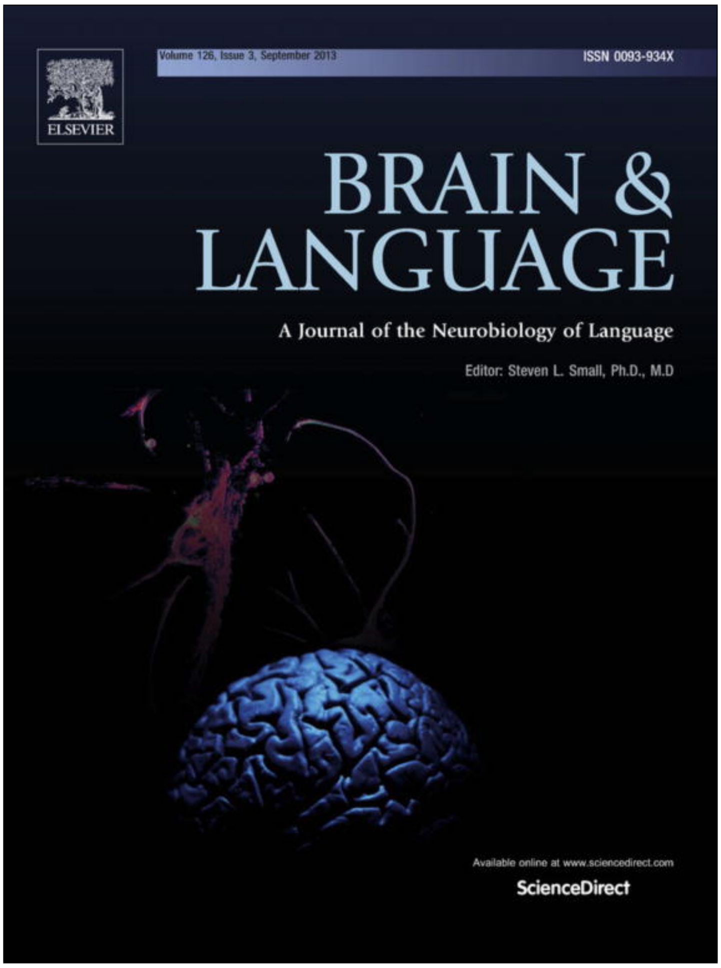
Volume 126, Issue 3, September 2013
ISSN 0093-934X
ELSEVIER
BRAIN &
LANGUAGE
A Journal of the Neurobiology of Language
Editor: Steven L. Small, Ph.D., M.D
Available online at www.sciencedirect.com
ScienceDirect

Brain & Language 126 (2013) 285–290

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Short Communication

Conditioned allophony in speech perception: An ERP study ☆

Sandra Miglietta [a,*], Mirko Grimaldi [b], Andrea Calabrese [c]

[a] Dipartimento Antichità, Medioevo e Rinascimento, Linguistica, Università di Firenze, Italy
[b] Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL), Dipartimento di Studi Umanistici, Università del Salento, Italy
[c] Department of Linguistics, University of Connecticut, USA

ARTICLE INFO

Article history:
Accepted 20 June 2013
Available online 2 August 2013

Keywords:
Event-related potentials (ERPs)
Phoneme perception
Allophone perception
Native vowel contrasts
Mismatch Negativity (MMN)

ABSTRACT

A Mismatch Negativity (MMN) study was performed to investigate whether pre-attentive vowel perception is influenced by phonological status. We compared the MMN response to the acoustic distinction between the allophonic variation [ε–e] and phonemic contrast [e–i] present in a Southern-Italian variety (Tricase dialect). Clear MMNs were elicited for both the phonemic and allophonic conditions. Interestingly, a shorter latency was observed for the phonemic pair, but no significant amplitude difference was observed between the two conditions. Together, these results suggest that for isolated vowels, the phonological status of a vowel category is reflected in the latency of the MMN peak. The earlier latency of the phonemic condition argues for an *easier* parsing and encoding of phonemic contrasts in memory representations. Thus, neural computations mapping auditory inputs into higher perceptual representations seem 'sensitive' to the contrastive/non-contrastive status of the sounds as determined by the listeners' knowledge of the own phonological system.

1. Introduction

Traditionally, sound units in a language can be divided into phonemes, i.e., linguistically contrastive units that cannot be substituted for one another without a change in meaning at the word-level, and allophones, i.e., linguistically non-contrastive variants of a phoneme that appear regularly in specific phonological contexts (Trubetzkoy, 1969). One type of phonological process that generates conditioned allophones is assimilation, by which a sound is changed to resemble a nearby sound.

The focus of the current study is the assimilatory process present in a Southern-Italian dialect (Tricase; Southern Apulia). This variety shows a phonemic five-vowel system (/i, ε, a, ɔ, u/) and presents a vowel-to-vowel assimilatory process that raises the stressed low-mid front vowels to their high-mid counterparts before a high front vowel. Previous studies (Grimaldi, 2006; Grimaldi, Calabrese, Sigona, Garrapa, & Sisinni, 2010) based on acoustic and articulatory data showed that this process produces the allophonic variation [ε–e], among other adjustments, by spreading the feature specification [+ATR] from the final high front [i] vowel ([+high, +ATR]) to the stressed mid vowel [ε] ([−high, −ATR]) in both open and closed syllables: [ˈmεte]/[ˈmeti] *I/he reap(s)*; [ˈdεnte]/[ˈdenti] *tooth/teeth*, etc.

There are few behavioral studies that explicitly address the perception of conditioned allophones. For example, Peperkamp, Pettinato, and Dupoux (2003) showed that French listeners had difficulty in perceiving the allophonic variation [ʁ–χ] compared to the phonemic contrast [m–n]. Similarly, Boomershine, Hall, Hume, and Johnson (2008) analyzed the attentive discrimination of the distinction between [d–ð] (phonemic in English but allophonic in Spanish) and [d–ɾ] (in which the reverse holds). Both language groups judged the phonemic pair as more different. The identical scenario is observed for vowels in Pallier, Bosch, and Sebastián-Gallés (1997) who observed that bilingual Spanish–Catalan speakers with native Spanish-speaking parents did not discriminate the vowel pair [ε–e] (phonemic in Catalan but allophonic in Spanish), whereas bilingual speakers with native Catalan-speaking parents accurately discriminated the vowel pair for the complex distribution of mid-vowel allophones in Spanish (see Navarro Tomás (1918).

Recently, several Mismatch Negativity (MMN) studies have been performed to further investigate speech perception. The MMN is elicited in an oddball paradigm. Frequent *standard* stimuli that share certain regularities are presented, and infrequently, sounds that *deviate* from the regularity of the standard events are interspersed (for a recent review Näätänen, Kujala, & Winkler, 2011). The MMN (and its magnetic counterpart MMNm) is an early response component of the auditory event-related potential (ERP) that reflects changes in a regular auditory sequence. Most MMN studies that have investigated phonemic and phonetic speech perception are cross-linguistic studies. For instance, Phillips et al. (1995) investigated the perception of the acoustic distinction

☆ The authors contributed equally to the overall experimental design, execution of the experiments, analysis of data and writing of the paper.

* Corresponding author. Address: University of Florence, Dipartimento di Scienze dell'Antichità, Medioevo e Rinascimento e Linguistica, Piazza Brunelleschi, 3-4, 50121 Firenze (FI), Italy. Fax: +39 55 2476808.
E-mail address: sandra.miglietta@unifi.it (S. Miglietta).

0093-934X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bandl.2013.06.001

between [r] and [l] in English and Japanese speakers. The two sounds are phonemes in English but not in Japanese, which only contains [r] in its phonemic system. The authors showed that only English speakers elicited an MMNm. Similarly, Winkler et al. (1999) observed that Hungarians did not show an MMN for the Finnish vowel contrast [e–æ] not present in their language. Briefly, the phonetic variations analyzed in these previous studies involved sounds that were alien to the phonological systems of the subjects and not language-specific allophones.

To our knowledge, only two cross-linguistic MMN studies have addressed the perception of conditioned allophones. Hacquard, Walter, and Marantz (2007) addressed (not explicitly) the pre-attentive perception of a conditioned allophonic variation by comparing the perception of the vowel pair [ɛ–e] in Spanish and French listeners, in which [ɛ–e] is a phonemic contrast in French but involves allophonic variation in Spanish. Notably, the results showed a similar MMNm response for both language groups. However, the study of Hacquard et al. (2007) was based on an oddball-design, in which standard and deviant stimuli corresponded to a single token for each sound type. In contrast, Kazanina, Phillips, and Idsardi (2006) used a multiple-token design with acoustic varying tokens for each of the stimuli to analyze the sound pair [t–d], which has allophonic status in Korean ([d] between voiced sounds and [t] elsewhere), whereas the sound pair has a phonemic status in Russian. The results revealed an MMNm response for the Russian listeners but no response for the Korean listeners. The authors concluded that the phonemic representations are immediately computed from speech. However, as noted in Calabrese (2012), if the perceptual representations computed from speech and encoded in memory representations are only phonemic, language-specific conditioned allophones could not be acquired in first or second language acquisition because of language-specific phonological rules and not universal co-articulation adjustments (Jakobson, 1968; Kuhl et al., 2008). Therefore, access to contrastive and non-contrastive sound variants must be possible to acquire full knowledge of the distributional sound pattern of a language. However, normal linguistic interactions in a speech community require the identification of the non-contrastive phonological patterns that characterize different dialects/accents/social registers of this community. If perceptual representations were only phonemic, the acquisition of this knowledge would be impossible, which is contrary to reality. Hence, listeners must be able to access non-contrastive speech sounds that are present in the speech signal. Strange (2011) suggests that listeners normally use a phonological mode of perception to detect phonologically contrastive information to retrieve word meaning. If non-contrastive information must be accessed, such as allophonic variation, a phonetic mode of perception is implied, through which the speaker accesses knowledge of the phonetic and phonotactic patterns necessary to produce the appropriate phonetic sequences (see also Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, 1974; Werker & Logan, 1985).

The general aim of our study was to further investigate perception at the attentive and pre-attentive levels of the conditioned allophonic variation [ɛ–e] present in the Southern-Italian variety of Tricase to compare our results with the above-mentioned evidence. Similar to Calabrese (2012), we also hypothesize that non-contrastive speech sounds produced by a vowel-to-vowel assimilation process may be successfully parsed and encoded in memory representations. We used a same/different AX discrimination task to assess the attentive discrimination of such allophonic variation, whereas we recorded ERPs in an MMN design to investigate phonological and allophonic perceptions at the pre-attentive level. By using a multiple-token design (see Section 4), we compared responses to acoustic distinctions generated by the assimilation process with responses to acoustic distinctions associated with phonemic contrasts. By selecting the relevant distinctions, we carefully considered the acoustic distances between the possible stimuli pairs. We used the variant [e] of the mid vowel for the allophonic and phonemic conditions to reduce the difference in acoustic distance between the two conditions. More specifically, we decided to pair the phoneme /i/ with the variant [e] to generate a mean acoustic distance (88 mel) more similar to the allophonic [ɛ–e] (130 mel) than alternative pair [i–ɛ] (212 mel) (see Section 4.2 for more details). This set-up permitted us to significantly match the acoustical deviance between stimuli because this parameter is well known to influence MMN amplitude and latency (Näätänen, Paavilainen, Rinne, & Alho 2007; Näätänen et al., 1997). Lastly, to control for the effects because of physical differences between the stimuli, we calculated the identity mismatch negativity (iMMN) to obtain a pure contribution of the memory network by reducing bottom-up perceptual process contributions (Pulvermüller & Shtyrov, 2006).

2. Results

2.1. Behavioral

The percentage analysis of the participants' 'same' and 'different' responses indicated that the vowels that composed the allophonic pair [ɛ–e] were judged as "different" at a high rate (94%), whereas both the vowel pairs composed of the identical vowel type showed a high percentage of 'same' responses ([ɛ–ɛ] = 90%; [e–e] = 80%). A paired-samples *t*-test confirmed that the percentages of the 'same-different' responses were significantly different for the three tested vowel pairs ($p = .000$). Because a percentage analysis alone was not a meaningful measure of discrimination, we also performed a d-prime analysis to investigate the listener's tendency to respond 'same' or 'different' (cf. Macmillan & Creelman, 2005). The mean d' score was 2.55, which indicated an accurate discrimination between the allophones (see Fig. 1).

2.2. ERPs

Significant MMNs were elicited in the allophonic (probability: $F(1,66) = 14.592$, $p < 0.001$) and phonemic conditions (probability: $F(1,66) = 6.047$, $p < 0.05$).

The amplitude ANOVA revealed no significant main effect (vowel pair status: $F(1,66) = 1.052$, $p = 0.31$; electrode: $F(2,66) = 0.191$, $p = 0.83$) or interaction (vowel pair status × electrode: $F(2,66) = 0.204$, $p = 0.82$).

In contrast, the ANOVA on the latency of the MMN revealed a significant main effect for vowel pair status ($F(1,66) = 6.017$, $p < 0.05$), which indicated that the latency of the phonemic condition was significantly earlier. No main effect was observed for the factor electrode ($F(2,66) = 0.283$, $p = 0.76$) or interaction vowel pair status × electrode ($F(2,66) = 0.193$, $p = 0.83$) (see Table 1 and Fig. 2).

The laterality ANOVAs revealed no significant amplitude difference between the electrodes positioned at the right and left

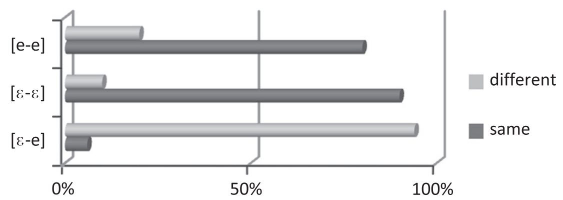

Fig. 1. Percentages of same/different responses of the AX discrimination task.

Table 1
The grand-average MMN mean (Fz, Cz, FCz) peak amplitudes and latencies.

	Amplitude	Latency
Allophonic condition	−1.898 (±0.679)	182 (±37)
Phonemic condition	−2.004 (±0.895)	154 (±41)

hemisphere for either the allophonic (F (2,69) = 0.236, p = 0.79) or phonemic condition (F (2,69) = 0.518, p = 0.60) (see Fig. 3).

3. Discussion

Both the behavioral and electrophysiological responses indicated that the allophonic pair [ɛ–e], produced by a vowel-to-vowel assimilation process, was detected by the listeners at the attentive and pre-attentive levels.

Our behavioral results contrast with Pallier et al. (1997) who showed that the allophonic variation [ɛ–e] was not discriminated. However, Pallier's study used synthesized stimuli in the discrimination task that only differed in the frequency of the first formant, whereas the stimuli in the present study were natural speech tokens produced by different speakers. Thus, these contrasting results may be because of the different nature of the stimuli.

With respect to the pre-attentive level, the MMN amplitude analysis revealed no significant difference between the allophonic and phonemic conditions, suggesting that the contrastive and non-contrastive vowel pairs, and relevant features [high] and [ATR] are equally computed in early speech processing and encoded in memory representations. This result contrasts with Kazanina et al. (2006), who showed that the consonant allophonic condition (pair [t–d] in Korean) did not elicit an MMN response. These contrasting results could be imputed to the different perceptual status of consonants and vowels. Consonants tend to be categorically discriminated, whereas vowels have been observed to be perceived continuously, i.e., within-category (Liberman, Harris, Hoffman, & Griffith, 1957; Pisoni, 1973; Repp, Healy, & Crowder, 1979). Accordingly, Hacquard et al. (2007) observed that Spanish and French listeners MMNm responses to the vowel pair [ɛ–e] did not significantly differ, although the vowel pair is phonemic in French and allophonic in Spanish. Nevertheless, several other factors can be implicated such as methodology (behavioral or neurophysiological) and the nature of the stimuli and their typology (natural or synthetic, syllables, words or non-words, voice-onset time continuum, phonemic pairs vs. within-category free variants or phonemic pairs vs. non-prototypical segments, etc.). Therefore, further research is necessary to assess the different perceptual statuses of consonants and vowels. Notably, we used multiple acoustic varying natural stimuli tokens for each vowel type and performed an identity MMN analysis so the MMN response did not purely reflect an acoustical analysis, which was different from Kazanina et al. (2006) and Hacquard et al. (2007).

In contrast to previous studies, we crucially observed a shorter latency for the phonemic contrast, which indicated that phonological knowledge was accessed and listeners' MMN responses to the phonemic/allophonic distinctions was not elicited by an acoustic distance analysis. Generally, the MMN peak latencies are attributed to the acoustic distances between the stimuli. In particular, the MMN latency steadily decreased with increasing acoustic deviation (Näätänen et al., 1997). However, as above noted, the Euclidean distance of our phonemic contrast [i–ɛ] was 88 mel, whereas the Euclidean distance of the allophonic contrast was 130 mel. Thus, if the difference in the peak latencies elicited by our stimuli were purely acoustic, we would expect that the allophonic contrast elicited a shorter latency then the phonemic one.

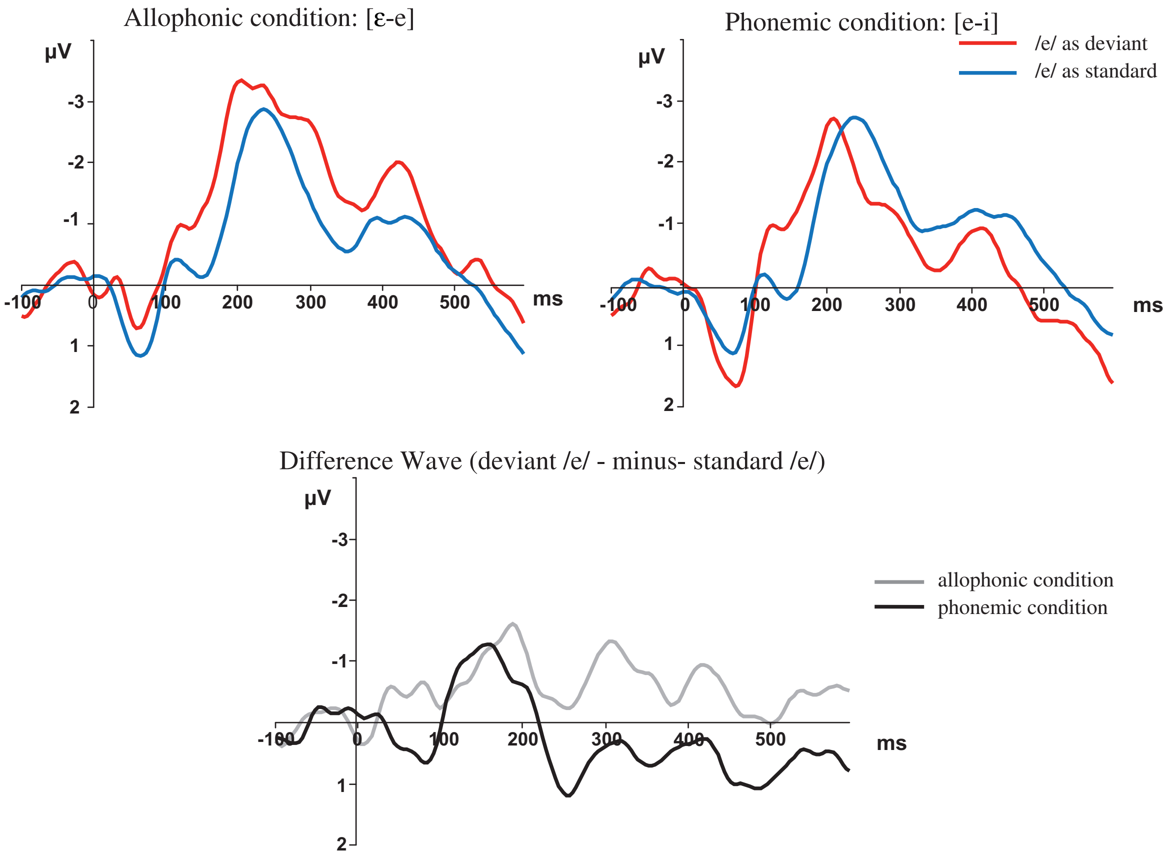

Fig. 2. Top panel: the group-averaged ERP responses to standards (blue lines) and deviants (red lines) at Fz for the allophonic (left) and phonemic conditions (right). Bottom panel: the deviant-minus-standard difference wave for the allophonic (gray solid line) and phonemic conditions (black solid line) at Fz.

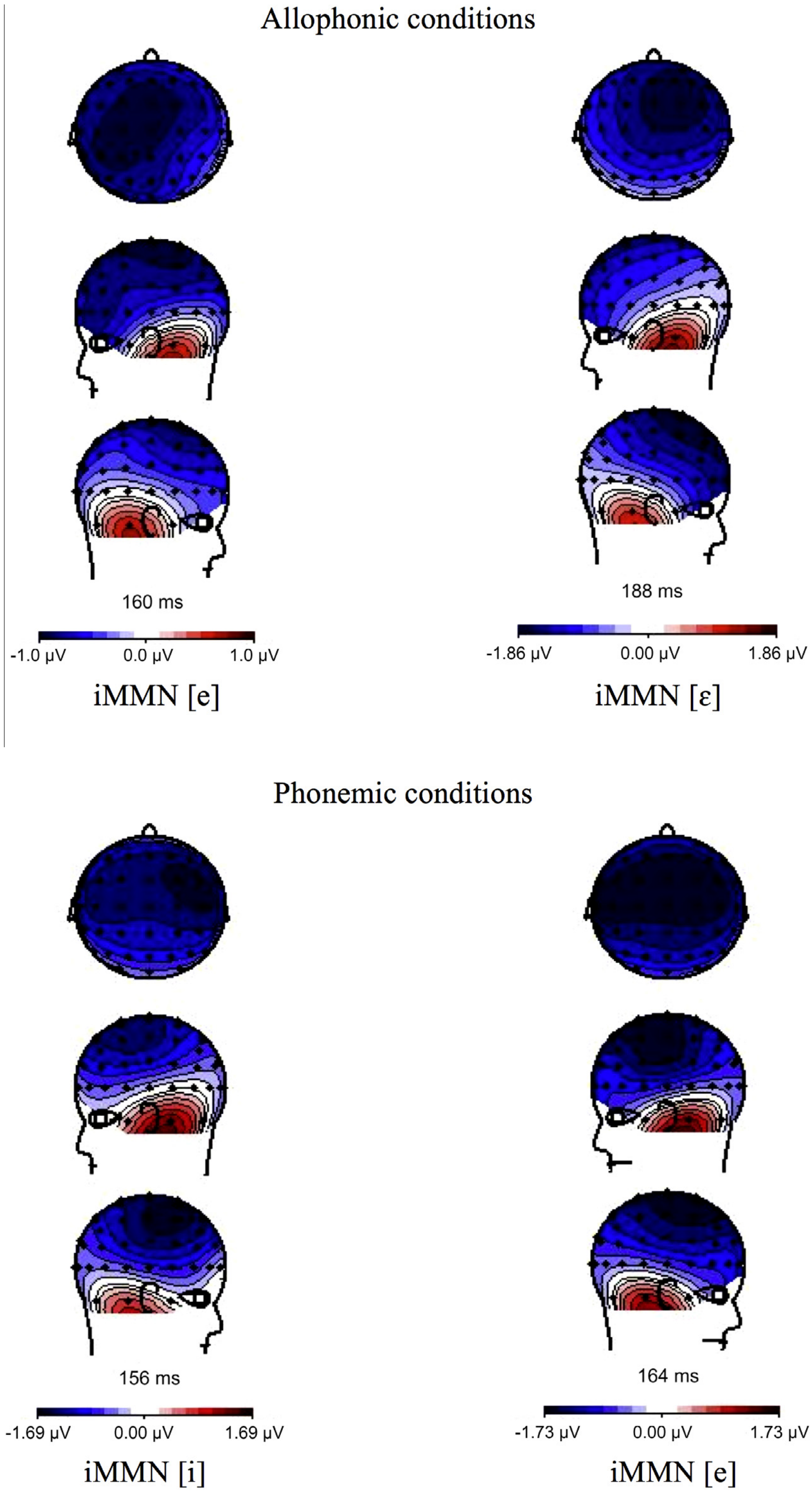

Fig. 3. Scalp distribution maps showing the iMMN peak latency activation viewed from above (top), the left side (middle) and right side (bottom). Top panel: allophonic conditions – iMMN [e] (left) and iMMN [ɛ] (right). Bottom panel: phonemic conditions – iMMN [i] (left) and iMMN [e] (right). The electrode positions are indicated by filled circles. The blue color marks negativity, and red marks positivity.

Thus, our results appear to suggest that two perceptual modes are available for speech perception: a faster phonological (categorical) mode and a slower phonetic (sensory) mode. We have used the term "perceptual mode" to refer to the set of perceptual computations that transform the continuously varying acoustic waveforms into discrete representations. We do not want to imply that the two modes are qualitatively different (that is, one is 'prior' in a chain of processing steps), nor do we want to suggest that they are derived from different neural processes. Rather, our idea is that there is a single neural computation—the mapping of auditory inputs into higher perceptual representations—that is 'sensitive' to the contrastive/non-contrastive status of the sounds as determined by the listeners' knowledge of the phonological system of their own language. This knowledge is apparently reached early by infants (see the studies discussed in Seidl & Cristia, 2012).

We suppose that these two perception modes occur simultaneously when processing speech-sound contrasts. In both modes, an acoustical parameter analysis of the incoming speech sounds is performed, and short term-memory traces are formed to process the signal. If the perceived sounds are phonemes, short-term memory trace formation is facilitated (Houtilainen, Kujala, & Alku, 2001; Näätänen et al., 1997, 2007), most likely because the fundamental goal of normal perception is to detect phonologically contrastive information and retrieve word meaning. Thus, for the phonological mode of perception, the restriction of the search to only contrastive sound properties may result in faster, less effortful cognitive processing. However, in the phonetic mode of perception, in which both contrastive and non-contrastive sound properties must be accessed and parsed, the processing requires additional computational operations and related supplementary neural activations. Most likely, the pattern observed in attentive discrimination is similar to that reflected by MMN responses at the pre-attentive level, which is likely because the listeners rely on two different perceptual modes at both the attentive and pre-attentive level.

From a general perspective, our results show that listeners recognize and detect sound alternations conditioned by the linguistic system environment. For Tricase speakers, learners converge on five phoneme categories and a process that predictably shifts the target realization of the phoneme /ɛ/ when followed by a high front vowel. This phonological knowledge provides the language users with the necessary information to produce the appropriate vowel token in a given context. However, it is notable that our experiment involves the perception of isolated vowels. Previous behavioral research has noted that within-category variations presented out of context are easier to discriminate than within context. For instance, Peperkamp et al. (2003) tested the perception of the French allophonic variants [ʁ–χ] in isolation and within context and observed that the allophonic consonants are well discriminated when presented in isolation but difficult to discriminate when embedded in context. The authors suggested that the absence of context increases the access to acoustic correlates of the contrast. This interpretation is consistent with the hypothesis of a phonetic mode for the computation of allophonic variants; however, further research is necessary to determine whether the facilitative effect of the absence of context can be observed for vowels at the attentive and pre-attentive level.

Overall, our results generate the assumption that predictable vowel allophonic alternations (out of context) pattern with phonemic contrasts for auditory perception. This assumption implies a model of phonology in which the acquisition of phonemic categories occurs with the learning of phonetic distribution patterns and their relationships.

4. Materials and methods

4.1. Participants

Twelve right-handed (Oldfield, 1971) students from the University of Salento (7 females; mean age 21.2, range: 19.1–26.0, s.d. ± 2) participated in both sessions. No participants reported a history of neurological illness. All of the participants were native speakers of the Tricase dialect, provided written informed consent and participated in both the experiments. The discrimination task and EEG recordings were run within one session with the discrimination task preceding the EEG recordings to prevent the stimuli from being attentively processed before the ERP measurements. The experimental procedure received the approval of the local ethics committee.

4.2. Stimuli

We used the vowels [ɛ, e, i] present in the stressed vowel system of the Tricase dialect as experimental stimuli. The introduction noted that the mid-vowel /ɛ/ and high-vowel /i/ have phonemic statuses, whereas the mid-high vowel [e] has an allophonic status. We used three natural speech tokens for each stimulus type to introduce acoustic variability and ensure that the acoustically different tokens were grouped together in a more abstract representation of the speech sound category.

A male speaker of Tricase produced a total of 30 pseudowords (10 for each vowel type). The vowels were inserted in the context b[V]b[V] and embedded in the carrier sentence *Ieu ticu__moi (I say__now)*. The speech signal was recorded in a soundproof room with CSL 4500 (hardware/software speech analysis system) and a Shure SM58-LCE microphone with a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits. The acoustic analysis was performed using Praat 5.2 (Boersma & Weenink, 2011). The fundamental frequency (F0), first formant (F1) and second formant (F2) formant values were measured in the vowel steady tract (0.25 s) centered at the vowel midpoint. For every vowel category, we selected three acoustic varying tokens with comparable pitch ([ɛ] = 174 (±3); [ɛ] = 174 (±7); [i] = 182 (±7). The F1/F2 average formant values in Hz of the three exemplars were the following: [ɛ]: F1 = 519 (±11), F2 = 1906 (±23); [e]: F1 = 389 (±7), F2 = 1967 (±20); [i]: F1 = 327 (±6); F2 = 2108 (±51). The mean Euclidean distances in mel in the F1–F2 plane (at the vowel mid-point) between all combinations of the three vowel types were [ɛ–e] 130, [e–i] 88 and [ɛ–i] 212 mel. Lastly, portions containing only the steady-state vowel signal were eliminated from the selected words. All nine stimulus audio files were ramped with 10 ms Gaussian on- and offsets and normalized for duration (200 ms) and peak amplitude (70 dB/SPL).

4.3. Behavioral test

In an AX (same-different) discrimination task, we assessed the attentive discrimination of the allophonic variation [ɛ–e]. Each of the three variants of the two vowel categories composing the allophonic variation were combined with one another and the three tokens of the other vowel category that composed the pair. Thus, three pair types were tested in all: [ɛ–ɛ], [ɛ–e] and [e–e]. The inter-stimulus interval was 800 ms, and the trial's initial silence was 500 ms. Each of the 54 stimulus pairs occurred twice. The complete set of 108 stimuli pairs was presented in random order. The listeners indicated whether the sounds of a pair were identical or different. The experimental arrangement provided by Praat 5.2 was applied in these tests. The subjects were tested in a quiet room with a laptop and headphones. The discrimination accuracy of allophonic variation was assessed using a percentage and d-prime analysis (Macmillan & Creelman, 2005).

4.4. ERP recordings and data analysis

In an oddball paradigm, the MMN responses to the phonemic contrast [e–i] and allophonic pair [ɛ–e] were recorded. Together, stimulus sequences of 1000 trials were randomly presented in each block. In each sequence, one vowel type served as the standard (85% of the trials) and the remaining vowel of the sound pair was the deviant. For both vowel pairs, the roles of standard and deviant were reversed in separate blocks. Therefore, a total of four blocks were recorded. Stimulus sequences were presented with a variable inter-stimulus interval of 500–550 ms. During the recording, the subject sat in an acoustically shielded room watching a silent movie. The subject was instructed to disregard the sounds presented via loudspeakers. The stimuli were presented using ePrime 2.0, and the order of the blocks was counterbalanced between participants.

EEGs were recorded (0.1–100 Hz, −2 dB points, sampling rate 250 Hz) with a 64-channel ActiCap system (Brain Products). Vertical eye movements were monitored using electrodes attached above and below the right eye and horizontal movements with electrodes attached to the outer canthi of each eye. The online reference electrode was FCz. Impedance was maintained under 5 Ω.

Off-line signal processing was performed with the Brain Vision Analyzer (Brain Products) software package. The EEG was filtered with a bandpass of 1–25 Hz (12 dB/oct), and the raw data were re-referenced against the average of the left and right mastoids. Epochs began 100 ms before until 600 ms after stimulus onset. Standard and deviant epochs were averaged and included a pre-stimulus baseline of 100 ms. The ERP responses to the initial three standard stimuli of each block and standard stimuli that immediately followed the deviants were not included in the analyses. The averaged data were baseline corrected over a pre-stimulus interval of 100 ms. Epochs with an amplitude change exceeding 75 μV at any of the electrodes were rejected. All remaining standard and deviant epochs were included in the identity MMN analysis. For the amplitude and peak latencies analysis of the MMN component, we selected a time window based on a visual inspection of the grand average data across all of the subjects.

For assessing the MMN component, we adopted the identity MMN (iMMN) approach. The MMN is reflected in a difference waveform calculated by subtracting the ERP response to standard stimuli from the deviant present in the identical block. In contrast, the iMMN is calculated using the recordings of two corresponding blocks. For instance, the standard [e] (of the block [e] standard and [i] deviant) was subtracted from the [e] deviant of the reverse block ([i] standard and [e] deviant). The iMMN approach eliminates variation in ERP morphology that may result from purely acoustic differences and therefore permits the observation of memory representation contributions (e.g., Pulvermüller & Shtyrov, 2006). The MMN latency corresponded to the time at which the highest negative amplitude peak in the MMN time window occurred (120–200 ms). This time window was selected based on the grand average across all subjects and was motivated by the expectation to observe the MMN response 100–200 ms after the onset of the deviant sound. The MMN amplitude was obtained by measuring the mean amplitude (μV) contained within a 50 ms time window centered at the MMN latency peak. The analysis was based on the electrodes Fz, Cz and FCz.

4.5. Statistical analyses

Separate one-way ANOVAs were performed with mean amplitude (μV) as the dependent measure and probability (standard vs. deviant) as the independent measure to assess that reliable MMN components were elicited. Separate two-way ANOVAs were performed with latency (ms) and mean amplitude (μV) as the dependent measures to analyze the MMN component. The independent variables were vowel pair status (allophonic vs. phonemic) and electrode (Fz, Cz and FCz). Additionally, the interaction vowel pair status × electrode was included in the analyses. Lastly, two separate one-way ANOVAs with an electrode grid of six electrodes (C3, Cz, C4, F3, Fz and F4) were performed to analyze whether hemispheric asymmetries could be observed. The dependent measure was mean amplitude and independent variable laterality (3-line (C3, F3), *z*-line (Fz, Cz), and 4-line (C4, F4)).

References

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Computer program), Version 5.2. <http://www.praat.org/> Retrieved 15 February.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, E. Dresher, & K. Rice (Eds.), *Contrast in phonology* (pp. 143–172). Berlin: Mouton de Gruyter.

Calabrese, A. (2012). Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception. *Journal of Neurolinguistics, 25*, 355–381.

Grimaldi, M., Calabrese, A., Sigona, F., Garrapa, L., & Sisinni, B. (2010). Articulatory grounding of southern Salentino harmony processes. In *Proceedings of the 11th annual conference of the International Speech Communication Association (Interspeech 2010)* (pp. 1561–1564). Japan: Makuhari.

Grimaldi, M. (2006). Acoustic correlates of phonological microvariations. The case of unsuspected highly diversified metaphonetic processes in a small area of Southern Salento (Apulia). In D. Torck & W. L. Wetzels (Eds.), *Romance languages and linguistic theory* (pp. 89–109). Amsterdam/Philadelphia: John Benjamins.

Hacquard, V., Walter, M. A., & Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: An MEG study. *Brain and Language, 100*, 295–300.

Houtilainen, 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.

Jakobson, R. (1968). *Child language, aphasia and phonological universals*. Paris: Mouton.

Kazanina, N., Phillips, C., & Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 1138–1186.

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 negative language magnet theory expanded. *Philosophical Transactions of the Royal Society B, 363*, 979–1000.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431–461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.

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.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., & Iivonen (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

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

Navarro Tomás, T. (1918). *Manual de pronunciación española*. Madrid: CSIC.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*, 9–17.

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th annual boston university conference on language development* (pp. 650–661). Somerville, MA: Cascadilla Press.

Phillips, C., Marantz, A., McGinnis, M., Pesetsky, D., Wexler, K., & Yellin, E. (1995). Brain mechanisms of speech perception: A preliminary report. In T. Schütze, J. B. Ganger, & K. Broihier (Eds.), *Papers on language processing and acquisition* (pp. 153–191). Cambridge, MA: MIT Press.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253–260.

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.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology, 5*, 129–145.

Seidl, A., & Cristia, A. (2012). Infants' learning of phonological status. *Frontiers in Psychology, 3*, 448.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*, 456–466.

Studdert-Kennedy, M. (1974). The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics, vol. 12* (pp. 2349–2385). The Hague: Mouton.

Trubetzkoy, N. S. (1969). *Principles of phonology. (English Translation of Grundzüge der Phonologie Travaux du Circle Linguistique de Prague), vol. 7*. Berkeley: University of California Press. 1939.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*, 35–44.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*, 638–642.