Cross-linguistic Perceptual Differences Emerge from the Lexicon

Keith Johnson
Ohio State University

1. Introduction

The anatomy and function of the peripheral auditory system differs from person to person (sometimes to a clinical degree), but such differences are probably not correlated with linguistic knowledge. For example, the distribution of French speakers' peripheral auditory sensitivities are probably no different from the distribution of auditory sensitivities of speakers of any other language. That is, there is no reason to expect psychophysical thresholds for simple or complex stimuli to vary from language to language. So, there is probably a substrate of auditory-perceptual capacity underlying the perception of speech that is the same for all languages. Such a language universal perceptual map plays an important role in some approaches to understanding how processes of speech perception may influence language sound systems (Steriade, 2001). Mismatched negativity studies also show both language-specific and language-nonspecific patterns of response (Dehaene-Lambertz et al., 2000).

However, it is commonly observed that people who speak different languages process speech sounds differently. There are language-specific phonetic prototypes which function as "perceptual magnets" (Kuhl, et al., 1992), and which may arise from self-organizing perceptual maps (Guenther & Gjaja, 1996). And there are effects of native language knowledge in the perception of a second language (Best, 1995; Flege, 1995). And babies show signs of perceptual reorganization during early prelinguistic development (Werker & Tees, 1984). This all suggests that linguistic experience "tunes" speech perception, so that the language-universal auditory capabilities are warped to serve the contrasts of the native speaker's language.

The theme of this paper is that the language-specific linguistic tuning of perception is fundamentally compatible with a language-universal perceptual map for speech and that the perceptual map need not be altered at all to account for cross-linguistic differences in speech perception. I will demonstrate in a set of simple simulated speech perception experiments how language-specific perceptual effects could arise from lexical knowledge without altering the universal perceptual map.

It may be that linguistic experience operates both through contact with lexical memory, as I will suggest below, and by warping low-level auditory processing (Guenther, et al., 1999). By emphasizing the lexical component in this paper I hope to provide some diagnostics and predictions as to how to separate these two kinds of effects. The empirical study reported in section 5 suggests that, in rough outline at least, this "lexical distance" model of cross-linguistic speech perception is correct.

2. An illustration of the approach: /r/-/l/ confusions by Japanese listeners.

Japanese speakers' perception of the American English /r/-/l/ distinction is a frequently cited case of cross-linguistic differences in speech perception. While American English speakers have very good discrimination of words that start with /r/ and /l/ (e.g. *rake* versus *lake*), Japanese speakers cannot discriminate these very well (Miyawaki, et al., 1975; MacKain et al., 1981; Strange and Dittmann, 1984; Logan et al., 1991).

* This research was supported by Grant No. 5 R01 DC04421 from the National Institute on Deafness and Other Communication Disorders.

 Proceedings of the 2003 Texas Linguistics Society Conference, ed. Augustine Agwuele, Willis Warren, and Sang-Hoon Park, 26-41. Somerville, MA: Cascadilla Proceedings Project.

This effect can be simulated by assuming that the only difference between English and Japanese listeners is that they have different lexicons (see Yamada et al. 1992 and Flege et al., 1996 on the role of the L2 lexicon on perception). The auditory-perceptual difference between [r] and [l] was the same for both of the simulated listeners, and the mechanism of perception was also held constant. The only difference is that the simulated English speaker has some words that start with /r/ and some that start with /l/, while the simulated Japanese speaker does not have any words distinguished by these sounds.

The estimated universal perceptual distances that I used in these simulations is shown in table 1. A more rigorous study would find a way to measure these values exactly (Dooling et al., 1995, see also section 5 below), but it is adequate for purposes of demonstrating the possible lexical locus of cross-linguistic perceptual differences (as I am seeking to do here) to offer a reasonable approximation based on acoustic dissimilarity and hold this table of distances constant in the simulations.

Table 1. Auditory-perceptual distances (d_a) assumed in the /r/-/l/ perception simulation.

	w	r	d	l	ɾ
w	-	0.5	0.8	0.6	0.7
r		-	0.6	0.3	0.5
d			-	0.7	0.3
l				-	0.4
ɾ					-

To simulate the influence of linguistic knowledge on speech perception, the model combines two measures of distance - auditory distance and lexical distance (1). Auditory distance (d_a) is read directly off of table 1, and lexical distance (d_l) is calculated over the lexicon. Note that the constant k gates the influence of lexical distance, so that we can simulate conditions in which perceptual response is determined primarily by auditory-perceptual distance (e.g. minimal uncertainty discrimination tasks), as well as conditions in which perceptual response is based more on linguistic knowledge (e.g. high uncertainty discrimination tasks).

$$d = d_a + (d_l \cdot k) \qquad \textbf{(1)}$$

Lexical distance (d_l) is a function of the difference in the lexical activation pattern caused by one of the stimuli (j) and the lexical activation pattern caused by another of the stimuli (i). If the two stimuli provoke basically the same response from the lexicon then the listener will not have much evidence that the two stimuli are "different". However, if the lexical response is quite different for the two then the listener has good evidence that they are "different".

Lexical distance is written in (2) with a_{li} indicating the activation of lexical item l in response to stimulus i and a_{lj} indicating the activation of that same lexical item by stimulus j. Equation (2) indicates that the lexical distance calculation involves taking the sum over all lexical items of activation differences $|a_{li} - a_{lj}|$ scaled by the sum of the activations $(a_{li} + a_{lj})$. Scaling by activation level reduces the influence of lexical differences if the overall match to the lexicon is low. Dividing the sum of scaled differences by the number of lexical items (n) gives us the average scaled difference in lexical activation which is useful in comparisons across lexicons of different sizes.

$$d_{l(ij)} = \frac{\sum_{l}^{n}\left[\left|a_{li} - a_{lj}\right| \cdot (a_{li} + a_{lj})\right]}{n} \qquad \textbf{(2)}$$

Activation of a lexical item by a stimulus (a_{lj}) is computed (3) from the auditory distance between them $d_{a(lj)}$ using the exponential relationship between psychological distance and psychological similarity that was proposed by Shepard (1974). Lexical activation then is the similarity of the stimulus to the lexical item. The constant c is used to widen or sharpen the similarity function.

$$a_{ij} = e^{d_{a(ij)}c} \quad \textbf{(3)}$$

I conducted two simulations of [r]/[l] perception. In the first, the model had a simplified American English lexicon of 12 CV "words" (ri ra ru li la lu wi wa wu di da du) and in the second the lexicon was "Japanese" (ɾi ɾa ɾu wi wa wu di da du) with words that start with a flap, and no words with [l] or [r]. Simulations of an AX discrimination task contrasted all combinations of the stimuli [wa, ra, da, la, and ɾa]. The parameters of the model were k = 9 and c = 2. A fairly wide range of parameter values gave the same pattern of data.

Table 2. Simulated American English perceptual distances between stimuli in a discrimination task calculated according to equation (1). The lexicon for this simulation (ri ra ru li la lu wi wa wu di da du) has "words" that contrast [r] and [l] and no words beginning with [ɾ]. Values for lexical contrasts are underlined.

	wa	ra	da	la	ɾa
wa	-	2.06	2.33	2.17	1.62
ra		-	2.22	1.59	1.41
da			-	2.32	1.01
la				-	1.30
ɾa					-

Table 3. Simulated Japanese perceptual distances (d) produced for an /r/-/l/ discrimination task calculated according to equation (1). The lexicon for this simulation (ɾi ɾa ɾu wi wa wu di da du) has "words" with a flap, but no words contrasting [r] and [l]. Values for lexical contrasts are underlined.

	wa	ra	da	la	ɾa
wa	-	1.52	2.96	1.67	2.85
ra		-	1.74	0.38	1.64
da			-	1.79	2.00
la				-	1.49
ɾa					-

Results for the American English and Japanese lexicons are shown in tables 2 and 3, respectively. The distance values in these tables are on an arbitrary scale that is comparable across the two tables. Notice in table 2, the simulation of an American English listener, that the non-contrastive flap initial stimulus [ɾa] shows generally low distance values when paired with the other stimuli. The pair [da]/[ɾa] has the lowest perceived distance, which accords with the raw perceptual distance value for contrast in table 1. Among words that exist in the lexicon, the pair [ra]/[la] have the lowest distance reflecting the lower auditory distance.

The predicted distance values for Japanese listeners (table 3) are different from those given for American English listeners. In particular, the [ra]/[la] distinction shows a very low distance value, while the [da]/[ɾa] distance is twice as large as it was in the English simulation. The first difference, lower [ra]/[la] discrimination, accords with the published research findings, while the second difference is a prediction given by this model.

3. Noteworthy aspects of the lexical distance model.

What is interesting about these simulations is that we get something comparable to the cross-linguistic perceptual result (reduced [ra]/[la] discrimination for Japanese listeners) with a model in which listeners differ only in terms of their lexicons. In these simulations, the raw auditory perceptibility of the stimulus contrasts did not differ.

There are four key noteworthy features of this "lexical distance" model. First, in this approach, linguistic phonological knowledge is held in a phonetically detailed lexicon. An alternative to this way of representing linguistic knowledge is to list phonemes, their phonetic realizations, phonotactic rules, allophonic rules, and so on, as knowledge bases which are separate from an abstract lexicon. However, because phonetic and phonological generalization are language-specific, it is necessary at some stage during acquisition for there to be a non-abstract lexicon over which such generalizations can be computed, i.e. the lists of language-specific generalizations must be derived from a phonetically detailed store of lexical items. These generalizations have not stabilized for children as old as 10 years (Nittrouer, 1992) and temporary phonological patterns followed for a set of experimental stimuli can effect speech error patterns in adults (Dell, et al., 2000). The lexical distance model uses detailed lexical knowledge directly rather than rely on the results of generalization routines. This is a simpler strategy. It may be that listeners actually use lists of generalizations which encode linguistic knowledge, but these simulations demonstrate that this is not necessarily the case.

Second, the lexical distance model makes use of a perceptual foundation which is language universal and unchanged by linguistic experience. An alternative to this approach would suggest that linguistic experience tunes the auditory/perceptual foundation of speech perception. Again this could be a more correct view. However, things get tricky for such an auditory tuning model when we consider "language set" effects in perception in which bilingual listeners demonstrate different perceptual spaces depending on the language in which they perform a perceptual task (Caramazza, et al., 1973; Elman, et al., 1977; Grainger and Beauvillain, 1987; Dijkstra, et al., 1998).

Third, some features of the lexical distance model seem to be appealing. For example, this model makes predictions beyond the particular linguistic contrast under investigation, [r] versus [l] in this case. The simulation results in table 3 predict generally lowered perceptual contrast for all pairs involving [r] and [l] in Japanese - with the most extreme case [r]/[l] being the most easily testable. In addition, the inclusion of a "lexical importance" constant permits us to model results of experimental tasks which would tend to reduce the listener's reliance on lexical activation thus permitting us to model the reduction of language differences in low-uncertainty discrimination tasks (Strange & Dittmann, 1984; see also Lively et al., 1993; and see also Hickok & Poeppel, 2000).

Fourth, the model incorporates an important distinction between discrimination performance and categorization performance (Guenther, et al., 1999). Contrast is the key contribution of the peripheral auditory system (table 1) while categorization is the key contribution of the lexical activation procedure (equation 1).

4. A model test: simulating "Perceptual assimilation" of three Zulu contrasts.

Best et al. (2001) reported a very interesting test of American English listeners' discrimination responses to three Zulu consonant contrasts that exemplify three types of perceptual assimilation in Best's (1995) perceptual assimilation model.

For American English listeners the Zulu contrast between [b] and [ɓ] is a "single category" assimilation because when they are asked to write down what they hear, American English listeners tend to write the letter "b" with no notation suggesting that one is a better "b" than the other. As would be predicted by the phenomenon of categorical perception, these sounds which get the same identification label should not be very discriminable at least in a "linguistic" discrimination task. Best et al. (2001) found this to be the case. Percent correct discrimination by American English listeners in an AXB task was 66%.

The Zulu contrast between [k^h] and [k'] is similar to this in that American English listeners think of them both as a kind of /k/ but they clearly hear the difference between the aspirated and ejective

tokens, and indicate the difference with some additional mark like an apostrophe or dash. The categorical perception prediction for this contrast is that listeners will be better able to discriminate these two Zulu sounds because, though they are the same at one level of labeling, they differ in their degree of fit to the English /k/ category - a "category goodness" difference. American English listeners detected this difference in an AXB discrimination task 89% of the time.

The voicing contrast in Zulu lateral fricatives, [ɬ] versus [ɮ], is of a different type entirely. Listeners use voiceless fricative symbols to transcribe the Zulu voiceless lateral fricative, and use voiced fricative or "l" to transcribe the Zulu voiced lateral fricative. In a categorical perception model then, AE listeners having mapped the Zulu contrast onto different American English sounds will be able to distinguish them. AXB discrimination results conformed to these predictions. Listeners correctly discriminated the lateral fricatives 95% of the time.

As indicated in this summary, the predictions of the perceptual assimilation model derive primarily from the main observation of categorical perception - that within category discriminations are difficult and between category discriminations are easy. In this approach to cross-linguistic speech perception, the categories are "phones", and because different languages have different phones, the mapping of foreign sounds onto native phones is the key to predicting nonnative speech perception performance.

The lexical distance approach outlined in section 2 above suggests that a universal auditory-perceptual map guides the "assimilation" process (the process that maps foreign sounds onto the native phonology), and that language differences arise solely through the process of lexical activation.

Table 4. Auditory-perceptual distances (d_a) that were used in the Zulu perception simulations.

	p	v	m	b	ɓ
p	-	0.8	0.8	0.6	0.7
v		-	0.4	0.4	0.5
m			-	0.5	0.6
b				-	0.165
ɓ					-

	g	tʃ	p	t	k^h	kʻ
g	-	0.8	0.8	0.9	0.6	0.8
tʃ		-	0.6	0.3	0.4	0.6
p			-	0.4	0.4	0.6
t				-	0.5	0.7
k^h					-	0.21
kʻ						-

	s	z	ʃ	ʒ	l	h	ɬ	ɮ
s	-	0.5	0.4	0.7	1.0	0.6	0.4	0.8
z		-	0.7	0.4	0.8	1.0	0.8	0.4
ʃ			-	0.5	1.0	0.6	0.2	0.8
ʒ				-	0.8	0.8	0.6	0.4
l					-	1.2	0.4	0.2
h						-	0.4	0.8
ɬ							-	0.17
ɮ								-

In order to carry out a simulation of the Best et al. (2001) results I established three auditory-perceptual distance matrices for consonants that are similar to [b]/[ɓ], [k^h]/[k'], and [ɬ]/[ɮ]. These three matrices are shown in table 4. For the simulation, the distance matrices in table 4 were combined into

a single larger consonant distance matrix with all of the cells not shown in the table filled with a distance value of 2.

The values in table 4 are intended to be estimates of universal auditory-perceptual distance. In the absence of objective measurements of these distances, I estimated auditory-perceptual distance from phonetic descriptions of the sounds. The simulation results depend on the values given in the table - if they are realistic, then the simulation may be a reasonable, though simplified, picture of cross-linguistic speech perception. Appendix A contains the source code for the simulation, so the reader can explore the effects of changing the model characteristics.

Note that the auditory distance between [b] and [ɓ] in table 4 (0.165) is almost the same as that for [ɬ] and [ɮ] (0.17) while the distance between [k^h] and [k'] is a little greater (0.21). I adjusted these values by trial and error from the original values that I had assigned to them (0.2, 0.2, and 0.3, respectively) to provide a close fit to the Best et al. (2001) discrimination results. Also note that the distance vectors for the bilabial and velar sounds are correlated with each other. For example, [b] and [ɓ] are both counted as being more similar to [v] than to [p], and [k^h] and [k'] are both counted as being more similar to [p] than to [g]. The distance values for the lateral fricatives [ɬ] and [ɮ] are less well correlated. [ɬ] is counted as being more similar to other voiceless fricatives while [ɮ] is more similar to [l] and the other voiced fricatives.

The lexicon used in this simulation contains CV "words" that exhibit lexical contrasts found in English [ba pa be pe bu pu va ve vu ma me mu ka ke ku ga ge gu tʃa tʃe tʃu sa se su ta te tu za ze zu ʃa ʃe ʃu ʒa ʒu la le lu ha he hu]. Note, that I left out [ʒe]. This models the relatively lower type frequency of [ʒ] in English. The lexical importance constant *k* (in formula 1) was set to 10 and the distance-to-similarity scale factor *c* (in formula 3) was set to 2.

Table 5. Simulated American English perceptual distances (*d*), calculated by equation (1), for stimulus sets that include Zulu contrasts.

	pu	vu	mu	bu	ɓu
pu	-	1.53	1.51	1.30	1.25
vu		-	0.85	0.86	0.85
mu			-	0.98	0.97
bu				-	0.33
ɓu					-

	ga	tʃa	pa	ta	k^ha	k'a
ga	-	1.42	1.44	1.54	1.19	1.18
tʃa		-	1.15	0.72	0.91	0.96
pa			-	0.94	0.93	0.98
ta				-	1.00	1.11
k^ha					-	0.45
k'a						-

	se	ze	ʃe	ʒe	le	he	ɬe	ɮe
se	-	1.03	0.85	1.03	1.62	1.15	0.79	1.26
ze		-	1.25	0.68	1.37	1.56	1.28	0.77
ʃe			-	0.85	1.61	1.13	0.43	1.27
ʒe				-	1.18	1.16	0.87	0.57
le					-	1.78	0.83	0.42
he						-	0.80	1.26
ɬe							-	0.48
ɮe								-

The results are shown in table 5. The main observation to make from this table is that with approximately the same level of auditory contrast (d_a) for [b] versus [ɓ] as for [ɬ] versus [ɮ], the lateral fricative contrast shows a larger perceptual distance value (d) than does the bilabial contrast. Additionally, the velar contrast [k^h] versus [k'] which had a higher auditory distance value than either of these, shows in the perceptual distance results a smaller perceptual distance value than does the lateral fricative distinction. In this case, the raw auditory discriminability pattern was reversed in the context of lexical support for one contrast and the lack of lexical support for the other.

The results in table 5 are compatible with the AXB discrimination results reported by Best et al. (2001). If we allow perceptual distance of 0.5 or higher to produce perfect performance in the AXB task, and calculate the proportion of correct responses as perceptual distance divided by the perfect performance threshold, then we have the comparison of results shown in table 6.

In addition, note that the simulation of [ɬ] and [ɮ] perceptual distances provides a mechanism for predicting patterns of "perceptual assimilation". The token [ɬe] was most similar to [ʃe] while [ɮe] was most similar to [le]. This result is built into the universal perceptual distances matrix (table 4) so I am not making any special claim for it other than that when we do have a set of accurate estimates of auditory distances we will be able to compute patterns of perceptual assimilation and then compare the predictions to empirical data gathered from listeners.

Table 6. A comparison of the simulation perceptual distances (d), a conversion of these distances into probability of a correct discrimination response, and the % correct discriminations found by Best et al. (2001).

pair	d	d/0.5	%correct
ɓu bu	0.33	0.66	66%
k'a k^ha	0.45	0.9	89%
ɮe ɬe	0.48	0.96	95%

There is one other aspect of the Best et al. (2001) results that seems to be captured in this simulation. In the AXB task, listeners are asked to judge whether X is the same as A or B. Best et al. found that listeners' performance was better when the tokens judged to be the "same" were more English-like (bu, ka, or ɬe) than when they were the less English-like member of the pair (ɓu, k'a, or ɮ). That is, listeners were more accurate in trials like [bu][bu][ɓu] than they were in trials like [ɓu][ɓu][bu]. One possible explanation for this is that the more English-like stimuli give rise to higher lexical activations (as shown in table 7). It may be that a stronger lexical response gives listeners more information upon which to base their "same" responses.

Table 7. Average lexical activation, and maximum lexical item activation for Zulu English-like and non English-like stimuli. Lexical activation was calculated using equation (3).

	average activation		*maximum activation*	
pair	*not native-like*	*native-like*	*not native-like*	*native-like*
ɓu bu	0.06	0.08	0.72	1.00
k'a k^ha	0.06	0.9	0.66	1.00
ɮe ɬe	0.08	0.09	0.67	0.67

Finally, the relative success of this simulation of American English listeners' perception of Zulu consonant contrasts makes a very specific testable prediction. This simulation was based on auditory-perceptual distances (table 4) in which it was assumed that the [b]/[ɓ] distance is about the same as the [ɬ]/[ɮ] distance, and that the Zulu [k^h]/[k'] auditory-perceptual distance is greater than these. The

prediction then is that this pattern of auditory discriminability will be obtained in a minimal-uncertainty discrimination task (such as the fixed AX discrimination task) in which "context" effects such as linguistic knowledge are minimized (Watson & Kelly, 1981; Braida & Durlach, 1988; Macmillan, 1987).

5. An empirical test: Dutch and English perception of fricatives.

This section reports the results of an experiment that was designed to test a couple of key assumptions of the lexical distance model. The first aspect of the model under test is the claim that there exists a language-universal auditory perceptual space that is unmodified by linguistic experience. As mentioned in section 2, this claim may be too strong because there is reason to believe that the auditory cortex mapping of sound is altered by experience (Guenther et al., 2001), however the results of the experiment reported here do provide support for the idea of a universal perceptual space.

To test for a universal auditory/perceptual base in speech perception I used a speeded AX discrimination task. In most speech perception research it is assumed that a "linguistic" speech perception task should be used. I am here putting "linguistic" in quotes because I'm not sure what exactly is meant by the researchers who use this descriptor (e.g. Best, et al., 2001). The tasks typically used to measure auditory discriminability of speech sounds have a fairly high memory load (the ABX and AXB paradigms) and stimuli are presented with relatively long interstimulus (ISI) intervals. The speeded discrimination task that I used in this experiment has a low memory load (the AX task) with a short 100 ms ISI. Additionally, Fox (1984) showed that lexical effects in speech perception are eliminated by fast responding. When listeners responded within 500 ms. of the onset of a stimulus there was no lexical effect in continuum labeling (Ganong, 1980). Therefore, one condition in the experiment reported here had listeners responding in 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 language-independent auditory perceptual representation of speech, as hypothesized in the lexical distance model.

In a second condition, listeners were asked to rate the subjective similarity of the same stimuli that had been presented in the fixed discrimination task. The lexical distance model predicts that the listener's native language will not influence response patterns in the fixed discrimination task, and will influence response patterns in the similarity rating condition.

The second aspect of the lexical distance approach that was tested in this experiment is the assumption that the lexicon contains forms that are fully phonetically specified, as opposed to abstract phonological representations in terms of phonemes or underspecified feature bundles. The experiment doesn't actually decide between these two approaches, but rather only shows that language-specific perceptual performance is sensitive to phonetic detail.

5.1 Method

Participants. Sixteen American English speakers (5 male, 11 female) participated in the similarity rating task. Nineteen American English speakers (7 male, 12 female) participated in the fixed discrimination task. Data from two participants (1 male, 1 female) in the fixed discrimination task were removed because English was not their native language. 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.

Nine Dutch speakers (4 male, 5 female) participated in both the rating task and the discrimination task. Their ages and number of years in the US are listed in table 8. The Dutch participants were paid $20 for their participation and none of them reported any past or present speech or hearing disorder.

Stimuli. Eighteen bisyllabic 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]. I recorded myself saying multiple instances of the eighteen bisyllabic sequences that result from placing each of the six fricatives into the three vowel environments, and selected for use in the experiment some fluently produced instances that were matched on intonation pattern (H* accent on

the first syllable and LL% over the second syllable), and duration. Table 9 shows the vowel and fricative durations of the stimuli.

Table 8. Characteristics of the Dutch listeners. AOA = Age of arrival in US.

listener	age	AOA	Gender	years in US
501	60	60	F	<1
502	58	58	M	<1
503	64	24	F	40
504	27	23	M	4
505	25	21	M	4
506	63	16	F	43
507	41	24	M	17
508	19	15	M	4
509	24	19	F	5

Table 9. Durations 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. In the speeded AX discrimination condition the listener's task was to respond as quickly and accurately whether the last member of a pair of stimuli (X) presented on that trial was the "same" (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 is more difficult to hear the difference between the stimuli, than when the reaction time is short. Reaction time was measured from the onset of the

fricative noise in the X stimulus. Statistical analyses also for RTs measured from the onset of V1 and from the offset of the fricative noise showed the same pattern of results found in the analysis reported below.

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). Each trial was presented twice for a total of (42 * 3 vowels * 2 repetitions) = 252 trials in the speeded discrimination task. The trials were blocked by vowel.

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). I asked listeners to keep their reaction times to 500 ms or less and to not make "too many" mistakes.

Listeners in the similarity rating task heard each of the AX trials (42 trials per vowel) three times for a total of 378 trials. As in the speeded discrimination task the interstimulus interval was 100 ms. However, in this task the stimuli were not blocked by vowel. 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.

5.2 Results

Speeded discrimination. The arcsin transform of the proportion of a correct "different" responses for the DIFFERENT pairs was analyzed in a repeated measures analysis of variance. Because each pair was presented only twice to each listener the pairs were collapsed across order of presentation (e.g. responses to afa/atha were collapsed with responses to atha/afa) for this analysis. The between listeners factor was native language (English vs. Dutch) and the within listeners factors were vowel (/i/, /a/, or /u/) and fricative pair (15 comparisons). The vowel main effect was significant ($F[2,52] = 9.48$, $p< 0.01$) and for fricative pair ($F[14, 364] = 35.4$, $p<0.01$). The fricative pair by vowel interaction was also reliable ($F[28, 728] = 9.2$, $p < 0.01$). No other main effects or interactions were significant (all $Fs < 1$).

The patterns in the RT data mirror those found in the proportion correct data. The same three effects that were statistically reliable in the proportion correct data were also reliable in two separate analyses of the RT data. In a repeated measure analysis of variance of reaction times to the DIFFERENT trials there was a vowel main effect ($F[2, 52] = 69.8$, $p<0.01$), a fricative pair main effect ($F[29, 754] = 8.83$, $p < 0.01$) [note that this analysis does not collapse across pair order], and a pair by vowel interaction ($F[58, 1508] = 2.5$, $p< 0.01$).

The reaction time data were also transformed into a measure of psychological distance (Takane & Sergent, 1983) by taking the inverse of the natural log of the reaction time (1/ln[RT]). As with the analysis of raw reaction time I included both "different" responses and "same" responses to the DIFFERENT trials (see Takane & Sergent, 1983). Again three factors emerged as significant in the analysis. There was a vowel main effect ($F[2, 52] = 75.4$, $p<0.01$), a fricative pair main effect ($F[29, 754] = 7.8$, $p < 0.01$), and a pair by vowel interaction ($F[58, 1508] = 2.6$, $p< 0.01$). Reaction time to /a/ (458 ms) was significantly shorter than to /i/ (508 ms) or /u/ (523 ms), which did not differ from each other in planned comparisons.

With 30 different fricative pairs for each vowel (90 comparisons overall) posthoc tests exploring the fricative pairs main effect and the pairs by vowels interaction are unwieldy. Figure 1 shows the result of some data aggregation. For each fricative, in each of the vowel environments, the figure shows the average distance (1/ln[RT]) for all AX pairs involving that fricative. These aggregated data show that fricatives were generally more perceptually distinct when preceded and followed by /a/, and that in the /a/ and /u/ environments the strident fricatives were generally more discriminable than the nonstrident fricatives.

Similarity rating. The average rating score given by listeners was analyzed in a repeated measures analysis of variance again 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 ($F[2,52] = 117.5$, $p< 0.01$) as was the fricative pair main effect ($F[29, 364] = 71.3$, $p<0.01$). The fricative pair by vowel interaction was also reliable ($F[58, 728] = 5.46$, $p < 0.01$).

These effects were also seen in the reaction time data. However, unlike the reaction time results in these subjective rating data we find also a fricative pair by language interaction ($F[29,364]=2.8$, $p < 0.01$). No other main effects or interactions were significant (all $Fs < 1$).

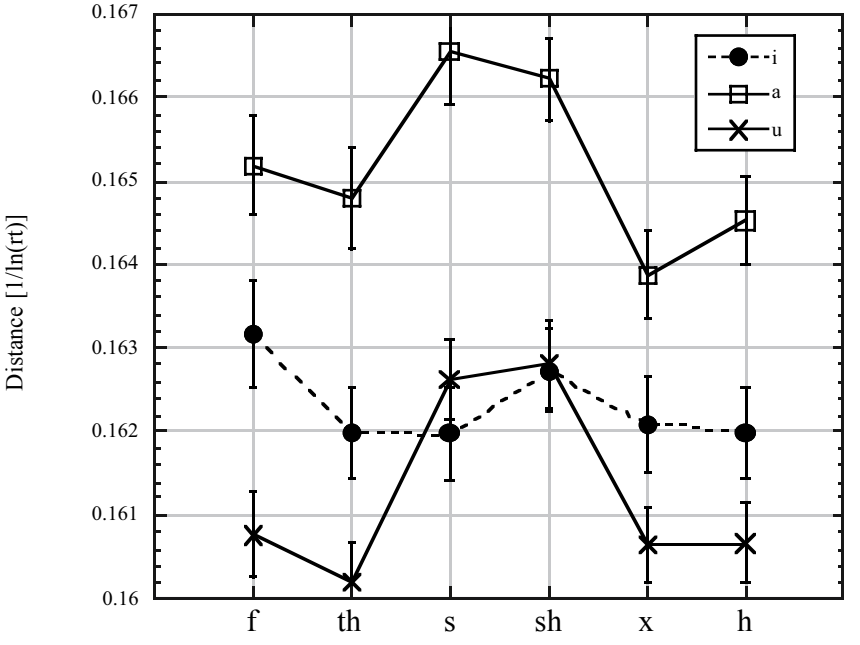

Figure 1. Average perceptual distinctiveness of six fricatives as measured by reaction time in a speeded discrimination task. Data plotted with open squares are for fricatives with /a/ preceding and following. Filled circles show data for the /i/ environment, and x's show data for the /u/ environment.

A similar repeated measures analysis of variance done on the z-scores of the ratings (normalizing the rating responses by each listener's mean and standard deviation) found that the same four factors - vowel, fricative pair, vowel*pair, and language*pair - were significant. The F-values were 122, 72.8, 5.9 and 3.0 respectively.

As with the reaction time data, the large number of different fricative pairs for each vowel presents a complicated post-hoc analysis situation. Figure 2 presents a summary of the listeners' subjective ratings of fricative distinctiveness by aggregating in the fashion of figure 1 over all paired comparisons that involve the fricative listed on the horizontal axis of the figure. These data show the difference between American English (AE) listeners and Dutch (D) listeners. Pairs involving [ʃ] tended to be judged as more "different" for American English listeners, while pairs involving [x] tended to be judged as more "different" by Dutch listeners.

5.3 Discussion

The key finding of this study is that with the same group of listeners, hearing the same pairs of stimuli, we have differing results depending on the task. In speeded discrimination, in which we hypothesize that the mental lexicon is not consulted, there was no reliable difference between Dutch and American English listeners, while in similarity rating, which does involve lexical knowledge, Dutch and American English listeners showed different patterns of responses. This is the result that was predicted by the lexical distance model.

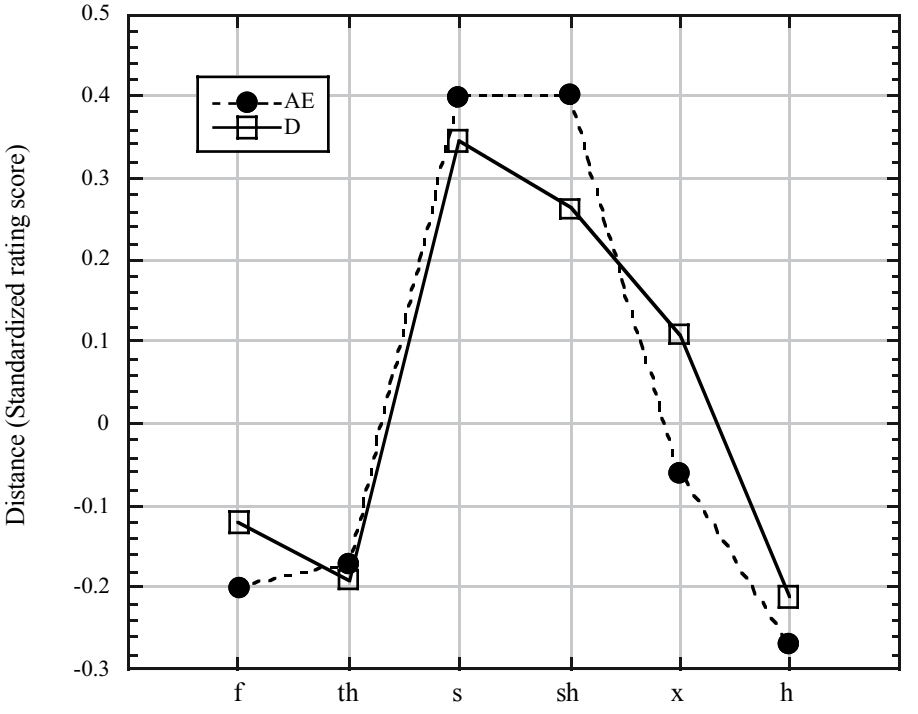

Figure 2. Average perceptual distinctiveness of six fricatives as measured by subjective rating in a perceptual distance rating task. Data plotted in filled circles show aggregate rating scores for American English listeners, and data plotted with open squares are for Dutch listeners.

The language-specific pattern of responding in the similarity rating task also largely accords with what the lexical distance model predicts. Pairs involving [ʃ] were judged to be more distinct by American English listeners while pairs involving [x] were judged to be more distinct by Dutch listeners. These effects make sense given that [ʃ] is not a contrastive sound in the Dutch lexicon and [x] is not a contrastive sound in the English lexicon, and assuming that if sounds that are involved in lexical distinctions are heard by listeners as being more distinct than sounds that do not distinguish lexical items. However, on this basis we might also have expected [θ] to be rated as more distinct by English listeners than it was by Dutch listeners. This lack of difference may reflect a floor effect driven by the somewhat low auditory perceptual salience of [θ], or it may be due to the bilingual experience of the Dutch speakers who participated in this experiment.

6. Conclusion

Best (1995) posited that incoming foreign sounds are compared with the sound inventory of the speaker's native language and "assimilated" to the native inventory in one of several ways. The lexical distance simulations presented in this paper also use a mechanism to compare foreign stimuli with the native language phonology.

In both approaches the initial comparison relies on a mechanism that permits foreign sounds and native sounds to be compared. Best suggests that this is done by direct perception of gestural structures (Fowler, 1986), which presumably would then be compared with records in memory of native language gestural structures during the process of perceptual assimilation. The approach that I outlined here posits that memory for language sound patterns is lexical. By this I mean that there is

no separate list of native gestural structures, but only an inventory of words. Also, clearly the names that I chose for tables 1 and 4 ("auditory-perceptual") indicate that I am not conceiving of this comparison mechanism in gestural terms, though this need not be a central theoretical difference.

One kind of evidence might indicate whether we should adopt a lexical approach or a phone list approach. Language sound pattern differences extend far beyond phoneme inventory differences. For example, while Mandarin and English both have /p/, /t/, and /k/, these may occur in syllable codas in English, but not Mandarin. So, although /p/, /t/, and /k/ may be familiar to both groups of listeners, it may be that Mandarin listeners are not as good at detecting coda stop place as English listeners are. Similar examples of word position effects can be cited for numerous languages. The point is that knowledge of phonetic detail, whether auditory or gestural is contextual. A lexicon-based approach captures contextually-conditional knowledge without further elaboration while a phone list approach must posit that positional variants are computed and stored separately.

Another, possibly key difference between the lexical distance approach and Best's perceptual assimilation model is that the lexical distance approach permits gradient patterns of cross-linguistic similarity. The perceptual assimilation model posits an inventory of discrete types of cross-linguistic differences, including single-category assimilation, category goodness assimilation and two-category assimilation as illustrated in the Zulu consonant contrasts above. The implicit claim of such a categorization is that cross-linguistic perceptual effects will fall neatly into these categories. A lexicon-based model suggests that the type frequency of phonological patterns, or the token frequency of words that exhibit a pattern, or perhaps also the neighborhood densities of words that exhibit a pattern, can play a role in cross-linguistic speech perception. So, we might expect a range of different assimilation patterns within the broad perceptual assimilation categories envisioned by Best. Whether or not this happens is an empirical question.

Appendix: Source code of the lexical distance model

```
#!/usr/bin/perl
#this file is zululex.prl    -    simulating results from Best, McRoberts & Goodell (2001) JASA.

# here is a set of syllables to use as stimuli in an experiment
# L = voiced lateral fricative, & = voiceless lateral fricative,
# k = voiceless aspirated dorsal stop, K = voiceless dorsal ejective,
# b = voiced bilabial stop, B = voiced implosive stop
@stim = ("&e","Le","ka","Ka","bu","Bu");

# here is a lexicon of relevant "English words" that will be used to calculate a lexically-based
#  perceptual response for each stimulus. C = tS, S = esh, Z = voiced esh
@lexicon = ("ba","pa","be","pe","bu","pu", "va", "ve", "vu", "ma",
          "me", "mu", "ka", "ke","ku", "ga", "ge", "gu",
           "Ca","Ce","Cu","sa","se","su","ta","te","tu",
          "za","ze","zu","Sa","Se","Su","Za","Zu",
           "la","le","lu","ha","he","hu");

# model parameters.
$leximp = 10;  # how much more weight to put on lexicon versus raw auditory distance?
$c = 2;              # governs selectivity of distance-to-similarity mapping.

open(DIST,"zuludist.txt");  # The consonant distance matrix (table 4) is stored in this file
while (<DIST>) {
   chomp;
    ($s1,$s2,$d) = split(/,/,$_);
    $cdist{$s1}{$s2}=$d;
}
close(DIST);
```

```
$vdist{"e"}{"a"} = $vdist{"a"}{"e"} = 0.8;
$vdist{"e"}{"u"} = $vdist{"u"}{"e"} = 0.7;
$vdist{"u"}{"a"} = $vdist{"a"}{"u"} = 0.8;

# Now the main loop of the experiment simulation program.  The basic idea is to compare
#  successive stimulus with each other. The difference between stimuli is a function of the
#  difference in lexical activation caused by them, plus the raw auditory distance between them.

print "Lexicon is: @lexicon\n";
while ($A = pop(@stim)) {  # item A in the discrimination pair
   $X = pop(@stim); # item X in the discrimination pair

   $lex_diff = $actA = $actX = 0;  # summation variables

   #here's the "lexical" distance loop
   foreach $lex (@lexicon) { # look at each item in the lexicon
      $aA = activation($A,$lex);          # equation (3)
      $aX = activation($X,$lex);

      $actA += $aA;  # keep a sum of the lexical activations
      $actX += $aX;  #  for each stimulus

      # so, the lexical distance between A and X is the sum over  lexical items of:
      #  the difference in lexical activation produced by A and X [abs($aA - $aX)]
      #  which has been scaled by the activation levels, [($aA + $aX)]
      #  and normalized by lexicon size [(1/@lexicon)]

      $lex_diff += abs($aA - $aX) * ($aA + $aX) * (1/@lexicon);  # equation (2)
   }

   # here we look up the "raw" auditory distance
   $phondist = distance($A,$X);

   # so the total distance between A and X is a combination of "lexical"  distance and
   #       raw auditory distance.

   $diff = $phondist + $lex_diff*$leximp;       # equation (1)
   $actA /= @lexicon;  # average activation
   $actX /= @lexicon;

   print "$A $X\t";
   printf("%3.2f\t%3.2f\t%3.2f\n",$diff,$actA,$actX);
}

# This subroutine takes two CV words and sums up the perceptual distance between the words
#       from the consonant and vowel distance tables given at the top of this file.
sub distance {
      my ($s1, $s2) = @_;

      my $d = 0.0000001;
      @st1 = split (//,$s1);
      @st2 = split (//,$s2);
```

```
        foreach $p1 (@st1) {  # phone-by-phone matching
            $p2 = shift(@st2); #pull off the corresponding one from s2

            if ($cdist{$p1}{$p2}) {$d+= $cdist{$p1}{$p2};}
            elsif ($cdist{$p2}{$p1}) {$d+= $cdist{$p2}{$p1};}
            elsif ($vdist{$p1}{$p2}) {$d+= $vdist{$p1}{$p2};}
        }
        return ($d);
}

# This subroutine returns the "activation" of a lexical item, which is defined here as simply
#    similarity, using Shepard's rule - similarity is an exponential function of distance.
sub activation {
        my ($s1, $s2) = @_;
        my $d;

        $d = distance($s1,$s2); # look up distance
        return (exp(-$d*$c));   # equation (3)
}
```

References

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. and Goodell, E. (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. J. Acoust. Soc. Am. **109**, 775-794.

Braida, L.D. and Durlach, N.I. (1988) Peripheral and central factors in intensity perception. In *Functions of the Auditory System*, edited by G.M. Edelman, W.E. Gall and W.M. Cohen (New York: Wiley).

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.

Dehaene-Lambertz, G., Dupoux, E. & Gout, A. (2000) Electrophysiological correlates of phonological processing: A cross-linguistic study. *J. Cognitive Neuroscience,* **12**, 635-647.

Dell, G.S., Reed, K.D., Adams, D.R., and Meyer, A.S. (2000) Speech errors, phonotactic constraints and implicit learning: A study of the role of experience in language production. *J. Exp. Psych: Learning, Memory & Cognition* **26**, 1355-1367.

Dijkstra, T., Van Jaarsveld, H. and Ten Brinke, S. 1998. Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition* 1:51-66.

Dooling, R.J., Best, C.T. and Brown, S.D. (1995) Discrimination of synthetic full-formant and sinewave /ra - la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *The Journal of the Acoustical Society of America* **97**, 1839-1846.

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..

Fowler, C.A. (1986) An event approach to the study of speech perception from a direct realist perspective. *J. Phonetics,* **14**, 3-28.

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

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

Grainger, Jonathan, and Cecile Beauvillain. 1987. Language blocking and lexical access in bilinguals. *Quarterly Journal of Experimental Psychology,* **39A**, 295-319.

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.

Guenther, F. & Gjaja (1996) The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America,* **100**, 1111-1121.

Hickok, G. and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* **4**, 131-138

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65-70.

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.

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.

Macmillan, N.A. (1987) Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In *Categorical Perception*, edited by S. Harnad (New York: Cambridge), pp. 53-87.

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.

Nittrouer, S. (1992) Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* **20**, 351-82.

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

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

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.

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.

Watson, C.S. & Kelly, W.J. (1981) The role of stimulus uncertainty in the discrimination of auditory patterns. In *Auditory and Visual Pattern Recogntion*, edited by D.J. Getty and J.H. Howard (Hillsdale: LEA), pp. 37-59.

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.

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).

Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception

edited by Augustine Agwuele, Willis Warren, and Sang-Hoon Park

Cascadilla Proceedings Project Somerville, MA 2004

Copyright information

ISBN 1-57473-402-4 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Johnson, Keith. 2004. Cross-linguistic Perceptual Differences Emerge from the Lexicon. In *Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele et al., 26-41. Somerville, MA: Cascadilla Proceedings Project.

or:

Johnson, Keith. 2004. Cross-linguistic Perceptual Differences Emerge from the Lexicon. In *Proceedings of the 2003 Texas Linguistics Society Conference*, ed. Augustine Agwuele et al., 26-41. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1065.