

Variation in the perception of an L2 contrast: A combined phonetic and phonological account¹

Silke Hamann

The present study argues that variation across listeners in the perception of a non-native contrast is due to two factors: the listener-specific weighting of auditory dimensions and the listener-specific construction of new segmental representations. The interaction of both factors is shown to take place in the perception grammar, which can be modelled within an OT framework. These points are illustrated with the acquisition of the Dutch three-member labiodental contrast [v v f] by German learners of Dutch, focussing on four types of learners from the perception study by Hamann and Sennema (2005a).

1. Introduction

Dutch has a cross-linguistically very unusual labiodental contrast between an approximant /v/, a voiced fricative /v/, and a voiceless fricative /f/ (see e.g. Booij 1995 and Gussenhoven 1999). Minimal triplets illustrating the contrast are given in (1).

(1)

| | | | | |
|------|------|------|--------------------|------------------------------|
| wee | vee | fee | [ve:, ve:, fe:] | ‘contraction, cattle, fairy’ |
| wijj | vijl | feil | [vɛil, vɛil, feil] | ‘while, rasp, error’ |

The occurrence of this contrast is restricted to word-initial position between voiced segments. Many speakers of Standard Dutch, especially those from the western part of the Netherlands, neutralize the /f/ - /v/ contrast (Booij 1995: 7f.). According to Mees and Collins (1982: 5), no neutralization occurs in careful speech and in more formal registers. Very Southern Dutch and Flemish speakers realise the approximant as bilabial (Booij 1995: 8), which can be transcribed as /β/ (see Gussenhoven 1999: 75).² This regional variance is not further considered in the following study.

2 *Silke Hamann*

German learners of Dutch have problems acquiring the three-member contrast since their native language differentiates only a voiced and a voiceless labiodental fricative (Kohler 1999; Wiese 1996), see the minimal pairs in (2).³

(2)

| | | | |
|------|------|--------------|--------------|
| Wein | fein | [vain, fain] | ‘wine, fine’ |
| Wort | fort | [vɔət, fɔət] | ‘word, away’ |

In the phonetic literature on German, the voiced labiodental sound is usually described as a fricative (see e.g. Jessen 1998, Kohler 1999, Wängler 1974). However, Kohler (1995: 154) mentions that German /v/ can turn into an approximant, especially in initial position, and Viëtor (1897: 229, 231) writes that friction in German /v/ is very little compared to English and French /v/.

Several studies on second language (L2) acquisition have shown that the existence of the same phonological category in the first language (L1) does not necessarily result in perfect L2 performance (see e.g. Best and Strange 1992; Flege 2002; Iverson et al. 2003). Instead, the phonetic realisations of the respective categories play a major role in determining the categorization. Hamann and Sennema (2005a) add to this evidence with a categorisation experiment that tested the perception of the Dutch labiodental contrast by German learners of Dutch. They found that the German learners had problems perceiving the Dutch /v/ correctly, despite the fact that German has the same phonological category.

The present article gives an account of the data acquired in Hamann and Sennema’s study, describing the performance of four types of speakers. The data is used to illustrate that L2 perception has to be accounted for with the fine-grained auditory differences between L1 and L2 segments *and* the influence of the native phonology (see e.g. Flege and Hillenbrand 1986). This interaction of phonetics and phonology can be modelled with language-specific perception grammars, which map auditory forms onto surface phonological forms. Proposals for such perception grammars have been made by Boersma (1998 et seq.) and Pater (2004) in the framework of Optimality Theory (McCarthy and Prince 1993, Prince and Smolensky 1993; henceforth: OT). In the present article, the perception grammars of the L1 and L2 native speakers and of the L2 learners are formalised. Furthermore, the article illustrates that the differences between the native and the L2 perception grammar allow different acquisition strategies, mirrored in the varied perceptual performance of

L2 learners. Thus, variation in the perception of an L2 contrast is accounted for by differences in the L2 perception grammars.

The article is structured as follows. Section 2 presents data on the performances by four groups of Dutch learners from Hamann and Sennema's perception study. In section 3, acoustic differences between the labiodentals in Dutch and German are discussed. Section 4 formalises the perception grammar of Dutch and German native listeners, and section 5 the perception grammars of the four types of learners of Dutch, illustrating different acquisition strategies. Section 6 summarises and concludes.

2. Variation in the perception of a second language: four types of learners

In a categorisation experiment, Hamann and Sennema (2005a) tested the perception of the Dutch labiodental contrast by three groups: six German listeners without any knowledge of Dutch (German L1), twenty-one German learners of Dutch (Dutch L2), and six Dutch native listeners (Dutch L1). The test materials did not only contain the three labiodentals but all Dutch obstruents /p, b, t, d, k, x, f, v, v, s, z, ç/ (see Mees and Collins 1982).⁴ These consonants were followed by the vowel /a/ and were read eight times by a male speaker in the sentence "Hoor je __", 'Do you hear __'. The token sentences were repeated four times and the total set of 384 stimuli sentences were randomized. The participants were presented with one stimuli sentence at a time via headphones and had to click on orthographic representations of the consonant-vowel sequences. With respect to the labiodentals, the German L1 listeners had to classify the three Dutch labiodental sounds as one of their two native German sounds /f/ and /v/ or any other of the ten obstruents /p, b, t, d, k, g, s, z, ç, ʃ/ in German (Kohler 1999). The Dutch native and L2 listeners had as answer categories the 12 Dutch consonants /p, b, t, d, k, x, f, v, v, s, z, ç/. The categorisation results for the labiodentals, split by the three participant groups, are given in table 1 on the next page.

Based on the phonological descriptions of the categories alone, we would expect the German listeners to equate the Dutch /f/ with their /f/, the Dutch /v/ with their /v/, thus to perform well for these two known categories, and to show problems with Dutch /v/, since German does not have an equivalent category for this sound. The results of the experiment in table 1 illustrate

4 Silke Hamann

Table 1: Mean identification scores (percent correct) of the three test groups for the labiodentals in the perception experiment by Hamann and Sennema (2005a), with stimuli in rows, and responses sorted by language group in columns. The numbers in each row per language group do not add up to 100 percent, because miscategorisations involving non-labiodental sounds are not included.

| | | German L1 | | Dutch L2 | | | Dutch L1 | | |
|----------|-----|-----------|------|----------|------|------|----------|------|------|
| | | /f/ | /v/ | /f/ | /v/ | /v/ | /f/ | /v/ | /v/ |
| stimulus | /f/ | 99.5 | 0 | 79.0 | 17.7 | 2.1 | 94.8 | 5.2 | 0 |
| | /v/ | 16.7 | 82.8 | 5.2 | 74.6 | 18.5 | 5.2 | 94.8 | 0 |
| | /v/ | 0 | 99.5 | 0.1 | 6.1 | 92.6 | 0 | 0 | 99.5 |

that the categorization of the Dutch sounds by German listeners departs from these expectations.

German native listeners, without knowledge of Dutch, perceived the Dutch voiced fricative as their voiced fricative in 82.8 percent of the cases, and the Dutch labiodental approximant /v/ as their voiced fricative /v/ in almost all of the cases, converse to the phoneme-based prediction. Unsurprisingly, Dutch /f/ was perceived as German /f/ in 99.5 percent of the cases. The *German learners of Dutch* were also not in line with the expectations. They succeeded in categorising Dutch /v/ correctly, but miscategorised /f/ in 21 percent of the cases and /v/ in 25.4 percent of the cases.

The performance of the listeners in this study varied depending on the speakers and on their level of proficiency. Whereas certain mis-categorisations almost never occurred (e.g. the categorisation of /f/ as /v/ and reverse), others were made by a large number of participants (e.g. confusing /v/ and /v/). A detailed examination of the data showed that recurring patterns in the performance of the participants allow us to group them into four types of learners (miscategorisations below 10 percent are not taken into consideration).⁵

The first type of learners, termed Learner A in the following, does not discriminate Dutch /v/ and /v/, but categorises tokens of both types of sound as approximant, see table 2. Though this pattern could only be observed for two of the twenty-one participants (PP5 and PP11)⁶, we will see below that it is of great interest for modelling the learning process because it shows the application of a German perception grammar to the Dutch sounds, i.e. the initial stage of L2 learning. Three participants (PP1, PP20, PP21) are summarized here as learner type B because they show almost identical patterns of

confusing /f/ and /v/.⁷ A similar mis-categorisation of /f/ as /v/ could also be observed for seven other participants (PP7, PP11, PP12, PP14, PP15, PP17 and PP18). A third group of participants (PP3, PP6, PP9) confused /v/ and /v/. This type of learning is summarized as Learner C. Again, we find part of their miscategorisation patterns in other participants: the misperception of /v/ as /v/ by nine participants (PP3, PP5-PP9, PP13, PP15 and PP17), and the misperception of /v/ as /v/ by two participants (PP8 and PP14). Five participants (PP2, PP4, PP10, PP16 and PP19) performed native-like in making almost no mistakes in the categorisation of all three labiodentals. These participants are represented by learner type D. A correct identification of /f/ could be observed for further five participants (PP3, PP5, PP6, PP9 and PP13), a correct identification of /v/ for three (PP12, PP14 and PP18), and a correct identification of /v/ for ten participants (PP5, PP7, PP11-PP13, PP15, PP17, PP18, PP20 and PP21).

These four types of learners represent 13 of the 21 participants. The remaining eight listeners are covered by combinations of two learning types, with one exception: participant PP8 categorised /f/ as /v/ (in 40.6 percent of the cases), a miscategorisation that none of the other listeners showed and that is not further considered in the present study.

Table 2: Identification scores (percent correct) for four different types of learners in the study by Hamann and Sennema (2005a: 166).

| | | Learner A | | | Learner B | | |
|----------|-----|-----------|-----|------|-----------|------|------|
| | | /f/ | /v/ | /v/ | /f/ | /v/ | /v/ |
| stimulus | /f/ | 100% | 0% | 0% | 40% | 40% | 0% |
| | /v/ | 0% | 0% | 100% | 60% | 60% | 0% |
| | /v/ | 0% | 0% | 100% | 0% | 0% | 100% |
| | | Learner C | | | Learner D | | |
| | | /f/ | /v/ | /v/ | /f/ | /v/ | /v/ |
| stimulus | /f/ | 100% | 0% | 0% | 100% | 0% | 0% |
| | /v/ | 0% | 75% | 25% | 0% | 100% | 0% |
| | /v/ | 0% | 25% | 75% | 0% | 0% | 100% |

The performance of these four types of learners cannot be accounted for by a comparison of the phonemic categories in both languages, because in that case we would expect the German listeners to have problems with the Dutch

labiodental approximant, a category that does not exist in German. However, only Learner C showed this pattern. Conversely, three of the four learner types had problems categorising the Dutch voiced labiodental fricative /v/ correctly, although this category exists in German. An alternative explanation for the performance of the German listeners is the (dis)similarity in the *phonetic realisations* of these categories, which is the topic of the following section.

3. Acoustic and auditory differences between German and Dutch labiodentals

The labiodentals in Dutch and German contrast in voicing (voiced versus voiceless fricatives), and the Dutch ones additionally in manner (fricative versus approximant). Acoustically, fricatives differ from approximants in their presence of friction noise, absence of continuous formants and a longer duration. Voiced fricatives can be distinguished from voiceless ones by the presence of a voicing bar and periodicity of the signal. Usually, voiced fricatives are also shorter than their voiceless counterparts, see Stevens et al. (1992) for English, Mees and Collins (1982) for Dutch and Jessen (1998) for German. *Duration* is thus employed both for distinguishing fricatives from approximants and voiced from voiceless fricatives. Furthermore, duration is a well-known auditory dimension. Humans are able to perceive small durational differences in the speech signal due to the high temporal resolution of their auditory system (Plack 2004: 19).

The presence or absence of a voicing bar and of continuous formants, on the other hand, are acoustic characteristics but not necessarily relevant auditory dimensions. Instead, listeners seem to pay attention to the periodicity versus aperiodicity of a signal (Faulkner and Rosen 1999). The presence of friction noise is closely related to the auditory dimension of periodicity, since a segment with a large friction component is less periodic. An acoustic measure for this relation is the harmonics-to-noise ratio, or *harmonicity median*, which indicates the ratio of periodicity to friction in a sound. A signal with a harmonicity median of 0 dB, for instance, has equal energy in the harmonics and in the noise, and a signal with a harmonicity median of 20 dB has almost 100% of the energy in the periodic part (Boersma 1993). In the following, we employ harmonicity median as auditory dimension of periodicity and friction.

In an acoustic study with five Dutch and five German female speakers, Hamann and Sennema (2005b) compared, among other parameters, the duration and harmonicity median of the German labiodentals /f, v/ with those of the Dutch labiodentals /f, v, v/. The results of this study show that both German /v/ and Dutch /v/ share a short duration and a high harmonicity median (i.e., they have little friction and are voiced). Dutch /v/ differs in both parameters; it is of medium duration and has a low harmonicity median (i.e., has considerable friction and is voiced). Dutch and German /f/ are both of long duration and have a negative harmonicity median (indicating considerable friction and voicelessness). These results are presented in figure 1.

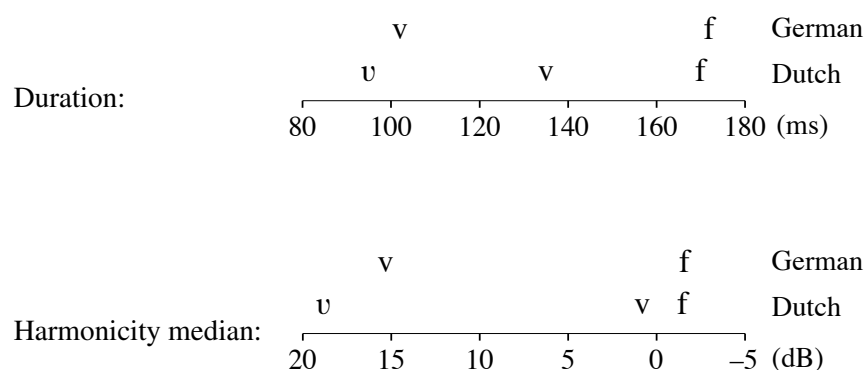


Figure 1: Scales comparing the realisations of the two German and the three Dutch labiodentals in the acoustic study by Hamann and Sennema (2005b) along the dimensions of duration (top) and harmonicity median (bottom).

The German labiodental fricative /v/ is acoustically closer to the Dutch labiodental approximant /v/ than to the corresponding Dutch labiodental fricative /v/. These findings are in line with the observations made by Viëtor (1897) and Kohler (1995) that German /v/ is not very fricative-like and can be realised as an approximant in initial position.

The actual spreading of the labiodental tokens from Hamann and Sennema's data along duration and harmonicity median are shown in figure 2, where we can see a clear overlap of Dutch /v/ and German /v/, and of Dutch and German /f/. The large variation in the realisation of Dutch /v/ and its partial overlap with Dutch /f/ is probably due to the often occurring neutral-

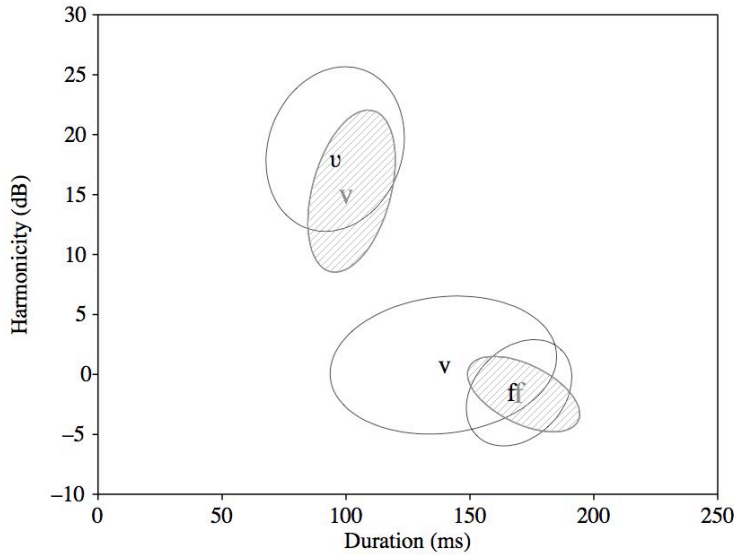


Figure 2: Distributions of German labiodentals (grey lines and hatched) and Dutch labiodentals (black lines) and their average values (position of symbols) along the dimensions of duration in ms (horizontal) and harmonicity median in dB (vertical).

isation of /v/ to /f/ described in section 1, though the speakers in the study were from the area of Nijmegen, where a contrast is supposedly made.

In contrast to the average values for the five Dutch speakers in figure 2, the stimuli produced by the Dutch male speaker (also from the area of Nijmegen) that were used in the perception experiment, given in figure 3, did not show a three-way distinction in duration for the labiodentals. Instead, the realisations of /v/, in the lower middle of figure 3, and those of /v/, in the upper part of figure 3, are almost of equal duration for this speaker.

The comparison of the average distributions for both languages (figure 2) offers an explanation for the perceptual misclassifications of Dutch labiodentals by the German L2 learners described in section 2. As we can infer from the grey ellipsis on the upper left in figure 2, German listeners are used to encounter tokens of /v/ with a duration between 85 and 120 ms and a harmonicity median between 9 and 22 dB. The realisations of Dutch /v/ are very similar to those. Likewise, the realisations of Germans /f/ with a du-

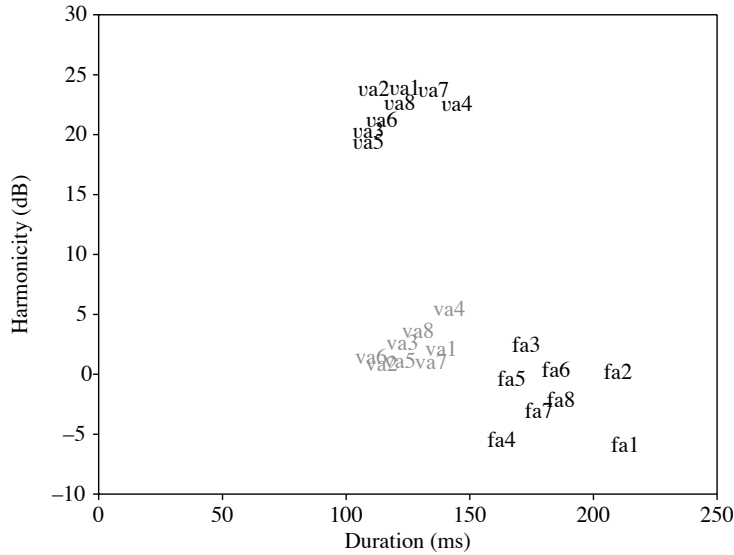


Figure 3: Dutch labiodental tokens (by one speaker) used in the perception experiment along the dimensions of duration (ms) and harmonicity (dB): tokens of /v/ in the upper middle, tokens of /v/ in the lower middle, and those of /f/ in the lower right corner.

ration between 150 and 200 ms and a harmonicity median between -5 and 2 dB are very similar to those of Dutch /f/. Thus Germans can employ the same auditory dimensions and very similar values along these dimensions for categorising Dutch /v/ and /f/ than they use for categorising their native labiodentals. However, tokens of Dutch /v/ pose a problem for German listeners because, although these tokens partly overlap with the two categories existing in German, they also provide values in between that the German listeners are not acquainted with.⁸ More importantly, the German listeners do not have a third category to associate these values with. This accounts for the problems German listeners have with correctly categorizing the Dutch voiced fricative. German learners of Dutch have to create a new segmental category for the Dutch /v/ and to associate the relevant auditory values with it, before they are able to classify the Dutch labiodentals correctly. The following section shows how the association of (new) segmental categories with auditory dimensions and specific values along these dimensions can be formalised.

4. Perception grammars for German and Dutch listeners

In speech perception, an auditory input is mapped onto an abstract phonological surface form. This process is largely language-specific, as both the abstract categories and their concrete auditory realisations can differ from language to language. The involvement of language-specific, i.e. grammatical, knowledge in speech perception has led a number of researchers to model this process with a *perception grammar* within an OT framework, see Boersma (1998 et seq.) and Pater (2004) for L1, and Hayes (2002), Escudero and Boersma (2003, 2004) and Escudero (2005) for L2.

The mapping of auditory forms onto phonological forms can be formalised in an OT perception grammar by means of negative *cue constraints*, as proposed by Boersma (1998) and Escudero and Boersma (2003):

- (3) *Cue constraints (or perceptual construction constraints):*
 “A value x on the auditory continuum f should not be perceived as the phonological category y ” or short “ xf not $/y/$ ”.

In our case, we use for instance the cue constraint “110 ms not $/f/$ ” to formalise the fact that a segment of 110 ms duration is not a voiceless labiodental fricative (at least at normal speaking rate)⁹, neither in German nor in Dutch.

Every cue constraint that maps an auditory value onto a phonological category has antagonistic cue constraints that map the same value onto other phonological categories in the respective language. Staying with our example and restricting our description to labiodentals, we also have, e.g., “110 ms not $/v/$ ” for German. Since 110 ms is a perfectly acceptable duration for a $/v/$ (but not for a $/f/$) in German, the constraint “110 ms not $/v/$ ” is much lower ranked than its antagonist “110 ms not $/f/$ ”, and therefore plays only a small role in the decision on the perceived category (due to the nature of OT as a decision-by-exclusion mechanism). Dutch has the additional constraint “110 ms not $/v/$ ”. From the data in figure 2 we can see that a $/v/$ with a duration of 100 ms does occur in Dutch, but is a marginal case of this category. For this reason, “110 ms not $/v/$ ” is higher ranked than “110 ms not $/v/$ ”, but lower than “110 ms not $/f/$ ”, to mirror the fact that a token with 110 ms is usually classified as $/v/$, sometimes as $/v/$, but never as $/f/$. The actual decision on the classification hinges then on the evaluation noise and the harmonicity value of this token interact with the durational constraints.

Cue constraints have to be formalised negatively for the following reason: positively formulated cue constraints fail to predict the correct categorisation in OT if more than two categories are involved, because the highest-ranked cue constraint will always determine the output (for an illustration of this point, see Boersma and Escudero 2008). The advantage of the negatively-formulated cue constraints is that they can also be employed in the modelling of production, where they ensure that a phonological form is realised with the corresponding perceptual cue values, see the bidirectional-phonology-and-phonetics model by Boersma (2006, 2007).

We postulate cue constraints for every possible incoming value along an auditory dimension that is used as a cue in the respective language. But instead of listing large numbers of cue constraints that refer to single values, the present article summarizes these as cue constraints that cover a certain interval along the auditory continuum, and gives them the following form: “ $x_1 .. x_2 f$ not /y/” (where x_1 and x_2 are values forming an interval on the auditory continuum f). Furthermore, the simplifying assumption is made that both Dutch and German listeners employ segmental duration and harmonic-ity median as the only auditory dimensions for labiodentals. In reality, speech perception involves several perceptual cues, which are weighted language-specifically (see e.g. Bradlow 1995). In addition, the present study does not include constraints necessary to identify the segments as labiodentals, though they are of course assumed to apply.¹⁰

The cue constraints as formalized by Escudero and Boersma (2003) are modified here to map continuous auditory dimensions onto features, and not directly onto segmental categories. By this, we can identify classes that share certain auditory qualities, thereby simplifying the identification of sounds. Furthermore, the mapping of auditory values onto phonological features allows for the normalisation of speech rate (see endnote 9).

The features used in the present study are of a more phonetic nature than those traditionally employed in phonology (e.g., Chomsky and Halle 1968 and Clements 1985). They are not part of a universal set, and thus can differ from language to language, and they do not necessarily provide minimal representations, as we will see below. Instead, the features employed here mirror as closely as possible the phonetic distributions of the classes they describe. This does not prevent their use to describe phonological processes, a point that cannot be illustrated here for lack of space.

The two German labiodentals show a distinction in their duration, which can be mapped onto a binary feature such as [\pm long].¹¹ The average produc-

tion of the Dutch labiodentals in figure 2 shows that all three are (at least partially) differentiated by duration, and therefore cannot be described with a binary feature [\pm long], alone. Either a ternary feature [long, medium, short] or two binary features [\pm long] and [\pm short] have to be used. The present study employs the latter solution for the simple reason that it reduces the number of cue constraints, no theoretical implication hinges on this decision. With respect to the dimension of harmonicity median, the German labiodentals show again a binary distinction, which can be mapped onto the feature [\pm noise], with /f/ being [+noise] and /v/ [-noise]. The Dutch labiodentals can be argued to show a three-way distinction in harmonicity (though with large overlap of /f/ and /v/), which would require an additional specification with a feature such as [periodic], where /v/ and /v/ are [+periodic], and /f/ is [-periodic]. The resulting feature specifications of the labiodentals are summarised in (4).

(4)

a. *Specification of labiodentals in German*

| | /f/ | /v/ |
|---------|-----|-----|
| [long] | + | - |
| [noise] | + | - |

b. *Specification of labiodentals in Dutch*

| | /f/ | /v/ | /v/ |
|------------|-----|-----|-----|
| [long] | + | - | - |
| [short] | - | - | + |
| [noise] | + | + | - |
| [periodic] | - | + | + |

The specifications in (4) are not minimal. The labiodentals in German could be distinguished by either [\pm long] or [\pm noise] alone. Similarly, the distinction of the Dutch sounds would require only two features, e.g. [\pm long] and [\pm short], or [\pm long] and [\pm noise]. Such minimal representations are possible, and actually employed by some speakers, as we will see below. However, they do not cover all perceptual data. The German L1 listeners in Hamann and Sennema's (2005a) perception experiment, for instance, categorised the Dutch stimuli in figure 3 mainly by duration (thus seem to employ a feature [long]), which accounts for a categorisation of /v/ and /v/ as their native /v/ in respectively 99.5 and 82.8 percent of the cases, see table 1 above. How-

ever, further 16.7 percent of Dutch /v/ were categorised as /f/, which is only possible if at least some of the listeners included the dimension of harmonicity (or another measure for noisiness, and thus a feature like [noise]) in their decision. This example illustrates that listeners of the same native language can differ in their use of perceptual cues, and as a consequence, several perception grammars (with different segmental specifications and usage of cue dimensions) are possible for the same language. The specifications proposed in (4) are what is maximally necessary for distinguishing the labiodentals in Dutch and German by duration and harmonicity median.

Based on the feature specifications in (4) and on the average values for the labiodentals in Hamann and Sennema (2005b), the following cue constraints are postulated for German (5a) and Dutch (5b).

(5)

- | | |
|--------------------------------------|-----------------------------|
| a. <i>Cue constraints for German</i> | |
| 70 .. 135 ms not [+long] | 70 .. 135 ms not [-long] |
| 136 .. 200 ms not [-long] | 136 .. 200 ms not [+long] |
| -6 .. 4 dB not [-noise] | -6 .. 4 dB not [+noise] |
| 5 .. 26 dB not [+noise] | 5 .. 26 dB not [-noise] |
| b. <i>Cue constraints for Dutch</i> | |
| 70 .. 155 ms not [+long] | 70 .. 155 ms not [-long] |
| 156 .. 200 ms not [-long] | 156 .. 200 ms not [+long] |
| 70 .. 114 ms not [-short] | 70 .. 114 ms not [+short] |
| 115 .. 200 ms not [+short] | 115 .. 200 ms not [-short] |
| -6 .. 8 dB not [-noise] | -6 .. 8 dB not [+noise] |
| 9 .. 26 dB not [+noise] | 9 .. 26 dB not [-noise] |
| -6 .. -1 dB not [-periodic] | -6 .. -1 dB not [+periodic] |
| 0 .. 26 dB not [+periodic] | 0 .. 26 dB not [-periodic] |

The constraints in the left column of (5) are crucial for the correct classification in both languages and are ranked high, whereas the antagonistic constraints on the right are ranked low. These antagonist cue constraints are not included in the following tableaux for lack of space.

A perception grammar that can correctly categorize the German data from figure 2 above is given in (6). Both tableaux have input auditory forms based on the average values for the German segments obtained in Hamann and

Sennema's (2005b) acoustic study. Candidates are the two phonological segments /f/ and /v/ since these are the only labiodentals in German.

(6) *A German perception grammar*

| | | | | | |
|-----|-----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [102 ms, 15 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| [v] | /f/ | | *! | | * |
| | /v/ | | | | |

| | | | | | |
|-----|-----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [172 ms, -2 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| [f] | /f/ | | | | |
| | /v/ | *! | | * | |

As we can see from these tableaux, the ranking among the four constraints is irrelevant for the perception of the German labiodentals, since any ranking will result in the same winning candidates, as long as these four constraints are ranked above their not-included antagonist constraints. Furthermore, we can see that either the two constraints referring to the durational contrast or those referring to the noisiness would be sufficient to decide on the winning candidates. This shows that there is not only one possible perception grammar for German labiodentals but several. These possible grammars differ in the number of cue constraints, depending on which cues the individual listener uses to distinguish the sounds, and/or in the ranking of the constraints, illustrating the different weight individual listeners give to certain cues. We assume the ranking as given in (6), because it predicts the correct naïve perception of Dutch labiodentals (see section 5.1 below).

Similarly, there are several possible perception grammars for the Dutch labiodentals. The most complex one would use all specifications in (4b) and therefore all cue constraints listed in (5b). In (7), we see a perception grammar for Dutch that is less complex and uses only the specifications [\pm long] and [\pm noise] and the respective cue constraints, modelling a listener who employs only a binary distinction on the dimensions of duration and harmonicity. The input auditory forms are based on the average values for the Dutch segments from Hamann and Sennema's (2005b) acoustic study. Output candidates are the three labiodental categories in Dutch.

(7) *A Dutch perception grammar*

| | | | | | |
|-----|----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [97 ms, 19 dB] | 156..200ms not [-long] | 70..155ms not [+long] | -6..8dB not [-noise] | 9..26dB not [+noise] |
| [v] | /f/ | | *! | | * |
| | /v/ | | | | *! |
| | /v/ | | | | |

| | | | | | |
|-----|----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [133 ms, 1 dB] | 156..200ms not [-long] | 70..155ms not [+long] | -6..8dB not [-noise] | 9..26dB not [+noise] |
| [v] | /f/ | | *! | | |
| | /v/ | | | | |
| | /v/ | | | *! | |

| | | | | | |
|-----|-----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [170 ms, -2 dB] | 156..200ms not [-long] | 70..155ms not [+long] | -6..8dB not [-noise] | 9..26dB not [+noise] |
| [f] | /f/ | | | | |
| | /v/ | *! | | | |
| | /v/ | *! | | | |

The use of additional constraints or a different constraint ranking would yield different perception grammars with the same categorisation results. Different results can only occur if the ranking between the constraints in (7) and their antagonists changes, if the feature specification of the segments in (4) is changed, or if we employ different cues than harmonicity and duration.

5. The perception grammars of four types of learners

With the formalisation of the differences in perception grammars between Dutch and German listeners in the previous section, we can now account for the performance of the four types of L2 learners that were introduced in section 2. Since these learners had to categorise the tokens of the speaker given in figure 3, the input to the following tableaux are the average values for these tokens.

Learners of type D do not differ in performance from Dutch speakers and we can conclude that this group has acquired L2 perception grammars identical to that of native listeners. If we assume the Dutch grammar in (7), then the change in input (from the values averaged over 5 female speakers to the average values for the one male speaker) still yields the same results, which is not

further illustrated here. Learner types A to C differ in various degrees from native-like performance, as described in detail in the following subsections.

5.1. The L2-perception grammar of Learner A

Learners of type A identified the tokens with /f/ and /v/ correctly and categorised items of the voiced fricative as /v/. This performance is remarkable in as far as the difference in harmonic median between the tokens of Dutch /v/ and /ʋ/ is not employed as a cue for their categorisation. Learners of type A seem to simply apply their German perception grammar to the Dutch tokens and to equate both Dutch /v/ and /ʋ/ with their native /v/. This indicates that the learners are at an early stage of their acquisition of Dutch and have not created a separate phonological category for a third labiodental in Dutch, yet.

Following Escudero and Boersma (2004) and Escudero (2005), who argue that an L2 learner starts off by copying the native perception grammar, i.e. the phonological categories and cue constraints from the native language (based on the *Full Transfer Theory* by Schwartz and Sprouse 1996), we can assume that learners of type A simply use a copy of their German perception grammar for the Dutch sounds. This is illustrated with the perception tableaux in (8), which contain the constraints and rankings of the German perception grammar in (6). Input are the Dutch tokens, and candidates are the Dutch phonemes /f/ and /v/ (they could also be labelled German /f/ and /v/, respectively). Dutch /v/ is not included, since learners of type A seem to have no segmental representation for this sound.

(8) *Learner A: A German perception grammar (with Dutch labels) on Dutch inputs*

| | | | | | |
|-----|-----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [120 ms, 22 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| /v/ | /f/ | | *! | | * |
| | /ʋ/ | | | | |

| | | | | | |
|-----|----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [125 ms, 3 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| /v/ | /f/ | | *! | | |
| | /ʋ/ | | | * | |

| | | | | | |
|-----|-----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [185 ms, -2 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| /f/ | /f/ | | | | |
| | /v/ | *! | | * | |

The application of a German perception grammar on the Dutch input can account for the performance of naïve monolingual Germans, too. The results of the German L1 listeners in table 1 show that not all tokens of Dutch /v/ were heard as German /v/. 16.7 percent have been categorised as /f/. This is not surprising, if we look at the actual realisations of the tokens in figure 3. Token *va4*, for instance, has a duration of 141 ms which will be categorised by the German perception grammar as /f/:

(9) *A German perception grammar (with Dutch labels) on token “va4”*

| | | | | | |
|------------|----------------|---------------------------|--------------------------|-------------------------|-------------------------|
| | [141 ms, 5 dB] | 136..200ms not [-long] | 70..135ms not [+long] | -6..4dB not [-noise] | 5..26dB not [+noise] |
| <i>va4</i> | /f/ | | | | * |
| | /v/ | *! | | | |

Token *va1* undergoes a similar misperception as formalised in (9), because it has as a duration of 138 ms, which is categorised as [+long]. All other tokens of /v/ have durations below 136 ms and are therefore categorised as /v/.

The categorisation of some Dutch /v/ tokens as German /f/ could be used to explain why one of the L2 listeners (PP5) categorised only 75 percent of /v/ as Dutch approximant and 25 percent incorrectly. This explanation requires, however, that this listener has separate categories for /f/, /v/ and /v/ and has associated some durational values with /v/. A more likely explanation for this performance is the experimental setup, where three answer categories were provided for the labiodentals, and the participant felt obliged to use the third one (/v/) from time to time, although she does not have three labiodental phonemes.

The second L2 learner that belongs to the group of Learner A showed some additional miscategorisation of /f/ as /v/, a pattern that is covered by type B below.



5.2. The L2-perception grammar of Learner B

Learners of type B perceived the Dutch labiodental approximant /v/ always correctly, but frequently confused the two fricatives. These results can be

interpreted in the following way. The learners of type B have only two labiodental categories in their L2 grammar, namely one (the German voiced fricative) for Dutch /v/ and one (the German voiceless fricative) for /v/ and /f/ together. This latter category has two corresponding graphemes, <v> and <f>. ¹² Such a grammar is very likely to emerge in an environment where all speakers neutralize the distinction between voiced and voiceless fricatives, something that is reported for a large number of Dutch speakers, recall the description in section 1. As a result, learners receive no input with separate distributions for the labiodental voiced and voiceless fricatives, and thus have no perceptual dimension (neither harmonicity median nor duration) that yields a correct classification, but have to base their decision on chance. We could not control for previous L2 input in our perception experiment, but it seems very likely that some of the learners received only ‘neutralising’ input.

The learners of type B had to learn to collapse or ignore certain cues and cue values in order to achieve this performance. The cue of duration clearly distinguishes the tokens of /v/ and /f/ in the perception experiment, and therefore the uniform treatment of these tokens can only be attained if duration is not considered as a cue. Furthermore, the tokens of /v/ and /v/ are not distinguished via duration by our speaker, but learners of type B have a clear distinction between these two classes, giving further evidence to the hypothesis that these learners do not employ durational cues. These two observations can be modelled with a perception grammar that does not have constraints on the duration of labiodentals. Furthermore, such a perception grammar does not employ separate representations for /v/ and /f/, see the tableaux in (10).

(10) *A Dutch perception grammar of Learner B*

| | | | |
|------------|---|-------------------------|-------------------------|
| /v/ | [120 ms, 22 dB] | -6..8ms not [-noise] | 9..26ms not [+noise] |
| | /f,v/ | | *! |
| |  /v/ | | |
| /v/ or /f/ | [125..185 ms, -2..3 dB] | -6..8ms not [-noise] | 9..26ms not [+noise] |
| |  /f,v/ | | |
| | /v/ | *! | |

Two of the three learners in group B actually showed some additional mis-categorisation of /v/ as /v/. This performance can be explained if duration interacts with harmonicity, and all tokens shorter than for instance 120 ms

are categorised as /v/, see the pattern of learner type C described below. The durational difference between German /v/ and /f/ is “unlearnt” for the Dutch perception grammar of these listeners, otherwise we would not find the observed confusion patterns.

An alternative explanation for the learners of type B is that they have created a third category, and distinguish all three labiodentals by harmonicity, only. This requires a specification along the harmonicity dimension with two features such as [\pm noise] and [\pm periodicity], recall (4b). Such an account works if we assume as input the data from figure 2, where the categories of /f/ and /v/ are largely overlapping on the harmonicity dimension. The same categories show much less overlap in the tokens in figure 3, however, and we would therefore run into problems when trying to predict the observed confusions with a three-member harmonicity contrast and the input from our male speaker in figure 3.

5.3. The L2-perception grammar of Learner C

Listeners of type C correctly identified Dutch /f/ and categorised some tokens of /v/ as /v/ and vice versa. As the majority of the tokens are classified correctly, we can conclude that these learners have created a third labiodental category for Dutch. A possible explanation for the observed confusions is that tokens of the approximant with a long duration are perceived as /v/, whereas tokens of /v/ with a high harmonicity value are perceived as /v/. Hence both duration and harmonicity are used as cues, and they are traded in a specific relation. Consequently, the resulting perception grammar is more complex than the one formalised in (7) for Dutch listeners. A modelling of this grammar and the cue trading requires a specification of the labiodentals as both [\pm short] and [\pm long]. For the features [\pm short] and [\pm noise], the boundaries and thus the respective constraints have to be slightly different from those postulated in (4b) and (5b): The learners of type C set the boundary between [+short] and [-short] at 124 ms, not 114 ms, because they categorise tokens with a duration of 114-124 ms as [+short], i.e., /v/. Furthermore, the boundary between [+short] and [-short] lies at 4 dB, not 8 dB, because tokens with harmonicity values between 4 and 8 dB are categorised as [-noise], thus /v/, independent of their length.¹³

To illustrate the perception of such tokens, the input Dutch /v/ is split into two categories in the following grammar; tokens with a duration of 70-124 ms

(i.e., /v/ tokens from the five speakers in figure 2), and tokens with a longer duration of 125-155 ms (which are still tokens of /v/ for our male speaker). Both categories have the same harmonicity values of 9-26 dB. Similarly, the input of Dutch /v/ is split into two categories; tokens with a harmonicity value of 1-3 dB and tokens with a harmonicity value of 4-8 dB, both categories having the same durational values.

(11) *A Dutch perception grammar for Learner C¹⁴*

| | [70..124 ms, 9..26dB] | 156..200 ms not [-long] | 70..155 ms not [+long] | -6..3dB not [-noise] | 70..124 ms not [-short] | 125..200 ms not [+short] | 4..26dB not [+noise] |
|-------------------|--------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|--------------------------------|----------------------------|
| /v ₁ / | /f/ | | *! | | * | | * |
| | /v/ | | | | *! | | * |
| | /v/ | | | | | | |

| | [125..155 ms 9..26dB] | 156..200 ms not [-long] | 70..155 ms not [+long] | -6..3dB not [-noise] | 70..124 ms not [-short] | 125..200 ms not [+short] | 4..26dB not [+noise] |
|-------------------|--------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|--------------------------------|----------------------------|
| /v ₂ / | /f/ | | *! | | | | * |
| | /v/ | | | | | | * |
| | /v/ | | | | | *! | |

| | [70..155 ms, 1..3dB] | 156..200 ms not [-long] | 70..155 ms not [+long] | -6..3dB not [-noise] | 70..124 ms not [-short] | 125..200 ms not [+short] | 4..26dB not [+noise] |
|-------------------|-------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|--------------------------------|----------------------------|
| /v ₁ / | /f/ | | *! | | * | | |
| | /v/ | | | | (*) | | |
| | /v/ | | | *! | | (*) | |

| | [70..155 ms, 4..8dB] | 156..200 ms not [-long] | 70..155 ms not [+long] | -6..3dB not [-noise] | 70..124 ms not [-short] | 125..200 ms not [+short] | 4..26dB not [+noise] |
|-------------------|-------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|--------------------------------|----------------------------|
| /v ₂ / | /f/ | | *! | | * | | * |
| | /v/ | | | | *! | | * |
| | /v/ | | | | | (*) | |

| | [156..200 ms -6..3dB] | 156..200 ms not [-long] | 70..155 ms not [+long] | -6..3dB not [-noise] | 70..124 ms not [-short] | 125..200 ms not [+short] | 4..26dB not [+noise] |
|-----|--------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|--------------------------------|----------------------------|
| /f/ | | | | | | | |
| /v/ | | *! | | | | | |
| /ʋ/ | | *! | | * | | * | |

The actual data from the perception experiment supports the analysis formalised in the perception grammar in (11): A miscategorisation of /v/ as /ʋ/ occurred mainly for tokens of /v/ that had a harmonicity median around 3 dB or above (*va4*: 5.6 dB, *va8*: 3.9 dB, *va3*: 2.9 dB), and there was a tendency to miscategorise /v/ tokens that were rather long (*va4*: 144 ms, *va8*: 122ms, *va1*: 124ms).

Like the native grammar in (7), the learner's grammar in (11) predicts the correct output for the average values of the male speaker in the perception experiment and the average values for the five speakers from the acoustic study. Thus, this grammar is able to deal with prototypical tokens of the three labiodentals, but departs from the L1 grammar in (7) with respect to the categorisation of marginal tokens.

6. Summary and conclusion

The four types of German learners described in section 5 showed very different perceptual strategies for categorising the labiodentals in Dutch. Learner type A simply applied a copy of the German perception grammar, with German representations of the two labiodentals /f/ and /v/ and their respective cue constraints. The learners of type B also had only two labiodental categories for the three Dutch sounds, but had changed their German representations and cue constraints to mirror a Dutch environment that neutralises /f/ and /v/. The learners summarised in Type C were capable of classifying prototypical tokens of the Dutch labiodentals, hence had constructed a category and representations for the voiced fricative and had associated them with the relevant cues. However, these representations and cue constraints proved to be different from those of a native speaker when dealing with non-prototypical tokens. Lastly, learners of type D showed native-like performances, indicating that they acquired representations and cue constraints identical to those of L1 speakers.

One could argue that these four types of learners represent four different stages in the L2 acquisition of Dutch. This is certainly the case for type A

and type D, with type A representing the very beginner's stage and type D the end stage. However, there is no evidence for assuming that the increasing correct performance from type B to D indicate three successive developmental stages. The assumption of three successive stages implies that learners first have to have only one representation that combines both labiodental fricatives in Dutch, and employ only harmonicity as a cue to differentiate them from the labiodental approximant. Then they have to move on to construct a third labiodental category and durational cue constraints that distinguish three levels and that interact with harmonicity cues. And lastly they have to get rid again of the third durational category and the interaction of harmonicity and duration constraints.

The present article assumes instead that the perception grammars for types B to D are possible variants of dealing with the same categorisation task, provided the learners had a certain amount of L2 input. These different grammars reflect perception strategies that the learners developed based on their L2 input and on their L1 perception grammar. The role of the received input was elaborated already in section 5.2 for Learner B, who seems to have encountered only neutralised, i.e. voiceless, labiodental fricatives. Learner C and D had received less or no neutralised input, but differ insofar as Learner C probably had a less varied input than Learner D, since Learner C could not deal with all tokens of the Dutch labiodentals provided in the experiment. With respect to the L1 grammar, the present article illustrated that there are several possible perception grammars for German and Dutch that all provide the same categorisation for the native labiodental categories. The idea of learners ending up with different grammars that result in the same output is not new in the OT acquisition literature, see Apoussidou's (2006) work on simulated learning of stress in Pintupi. The role of speaker-specific differences in the L1 grammar for the acquisition of an L2 perception grammar, however, could not be further elaborated here for lack of data. Future research is necessary where detailed perception experiments on L2 and L1 of the same speakers (and possibly also developmental studies on their L2 perception) allow a comparison between the L1 and (several stages of) the L2 grammar.

We saw in the present study that two factors play a role in the account of L2 perception: phonetic factors such as auditory dimensions and the weightings thereof, and phonological factors such as the segmental categories of the native language and their specifications. In this respect, the present study supports earlier L2 studies such as Flege and Hillenbrand (1986) who argue that both phonetic differences between L1 and L2 and the phonology of L1

are crucial in the account of L2 perception. Furthermore, the present study provides a formalisation of how these two factors interact with the help of perception grammars and cue constraints, following Escudero and Boersma's (2003, 2004) groundbreaking work. This enables us to move from a mere description of the data (by comparing phonemes and perceptual cues employed in L1 and L2) to a prediction of L2 perception based on language-specific perception and phonological representations.

The present study is innovative in proposing listener-specific variation in both phonetic and phonological factors of L2 perception: individual choices of auditory dimensions and weighting of these dimensions, as well as individual specifications of phonological contrasts that are not present in the native language.

Notes

1. Parts of the research reported here were presented at the *Phonetik und Phonologie 2* meeting, Tübingen, July 19, 2005; at the Universities of Amsterdam, Kiel, and Utrecht; and at the ZAS Berlin. I thank the audiences on all of these occasions for their comments. I have also received helpful suggestions and comments by Paul Boersma, Mirjam Ernestus, Frank Kügler, Aditi Lahiri, Anke Sennema, and two anonymous reviewers. I gratefully acknowledge the support of grant GWZ 4/8-1-P2 by the German Science Foundation (DFG) and a VENI postdoctoral fellowship 16.064.057 by the Dutch Science Foundation (NWO).
2. Whereas the symbol /β/ stands for a bilabial fricative, the underscore /_/ indicates that it is lowered, i.e., an approximant. The symbol /w/ is sometimes also used for the Southern Dutch/Flemish sound (e.g. Cohen et al. 1961), but incorrectly implies a secondary velarisation.
3. German has the grapheme <v>, which is used both for /v/ and /f/, see e.g. *Vase* [va:zə] 'vase' and *Vieh* [fi:] 'cattle', respectively.
4. The Dutch alveolo-palatal sibilant [ç] is usually considered an allophone of /s/, see Mees and Collins (1982: 2).
5. A statistical analysis of the data that would yield a clustering of speakers is unfortunately not possible. The most appropriate statistical method for this purpose is a generalised linear mixed-effects model (Lindstrom and Bates 1990). However, at the time this paper was finished such a model could not yet deal well with categorical dependent variables.
6. 5. The performance of these two participants is simplified here for reasons that become obvious in §5.1: PP11 showed some additional miscategorisation of /f/ as /v/, a pattern that is covered by learner type B. PP5 categorized 25% of /v/ incorrectly, as accounted for in §5.1.
7. PP20 and PP21 actually categorize /v/ as /v/ in 30 percent of the cases. This confusion pattern is covered by Learner C.
8. The speaker that was used in the perception experiment did not distinguish /v/ and /v/ by duration (cf. figure 3), and the German listeners therefore could not employ duration

alone for categorizing these sounds in the experiment. Since the average distributions of the three labiodentals in figure 2 (from the five speakers in Hamann and Sennema 2005b) show that there is large overlap of the tokens of all *three* categories along this dimension, duration alone does not seem to be a reliable cue for the Dutch labiodentals in general.

9. Normalisation for speaking rate is possible if we employ durational cue constraints that map auditory values onto phonological features such as [\pm long], see (5). This allows a relativization of values according to overall speech rate. The actual values of a constraint like “x–y ms is not [+long]” in the perception of a labiodental, for instance, can be determined by the duration of adjacent vowels or the larger context (word, utterance) in which the segment appears. Similar normalization of speaker-dependent differences such as differences in formant values can be performed via contextual information (like the pitch range within the word or utterance).
10. Cues employed for identifying place of articulation are for instance the spectral features of the friction noise (e.g. Whalen 1991) and vowel transitions (e.g. .Bladon et al. 1987; see also Nowak 2006 who illustrates the interaction of both cues in the perception of the Polish sibilants).
11. A larger contrast on the durational dimension might be necessary for other than labiodental contrasts in German, see e.g. Hamann’s (2003) proposal to account for the difference between long tense vowels, short tense vowels and glides of the same place of articulation in German, such as /i:, i, j/, by a three-way durational contrast.
12. This situation is comparable to German, where <f> corresponds to /f/, and both <v> and <w> correspond to /v/, see endnote 3.
13. As the constraints employed in this article cover certain intervals on auditory dimensions, the change of boundary values actually means the following. The high-ranked “115 ms not [+short]”, which was integrated in the constraint “115..200 ms not [+short]”, is now ranked lower than its antagonist “115 ms not [–short]”, which is now integrated in “70..124 ms not [–short]”. The same holds for the durational constraints for 116–124 ms. The change is thus one in ranking, not in values. This also applies to the ‘boundary’ change of the harmonicity constraints.
14. The violation marks in brackets indicate that the respective constraints are violated by part of the input values. Due to the low ranking of these constraints, the input is not further split to illustrate these violations.

References

- Apoussidou, Diana
2006 Learning to be insensitive to weight in Pintupi. Ms., ROA 799.
- Best, Catherine T., and Winifred Strange
1992 Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305-330.
- Bladon, Anthony, Christopher Clark, and Katrina Mickey
1987 Production and perception of sibilant fricatives: Shona data. *Journal of the International Phonetic Association* 17 (1): 39-65.
- Boersma, Paul
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.
1998 *Functional Phonology: Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.
2006 Prototypicality judgments as inverted perception. In *Gradience in Grammar*, Gisbert Fanselow, Caroline Féry, Ralf Vogel and Matthias Schlesewsky (eds.), 167-184. Oxford: Oxford University Press.
Some listener-oriented accounts of h-aspiré in French. *Lingua* 117: 1989-2054.
- Boersma, Paul, and Paola Escudero
2008 Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. In *Contrast in phonology: theory, perception, acquisition*, Peter Avery, Elain Dresher and Keren Rice (eds.), 271-301. Berlin and New York: Mouton de Gruyter.
- Booij, Geert
1995 *The Phonology of Dutch*. Oxford: Oxford University Press.
- Bradlow, Ann R.
1995 A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97: 1916-1924.
- Chomsky, Noam, and Morris Halle
1968 *The Sound Pattern of English*. New York: Harper and Row.
- Clements, George N.
1985 The geometry of phonological features. *Phonology Yearbook* 2: 225-252.
- Cohen, A., C.L. Ebeling, K. Fokkema, and A.G.F. van Holk
1961 *Fonologie van het Nederlands en het Fries*. 'S-Gravenhage: Martinus Nijhoff.
- Escudero, Paola
2005 *Linguistic perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: LOT.
- Escudero, Paola, and Paul Boersma
2003 Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In *Proceedings of the 25th Annual Penn Linguistics Colloquium (Penn Working Papers in Linguistics 8.1)*, Sundha Arunachalam, Elsi Kaiser and Alexander Williams (eds.), 71-85.

- 2004 Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551-585.
- Faulkner, Andrew, and Stuart Rosen
 1999 Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception. *Journal of the Acoustical Society of America* 106 (4): 2063-2073.
- Flege, James
 2002 Interactions between the native and second-language phonetic systems. In *An integrated view of language development: Papers in honor of Henning Wode*, P. Burmeister, Thorsten Piske and A. Rohde (eds.). Trier: Wissenschaftlicher Verlag.
- Flege, James Emil, and James Hillenbrand
 1986 Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America* 79 (2): 508-517.
- Gussenhoven, Carlos
 1999 Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association* 74-77. Cambridge: Cambridge University Press.
- Hamann, Silke
 2003 German glide formation functionally viewed. *ZAS Working Papers in Linguistics* 32: 137-154.
- Hamann, Silke, and Anke Sennema
 2005a Voiced labiodental fricatives or glides - all the same to Germans? In *Proceedings of the Conference on Plasticity in Speech Processing*, Valery Hazan and Paul Iverson (eds.), 164-167. London: UCL.
 2005b Acoustic differences between German and Dutch labiodentals. *ZAS Working Papers in Linguistics* 42: 33-41.
- Hayes, Rachel
 2002 The perception of novel phoneme contrasts in a second language: A development study of native speakers of English learning Japanese singleton and geminate consonant contrasts. In *University of Arizona Coyote Working Papers 12: Language in Cognitive Science*, Rachel Hayes, W.D. Lewis, E.L. O'Bryan and Tania Zamuner (eds.), 28-41. Tuscon: University of Arizona Press.
- Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ich Tohkura, Andreas Kettermann, and Claudia Siebert
 2003 A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: 47-57.
- Jessen, Michael
 1998 *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.
- Kohler, Klaus
 1995 *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
 1999 Illustrations of the IPA: German. In *Handbook of the International Phonetic Association* 86-89. Cambridge: Cambridge University Press.

- Lindstrom, Mary J., and Douglas M. Bates
1990 Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics* 46 (3): 673-687.
- McCarthy, John J., and Alan Prince
1993 Prosodic morphology I: Constraint interaction and satisfaction. Ms., *University of Massachusetts and Rutgers University*.
- Mees, Inger, and Beverly Collins
1982 A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association* 12 (1): 2-12.
- Nowak, Pawel M.
2006 The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 43 (2): 139-152.
- Pater, Joe
2004 Bridging the gap between perception and production with minimally violable constraints. In *Constraints in phonological acquisition*, René Kager, Joe Pater and Wim Zonneveld (eds.). Cambridge: Cambridge University Press.
- Plack, Christopher J.
2004 Auditory Perception. In *Psychology: An international perspective*, Michael W. Eysenck (ed.) 1-19. New York: Psychology Press.
- Prince, Alan, and Paul Smolensky
1993 Optimality Theory: Constraint Interaction in Generative Grammar. Ms., *Rutgers University*. Brunswick.
- Schwartz, Bonny, and R.A. Sprouse
1996 L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12: 40-72.
- Stevens, Kenneth N., Sheila Blumstein, Laura B. Glicksman, M. Burton, and Kathleen Kurowski
1992 Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America* 91: 2979-3000.
- Viëtor, Wilhelm
1897 (5th edition 1904) *Elemente der Phonetik des Deutschen, Englischen und Französischen*. Leipzig: Reisland.
- Wängler, Hans-Heinrich
1974 *Grundriss einer Phonetik des Deutschen*. Marburg: N.G. Elwert Verlag.
- Whalen, D. H.
1991 Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America* 90 (4): 1776-1785.
- Wiese, Richard
1996 *Phonology of German*. Oxford: Oxford University Press.