PERCEPTION OF MALAYALAM THREE-WAY STOP CONTRAST AMONG AMERICAN ENGLISH SPEAKERS

Meghavarshini Krishnaswamy, Natasha Warner

University of Arizona
mkrishnaswamy@arizona.edu, nwarner@arizona.edu

ABSTRACT

Acoustic proximity of novel segments to existing categories affects perception. Competition also erodes relationships between a category and familiar segments; affecting identification. This study examines the pattern of perception of three novel Malayalam coronal stops (dental, alveolar, retroflex) by American English listeners. Participants responded to three-way forced-choice questions on consonant identity in VCV sequences. Accuracy and proportion of responses were analysed before and after exposure to the contrast.

Alveolars were more likely to be identified correctly post-exposure, but not retroflexes, and the trend for dentals was significantly different from alveolars. Accuracy trends were not significant post-exposure. Retroflexes were significantly less confusible, and participants responded ‘retroflex’ fewer times for audio without them after exposure.

Participants do not start with a biased towards any stop, they could tell the difference between retroflexes and the other two stops- ‘coronal stop’ exist as one category, its perception has directionality and an observable trend.

Keywords: Novel contrast, Malayalam, speech perception, three-way contrast

1. INTRODUCTION

This study examines the perception of the Malayalam three-way geminate coronal stop contrast (dental /t̪ː/, alveolar /tː/ and retroflex /ʈː/) among adult speakers of American English with no prior exposure.

1.1. Coronal three-way stop contrast in Malayalam

Malayalam is a Dravidian language with a large phonemic inventory, mainly spoken in the Indian State of Kerala [1]. It exhibits a three-way contrast intervocalically among geminate stops involving the tongue-tip, blade and dorsum: dental /t̪ː/, alveolar /tː/ and retroflex /ʈː/ [2]. In this paper, “coronal contrast” refers to this three-way contrast. It occurs only with certain neighboring vowels, and there are few minimal triplets (example: /pət̪ːi/ ‘snake’s hood’, /pət̠ːi/ ‘what was said’, /pəʈːi/ ‘dog’).

Articulatory differences among coronal stops are well-researched in this and other Indic languages. Dart and Nihalani show that dental stops are articulated as ‘interdentals’ or ‘dentialveolars’; alveolars make a constriction in the region behind the upper teeth and about 5mm away from it, and retroflexes make a constriction well away from the alveolar ridge, with both tongue tip and underblade contact [2]. Dentals also show longer constriction duration than alveolars. Thus the dental-alveolar stop distinction is one of constriction placement and length. Retroflexes stand out in terms of their place of articulation and tongue region contact, and have been reported to show tongue curling (constriction involving the sublaminal region) in other Dravidian languages [3]. At the acoustic level there are differences in Voice Onset Time (VOT) [4, 5], F3 lowering near retroflex stops (in Tamil, Hindi and Punjabi), and F2 raising (Tamil) [3]. Further, preceding vowels also show colouring and lengthening [3]. Dutta & Redmon et al. also show differing VCV coarticulatory patterns [6]. Thus, there are potential acoustic cues to the three-way distinction.

1.2. Coronals in English

Among British and American varieties of English, word-medially, /t/ is realised as a flap (‘water’), an oral stop (‘attention, stem’), or a glottal stop, depending on stress and environment. Dental stops only occur before an interdental fricative (‘width’, ‘breadth’), and are rather rare [7]. Thus English speakers frequently hear alveolar oral stops and rarely hear a context-sensitive dental allophone.

Based on acoustic correlates and point of articulatory contact, the closest candidate for retroflection in English would be the rhotic and the rhotic schwa. While rhotics in American English can be bunched or retroflex [8], both have third formant transitions that are similar to Indic retroflection [3]. Coronal stops may coarticulate with a following rhotic (‘attraction’, ‘truck’, ’trash’), creating allophonically retroflex stops.

The three coronal stops would thus constitute a novel contrast for native English listeners that uses familiar acoustic cues. This experiment investigates whether the three-way distinction can be perceived by naive English listeners at all, and whether they have a bias toward any of the categories.

1.3. Novel Contrast Perception and the PAM Model

The Perceptual Assimilation Model (PAM) [9] discusses the perception and categorization of novel sounds outside of the native sound inventory of a language, based on proximity to native sound categories. A novel sound may be perceived as a ‘good’ or ‘bad example’ of a native category, it may fall outside of any native sound category, or be perceived as outside the speech stream.

PAM proposes four possibilities for how a distinction between two sounds of another language might relate to native sound categories:

1. ‘Single category (SC)’: novel sounds match one native category equally well. For example, Japanese listeners perceive both English /l/ and /r/ as poor examples of the Japanese category /r/ [10].
2. ‘Different category goodness (CG)’: one novel sound is a better example of a native category than the other. Example, Japanese perception of the English /w/-/r/ contrast shows better recognition of /r/ than of /w/ [11].
3. ‘Two categories (TC)’: novel sounds are binned in different categories, thus improving identification of both. For example, voicing contrast in Zulu lateral fricatives is perceived as two different categories by English speakers [10].
4. ‘Undefined space (UC/UU)’: one or both segments may fall outside any perceptual category. For example, perception of Zulu clicks by English listeners [12].

Among the first three, the TC shows the highest level of discrimination, while SC shows the lowest. CG falls somewhere in between. [10].

Training on Indic two-way coronal contrasts has shown SC or CG type of perception. Experiments on Hindi dental-retroflex stop contrast comparing infants and adults from an English-speaking background showed that adults performed only at chance, both with and without training [13]. Comparison between performance of American English monolinguals, strong Bengali-English and Spanish-English bilinguals showed that bilinguals exhibited better levels of discrimination (that is, more participants moved to the CG and TC type of perception) for alveolar-retroflex contrast among laterals, rhotics and nasal stops after long training sessions (for non-native contrasts) [14].

Previous work on phonemic perception has been limited to two-way contrasts (e.g. dental vs. retroflex) and did not include Malayalam alveolar stops. The current study analyzes a three-way contrast within a small articulatory space.

1.4. Research Questions

The first research objective is to test whether the three Malayalam stops will be perceived differently at all. If the contrast is perceived, the second question is whether any of the three stops is more detectable than another. Third, we can examine whether the listeners exhibit bias toward any of the categories.

2. MATERIALS AND METHODS

2.1. Materials

This perception experiment used 65 nonsense minimal triplets, in the pattern V_1CːV$_2$ (vowel-coronal geminate-vowel). For example, the stimuli for the vowel context ‘iCːo’ are [it̪ːo, itːo, iʈːo]. The vowel set used in this experiment is [ə, i, o, u, e, æ, ɔ]. In this paper, ‘ə’ symbol is used for the central mid vowel, which can be both stressed and unstressed in Malayalam. The last two are not phonemic in Malayalam, but were included to ensure that the study had a sufficient number of trials. The stimuli were recorded by a phonologist whose L1 is Malayalam to ensure that both the coronal contrast and the [æ-a, ɔ-o] distinction could be produced, while keeping the prosody of all words as similar as possible. The list of words included thirteen real words. Since the listeners have no knowledge of Malayalam this is not expected to affect results.

2.2. Participants

37 Native speakers of American English were recruited through Mechanical Turk (MT) (9 female, 15 male, others unknown), with a 90% or above rating on previous studies, and previous participation in at least 10 MT tasks. Of these 13 reported speaking only English, 16 reported studying Spanish or other languages in school, and 7 chose not to answer. One participant reported Malayalam and English as their L1 and their data was excluded.

2.3. Procedure

The experiment was run using Finding Five [15]. It began with a description of tongue shapes for the three-way contrast (see Appendix) and the symbols used for them, and also included instructions for keypress responses. The description was included because the native English listeners would otherwise have no information on which response symbol meant what. After the introduction, three trials tested their retention of the three symbols, with questions on tongue gestures associated with each symbol (ex.“choose the symbol for the sound made when the tongue cuts off the flow of air behind the teeth”). The next three trials were practice trials identical to the test trials discussed below.

The task was three-way alternative forced choice (AFC), in which the participant listened to a stimulus, and chose one of three options displayed on the screen. Listeners first responded to 60 pre-exposure test items, after having received only the written description of the categories described above. A filler stimulus containing an unrelated consonant was included after every 6 stimuli to make the task less repetitive. (This pre-exposure phase was included to test another hypothesis, but will not be discussed further here, as listeners were not able to identify the categories well based on abstract descriptions and the number of pre-exposure items gave low power.) Listeners pressed a key on the keyboard to indicate their response.

Listeners were then exposed to the three-way contrast by hearing 30 items of the same type as the test stimuli, with the correct response symbol shown on the screen while they heard it, and no response required. The first 15 items were presented as triplets in the order dental, alveolar, retroflex. The second 15 items consisted of the same minimal triplets with order within each triplet being random.

After this brief exposure phase, the listeners were presented with an additional 132 post-exposure test trials in the same format as the pre-exposure trials. All vowel neighbourhoods were included for this phase, resulting in 65×3 trials. The post-exposure test phase was followed by a language background questionnaire.

3. RESULTS

Table-1 shows the post-exposure results (36 participants, 44 minimal triplets) as a confusion matrix (accuracy in the diagonal, in bold). Chance is 33%. All statistical analyses on this data was conducted using the *lme4* package in R [16]. Maximal models was attempted; i.e. if a factor was within subjects, random slopes were included for that factor by subjects, and if a factor was within-minimal-triplets, random slopes were included for that factor by minimal triplets. If the maximal model failed to converge, random slopes were removed until convergence was achieved.

Target→ *Response*↓	**Dental**	**Alveolar**	**Retroflex**	**Response Given**
Dental	**612 (38.64%)**	550 (34.72%)	508 (32.07%)	1670 (**35.41%**)
Alveolar	561 (35.42%)	**635 (40.09%)**	525 (33.14%)	1721 (**36.22%**)
Retroflex	411 (25.95%)	399 (25.19%)	**551 (34.79%)**	1361 (**28.64%**)
Total	1584	1584	1584	4752

Table 1: Confusion Matrix for post-exposure trials

3.1. Accuracy

A GLME was fitted to examine the effect of Place on accuracy (Equation (1)). Accuracy for alveolars was not significantly different from accuracy for dentals and retroflexes ($p > 0.1$ for both dentals and retroflexes). If listeners thought almost all the stimuli sounded like alveolars (bias toward alveolar), they might respond with, say, 80% alveolar, regardless of what the stimulus was. This would be reflected as high accuracy for alveolar stimuli and low accuracy for the other two categories, but would only reflect bias, not ability to perceive a difference. The lack of a significant effect of Place on accuracy suggests that any bias is small.

$$\begin{aligned} &Accuracy \sim Place\ Of\ Articulation\ + \\ &(1 + Place\ Of\ Articulation \mid Participant) + \\ &(1 + Place\ Of\ Articulation \mid Minimal\ Triplet) \end{aligned} \quad (1)$$

However this does not show if listeners can perceive the difference among the three coronal categories. For this, the next subsection investigates detectability of the three-way distinction by examining proportion of a particular response instead of accuracy.

3.2. Detectability of distinctions among coronal stops

In order to assess whether listeners can detect the difference among the coronals, the analysis assessed the number of times a listener responded ‘dental’ when the stimulus really had a dental stop versus when it had an alveolar or a retroflex stop. This was done correspondingly for the ‘alveolar’ and ‘retroflex’ responses. These results appear in Table-1 (distributed row-wise). For example, participants responded ‘alveolar’ 635 times when presented with alveolar stops, 561 times when presented with dental stops, and 525 times when presented with retroflex stops.

For the dependent variable ‘alveolar response’ (yes/no), a GLME was fit with place of articulation of the stimulus as the fixed effect. Random intercepts were included for each triplet and place of articulation, and random slopes were included for each triplet. The same was done with 'dental response' and 'retroflex response' as dependent variables. In each case the category that represented the dependent variable was also used as the reference level of the factor 'place of articulation'. The model used is represented in (2):

$$Response \sim Place\,Of\,Articulation + (1 + Place\,Of\,Articulation \mid Participant) + (1 \mid Triplet) \quad (2)$$

Listeners were significantly more likely to respond 'alveolar' to alveolar stimuli than to retroflex stimuli ($p < 0.02$). However, they were not significantly more likely to respond 'alveolar' to alveolar stimuli than to dental stimuli ($p > 0.05$). This suggests that the listeners were able to hear some difference between alveolar and retroflex stimuli, but not between alveolar and dental stimuli. With 'dental' response as the dependent variable, alveolar and dental stimuli did not differ significantly ($p > 0.1$), but retroflex stimuli were significantly less likely to be identified as dental than dentals stimuli ($p < 0.05$). With 'retroflex' response as the dependent variable, participants were significantly more likely to identify retroflex stimuli as retroflex than to identify either other category as retroflex. Retroflex stimuli was also misidentified significantly fewer times as alveolars stops ($p < 0.005$) and dentals stops ($p < 0.02$). This analysis indicate that listeners could detect some difference between retroflexes and the other two categories, but could not hear a difference between alveolars and dentals.

4. DISCUSSION

This experiment was designed to examine whether a novel three-way contrast could be perceived by speakers with access to allophonic acoustic cues associated with alveolar stops, interdental stops, and retroflection in their language; and whether a short exposure to the contrast was enough to allow them to detect differences among the three categories.

The analysis with response accuracy as the dependent variable verifies that listeners do not have a strong bias either toward or away from any one of the three categories. They do not, for example, respond 'alveolar' predominantly because of that sound being most similar to their native category.

The analysis of detectability shows that even after a brief exposure to this novel three-way distinction, listeners have some significant ability to distinguish retroflexes from the other two categories. Listeners respond in the same way regardless of whether the stimulus is alveolar vs. dental, showing that they do not detect any difference between these two categories. In terms of detectibility of the three-way distinction, American English listeners show some ability to distinguish retroflexes from the other two categories, collapsing this three-way distinction to a two-way one. This could be due to the availability of robust acoustic cues that speakers are already sensitive to in the retroflection of rhotics (since low F3 is a strong cue for American English /r/). Familiarity with retroflection among sonorant rhotics may be helpful in the identification of retroflection among stops. Alternatively, assimilation of English /t/ to a following rhotic (as in 'attraction, truck, trash') might be providing enough experience with a rhotic stop to allow listeners to detect that category, even though that assimilation is allophonic rather than phonemic in American English.

The three-way contrast appears to show the 'different category goodness' (CG) pattern of assimilation, with all three stops mapping to one category, but retroflexes being less confusable. One possibility here is that retroflex stops are perceived to be closer to the exemplar 'coronal' category, even with a brief exposure. The CG prediction and results are further supported by the low accuracy results for each category, and lower detectablility for alveolars and dentals than for retroflexes. This result is consistent with previous research on Indic two-way coronal contrasts (dentals vs retroflexes, alveolars vs retroflexes).

There are interesting acoustic and lexical effects that could be playing a role here. The low lexical frequency of allophonic interdentals in English (e.g. 'width, breadth' and very few other words) could make the alveolar/dental distinction more difficult to learn. Acoustically, the English /tr/ sequence includes strong perceptual cues to retroflection, including increased VOT/aspiration along with low F3. This may contribute to the listeners' ability to distinguish retroflex from other categories. Future research on novel phoneme perception, especially when expanded to multiple choices for a given small perception space as in this three-way distinction, can provide us with greater insights into the interaction between acoustic cues, allophonic patterns, and amount of past exposure to a sound type.

5. REFERENCES

[1] C. Chandramouli, "Census of India 2011," *Language, States and Union Territories. New Delhi: Government of India*, pp. 13–14, 2018.

[2] S. N. Dart and P. Nihalani, "The articulation of Malayalam coronal stops and nasals," *Journal of the International Phonetic Association*, vol. 29, no. 2, pp. 129–142, 1999.
[3] S. Bakst, "Rhotic and Retroflexes in Indic and Dravidian," Master's thesis, Trinity College, University of Cambridge, June 2012.
[4] E. Chodroff, A. Golden, and C. Wilson, "Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. EL109–EL115, 2019.
[5] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20, no. 3, pp. 384–422, 1964.
[6] I. Dutta, C. Redmon, M. Krishnaswamy, S. Chandran, and N. Raj, "Articulatory complexity and lexical contrast density in models of coronal coarticulation in malayalam," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds., 2019, pp. 1992–1996.
[7] D. Eddington, "Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions," *Cognitive Linguistics*, vol. 18, no. 1, pp. 23–46, 2007. [Online]. Available: https://doi.org/10.1515/COG.2007.002
[8] X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4466–4481, 2008.
[9] C. T. Best, "Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development," in *Developmental neurocognition: Speech and face processing in the first year of life*. Springer, 1993, pp. 289–304.
[10] C. T. Best, G. W. McRoberts, and E. Goodell, "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *The Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 775–794, 2001. [Online]. Available: https://doi.org/10.1121/1.1332378
[11] C. T. Best and W. Strange, "Effects of phonological and phonetic factors on cross-language perception of approximants," *Journal of phonetics*, vol. 20, no. 3, pp. 305–330, 1992.
[12] C. T. Best, G. W. McRoberts, and N. Sithole, "The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by english-speaking adults and infants," *Journal of Experimental Psychology: Human perception and performance*, vol. 14, no. 3, pp. 345–360, 1988.
[13] J. F. Werker, J. H. Gilbert, K. Humphrey, and R. C. Tees, "Developmental aspects of cross-language speech perception," *Child development*, pp. 349–355, 1981.
[14] D. V. Gogoi, *Acquisition of novel perceptual categories in a third language: the role of metalinguistic awareness and feature generalization*. University of Florida, 2009.
[15] Team FindingFive, "Findingfive: A web platform for creating, running, and managing your studies in one place. findingfive corporation (nonprofit), NJ, USA," 2019.
[16] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

6. APPENDIX

The text used to introduce the three-way contrast to participants is as follows:

"Malayalam language has three types of stops produced with the tongue tip and tongue body. They are: Dental (tongue tip stops air-flow at the back of the teeth) , Alveolar (tongue front stops air-flow at the alveolar ridge,bumpy part behind the teeth), Retroflex (the tongue is curled such that the tongue tip stops air-flow behind the alveolar ridge)"

The following table lists words used as stimuli in the post-exposure condition. Real words are highlighted in grey.

Exposure Condition	Triplet ID	Triplet Template	Real Word?	Repeated in both exposure conditions?
post-exposure	1	ɔCə	N	N
post-exposure	2	ɔCo	N	N
post-exposure	3	ɔCe	N	N
post-exposure	4	ɔCu	N	N
post-exposure	5	ɔCi	N	N
post-exposure	6	ɔCɔ	N	N
post-exposure	7	ɔCæ	N	N
post-exposure	8	əCɔ	N	N
post-exposure	9	oCɔ	N	N
post-exposure	10	eCɔ	N	N
post-exposure	11	uCɔ	N	N
post-exposure	12	iCɔ	N	N
post-exposure	13	æCæ	N	N
post-exposure	14	æCə	N	N
post-exposure	15	æCo	N	N
post-exposure	16	æCe	N	N
post-exposure	17	æCu	N	N
post-exposure	18	æCi	N	N
post-exposure	19	æCɔ	N	N
post-exposure	20	æCa	N	N
post-exposure	21	əCæ	N	N
post-exposure	22	oCæ	N	N
post-exposure	23	eCæ	N	N
post-exposure	24	uCæ	N	N
post-exposure	25	iCæ	N	N
post-exposure	26	əCa	N	N
post-exposure	27	oCa	N	N
post-exposure	28	eCa	Y	N
post-exposure	29	uCa	N	N
post-exposure	30	iCa	N	N
post-exposure	31	ɔCa	N	N
post-exposure	32	aCa	N	N
post-exposure	33	aCə	N	N
post-exposure	34	aCe	N	N
post-exposure	35	aCu	Y	N
post-exposure	36	aCi	N	N
post-exposure	40	əCa	N	N
post-exposure	41	oCa	N	N
post-exposure	42	eCa	Y	N
post-exposure	43	uCa	N	N
post-exposure	44	iCa	N	N