CORRELATION BETWEEN CONSONANTAL VC TRANSITIONS AND DEGREE OF PERCEPTUAL CONFUSION OF PLACE CONTRAST IN HINDI

Manjari Ohala and John J. Ohala+*

*San Jose State University
+University of California, Berkeley

ABSTRACT

A previous study of the VC transitions characteristic of the five places of articulation in Hindi stops (labial, dental, retroflex, palatal, and velar) showed considerable overlap between different places correlated with the quality of the preceding vowel. For example, [labial] and [dental] had similar transitions following /i/; [labial] and [velar], following /u/; and [dental] and [palatal] following /u/ and /a/. To find out whether these patterns of acoustic overlap correspond to perceptual neutralization and to evaluate the role of the VC transitions independent of the final stop release, a perception experiment was run: 21 native Hindi listeners identified the final stop in /pVC/ syllables where V was /i a u/ and C was /p t̪ ʈ tʃ k/ and where the final stop release was either left in or gated out. The overall probability of correct place judgments was .86 with the final stop release but only .63 when it was gated out. In general in the gated condition the confusions were well predicted by the similarity of the formant transitions.

1. INTRODUCTION

Speech sounds must be (a) different from their neighboring sounds (a syntagmatic requirement) and (b) different from all the other sounds that might have appeared in the same context (a paradigmatic requirement). These two constraints can interact with each other. For example, in English (and many other languages [5]) in native vocabulary the labial-velar approximant [w] fails to occur as C2 in syllable initial consonant clusters where C1 is [labial]. This is presumably motivated by the fact that the consonantal transitions from a labial consonant create a formant trajectory that is too similar to that of a [w] to be reliably differentiated from it. Thus, although /w/ is otherwise paradigmatically distinct from other approximants that can appear in C2 position, i.e., /j l r/, it is syntagmatically too close to other labials.

One might suppose that paradigmatic and syntagmatic similarity would especially be a problem in languages that supported a relatively large number of contrasts. Hindi, for example, has 34 consonant distinctions, not counting singleton (short) vs. geminate (long) contrasts. Among stops (including affricates) it has four laryngeal distinctions (voiced, voiceless, voiceless aspirated, and breathy-voiced) and five place distinctions (labial, dental, retroflex, palatal, velar) [7]. Table 1 presents a minimal set showing the five place contrasts in coda position.

In an earlier study [6] of the shape of the VC formant transitions characteristic of the stops with the five place distinctions, it was found that many transitions were visually quite similar in specific vowel contexts. Specifically, the following were similar:

- Labials and Dentals have similar transitions after /i/.
- Retroflexes and velars have similar transitions after /i/.
- Dentals and palatals are similar after /a/ and /u/.
- Labials and velars are similar after /u/.
- Dentals and velars are similar after /a/.

The question arises: are there other parameters which carry the necessary place cues or are these places in fact highly confusable with their differentiation requiring a stop release with its rich place cues present?

pip	*pus*
pit̪	*yellow*
piʈ	*beat (verb)*
pitʃ	*raised garden row*
pik	*spit (betel juice)*

Table 1: Hindi words (and their English translations) showing contrast of five places of articulation in stops in coda position.

Another issue we sought to explore was a claim by Steriade [9] regarding the context where retroflex place distinction is most robust. Steriade starts with the view, similar to that of Ohala & Kawasaki [4] and Ohala & Kawasaki-Fukumori [5], that contrasts are "...permitted (or licensed) in positions that are high on a scale of perceptibility." Most of her arguments are based on neutralization of certain laryngeal contrasts but, referring to a study by Dave [2], she also considers the case of retroflexes and indicates that since the VC transitions of retroflexes are more prominent than their CV transitions, the position that should show more neutralization involving retroflexion is the onset (CV) position (i.e. word-initial or post-consonantally). (This is in contrast to other segment types where CV cues are said to be more prominent than VC.) Anderson [1] finds support for Steriade's claims based on perceptual data from Western Arrernte intervocalic dental/alveolar/retroflex contrasts. She found that the vowel preceding the retroflex shows the most prominent formant transitions and these formant transitions increase correct identification more than is the case with alveolar stops.

A perception experiment was designed to elucidate these issues.

2. THE EXPERIMENT

In a previous study [6] CVC syllables uttered in the frame /vo __ aja/ ("he, ___, came") were acoustically analyzed from 3 male native speakers of Standard Hindi. For the present perceptual study again three speakers were used, two of which were the same as those from the earlier study. (One of the speakers from the earlier study often released his final stops into the vowel of the following word and thus another speaker whose stop releases could be better isolated was selected.) As before, the stimuli were words (most of them nonsense words) of the form /pVC/ where V = [i a u] and C = a voiceless unaspirated stop that was bilabial, dental, retroflex, palatal, or velar. Thus the resulting tokens were of the sort [puk], [pap], [pitʃ], etc. (In the earlier study on formant transitions the final stops were voiced. Voiceless stops were used here because it facilitated the isolation of the final stop release. We don't believe the voicing of the final stops would have any significant effect on the overall trends in the results.) The data, digitized at 16 kHz after low-pass filtering at 8 kHz, were further processed to yield two versions of each word type: the "whole word" version, and a gated version that excluded the signal after the halfway point in the silence of the final stop, i.e., excluding the final burst. No part of the frame sentence was included. Thus the stimulus list contained 90 items: 3 vowels x 5 stops x 2 conditions x 3 subjects = 90. This 90 item word list was randomized and recorded on a tape with an interstimulus of 4 s. After every 5 tokens a 10 second gap was given (to allow subjects to find their way in case they lost their place on the answer sheet -- although none did). Preceding the 90 item word list was a 10 item practice session to familiarize the subjects with the way the stimuli would sound, how long they would have to respond, and with the way they were to mark their answer sheet (giving the candidate answers in the Devanagari script). These 10 tokens were similar to the words for the main test except they were prepared from /pVC/ tokens where the C was voiced. Five were in the "whole word" format and 5 in the gated format. For these, too, there was a 4 second gap between tokens. Between the 10 tokens of the practice session and the tokens of the main test there was a 20 second gap. This test was administered over headphones via a high-quality portable tape playback system in 21 subjects' homes in India.

i-gated	**p**	**t̪**	**ʈ**	**tʃ**	**k**	**i-whole**	**p**	**t̪**	**ʈ**	**tʃ**	**k**
p	**90.5**	7.9	0.0	0.0	1.6		**95.2**	4.8	0.0	0.0	0.0
t̪	54.0	**36.5**	6.3	0.0	1.6		4.8	**71.4**	23.8	0.0	0.0
ʈ	19.0	39.7	**36.5**	1.6	3.2		1.6	0.0	**92.1**	4.8	1.6
tʃ	20.6	39.7	11.1	**11.1**	17.5		0.0	1.6	17.5	**76.2**	4.8
k	3.2	11.1	0.0	0.0	**81.0**		0.0	0.0	0.0	0.0	**95.2**

Table 2: Confusion matrix (in percent) for /iC/; left: with release gated; right: with release intact. Correct response along diagonal.

u-gated	**p**	**t̪**	**ʈ**	**tʃ**	**k**	**u-whole**	**p**	**t̪**	**ʈ**	**tʃ**	**k**
p	92.1	6.3	0.0	0.0	1.6		**84.1**	6.3	9.5	0.0	0.0
t̪	3.2	**90.5**	4.8	1.6	0.0		31.7	**65.1**	3.2	0.0	0.0
ʈ	9.5	17.5	**66.7**	4.8	1.6		0.0	0.0	**95.2**	4.8	0.0
tʃ	3.2	50.8	3.2	**38.1**	4.8		0.0	0.0	0.0	**95.2**	4.8
k	49.2	0.0	3.2	1.6	**42.9**		1.6	0.0	1.6	0.0	**92.1**

Table 3: Confusion matrix (in percent) for /uC/; left: with release gated; right: with release intact. Correct response along diagonal.

a-gated	**p**	**t̪**	**ʈ**	**tʃ**	**k**	**a-whole**	**p**	**t̪**	**ʈ**	**tʃ**	**k**
p	**92.1**	4.8	0.0	1.6	1.6		**93.7**	4.8	1.6	0.0	0.0
t̪	9.5	**81.0**	6.3	0.0	1.6		0.0	**90.5**	9.5	0.0	0.0
ʈ	1.6	6.3	**87.3**	4.8	0.0		0.0	0.0	**95.2**	4.8	0.0
tʃ	3.2	54.0	7.9	**25.4**	9.5		0.0	0.0	0.0	**95.2**	4.8
k	3.2	19.0	0.0	0.0	**73.0**		0.0	0.0	0.0	0.0	**95.2**

Table 4: Confusion matrix (in percent) for /aC/; left: with release gated; right: with release intact. Correct response along diagonal.

3. RESULTS

The results are given in Tables 2-4 in the form of confusion matrices (in per cent). The results from the gated condition are on the left and those from the whole word condition on the right. The original consonant categories are listed on the leftmost column and the response categories on the top row. Thus in Table 2, the intersection of the /t̪/ row and the /p/ column (on the left side) shows that 54% of the responses to the gated /t̪/ were /p/.

4. DISCUSSION

Overall, the rate of correct place identification was 63% when the release burst was gated out but 86% for the whole word condition (χ^2 = 41.5, df = 1, p< .001). Only in the case of /pu/ and /t̪u/ were the identification scores less in the whole word condition than in the gated. The rate of 63% is far above chance but it amounts to greater than 1 error for every 3 stimuli. The stimuli here represent careful “laboratory” speech; one may guess that casual connected speech would have a much lower rate of correct identification of place were these sequences presented in isolation. Insofar as connected speech is intelligible, one must credit this to higher order redundancies, e.g., lexical, grammatical, semantic, pragmatic. We do not know what an “acceptable” level of intelligibility should be in a redundancy-free context such that communication will succeed in ordinary redundant situations.

The 37% errors in the gated stimuli were not distributed evenly. The /a/ context had the least errors (72% correct), then the /u/ context (66%) and the /i/ context had the most errors (51%). This difference is significant (χ^2 = 11.4, df = 2, p<.01) The major particular confusions in the gated condition, in descending order, were the following:

it̪	>	ip	54%
atʃ	>	at̪	54%
utʃ	>	ut̪	50.8%
uk	>	up	49.2%
itʃ	>	it̪	39.7%
iʈ	>	it̪	39.7%
itʃ	>	ip	20.6%
iʈ	>	ip	19%
ak	>	at̪	19%
uʈ	>	ut̪	17.5%

Understandably, the palatal stop (the affricate) is very poorly identified without its affricated release. It is most often confused with the dental stop. Confusion between /p/ and /t̪/ after /i/ is high as has previously been found by others [8, 10] studying CV sequences, although in our data the direction of the confusion is the reverse from these earlier studies where it was the sequence /pi/ most often confused with /t̪i/. The confusion between /k/ and /p/ after /u/ is high and this, too, parallels previous findings [10] (although in their case the confusion was largely symmetrical). From the listing above one sees that the most common responses in the errors were /p/ and /t̪/.

This asymmetry in the errors may be due to listeners expecting to hear a burst and, when they didn’t, deducing that the stops must have had very low intensity bursts that were inaudible. Listeners presumably know from experience that /p/ and /t̪/, among all the stops, have the weakest bursts (due to the fact that they, unlike stops at the other places of articulation, have little or no downstream resonator serving to reinforce and amplify the noise burst). In general, the greater percentage of correct responses for the labial and dental place may reflect a kind of response bias because labial or dental were also the most common erroneous responses for the other places.

Of the five noted similarities in the shape of formant transitions from the earlier study [6], four of them corresponded to high rates of confusion in in the present study, namely, labial and dentals after /i/, dentals and palatals after /a/ and /u/, labials and velars after /u/, and velars and dentals after /a/. An expected confusion between retroflexes and velars after /i/ did not manifest itself -- possibly because the retroflexes, unlike the velars, tended to induce a characteristically low F3 throughout the preceding vowel.

There was also some confusions that were not predicted, at least based on a visual examination of the formant transitions, namely, palatals and retroflexes were confused with both dentals and labials after /i/ and retroflexes were confused with dentals after /u/.

Over all three vocalic environments, the rank order of percent correct identification of place was, from more to less: labial (91.6%), dental (69.3%), velar (65.6%), retroflex (63.5%), and palatal (24.9%).

The degree to which place perception was improved, on average, by the addition of the final stop release varied considerably:

palatals	64%
retroflex	31%
velar	29%
dental	6.3%
labial	- .57%

This ranking is roughly in the same order as the typical prominence and robustness of the releases of these stops. The palatal (affricate) is known to have an intense characteristic fricative release and the back-articulated velar and the retroflex stops would also tend to have relatively intense bursts by virtue of their having downstream resonators to amplify their sound.

Regarding Steriade’s [9] claim: it is true that the retroflex has robust VC transitions after [a], and corresponding to this the percent correct identification in the gated conditions was a high 87.3%. However, after [i] this is not true: the transitions are less robust [2, 6] and the rate of correct identification was only 36.5%. Thus claims about the robustness of the retroflex place distinction in VC context need to take into consideration the nature of the V. Additionally, another factor may have to be taken into account in evaluating the relative salience of retroflexes’ cues in some VC positions. It is possible that via sound change the presence of a syllable final retroflex consonant has yielded a different preceding vowel, e.g., in Dave’s [2] study the /ə/ before a retroflex may no longer be the same /ə/ before other consonant places. The data presented by Dave [2] for Gujarati retroflexes suggests, for example, that the preceding vowel itself, not just the transitions, may differ from the "same" vowel before dentals. Similarly, in the history of English a post-vocalic /r/ has had major influences on preceding vowels [3]. One example is that after metathesis changed *brid* to 'bird' the vowel underwent lowering and centralization. (We cite this as an example of how a consonantal contexts can influence the phonological quality of vowels. It is not necessarily directly relevant to retroflexes' influence since the original /r/ in English may not have been retroflex.) If such a vowel change is involved, the identification of final retroflexes may in some cases be helped by a correlated difference in vowel quality, not just by the consonantal transitions *per se*.

One final note of caution: although it seems reasonable to think that VC intervocalically (the position reported on by Anderson [1], Dave [2], and Steriade [9]) and VC in final position (used in the present study) would behave similarly for formant transitions and how they are perceived, this might not be so and still needs to be examined.

5. REFERENCES

1. Anderson, V., "The perception of coronals in Western Arrernte," *Proc. Eurospeech, Rhodes, , Greece, 22-25 September 1997. 1*.389 - 392, 1997.

2. Dave, R., "Retroflex and dental consonants in Gujarati. A palatographic and acoustic study," *Annual Rep. Inst. Phonetics, Univ. Copenhagen, 11*.27-156, 1977.

3. Dobson, E. J., *English pronunciation 1500-1700*. 2nd ed. Oxford: Oxford University Press, 1968.

4. Ohala, J. J. & Kawasaki, H., "Prosodic phonology and phonetic," *Phonology Yearbook, 1*.113 - 127, 1984.

5. Ohala, J. J. & Kawasaki-Fukumori, H., "Alternatives to the sonority hierarchy for explaining segmental sequential constraints," In S. Eliasson & E. H. Jahr (eds.), *Language And Its Ecology: Essays In Memory Of Einar Haugen*, Berlin: Mouton de Gruyter, 343-365, 1997.

6. Ohala, M., "Acoustic study of VC transitions for Hindi stops," *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, August 13-19, 1995*, vol *4* pp 22-25, 1995.

7. Ohala, M., "Hindi," *Journal of the International Phonetic Association, 24*:1.35-38, 1994.

8. Plauché, M., Delogu, C., & Ohala, J. J., "Asymmetries in consonant confusions" *Proc., Eurospeech 97. Rhodes, Greece, 22-25 Sept. 1997, 4*.2187 - 2190, 1997.

9. Steriade, D., *Phonetics in phonology: The case of laryngeal neutralization.* UCLA. Ms., 1997.

10. Winitz, H., M. Scheib, and J. Reeds, "Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech," *J. Acoust. Soc. Am., 51,* 1309-1317, 1972.