Title
Perception of Fricatives by Dutch and English Speakers

Journal
UC Berkeley PhonLab Annual Report, 3(3)

ISSN
2768-5047

Authors
Johnson, Keith
Babel, Molly

Publication Date
2007

DOI
10.5070/P79sf875pc

Perception of fricatives by Dutch and English speakers

Keith Johnson and Molly Babel
University of California, Berkeley

Abstract. Two speech perception experiments explored the physical and linguistic bases of phonetic similarity. Because phonological features are grounded in phonetic similarity, these experiments were thus an exploration of the phonetic bases of distinctive features. Experiment 1 found that Dutch listeners rated [s] and [ʃ] as more similar to each other than did American English listeners. We attributed this to a pattern of alternation in Dutch phonology involving these sounds. Experiment 1 also found that Dutch listeners rated [θ] and [s], and [θ] and [ʃ] as more similar to each other than did American English listeners. We attributed this to the lack of [θ] in the Dutch inventory of native sounds. Experiment 2 found that Dutch and American English listeners did not signficantly differ from each other in a speeded discrimination task with the same stimuli as experiment 1. Reaction times in experiment 2 were highly correlated with the rating data of experiment 1 (r = -.86 to -.96) indicating that the general pattern of response in experiment 1 was based on auditory similarity, with language-specific effects superimposed on the general pattern. We conclude that phonetic similarity, a base of phonological features, is comprised of three components: (1) auditory similarity, (2) phonetic inventory, and (3) language-specific patterns of alternation.

1. Introduction

This is a paper about phonetic similarity. By "phonetic similarity" we mean a psychological phenomenon - something like "the subjective experience of linguistic sound similarity" - which is no doubt related to physically measureable acoustic, visual, or motor aspects of speech sounds; but which also may be derived from the listener's experience of the sounds of language. We will argue on the basis of data from two cross-linguistic speech perception experiments, that phonetic similarity is a product both of the physical aspects of sounds and of the linguistic patterning of sounds. This research is relevant to distinctive feature theory because phonological features are grounded in phonetic similarity. Phonetic similarity is relevant in the poetic use of language to determine near-rhymes (Steriade, 2003), sound change is influenced by the perceptibility of the sounds involved in the change (Hume & Johnson, 2001), and synchronic phonological patterns are influenced by phonetic similarity (e.g. Hanson, 2001; Frisch et al. 2004; Rose & Walker, 2004).

It makes sense then to suppose that distinctive features are grounded (Archangeli and Pulleyblank, 1994; Steriade, 2001) in the physical phonetic properties of speech sounds - acoustic, visual, and motor. However, a universal phonetic (Chomsky & Halle, 1968) grounding is implausible because phonetic realization is language-specific (Ladefoged, 1984) and thus the phonetic basis of particular phonological patterns are language specific (Johnson, 1994), and the particular system of contrasts within a language influences phonetic/phonological similarity (MacKenzie, 2005).

Our results suggest that phonetic similarity, as reflected in listeners' explicit judgements of sound similarity, derives from at least three sources of subjective similarity. The first of these is the psychophysical similarity of actual speech sounds in the human auditory, visual and motor sensory systems. The second source of similiarity derives from the inventory of contrasts in the listener's native language, where the range of phonetic variation covered by a sound in the inventory may be more than one sound in a second language (see Werker & Tees, 1984; Kuhl et al, 1992; Best et al., 2001). For example, the Polish fricatives [ʂ] and [ç] both sound like [ʃ] to speakers of American English and this influences the relative perceptibility of the contrast for Polish and American English listeners (McGuire, 2007). Lastly, the pattern of phonological alternations in the listener's native language leads to greater or lesser subjective phonetic separation of sounds as these sounds participate in phonological alternations. For example, [d] and [ð], while perfectly distinguishable are rated as sounding more similar to each other by Spanish speakers (for whom [ð] is related to [d]) than they are by English speakers (Boomershine, et al., in press). Before we turn to the experiments that led us to this view of phonetic similarity let us review some of the speech perception literature that serves as background for this work.

Many studies of cross-linguistic speech perception have examined the consequences of a lack of contrast in a language. For example, Japanese listeners' behavior with the English /r/-/l/ contrast has been the focus of numerous studies (Goto, 1971; Miyawaki et al. 1975; McKain et al. 1981; Strange & Ditman, 1984; Logan et al., 1991; Yamada et al., 1992; Lively et al., 1993; Flege et al., 1996). Perception of the English voicing dimension (cued by aspiration noise in the onset of stressed syllables) has also been studied in French/English bilinguals (Caramazza et al., 1973) and Spanish speakers (Elman et al., 1977; Williams, 1977). Similarly, Terbeek (1977) studied cross-linguistic differences in vowel perception in a classic comparison of inventory effects on speech perception.

A few other studies have also explored the perceptual consequences of allophony (Gandour, 1983; Dupoux et al., 1997; Harnsberger, 2001; Hume & Johnson, 2003; Huang, 2004; Boomershine et al., in press). For example, Boomershine, et al. (in press) investigated the perception of [d], [ð], and [ɾ] by speakers of Spanish and English. In English, [ɾ] and [d] may alternate with each other (e.g. [d] in "odd" and [ɾ] in "odder") while [ð] does not alternate with either.[1] The situation is reversed in Spanish, where [ð] and [d] alternate with each other (*un [d]edo* "a finger", *tu [ð]edo* "your finger") while [ɾ] doesn't alternate with either. Boomershine et al. found that English listeners rated [ɾ] and [d] as more similar to each other than Spanish listeners did, while [ð] and [d] were more similar for Spanish listeners than for English listeners. Interestingly, Boomershine et al. also found these cross-linguistic patterns of perceptual similarity in reaction time in a speeded discrimination task, like the one that we report here in experiment 2.

The present paper presents the results of two experiments that explore the language-specificity of speech perception, looking at similarities and differences in how American English- and Dutch-speaking listeners discriminate the voiceless fricatives [f θ s ʃ x h] in a subjective similarity rating task (experiment 1) and in a speeded discrimination task (experiment 2).

The phonemic inventory of Dutch (Booij, 1995; Cohen et al., 1972; De Groot, 1968) includes the velar fricative [x] and does not include [ʃ] or [θ]. English on the other hand does use [ʃ] and [θ] to contrast words but does not include [x]. We expect then, given these descriptions of the inventories of contrastive sounds and previous research on second language speech perception, that Dutch speakers will have greater sensitivity to [x] than will English speakers and that English listeners will be more sensitive to the phonetic properties of [ʃ] and [θ]. These predictions depend though on whether listeners identify nonnative fricatives with fricatives that they are familiar with. For example, we might expect that English listeners would hear [x] as a variant of [h] because of the acoustic similarities between these sounds. The possibility that English /h/ is sometimes said with velar frication, e.g. in "who", may also be a factor. If English

[1] Manuel & Wyrick (1999) found that [d] and [ð] do alternate with each other in connected speech in the sense that many words with [ð] in their citation form are pronounced with stop closure and a noticable release burst, i.e. they are pronounced with [d]. An important question for future research is whether the presence of alternations of this type that is found only or primarily in connected speech impact perception. Here we confine the discussion and subsequent interpretation to alternations that are apparent in carefully produced forms.

listeners make this association then we would expect that [x] and [h] would be more phonetically similar to them than they are for Dutch-speaking listeners (see Best, 1995; and Flege, 1995 for detailed discussion of these issues). Similarly, we might expect that Dutch listeners would have lowered perceptual sensitivity to [θ] particularly in contrast with [f] which they do have.

The case of [ʃ] in Dutch is different from [x] in English or [θ] in Dutch where the fricatives do not appear in one of the languages' phonological inventory. This is because phonetic [ʃ] is found in Dutch just as it is in English. For example, the CELEX Dutch lexicon encodes [ʃ] along with [f s x h] as one of the phonetically transcribed fricatives of Dutch. Indeed, Nooteboom & Cohen (1976, p. 144) transcribe Dutch with [ʃ], arguing that words such as *meisje* [ʃ] "girl", *sjaal* [ʃ] "shawl", *chef* [ʃ] "chef, boss", and *sjouwen* [ʃ] "carry" prove that [ʃ] is a contrastive sound in Dutch - though only found now in borrowed words or words of Frisian origin (e.g. *sjouwen*). However, other analysts (Booij, 1995; Cohen et al., 1972; De Groot, 1968) have noted that [s] and [ʃ] do alternate in diminutive forms like "girl" above (*poes* [s] "cat" - *poesje* [ʃ] "kitten", *tas* [s]"bag" - *tasje* [ʃ] "small bag") and in connected speech when [s] and [j] are adjacent across word boundaries (*was je* [ʃ] "were you", *zes januari* [ʃ] "January the 6th"). So, on the one hand, Dutch speakers do have extensive experience hearing [ʃ] and in some words [ʃ] does not alternate with [s] making it reasonable to simply represent the word with an underlying phonemic [ʃ]. However, on the other hand, some instances of [ʃ] do alternate with [s] as if [ʃ] is a contextual variant of [s]. One aim in conducting the experiments reported was to determine the perceptual consequences of such phonological ambiguity[2] .

2. Experiment 1: Rated perceptual similarity.

Experiment 1 used a perceptual similarity rating task to study the role of linguistic structure in phonetic similarity without requiring that nonnative speakers be "learners" in the process of acquiring a second language (L2). By not requiring that listeners identify sounds we avoid dependence upon listeners' skill in using the spelling system of a new language and also avoid different levels of familiarity or study of the L2. Previous research (Huang, 2004; Boomershine

[2] Smits et al. (2003) provided part of the answer. They found an asymmetry in [ʃ] and [s] labeling that seems to be related to their phonological status - 14.5% of [ʃ] tokens (in VC diphones when the vowel and only 1/3 of the fricative was played in a gating task) were identified as "s", while only 3.3% of [s] tokens were identified as "sj". Interestingly, a similar though less striking asymmetry was also observed for [x] and [h] - 5.9 of [x] were identified as "h" while less than 1% of [h] were identified as "x".

et al. in press) has shown that the perceptual similarity rating task does reveal cross-linguistic differences that relate to phonological differences between languages. Phonetic similarity has also been studied using an “implicit” labeling task in which listeners hear three tokens and indicate whether the middle token is more similar to the first or the third in the triad (e.g. Harnsberger, 2001). We feel that the similarity rating task is preferable to this AXB identification task because it is less memory intensive and is more efficient in that listeners provide a gradient, more informative, response on each trial.

2.1. Method

Participants. Sixteen American English speakers (5 male, 11 female) participated in experiment 1. These participants were college undergraduate students who received partial course credit for their participation in the experiment, and none of them reported any past or present speech or hearing disorders.

Twelve Dutch speakers (6 female, 6 male) participated in experiment 1. Their ages and number of years in the US are listed in table 1. The Dutch participants were paid $10 for their participation and none of them reported any past or present speech or hearing disorders.

Table 1. Characteristics of the Dutch listeners in Experiment 1. AOL = Age when listener started English instruction.

listener	age	AOL	Gender	years in US
501	60	NA	F	<1
504	27	12	M	4
505	25	NA	M	4
508	19	NA	M	4
509	24	NA	F	5
001	18	8	M	2
002	24	8	M	<1
003	40	12	F	11
004	29	12	F	4
005	27	10	F	<1
006	18	6	F	12
007	19	7	M	<1

Stimuli. Eighteen disyllabic vowel-fricative-vowel stimuli were used in this experiment. They were composed of the six fricatives [f θ s ʃ x h] embedded in three vowel environments [ɑ_ɑ], [i_i], and [u_u]. The first author was recorded saying multiple instances of the eighteen disyllabic sequences that result from placing each of the six fricatives into the three vowel environments. We then selected for the experiment one instance of each VCV so that the tokens were approximately matched on intonation pattern (H* accent on the first syllable and LL% over the second syllable), and duration. Table 2 shows the vowel and fricative durations of the stimuli.

Table 2. Durations (in milliseconds) of the first vowel (V1), the fricative (Fric), second vowel (V2), and the total duration of the stimulus, for each of the stimuli used in the rating and discrimination conditions.

	V1	Fric	V2	total
afa	186	141	200	527
ifi	235	144	154	533
ufu	198	171	209	578
atha	185	157	148	490
ithi	172	204	224	600
uthu	175	140	224	539
asa	174	187	223	584
isi	184	192	217	593
usu	195	165	189	549
asha	180	166	206	552
ishi	181	171	241	593
ushu	175	177	218	570
axa	156	168	157	481
ixi	189	186	222	597
uxu	174	161	205	540
aha	180	148	212	540
ihi	215	144	221	580
uhu	163	160	206	529

Procedure. Taking all pairs of the six fricatives gives 6^2=36 pairs, with 6^2 - 6=30 DIFFERENT pairs and 6 SAME pairs for each of the 3 vowels [i], [a], and [u]. The SAME pairs were each presented twice so that there were 42 trials per vowel (30 DIFFERENT pairs and 12 SAME pairs). Listeners heard each of the AX trials (42 trials per vowel) three times for a total of 378 trials, with an interstimulus interval of 100 ms between the A and X stimuli. The 378 trials were randomized separately for each listener. Listeners had 5 seconds to respond with a button press rating the pair on a 5 point scale from “very similar” (1) to “very different” (5). They were not given feedback. The session began with four practice trials after which listeners were given a chance to ask questions about the task before proceeding to the test trials.

2.2. Results

The rating scores were analyzed in a repeated measures analysis of variance with the between-listeners factor native language (English vs. Dutch) and the within-listeners factors vowel (/i/, /a/, or /u/) and fricative pair (15 comparisons). The vowel main effect was significant ($F[2,52] = 65.3$, $p < 0.01$) as was the fricative pair main effect ($F[14, 364] = 94.5$, $p<0.01$). The fricative pair by vowel interaction was also reliable ($F[28, 728] = 15.5$, $p < 0.01$). We also found a fricative pair by language interaction ($F[14,364]=3.8$, $p < 0.01$), and the three-way, pair by vowel by language, interaction was also significant ($F[28,728]=1.6$, $p<0.05$).

Figure 1 shows the fricative pair by language interaction, and in this figure we see that the main points of difference between the Dutch and English listeners were with pairs that contrasted [s] and [ʃ] (labeled “s_sh” on the horizontal axis of the graph), pairs that contrasted [s] and [θ] (labeled “s_th”), and pairs that contrasted [ʃ] and [θ] (labeled “sh_th”).

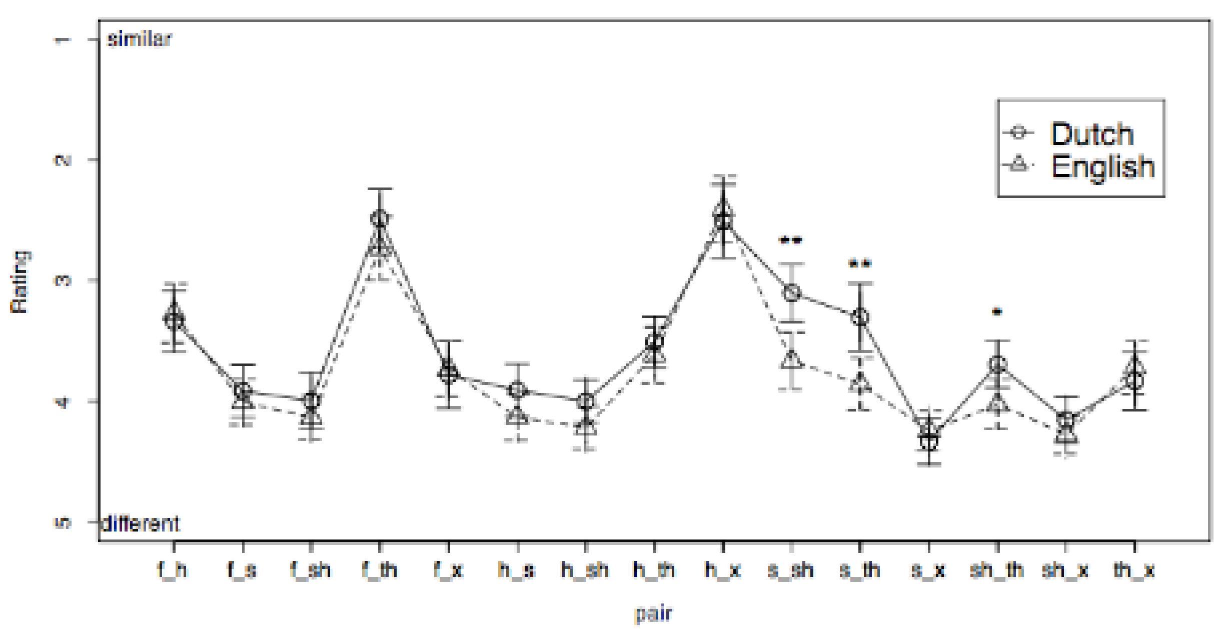
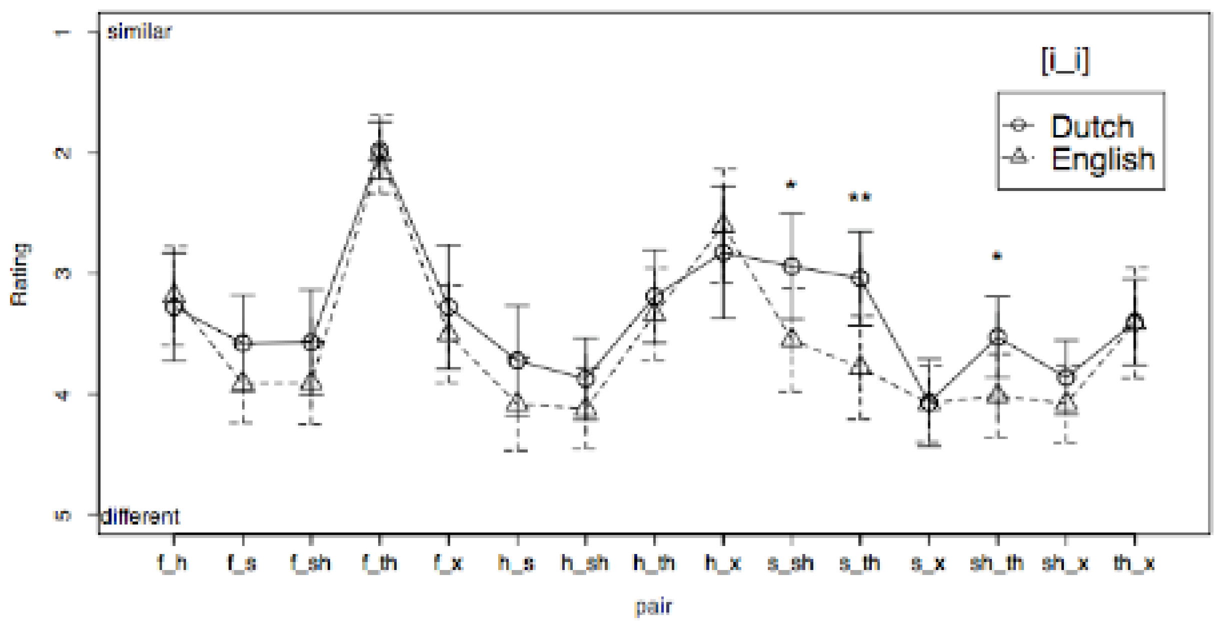


Figure 1. Results of Experiment 1 showing the language by fricative pair interaction. Rating scores given by Dutch-speaking listeners are plotted with a solid line and an open circle, while scores given by English-speaking listeners are plotted with a dashed line and open triangles. In the pair labels, "sh" is used to indicate [ʃ] and "th" is used for [θ]. Pairs marked with "**" or "*" showed a significant language difference in a planned comparison at $p<0.01$ and $p<0.05$ respectively.

Figure 2 shows the pair by vowel by language interaction. The three pairs for which Dutch and English listeners differed in the pair by language interaction also differed in the [i_i] environment. In the [u_u] environment the only language effect was for the s/θ pair. While in the [a_a] environment Dutch and English listeners differed only for the s/ʃ pair.



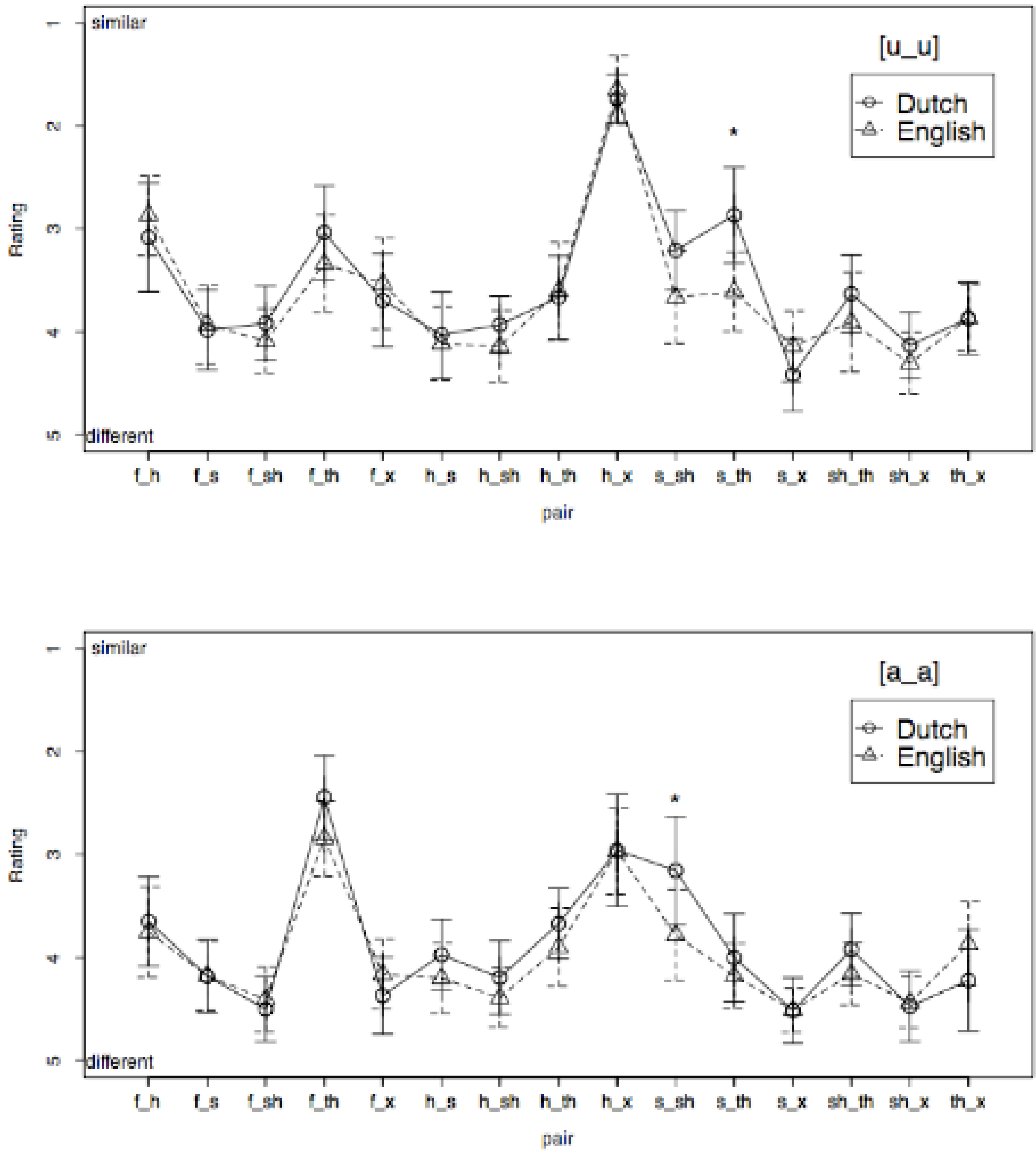


Figure 2. Results of Experiment 1 for the vowel by pair by language interaction. Pairs for which Dutch and English listeners differed are marked as in Figure 1.

2.3. Discussion

In this experiment, we found that Dutch listeners rated [s] and [ʃ] more similar to each other than did English listeners. This pattern was reliably present for [isi]/[iʃi] and [asa]/[aʃa] and we observed a trend in this direction for [usu]/[uʃu]. This finding accords with a growing number of studies that suggest that phonological alternation influences speech perception by reducing contrast between phones that stand in alternation with each other. Of interest here is the phonological ambiguity of [s] and [ʃ] in Dutch, because they contrast in some lexical pairs and alternate in others. The results of this similarity rating experiment suggest that phonological alternation has a powerful impact on speech perception. For our Dutch listeners, we found the perceived phonetic difference between [s] and [ʃ] to be reduced, relative to the phonetic difference reported by English speakers, both in a context where [s] becomes [ʃ] (e.g. before [i]) and in environments where [s] and [ʃ] may contrast (e.g. before [a] in *sjaal* "shall" and *saal* a pronouncable nonword). It is interesting that perceptual distance was reduced even though there are forms in Dutch where [s] and [ʃ] do not alternate.

In addition to this effect of phonological alternation, the different phonological inventories of Dutch and English are also related to differences between Dutch and English listeners' perception of [θ]. Dutch listeners rated [θ] as more similar to [s] than did English listeners both for the [isi]/[iθi] and the [usu]/[uθu] pairs. Dutch listeners also rated [iʃi]/[iθi] as sounding more similar than did English listeners. In these pairs, Dutch listeners found [θ] to be less distinct than did English listeners. It may be that our Dutch listeners were attending more to the vowel formant transitions of [θ] than to the fricative noise (McGuire, 2007). Interestingly, this effect of phonological inventory was not found across the board for all pairs involving [θ]. For example, we might have expected [θ] and [f] to be more confusable for Dutch than for English listeners and though the average ratings do trend in this direction in [ufu]/[uθu] and [afa]/[aθa] these differences were not significant. It might be that for highly confusable pairs such as [f]/[θ] the rating task is not sensitive to language differences because a floor effect on rating scores obscures any differences between the Dutch and English listeners. Note in support of this interpretation that the [f]/[θ] pairs that show the largest (though still nonsignificant) Dutch vs. English differences have higher, more "different" average ratings.

Another inventory difference that led us to expect perceptual differences did not have an effect. This is the presence of [x] in Dutch and the lack of a velar fricative in English. The pattern of [x] in the perception results is different from [θ]. While we saw no language difference for [f]/[θ] there were indications that the Dutch and English listeners differed in how they perceive [θ]. With [x] we found that none of the pairs involving [x] showed any difference between the Dutch and English listeners - not the highly similar pair [x]/[h] or any of the the less similar pairs. This suggests that the two groups of listeners were on an even footing when faced with these stimuli, and this may have been due to their non-nativeness for both group of listeners, because the test tokens were produced by a native speaker of American English. Thus, while the [θ] productions were produced by a native speaker of English the [x] productions were not produced by a native speaker of Dutch.

We explored the perceived nativeness of the stimuli in a small additional study. We asked four Dutch speakers and four American English speakers (drawn from the same populations as in experiment 1) to judge the “nativeness” of each of the stimuli used in experiment 1. Listeners used a five-point scale where a rating of 5 indicated that the stimulus was a native sound produced by a native speaker and 1 indicated that the stimulus was not native at all. Each listener judged each stimulus token once.

The average “nativeness” ratings are shown in table 3. Overall, Dutch listeners gave lower average ratings than did American English listeners (2.8 versus 3.3 respectively). For Dutch listeners, both [θ] and [ʃ] received low nativeness ratings while for English listeners only [x] had particularly low nativeness ratings, and even here one token [uxu] sounded native to these listeners - probably being taken as an instance of /uhu/. For the Dutch listeners [x] and [h] had about the same degree of middling non-nativeness - not as native-sounding as the [s] or [f] stimuli, but also not as clearly non-Dutch as [θ]. Perhaps some explanation for the lack of a language difference for [x] lies in the fact that the speaker of these tokens is not a native speaker of Dutch, but listeners’ explicit judgements of how “native” these sounds are doesn’t support this conclusion. We have noted that [x] didn’t sound particularly native to Dutch listeners, but it is also apparent that [θ] didn’t sound particularly native to English listeners. So there is little reason to conclude that the native language of the speaker who produced these tokens explains the lack of a difference between Dutch and English listeners for [x].

Table 3. Average "nativeness" ratings for stimuli of experiment 1 as provided by four speakers of Dutch and four speakers of English. A rating of 5 indicates the token is a native sound produced by a native speaker, and a rating of 1 indicates that the token is not native at all.

		f	h	s	ʃ	θ	x
Dutch	a	4.25	3.25	4.25	2.5	1.5	3.75
	i	3	2.25	3.75	2.25	1.5	2.75
	u	3.25	2.25	3.5	2.75	1	2.33
English	a	5	3.5	4	4	2.5	1.25
	i	2.75	4.5	4.5	3.5	3.5	1.25
	u	3.25	4	4	3	2	3.25

3. Experiment 2 - Speeded Discrimination.

The main results of experiment 1 are clear. Dutch and English listeners judge the phonetic similarity of fricatives differently and these differences are tied to the differing phonologies of Dutch and English - particularly the absence of [θ] in the Dutch phoneme inventory and the presence of phonological alternation between [s] and [ʃ] in Dutch. These are effects observed when listeners are allowed to introspect about the sounds presented in an experimental trial. Experiment 2 was designed to by-pass this level of conscious introspection in an attempt to observe possible language effects in a less explicitly linguistic task.

There is evidence suggesting that linguistic experience affects speech perception by warping low-level auditory processing (Guenther, et al., 1999). For example, Boomershine et al. (in press) found that Spanish-speaking and English-speaking listeners showed reaction time differences in a speeded discrimination task that mirrored differences found in a similarity rating task. As we argue below, response patterns in the speeded discrimination task are much more likely to show the effects of auditory/perceptual warping than are responses in the similarity rating task. Similarly, Huang (2004) found language differences for tone perception in a low-uncertainty fixed discrimination task. Krishnan et al. (2005) have also found that the brain stem "frequency following response" to Mandarin tones is different (more accurate and less variable) for speakers of Mandarin than for speakers of English. Together, these findings suggest that linguistic experience may influence speech perception by altering auditory response to speech by tuning the auditory path way to be particularly sensitive to the phonetic patterns that are used in one's native language.

Experiment 2 was designed to test for linguistic differences in lower-level auditory processing in fricative perception by Dutch and English-speaking listeners. Additionally, we noted in experiment 1 that Dutch and English listeners responded with very similar rating judgments for pairs ([f]/[θ] and [x]/[h]) which are contrastive in one language and not in the other. So this experiment was designed to explore the auditory basis of the similarities of the Dutch and English listeners of experiment 1.

3.1. Methods

Participants. Nineteen American English speakers (12 female, 7 male) participated in experiment 2. These participants received partial course credit for their participation in the experiment, and none of them reported any past or present speech or hearing disorders. Data from two participants (1 male, 1 female) were removed because English was not their native language.

Fifteen Dutch speakers (9 male, 6 female) participated in experiment 2. Their age at the time of the experiment, age at the onset of English language instruction, and number of years in the US are listed in table 4. The Dutch participants were paid $10 for their participation and none of them reported any past or present speech or hearing disorders. Nine of the Dutch participants in this experiment also participated later in experiment 1 - with a gap of about 3 months between participation in experiment 2 and experiment 1. Dutch listener number 10 was born and grew up in the US speaking Dutch in the home. This listener rated her proficiency in speaking and hearing Dutch as four on a four-point scale, and three for reading and writing.

Table 4. Characteristics of the Dutch listeners in Experiment 2. AOL = Age when the person started English instruction. * = these listeners also participated in Experiment 1.

listener	age	AOL	Gender	years in US
501*	60	NA	F	<1
504*	27	12	M	4
505*	25	NA	M	4
508*	19	NA	M	4
509*	24	NA	F	5
001*	18	8	M	2
002*	24	8	M	<1
003	30	NA	M	4
004	22	10	M	<1
005*	22	12	F	<1
006*	29	12	F	4
007	19	7	M	<1
008	18	7	M	<1
009	27	10	F	<1
010	21	0	F	21

Stimuli. The same stimuli that were used in Experiment 1 were also used in this experiment.

Procedure. The speeded AX discrimination task that we used in experiment 2 has a low memory load with a short 100 ms ISI (Pisoni & Tash, 1974). Additionally, Fox (1984) showed that lexical effects in speech perception are eliminated by fast responding. He found that when listeners responded within 500 ms. of the onset of a stimulus there was no lexical effect in continuum labeling (Ganong, 1980). Therefore, in this experiment listeners were given an AX discrimination task, with a 100 ms ISI and a 500 ms. response deadline. This condition was designed to elicit responses that tap a low-level auditory/perceptual representation of speech.

The listener's task was to identify the last member of a pair of stimuli (X) as the "same" as (physically identical) or "different" from the first member (A) of the pair. The stimuli were presented in the clear (no added background noise) at a comfortable listening level with an interstimulus interval of 100 ms. For most comparisons listeners could achieve almost perfect performance in this task, and the overall performance across all listeners and stimulus pairs was 95% correct. Reaction time was measured for each response, and following Shepard et al. (1975), Nosofsky (1992) and others, reaction time was taken as a correlate of perceptual distance, where longer responses to "different" pairs are taken as an indication that it was relatively difficult to hear the difference between the stimuli, and short reaction times indicate that the difference was more salient. Reaction time was measured from the onset of the fricative noise in the second VCV stimulus of each pair.

Listeners in the speeded discrimination task were given feedback on every trial. They were shown their reaction time (from the onset of the X stimulus) and their overall percent correct score). We emphasized speed of responding, asking listeners to keep their reaction times to 500 ms. or less.

3.2. Results and Discussion

In a repeated measures analysis of variance of the log reaction times for correct "different" responses, there was a vowel main effect ($F[2, 11] = 18.9$, $p<0.01$), a fricative pair main effect ($F[14, 393] = 15.8$, $p < 0.01$), and a pair by vowel interaction ($F[28, 888] = 3.4$, $p< 0.01$). None of the interactions involving the language of the listener reached significance, though the pair by language interaction was close ($F[14, 393]=1.6$, $p=0.071$). Visual inspection of the pair by language interaction indicated that this marginally significant interaction was due to longer reaction times for American English listeners for the hardest pairs [f]/[θ] and [h]/[x], rather than any pattern that matches the language by pair interaction that was found in experiment 1 (Figure 1). This nominal difference between language groups may be an indication that the Dutch listeners were trying harder to keep their reaction times under 500 ms. even when the discrimination judgment was difficult to make.

Some recent research has found effects of the listener's language in both a similarity rating task and in an AX discrimination task (Boomershine et al., in press; Huang, 2004), but in this study we found that the listener's language only affected the results of our similarity rating task - the AX discrimination reaction times showed no significant effect of listener's language. One difference that may be important for understanding the difference between these earlier studies and this one is that in this speeded discrimination experiment we emphasized speed of responding. The observed reaction times (figure 3) were between 500 and 800 ms. This is faster than the reaction times seen in Boomershine et al. for consonant discrimination where a language effect was found in AX discrimination. On the other hand, Huang's listeners also produced reaction times between 500 and 800 ms. but nonetheless showed a language effect. The fact that highly salient lexical tones were being distinguished may have something to do with her finding a language effect even for fast responding, but more research is needed to understand the experimental situations that will result in language effects in experiments like these.

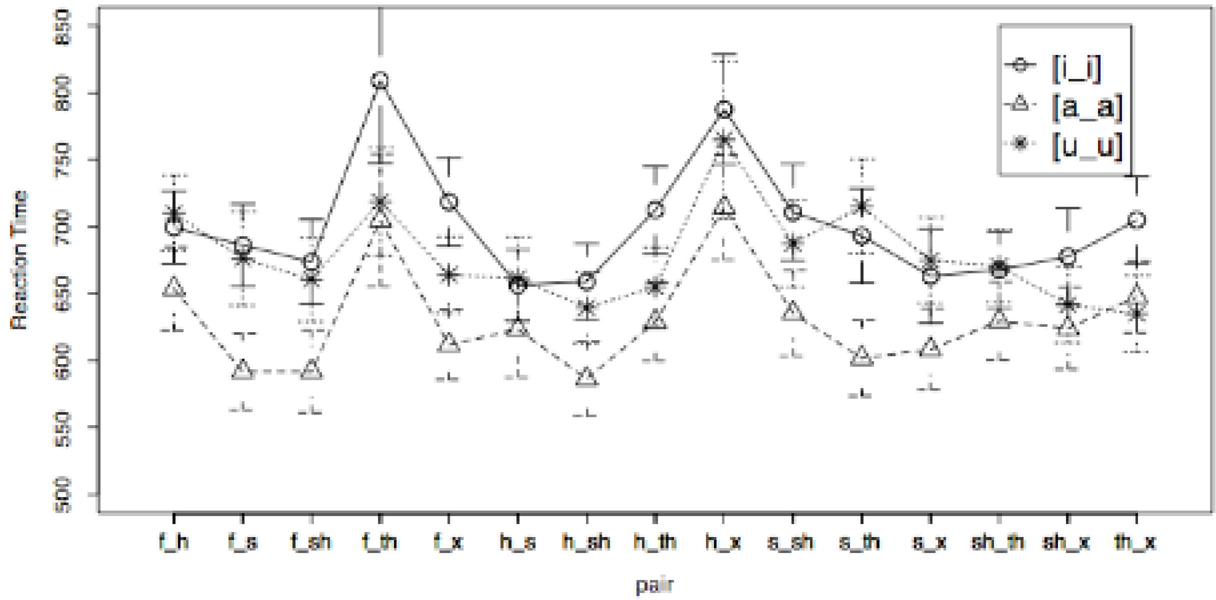


Figure 3. Reaction times for correct "different" responses in experiment 2 averaged over the native language of the listener. Reaction times for different consonant pairs (regardless of which consonant was in the X position) is shown on the horizontal axis, and pairs in the [i_i] environment are plotted with open circles, pairs in the [a_a] environment are plotted with open triangles, and pairs in the [u_u] environment are plotted with asterisks.

The experiment 2 "pair by vowel" interaction is shown in Figure 3. The parallels between the reaction time data shown in this figure and the average similarity ratings that were obtained in experiment 1 are striking. For example, in the rating data (Figure 2) we found that the phonetic difference between [uhu] and [uxu] was smaller than the phonetic difference [ufu] between [uθu], while in the [i_i] environment the opposite trend was found - the difference between [ihi]/[ixi] was larger than the difference between [ifi]/[iθi]. This pattern is also apparent in the reaction time data in figure 3. RT for "different" responses to [uhu]/[uxu] were longer than to [ufu]/[uθu] while RT for "different" responses to [ifi]/[iθi] was longer than to [uhu]/[uxu]. To quantify the similarity between the results of experiments 1 and 2 we calculated correlations between the RT data in figure 3 and the rating data in figure 2 comparing data from vowel environments across the experiments. These correlations are shown in Table 5. This table shows that, for each vowel environment, the similarity rating data that best fits the "different" reaction time data is the data found in the same vowel environment.

Table 5. Correlations between (log) reaction time in Experiment 2 and rated similarity in Experiment 1 for pairs in different vowel environments.

		Phonetic similarity		
		a_a	i_i	u_u
	a_a	-0.86	-0.80	-0.77
Reaction time	i_i	-0.90	-0.96	-0.80
	u_u	-0.68	-0.63	-0.87

In summary, Experiment 2 has two main results. The first is that we found no consistent effect of the language of the listener. Dutch and English listeners showed similar reaction time patterns in this experiment, suggesting that this experiment was successful in measuring the raw auditory distances between these stimuli without influence from the listener's native language. The second main finding is that the reaction time patterns found in this experiment are highly correlated with the phonetic similarity ratings found in experiment 1. These findings suggest that (1) the language differences found in experiment 1 are not due to low-level linguistic warping of the auditory system, though previous research clearly demonstrates that such warping does occur in some circumstances, and (2) that responses in the phonetic similarity rating task of experiment 1 were highly influenced by raw auditory similarity.

4. Conclusion

Best (1995) posited that during perception foreign sounds are compared with the phonetic inventory of the speaker's native language and "assimilated" to the native inventory in one of several ways. For example, the foreign [θ] for Dutch listeners appears to have been assimilated to [s], perhaps on the basis of similarity in their vowel formant transitions, even though for both Dutch and English listeners [θ] is more auditorily similar to [f]. (In addition to similarity in the vowel formant transitions of [θ] and [s], here it may also be important to note that phonetic similarity may encompass visual, motor, and auditory properties.) The perceptual assimilation process by which phonetic input is related to the listener's linguistic knowledge involves an initial comparison (in this or any other plausible model) that must rely on a mechanism that permits foreign sounds and native sounds to be phonetically compared. Whether the comparison mechanism operates using an auditory or gestural code (as Best suggests) is open to debate, but the findings of this study contributes to the emerging consensus that linguistic/phonological knowledge is phonetically very rich.

However, cross-linguistic phonological differences extend far beyond basic differences in phoneme inventories. For example, while Dutch and English have an [ʃ] sound, [s] and [ʃ] are always contrastive in English, but in Dutch [ʃ] is often derived from [s] and alternates with [s] in related forms such as [tas] "bag" and [taʃə] "little bag". This linguistic difference is reflected by listeners' similarity rating judgments in experiment 1, thus emphasizing that phonetic similarity is related to phonological patterning. So in addition to accounting for cross-linguistic perceptual differences that arise from the inventory of phonological contrasts in a language we must be able to account for effects that show sensitivity to more complex phonological patterns of contrast and alternation.

The experiments described in this report highlight three components of phonetic similarity - auditory distance, inventory of contrasts, and patterns of alternation. These components of phonetic similarity, which arise from the physics of speech and the deployment of speech sounds in a language's lexicon, form a phonetic basis for phonological features.

We found in these experiments that the perceptual distance between tokens in a similarity rating experiment is influenced by the <u>raw auditory contrast</u> between the two stimuli. The fact that rating judgments from experiment 1 were so highly correlated with reaction times in experiment 2 (together with the assumption that reaction time was a prelinguistic measure of auditory contrast) lends support to the notion that the phonetic similarity judgments in experiment 1 were largely based on raw auditory similarity. Though English does not have [x] as a phonological category, English native-speakers, like Dutch native-speakers rated it as more similar to [h] than to other sounds. Similarly, the Dutch speakers' lack of [θ] in their native inventory did not keep them from patterning with English listeners in rating [θ] as more similar to [f] than to other sounds. This component of phonetic similarity may be considered linguistically universal because it provides a prelinguistic basis for phonetic similarity.

Beyond investigating the acoustic basis for distinctive features, the experiments here also lead to the conclusion that the presence or absence of a sound in the words of a language also influences phonetic similarity. We saw this particularly with [θ] which is not used in Dutch. Although [θ] and [f] were judged to be equally similar by English-speaking and Dutch-speaking listeners (compare this with findings for English [r]/[l] presented to Japanese speakers) we did find that Dutch speakers judged [θ] to be more similar to [s] and [ʃ] than did English speakers. Although a definitive explanation of this result cannot be given at this time, we could speculate that our Dutch listeners as bilinguals in English were aware of [θ] and called up visual and perhaps motor phonetic experience to supplement their limited auditory experience with this sound - reporting a combination of visual, motor, and auditory phonetic similarity. Another speculation about this finding is that our Dutch listeners are aware of some conventionalized non-native pronunciations of English words in which [θ] is replaced by [s] and thus report as a part of phonetic similarity a conventional relationship between the sounds. Regardless of what explanation is correct, the results of the experiments make it clear that the <u>inventory of contrastive sounds</u> in the native language influences phonetic similarity.

The third component of phonetic similarity that we have seen in these experiments is related to the <u>patterns of alternation</u> of sounds in language. The observation here, which has been made before (Harnsberger, 2001; Hume & Johnson, 2003; Boomershine et al., in press), is that even when listeners have experience with sounds - having to recognize them and produce them in their native language - when the sounds alternate with each other in the language's phonology they are phonetically more similar than when they don't. Dutch listeners judged [s] and [ʃ] to be more similar to each other than did English listeners. This case is interesting also because of the phonological ambiguity of [ʃ] in Dutch. It alternates with [s] in morphologically related words and in fast speech pronunciation, but it also appears in non-alternating forms in borrowed words and dialectal variants. Our result is that people definitely do not perceive speech in terms of phonemes (a view that has been attributed to the American structuralists). For example, Dutch listeners do hear the difference between [s] and [ʃ], but phonological alternation reduces the phonetic separation of the sounds.

Our conclusion leads us to a somewhat circular view of the relationship between phonetics and phonology (see also Hume & Johnson, 2001). We conclude that phonetic similarity, while largely determined by physical properties of sounds (here particularly acoustic properties), is also partly determined by the listener's linguistic experience of sounds in language - the inventory of contrastive sounds and the patterns of phonological alternation in his/her native language. The circularity is that this phonetic similarity which is partly determined by phonological facts is in turn a basis for phonological patterns. Seen from a purely synchronic point of view it makes no sense to speak of phonetic similarity as simultaneously the result *and* cause of phonological patterning, but when phonology is viewed as the product of a complex diachronic process the circularity of our perspective is just a reflection of the cycle of language acquistion and transmission. Phonological patterns contribute to phonetic similarity for the present generation, which is then a basis for the development of new phonological processes in a later generation. We have shown here that phonetic similarity is at least partly determined by phonological patterning, and we have gone some way toward measuring the degree to which phonetic similarity is determined by three different components of similarity. If distinctive features (and the natural classes they capture) are grounded at all in phonetic similarity and if phonetic similarity does at all derive from phonological patterning, then the circularity of our view is correct, the system is circular. Thus, to understand phonology we must study it from a diachronic perspective.

Acknowledgements

This research was supported by NIH grant R01 DC004421-05. Earlier versions of this project were presented to the Texas Linguistics Society in 2003, and the Fall 2006 meeting of the Acoustical Society of America, and to the Department of Linguistics and the Department of Psychology at UC Berkeley and the Department of Linguistics at Stanford University. We appreciate very much the comments and suggestions given by colleagues at these presentations. Discussions with Sheila Blumstein, Grant McGuire, and Sharon Inkelas have also improved the paper, though of course the remaining mistakes are ours. A preliminary report of a portion of these results was published in Johnson (2004).

References

Archangeli, D. & Pulleyblank, D. (1994) *Grounded Phonology*. Cambridge, MA: MIT Press.

Best, C.T. (1995) A direct realist perspective on cross-language speech perception. In *Speech perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, edited by W. Strange (York: Timonium, MD), pp. 167-200.

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 (2)**: 775-794.

Booij, G. (1995) *The Phonology of Dutch*. Oxford: Oxford University Press.

Boomershine, A., Hall, K. C., Hume, E., Johnson, K. (in press). The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher & K. Rice (eds.), *Phonological Contrast: Perception and Acquisition*. (New York: Mouton de Gruyter).

Caramazza, A., Yeni-Komshian, G., Zurif, E, and Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* **54**, 421-428.

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English*. New York: Harper & Row.

Cohen, A., Ebeling, C.L., Fokkema, K. & van Holk, A.G.F. (1972): *Fonologie van het Nederlands en het Fries: inleiding tot de moderne klankleer*. 's-Gravenhage: Martinus Nijhoff.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language* **36**, 406-421.

Elman, Jeffrey, Randy Diehl, and Susan Buchwald. 1977. Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America* **62**, 971-974.

Flege, J.E. (1995) Second language speech learning: Theory, findings, and problems. In *Speech perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, edited by W. Strange (York: Timonium, MD), pp. 167-200.

Flege, J.E., Takagi, N. and Mann, V. (1996) Lexical familiarity and English-languge experience affect Japanese adults' perception of /r/ and /l/. *The Journal of the Acoustical Society of America* **99**, 1161-1173..

Fox, R.A. (1984) Effect of lexical status on phonetic categorization. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 526-540.

Gandour, J. T. 1983. Tone perception in Far Eastern languages. In *Journal of Phonetics,* 11: 149-176.

Ganong, W.F. (1980) Phonetic categorization in auditory word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 110-125.

Goto, Hiromu (1971) Auditory perception by normal Japanese adults of the sounds "l" and "r." *Neuropsychologia, 9*, 317-323.

De Groot, A.W. (1968): *Inleiding in de algemene taalwetenschap*. Groningen: Wolters Noordhof.

Frisch S., Pierrehumbert, J. & Broe, M. (2004) Similarity Avoidance and the OCP. *Natural*

Language and Linguistic Theory 22: 179-228.

Guenther, F., Husain, F.T., Cohen, M.A. and Shinn-Cunningham, B.G. (1999) Effects of categorization and discrimination training on auditory perceptual space. *The Journal of the Acoustical Society of America,* **106**, 2900-2912.

Hansson, G. (2001) *Theoretical and Typological Issues in Consonant Harmony*. Ph. D. dissertation, UC Berkeley.

Harnsberger, James (2001) The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics, 29*, 303-327.

Huang, T. (2004) *Language specificity in auditory perception of Chinese tones*. Ph. D. dissertation, Ohio State University.

Hume, E. & Johnson, K. (2001) A model for the interplay of speech perception and phonology. In Hume, E. & Johnson, K. (eds) *The Role of Speech Perception in Phonology*. New York: Academic Press. pp. 3-26.

Hume, E. & Johnson, K. (2003) The impact of partial phonological contrast on speech perception. *Proceedings of the Fifteenth International Congress of Phonetic Sciences* (pp. 2385-2388).

Johnson, K. (1994) Phonetic arbitrariness and the input problem. In P. Keating (ed.) *Phonological structure and phonetic form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press. pp. 325-330.

Johnson, K. (2004) Cross-linguistic perceptual differences emerge from the lexicon. In Augustine Agwuele, Willis Warren, and Sang-Hoon Park (eds.) *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*. Sommerville, MA: Cascadilla Press. pp. 26-41.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research* **25**, 161-168.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992) Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science* **255**, 606-608.

Ladefoged, P. (1984) 'Out of chaos comes order': Physiological, biological, and structural patterns in phonetics. In M.P.R. Van den Broeke and A. Cohen (eds.) *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications (pp. 83-

95).

Lively, S.E., Logan, J.S. and Pisoni, D.B. (1993) Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America,* **94**, 1242-1255.

Logan, J.S., Lively, S.E. and Pisoni, D.B. (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America,* **89**, 874-886.

MacKain, K.S., Best, C.T. and Strange, W. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics,* **2**, 369-390.

MacKenzie, S. (2005) Similarity and contrast in consonant harmony systems. In Chiara Frigeni, Manami Hirayama, Sara Mackenzie (eds.) *Toronto Working Papers in Linguistics (24): Special Issue on Similarity in Phonology*. pp. 169–182.

Manuel, S.Y. & Wyrick, G.C. (1999). Casual Speech: A rich source of intriguing puzzles. *Proceedings of the XIV International Congress of Phonetic Sciences*.

McGuire, G. (2007) English listeners' perception of Polish alveopalatal and retroflex voiceless sibilants: a report of some pilot data. Unpublished manuscript. Ohio State University.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., and Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, **18**, 331-340.

Nooteboom, S.G. & Cohen, A. (1984): *Spreken en verstaan. Een nieuwe inleiding tot de experimentele fonetiek*. Assen: Van Gorcum.

Nosofsky, R.M. (1992) Similarity scaling and cognitive process models. *Ann. Rev. Psychol*. **43**, 25-53.

Pisoni, D. B. & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* **15**, 285-290.

Rose, S. & Walker, R. (2004) A typology of consonant agreement as correspondence. *Language 80:3,* 475-531.

Shepard, R.N., Kilpatric, D.W. and J.P. Cunningham (1975) The internal representation of numbers. *Cognitive Psychology* **7**, 82-138.

Smits, R., Warner, N., McQueen, J.M. & Cutler, A. (2003) Unfolding of phonetic information over time: A database of Dutch diphone perception. *Journal of the Acoustical Society of America* **113**, 563-574.

Steriade, D. (2001) Directional asymmetries in place assimilation: A perceptual account. In *The*

Role of Speech Perception in Phonology, edited by E. Hume and K. Johnson. (New York: Academic Press), pp. 219-250.

Steriade, D. (2003). Knowledge of perceptual similarity and its phonological uses: evidence from half-rhymes. In M. J. Solé, D. Recasens & J. Romero (eds). *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Universitat Autònoma de Barcelona. 363–366.

Strange, W. and Dittmann, S. (1984) Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics* **36**, 131-145.

Takane, Y. and Sergent, J. (1983) Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika* **48**, 393-423.

Terbeek, D. (1977). *A cross-language multidimensional scaling study of vowel perception*. PhD dissertation, UCLA. (UCLA Working Papers in Phonetics #37).

Werker, J.F. & Tees, R.C. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* **7**, 49-63.

Williams, L. (1977) "The perception of stop consonant voicing by Spanish-English bilinguals", *Perception and Psychophysics* **21**, 289-297.

Yamada, R., Tohkura, Y., and Kobayashi, N. (1992) Effect of word familiarity on non-native phoneme perception: Identification of English /r/,/l/ and /w/ by native speakers of Japanese. In *Second Language Speech*, edited by A. James and J. Leather (The Hague: Mouton de Gruyter).