Brain and Language 176 (2018) 26–35

Contents lists available at ScienceDirect

Brain and Language

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

Short communication

Using rotated speech to approximate the acoustic mismatch negativity response to speech

Ellen Marklund*, Francisco Lacerda, Iris-Corinna Schwarz

Stockholm University, Department of Linguistics, SE-106 91 Stockholm, Sweden

ABSTRACT

The mismatch negativity (MMN) response is influenced by the magnitude of the acoustic difference between standard and deviant, and the response is typically larger to linguistically relevant changes than to linguistically irrelevant changes. Linguistically relevant changes between standard and deviant typically co-occur with differences between the two acoustic signals. It is therefore not straightforward to determine the contribution of each of those two factors to the MMN response. This study investigated whether spectrally rotated speech can be used to determine the impact of the acoustic difference on the MMN response to a combined linguistic and acoustic change between standard and deviant. Changes between rotated vowels elicited an MMN of comparable amplitude to the one elicited by a within-category vowel change, whereas the between-category vowel change resulted in an MMN amplitude of greater magnitude. A change between rotated vowels resulted in an MMN ampltude more similar to that of a within-vowel change than a complex tone change did. This suggests that the MMN amplitude reflecting the acoustic difference between two speech sounds can be well approximated by the MMN amplitude elicited in response to their rotated counterparts, in turn making it possible to estimate the part of the response specific to the linguistic difference.

1. Introduction

Mismatch negativity (MMN) is an event-related component that reflects pre-attentive change detection (e.g., Näätänen, Paavilainen, Rinne, & Ahlo, 2007). Its amplitude reflects the magnitude of acoustic difference between the standard and deviant sounds (e.g., Tiitinen, May, Reinikainen, & Näätänen, 1994). The MMN amplitude also reflects the presence of linguistic relevance in the change between standard and deviant (e.g., Sharma & Dorman, 1999). Since a linguistically relevant change between standard and deviant in most cases co-occur with a difference between the two acoustic signals, it is not straightforward to determine the relative contribution of each of those two factors to the MMN response.

The terms and definitions for making the distinction between linguistically relevant and linguistically irrelevant information in the speech signal vary. In this paper, the terms *linguistic* and *acoustic* will be used. *Linguistic* refers to the abstract content conveyed by the speech signal, whereas *acoustic* refers to the properties of the speech signal itself. The term acoustic thus includes both acoustic information that may function as a cue to linguistic content (e.g., spectral differences that results in two sounds being perceived as two different vowels) and acoustic information that does not (e.g. spectral differences that only signal allophonic variation). Importantly, in the case of a linguistic difference between two sounds, there is in most cases also an acoustic difference between them. In some rare cases, a linguistic difference can be perceived without altering the acoustic signal (by modifying the acoustic context or the listener's expectations instead), but as a rule, a linguistic difference always co-occurs with (and is conveyed by) an acoustic difference.

1.1. *MMN responses to linguistic and acoustic differences between standard and deviant*

A growing body of evidence shows that the amplitude of the MMN is influenced by the type of difference that is present between standard and deviant. Specifically, a combination of a linguistic and an acoustic difference results in a greater amplitude than a difference that is acoustic only, provided that the magnitude of the acoustic differences are matched (see Section 1.2).

When contrasting the MMN responses to speech and non-speech stimuli, it is likely that the speech condition includes linguistic differences, whereas the non-speech condition by definition includes only acoustic differences (e.g., Christmann, Berti, Steinbrink, & Lachmann, 2014; Sorokin, Alku, & Kujala, 2010). In a study by Christmann et al.

* Corresponding author.
E-mail addresses: ellen.marklund@ling.su.se (E. Marklund), francisco.lacerda@ling.su.se (F. Lacerda), iris-corinna.schwarz@ling.su.se (I.-C. Schwarz).

https://doi.org/10.1016/j.bandl.2017.10.006
Received 2 April 2015; Received in revised form 31 August 2017; Accepted 21 October 2017
Available online 22 November 2017
0093-934X/

(2014), the speech condition consisted of changes between vowels, making the difference between standard and deviant both acoustic and linguistic in nature. In their non-speech conditions on the other hand, differences between standard and deviants were of course acoustic-only. The authors found larger MMNs in their speech condition than in their two non-speech conditions (Christmann et al., 2014). In the same vein, Sorokin et al. (2010) found that their speech condition overall had a larger MMN than their non-speech condition. Further, differentiating between changes of linguistic relevance and other changes *within* the speech condition, they found that for the most part, changes between standard and deviant that were comprised of both an acoustic and a linguistic change had enhanced MMN responses compared to those that were comprised of an acoustic-only change (Sorokin et al., 2010). On the whole, making the distinction between linguistic and acoustic changes is more informative than merely differentiating between speech stimuli and non-speech sound stimuli, since both types of stimuli invariably entail acoustic changes.

Zeroing in on speech stimuli, between-category changes between standard and deviant comprise a combination of a linguistic and acoustic change, whereas within-category changes by design are comprised of an acoustic-only difference. As expected in studies making this distinction between conditions, the amplitude of the MMN is greater in response to between-category changes than to within-category changes (e.g., Dehaene-Lambertz, 1997; Sharma & Dorman, 1999).

When comparing the MMN responses to native versus non-native contrasts, studies are often designed in such a way that the two sounds in the non-native contrast will be perceived as different exemplars of the same speech sound in the native language (e.g., Sharma & Dorman, 2000; Winkler et al., 1999). In essence, the distinction between native versus non-native contrasts is the same distinction as that between between-category changes and within-category changes, but set in the context of the language background of the listeners. This means that the difference between native speech sounds is both linguistic and acoustic, whereas the difference between non-native speech sounds is acoustic-only. The study by Winkler et al. (1999) illustrates this point beautifully: a pair of vowels that to Hungarian speakers belongs to different categories is considered a variation of a single vowel by speakers of Finnish, and vice versa. The difference between the same two vowels is thus either both linguistic and acoustic or acoustic-only, depending on the native language of the listener. In the study, the combination of linguistic and acoustic difference elicited larger MMN responses than the acoustic-only difference (Winkler et al., 1999). Similarly, Sharma and Dorman's (2000) difference between the sounds in a Hindi contrast was linguistic and acoustic to Hindi speaking participants. To the English speaking participants, the difference was acoustic-only, since the two stimuli were different exemplars of the same speech sound. Consequently, the Hindi speakers showed larger MMN responses than the English speakers (Sharma & Dorman, 2000).

In sum, three patterns have been observed regarding the MMN response to acoustic and linguistic changes:

(1) Speech stimuli result in larger MMNs than non-speech stimuli (Christmann et al., 2014; Sorokin et al., 2010).
(2) Native contrasts result in larger MMNs than non-native contrasts (Sharma & Dorman, 2000; Winkler et al., 1999).
(3) Between-category changes result in larger MMN responses than within-category changes (Dehaene-Lambertz, 1997; Sharma & Dorman, 1999).

These patterns can be collapsed into one single distinction: combined linguistic and acoustic differences between standard and deviant result in larger MMN responses than acoustic-only differences (of corresponding acoustic magnitude, see Section 1.2 below) between standard and deviant.

1.2. Effect of difference magnitude between standard and deviant on MMN amplitude

The magnitude of the acoustic difference between standard and deviant influences the amplitude of the MMN response, with a greater acoustic difference resulting in a higher amplitude. This pattern is found for:

(1) Non-speech stimuli, when the difference between standard and deviant is consequently acoustic-only (e.g., Nikjeh, Lister, & Frisch, 2009; Novitski, Tervaniemi, Huotilainen, & Näätänen, 2004; Tiitinen et al., 1994).
(2) Speech stimuli, when the difference between standard and deviant is acoustic-only (Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997; Marklund, Schwarz, & Lacerda, 2014; Pakarinen et al., 2013).
(3) Speech stimuli, when the difference between standard and deviant is both acoustic and linguistic (Deguchi et al., 2010; Näätänen et al., 1997; Pakarinen et al., 2013; Sittiprapaporn, Tervaniemi, Chindaduangratn, & Kotchabhakdi, 2005).

The pattern of greater MMN amplitude in response to greater difference between standard and deviant is found as long as the difference between standard and deviant is of the same kind (acoustic-and-linguistic or acoustic-only), and stimuli are of the same type (non-speech or speech). The pattern is *not* found when comparing conditions with different types of differences; the combination of a linguistic and an acoustic difference between standard and deviant results in an overall higher MMN amplitude than an acoustic-only difference (see Section 1.1).

Considering the overall difference between standard and deviant instead of only the acoustic, this is however not surprising, assuming the two acoustic differences are equal. In the first condition, a linguistic difference of some magnitude greater than zero is added, whereas in the second condition, there is a linguistic difference of zero (no linguistic difference). There is then a greater overall difference in the first condition than in the second. Following the pattern established above, the first condition has a greater difference magnitude than the second and will thus yield a larger MMN response.

1.3. Variations in MMN amplitude in response to linguistic differences

In general, not much is known about the ways in which a linguistic difference influences the amplitude of the MMN response, except for the fact that its mere presence enhances the response (see Section 1.2 above). However, asymmetries in speech sound discrimination have been demonstrated in the MMN amplitude for both vowels (e.g., Eulitz & Lahiri, 2004) and consonants (Scharinger, Merickel, Riley, & Idsardi, 2011). This type of asymmetries entails greater MMN amplitude when the change between two sounds occurs in one direction than when it occurs in the other direction. For example, the MMN is more negative when /y/ is presented as a deviant among /u/ standards than when /u/ is presented as deviant among /y/ standards (de Jonge & Boersma, 2015). The relative size of these asymmetries has been proposed to reflect different degrees of specifications of phonemic characteristics of speech sounds (de Jonge & Boersma, 2015; Schluter, Politzer-Ahles, Al Kaabi, & Almeida, 2017). However, studies of this kind assume that the influence of the acoustic difference between standard and deviant can be eliminated by measuring the MMN in a difference wave created by subtracting the ERP responses to the same physical stimulus presented both as deviant and standard (in different blocks). This is a problematic assumption, since the exact morphology of the ERP response to the deviant sound depends not only on its own characteristics, but also on the degree to which the same neuronal populations are activated by the deviant and the standard sound (May & Tiitinen, 2010). That is, the deviant-among-standards ERP waveform is inevitably influenced by the characteristics of the standard sound, and directional asymmetries in the MMN amplitude may be related to acoustic properties of the sounds rather than linguistic ones. For example, focalization of formant frequencies has been proposed as a potential explanation to directional asymmetries in vowel perception (Polka & Bohn, 2011).

It has also been investigated how different levels of linguistic information associated with the change between standard and deviant influence the MMN response. The results indicate that the presence of a difference in an additional linguistic dimension (lexical) on top of the phonemic difference (isolated vowels vs. words) actually results in a lower MMN amplitude (Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). However, the interstimulus interval (ISI), which also has an effect on the MMN amplitude (Näätänen et al., 2007), differed between the conditions. This makes the comparison of MMN amplitude between conditions problematic, since it is not clear to what extent the higher amplitude in the isolated vowel condition can be explained by the shorter ISI in that condition.

It is thus not clear whether the linguistic part of the MMN in itself varies with some measure of linguistic distance between standard and deviant, or if it is a binary response. To further investigate this, it would be very valuable to be able to isolate the part of the MMN response that is specific to the linguistic difference between standard and deviant.

1.4. Isolating the linguistic parts of the MMN response

The optimal way of isolating the MMN response to the linguistic difference from that to the acoustic difference would of course be to present the linguistic difference only. That is however not feasible, since it is to a great extent the acoustic signal that provides the cues to the linguistic message, and the linguistic information in the signal is therefore greatly intertwined with the acoustic information.

In some sense, several studies have isolated the MMN response to the linguistic difference, by showing that MMN is elicited in response to linguistic differences between standard and deviant in the presence of various amounts of acoustic variation within both standard and deviant stimuli (Aulanko, Hari, Lounasmaa, Näätänen, & Sams, 1993; Deguchi et al., 2010; Jacobsen, Schröger, & Alter, 2004; Phillips et al., 2000; Shestakova et al., 2002). It can be argued that the acoustic variability in the stimuli ensures that no systematic acoustic differences are present that would elicit an MMN on their own, and that the activation therefore must be in response to the linguistic difference only. However, this does not provide much information about the characteristics of the MMN response specific to the linguistic difference, other than the fact that it can be elicited amongst high variability in the acoustic differences.

An alternative way to isolate the linguistic part of the MMN is to determine the impact of the acoustic difference on the MMN response, and subtract it from the total MMN response. The remainder then arguably represents the part of the MMN response tied specifically to the linguistic difference.

Determining the impact of the acoustic difference on the MMN response to a combined linguistic and acoustic difference can be done by comparing two conditions in which the acoustic difference between standard and deviant are equal, and in which a linguistic difference between the two is present only in one condition. This is a fairly typical setup when comparing within-category changes to between-category changes (e.g., Sharma & Dorman, 1999), or native contrasts to non-native contrasts (e.g., Sharma & Dorman, 2000). However, achieving matched acoustic differences in the two conditions often requires limiting the differences to a few acoustic dimensions, and in most cases the stimuli will need to be synthesized.

Using a non-speech condition that approximates the acoustic difference between two speech sounds is an alternative option. Both pure and complex tones have previously been used as non-speech conditions in MMN experiments, for example by matching the duration of the tone to the duration of the speech stimuli (e.g., Kasai et al., 2001) or by matching formant values to the partials in complex tones (e.g., Jacobsen, Schröger, & Sussman, 2004). However, for the purposes of approximating the acoustic difference between speech sounds, the complexity of the non-speech sounds should preferably be better matched to the acoustic speech signal. There are also non-speech sounds that to some extent preserve the linguistic content while distorting the acoustic signal, for example vocoded speech (e.g., Bernstein, Auer, Eberhardt, & Jiang, 2013) and sine wave speech (e.g., Stekelenburg & Vroomen, 2012). This approach is useful for showing that there is a response to a linguistic difference, just like the studies with high acoustic variability in both standards and deviants mentioned earlier. With those types of non-speech stimuli, it would also be fairly straightforward to create acoustically matched differences that are not perceived as linguistic differences. However, this type of non-speech is not suitable for approximation of the acoustic difference between two natural sounding speech sounds since the acoustic signal is so dissimilar to speech.

A non-speech signal that is acoustically similar to speech can be created by manipulating individual speech sounds (or parts of speech sounds) in order to remove the linguistic content. Kuuluvainen et al. (2014), for example, replaced the burst of a plosive with random noise, and created pseudo-vowels with only one “formant”. The resulting signal is acoustically similar to speech, but does not convey linguistic content. The same can be accomplished by spectrally rotating the speech signal (Blesser, 1972); the fundamental acoustic characteristics of the signal are similar to speech in many ways, but they do not provide cues to any linguistic content.

1.5. Spectrally rotated speech

Spectrally rotated speech has previously been used as a non-speech condition in both behavioral (e.g., Vandermosten et al., 2011) and brain imaging studies (e.g., Christmann et al., 2014; Narain et al., 2003). Rotated speech is created by multiplying the original speech waveform with a high-frequency carrier wave and then applying a filter to the signal in order to extract the lower sideband of the modulated signal’s spectrum (Blesser, 1972). Multiplying the waveform with a carrier wave centers the spectrum on the frequency of that wave, resulting in the negative frequencies shifting up to above zero. Applying a low-pass filter with the frequency of the carrier wave as cut-off frequency removes the positive part of the spectrum preserving only the negative part. Since this mirrors the positive part of the spectrum, the remaining signal is a spectrally rotated version of the original signal. If the original waveform at one point in time has a high level of energy at low frequencies, the rotated signal will at the corresponding time-point have a high level of energy at high frequencies.

The spectral similarity of rotated speech to speech is dependent primarily on the frequency of the carrier wave used for the transformation, since this is what determines the total frequency span to be included in the “rotation”. The higher the frequency of the carrier wave, the wider the span that is rotated (i.e. low formants would transform into higher frequency pseudo-formants with higher carrier wave frequency). When trying to keep the acoustic difference between two rotated sounds as similar as possible to the acoustic difference between the two original speech sounds, it is important to use a carrier wave that places the pseudo-formants in the rotated speech within a reasonable formant frequency range.

Importantly, the basic acoustic form of spectrally rotated speech is roughly the same as that of the corresponding speech signal. As continuous signals, both speech and rotated speech can be described as a sequence of segments that are either periodic in nature (vowels and other voiced segments in speech), filtered noise (fricatives and plosive releases) or silence (plosive closures). A crucial similarity between speech and its rotated counterpart is that order, duration and type of the original speech sounds is preserved in the transformation. For instance, spectral rotation of a vowel followed by a plosive will result in a periodic segment followed by silence, then by a noise burst. That is, both durational information and the general basic acoustic form from the original speech waveform are preserved in the rotated signal. This is of course applicable to isolated segments as well; the basic acoustic characteristics will be the same (e.g., a rotated fricative will consist of noise whereas a rotated vowel will consist of periodic amplitude variations and rich spectral information), and durational dynamics will transfer to the rotated signal (e.g., formant transitions will be preserved, albeit spectrally rotated). Additionally, the fundamental frequency of voiced segments is preserved, since the basic harmonic spacing of the waveform is not altered when multiplying the complex speech waveform with the carrier wave.

The major points of dissimilarity between rotated speech and speech are the spectral tilt and the distribution of formants/pseudo-formants. In speech, the amplitude of the harmonics is in general more dampened at higher frequencies (Fant, 1970), whereas in rotated speech, the amplitude of harmonics is instead greater at higher frequencies, since intensity information is preserved in the transformation. That is, the spectral tilt of rotated speech is opposite and thus characteristically different from that of speech. Regarding formant frequencies, the same physical frequency difference (measured in Hz) will result in different perceptual frequency differences (measured in Bark) depending on the frequency. The same physical difference will be more salient at lower frequencies than at higher frequencies. That is, a fairly small difference between the first formants of two sounds can be perfectly perceivable, but when the sounds are rotated and that difference is between higher frequencies, it might not be as easily perceived.

Whereas the acoustic difference between two spectrally rotated speech sounds is arguably approximately of the same magnitude as the acoustic difference between the two original speech sounds, the differences are not, and cannot be, completely equal. Nevertheless, rotated speech is still, by and large, a non-speech signal that is acoustically similar to speech.

1.6. The present study

This study investigates whether spectrally rotated speech can be used to determine the impact of the acoustic difference on the MMN response to a combined linguistic and acoustic change between sounds. Rotated speech is investigated because it is arguably one of the most viable options when trying to approximate the MMN response to the acoustic difference between two speech sounds. Two experiments were conducted.

In Experiment 1, speech stimuli and non-speech stimuli, in the form of spectrally rotated speech, were presented in different blocks. The speech stimuli were synthesized vowels, and the non-speech stimuli were rotated versions of those same vowels. In the speech block, the standard vowel was /e/. In the between-category change condition (Sp-BC), the deviant vowel was /i/, and in the within-category change condition (Sp-WC), the deviant was a second exemplar of /e/. Both deviants were equally distant acoustically from the standard in terms of the first four formants. The design of the non-speech block mirrored the speech block. Importantly however, although the non-speech conditions will be denoted as between-category (Ro-BC) and within-category (Ro-WC) for ease of comparison with the speech block, no perceptual categories are expected for the rotated vowels. Consequently, Ro-BC refers to the condition in which the *non-rotated* versions of standard and deviant belong to different vowel categories.

In the Sp-BC condition, the difference between standard and deviant is both acoustic and linguistic in nature. In the Sp-WC condition, the Ro-BC condition, and the Ro-WC condition, the type of change is acoustic only. Based on this distinction, the Sp-BC condition is hypothesized to yield a larger MMN response than the other three conditions. Further, the acoustic difference between standard and deviant is of approximately the same magnitude in all three conditions with only acoustic changes, so no differences in MMN amplitude are expected.

In Experiment 2, the MMN response to rotated speech is compared to another type of non-speech stimuli, namely complex tones, in how well they compare to a within-category vowel change. In the rotated speech condition (Ro), stimuli were two rotated vowels that before rotation belonged to different categories (the same stimuli as in condition Ro-BC in Experiment 1). In the speech condition (Sp), two exemplars of the same vowel were used, crucially with equal acoustic distance between them as between the two vowels in the Ro condition, before rotation (the same stimuli as in condition Sp-WC in Experiment 1). In the complex tone condition (To), the tones were created to be as similar to the stimuli in the Ro condition as possible. Sine tones with frequencies matching the pseudo-formant values of the rotated stimuli were combined to create the complex tones.

All three conditions can be used to approximate the acoustic part of the MMN in response the two vowels in the Sp-BC condition in Experiment 1. Since a within-category speech sound change is the most commonly used strategy, the Sp condition functions as a gold standard for the other two conditions. Comparing how close the Ro and To conditions come to the Sp condition will shed light on the relative merits of rotated speech and complex tones for approximating the acoustic part of the MMN.

If rotated speech can be used in this way it would greatly facilitate the design of studies in which linguistic differences are investigated using the MMN. Controlling for the acoustic difference between standard and deviant could then be relatively easily accomplished by including a condition with rotated versions of the relevant speech stimuli.

2. Results

For ERP-traces of standard and deviant in each condition of Experiment 1, see Fig. 1. The MMN response can be seen in the difference waves (deviant minus standard) for all four conditions (see Figs. 2 and 3).

Individual one-sample two-tailed t-tests ($H_0 = 0$) showed a significant difference from zero for all four conditions in Experiment 1, confirming that all conditions elicited an MMN (see Table 1).

A repeated measures ANOVA with speech type (normal vs. rotated) and change type (between-category vs. within-category) as within-subject factors revealed a significant effect of speech type ($F(1,11) = 5.527$, $p = .038$), but not for change type ($F(1,11) = 0.855$, $p = .375$). There was a significant interaction between change type and speech type ($F(1,11) = 10.318$, $p = .008$). Post-hoc LSD-tests (conditions treated independently) showed that Sp-BC differed significantly from Sp-WC, Ro-BC and Ro-WC ($p = .046$, $p = .026$, and $p = .044$ respectively), but comparisons between all other conditions yielded non-significant results.

To test whether an MMN had been elicited in each of the three conditions in Experiment 2, three two-tailed one-sample t-tests ($H_0 = 0$) were run on the MMN amplitude (Table 2). All conditions differed significantly from zero.

The ERP-traces from Experiment 2 can be seen in Fig. 4, and the difference waves in Fig. 5. A repeated measures ANOVA revealed a significant effect of condition (Sp, Ro and To; $F(2,22) = 29.097$, $p < .001$), and within-subject contrasts showed that both the difference between Sp and Ro ($F(1,11) = 14.615$, $p = .003$) and the difference between Sp and To ($F(1,11) = 19.474$, $p = .001$) were significant. Lastly, the amplitude difference between Sp and Ro (Sp-Ro) was compared to the amplitude difference between Sp and To (Sp-To). A paired-samples *t*-test showed that the Sp-Ro difference was significantly different from the Sp-To difference ($t(11) = -6.614$, $p < .001$). The mean difference for Sp-Ro was $-1.12\,\mu V$, and for Sp-To it was $2.97\,\mu V$.

3. Discussion

The MMN amplitude in the condition with a combined acoustic and linguistic difference between standard and deviant (Sp-BC) was higher than in the three conditions in which there was only an acoustic

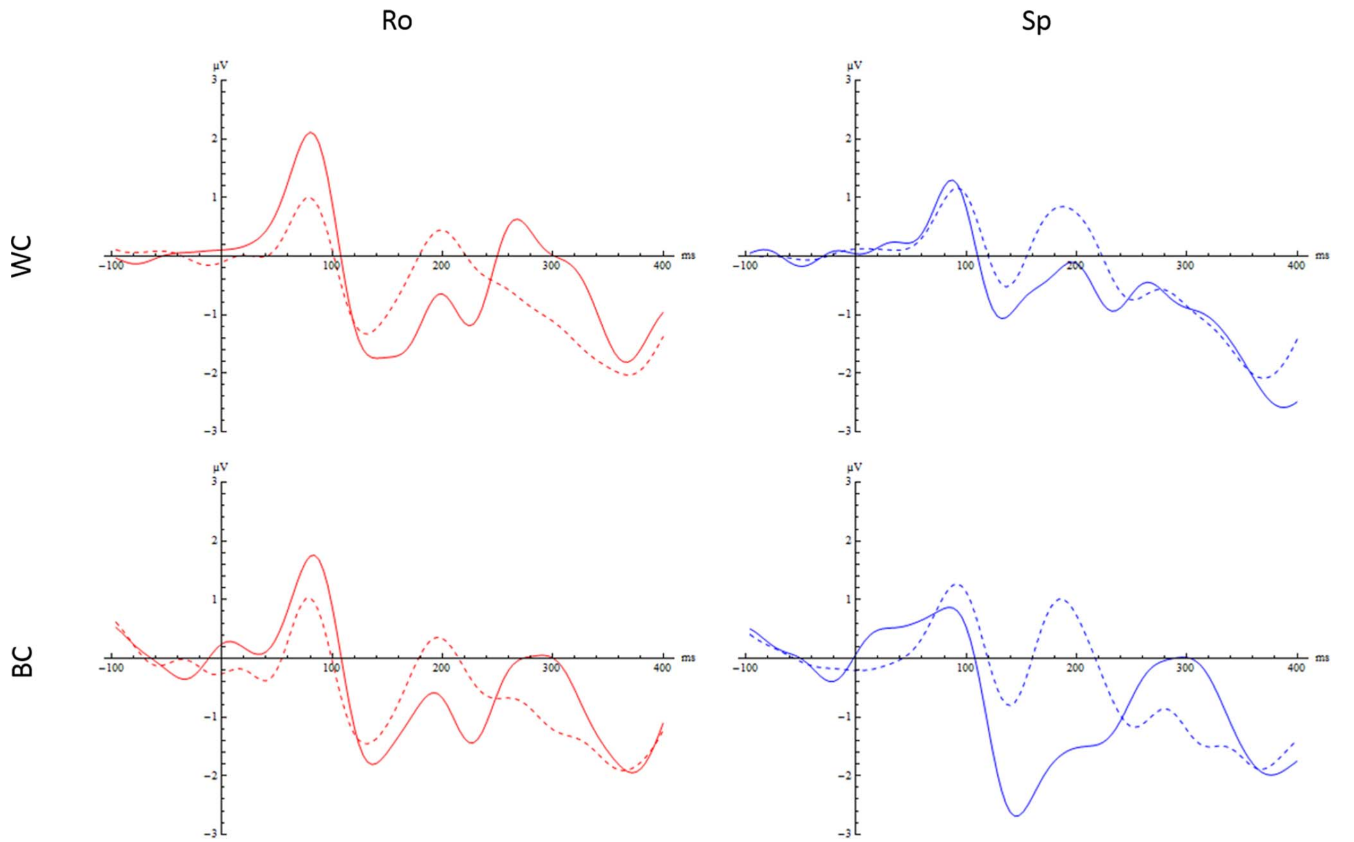

Fig. 1. Grand average ERP-traces from Fz for all four conditions in Experiment 1. Solid lines indicate deviant ERPs and dashed lines indicate standard ERPs.

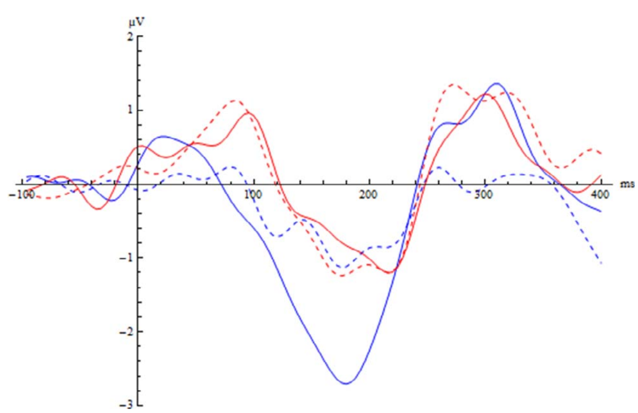

Fig. 2. Difference waves at Fz for all four conditions in Experiment 1. Negativity is plotted downwards. The blue solid line represents the Sp-BC condition, the blue dashed line the Sp-WC condition, the red solid line represents the Ro-BC condition, and the red dashed line represents the Ro-WC condition.

difference of equal (Sp-WC) or approximately equal (Ro-BC and Ro-WC) magnitude. Further, the MMN amplitude in response to a rotated vowel pair (Ro) better approximates the acoustic part of the MMN response (gold standard being a within-category change with equal acoustic difference magnitude, i.e., Sp) to a speech sound difference than do complex tones (To) with comparable spectral characteristics. In general terms, the MMN response to the acoustic difference between speech sounds can be fairly well approximated by the MMN response to their rotated counterparts. This in turn makes it possible to estimate the part of the MMN response specific to the linguistic difference.

Being able to isolate the part of the MMN response specific to the linguistic difference between standard and deviant can be useful in various ways. For example, the transition from acoustic to linguistic perception of speech that has been demonstrated to take place during infancy (e.g., Cheour et al., 1998; Kuhl, 2004) could be explored in more detail with the MMN, if the linguistic response can be separated out from the acoustic response. When comparing different age groups in infancy, potential confounds are present in the MMN responses as a result of the rapid growth of the brain as well as physical changes to the cranium, such as closure of the fontanels. This means that the measurable response to acoustic differences cannot be assumed to be consistent across ages, which in turn means that it is difficult to ascertain to what extent an observed difference between ages is related to language development. Using rotated speech to control for the acoustic difference between linguistic stimuli simultaneously controls for developmental differences, which means that changes in the MMN response particular to language development could potentially be more easily detected across age groups. Further, being able to isolate the linguistic part of the MMN response could speculatively be used to study the processing of various measures of linguistic difference, such as semantic distance or phonological neighborhood density.

Although the present study has provided the first support for the notion that spectrally rotated speech can be used to approximate the MMN response to the acoustic difference between two speech sounds, further research is required to map out the scope and limitations of its usefulness. The present study focused specifically on the amplitude of the MMN, since that measure has consistently been shown to reflect both magnitude of acoustic difference between standard and deviant (e.g., Tiitinen et al., 1994), and the presence of a linguistic difference between the two (Sharma & Dorman, 1999). However, latency and topography measures may also be of interest when attempting to isolate the response to a linguistic difference. Since a combined acoustic and linguistic difference often results in a more left-lateralized response than an acoustic-only difference (e.g., Kasai et al., 2001; Kuuluvainen et al., 2014), looking specifically at hemispheric lateralization of the MMN response and subtracting the response to the acoustic-only difference could potentially be key to describe the isolated linguistic

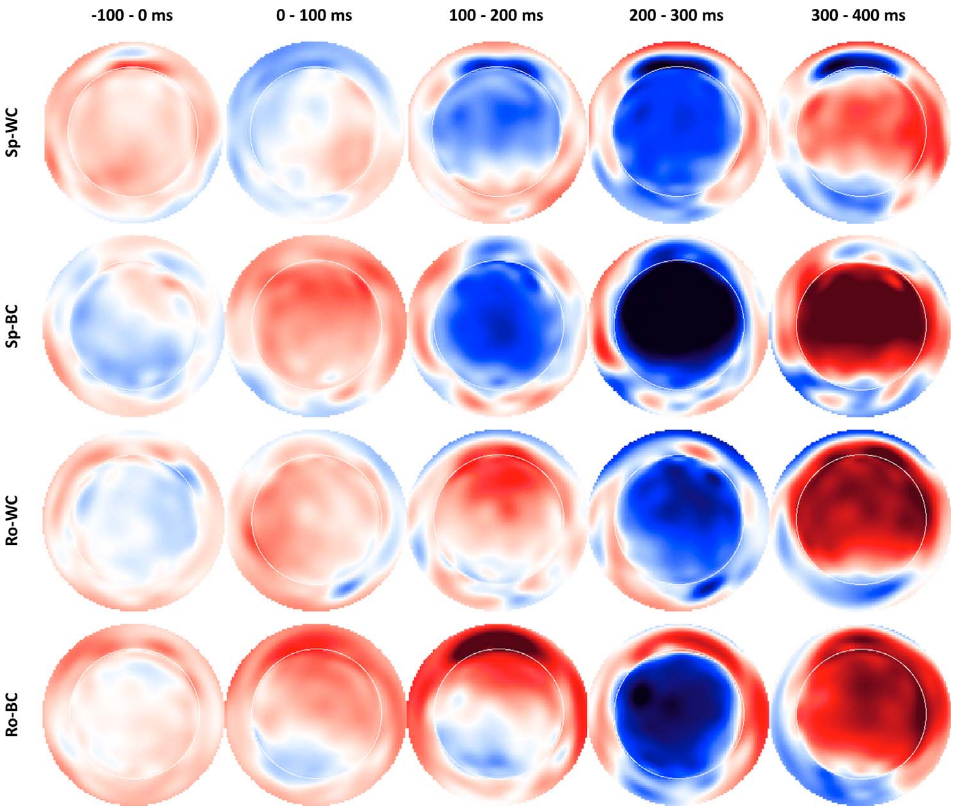

Fig. 3. Topographies of the difference waves for all four conditions, in time windows of 100 ms. Nose is up.

Table 1
Mean MMN amplitude for all conditions, and results of the four one-sample *t*-tests. In all conditions, mean amplitude of the MMN was less than zero. Significance is marked by an asterisk.

Condition	Mean amplitude (μV)	SD	*t*	df	*p*
Sp-BC	−2.64	2.18	−4.196	11	.001*
Sp-WC	−1.25	1.19	−3.626	11	.004*
Ro-BC	−1.08	1.70	−2.200	11	.050*
Ro-WC	−1.24	1.41	−3.047	11	.011*

Table 2
Mean MMN amplitude for all conditions in Experiment 2, and results of the three one-sample *t*-tests. In all conditions, mean amplitude of the MMN was less than zero. Significance is marked by an asterisk.

Condition	Mean amplitude (μV)	SD	*t*	df	*p*
Sp	−2.60	1.67	−5.407	11	.001*
Ro	−1.48	1.29	−3.965	11	.002*
To	−5.58	1.97	−9.832	11	.001*

response. Additionally, it should be noted that magnitude of difference between standard and deviant is of course not the only variable influencing the amplitude of the MMN. The proportion of standards to deviants, number of standards preceding a deviant, inter-stimulus interval, acoustic complexity of the stimuli, attention to stimuli, and various aspects of familiarity with the stimuli have all been demonstrated to impact the size of the MMN response (see Kujala & Näätänen, 2010; Näätänen et al., 2007, for reviews). While most of those can be controlled and kept constant within a single experiment, overall familiarity with the stimuli is, for obvious reasons, difficult to match between speech and rotated speech conditions. It would certainly be of interest to explore the possible implications of this for the use of rotated speech as a control condition for acoustic difference between standard and deviant. However, it is likely that constructing exactly matched conditions in terms of familiarity is an unattainable target, just like in the case of matching the acoustic difference magnitude between conditions, and that a sufficient approximate match may be a more appropriate goal for future forays into this topic.

The present study compared the responses to acoustic differences of a spectral nature, and contrasted speech and non-speech stimuli. First and foremost, the responses to other measures of acoustic difference need to be compared between rotated and un-rotated conditions before the results of the present study can be generalized in a broader sense. Second, the spectral dimension is one in which the perceptual space (in terms of speech) has been consistently reported to be warped relative to the acoustic space (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Masapollo, Polka, & Ménard, 2015). This means that a within-category variation of a spectral nature may not be perceived to be as different as the acoustic difference suggests it to be, thus confounding the comparison between responses to acoustically matched differences in speech and non-speech stimuli. To rule out effects of this nature, again other measures of acoustic difference should be explored. For example, matched intensity differences have been shown to yield similar MMN responses in speech and non-speech conditions (Sorokin et al., 2010), and could be a viable option. Another option to make sure that perceived differences between rotated versions of sounds as accurately as possible reflect the perceived differences between the un-rotated versions of the same sounds, is to use non-speech sounds in both the rotated and the non-rotated conditions.

Future studies exploring the use of spectrally rotated speech as a way to approximate the MMN response to an acoustic difference between speech sounds, and thus isolating the response to the linguistic difference, may benefit from a modified procedure for creating rotated speech introduced by Christmann et al. (2014). Instead of filtering out frequencies above the rotated section, the original spectral information was preserved at higher frequencies. This modification should be especially useful when using natural speech stimuli and rotated counterparts, as it preserves the naturalness of the un-rotated stimuli while maintaining the overall approximate match in acoustic difference

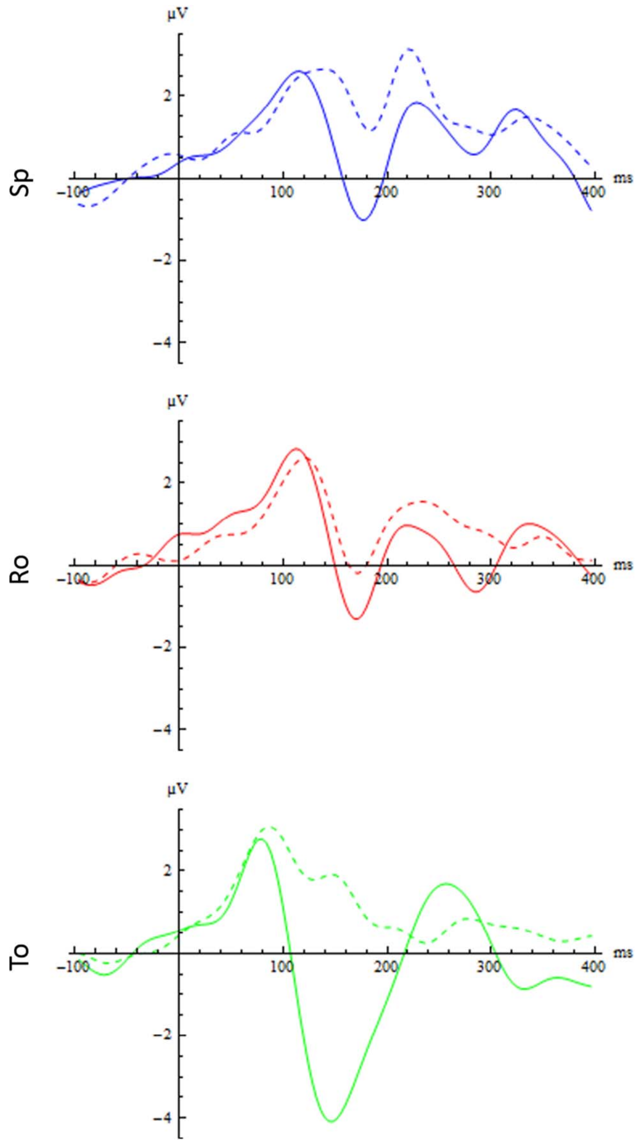

Fig. 4. Grand average ERP-traces for the three conditions in Experiment 2, from Fz. Solid lines indicate deviant ERPs and dashed lines indicate standard ERPs.

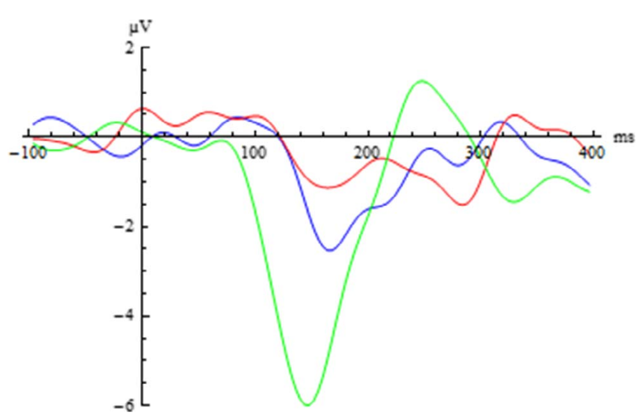

Fig. 5. Difference waves for all three conditions at Fz. Negativity is plotted downwards. The blue line represents the Sp condition, the red line the Ro condition, and the green line the To condition. A clear MMN response can be seen for all three conditions.

between the rotated and un-rotated conditions.

In conclusion, this study has shown that the amplitude of the MMN response to the acoustic difference between two vowels can be approximated by the amplitude of the response to their spectrally rotated counterparts. This makes it possible to calculate the amplitude of the MMN response to the linguistic difference between the two vowels. Isolating the response to the linguistic difference opens up numerous avenues for future research, for example a detailed exploration of the transition from acoustic to linguistic perception of speech in early development.

4. Method

4.1. Experiment 1

Participants: 12 right-handed native speakers of Swedish participated in the study (6 male and 6 female, mean age 30 years, age range 24–39 years). Two additional recordings were performed but were excluded due to technical failure or multiple interruptions during recording. The subjects gave informed consent before participation and were given movie vouchers as compensation for their contribution. The study has been approved by the Ethical Review Board at Karolinska Institutet (2011/955-31/1).

Stimuli: The speech stimuli were three synthesized vowels on a continuum of nine tokens (V1-V9) ranging from /i/ (V1) to /e/ (V9), created in Praat 5.3.13 (Boersma & Weenink, 2012) and their rotated counterparts. All vowels were 200 ms long with 50 ms fade in/out. The fundamental frequency was 200–220 Hz and had a symmetrical low-high-low contour with linear rise and fall. The standard vowel was an /e/ token from the mid-part of the continuum (V6), and the deviants were an /i/ token (V3; BC conditions) and an /e/ token (V9; WC conditions). Both deviants were equally distant acoustically from the standard in terms of formant frequencies. The formant frequencies of the three vowels in the Sp conditions can be seen in Table 3. Rotated speech was created by multiplying the original speech waveform with a high-frequency carrier wave and then applying a low-pass filter to the signal (Blesser, 1972). Mathematica 9 (Wolfram Research Inc., Champaign, Illinois, USA) was used for the transformation and 4602 Hz was used as carrier wave frequency and filter cut-off frequency. Participants were not systematically asked about how, in general, they had perceived the sounds in the experiment, but a number of participants spontaneously commented on the stimuli after the end of the experiment. The speech stimuli were reported to be recognizable as vowels, albeit clearly synthesized and not naturally produced, whereas the rotated versions were reported to sound like mechanical noise. Spectrograms of the deviant stimuli can be seen in Fig. 6.

Experiment design: The experiment consisted of four blocks; two speech conditions (Sp and Ro) by two change conditions (BC and WC). The order of the speech conditions was counterbalanced between participants, but within each speech condition, the order of the two types of changes was fixed, with the WC block always presented before the BC block. The presence of a linguistically relevant change has been shown to suppress the MMN response to linguistically irrelevant changes (Lipski & Mathiak, 2008), so in order to not risk suppressing the MMN response to a linguistically irrelevant change, the WC block was always

Table 3
Formant values of the synthesized vowels in the Sp conditions (top), and values for the "pseudo-formants" of the Ro conditions (bottom).

Speech		F1	F2	F3	F4
Standard	/e/	376 Hz	2557 Hz	3630 Hz	4226 Hz
WC deviant	/e/	476 Hz	2607 Hz	3580 Hz	4126 Hz
BC deviant	/i/	276 Hz	2507 Hz	3680 Hz	4326 Hz
Rotated speech		"F1"	"F2"	"F3"	"F4"
Standard		376 Hz	972 Hz	2045 Hz	4226 Hz
WC deviant		476 Hz	1022 Hz	1995 Hz	4126 Hz
BC deviant		276 Hz	922 Hz	2095 Hz	4326 Hz

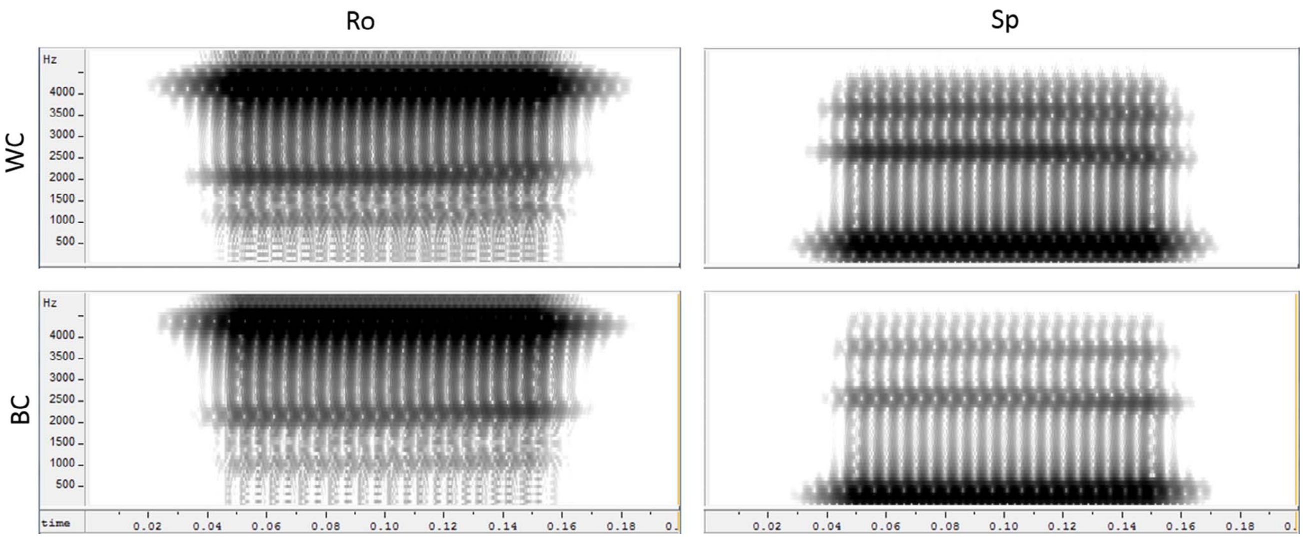

Fig. 6. Spectrograms of the deviant stimuli, to illustrate the spectral characteristics of speech and rotated speech stimuli.

presented before the BC block in the Sp condition. To keep the two speech conditions comparable, the same order was used in the Ro condition. Each block consisted of 750 trials, in 150 of which the deviant sound was played. The trials were presented in random order except that the first twelve trials of each block always contained standards, and each deviant trial was preceded by at least two standard trials. The stimulus onset asynchrony (onset-to-onset) was 500 ms. E-Prime 2.0 (Psychology Software Tools, Sharpsburg, Pennsylvania, USA) was used for stimulus presentation, and recording of the EEG was done in EGI NetStation 4.4.2 (Electrical Geodesic Inc., Eugene, Oregon, USA).

Procedure: The EEG-net was applied and electrode impedance measured to ensure that the impedance did not exceed 100 kΩ. The net was a Hydro-Cel 128-electrode high-impedance net (Electrical Geodesics Inc.), which includes electrodes for electrooculogram recording at the side of and under each eye. Reference during recording was Cz. In order to direct attention away from the auditory stimuli during the experiment, participants were asked to count the number of occurrences of specified objects in a silent film that they were watching on a screen during the experiment. Stimuli were presented via two loudspeakers (placed in front of the participant at a distance of 100 cm) at approximately 65 dB SPL, measured at the participant's head. The duration of the experiment (excluding net application and preparations) was approximately 30 min. After the MMN experiment, the EEG-net was removed and the subjects took part in a short behavioral perceptual experiment, which is not relevant to the present paper. The duration of the entire session was approximately one hour.

Preprocessing and analysis: An off-line band-pass filter of 1–20 Hz was applied to the continuous EEG-recording, and it was re-referenced to the average of the mastoid electrodes. It was then segmented into single-trial epochs of 500 ms with a 100 ms pre-stimulus baseline period. Epochs were excluded from further analysis if they either contained eye-artifacts or if the signal range exceeded 55 μV in 20 channels or more. In both cases, the signal was averaged within a moving window of 80 ms. Eye-artifacts were defined as regions in which range within consecutive 100 ms windows exceeded 140 μV in the horizontal electrooculogram channels (eye-movements), or the range within the entire segment exceeded 55 μV in the vertical electrooculogram channels (eye-blinks). First, subject average waveforms were created for standard and deviant stimuli in each condition, then difference waveforms were calculated for each subject and condition. Channel 11 in the Hydro-Cel net, corresponding to Fz in the 10–20 system, was selected for statistical analysis of the MMN amplitude. The negative peak within 130–210 ms after stimulus onset was identified in the individual subject difference waveforms, and the average of a 40 ms time window centered at the peak was extracted for statistical analysis. The wider time window was chosen based on visual inspection of where standard and deviant grand average waveforms (blind to experimental conditions) deviated from each other. EEG-data was preprocessed in EGI NetStation 4.4 (Electrical Geodesic Inc). Statistical tests were performed in SPSS 21 (International Business Machines Corp., Armonk, New York, USA).

4.2. Experiment 2

Participants: 12 right-handed native speakers of Swedish participated in the study (6 male and 6 female, mean age 33 years, age range 23–45 years). One additional session was completed but was excluded due to technical failure during recording. The subjects gave informed consent before participation and were given movie vouchers as compensation for their contribution. The study has been approved by the Regional Ethical Review Board (2015/63-31).

Stimuli: Stimuli from the Sp-WC and Ro-BC conditions from Experiment 1 were used also in Experiment 2. In addition, two complex tones were used, created in Praat 6.0.21 (Boersma & Weenink, 2012), by combining sine tones with frequencies corresponding to the pseudo-formant values of the rotated standard sound and the rotated BC deviant sound (Table 3). The duration of the complex tones was 200 ms, with 50 ms rise/fall time. Intensity was normalized across all six files.

Experiment design: The experiment had three conditions: one speech condition (Sp), one rotated speech condition (Ro), and one tone condition (To). The order of presentation of the conditions was counterbalanced between participants. Each condition consisted of five blocks of stimuli presentation with 15 s pauses in between. In each block, 200 sounds were presented, 40 of which were deviants (800 standards and 200 deviants in total per condition). The sounds were presented in random order except that the first ten sounds of each block were always standards, and each deviant was preceded by at least one standard. The stimulus onset asynchrony (onset-to-onset) was 500 ms. E-Prime 2.0 (Psychology Software Tools, Sharpsburg, Pennsylvania, USA) was used for stimulus presentation, and EEG-recording was done using the BioSemi ActiveTwo system and ActiView 7.06 software (BioSemi, Amsterdam, The Netherlands). The sampling rate during recording was 2048 Hz, and a driven-leg reference (a CMS/DRL loop with voltage recorded relative to the CMS electrode) was used.

Procedure: Electrodes were fastened above and below the left eye of the participant, as well as at the outer side of each eye and behind the ears at the mastoid bones. Participants then donned a head-cap to which 16 electrodes were fastened (Fp1, Fp2, F3, F4, Fz, T7, T8, C3, C4, Cz, P3, P4, Pz, O1, O2 and Oz). During the experiment, participants watched a self-selected movie (muted with Swedish subtitles), and were instructed to remain as still as possible during stimuli presentation and move around and stretch during the pauses. The experiment was 30 min long including pauses, and each recording session (including electrode application) lasted approximately one hour. Stimuli were presented via loudspeakers (placed in front of the participant at a distance of 130 cm) at approximately 56 dB SPL, measured at the participant's head.

Preprocessing and analysis: The same preprocessing procedure was applied to the EEG data as in Experiment 1, except artifact detection and rejection. Independent component analysis was used to identify and remove eye-movement and eye-blink artifacts, and the threshold for inclusion of trials was ± 50 μV in any channel. As in Experiment 1, visual inspection of where standard and deviant grand average waveform differed determined the time window within which the negative peak of the difference wave was identified. The time window was 110–210 ms, and like in Experiment 1 the average of 40 ms centered around the negative peak was used in the analysis. Software used for data preprocessing and analysis were MATLAB R2014b (MathWorks Inc., Natick, Massachusetts, USA), EEGLAB 13 (Delorme & Makeig, 2004), Mathematica 9 (Wolfram Research Inc., Champaign, Illinois, USA), and SPSS 21 (International Business Machines Corp., Armonk, New York, USA).

Conflicts of interest

None.

Statement of significance to the neurobiology of language

It is important to differentiate between processing of linguistically relevant information and acoustic characteristics of the speech signal. Our results indicate that processing of spectrally rotated speech is comparable to processing of the acoustic aspect of speech, making it a useful tool for separating acoustic and linguistic processing.

Acknowledgements

This work was supported by Stockholm University. The authors would like to thank all participants for their contribution, and the anonymous reviewers for many helpful comments.

References

Aaltonen, O., Eerola, O., Hellström, Å., 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*(2), 1090–1105. http://dx.doi.org/10.1121/1.418031.

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R., & Sams, M. (1993). Phonetic invariance in the human auditory cortex. *NeuroReport, 4*(12), 1356–1358.

Bernstein, L. E., Auer, E. T., Jr., Eberhardt, S. P., & Jiang, J. (2013). Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Frontiers in Neuroscience, 7*, 34. http://dx.doi.org/10.3389/fnins.2013.00034.

Blesser, B. (1972). Speech perception under conditions of spectral transformation: I∗ Phonetic characteristics. *Journal of Speech, Language & Hearing Research, 15*(1), 5–41. http://dx.doi.org/10.1044/jshr.1501.05.

Boersma, P., & Weenink, D.. Praat: Doing phonetics by computer [Computer program]. Version 5.3.13. (2012). <http://www.praat.org/> Retrieved from 18 April 2012.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Ahlo, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351–353. http://dx.doi.org/10.1038/1561.

Christmann, C. A., Berti, S., Steinbrink, C., & Lachmann, T. (2014). 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). *Brain and Language, 136*, 8–18. http://dx.doi.org/10.1016/j.bandl.2014.07.004.

de Jonge, M. J., & Boersma, P. (2015). French high-mid vowels are underspecified for height. *Proceedings of the 18th international congress of phonetic sciences*.

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. http://dx.doi.org/10.1016/j.brainres.2010.09.104.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*(4), 919–924. http://dx.doi.org/10.1097/00001756-199703030-00021.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*(4), 577–583. http://dx.doi.org/10.1162/089892904323057308.

Fant, G. (1970). *Acoustic theory of speech production.* Paris: Mouton & Co. N.V., Publishers, The Hauge.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*(4), 654–659. http://dx.doi.org/10.1111/1469-8986.2004.00175.x.

Jacobsen, T., Schröger, E., & Sussman, E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research, 20*(3), 473–479. http://dx.doi.org/10.1016/j.cogbrainres.2004.03.021.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., ... Kato, N. (2001). Brain lateralization for mismatch response to across-and within-category change of vowels. *NeuroReport, 12*(11), 2467–2471. http://dx.doi.org/10.1097/00001756-200108080-00036.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*, 831–843. http://dx.doi.org/10.1038/nrn1533.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. E. F. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606–608. http://dx.doi.org/10.1126/science.1736364.

Kujala, T., & Näätänen, R. (2010). The adaptive brain: A neurophysiological perspective. *Progress in Neurobiology, 91*(1), 55–67. http://dx.doi.org/10.1016/j.pneurobio.2010.01.006.

Kuuluvainen, S., Nevalainen, P., Sorokin, A., Mittag, M., Partanen, E., Putkinen, V., ... Kujala, T. (2014). The neural basis of sublexical speech and corresponding nonspeech processing: A combined EEG–MEG study. *Brain and Language, 130*, 19–32. http://dx.doi.org/10.1016/j.bandl.2014.01.008.

Lipski, S. C., & Mathiak, K. (2008). Auditory mismatch negativity for speech sound contrasts is modulated by language context. *NeuroReport, 19*(10), 1079–1083. http://dx.doi.org/10.1097/WNR.0b013e3283056378.

Marklund, E., Schwarz, I. C., & Lacerda, F. (2014). Mismatch negativity at Fz in response to within-category changes of the vowel/i. *NeuroReport, 25*(10), 756–759. http://dx.doi.org/10.1097/WNR.0000000000000168.

Masapollo, M., Polka, L., & Ménard, L. (2015). Asymmetries in vowel perception: Effects of formant convergence and category "goodness". *Journal of the Acoustical Society of America, 137*, 2385. http://dx.doi.org/10.1121/1.4920678.

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology, 47*(1), 66–122. http://dx.doi.org/10.1111/j.1469-8986.2009.00856.x.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... Ahlo, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432–434. http://dx.doi.org/10.1038/385432a0.

Näätänen, R., Paavilainen, P., Rinne, T., & Ahlo, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590. http://dx.doi.org/10.1016/j.clinph.2007.04.026.

Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex, 13*(12), 1362–1368. http://dx.doi.org/10.1093/cercor/bhg083.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: Influence of music training. *Ear and Hearing, 30*(4), 432–446. http://dx.doi.org/10.1097/AUD.0b013e3181a61bf2.

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*(1), 26–36. http://dx.doi.org/10.1016/j.cogbrainres.2003.12.011.

Pakarinen, S., Teinonen, T., Shestakova, A., Kwon, M. S., Kujala, T., Hämäläinen, H., ... Huotilainen, M. (2013). Fast parametric evaluation of central speech-sound processing with mismatch negativity. *International Journal of Psychophysiology, 87*(1), 103–110. http://dx.doi.org/10.1016/j.ijpsycho.2012.11.010.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., ... Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*(6), 1038–1055. http://dx.doi.org/10.1162/08989290051137567.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467–478. http://dx.doi.org/10.1016/j.wocn.2010.08.007.

Scharinger, M., Merickel, J., Riley, J., & Idsardi, W. J. (2011). Neuromagnetic evidence for a featural distinction of English consonants: Sensor-and source-space data. *Brain and Language, 116*(2), 71–82. http://dx.doi.org/10.1016/j.bandl.2010.11.002.

Schluter, K. T., Politzer-Ahles, S., Al Kaabi, M., & Almeida, D. (2017). Laryngeal features are phonetically abstract: Mismatch negativity evidence from Arabic, English, and Russian. *Frontiers in Psychology, 8*. http://dx.doi.org/10.3389/fpsyg.2017.00746.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106*(2), 1078–1083. http://dx.doi.org/10.1121/1.428048.

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*(5), 2697–2703. http://dx.doi.org/10.1121/1.428655.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., ...

Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport, 13*(14), 1813–1816. http://dx.doi.org/10.1097/00001756-200210070-00025.

Sittiprapaporn, W., Tervaniemi, M., Chindaduangratn, C., & Kotchabhakdi, N. (2005). Preattentive discrimination of across-category and within-category change in consonant vowel syllable. *NeuroReport, 16*(13), 1513–1518. http://dx.doi.org/10.1097/01.wnr.0000175618.46677.07.

Sorokin, A., Alku, P., & Kujala, T. (2010). Change and novelty detection in speech and non-speech sound streams. *Brain Research, 1327*, 77–90. http://dx.doi.org/10.1016/j.brainres.2010.02.052.

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological evidence for a multisensory speech-specific mode of perception. *Neuropsychologia, 50*(7), 1425–1431. http://dx.doi.org/10.1016/j.neuropsychologia.2012.02.027.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*(6501), 90–92. http://dx.doi.org/10.1038/372090a0.

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Wouters, J., & Ghesquière, P. (2011). Impairments in speech and nonspeech sound categorization in children with dyslexia are driven by temporal processing difficulties. *Research in Developmental Disabilities, 32*(2), 593–603. http://dx.doi.org/10.1016/j.ridd.2010.12.015.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., ... Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*(3), 357–369. http://dx.doi.org/10.1016/S0926-6410(98)00039-1.

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*(1), 175–185. http://dx.doi.org/10.1016/j.brainres.2005.12.004.