Language-specific differences in the weighting of perceptual cues for labiodentals

Silke Hamann, Paul Boersma, Małgorzata Ćavar

University of Düsseldorf, University of Amsterdam, University of Zadar
hamann@phil-fak.uni-duesseldorf.de, paul.boersma@uva.nl, mecavar@unizd.hr

ABSTRACT

Cross-language perception provides insight into the use of perceptual cues to native segments and their application to segments in a different language. In the present study we test the perception of the three Dutch labiodentals /f, v, ʋ/ by listeners of German, English, Croatian and Polish in a forced-choice identification task. We test whether the perceptual boundaries on the auditory dimensions of harmonics-to-noise ratio and duration are more similar for listeners from the same language family (German and English versus Croatian and Polish) or whether these boundaries are more similar for listeners with the same number of labial categories in their native languages (German and Croatian with four labials versus English and Polish with five). Our findings show that the same number of labial categories results in similar perceptual boundaries along the two auditory dimensions, and that language family does not influence the location of the boundaries.

Keywords: cross-language perception, perceptual cues, labiodental, fricative, approximant

1. INTRODUCTION

Adult listeners often have difficulties perceiving a category contrast in a second or foreign language (L2) that does not exist or differs from the contrasts in their native language (L1). Several factors have been held responsible for this phenomenon, among them the auditory similarity between L1 and L2 categories (e.g., Best et al. 2001), which is often determined by the number of categories on the relevant phonetic dimension. Little attention has been paid to the role of language family in L2 perception.

In this study we investigate the question whether a genetic relationship between L1 and L2 influences L2 speech perception, or whether the perception depends solely on the number of phonological categories. We test the perception of the Dutch labiodentals /f, v, ʋ/ by native listeners of four languages from two language families: two Germanic languages, German and English, and two Slavic languages, Polish and Croatian. All four languages have the two labiodentals /f, v/ and the two bilabial plosives /p, b/. They differ, however, with respect to the presence of the labiovelar approximant /w/: while this sound is absent from German and Croatian, it is present in English and Polish. The labiodentals and further labials in the inventories of these four L1 languages and the L2 language Dutch are summarized in Table 1.

Table 1: Nonnasal labial consonants of the five languages investigated in the present study.

Language family	Language	Labiodentals	Other labials
Germanic	Dutch	f, v, ʋ	p, b
	German	f, v	p, b
	English	f, v	p, b, w
Slavic	Croatian	f, v	p, b
	Polish	f, v	p, b, w

This choice of languages allows us to make a preliminary comparison between the influence of the number of labial categories (four versus five) and the influence of language family (Germanic versus Slavic) on the perceptual cue weighting for labiodentals. We expect the influence of the labial inventory to be more important than the influence of language family, i.e. listeners of languages with a similar inventory (German and Croatian versus English and Polish) to behave more similarly than listeners of languages from the same language family (i.e. German and English versus Croatian and Polish) in the perception of the Dutch labiodentals. This expectation is based on the findings in the study by Boersma & Hamann (2008), where it was illustrated with computer simulations of diachronic data that any phoneme inventory with the same number of categories (one to five) along one auditory dimension must end up with the same stable system, i.e. with the same location of category boundaries on this dimension. Though Boersma & Hamann's study only looked at sibilant inventories along the auditory dimension of spectral mean, we transfer those findings in the present study to labiodental inventories that can be characterized by multiple auditory dimensions.

Acoustically, labiodentals are characterized by low-amplitude friction noise that is spread over the frequency range in a spectrum. The perception of this friction noise is influenced by the presence or absence of periodicity caused by vocal fold vibration. The more voiceless a sound is, the more fricated it sounds, and the more voiced it is, the more periodic its frication is and the less fricated it sounds. This perceptual correlation can be captured by the acoustic measure of harmonics-to-noise ratio (Yumoto et al. 1982, Boersma 1993). Hamann & Sennema (2005a) illustrate that the harmonics-to-noise ratio clearly distinguishes between the three Dutch labiodentals and between the two labiodentals in German. Their study further shows that Dutch and German /f/ have almost identical harmonics-to-noise ratios (around −1.5 dB), but German /v/ has a much higher ratio than Dutch /v/ (German 15.3 dB, Dutch 0.8 dB), coming close to the ratio for Dutch /ʋ/ (18.8 dB). In a perception study, Hamann & Sennema (2005b) find that German naïve listeners perceive Dutch /ʋ/ as their native /v/ in all of the cases, and Dutch /v/ as their /f/ in a considerable number of cases. In a perception experiment on the boundary differences between Dutch and German labiodentals along the dimension of harmonics-to-noise ratio, Hamann et al. (2007) find that the location of the perceptual boundary between the labiodentals /f/ and /v/ in Dutch and German differs. This leads us to use the harmonics-to-noise ratio as one of the auditory dimensions in the present study.

A further acoustic and auditory difference between fricatives and approximants, and between voiced and voiceless fricatives, is duration: fricatives are longer than approximants (e.g. Romero Gallego 1995 for Spanish) and voiceless fricatives are longer than voiced ones (e.g. Stevens et al. 1992 for English; Mees and Collins 1982 for Dutch; Jessen 1998 for German; Hamann & Sennema 2005a for the Dutch labiodentals).

Though formant transitions of the preceding and following vowels and the intensity of the consonant yield further possible auditory cues for the distinction of the three labiodentals, we concentrate in the present study on the two dimensions of harmonics-to-noise ratio and duration.

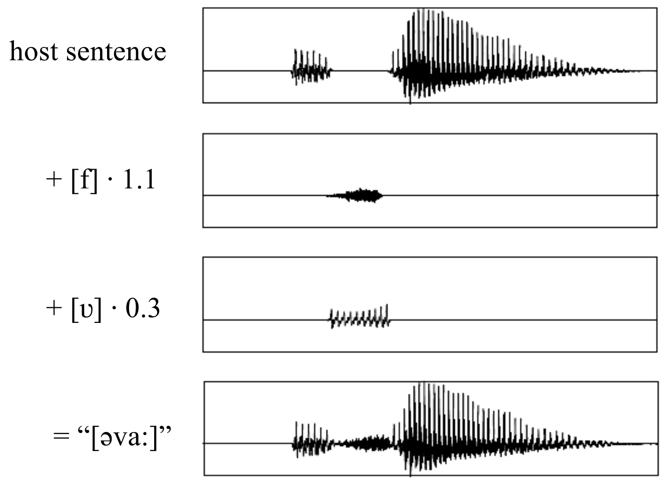
2. METHOD

2.1. Stimuli

The stimuli were synthesized on the basis of natural speech recordings from one male Dutch native speaker from the South of the Netherlands. Recordings were made in a sound-proof booth to a Pioneer PDR-555 CD recorder, using a Sennheiser MKH-105 microphone. The recordings included one token each of the three Dutch labiodentals /f, v, ʋ/ in a VCV context with a preceding [ə] and a following [aː] (from the sentence "Hoor je _a", *Do you hear _a*), with stress on the last vowel.

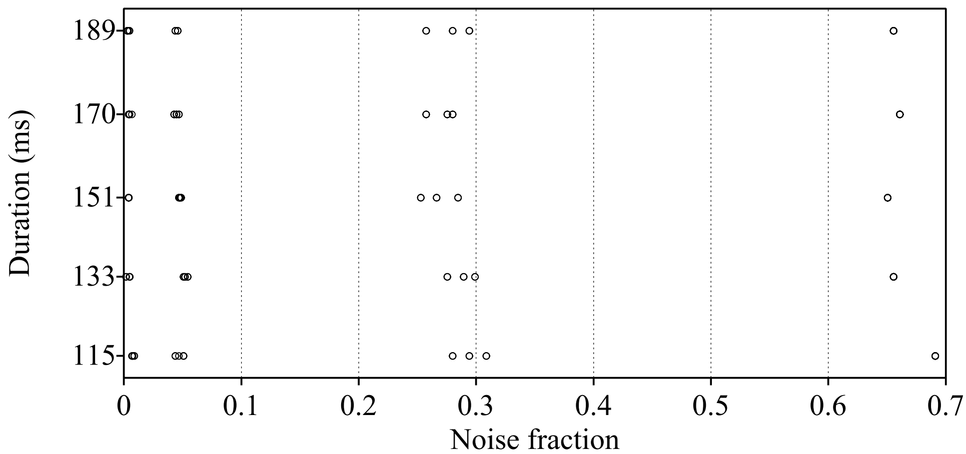
We manipulated the acoustic parameters of harmonics-to-noise ratio, duration, intensity and transition with the help of the Praat program (Boersma & Weenink 2009) in the following way. We cut the two labiodental tokens [f] and [ʋ] from their surrounding. The [f] was multiplied by the factor 0, 0.005, 0.25, 0.5, 0.75, 0.9, 1, 1.1, 1.8, 2.3, or 3. The [ʋ] was multiplied by the factor 0, 0.22, 0.3, 0.6, 0.7, 1, or 1.1. The two resulting sounds were added together and added to a host sentence, which we had created by using a recording of [əʋaː] and replacing the [ʋ] by silence. An illustration of this manipulation process is given in Figure 1.

Figure 1: Example of stimulus creation.



The stimuli resulting from this manipulation have the following values for harmonics-to-noise ratio: around −3.5 dB (close to a natural [f]), +3.5 dB (close to the natural recording of [v]), +13 dB, +22 dB (close to natural [ʋ]) and no ratio (close to natural [p]). Figure 2 shows how these values are dispersed in a two-dimensional plane. To achieve a moderate degree of perceptual equidistance the logarithmic harmonics-to-noise ratio in dB was first converted to its nonlogarithmic counterpart (HNR), then to a *noise fraction* between 0 and 1 (= HNR / (HNR + 1)).

Figure 2: Distribution of stimuli on the acoustic dimension of noise fraction (horizontal) and duration (vertical).



The duration of the stimuli was manipulated by lengthening the [ʋ] (by duplicating part of the original signal), shortening the [f] (by removing part of the original signal), and adjusting the duration of the silence phase in the frame sentence. Resulting duration values of the stimuli are 115 ms (the duration of the natural [ʋ]), 133 ms (natural [v]), 151 ms, 170 ms, and 189 ms (natural [f]).

The intensity of the stimuli was manipulated identical to the harmonics-to-noise ratio (the two could not be manipulated independently from each other). Intensity values of the stimuli are: 45 dB (identical to the natural recording of [f]), 50 dB (natural [v]), and 55 dB (natural [ʋ]).

The transitions were manipulated by removing two glottal waves from the transitions in the host sentence, yielding two transitional values: long transitions (from the recording of [əʋa:]) and short transitions (with glottal waves removed, mimicking an [f]-context).

The combination of all these parameter values results in 120 stimuli (four for voicing-to-friction ratio × five for duration × three for intensity × two for transition). Approximately 18 of these 120 stimuli do not sound like a labiodental fricative or approximant (all of them have a factor of 0.005 for [f] and a low factor for [ʋ]) but rather like a bilabial plosive [b] or [p]. This led us to include the bilabial plosives as possible answer categories (see 2.3 below). To avoid a range effect, we added 10 further stimuli with silence in the host sentence (five durational values × 2 transitions). This results in a total of 130 stimuli.

In the following we only report on the parameters of noise fraction and duration.

2.2. Listeners

The listeners of this experiment were a total of 94 participants with either German, English, Croatian or Polish as their native language. We tested 31 German listeners at the University of Potsdam. They were 20–41 years old. The English group consisted of 20 participants, 18–47 years of age, tested at University College London. The Polish group consisted of 23 participants, 20–36 years of age, and tested at the University of Warsaw. The Croatian group consisted of 20 listeners between 19 and 29 years of age, who were tested at the University of Zadar. The participants were mostly university students, though some were faculty members. No participants had lived outside of their country for longer than six months and none of them reported any hearing impairment.

2.3. Task

The task was a forced-choice identification task. Each of the 130 stimuli was repeated once, giving a total of 260 stimuli. This total set was randomized for each listener and presented via headphones. The set of answers included the labiodental fricatives (voiced and voiceless) and the labial plosives for all groups. The English and Polish groups had in addition the labiovelar approximant. German and Croatian listeners thus had four answer categories, the English and Polish listeners five. Orthographic representations of these labials in the native language of the listeners were presented on a computer screen. These answer categories are given in Table 2.

Table 2: Answer categories (in orthographic representation) for the five language groups. German and Polish <w> is /v/, and Polish <ł> is /w/.

Language	Answer categories
German	f w p b
English	f v p b w
Croatian	f v p b
Polish	f w p b ł

Participants heard one stimulus at a time and had to click on the consonant they thought they had heard. They could not listen to the stimulus a second time.

3. RESULTS

Of the 94 listeners we remove two outliers, one English and one Polish native speaker. These are the listeners who show a great lack of consistency in their answers: for more than 50 percent of the 120 target stimuli they gave a different response to the two replications.

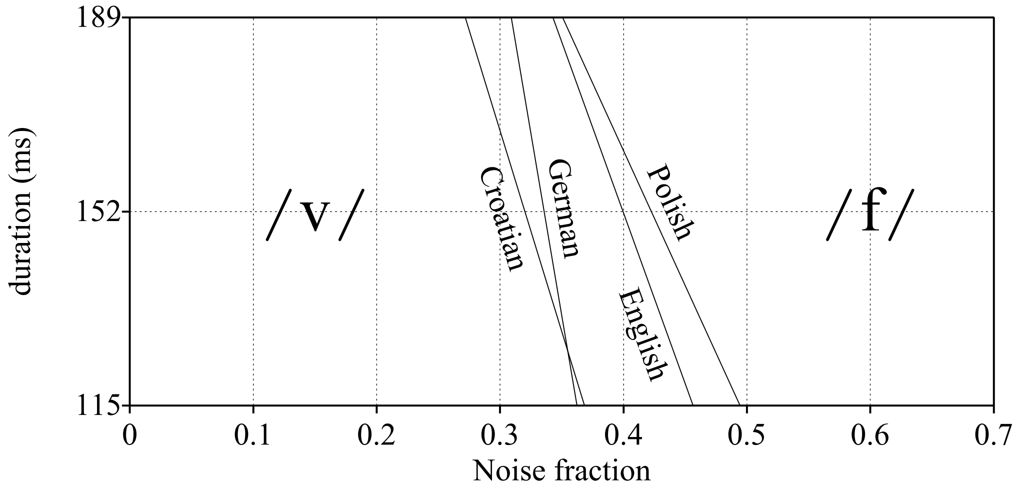
On the basis of the results of the identification test we perform a logistic regression analysis on the responses of each of the 92 speakers, with noise fraction and duration as the factors and the /f/–/v/ choice as the dependent variable. From the logistic regression coefficients of a speaker we define that speaker's *boundary location for noise fraction* as - (*intercept* + *duration coefficient* × 151 ms) / *noise fraction coefficient*, and that speaker's *angle between noise fraction and duration* as arctan2 (- *noise fraction coefficient*, *duration coefficient*). The four /f/–/v/ *boundary lines* in Figure 3 reflect the medians (over the speakers) of these two quantities for the four languages, as follows.

The boundary location for noise fraction is shown in Figure 3 as the noise fraction value where a boundary line intersects the horizontal line at the middle duration of 151 ms: for Croatian listeners it lies at 0.320, for German listeners at 0.336, for English listeners at 0.400, and for Polish listeners at 0.423. A Kruskal-Wallis test reveals that these boundary locations are not the same for all four languages ($\chi^2 = 25.663$, $df = 3$, $p = 0.000011$); in fact, the four languages seem to divide into two groups, namely Croatian and German on the one hand, and English and Polish on the other: the difference between the boundary locations for Croatian and English, as measured by Wilcoxon's rank sum test, is significant ($p = 0.013$), and so are the differences between Croatian and Polish ($p = 0.010$), German and English ($p = 0.004$), and German and

Polish ($p = 0.0009$), while the differences in boundary location between Croatian and German listeners and between English and Polish listeners are entirely nonsignificant ($p = 0.685$ and 0.418, respectively).

The angle between noise fraction and duration is shown in Figure 3 as the slope of the boundary line: for Polish listeners the slope is -514 ms, for English listeners -651 ms, for Croatian listeners -767 ms, and for German listeners -1390 ms (the steepest). The difference in angles is quite significant between Poles and Germans ($p = 0.003$), and less significant between Croatians and Poles ($p = 0.044$) and between Croatians and Germans ($p = 0.056$).

Figure 3: Boundaries between labiodental categories in German, English, Croatian and Polish.



As the angles of the boundaries in Figure 3 indicate, all four languages use noise fraction as a major cue and duration as a secondary cue to distinguish between /f/ and /v/: noisier and longer stimuli are more likely to be identified as /f/.

4. DISCUSSION AND CONCLUSIONS

The present study suggests that the size of the native labial inventory influences the perception of Dutch labiodental sounds to a greater extent than the genetic relationships between languages do. These findings are in accordance with the hypothesis that the size of the inventory determines the location of the category boundary between /f/ and /v/. This hypothesis was based on Boersma & Hamann's (2008) results for sibilants, which were differentiated on a single acoustic/auditory dimension. In the present study, at least two dimensions seem to play a role in the perception of labiodentals, namely noise fraction and duration. For the listeners in the four languages we tested here, the dimension of noise fraction was a major cue, and duration only a minor cue. This is probably due to the fact that noise fraction is a more salient and static (i.e. non-durational) cue. We therefore expect a similar perceptual preference for noise fraction in other languages as well. For segmental contrasts with several auditory dimensions and no phonetically-based preference for one of these dimensions, we expect languages to differ much more in their cue weighting and in the location of their perceptual boundaries, even if the segmental inventories in these languages have the same size. These expectations have to be tested in future studies.

Four languages from two language families, as well as one particular type of contrast, are of course not sufficient to make larger generalizations on the influence of language family on the perception of L2 sounds. Extensive future studies are required to support the conclusions of the present study.

5. ACKNOWLEDGEMENTS

This research was supported by NWO grant No. 016-064-057 to Silke Hamann, NWO grant No. 016-094-610 to Paul Boersma, and grant No. 269-2120920-0896 by the Croatian Ministry of Science and Education to Małgorzata Ćavar. We are thankful to Olga Głębicka, Paul Iverson, Melanie Preece-Pinet, Paweł Rydzewski, Anke Sennema, and Kordula Witjes for their support and/or help in testing the subjects.

6. REFERENCES

Best, C. T., McRoberts, G. W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109. 775–794.

Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences (University of Amsterdam*) 17. 97–110.

Boersma, P., Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270.

Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer (version 5.1.04) [Computer program].

Hamann, S., Sennema, A. 2005a. Acoustic differences between German and Dutch labiodentals. *ZAS Working Papers in Linguistics* 42. 33–41.

Hamann, S., Sennema, A. 2005b. Voiced labiodental fricatives or glides — all the same to Germans? In: Hazan, V., Iverson, P. (eds.) *Proceedings of the Conference on Plasticity in Speech Processing*. London: UCL. 164–167.

Hamann, S., Sennema, A., Boersma, P. 2007. Dutch labiodental perception in German L2 learning. Talk presented at *New Sounds 2007*, Florianópolis.

Jessen, M. 1998. *Phonetics and phonology of tense and lax obstruents in German.* Amsterdam: Benjamins.

Mees, I. & Collins, B. 1982. A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association* 12(1). 2–12.

Romero Gallego, J. 1995. Gestural organization in Spanish. An experimental study of spirantization and aspiration. PhD. diss. University of Connecticut.

Stevens, K, Blumstein, S., Burton, M., Kurowski, K. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustic Society of America* 91. 2979–3000.

Yumoto, E., Gould, W.J., Baer, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustic Society of America* 71. 1544–1550.