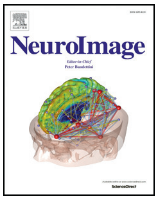
NeuroImage xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations

Mathias Scharinger [a,d,e,*], Philip J. Monahan [b,c], William J. Idsardi [d]

[a] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany
[b] Centre for French and Linguistics, University of Toronto Scarborough, Canada
[c] Department of Linguistics, University of Toronto, Canada
[d] Department of Linguistics, University of Maryland, College Park, MD, USA
[e] Biological incl. Cognitive Psychology, Institute for Psychology, University of Leipzig, Germany

ARTICLE INFO

Article history:
Received 4 May 2015
Accepted 2 January 2016
Available online xxxx

Keywords:
Speech sound perception
MEG
Mismatch negativity
Cortical oscillations
Beta-band power
Underspecification
Predictive coding

ABSTRACT

While previous research has established that language-specific knowledge influences early auditory processing, it is still controversial as to what aspects of speech sound representations determine early speech perception. Here, we propose that early processing primarily depends on information propagated top–down from abstractly represented speech sound categories. In particular, we assume that mid-vowels (as in 'bet') exert less top–down effects than the high-vowels (as in 'bit') because of their less specific (default) tongue height position as compared to either high- or low-vowels (as in 'bat'). We tested this assumption in a magnetoencephalography (MEG) study where we contrasted mid- and high-vowels, as well as the low- and high-vowels in a passive oddball paradigm. Overall, significant differences between deviants and standards indexed reliable mismatch negativity (MMN) responses between 200 and 300 ms post-stimulus onset. MMN amplitudes differed in the mid/high-vowel contrasts and were significantly reduced when a mid-vowel standard was followed by a high-vowel deviant, extending previous findings. Furthermore, mid-vowel standards showed reduced oscillatory power in the pre-stimulus beta-frequency band (18–26 Hz), compared to high-vowel standards. We take this as converging evidence for linguistic category structure to exert top–down influences on auditory processing. The findings are interpreted within the linguistic model of underspecification and the neuropsychological predictive coding framework.

Introduction

Neuroimaging methods have been increasingly used to probe the mechanisms that underlie speech sound processing. Recently, a number of studies have demonstrated that linguistic category structure has specific modulatory effects on early stages of auditory perception (Bien and Zwitserlood, 2013; Cornell et al., 2011, 2013; Eulitz and Lahiri, 2004; Friedrich et al., 2008). Linguistic category structure allows speech sound classification according to their acoustic and articulatory properties, often described in terms of a deviation from the neutral, resting position of the mouth. For example, high-vowels (e.g., [ɪ] as in 'bit') with a relatively high tongue position during production can be distinguished from low-vowels (e.g., [æ] as in 'bat') with a relatively low tongue position during production. Some theories assume that vowels that fall between high and low-vowels (e.g., [ɛ] as in 'bet') are neither high nor low, and, being produced with a neutral tongue position, have no descriptive feature for tongue height (Lahiri and Reetz, 2002, 2010; Scharinger and Idsardi, 2014). The production of mid-vowels in English does not necessarily lead to a larger spread of individual vowel tokens, but rather to greater overlap with neighboring vowel category tokens (Hillenbrand et al., 1995). These vowels are assumed to be underspecified and may refer to a rather unspecific motor plan regarding their tongue height.

Recently, it has been proposed that less specific, underspecified vowels have less intrinsic "predictive" value compared to more specific, specified vowels (Eulitz and Lahiri, 2004; Scharinger et al., 2012a, 2012b). Scharinger et al. (2012b) demonstrated that the unspecific category structure of the American English vowel [ɛ] influenced processing, as indexed by the mismatch negativity, an automatic change and prediction error response of the brain (Näätänen and Alho, 1997; Schröger, 2005; Winkler, 2007). In a passive oddball design, the authors contrasted the high- and low-vowels [ɪ] and [æ] in standard position with the low- and high-vowels [æ] and [ɪ] in deviant position. This condition showed a relatively large acoustic distance of the first resonance frequencies (first formant, F1) between the vowels and was compared to a condition in which the acoustic F1 distance was relatively small, i.e., in which the standard was either specific ([æ]) or unspecific ([ɛ]), contrasting with the deviants [ɛ] and [æ]. The results showed similar

---

* Corresponding author at: Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, Frankfurt, 60322, Germany.
E-mail address: mathias.scharinger@aesthetics.mpg.de (M. Scharinger).

http://dx.doi.org/10.1016/j.neuroimage.2016.01.003
1053-8119/© 2016 Published by Elsevier Inc.

Please cite this article as: Scharinger, M., et al., Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscil..., NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.01.003

symmetric mismatch responses in the large F1 distance condition, while the small F1 distance condition showed asymmetric MMN differences: the condition with unspecific [ɛ]-standards yielded significantly reduced MMN amplitudes compared to the condition with specific [æ]-standards. This result is consistent with other electrophysiological studies (Cornell et al., 2011, 2013; Eulitz and Lahiri, 2004). Within the framework of predictive coding (Friston, 2005; Garrido et al., 2009), this pattern was interpreted as evidence for [ɛ] being inherently less predictive, such that the prediction error upon encountering the deviant [æ] was reduced.

While this study suggests that linguistic category structure may indeed influence early auditory processing, generalization to further vowel contrasts was impossible (e.g., between [ɛ] and [ɪ]). Moreover, there was no measure with a closer relation to the assumed top–down propagation of category information (strong for [ɪ] and [æ], weak for [ɛ]). In this regard, recent research suggests that cortical oscillations index directional message passing between different levels of the cortical hierarchy (Arnal and Giraud, 2012; Arnal et al., 2011; Engel and Fries, 2010; Fontolan et al., 2014). In particular, cortical oscillations within the beta-band (15–30 Hz) are assumed to reflect endogenous top–down processes that are interpreted within the predictive coding framework (Wang, 2010). In this framework, beta-power scales with prediction strength propagated downward from representational units to lower processing levels. This mechanism should also operate on speech sound category representations, such that differences in linguistic structure lead to differences in cortical beta-power, which should arise prior to stimulus presentation in an MMN paradigm.

Thus, the current magnetoencephalography (MEG) study has two primary goals: (1) to examine cortical oscillations as a means to further elucidate the mechanisms by which linguistic category structure exerts influence on lower-level auditory processing, and (2) to extend the MMN findings from Scharinger et al. (2012b) to the contrast between the vowels [ɛ] and [ɪ]. We expect (1) beta-power to differ between [ɛ] and [ɪ] presented as standards, where predictions build up (Winkler et al., 1996a) and should most strongly be influenced by linguistic category structure, and (2) the MMN to be reduced or absent if deviant [ɪ] follows the standard [ɛ].

Methods

Participants

Thirteen students, all native speakers of American English, were recruited from the University of Maryland (9 females, 4 males, mean age 21 ± 1.3 years). They had no reported history of hearing or neurological problems and participated for class credit or monetary compensation ($10 per hour). All participants provided informed written consent and tested strongly right-handed (>80%) on the Edinburgh Handedness Inventory (Oldfield, 1971). The study was approved by the Institutional Review Board of the University of Maryland and in accordance with the Declaration of Helsinki.

Materials

Stimulus material was similar to that used in Scharinger et al. (2012b) and involved 10 renditions of each of the vowels [æ], [ɛ] and [ɪ], produced by a female native speaker of American English, who made a robust three-way height distinction (see Fig. 1). All vowels were recorded embedded in the carrier sentence “I will say h__d again”. This was repeated 20 times for each vowel. The phonetically trained speaker ensured that vowels had the quality of short vowels. The speech material was digitized at 44 kHz with 16 bit amplitude resolution using the phonetic sound application PRAAT (Boersma and Weenink, 2011). We then spliced 100 ms out of the steady-state portion of the respective vowels from the carrier sentences and selected a final set of 10 vowels on the basis of intensity and pitch. The first 10 ms of each vowel was multiplied with the first half period of a $(1 - \cos(x)) / 2$ function and the last 10 ms with the first half period of a $(1 + \cos(x)) / 2$ function to reduce acoustic artifacts. Stimulus intensity was normalized to 70 dB within PRAAT, which corresponds to sound pressure level (SPL, Boersma and Weenink, 2011). We further set up the sound delivery system in the MEG scanner such that participants would hear the stimuli at a level of 60 dB SPL, which was confirmed using a sound pressure level meter in the MEG cabin. Finally, we obtained three independent opinions regarding the perceived loudness of the three vowel types in the scanner. Since no differences in perceived loudness were reported, no further loudness modifications were deemed necessary. Detailed acoustic measures of the vowel stimuli are provided in Table 1 and illustrated in Fig. 1.

Spectral analyses of the vowel stimuli involved a linear predictive coding (LPC) formant analysis and estimated the first three resonant frequencies (formants, F1–F3). As the three vowels mainly differ in tongue height, which is inversely correlated with F1 frequency (Stevens, 1998), we had defined that the opposition of [æ] and [ɪ] constitutes the large F1 distance condition, while the opposition of [ɛ] and [ɪ] represents the small F1 distance condition. This definition was corroborated by the Euclidean F1 distances which were larger between [æ] and [ɪ] (491.8 Hz) than between [ɛ] and [ɪ] (269.5 Hz; $t\ (18) = 17.88$, $p < 0.001$).

Design

Vowel stimuli were presented in a passive standard/deviant many-to-one oddball paradigm (Winkler et al., 1999): the vowels [æ]/[ɪ] (large F1 distance) and [ɛ]/[ɪ] (small F1 distance) were distributed over four blocks (the order permutated across participants) in which they occurred in either standard ($p = 0.875$, $N = 700$) or deviant position ($p = 0.125$, $N = 100$, for details, see Fig. 1C). We referred to the direction of standard/deviant presentation as “F1 increasing” if the deviant had a higher F1 (lower tongue position) than the standard (i.e. [ɪ]–[æ]; [ɪ]–[ɛ]) and as “F1 decreasing” if the deviant had a lower F1 (higher tongue position) than the standard (i.e. [æ]–[ɪ]; [ɛ]–[ɪ]). The distribution of vowel stimuli over the factor levels is illustrated in Table 2.

The 10 different vowel renditions for standards and deviants had the same probability of occurrence. Note that using different renditions for standards is beneficial for activating memory traces not bound to a particular phonetic realization but rather referring to an abstract representation (see Phillips et al., 2000). The number of consecutive standards pseudo-randomly varied between 3 and 10, and inter-stimulus intervals (ISIs) were jittered between 500 and 1000 ms (in steps of 1 ms, random selection from a uniform distribution) to prevent participants from entraining to a specific presentation rhythm. A total of 800 vowel stimuli were presented in each block, leading to block durations of approximately 15 min and a total experiment duration of about 90 min. This included participant preparation and debriefing. Stimulus presentation was controlled by the *Presentation* software (Neurobehavioral Systems, Albany, CA); delivery of auditory stimuli into the shielded MEG chamber was achieved by air conduction transduction and non-magnetic earphones (Etymotic Research Inc., IL, USA), resulting in a binaural, comfortable listening level at 60 dB (SPL). Earphones (Etymotic ER3A insert) were calibrated to have a flat frequency response between 50 Hz and 3100 Hz within the shielded room. This guaranteed an optimal acoustic delivery of the first three vowel formant frequencies (Stevens, 1998).

MEG recording

MEG activity was recorded from 157 axial gradiometers (whole-head system, Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 500 Hz. Data were online filtered between DC and 200 Hz, together with a notch filter of 60 Hz to reduce ambient electrical

Please cite this article as: Scharinger, M., et al., Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscil..., NeuroImage (2016), http://dx.doi.org/10.1016/j.neuroimage.2016.01.003