THE PERCEPTUAL BASIS OF THE FEATURE VOWEL HEIGHT

Kateřina Chládková[1], Paul Boersma[1], Titia Benders[2]

[1]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands
[2]School of Psychology, Newcastle University, Newcastle, NSW, Australia
k.chladkova@uva.nl, paul.boersma@uva.nl, titia.benders@newcastle.edu.au

ABSTRACT

The present study investigated whether listeners perceptually map phonetic information to phonological feature categories or to phonemes. The test case is a phonological feature that occurs in most of the world's languages, namely vowel height, and its acoustic correlate, the first formant (F1).

We first simulated vowel discrimination in virtual listeners who perceive speech sounds through phonological features and virtual listeners who perceive through phonemes. The simulations revealed that feature listeners differed from phoneme listeners in their perceptual discrimination of F1 along a front-back boundary continuum as compared to a front (or back) continuum.

The competing predictions of phoneme-based versus feature-based vowel discrimination were explicitly tested in real human listeners. The real listeners' vowel discrimination did not resemble the simulated phoneme listeners, and was compatible with that of the simulated feature listeners. The findings suggest that humans perceive vowel F1 through phonological feature categories like /high/ and /mid/.

Keywords: speech perception, vowel discrimination, vowel height, distinctive features

1. INTRODUCTION

Distinctive features [8] are phonological representations supposedly related to observable phonetic properties of sounds: articulatory gestures, auditory cues, or both [5, 8, 22]. For instance, the feature vowel height corresponds to the first formant dimension (F1) phonetically. Accordingly, a language that uses F1 to contrast some of its vowels phonetically, is described as having the vowel height feature in its phonology.

It has long been debated whether the units through which listeners perceive speech are indeed distinctive features, or whether they are phonemes (for a review see [16]). The findings of a variety of studies suggest that listeners might perceive vowels through feature categories [7, 14, 23, 9, 2, 19, 20, 13]. To the best of our knowledge, however, no previous study tested directly the two contrasting hypotheses: whether listeners map phonetic information onto feature or onto phoneme categories. This question is explicitly addressed by the present study. We investigate whether listeners perceptually map vowels' F1 onto height categories such as mid and high, or onto unanalysed phonemes such as /e/ and /i/.

The mapping between F1 and the vowel height feature is tested in a language with a typical 5-vowel inventory /i e a o u/. In the upper central region (i.e. halfway between non-low front and back vowels), typical 5-vowel languages do not have any phonemes. When listeners with such a 5-vowel inventory are forced to label stimuli from the upper central region in terms of their native phonemes, they assimilate [ɨ]-like sounds into their native high vowel categories, /i/ or /u/, and [ɘ]-like sounds into their native mid vowel categories, /e/ or /o/ (see e.g. [18]). Experiments that test vowel identification thus suggest that, in 5-vowel languages, the upper central region of [ɨ]-like and [ɘ]-like vowels contains perceptual /i/–/u/ and /e/–/o/ boundaries.

Labelling experiments tacitly assume that listeners map sounds initially to phonemes. If listeners map sound rather to phonological features, they do not actually associate the [ɨ]-region with the /i/–/u/ boundary, or the [ɘ]-region with the /e/–/o/ boundary. Instead, the 5-vowel feature listener will initially perceive sounds from the [ɨ]-region as the feature categories /high/ and /central/ and sounds from the [ɘ]-region as the feature categories /mid/ and /central/. These phonological feature categories are unlike phonemes in that they do not have labels that would be known to an ordinary language user. For that reason, it is virtually impossible to use vowel *identification tasks* to test whether listeners map sound initially to phonemes or to features. Therefore, tasks that do not involve explicit labels are better suited for testing the nature of the initial perceptual categories; among these are vowel discrimination tasks.

2. MODELED VOWEL DISCRIMINATION

We first simulated a virtual 5-vowel listener who perceives speech sounds through phoneme categories and a listener who perceives speech sounds through feature categories. We tested how these two listeners perceptually divide the vowel space in a *discrimination task*, which does not force them to label the stimuli with the conventional phoneme labels they have established when learning the alphabet.

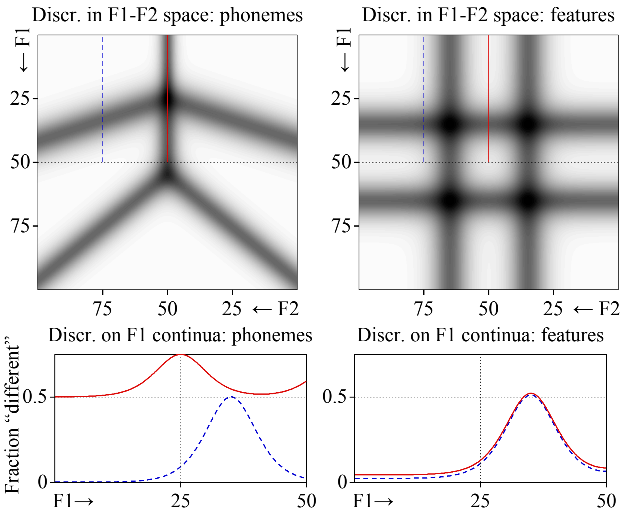
A discrimination task can reveal that listeners perceptually categorize a given auditory continuum if they report to hear a difference between sounds from some parts of the continuum but not between sounds from other parts [15]. Data obtained in a discrimination task yield a discrimination function, which is the number of ‘different’ responses as a function of the location along the stimulus continuum. A peak in the discrimination function (i.e., a larger number of ‘different’ responses in a small part of the stimulus continuum) corresponds to a boundary between two categories [12]. A valley in the discrimination function (i.e., a smaller number of ‘different’ responses in a small part of the stimulus continuum) corresponds to a centre of a perceptual category. Note that perception of stimuli that lie at phoneme boundaries tends to be associated with uncertainty [17]. Therefore, listeners are more likely to consider two *acoustically identical* stimuli to be *perceptually different* if they lie at or near their perceptual boundary than if they lie far from boundaries.

In our simulated same–different discrimination task listeners heard the same F1–F2 token twice, and had to respond whether the two stimuli were the same or different. A discrimination probability was obtained for tokens across the whole F1–F2 space. The results are plotted in Fig. 1. It is seen that a phoneme-based and a feature-based discrimination result in markedly different perceptual patterns.

As is seen in the top left graph of Fig. 1, the phoneme listener divides up the vowel space into five categories, which correspond to the five phonemes /i/, /e/, /a/, /o/, /u/. The phoneme listener has a vertical boundary in the upper central part of the vowel space, which can be interpreted as combining the /i/–/u/ and /e/–/o/ boundaries. This boundary is vertical because the categories /i/ and /e/ differ from /u/ and /o/, respectively, only in F2 but share F1. The phoneme listener also has four diagonal boundaries, which separate the phoneme pairs /i/–/e/, /e/–/a/, /a/–/o/ and /o/–/u/. These boundaries are diagonal because the two members of each phoneme contrast differ in both F1 and F2.

The feature listener, in the top right graph of Fig. 1, divides up the vowel space into nine categories, which correspond to the nine feature combinations /high front/, /mid front/, /low front/, /high central/, /mid central/, /low central/, /high back/, /mid back/, /low back/. The feature listener has two vertical boundaries along the entire F1 axis, which can be interpreted as the /front/–/central/ and the /central/–/back/ boundaries. These boundaries are vertical because all the /central/ categories differ from the respective /front/ and /back/ categories only in F2. The feature listener also has two horizontal boundaries along the entire F2 axis, which can be interpreted as the /high/–/mid/ and /mid/–/low/ boundaries. These boundaries are horizontal because all the /mid/ categories differ from the respective /high/ and /low/ categories only in F1.

Figure 1: Simulated vowel discrimination in phoneme and feature listeners. Top: discrimination in the whole F1-F2 space; darkness correlates with discrimination probability (black = 1, white = 0). Bottom: discrimination on a front, and a central F1 continuum at a front-back phoneme boundary (the continua are also highlighted in the top graphs).



The bottom graphs in Fig. 1 show that the phoneme listener, but not the feature listener, discriminates an F1 continuum in the central part of the vowel space differently than an F1 continuum in the front part of the vowel space. It is seen that on the front continuum phoneme listeners have a discrimination peak clearly separating two deep troughs, which can be interpreted as a category boundary between their phonemes /i/ and /e/. In contrast, on the central continuum phoneme listeners have a less well defined discrimination peak and less deep valleys, because the central continuum runs through their /i/–/u/ and /e/–/o/ boundaries, at which discrimination is already (by definition) high. On the other hand, feature listeners discriminate the front and the central continuum in a similar way. On both the front and the central continuum they have a discrimination peak clearly separating two deep troughs, which can be interpreted as a boundary between their feature categories /high/ and /mid/.

To sum up, the depth of discrimination valleys reflects how vowel perception differs between simulated phoneme and feature listeners. The competing predictions of a phoneme-based versus a feature-based vowel discrimination on a front and central continuum are tested here in real participants.

We report on two experiments. Exp. I is a vowel identification task that determined the location of the central boundary region, i.e. the region that in identification tasks appears to separate front and back phonemes. The predictions of the phoneme- and feature-based perception models differ crucially in how perception in this boundary region compares to perception in a front (and analogously in a back) vowel region. Exp. II consists of discrimination tasks along F1 continua in the front, back and central 'boundary' regions of the vowel space.

The listeners were native speakers of the Moravian variety of Czech (MC). MC has 5 monophthongal qualities (/i ɛ a o u/), all of which occur as phonemically short and long [21]. The vowels are defined by 3 height and 3 backness features: /i/ = /high front/, /ɛ/ = /mid front/, /a/ = low central, /o/ = /mid back/, /u/ = /high back/ [11].

3. EXPERIMENT I

Experiment I aimed to determine the identification boundary between front and back vowel phonemes.

3.1. Method

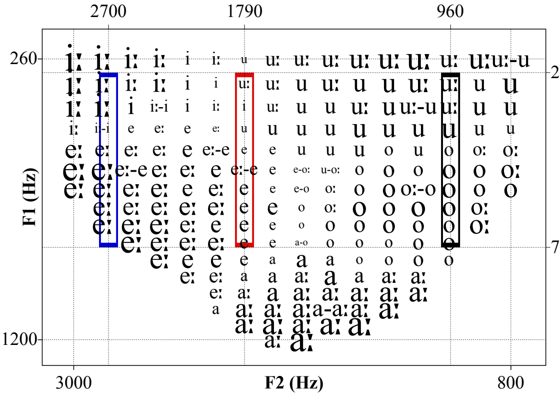
The participants were 50 monolingual native speakers of MC (aged 19–26). The stimuli were synthesized isolated vowels covering the whole vowel space. F1, ranging from 280 to 1200 Hz, and F2, ranging from 800 to 3000 Hz, were sampled in 16 auditorily equal steps. The F1–F2 grid contained 194 tokens (see Fig. 2). The tokens were synthesized with three F3 values: 2900 Hz, 3260 Hz and 3700 Hz, which yielded a total of 582 different stimuli. The duration of all tokens was 330 ms, and the fundamental frequency had a rise-fall pattern. The stimuli modelled a female voice and were made with a Klatt synthesizer [10] in the program Praat [4].

Vowel identification was tested in a multiple-forced-choice identification task with response labels in Czech orthography. The 582 stimuli were presented in random order, with a 600-ms inter-stimulus interval.

3.2. Results and Discussion

Fig. 2 plots the responses: for each F1–F2 pair, the label selected by most listeners is shown and its size reflects the proportion of listeners who chose it. Visual inspection suggests that the front–back identification boundary lies at an F2 of 1790 Hz.

Figure 2: Results of the vowel identification task.



To numerically determine the location of the front–back phoneme boundary, we submitted the data to binomial logistic regression (BiLR) analyses. First we searched for an F1 value that lies at the boundary between high and mid vowels. For each participant we fitted a BiLR on her /i/ vs. /e/ responses, and a BiLR on her /u/ vs. /o/ responses, with F1, F2 and F3 as the independent variables. From the BiLR coefficients we computed the F1 value of the /i/–/e/ and the /u/–/o/ boundaries using the formula

(1) $$x = -\frac{\beta_0+\beta_2 y+ \beta_3 z}{\beta_1}$$

where x is the F1 value of the /i/–/e/, or /u/–/o/, boundary, β_0 through β_3 are the BiLR coefficients, y is the F2 value of the /i/–/e/ or /u/–/o/ continuum (23.42 Erb for /i/–/e/, and 14.97 Erb for /u/–/o/), and z is the medium F3 value (24.97 Erb). The average high–mid boundary was found to lie at F1 = 9.832 Erb. In a second step, we searched for an F2 value that lies at the boundary between front and back vowels. For each participant, we fitted a BiLR on her /i/ vs. /u/ responses, and a BiLR on her /e/ vs. /o/ responses, with F1, F2 and F3 as the independent variables. The F2 value of the /i/–/u/ and the /e/–/o/ boundary was computed with the formula

(2) $$y = -\frac{\beta_0+\beta_1 x+ \beta_3 z}{\beta_2}$$

where y is the F1 value of the /i/–/u/ or /e/–/o/, boundary, β_0 through β_3 are the BiLR coefficients, x is the F1 value of the average high–mid boundary (9.832 Erb), and z is the medium F3 value. The average front–back boundary was found to lie at F2 = 19.542 Erb (99.9% c.i. = 19.029..20.055).

The visually observed boundary from Fig. 2 (1790 Hz = 20.005 Erb) lies within the 99.9% confidence interval of the numerically determined boundary. 20.005 Erb was thus considered as the representative F2 value of the identification boundary between front and back vowel phonemes, and was further used in Exp. II as the F2 value of the central vowel region.

4. EXPERIMENT II

Experiment II investigated listeners' perceptual categorization of an F1 continuum in the front, back and central regions of the vowel space. Perceptual categorization was tested by means of a same-different discrimination task.

Exp. II tests predictions that follow directly from our simulations (Sec. 2). If listeners map sound initially to features, we expect to find similar perceptual discrimination – namely, equally deep discrimination valleys – across the front and the back, as well as the central F1 continuum. If, on the other hand, listeners map sound initially to phonemes, we expect to find differences in discrimination – namely, differences in the depth of the discrimination valleys – between the front and back versus the central continuum.

4.1. Method

81 monolingual native speakers of MC participated (aged 18–30): 24 were tested on the front, 26 on the back, and 31 on the central continuum.

The stimuli were isolated vowels created with a synthesis procedure identical to Exp. I. Vowels were synthesized along three F1 continua: front, back, and central, which are highlighted in Fig. 2 in blue, black, and red, respectively. F1 ranged from 280 to 725 Hz. The F2 and F3 were 2700 Hz and 3300 Hz for the front 960 Hz and 2900 Hz for the back, and 1790 Hz and 3260 Hz for the central continuum. Per continuum, we synthesized 260 tokens that combined into 130 stimulus pairs. The F1 distance between the vowels within a stimulus pair was 0.9 Erb, and the distance between two neighbouring stimulus pairs was 0.039 Erb. Stimulus design followed that of [3].

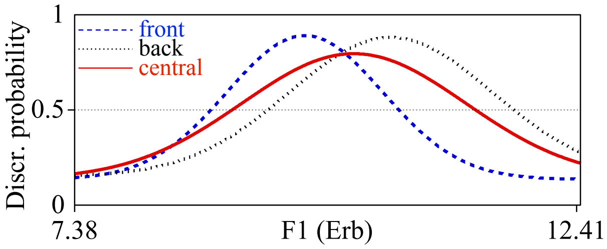
On a trial, participants heard the two sounds of a randomly chosen stimulus pair and answered whether the sounds were the same or different. The inter-stimulus interval was 500 ms. Each of the 130 stimulus pairs occurred twice.

4.2. Results and Discussion

Peaks and valleys in the discrimination function were assessed using the method described in [3]. On each of the three continua, the majority of listeners (namely, 11 on front, 14 on central, and 13 on back continuum) had one discrimination peak separating two deep valleys, which means that they discriminated two categories. The discrimination functions pooled across listeners are shown in Fig. 3.

As was shown in Fig. 1, the discrimination probability in the discrimination valleys for the modelled phoneme listener is 0.5 on the central and 0 on the front continuum: thus the difference between central and front in the phoneme listener is 0.5. In contrast, for the modelled feature listener, both the front and the central continuum have a discrimination probability in the valleys near 0: thus the difference between central and front in the feature listener is around 0.1. In order to assess whether our real listeners are compatible with the feature-listener model and/or with the phoneme-listener model, we computed from the data obtained in Exp. II the difference in discrimination probability in the valleys between front and central, and between back and central; central–front = 0.001 (95% c.i. = -0.134..0.137), central–back = -0.015 (95% c.i. = -0.150..0.120). Both these differences are very reliably different from 0.5, which means that the real listeners are significantly different from phoneme listeners in their discrimination patterns across the central versus front and back continua. The phoneme-listener model can therefore be rejected. By contrast, the conference intervals of both differences contain all values around 0.1, and are therefore compatible with the feature-listener model. We conclude that for the real listeners the phoneme-listening hypothesis can be rejected, and that the feature-listening hypothesis cannot be rejected.

Figure 3: Discrimination by real listeners.



5. CONCLUSION

Our simulations showed that a feature and a phoneme listener differ in their discrimination patterns across the front (and back) versus the central F1 continuum. Data obtained from real listeners indicate that humans do not resemble the simulated phoneme listeners, and that their behaviour is compatible with that of the simulated feature listeners. This suggests that the phonetic F1 dimension is perceptually mapped directly onto phonological feature categories.

The focus here was on one parameter that showed most obvious differences between modelled feature and phoneme listeners: the depth of discrimination valleys. Further research is needed into additional discrimination parameters such as the number, location or height of discrimination peaks.

6. REFERENCES

[1] Ashby, J., Sanders, L. D., Kingston, J. 2009. Skilled readers begin processing sub-phonemic features by 80 ms during visual word recognition: Evidence from ERPs. *Biological Psychology* 80, 84–94.

[2] Boersma, P., Chládková, K. 2011. Asymmetries between speech perception and production reveal phonological structure. *Proc. 17th ICPhS* Hong Kong, 328–331.

[3] Boersma, P., Chládková, K. 2013. Detecting categorical perception in continuous discrimination data. *Speech Commun* 55, 33–39.

[4] Boersma, P., Weenink, D. 1992–2010. Praat: doing phonetics by computer (Version 5.1.30). [Computer program], Retrieved December 15, 2009, from http://www.praat.org/.

[5] Chomsky, N., Halle, M. 1968. *Sound Pattern of English*. Cambridge, MA: MIT Press.

[6] Diehl, R. L. 1981. Feature detectors for speech: a critical reappraisal. *Psychologial Bulletin* 89, 1–18.

[7] Eimas, P. D., Corbit, J. D. 1973. Selective adaptation of linguistic feature detectors. *Cognitive Psychology* 4, 99–109.

[8] Jakobson, R., Fant, G., Halle, M. 1952. *Preliminaries to Speech Analysis: the Distinctive Features and their Correlates*. Cambridge, MA: MIT Press.

[9] Kingston, J. 2003. Learning foreign vowels. *Lang Speech* 46, 295–349.

[10] Klatt, D. H., Klatt, L. C. 1990. Analysis, synthesis and perception of voice quality variations among male and female talkers. *J Acoust Soc Am* 87, 820–856.

[11] Kučera, H. 1961. *The Phonology of Czech*. s' Gravenhage: Mouton & Co.

[12] 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 Psychol: HPP* 54 (5), 358–368.

[13] Lin, Y., Mielke, J. 2008. Discovering place and manner features: what can be learned from acoustic and articulatory data. *University of Pennsylvania Working Papers in Linguistics*, 14, 241–254.

[14] Miller, G. A., Nicely, P. E. 1955. An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 27, 338–352.

[15] Pisoni, D. B. 1973. Auditory short-term memory and vowel perception. *Memory Cognition*, 3 (1), 7–18.

[16] Pisoni, D. B., Luce, P. A. 1987. Acoustic-phonetic representations in word recognition. *Cognition* 25, 21–52.

[17] Pisoni, D. B., Tash, J. 1974. Reaction times to comparisons within and across phonetic categories. *Perc Psychophys*, 15 (2), 285–290.

[18] Savela, J. 2009. *Role of selected spectral attributes in the perception of synthetic vowels*. PhD thesis, University of Turku.

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

[20] Scharinger, M., Monahan, P. J., Idsardi, W. J. 2012. Assymetries in the processing of vowel height. *Journal of Speech, Language and Hearing Research* 55, 903–918.

[21] Šimáčková, Š., Podlipský, V. J., Chládková, K. 2012. Czech spoken in Bohemia and Moravia. *J Int Phon Assoc,* 42 (2), 225–232.

[22] Stevens, K. N. 1989. On the quantal nature of speech. *J Phon,* 17, 3–46.

[23] Studdert-Kennedy, M., Shankweiler, D. 1970. Hemispheric specialization for speech perception. *J Acoust Soc Am* 48, 579–594.