Phonemic and Phonetic Representation in Auditory Prediction

Speech sounds can be represented at several levels of abstractness, including high fidelity acoustic representations, gradient but language-modulated phonetic representations, and purely categorical phonemic representations. Each of these levels of representation has a use in the speech perception stream, and some levels are necessary precursors to others – acoustic precedes phonetic, phonetic precedes phonemic.

The current study focuses on the nature of representations generated in response to varying input, using measures of perceptual encoding and prediction to determine their content. By measuring involuntary brain responses to variations in the acoustic environment, we can determine what information is retained over time, as new exemplars are encountered which vary from those previously encountered. We used EEG to measure the Mismatch Negativity (MMN) brain response in a varying standards oddball paradigm. We predict modulations in MMN amplitude to correspond to the presence of phonetic information in the memory trace.

Mismatch Negativity gives us an insight into both perceptual encoding and predictive mechanisms in the auditory perceptual system (Näätänen, 1990). Previous studies (e.g. Phillips et al., 2000) have claimed that varying a phonetic parameter such as Voice Onset Time (VOT) of the standards in an oddball paradigm forces the auditory system to use phonemic representations in order to make predictions about incoming sounds. Phillips et al. found no MMN when the standards varied across the perceptual boundary separating voiced from voiceless. They argue that this prevents category assignment, because some standards fall into the voiceless category, and others the voiced category. In a previous study, we tested this claim by measuring MMNs to an across-category contrast with a phonetic distance manipulation – the two conditions differed with respect to the distance between standards and deviant. However, no difference in MMN amplitude was observed, supporting the phonemic memory trace account.

This study conducts a further test of these claims by presenting a within-category contrast with a phonetic distance manipulation. The reason for this design is twofold. First, the contribution of phonemic category assignment to MMN amplitude is quite large (see e.g. Sharma & Dorman, 1999), so the across-category contrast may have obscured the phonetic distance effect. Second, the within-category contrast necessitates a phonetically-detailed memory trace for any MMN to be elicited. If the standards are represented only with a phonemic category representation, there will be no clash with the deviant, thus, no MMN. The presence of an MMN will indicate the presence of phonetic information in the memory trace. The presence of a distance effect will mean that the phonetic information in the memory trace corresponds directly to the phonetic information of the standards.

Design and stimuli

The stimuli were synthesized CV syllables consisting of a coronal stop and low front vowel. There were two conditions: High Standards, and Low Standards. The High Standards condition consisted of 110, 115, and 120ms VOT standards. The Low Standards condition consisted of 95, 100, and 105ms VOT standards. The deviant in both conditions had a VOT of 50ms. Each experimental block consisted of 900 standard tokens and 100 deviant tokens. Participants watched a movie while passively listening to sounds.

Categorization task

Participants completed an offline categorization task to determine their perceptual boundary separating voiced and voiceless both before and after the EEG recording session. This was done to ensure that the perceptual boundary for each subject did not change after exposure to very

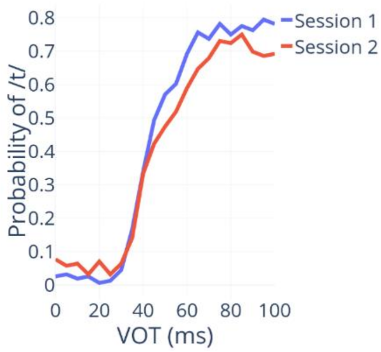

Figure 1. Categorization results.

high VOT stimuli (see e.g. Eimas & Corbit, 1973). A threshold analysis was applied to the categorization data to determine the perceptual boundary between voiced and voiceless for each participant. A logistic regression was built for each participant, and the model generated a VOT threshold value, defined by the interpolated VOT value when the participant would respond with exactly 50% that the sound belonged to the /t/ category. We find a very high average threshold (pre: 65.93ms, SD = 53.25; post: 73.2ms, SD = 53.37), but a t-test shows no significant change after the EEG recording session (see Figure 1).

MMN results

We find a significant MMN to the within-category deviant in both conditions (see Figure 2), but the amplitude of the effect is not modulated by phonetic distance between standards and the deviant. This indicates that the memory trace generated in response to the varying standards did contain phonetic information – but this phonetic information did not derive from the standards themselves. We conclude that the representation is indeed phonemic, but that the auditory system uses the phonemic-phonetic mapping to make a phonetically-detailed sensory prediction about upcoming sounds. The auditory predictive system relies on known categories when it encounters variance, and uses known mappings between phonemic categories and phonetic realizations to make detailed sensory predictions.

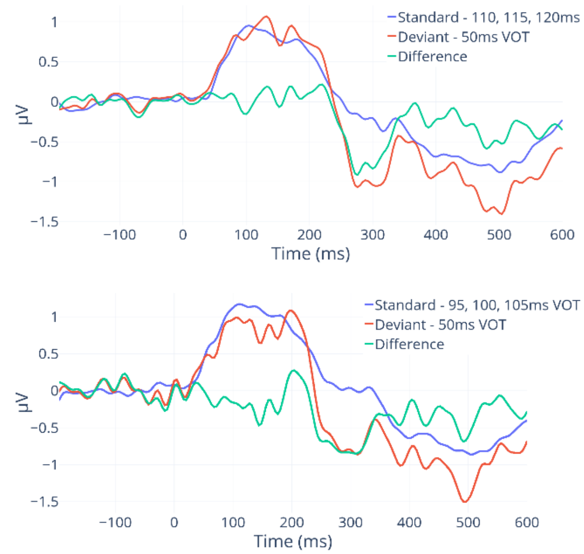

Figure 2. Standard, deviant, and difference waveforms in the High and Low conditions.

References

Eimas, P. D., & Corbit, J. D. (1973). Selective Adaptation of Linguistic Feature Detectors. *Cognitive Psychology*, (4), 99–109.

Näätänen, R. (1990). The role of attention in auditory by event-related potentials and other brain measures of cognitive function, *13*, 201–288.

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://doi.org/10.1162/08989290051137567

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078–83. http://doi.org/10.1121/1.428048