Internal Use Only
非公開

TR - H - 079 0001

The Perception of Hindi Dental and Retroflex Stop Consonants by Native Japanese and English Speakers I: Initial Perception

John S. PRUITT
(University of South Florida)
Reiko A. YAMADA

1994. 5. 31
(1995.6.14受付)

ATR人間情報通信研究所
〒619-02 京都府相楽郡精華町光台2-2 ☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

The Perception of Hindi Dental and Retroflex Stop Consonants by Native Japanese and English Speakers I: Initial Perception. *

John S. Pruitt[1] Reiko A. Yamada[2]

[1]Dept. of Psych., University of South Florida, Tampa, Florida 33620, U.S.A.
[2]ATR Human Information Processing Research Laboratories, Soraku, Kyoto, Japan

*A part of this study was done at ATR HIP Labs, when John S. Pruitt stayed at ATR as a visiting researcher from April 1st to May 31st, 1994.

Contents

1 Abstract

Previous research has shown that American English speakers have great difficulty distinguishing dental versus retroflex stop consonants of the Hindi language (which occur in five manner/voicing contexts). While both dental and retroflex consonants occur as allophones in English in several phonetic contexts, they do not occur phonemically. Furthermore, American English contains the rhotic /ɹ/, which frequently is produced as a retroflex. In spite of this rich allophonic experience with dental and retroflex consonant allophones, English speakers have great difficulty in distinguishing the Hindi sounds. Unlike English, the Japanese language includes a distinction which is similar to the Hindi distinction - the Japanese /d/ versus the flapped /r/. The Japanese /d/ is typically produced as an dental/alveolar stop while the flapped /r/ is either palatalized or retroflexed. However, no research has determined whether Japanese speakers can distinguish the Hindi contrast. The present research compared the ability of English, Japanese, and a control group of native Hindi speakers to distinguish Hindi dental versus retroflex stop-consonants in four contexts (voiced-aspirated, voiced-unaspirated, voiceless-aspirated, and voiceless-unaspirated). Subjects were presented consonant-vowel syllables in three vowel contexts (/a/, /e/, /o/) which were produced by two native Hindi speakers. Marked differences in identification scores were found between all three language groups which were partially dependent upon the manner/voicing context of the consonants. As expected, native Hindi speakers performed near perfectly while Japanese and English speakers performed less well. In general, however, Japanese speakers had a distinct advantage over English speakers in perceiving this contrast. Thus, it appears that the functionally similar contrast that Japanese has to Hindi provides a better experiential basis for perceiving the nonnative Hindi contrast than the allophonic experience of English speakers.

2 Introduction

Many studies have shown that adult speakers of any language have difficulty distinguishing some nonnative speech sounds (Flege, 1984, 1989 [14, 15]; Gillette, 1980 [16]; Goto, 1971 [17]; Jamieson & Morosan, 1986, 1989 [19, 20]; Lisker & Abramson, 1970 [27]; Polka, 1991 [33]; Tees & Werker, 1984 [42]; and Trehub, 1976 [43]; for a brief review, see Strange, in press). Problems occur in both production and perception, and tend to persist even after years of immersion in the nonnative language (Flege & Eefting, 1987; MacKain, Best, & Strange, 1981; Strange, 1972; Yamada, in press; Yamada et al., 1994). Other studies have shown that not all non-native contrasts are equally difficult to perceive or produce (Best, McRoberts, & Sithole, 1988 [4]; Tees & Werker,1984 [42]). For example, Tees and Werker (1984) [42] showed that a place-of-articulation contrast in Hindi was more difficult than a voicing contrast. Other studies have shown that the phonetic environment of a contrast affects its perceptual difficulty (Logan, Lively, & Pisoni, 1991 [29]; Polka, 1991 [33]; Sheldon and Strange, 1982 [38]). It is thought that acoustic, phonetic, and phonemic factors contribute to this variability (Polka, 1991 [33]; Werker & Logan, 1985 [45]). Still, little is known as to why these differential difficulties exist. Werker (1986) [44] did a study which investigated the effects of multilingualism on the perception of nonnative contrasts. Monolinguals, bilinguals, and trilinguals of assorted languages were tested on two different place-of-articulation contrasts which did not occur in any of the multiple languages - the Hindi dental versus retroflex contrast and the Thompson velar versus uvular contrast. Surprisingly, no advantage was found for being multilingual and no particular language combination showed increased sensitivity. However, each particular language combination was underrepresented in the sample (only 9 bilinguals and 8 trilinguals; two sets of three bilinguals spoke identical languages and one set of three trilinguals spoke the same languages). Further, Werker presented very little information about the phonetic relations of the languages involved. Thus, it is difficult to say anything about the benefits of a specific language experience on the two nonnative contrasts in her study.

To explore ideas regarding the effect of native language experience on second language perception, Polka examined the perception of a Salish velar versus uvular voiceless glottalized stop contrast by native speakers of English and Farsi. The Farsi language includes this contrast but it occurs in voiced stops, not voiceless stops. Farsi also has a voiceless uvular fricative, but it does not contrast with a voiceless velar fricative. English, on the other hand, does not have any phones produced in the uvular place of articulation. Thus, it was predicted that Farsi speakers would differentiate the Salish contrast better than English speakers.

Surprisingly, results showed that experience with a related contrast, but in a different voicing or manner class, did not aid in perceiving the nonnative contrast. In fact, English speakers had a slight advantage over Farsi speakers on this Salish contrast (approximately 85% versus 73% correct on an identification task). Polka also found a great deal of individual variability among Farsi speakers which was correlated with the perceptual assimilation performance of the speakers. That is, Farsi speakers who indicated that the Salish consonants were similar to Farsi consonants performed at much higher levels. This supports the theory of Best and colleagues that perceptual assimilation strategies may be predictive of non-native perception (Best, 1993, 1994 [1, 2]; Best, McRoberts, & Sithole 1988 [4]). Still, the acoustic, phonetic, and phonemic factors that lead to different assimilation strategies (and to different perceptual performance) are little understood.

The present study examined a similar relationship between three language groups - Hindi (the target language), English, and Japanese. Hindi contains a contrast (dental versus retroflex place of articulation) which does not occur in the phonemic inventory of English or Japanese (Best & Strange, 1992 [3]; Bhat, 1974 [6]; Bloch, 1950 [7]; Delattre, 1966 [9]; Fairbanks & Misra, 1966 [12]; Hattori, 1984 [18]; Jones, 1964 [21]; Jorden, 1963 [23]; Marten, 1962 [30]; Ogawa, 1982 [31]; Polka, 1991 [33]; Price, 1981 [35]; Shibatani, 1990 [39]; and Tees & Werker, 1984 [42]). However, English and Japanese differ in their phonetic relationship to Hindi with regards to this contrast. The relationship of these languages is such that a cross-language perceptual study may provide further insight into how, why, or which acoustic, phonetic, and phonemic factors are important to non-native perception.

2.1 Linguistic Descriptions

Hindi.

The Hindi language is one of the many languages of India and is also the official language of India. This language contains a place of articulation contrast which does not occur in English nor Japanese stops; the dental versus

Table 1: The Hindi Dental and Retroflex stop consonants.

Voicing	Manner of Articulation	Place of Articulation	
		Dental	Retroflex
Voiceless	Unaspirated	t̪	ṭ
	Aspirated	t̪^{h}	ṭ^h
Voiced	Unaspirated	d̪	ḍ
	Aspirated	d̪^{h}	ḍ^h
	Nasal	n̪	ṇ

retroflex contrast. A dental stop is produced by making a constriction of the vocal tract with the tongue at or right behind the top front teeth. A retroflex stop is produced by curling the tip of the tongue upwards towards the hard palate to make a constriction somewhere behind the alveolar ridge usually with the underside of the tongue (Ladefoged, 1982 [25]). In Hindi, the dental-retroflex contrast occurs in five pairs of stop consonants (as shown in Table 1); voiceless-unaspirated /t̪ - ṭ/, voiceless-aspirated /t̪^{h} - ṭ^h/, voiced-unaspirated /d̪ - ḍ/, voiced-aspirated /t̪^{h} - ṭ^h/, and nasal /n̪ - ṇ/ (Fairbanks & Misra, 1966 [12]). The nasal retroflex occurs only in syllable-medial or syllable-final position. Also, prevoicing is typical for the voiced Hindi stop consonants. See the Appendix for additional information on Hindi and retroflexion.

English.

The dental versus retroflex contrast does not occur phonemically in English (Jones, 1964 [21]). It does, however, alveolar stops which contrast with bilabial and velar stops. More importantly, some of the Hindi dental and retroflex stop consonants occur in English phonetically (i.e., as allophonic variations of English consonants). And, of course, aspirated and nonaspirated stop consonants occur regularly in English but aspiration is never distinctive (Delattre, 1966 [9]). As Polka (1991) [33] points out, English /t/ and

/d/ in syllable-initial position are usually produced as voiceless-aspirated and voiceless-unaspirated alveolar stops, respectively, but they may be produced as retroflex stops when produced in a consonant cluster with /r/ (e.g., "try" and "dry"). Also, English /d/, in syllable-medial position, may be produced as a prevoiced alveolar stop or as a prevoiced retroflex stop when preceding /r/ (e.g., "added" and "address"). Furthermore, American English contains the rhotic approximate /ɹ/, which frequently is produced as a retroflex. Thus, English speakers have some experience with dental and retroflex stops, but this experience occurs only in the presence of certain other phonemes.

Japanese.

Unlike English, the Japanese language includes a contrast similar to the Hindi dental versus retroflex contrast; the Japanese dental /d/ versus the flapped /r/ (Best & Strange, 1992 [3]; Price, 1981 [35]). However, there is some uncertainty concerning the articulation of the flapped /r/ in Japanese. Many linguists describe the flapped /r/ as an alveolar flap which is sometimes palatalized (Bloch, 1950 [7]; Jorden, 1963 [23]). Others note that it may be produced as a retroflex - "by lifting the tip of the tongue backwards" (Marten, 1962 [30]; Price, 1981 [35]). Still others note that the actual place of articulation of the flapped /r/ moves backwards towards palatal when produced with back vowels (Hattori, 1984 [18]; Ogawa, 1982 [31]). More importantly, while English and Japanese both contain flapped /r/, which may be articulatorially and acoustically similar in both languages, the flapped /r/ operates in very different phonological systems (Price, 1981 [35]). As noted by Price(1981) [35], the English flapped /r/ does not contrast with /d/ or /t/, particularly in syllable-initial position, rather the flapped / in an allophone of /d/ and /t/ in intervocalix contexts. In Japanese, however, the flapped /r/ occurs in as many environments as /d/ and /t/. Further, the Japanese flapped /r/ contrasts with /d/ but not with any lateral or retroflex phones, while English flapped /r/, in intervocalic position, contrasts with /l/ and /r/ but not /d/. Thus, /d/ and flapped /r/ should be considered a phonemic distinction in Japanese but not in English.

Interestingly, no study has examined the perception of the Hindi distinction by native Japanese speakers. Several studies have investigated the perception of the Hindi dental versus retroflex contrast by native English speakers. Tees and Werker (1984) [42] compared the perception of Hindi voicing contrasts and place contrasts using the Hindi voiceless-aspirated versus voiced-aspirated stops contrast pand the Hindi dental voiceless-unaspirated stops versus retroflex voiceless-unaspirated stops contrast. Performance on the dental-retroflex contrast was quite poor; particularly in comparison to performance on the Hindi voicing contrast.

Polka (1989, 1991) [32, 33] demonstrated that the Hindi dental-retroflex contrast in different voicing contexts yielded different levels of perceptual difficulty for native English speaking listeners. Using an AX same-different task without feedback, Polka tested subjects' ability to discriminate the Hindi dental-retroflex contrast in four of the five voicing/manner contexts (voiceless-unaspirated, voiceless-aspirated, voiced-unaspirated and voiced-aspirated - all were produced in syllable-initial position with the vowel /a/ by a single Hindi male). For Polka's subjects, the voiced-unaspirated contrast was most difficult to discriminate (about 51% correct discrimination), followed by the voiceless-aspirated contrast (55% correct) and the breathy-voiced contrast (62% correct), and finally by the voiceless-unaspirated contrast (72% correct). This order of difficulty could not be predicted by theories of phonetic similarity (as discussed by Best, McRoberts, and Sithole, 1988 [4]) or by the particular allophonic experience of the listeners. Instead, the number of acoustic cues differentiating these contrasts in the stimulus corpus predicted the relative difficulty of these contrasts best.

More recently, Pruitt (1992) [36] replicated and extended the work of Polka (1989; 1991) [32, 33]. Pruitt [36] investigated the difficulty experienced by American English listeners in identifying dental and retroflex consonants produced by two speakers in different voicing/manner and vowel contexts. The results were similar to Polka (1989, 1991) [32, 33]. Native English speakers had differential difficulty identifying dental versus retroflex consonants when they were produced in different voicing/manner contexts. The order of difficulty only partially replicated Polkas findings because the relative difficulty was affected by which speaker produced the contrasts and, to a lesser extent, by the vowel context. Overall, subjects in the study by Pruitt [36] had the most difficulty with the voiced-aspirated (about 52% correct identification; 50% is chance performance), while the voiced-unaspirated and voiceless-aspirated were about equally difficult (about 57% correct) and the voiceless-unaspirated was the least difficult (about 59% correct). Additionally, one of the speakers was less intelligible than the other and this difference interacted with both vowel and voicing/manner contexts.

2.2 The Study

This study was an attempt to replicate and extend both Polka's (1989, 1991, 1992) [32, 33, 34] and Pruitt's (1992) [36] studies because it assessed the relative difficulty of the four voicing/manner contexts in different vowel contexts by both native American and Japanese speakers. To investigate the role that this differential experience of native language has on nonnative perception, examples of dental and retroflex stop consonants were recorded and presented to speakers of all three languages. The examples were produced by two male, native Hindi speakers in four of the five manner/voicing contexts with three vowels (/a/, /e/, /o/) in a consonant plus vowel (CV) syllable. The present study was limited to only syllable-initial consonants because preliminary data with this contrast as well as studies with other place-of-articulation contrasts indicated that non-native perception was quite good for syllable-medial and final consonant contrasts (e.g., Lively, Logan, and Pisoni, 1993 [28]). This is possibly due to coloration of the preceding vowel (i.e., coarticulation). Retroflexed vowels appear to have a very salient /r/ quality to English listeners.

From the description of the phonemic and phonetic make-up of Hindi, English, and Japanese with regard to dental and retroflex consonants, at least two possible hypotheses posited regarding the perceptual performance of English and Japanese speakers:

1) English speakers will perform better because of their more extensive phonetic experience with retroflexion (including experience with a truly retroflexed /r/, and additionally, some phonetic experience with aspirated stops).

2) Japanese speakers will perform better because of their phonemic experience with /d/ and flapped /r/ (in spite of their lack of experience with retroflexed /d/ and /t/ due to coarticulation with /r/).

3 Methods

3.1 Subjects

Twenty native English, 20 native Japanese, and 10 native Hindi speakers were tested. There was an equal number of males and females in both the English and Japanese language groups. The Hindi group had 6 males and 4 females. English and Japanese subjects were monolingual [1] and all subjects were free of any hearing deficiencies as screened by self report (Japanese subjects were given a quick auditory threshold test.) Hindi subjects were all at least trilingual with Hindi and English being the common two languages [2]. English subjects ranged in age from 19 to 43 with a mean of 26. Japanese subjects ranged in age from 20 to 33 with a mean of 25. Hindi subjects ranged in age from 22 to 32 with a mean of 27.

3.2 Stimuli

All stimuli were produced by two male native Hindi speakers. Both speakers were born and raised in India and learned Hindi as their first language. The parents of the speakers were native Hindi speakers and spoke mostly Hindi at home. Both subjects began learning English as a second language between the ages 4 to 6 when formal education began. Speaker 1 was a 26 year old male who had lived in the United States for two and a half years. Besides English and Hindi, he also spoke Punjabi (another language of India). Speaker 2 was a 23 year old male who had lived in the United States for three years and also spoke Punjabi and Urdu (Urdu is also a language of India).

The speakers read aloud consonant-vowel (CV) syllables in a random order, which were written in standard Hindi manuscript. Some of the syllables were words in Hindi and some were not. At least 20 exemplars (10 dental and 10 retroflex) were produced by each speaker in each of the four manner/voicing contexts with each of the three vowels. The syllables were recorded onto Ampex 631 - 1/4 inch audio tape at 71/2 inches per second using a Tascam 22-2 two-track reel-to-reel tape recorder via a Shure Spher-O-Dyne (533SA) high impedance microphone. The stimuli were low-pass filtered at 9.7 kHz and then converted to digital waveform files at a 20 kHz sampling rate with 12-bit resolution using a pulse code modulation (PCM) hardware coding device (see Whalen, Wiley, Rubin, & Cooper, 1990) attached to a VAX station computer. The digitized stimuli were pre-emphasized at 12 dB per octave beginning at 800 Hz.

[1] Although many of the Japanese subjects had studied English for several years, few had any extensive spoken training by native English speakers. None considered themselves fluent.

[2] However, one female subject in the Hindi group did not actually speak Hindi but did speak Malay, Kannada, and Tamil which are all phonetically similar to Hindi.

The stimuli were then randomly sequenced (but blocked by speaker and vowel context) and recorded to tape. Four native Hindi speakers were then asked to identify this random presentation of all of the recorded tokens. Tokens that were not correctly identified by all four listeners were discarded (75 out of 480 tokens were discarded; 37 out of 240 for speaker 1; 38 out of 240 for speaker 2). Finally, for each voicing context in each vowel context, eight tokens per speaker (four dental and four retroflex) that were similar in amplitude, intonation, and duration were selected for the test. Thus, a total of 96 tokens per speaker were used in the test. This selection process ensured that nonphonemic variables (e.g., amplitude, intonation, duration) did not vary in relation to phonemic variables (e.g., burst amplitude, VOT, aspiration amplitude). The distribution of tokens is illustrated in Table 2.

The final selection of tokens were then recorded from the VAX station to a Teac DAT recorder (DA-P20) and then re-digitized to a Macintosh computer at a rate of 44.1kHz samples per second with a 16-bit amplitude quantization (due to this process, the tokens retained the characteristics of their original format of 20kHz sampling rate and 12 bit quantization). Finally, the tokens were edited to remove surrounding silence and any anomalies of the recording process.

3.3 Procedure

Upon arrival, each subject completed a consent form and a language background questionnaire. Subjects were tested individually on a Macintosh computer via headphones (STAX SR Lambda Professional) in a sound attenuated chamber. Prior to testing, each subject was given a brief explanation of the articulatory characteristics of the Hindi dental and retroflex consonants using a diagram of the human vocal tract. Then a complete description of the testing procedure was given. All instructions were given in Japanese for Japanese speakers and in English for English speakers.

Four tests were given; one for each voicing/manner context. Prior to the administration of each test, a familiarization sequence was presented to the subjects in which all 48 stimuli were blocked by speaker, vowel, and place-of-articulation. During this familiarization, subjects were instructed to listen carefully without responding in order to become familiar with the particular stimuli for that test. During the familiarization, the computer indicated the speaker, the vowel context, and the place-of-articulation for each stimulus.

Table 2: The Distribution of Tokens for the Pretest-Posttest.

	Speaker 1			Speaker 2			
Vowel	/a/	/e/	/o/	/a/	/e/	/o/	
Voicing Context							
Test 1 - Breathy-voiced							
Dental	4	4	4	4	4	4	=24
Retroflex	4	4	4	4	4	4	=24
Test 2 - Pre-voiced							
Dental	4	4	4	4	4	4	=24
Retroflex	4	4	4	4	4	4	=24
Test 3 - Voiceless-aspirated							
Dental	4	4	4	4	4	4	=24
Retroflex	4	4	4	4	4	4	=24
Test 4 - Voiceless-unaspirated							
Dental	4	4	4	4	4	4	=24
Retroflex	4	4	4	4	4	4	=24
							= 192

Table 3: The Order of the Voicing/Manner Context Tests.

Ss	Serial Position				
1-5	A	D	B	C	A = Voiced-Aspirated
6-10	B	A	C	D	B = Voiced-Unaspirated
11-15	C	B	D	A	C = Voiceless-Aspirated
16-20	D	C	A	B	D = Voiceless-Unaspirated

Subjects were given an opportunity to ask questions before the test began. During the test, the subjects' task was to listen to a single token twice, decide whether the token contained a dental or a retroflex consonant, and then click on the correct response using the mouse. For every trial, the particular speaker was indicated with a small picture representing that speaker. No feedback was given during each test, but after completing the test, subjects were paid either 1 cent (or 1 yen) per correct trial. The maximum possible was $3.84 (96 trials by 4 sub-tests).

The order of the administration of the four voicing/manner tests was counterbalanced using a Latin square as shown in table 3. This controlled immediate sequencing effects as well as order effects (See Bradley, 1958). No analysis of order effects was attempted. The entire test lasted approximately one to one and a half hours.

3.4 Analyses

To investigate the relative difficulty of the contrasts in different voicing/manner classes, vowel contexts, and speakers for each language group, the percent correct identification scores [3] were averaged across subjects for each context

[3]Due to the nature of identification tasks when subjects are unfamiliar with the categories to be labeled, subjects sometimes reverse the labels of the categories. This was not considered a perceptual error on the subjects' part because the labels were determined somewhat arbitrarily (although the categories were not). Following the procedure of Polka (1989), if a subject's score was significantly below chance on one of the voicing context tests, the labels were reversed, and thus, the score was transformed for that test only. For example, a score of 39% correct is significantly below chance (50%) on a test with 96 trials, so the score would become 61% correct (100 - 39). This occurred only one time for

condition. The resulting design was a 3 x 4 x 3 x 2 mixed model analysis of variance (i.e., 3 language groups x 4 voicing contexts x 3 vowel contexts x 2 speakers). However, the comparisons of interest were between the English and Japanese language groups. Thus, after this initial analysis, the Hindi data set was removed and a second analysis was done with only the English and Japanese groups. To support this decision, an analysis was done on the Hindi data alone to ensure that little or no differences existed between the various stimulus conditions.

All between-subjects subsequent testing was performed using the conservative Tukey-Kramer Honestly Significant Difference Test (HSD) with a being set at 0.05 (see Keselman & Rogan, 1977 [24], for a discussion of Tukeys procedure). For within-subjects subsequent testing, the standard procedure of contrast specification was used with type III sums of squares to determine the error term.

4 Results

As shown in Figure 1, averaging across all contexts, native Hindi speakers performed much better (96%), than native Japanese speakers (71%), who were better than native English speakers (59%); p=0.0001, MSE = 111133.398. Subsequent testing revealed that all three means were reliably different (the critical HSD for English versus Japanese was 5.56; for English versus Hindi and Japanese versus Hindi, the critical HSD was 6.81). The near perfect performance of the native Hindi speakers indicates that the stimuli used in the test were correctly produced and that the task did not tax the native subjects perceptual or cognitive abilities. The ANOVA on Hindi speakers data alone showed no differences among the various stimulus conditions. However, the effect of vowel context, the interaction of vowel context with voicing/manner context, and the interaction of vowel context with speaker did approach significance (p=0.0814, MSE = 152.832;p=0.0505, MSE = 86.209; and p=0.0677, MSE = 78.288). From informal observation of the Hindi data, most of the variance in phonetic conditions was due to the performance of one subject - the Indian who did not actually speak Hindi but spoke several other Indian languages performed less well than other Indian subjects (this speakers overall average was 85% versus the overall group average of 97%).

a Japanese subject in the voiceless-unaspirated context condition.

While her performance is somewhat different than the other true native Hindi speakers, her performance is markedly different from English and Japanese subjects.

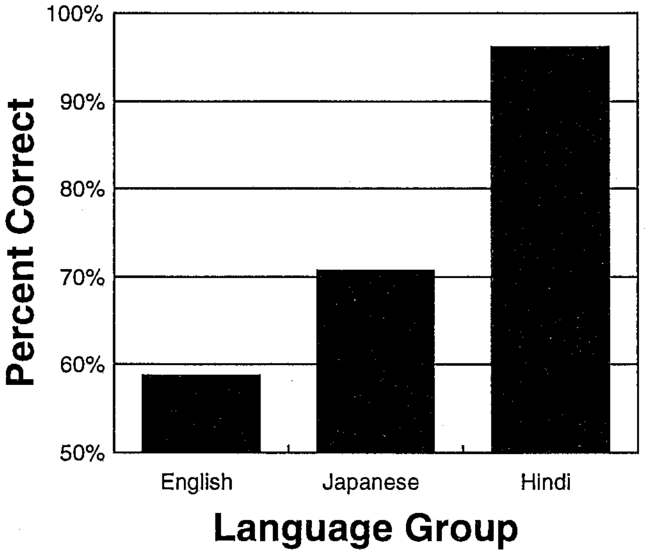

Figure 1: Overall Performance by Each Language Group.

All three language groups, overall, performed better on productions by speaker 1 than by speaker 2 (78% vs. 73% correct; p=0.0001, MSE = 8023.763). As shown in Figure 2, this difference in intelligibility between the speakers was greatest for American listeners (a difference of 9.1%; p=0.0001, MSE = 9969.076), slightly less for Japanese listeners (a difference of 5.5%; p=0.0017, MSE = 3588.867), and trivial for Hindi listeners (a difference of 1.8%; p=0.1620, MSE = 188.151). A significant interaction term supported these observed differences between the language groups (p=0.0112, MSE = 1130.273).

Perceptual differences as a function of vowel context were not as systematic. As shown in Figure 3, a distinct order of difficulty was found for Japanese listeners - from most difficult to least difficult, /a/, /e/, and /o/ (p=0.0001, MSE = 4068.929; contrast values are given in Table 4). No reliable differences in vowel context were found for either the English or

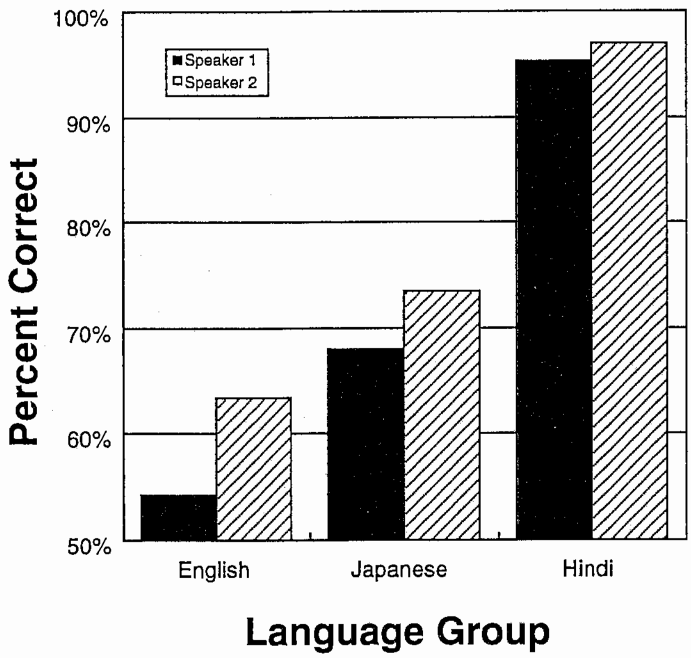

Figure 2: Performance by Speaker and Language Group.

Hindi groups (respectively, p=0.4113, MSE = 147.217, and p=0.0814, MSE = 152.832). Again, a significant interaction term supported these observed differences between the language groups (p=0.0001, MSE = 1364.160).

Finally, perceptual differences as a function of voicing/manner context were interesting and somewhat systematic. While no differences existed for Hindi speakers (p=0.3523, MSE = 238.064), large differences occurred for both English and Japanese speakers (respectively, p=0.0004, MSE = 2211.046; and p=0.0001, MSE = 6323.459). As illustrated in Figure 4, English speakers performed reliably worse on the voiced-aspirated context than the other three contexts (contrast values are given in Table 5). Similarly, Japanese speakers performed much worse on the voiced-aspirated context, but in contrast, also performed significantly better on the two unaspirated contexts than on the voiceless-aspirated context (contrast values are given in Table 6). Further contrast analysis showed that Japanese speakers performed reliably better on the two unaspirated contexts combined than with

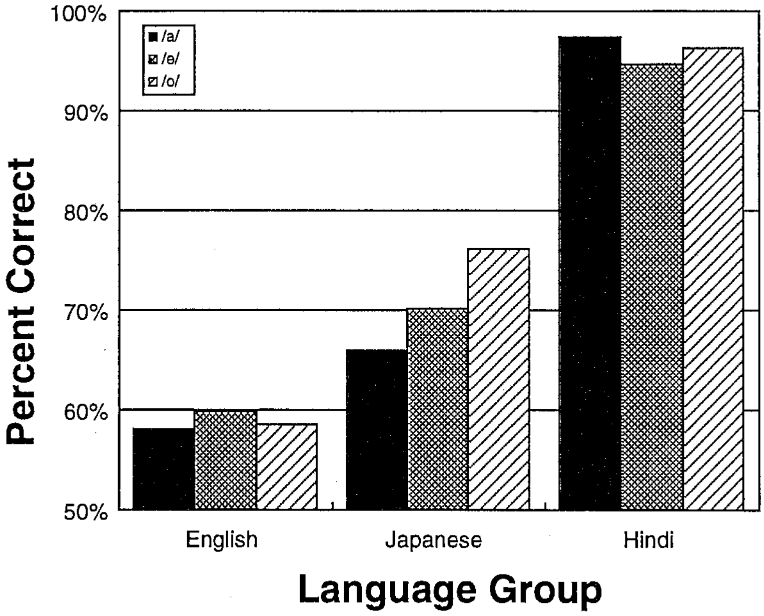

Figure 3: Performance by Vowel Context and Language Group.

the two aspirated contexts combined (p=0.0001, MSE = 16042.969). This was not the case for English speakers (p=0.0696, MSE = 1057.617).

5 Discussion

Several results are of interest in this study. First and most importantly, a clear difference emerged between the two groups of nonnative listeners. Japanese speakers showed a distinct advantage in perceiving the Hindi contrast. This advantage was found in all of the phonetic contexts and speaker conditions. In particular, while Japanese speakers performed best on the pair that is most similar to their native language phonemes (voiced-unaspirated dental versus retroflex), they still outperformed English speakers on the other voicing/manner contexts. This finding supports the hypothesis that phonemic (contrastive) experience with a place-of-articulation contrast similar to the Hindi dental-retroflex contrast leads to better performance on that non-

Table 4: Contrast Values between Vowels for Japanese Speakers.

	p Value		MSE	
Vowel	/e/	/o/	/e/	/o/
/a/	0.0008	0.0001	1397.583	8062.622
/e/		0.0001		2746.582

Table 5: Contrast Values for Voicing/Manner Context for English Speakers.

	p Value			MSE		
Voicing/Manner	VU	UVA	UVU	VU	UVA	UVU
Voiced-Asp.(VA)	0.0006	0.0001	0.0023	4114.746	5510.417	3144.694
Voiced-Unasp.(VU)		0.5685	0.6481		101.725	65.104
Voiceless-Asp.(UVA)			0.3063			329.590

UVU = Voiceless-Unasp.

Table 6: Contrast Values for Voicing/Manner Context for Japanese Speakers.

	p Value			MSE		
Voicing/Manner	VU	UVA	UVU	VU	UVA	UVU
Voiced-Asp.(VA)	0.0001	0.0406	0.0001	16047.526	2152.507	9769.694
Voiced-Unasp.(VU)		0.0006	0.2139		6445.475	774.902
Voiceless-Asp.(UVA)			0.0213			2750.651

UVU = Voiceless-Unasp.

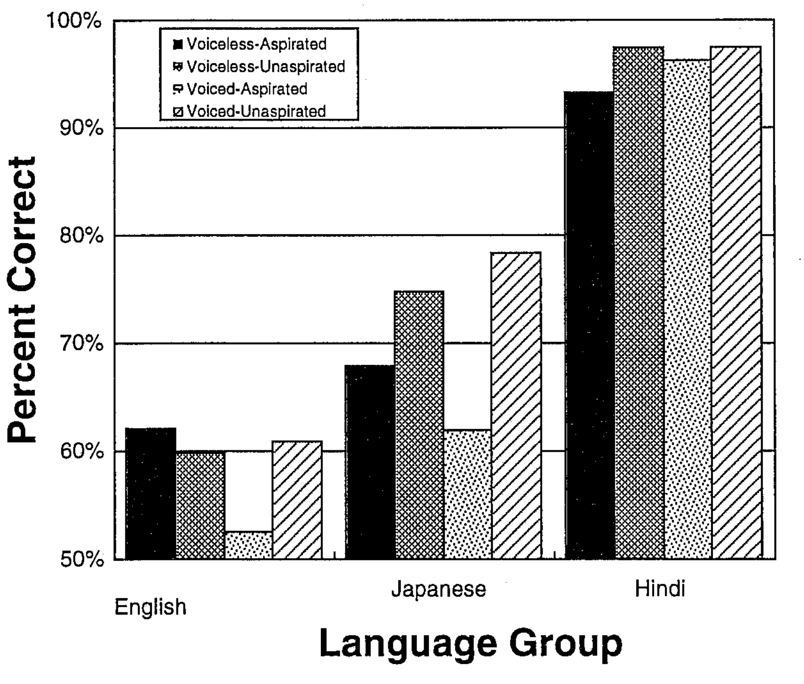

Figure 4: Performance by Voicing/Manner Context and Language Group.

native contrast than experience which is primarily non-contrastive but more allophonically robust. As mentioned earlier, both English and Japanese contain a flap, however, the English flapped /r/ does not contrast with /d/ or /t/, particularly in syllable-initial position. But, the Japanese flapped /r/ occurs in as many environments as /d/ and /t/ and contrasts with /d/ but not with any lateral or retroflex phones, while English flapped /r/, in intervocalic position, contrasts with /l/ and /r/ but not /d/ (Price, 1981 [35]).

Best has developed an assimilation model which can help to explain these data (Best, 1993,1994 [1, 2]). Within this framework subjects must categorize the target speech sounds (and possibly provide a goodness of fit value) according to their own native language speech catagories. Best has predicted varying degrees of difficulty based on the assimilation pattern that emerges. For example, when two non-native sounds fit nicely into two different native categories, the contrast is perceptually easy to make. On the other hand, if the two non-native sounds fit into only one native category, the contrast may be much more difficult perceptually. Several other patterns are possible, however, in the present study, the two afforementioned patterns are what emerged. From informal discussions with subjects following the pretest, many Japanese assimilated the Hindi dental stop to their Japanese dental stop. More importantly, they assimilated the Hindi retroflex stop to their flapped /r/ consonant. English subjects tended to categorize both Hindi consonants to either a /t/, /d/, /θ/, or /ð/ depending on the voicing/manner context. Occasionally, they distinguished the two as if they were a voicing contrast /t - d/ or /θ- ð/. These data are inconsequential at the moment due to the informality of the data collection process. Note however that Polka also found similar non-benefitial assimilation strategies among English listeners on these Hindi consonants (Polka, 1989,1991polka-89,polka-91). Part 2 of this technical report series reports data from a questionnaire regarding subjects' perceptions of assimilation.

Finally, in general, one of the speakers (Speaker 2) was more intelligible than the other. While speaker effects are common in cross-language perception studies (e.g., Logan, Lively, and Pisoni, 1991 [29]), the effect in this study was more pronounced for English speakers, less so for Japanese speakers, and virtually non-existent for Hindi speakers. One plausible explanation for this is that non-phonemic acoustic variance (e.g., intonation, voice quality, pitch, speaking rate, etc.) should have less of an effect on perception for native than for non-native speakers. Since the dental-retroflex contrast is less nonnative in some sense for Japanese speakers than English speakers because of their contrastive experience, this decreasing effect of speaker logically follows. On the other hand, vowel context only affected the performance of Japanese speakers - English speakers performed equally poorly in all three vowel contests. These vowel effects may have occurred because of the noted coarticulation effects for both Hindi retroflex consonants and the flapped /r/ of Japanese, as noted previously (Dave, 1977 [8]; Dixit, 1990 [10]; Dixit & Flege, 1991 [11]; Hattori, 1984 [18]; and Ogawa, 1982 [31]).

6 Conclusions

Clearly, Japanese speakers had an advantage over English speakers on the Hindi contrast. This finding indicates that the flapped /r/s relationship to

/d/ in Japanese is quite different than the relationship of flapped /r/ to /d/ in English. Further, the Japanese flapped /r/ may be similar to some of the retroflex stop-consonants in Hindi; acoustical analysis may shed further light on this relationship - a series of acoustical analysis are currently being conducted on the stimuli used in this study. Finally, the results of this study support the idea that phonetic experience which is primarily non-contrastive may not be sufficient to aid in perceiving certain non-native contrasts. It should be pointed out that the present study controlled for prior experience and motivation of the listeners by using a target language which neither language group had much exposure to or necessity to learn. Thus, this may have provided a purer measure of the role of native language phonology on perceptual difficulty than trying to compare across studies. For example, comparing the difficulty that Japanese have with /r-l/ to English speaker's difficulty with Hindi dental and retroflex is flawed due to motivation to learn and previous experience with the language at hand - Japanese students are exposed to English and are encouraged to learn English from an early age.

Part 2 of this technical report investigates further the role of native language phonology on non-native perception by providing training to the subjects in the present study. Subsequent to training, all subjects were given a posttest which assessed learning and generalization of learning in order to further understand the differences in perceptual ability on Hindi due to native language backgraound. Stay tuned!

7 Acknowledgment

The authors are grateful to Professors Winifred Strange and James J. Jenkins at University of South Florida, Dr. Yoh'ichi Tohkura, president of ATR HIP Labs, and Dr Hideki Kawahara, head of Department 1, ATR HIP Labs, for providing us the opportunity to fulfill this study. Although we conceived the idea, their efforts and support made it a reality. We also thank Ms. Masako Tanaka for running Japanese subjets.

8 Appendix

Additional Information on Hindi and Retroflexion

While English tends to use dental and alveolar somewhat interchangeably, some languages, like Malayalam, may contrast all three places; that is, they contrast dental, alveolar, and retroflex articulations (e.g., mu/t̪/t̺, muttu, muṭṭu; Ladefoged, 1982 [25]; see also Jongman, Blumstein, & Lahiri, 1985 [22]). The dental/alveolar place-of-articulation appears to be more widespread in its use, phonemically, than retroflexion. In the languages of the world, some places of articulation are preferred to others. For stop consonants, some 16 places of articulation are cited. The dental/alveolar place occurs in virtually every language - followed by velar (99%), bilabial, palatal (18%), uvular (14%), RETROFLEX (11%), labial-velar, glottal, and pharyngeal (Henton, Ladefoged, and Maddieson, 1992; Ladefoged & Bhaskarara o, 1983; Maddieson, 1980). The principal retroflexed consonants are ṭ, ḍ, ṇ, ḷ, ṣ, and ẓ (Ladefoged, 1982). In addition, vowels may also be retroflexed (Jones, 1964 [21]). Typically, any language that incorporates a retroflex consonant a lso incorporates its dental cognate (Bhat, 1973 [5]; Stevens & Blumstein, 1975 [40]). Thus, if there is a voiced retroflex fricative (/ẓ/) in a particular language, t here is likely to be a voiced dental fricative (/z̪/) in that language also. (Additionally, there are retroflex vowels but this will not be discussed further.) According to Bhat (1973, 1974) [5, 6], the use of retroflexion tends to occur in geographical clusters. There are four main areas of the world that use retroflexion phonemically; India, Australia-Southeast Asia, the Pacific coast of America, an d mid-Africa (Bhat, 1974 [6]).

There is some controversy as to the standard classification of retroflexion. Some argue that it is actually a manner of articulation and not a place (Bhat, 1974 [6]; Dixit, 1990 [10]). Some argue for a distinction between retroflexion and retraction (Bhat, 1974). Production studies of retroflexion have revealed that the place of articulation can vary greatly (from pre-alveolar to almost post-palatal) and so can the degree of curling of the tongue (Dave, 1977 [8]; Dixit, 1990 [10]; Dixit & Flege, 1991 [11]; Firth, 1957 [13]; Ladefoged & Bhaskararao, 1983 [26]; Svarny & Zvelebil,1955 [41]). This variation in production occurs across speakers, phonetic cont exts, dialects, and languages. For example, vowel context has been shown to affect the degree of retroflexion - front vowels lessen retroflexion while back vowels increase it (Dave, 1977 [8]; Dixit, 1990 [10]; Dixit & Flege, 1991 [11]). For this reason, the acoustic consequences of retroflexion are not well understood (Tees & Werker, 1984 [42]; Werker & Tees, 1984 [46]). Most acoustic studies show the differences between dental and retroflex as being primarily in the burst and the transition of the second, third, and fourth formants - a lower frequency burst spectrum, which is usually higher in amplitude, and a bunching or clustering of F2, F3, and F4 which rapidly moves towards the appropriate vowel (Dave, 1977 [8]; Ladefoged & Bhaskararao, 1983 [26]; Polka, 1991 [33]; Ramasubramanian & Thosar, 1971 [37]; Stevens & Blumstein, 1975 [40]; see also Jongman, Blumstein, & Lahiri, 1985 [22]).

References

[1] C. T. Best. Emergence of language-specific constraints in perception of non-native speech perception: A window on early phonological development. In P. Jusczyk P. MacNeilage B. de Boysson-Bardies, S. de Schonen and J. Morton, editors, *Developmental Neurocognition: Speech And Face Processing In The First Year Of Life*. Kluwer, Dordrecht, 1993.

[2] C. T. Best. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In H. Nusbaum and J. Goodman, editors, *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press, 1994.

[3] C. T. Best and W. Strange. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20:305–330, 1992.

[4] C. T. Best, McRoberts G. W., and N. N. Sithole. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click perception by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14:345–360, 1988.

[5] D. N. S. Bhat. Retroflexion: An areal feature. *Working Papers on Language Universals*, 13:27–67, 1973.

[6] D. N. S. Bhat. Retroflexion and retraction. *Journal of Phonetics,*, 2:233–237, 1974.

[7] B. Bloch. Studies in colloquial Japanese iv: Phonemics. *Language*, 26:86–125, 1950.

[8] R. Dave. Retroflex and dental consonants in Gujarati: A palatographic and acoustic study. Annual Report of the Institute of Phonetics 11, University of Copenhagen, 1977.

[9] P. Delattre. *Comparing the phonetic features of English, French, German, and Spanish.* George G. Harrap and Company Ltd., London, 1966.

[10] R. P. Dixit. Linguotectal contact patterns in the dental and retroflex stops of Hindi. *Journal of Phonetics*, 18:189–201, 1990.

[11] R. P. Dixit and J. E. Flege. Vowel context, rate, and loudness effects on linguopalatal contact patterns in Hindi retroflex /t/. *Journal of Phonetics*, 19:213–229, 1991.

[12] G. H. Fairbanks and B. G. Misra. *Spoken and written Hindi.* Cornell University Press, New York, 1966.

[13] J. R. Firth. Word palotagrams and articulation. In *Papers in Linguistics*, pages 1934–1951. Oxford University Press, London, 1957.

[14] J. E. Flege. The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast. *Folia Linguistica*, 18:117–138, 1984.

[15] J. E. Flege. Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, 86:1684–1697, 1989.

[16] S. Gillette. Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, 6:59–72, 1980.

[17] H. Goto. Auditory perception by normal Japanese adults of the sounds L and R . *Neuropsychologia*, 9:317–323, 1971.

[18] S. Hattori. *Onseigaku [Phonology].* Iwanami Shoten, Tokyo, 1984. (Text is written in Japanese).

[19] D. G. Jamieson and D. E. Morosan. Training non-native speech contrasts in adults: Acquisition of the English θ - ð contrast by francophones. *Perception & Psychophysics*, 40:205–215, 1986.

[20] D. G. Jamieson and D. E. Morosan. Training new non-native speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, 43:88–96, 1989.

[21] D. Jones. *An outline of English phonetics.* W.Heffer and Sons Ltd., Cambridge, 1964.

[22] A. Jongman, S. E. Blumstein, and A. Lahiri. Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, 13:235–251, 1985.

[23] E. H. Jorden. *Beginning Japanese, Part 1.* Yale University Press, New Haven, 1963.

[24] H. J. Keselman and J. C. Rogan. A comparison of the Modified-Tukey and Scheffe methods of multiple comparisons for pairwise comparisons. *Journal of the American Statistical Association*, 73:47–51, 1977.

[25] P. Ladefoged. *A course in phonetics.* Harcourt Brace Jovanovich, New York, 1982.

[26] P. Ladefoged and P. Bhaskararao. Non-quantal aspects of consonant production: A study of retroflex consonants. *Journal of Phonetics*, 11:291–302, 1983.

[27] L. Lisker and A. S. Abramson. The voicing dimension: Some experiments on comparative phonetics. In *the 6th International Congress of Phonetic Sciences.*, Prague: Academia., 1970.

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

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

[30] S. E. Marten. *Essential Japanese.* Charles E. Tuttle Co., Rutland, Vermont, 1962.

[31] Y. Ogawa. *Nihongo Kyoiku Jiten [Japanese Education Dictionary].* Taishuhkan Shoten, Tokyo, 1982. (Text is written in Japanese).

[32] L Polka. *The role of experience in speech perception: Evidence from cross-language studies with adults.* Unpublished doctoral dissertation, University of South Florida, 1989.

[33] L. Polka. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89:2961–2977, 1991.

[34] L. Polka. Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52:37–52, 1992.

[35] P. J. Price. *A cross-linguistic study of flaps in Japanese and in American English.* Unpublished Ph.D. Thesis, University of Pennsylvania, 1981.

[36] J. S. Pruitt. *Training native English speakers to identify Hindi retroflex-dental consonants.* Unpublished Masters Thesis, University of South Florida, 1992.

[37] N. Ramasubramanian and R. B. Thosar. Synthesis by rule of some retroflex speech sounds. *Language and Speech*, 14:65–85, 1971.

[38] A. Sheldon and W. Strange. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3:243–261, 1982.

[39] M. Shibatani. *The Languages of Japan.* Cambridge University Press, Cambridge, 1990.

[40] K. N. Stevens and S. E. Blumstein. Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, 3:215–233, 1975.

[41] O. Svarny and K. Zvelebil. Some remarks on the articulation of the cerebral consonants in Indian languages, especially in Tamil. *Archiv Orientalni*, 23:374–434, 1955.

[42] R. C. Tees and J. F. Werker. Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, 38:579–590, 1984.

[43] S. E. Trehub. The discrimination of foreign speech contrasts by adults and infants. *Child Development*, 47:466–472, 1976.

[44] J. F. Werker. The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics*, 7:141–156, 1986.

[45] J. F. Werker and J. S. Logan. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37:35–44, 1985.

[46] J. F. Werker and R. C. Tees. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75:1866–1878, 1984.