Listeners detect phonological features directly from the auditory signal: the perceptual basis of vowel height

Kateřina Chládková[1], Titia Benders[2], Paul Boersma[1]
[1]University of Amsterdam, [2]Radboud University Nijmegen

Since the 1950's, phonological theory has described the sound patterns of the world's languages in terms of distinctive features (Jakobson et al., 1952). Distinctive features are abstract phonological representations that were named after observable phonetic properties of speech sounds. For instance, the feature vowel height corresponds to the first formant dimension (F1). Accordingly, a language that uses F1 to contrast some of its vowels phonetically, is described as having the vowel height feature in its phonology. Thus, a direct relation has been traditionally assumed between a phonological feature and its phonetic correlate. The question remains whether such direct relation between features and phonetics exists also in the grammars of language users (this question has been brought up already by Ladefoged, 1980). Recent behavioral and neurolinguistic perception studies, as well as computer simulations, indicate that during speech comprehension, listeners seem to map phonetic information onto phonological feature categories (e.g. Kingston, 2003; Scharinger et al., 2011; Lin & Mielke, 2008). However, most of the previous studies did not explicitly compare a feature- and a phoneme-based model of perception. Therefore, their results do not provide unequivocal evidence for phonetic bases of features.

The present study addresses the feature vs. phoneme issue directly: we test whether listeners map phonetic information directly onto features or onto phonemes. We focus on a phonological feature that occurs in most of the world's languages, namely vowel height, and its acoustic correlate, the F1. We test whether in the regions of the vowel space that are not employed by the native vowel inventory, listeners still perceive the F1 dimension in terms of their native vowel height categories.

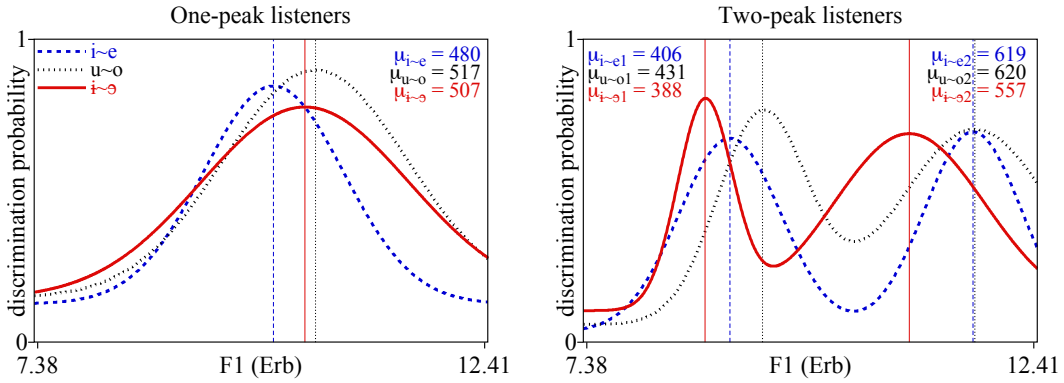
The test case here is a typical 5-vowel language (namely, Czech) that contains the phonemes /i e a o u/. This language employs a high-mid featural contrast in both front and back vowels: the /i-e/ and the /u-o/ contrast, respectively, but does not contain any high and mid vowels in the central region of the vowel space. We therefore use this phoneme-empty upper central vowel region to test whether listeners map phonetics (namely, F1) directly onto phonological features (namely, height). There are two competing predictions: (1) If listeners map the auditory F1 dimension onto features then they should generalize their phonological high–mid distinction from the front and back vowel regions to the phoneme-empty (or, uncolonized) central region. That is, even though they do not identify the F1–F2 combinations in the uncolonized region as phonemes of their language, they should still associate low F1 values with the feature high and medium F1 values with the feature mid. (2) If, on the other hand, the native speakers of our 5-vowel language map the auditory signal to phonemes and not to features, then they should not perceive the uncolonized region in terms of their native height categories.

We carried out two experiments. In Experiment 1, we determined the exact location of the uncolonized vowel region, i.e. the vowel region with which our listeners have no phonemic experience. Experiment 1 was thus a vowel identification task with stimuli sampled from the whole possible vowel space. In the upper part of the vowel space (between F1 of 260 and 725 Hz), we found that at an F2 of 1790 Hz, listeners did not consistently identify the stimuli in terms of their native vowels. We used this continuum (transcribed as [ɨ]-[ə]) in Experiment 2, which tested whether listeners perceive this continuum categorically despite the fact that they do not identify it in

terms of their phonemes. Experiment 2 comprised of three AX discrimination tasks: one with stimuli from the /i/-/e/ continuum, one with the /u/-/o/ continuum, and one with the [ɨ]-[ə] continuum. We thus compared listeners' perception of their existing front and back vowel continua to their perception of the uncolonized continuum. Note that using a discrimination task enabled us to assess categorical perception even with stimuli that are unidentifiable in terms of native categories: in line with Liberman et al. (1957), a peak in the discrimination data corresponds to a boundary between two categories.

To compare (the degree of) categorical perception across the three continua, we assessed the number of discrimination peaks (i.e. the number of perceived category boundaries), as well as the height, width, and location of the peaks (i.e. the crispness and location of the category boundaries). The results were similar across the three continua. First, perception on each continuum yielded significant discrimination peaks. Second, the number of peaks was comparable across the three continua: on each continuum half of the listeners had 1 peak, and about a third of the listeners had 2 peaks[1] (see the Figure below, which shows the discrimination functions of the 1-peak and 2-peak listeners on each continuum). Third, the peak locations, height and widths were mostly comparable across continua. In sum, the results showed that listeners perceived the [ɨ]-[ə] continuum categorically. Moreover, their perceptual categorization of the uncolonized [ɨ]-[ə] continuum resembled their categorization of the existing /i/-/e/ and /u/-/o/ continua. We therefore conclude that listeners map the auditory F1 dimension directly to the vowel height feature rather than to phonemes.

Our findings provide evidence for perceptual bases of phonological features. We argue that the phonological feature is the initial discrete representation onto which listeners map the incoming speech signal.



References: Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis: the Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

Kingston, J. (2003). Learning foreign vowels. *Lang Speech*, 46, 295–349.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language*, 56 (3), 485–502.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psych: HPP*, 54 (5), 358–368.

Lin, Y., & Mielke, J. (2008). Discovering place and manner features: what can be learned from acoustic and articulatory data. *UPenn Working Papers in Linguistics*, 14, 241–254.

Scharinger, M., Idsardi, W. J., & Poe, S. (2011). A comprehensive three-dimensional cortical map of vowel space. *J Cogn Neurosci*, 23 (12), 3972–3982.

[1] The difference in the number of peaks can be explained as follows: some listeners (i.e. those with 1 peak) perceived 2 height categories (i.e. high and mid) in the presented F1 stimulus range, while others (i.e. those with 2 peaks) attempted to fit all their 3 native height categories (i.e. high, mid and low) into that stimulus range.