# SPEECH SOUND PERCEPTION AND NEURAL REPRESENTATIONS

*Maija S. Peltola*

Department of Phonetics and the Centre for Cognitive Neuroscience, University of Turku, Finland
maija.peltola@utu.fi

## ABSTRACT

This commentary reviews some of the main findings in speech sound perception using the brain imaging techniques and comments briefly on the recent findings by the session contributors. The main emphasis is on the experimental settings used in these studies. The aim is to demonstrate how the search for the neural correlates for abstract linguistic units has resulted in various types of experimental designs and how the stimulus selection may play a crucial role in the findings. It seems that the experimental settings are becoming more and more elaborate thus offering an access to the abstract levels of representation.

**Keywords:** speech sound perception, abstract representations, experimental settings

## 1. INTRODUCTION

Speech sounds rarely appear without any coarticulatory context, but their perception is often studied so that discrete isolated vowel sound items are presented to the subjects. It may well be that a word or a syllable context would provide a more appropriate setting in terms of naturalness. Disregarding the naturalness domain (which may be of little consequence considering the unnaturalness of the laboratory setting in any case), a linguistically meaningful stimulus setting may be one key into the study of phonological phenomena, since by using minimally contrasting word pairs it may be possible to obtain information on the role of abstract phonological rules [1] or features. However, isolated vowel sounds also provide information on the manner in which speech is perceived (e.g. [7]), and even the abstract underspecified mental lexicon [3] can be studied when the selection of the stimuli is based upon a careful analysis of the phonological role of these vowels in the language to be studied.

One of the main questions in speech perception studies is connected with the level of representation that can be accessed using the methodology available (for a comprehensive review, see [11]). Language-specific speech sound perception has been studied by comparing the perception of vowel sounds in native speakers of different languages (e.g.[7]), and by studying the development of new representations in students of a foreign language (e.g. [8]). The question not often raised in connection with these reports is whether the perceptual differences between various speaker groups are based upon acoustic characteristics of the stimuli or on some more abstract level of representation. The problem may be in the definition of abstractness and phonemic level, which occasionally, but not always, refer to the same thing. In the following discussion, an attempt is made to explore the experimental settings of some studies on speech sound perception and to comment on recent work by the session contributors.

## 2. PREVIOUS FINDINGS

Several studies have addressed the language-specificity of speech sound perception, particularly after the result by Näätänen et al. [7], which showed that the phonological role of a vowel stimulus has a decisive role in the elicitation of the mismatch negativity (MMN or MMNm) auditory evoked response. The main idea in the experimental design was that speakers of two languages (Finnish and Estonian) were presented in the oddball paradigm with stimulus pairs deriving from both languages. The results demonstrated that, if a vowel contrast is phonological in the native language, the MMN response has larger amplitude and it is dominantly localized in the left hemisphere.

In contrast to most previous studies, Kazanina et al. [6] used many acoustically varying exemplars of a sound category as their stimuli. In this manner they were able to confirm the earlier finding by Shestakova et al. [12] that allophonic variation does not result in the elicitation of the MMN or MMNm, while a phonemically significant, boundary crossing stimulus pair does. The stimuli used by Kazanina et al. focused on the role of voicing in syllable initial plosives [t] and [d] which differentiate meaning in Russian, but not in

Korean. These plosives are in complementary distribution in Korean, so the stimulus pair [ta] is acceptable while [da] is not. Despite the apparent similarity with the study by Shestakova et al. [12], the stimuli are crucially different: In contrast to Kazanina et al., Shestakova used three stimulus categories all deriving from Russian to demonstrate that the category membership is perceptually of more significance than the acoustic variability within a category. These categories represented the three corner vowels /i/, /u/, /a/ which are acoustically very distant. Altogether, it would be interesting to see, whether the results would be the same if the stimuli in the contrastive Russian-Korean study were selected so that the phonemic sounds of one language were in free variation in the other, so that the potential role of rule violation would be removed. Also, the Russian study could be further implemented by using categories acoustically close to each other to allow for the within category allophonic variation to be at least equally extensive as the between category acoustic difference.

The contrastive setting where speakers of two languages are used as subjects has proven to be a fruitful experimental setting, and it has been used for non-native language learning research as well. Winkler et al. [13] showed that the MMN response to a Finnish vowel contrast is similar to a native response in fluent Hungarian learners of Finnish, while naïve Hungarians do not show a response. The subjects were immigrants who had lived in Finland for 3-12 years. The stimulus pair significant for both languages was /y/ - /e/ where both the tongue height and the lip rounding are different, whereas only one feature distinguishes between the contrast /Q/ - /e/ which is phonemic only in Finnish. In contrast, our study [8] showed an opposite result, since Finnish classroom learners of English did not show a native-like MMN response to the target language vowel pair /e/ - /I/ while a vowel contrast /i/ - /e/ with the same acoustic distance resulted in a large response, when the stimuli represented prototypical Finnish vowels. Finnish does not utilize the lax feature, which is of significance in the English phonology.

The age of acquisition is difficult to study, since it is questionable to compare adult and child brain responses, since differences in the neural maturity may distort the results. Therefore, Cheour et al. [2] followed the development of new neural representations in 3-6 year old Finnish children learning French in an immersion programme by measuring their MMN responses to both Finnish and French vowel pairs. The study showed that the response to a non-native contrast was significantly enlarged within two months of exposure. In contrast, our recent results [10] using the same experimental setting as in [8] indicated that Finnish children in an English speaking immersion do not acquire the native-like perceptual patterns, since their responses to an English vowel pair were not different from those of the monolingual control group. To further complicate the matter, we have also shown that the MMN response to a non-native French vowel contrast is larger in immersion children in comparison with the monolingual control group [9]. The differences are directly linked with the stimulus selection: In Cheour et al. [2] the stimulus labeled as /E/ actually falls within the Finnish category /O/ and therefore it seems surprising that Finnish children would not hear the contrast labeled as /E/ - /i/ without immersion education, since /O/ - /i/ contrast is phonemic in their native language. In our two conflicting studies, the main difference is that in the former we used prototypical representatives of each native and non-native category, whereas in the latter the stimuli were non-prototypes, which merely crossed the phoneme boundary. Therefore, it seems that a new phoneme boundary may develop prior to the prototype.

## 3. COMMENTARY

### 3.1. Idsardi: Distinctive features and MEG

Idsardi's present contribution [5] focuses on the possibilities offered by the technological advances in neuro-imaging. He begins his paper by presenting a comprehensive review on the three views of the feature specification, viz. the articulatory approach, the perceptual approach and the translational theories. He continues with a brief and informative account on the current magnetoencephalography (MEG) technique including a concise description of the mismatch field experimental setting. The idea of testing phonological natural classes with the mismatch field is of significance and it could be further developed so that such traditional linguistic concepts as markedness and phonological rules could also be studied in order to locate neural correlates for the phenomena sometimes seen as mere tools for abstract linguistic description.

Idsardi goes on to present MEG studies, which have focused on vowel height, nasality and plosive voicing. He presents their own work in progress on the role of F3, which correlates with the length of the vocal tract and may serve as a basis for the formant ratio representation of vowels, which may consequently serve as a perceptual invariant. Their results indicate that the vowels [E] and [«] behave differently when F3 values are systematically varied around the prototypical values. The finding that F1/F3 ratio changes affect the M100 latency for [E], but not for [«], is explained through the lack of other central vowels in English. This seems to be a plausible explanation, since it would suggest that the mental representation for [«] may be less specific due to the large acoustic area that could be occupied by this category in comparison with the acoustic space in the much more crowded front vowel continuum. With further back up from their follow-up study on the back vowel [o], the argument becomes even more convincing.

Altogether, Idsardi presents a concise description of the possible uses of the MEG technique in phonetic studies. He states that the oddball paradigm often used in the mismatch field approach appears to be a suitable tool for the study of phonemic phenomena. Our past and ongoing research has used the oddball paradigm to study the learning and acquisition of non-native speech sounds. Therefore, I am also convinced that, despite the obvious unnaturalness of the stimulus presentation, the oddball paradigm captures accurately the significant perceptual details relevant in speech sound perception.

### 3.2. Hertrich, Lutzenberger and Ackermann: Audiovisual perception

The contribution by Hertrich, Lutzenberger and Ackermann [4] deals with the temporal interaction and eventual fusion of visual and acoustic stimulation. They begin by an extensive overview of the audiovisual speech perception research including such established phenomenon as the McGurk effect. They continue by presenting models of audiovisual fusion showing how different studies have revealed various time windows exhibiting crucial cross-modal interaction. Their present study is a complicated MEG experiment, where an ambiguous synthesized speech stimulus could be perceived either as /pa/ or /ta/ depending on the visually presented articulatory motion. The effects elicited by these speech related audio-visual stimuli are then contrasted with the effects caused by non-speech stimuli. In addition, they present control stimuli to further strengthen their complex and systematic experimental setting. Altogether, the stimulus set in its complexity appears to control many variables and it enables the authors to draw firm conclusions.

The authors are able to show three stages where the two modalities of perception interact. Firstly, they show that the visual stimulus has an impact on the auditory M50 field both with AV and visual only stimuli and they suggest that this shows a preparatory baseline shift. Secondly, they report that speech and non-speech stimuli elicited different effects regarding the M100 with speech stimuli showing a hypoadditive enhancement. Lastly, the late mismatch field (MMF) showed a speech-specific lateralization. The conclusion is that the two sensory channels convert the phonetic information into a phonological feature code and only then these separate pieces of information are fused into a uniform phonetic percept.

The results are based upon a thorough consideration of the possibly optimal stimulus settings. They seem to be able to cover most bases and their arguments appear solid. Our laboratory has traditionally focused on auditory processing only, but the possibilities offered by audiovisual research seem tempting in the manner in which they give a more complex view to speech processing as a whole.

### 3.3. Eulitz: Underspecified representations

The commentary on the contribution by Eulitz is based upon a brief personal communication and his earlier work with Lahiri [3].

The access to abstract representations is one of the goals in linguistically oriented cognitive brain research. If it is possible to locate an abstract representation of e.g. feature specification of vowels or some phonological rules, it would mean that the brain does not store all acoustic characteristics of speech sounds, but rather that the representations carry only the relevant information. Eulitz and Lahiri [3] were able to show that vowels are not fully specified in the mental lexicon. The experimental setting was elaborate in its stimulus description. The fact that the isolated vowel stimuli contained acoustic variability within a category would be abstract enough for many approaches, since it would demonstrate that speech sounds function categorically. However, the feature description of the vowel stimuli is the key to the abstract level, and by reversing the stimulus order, the authors were able to show that the MMN response was larger and earlier when there was a conflict between the surface form and the underlying representation while the acoustic distance between the stimulus pairs was the same.

Altogether, the findings support the view that not all features are stored in the mental lexicon. It seems that it is possible to access something truly abstract when the experimental design is detailed enough. In a sense, this is connected with our ongoing study [1] where we show that the role of phonological vowel harmony rule (which is found in Finnish but not in Estonian) can be seen in the MMN responses. This study also appears to access a more abstract level of representation.

## 4. CONCLUSIONS

Previous findings have established the role of the native language speech sound system in the perception of native and non-native sounds, and it is also apparent that these maternal categories are so hard-wired that the perception does not demand conscious effort. More recent studies have also hinted towards the possibility that there are abstract representations for sound categories. This abstractness may indirectly be seen in contrastive experiments, where the same acoustic stimuli result in large or small MMN responses depending on whether or not the contrast is phonemic in the native language. The abstractness of the representation has recently been studied more directly and the results are encouraging. It will be interesting to see, whether more extensive research on the neural correlates for abstract phonemic rules will further strengthen the arguments in favor of the underspecified mental representations.

Lastly, it should be noted that one of the most crucial things to bear in mind both when designing an experiment and in reading reports is connected with the selection of stimuli. The importance of the choice of the stimulus set cannot be over-emphasized, since even apparently significant findings may be distorted by a restricted knowledge of the exact phonological role of the stimuli, or as our previous findings suggest [9,11], it is possible to obtain conflicting results by using different kinds of stimuli.

## 5. REFERENCES


[1] Aaltonen, O., Peltola, M.S., Savela, J., Tamminen, H., Lehtola, H. *In preparation.*

[2] Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., Näätänen, R. 2002. Mismatch negativity shows that 3-6-year-old children can learn to discriminate non-native sounds within two months. *Neuroscience Letters* 325, 187-190.

[3] Eulitz, C., Lahiri, A. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neuros.* 16, 577-583.

[4] Hertrich, I., Lutzenberger, W., Ackermann, H. 2007. Phonological aspects of audiovisual speech perception. This volume.

[5] Idsardi, W.J. 2007. Some MEG correlates for distinctive features. This volume.

[6] Kazanina, N., Phillips, C., Idsardi, W. 2006. The influence of meaning on the perception of sounds. *Proc. Nat. Acad. Sci.* 103, 11381-6.

[7] Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J. Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432-434.

[8] Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R. 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters* 352, 25-28.

[9] Peltola, M.S., Kuntola, M., Tamminen, H., Hämäläinen, H., Aaltonen, O. 2005. Early exposure to a non-native language alters prattentive vowel discrimination. *Neuroscience Letters* 388, 121-125.

[10] Peltola, M.S., Tuomainen, O., Aaltonen, O. 2007. The effect of language immersion education on the preattentive perception of native and non-native vowel contrasts. *Journal of Psycholinguistic Research* 36, 15-23.

[11] Phillips, C. 2001. Levels of representation in the electrophysiology of speech perception. *Cognitive Science* 25, 711-731.

[12] Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R.J., Näätänen, R. 2002. Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Neuroreport* 13, 1813-1816.

[13] Winkler, I., Kujala, T., Tiitinen, H., Sivonen, p., Alku, P., Lehtokoski, A., Cziegler, I., Csepe, V., Näätänen, R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638-642.