subcortical correlates of the somatic, auditory and visual vertex activites in man. I. Bipolar EEG responses and electrical stimulation. *Electroencephalography and Clinical Neurophysiology*, 61, 519–529.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649 –3658.

Wang, J., Friedman, D., Ritter, W., & Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology*, 42, 43–55.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Language and Cognitive Processes,* 10(5), 457–486.

Werker, J.F., & Lalonde, C.E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24, 672–683.

Werker, J. F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.

Werker, J.F., & Tees, R.C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Devopmental Psychobiology*, 46(3), 233–51.

Whiting, K.A., Martin, B.A., & Stapells, D.R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear & Hearing*, 19, 218–231.

Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, 21, 147–163.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Livonen, A., & Näätänen, R. (1999a). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7, 357–369.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R.J., & Näätänen, R. (1999b). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638–642.

Winkler, I., Czigler, I., Sussman, E., Horváth, J., & Balázs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, 17, 320–339.

Wolach, I., & Pratt, H. (2001). The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. *Clinical Neurophysiology*, 112, 186–197.

Woldorff, M., & Hillyard, S.A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, 79, 170–191.

Woldorff,M. G., Hackley, S. A., & Hillyard, S. A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28, 30–42.

Woldorff, M., Hillyard, S., Gallen, C.C., Hampson, S.R., & Bloom, F.E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, 35, 283–292.

Wolpaw, J.R., & Penry, J.K. (1977). Hemispheric differences in the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, 43, 99–102.

Woods, D.L. (1995). The component structure of the N1 wave of the human auditory evoked potential. In G. Karmos, M. Molnár, V. Csépe, & J.E. Desmedt (Eds.), *Perspective of Event-Related Potentials Research (EEG Suppl. 44)*, Amsterdam: Elsevier Science.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, 8, 1971–1974.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35, 615–619.

Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., & Kaneko, S. (2001a). Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Research*, 897, 222–227.

Yabe, H., Koyoma, S., Kakigi, R., Gunji, A., Tervaniemi, M., Sato, Y., & Kaneko, S. (2001b). Automatic discriminative sensitivity inside temporal window of sensory memory as a function of time. *Cognitive Brain Research*, 12, 39–48.

Yago, E., Escera, C., Alho, K. & Giard, M.H. (2001). Cerebral mechanisms underlying orienting of attention towards auditory frequency changes. *NeuroReport*, 12, 2583–2587.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Research*, 1072, 175–185.

Zatorre, R.J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953.

Zatorre, R.J., Belin, P., & Penhue, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1), 38–46.