Brain & Language 136 (2014) 8–18

Contents lists available at ScienceDirect

Brain & Language

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

Differences in sensory processing of German vowels and physically matched non-speech sounds as revealed by the mismatch negativity (MMN) of the human event-related brain potential (ERP)

Corinna A. Christmann [a,*], Stefan Berti [b], Claudia Steinbrink [a], Thomas Lachmann [a]

[a] Center for Cognitive Science, Cognitive and Developmental Psychology Unit, University of Kaiserslautern, Kaiserslautern, Germany
[b] Department of Clinical Psychology and Neuropsychology, Institute for Psychology, Johannes Gutenberg University Mainz, Mainz, Germany

ARTICLE INFO

Article history:
Accepted 17 July 2014

Keywords:
Event related brain potentials (ERPs)
Mismatch negativity (MMN)
Speech processing
Stimulus complexity
Auditory discrimination
Spectrally rotated speech sounds

ABSTRACT

We compared processing of speech and non-speech by means of the Mismatch Negativity (MMN). For this purpose, the MMN elicited by vowels was compared to those elicited by two non-speech stimulus types: spectrally rotated vowels, having the same stimulus complexity as the speech stimuli, and sounds based on the bands of formants of the vowels, representing non-speech stimuli of lower complexity as compared to the other stimulus types. This design allows controlling for effects of stimulus complexity when comparing neural correlates of processing speech to non-speech. Deviants within a modified multi-feature design differed either in duration or spectral property. Moreover, the difficulty to discriminate between the standard and the two deviants was controlled for each stimulus type by means of an additional active discrimination task. Vowels elicited a larger MMN compared to both non-speech stimulus types, supporting the concept of language-specific phoneme representations and the role of the participants' prior experience.

1. Introduction

1.1. Neurobiological specialization for speech sounds

The neurobiological system for auditory and speech processing has been shown to include hierarchically organized pathways, i.e. the dorsal pathway and the ventral pathway. Whereas the first one is responsible for spatial processing ("where path"), the latter one is associated with the identification of complex patterns or objects ("what path") (Rauschecker & Scott, 2009). Within this ventral pathway, the left anterior and middle superior temporal cortex is specifically activated for speech sounds when compared to non-speech sounds which are matched with respect to stimulus complexity (Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006). This finding is not limited to speech material incorporating semantic contents, as this area was also active during the processing of consonant-vowel (CV) syllables as compared to non-speech (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Obleser, Zimmermann, van Meter, & Rauschecker, 2007). Moreover, topographic maps which are ordered based on phonetic features of consonants (Obleser, Scott & Eulitz, 2006) and vowels (Obleser, Boecker, et al., 2006) could be identified.

1.2. Electrophysiological evidence for a specialization for native speech sounds

The existence of neuron populations which respond specifically to speech stimuli is in line with electrophysiological data based on the mismatch negativity. The mismatch negativity (MMN, Näätänen, Gaillard, & Mäntysalo, 1978) is an objective and reliable electrophysiological correlate of automatic sensory processing in the auditory domain (Näätänen, 2008). Within a classical oddball paradigm, the ERP of a frequently presented standard stimulus is subtracted from the ERP of an infrequently presented deviant stimulus. The resulting difference curve shows a negative peak between 150 and 250 ms, known as the MMN. In early studies applying the MMN, pure sinusoidal tones were used (e.g. Näätänen, 1979; Näätänen & Michie, 1979; Näätänen et al., 1978) but it has been shown that the MMN can also be elicited with stimuli of higher complexity (for a review see Näätänen, Paavilainen, Rinne, & Alho, 2007).

* Corresponding author. Address: Center for Cognitive Science, Cognitive and Developmental Psychology Unit, University of Kaiserslautern, Erwin-Schrödinger-Str. 57, D-67663 Kaiserslautern, Germany. Fax: +49 631/205 5034.
E-mail address: cchristm@rhrk.uni-kl.de (C.A. Christmann).

http://dx.doi.org/10.1016/j.bandl.2014.07.004
0093-934X/

The MMN has been shown to be sensitive to language-specific phoneme representations. This means that vowel contrasts representing distinct phonetic categories in one's native language elicit a larger MMN than non-native contrasts on deviant trials (Näätänen et al., 1997). This finding has been replicated in a number of studies using participant groups with various native-language backgrounds (e.g., Chládková, Escudero, & Lipski, 2013; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005; Peltola, Tamminen, Toivonen, Kujala, & Näätänen, 2012; Rinker, Alku, Brosch, & Kiefer, 2010). The role of prior experience with different stimulus types is, however, not limited to the speech domain; for example, the superiority of musicians as compared to non-musicians in processing musical sound material has been shown in several studies (e.g. Koelsch, Schröger, & Tervaniemi, 1999; Vuust et al., 2005).

1.3. Automatic auditory discrimination of speech and non-speech sounds

Taking into account the role of prior experience and language-specific phoneme representations, one would expect that the processing of speech stimuli would be accompanied by pronounced MMN when compared to non-speech stimuli, considering that enhanced sensitivity to differences is reflected in increased MMN amplitude (and partly in shorter peak latency of the MMN). Interestingly, this pattern of results has not been found in all MMN studies, in particular, there is an inconsistent pattern of speech vs. non-speech MMN results with respect to both spectral and durational manipulations. Concerning the spectral MMN, some studies have found a larger MMN for speech stimuli as compared to non-speech sounds (Sorokin, Alku, & Kujala, 2010; Čeponiené et al., 2002). In other studies, however, no difference between speech and non-speech processing (Davids et al., 2011; Jaramillo et al., 2001; Nikjeh, Lister, & Frisch, 2009; Wunderlich & Cone-Wesson, 2001) or even the contrary pattern of results has been reported (Lachmann, Berti, Kujala, & Schröger, 2005; Tervaniemi et al., 1999; Wunderlich & Cone-Wesson, 2001). Regarding duration decrements, the MMN for vowels has been reported to be larger compared to noise or single sinusoidal tones (Jaramillo, Alku, & Paavilainen, 1999; Jaramillo et al., 2001; Takegata, Alku, Ylinen, & Näätänen, 2008), whereas no MMN differences have been found for chords (Takegata et al., 2008) and complex analogues of speech sounds (Sorokin et al., 2010).

There are *three fact*ors typically affecting the MMN amplitude which may also explain the inconsistencies in the pattern of results in the summarized studies:

1. The relative importance of spectral and temporal cues might vary across languages. Therefore, temporal and spectral processing should both be taken into consideration.
2. The degree of deviation between standard and deviant: It has been demonstrated that the amplitude of the MMN is correlated with the degree of deviation between the standard and deviant stimuli (Amenedo & Escera, 2000; Berti, Roeber, & Schröger, 2004; Jaramillo, Paavilainen, & Näätänen, 2000; Näätänen, Syssoeva, & Takegata, 2004; Sams, Paavilainen, Alho, & Näätänen, 1985). For example, the MMN is larger in frequency deviants when the pitch of the deviant tone is 10% higher than the standard tone compared to a pitch difference of 5%. Importantly, in studies testing for differences between speech and non-speech sounds, this effect is not necessarily controlled for (see Lachmann et al., 2005).
3. The physical complexity of both stimulus classes: The idea that the size of the MMN depends on stimulus complexity is supported by the fact that single sinusoidal tones evoke a smaller MMN compared to harmonically rich tones (Takegata et al., 2008; Tervaniemi et al., 2000; Zion-Golumbic, Deouell, Whalen, & Bentin, 2007). Differences in the physical complexity of speech and non-speech signals are found, to the best of our knowledge, in all existing MMN studies, except for one (Davids et al., 2011).

To overcome these problems, temporal and spectral deviants were used in the present study. Moreover, all standards and deviants used in the passive oddball task were also presented within an active *same-different* task to control for the behavioral discrimination performance. Finally, we applied spectrally rotated speech (Blesser, 1972) and non-speech stimuli with lower complexity compared to the speech sounds (see Section 2.2.3). This approach controls for influences of stimulus complexity.

1.4. Using spectrally rotated speech

The creation of non-speech stimuli of the same complexity as speech stimuli is a real challenge, as the spectro-temporal pattern of an original speech sound has to be fitted. One effective solution has been presented by Blesser (1972), that is, the spectral rotation of speech signals. Spectrally rotated speech is created by inverting the spectrum around a center frequency. Starting with a study by Scott et al. (2000), this procedure is commonly used to compare speech to non-speech in behavioral (e.g. Vandermosten et al., 20,10, 2011) and imaging studies (e.g., Liebenthal et al., 2005; Obleser, Boecker, et al., 2006; Scott et al., 2006).

There are, however, still two shortcomings involving this procedure: The first one concerns the low-pass filtering of the original speech sound which is a technical precondition of the whole procedure. The most important frequencies of the speech signal are thought to lie between 500 and 4000 Hz (Wilmanns & Schmitt, 2002). Hence, 4000 Hz was used as the cut-off frequency for the low-pass filter in most studies using spectrally rotated speech stimuli (e.g., Davids et al., 2011; Evans et al., 2013; Scott et al., 2000, 2006). This approach ensures that the intelligibility of the speech sound is not reduced (Scott & Wise, 2004), however, its naturalness could be reduced severely (Moore & Tan, 2003). Therefore, we developed a new stimulus type for the present study: spectrally rotated speech stimuli with a complete spectrum. Note that this approach allows the use of unfiltered speech signals which sound natural.

The second shortcoming concerns the finding that some consonants, especially fricatives (e.g. /f/), nasals (e.g. /m/ and /n/) and plosives (e.g. /p/) are either not affected by the rotation, or their spectrally rotated counterparts evoke the impression of another consonant (see Blesser, 1972, for details). Thus, spectrally rotated speech sounds which include these consonants might evoke a speech-like impression. In contrast to the consonants, the formant structure of a vowel is completely inverted. Formants are the most intensive frequencies within the vowel (Carroll, 2004). For this reason, in the present study, we use single German vowels to ensure a non-speech like impression of the spectrally rotated sounds.

1.5. German vowels

There are seven pairs of monophthongs with differences in vowel length in German (Lühr, 1993). Each pair consists of a lax (or short) and a tense (or long) version of the respective vowel (Kohler, 1977; Moulton, 1962; Wiese, 2000). Their formants are systematically changed by moving the articulatory organs (Carroll, 2004). The first two formants are essential for the correct identification of a vowel (Peterson & Barney, 1952; Pols, van der Kamp, & Plomp, 1969). The region of frequencies with a power of at the most three dB beneath the power of the formant is defined as the bandwidth of the formant (Fant & Tatham, 1975).

In German, tense and lax vowels do not only differ with respect to their duration but also with respect to the position of their