MISMATCHED COARTICULATORY INFORMATION HINDERS LEXICAL ACCESS OF CORONAL STOPS IN MALAYALAM

Seema GP, Meghavarshini Krishnaswamy, Ramesh Mishra, Indranil Dutta

Technische Universität Dresden, University of Arizona, University of Hyderabad, Jadavpur University

gp.seema@gmail.com, mkrishnaswamy@arizona.edu, rkmishra@uohyd.ac.in, indranildutta.lnl@jadavpuruniversity.in

ABSTRACT

Coarticulatory information affects lexical access, more so among dense inventories. We explore the Malayalam dental, alveolar and retroflex stop contrast. In this eye-tracking study, VC:V words were used where 'C' was one of the three stops. Consonants were cross-spliced with either the same stop (match condition) or one of the other two from an identical vocalic environment (mismatch condition) to generate real word audio stimuli. Participants viewed two words, followed by an audio. Pupillary fixations were analysed for 3 conditions- target matched with audio; audio from the 'mismatch condition' with the the distractor and audio either sharing a stop or not.

Participants looked at the target and distractor at the same rate in the second condition. Coarticulatory information from two different consonants in the audio impacts lexical access. Presence of matched coarticulatory information results in higher looks at the target, even if the distractor is from the same phonological cohort.

Keywords: coarticulation, lexical access, dense coronal inventory, eye-tracking, Malayalam

1. COARTICULATORY INFORMATION AND LEXICAL ACCESS

One thread of research highlights the significance of segmental coarticulation as an aid to speech perception ([1, 2]). Coarticulatory information also effects lexical access and activation [3]. Research on coarticulatory resistance has also shown that segments exhibit variable degrees of coarticulatory resistance ([4, 5, 6]). While coarticulatory information is used successfully by listeners to both perceive and access lexical items, it has not been shown how coarticulatory resistance interferes in this process, especially when the segmental contrasts are relatively dense or comparably sparse. In this eye-tracking study, we show how coarticulatory information as a result of dynamic gestural overlap interacts with lexical access. This experiment is aimed to tease out the fine-grained interaction of contrastive and coarticulatory information in aiding lexical access.

Match-mismatch studies where auditory information is presented to subjects alongside visual stimuli within a modified visual world paradigm show that listeners pay keen attention to coarticulatory information for successful lexical access and activation. There are a number of studies that explore the subtle relationship between fine-grained phonetic information and lexical activation. For instance, listeners are found to use prosodic cues such as vowel duration to anticipate word length and constrain competing lexical choice [7]. In an eye-tracking paradigm, subjects look preferentially at those bisyllabic items that are consistent with auditory cues exhibiting vowel durations stemming from bisyllabic items. Similarly, when cross-spliced words where misleading coarticulatory cues for final consonants are tested against consistent patterns, subjects tend to show effects of these inconsistencies vis-a-vis lexical activation and competition (within the visual world paradigm) [8]. More recently, Beddor, McGowan, et al. show that listeners' moment-by-moment fixations on visual displays are mitigated by coarticulatory acoustic effects in real-time [3]. While these studies point to how coarticulatory information is successfully utilized for lexical access; the interaction of articulatory-motor and auditory-perceptual information, and the attendant effect on lexical access has not been studied. There is hardly any information about how varying inventory sizes could impact both coarticulatory influence and lexical access in complex ways.

In this study, we report on an eye-tracking study where cross-spliced matched and mismatched auditory stimuli are presented to the subjects in a modified visual world paradigm.

2. CORONAL STOPS IN MALAYALAM

Malayalam exhibits a number of coronal segments that makes it interesting to test the predictions of articulatory-motor and auditory-perceptual constraints in languages. Malayalam has a three-way contrast in place of articulation among plosives showing differences in tongue tip and tongue blade involvement (and sublaminal regions in the case of the retroflex), namely dental /t̪ː/, alveolar /tː/ and retroflex /ʈː/ [9]. This three-way contrast is only available intervocalically, and the plosives are always geminate in this context. Thus, Malayalam poses interesting questions on both these counts due to the fairly large number of contrasting segments in the dental/alveolar region. We are especially interested in determining how dynamic information (that is, the relationship between coarticulatory cues in the preceding and succeeding vowels, and the place cues in the articulation of the consonant itself) about the consonant place of articulation in the adjoining vowels can be successfully utilized by listeners to access the lexicon and what (if any) effect phonological cohorts may have in mitigating lexical access and activation. To that end, we present subjects with auditory stimuli under matched coarticulatory information, and mismatched coarticulation in a visual world experiment. Our results indicate that listeners gaze at printed words with coarticulatory information regardless of the presence or absence of a phonological cohort word. We argue that in dense consonantal systems, in addition to the activation of the cohort, coarticulatory information (ostensibly, the dynamics of variable articulatory overlap) is encoded in the lexicon.

3. MATERIALS AND METHODS

3.1. Materials

72 words were chosen for this experiment containing a V_1C:V_2 sequences, where V_1 belonged to the /ə, o, i, u, e/, V_2 was from the set /ə, aː, i, o, u, e/, and Cː is one of the three coronal geminate stops. The words chosen for this study are a subset of the ones used in [10], with four words excluded because the vowel neighbourhood only had one of the three stops.

First, three instances of each word were recorded in isolation by a female native speaker of Malayalam, using a fixed cardioid condenser microphone (Shure Beta 53), and a sampling rate of 22.5 kHz. Next, these recordings were used to create stimuli using using Praat [11] for two conditions: match and mismatch. For each condition, the stop closure and burst of the coronal stop were removed and replaced with the stop closure and burst from the audio of a different word. The stop closure was identified as the earliest point with low or no energy in the signal. The end of the burst was marked as the last point before the appearance of formant bands in the signal while ensuring that the burst was included completely in the extraction. This process separates the acoustic information of the stops' place of articulation observed in the stop (closure duration, burst amplitude) from that observed in the neighbouring vowels (formant transitions, F3 changes (retroflexes)).

For the match condition, closure and burst from one of the other instances of the same word were used for the cross-splicing (coarticulatory information before closure matched information of the consonant). For the mismatch condition, two stimuli were generated where possible, using a word with the same vocalic environment (i.e. identical V_1 and V_2), but a different coronal stop. This yielded 72 tokens for the match condition, and 105 tokens for the mismatch condition, with every word represented at least once in each condition. Figure-1 is an example of the original recording, the match condition, and the mismatch condition for the token [et̪ːi]. The stimuli in the second frame was generated by replacing the closure and burst from one of the other two recordings of [et̪ːi], while the stimuli in the third frame used the closure and burst from the token [eʈːi] (Figure 1). Following the cross-splicing, four native speakers of Malayalam listened to the stimuli in the mismatch condition, and provided qualitative feedback on the stimuli, to ensure that the words were still identifiable in the altered tokens.

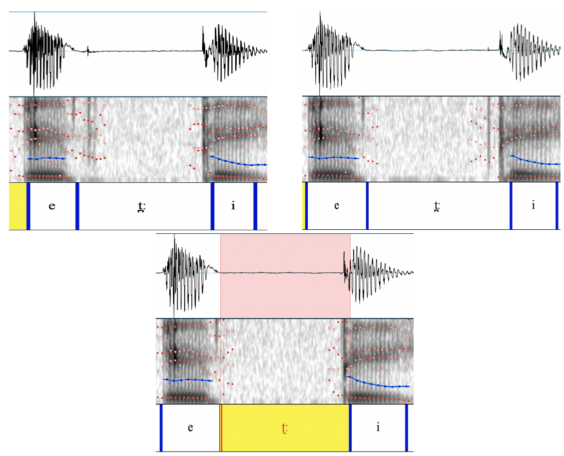

Figure 1: original recording, matched token and mismatched token for the word "et̪ːi"

3.2. Methods

We chose eye-tracking because it allowed us to study the effects of keeping or removing given acoustic cues in a stimuli in a task involving lexical

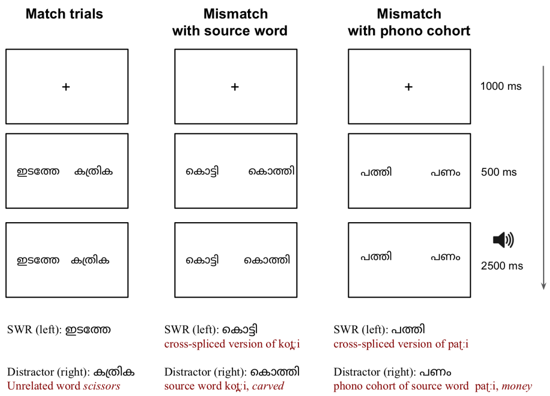

Figure 2: Sequence of events on a sample trial for the three experimental conditions. Note: SWR denotes spoken word referent

processing. 12 participants (4 female, mean age = 25 years, SD = 2.4) took part in the study. All participants were students at University of Hyderabad and had acquired Malayalam as their L1. All participants reported normal or corrected-to-normal vision. Eye movement data was recorded using a desktop mounted Eyelink 1000 eyetracker (SR research, Ontario) with a sampling rate of 1000 Hz under binocular viewing. Participants rested their head on a chin rest for stable viewing. A 9-point calibration was used for each participant. Stimuli was presented on a 19 inch LCD monitor with a refresh rate of 60 Hz placed at a distance of 60 cm from the participant.

Each trial began with a fixation cross for 1000 ms. A preview of the visual world display consisting of two printed Malayalam words was presented for 500 ms. Next, the Malayalam spoken word was presented and the visual world display was shown for 2500 ms longer. All the stimuli were presented in black against a grey background. There were four types of trials:

1. *Match trials* where the spoken word contained matching coarticulatory information to one of the printed words. The other word in the display was an unrelated distractor. These trials were included to serve as a baseline to demonstrate that our paradigm can capture the biased looks towards spoken word referents.
2. *Mismatch trials with source word, or "Mismatch (source)"* where the spoken word contained mismatched coarticulatory information with one of the printed words.The other word in the display was the source word that matched the coarticulatory information in the spoken word.
3. *Mismatch trials without source word "Mismatch (phono cohort)"* where the spoken word included mismatched coarticulatory information. The display contained the printed version of the spoken word and another word that was phonologically similar to the spoken word (but didn't match the coarticulatory information).
4. *Filler trials* where both the words in the display were unrelated to the spoken word. The filler trials were included to break anticipation and strategy.

Each participant was administered 10 match trials, 10 mismatch trials with source word, 6 mismatch trials without source word and 10 filler trials.

4. RESULTS

Fixations to the two printed words were extracted using SR research Dataviewer (SR research, Ontario). A square box of 5° by 5° around each word was considered an area of interest (AOI) for the eye movement analyses. The fixations falling within the AOIs from the onset of the spoken word till end of trial were considered for analyses. For each participant, the proportion of fixation to each word was calculated as the number of fixations to that word divided by the total number of fixations in the 1500 ms duration on each trial.

Statistical analyses on fixation data was performed using "lme4" package in R ([12]). Fixation data was logit transformed and analysed using mixed effects logistic regression models. *Glmer* function was used with family specified as "binomial" and "link logit". Participants and items were added as random effects in all analyses.

The fixation data on match trials was analysed to first investigate if spoken words elicited looks towards the referent word. Object type ("spoken word referent", "distractor") was added as a fixed effect with "Spoken word referent" as the baseline. Subject-wise and item-wise random slopes for Object type were also included in the model. Looks to the spoken word referent were significantly greater than looks to the unrelated distractor (t = - 3.6, p = 0.002) indicating that our paradigm was effective in capturing language-mediated eye movements.

Similar analyses was conducted on filler trials with object type ("distractor1", "distractor2") as a fixed effect. As expected, there was no difference in looks towards the two words (t = -0.33, p = 0.751).

The two mismatch conditions were then analysed by including condition (source word, phonological cohort of the source word) and object type (spoken word referent, distractor) as fixed effects with

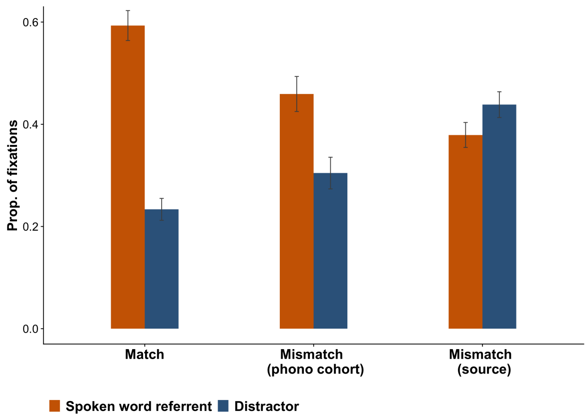

Figure 3: Proportion of fixations for the "Match" and the two "Mismatch" conditions. In the "Match" condition, the distractor was an unrelated word. In the "Mismatch (source)" condition, the distractor was the source word whereas in the "Mismatch (phono cohort)" condition, the distractor was the phonological cohort of the source word

"source word" and "spoken word referent" as baseline, respectively. Subject- and item-wise random slopes for both the fixed effects were also added to the model. There was no effect of object type (t = -0.18, p = 0.321) with descriptively greater looks to the spoken word referent compared to the distractor across both conditions. Crucial to our hypothesis, there was a significant interaction between condition and object type (t = -2.75, p = 0.015) indicating that participants looked at the spoken word referent more often only when the distractor word was a phonological cohort of the source word. Interestingly, when the distractor word was the source word, participants looked at the source word and the spoken word referent equally often because of the source-word related coarticulatory information present in the spoken word.

5. DISCUSSION

In this experiment, we investigated the influence of coarticulatory information on lexical access in a sample of Malayalam-English bilinguals in India using the visual world paradigm with printed words. We showed that, upon hearing a spoken word cross-spliced with coarticulatory information from a source word with a different coronal stop, participants gazed at this source word as often as they looked at the printed word referred to by the spoken word. We verified that this pattern of results was driven by the coarticulatory information in the spoken word (and not by the phonological similarity between the spoken word and the source word) by including a condition where the phonological cohort of the source word was presented on the screen along with the spoken word referent. In this condition, more looks were seen towards the spoken word referent, confirming that phonological similarity was not the driving our main results. We also had a baseline condition where the spoken-word referent was presented along with an unrelated word. More looks towards the referent word were seen as is typically observed in visual world studies.

These findings, in addition to general findings of Marslen-Wilson and Welsh [13], help us argue for encoding of the dynamic nature of articulatory overlap and the attendant acoustic variation in the lexicon. This, we feel, is more so the case in consonantal systems that are dense in terms of the potential for articulatory overlap. The Malayalam three-way coronal contrast serves as a testing ground for the nature of this encoding. In recent acoustic investigation of the formant dynamics of Malayalam, it has been found that in VC:V contexts, Malayalam alveolars offer the greatest coarticulatory resistance compared to the retroflex and dentals [10].

Previous research shows that despite the greater relative articulatory complexity of the retroflex (where both tongue tip and tongue dorsum are involved), alveolars resist coarticulation from the adjoining vowels, more so than the retroflexes and dentals, in that order [10, 14]. This seemingly unusual result has been argued to be due to the low neighborhood density of the alveolars in Malayalam. Alveolars appear only in the medial position, and without a voiced counterpart, compared to the ubiquitous retroflexes and (to a lesser degree) dentals.

The absence of the burst information in the mismatched stimuli also points to the strength of the acoustic cues relevant for place of articulation discrimination when compared to dynamics of the formants, as found in the offset of the preceding vowel and onset of the following vowel.

Our results, when read along with the acoustic study of Dutta, Redmon, et al.[10], underscore the need to understand encoding of coarticulatory information at a deeper level for successful lexical access – when listeners need to access a denser cohort. They also dovetail with the results of studies that find that coarticulatory information aids speech perception [1, 2], in addition to the need to encode lexicons with fine-grained phonetic detail.

6. REFERENCES

[1] J. L. McClelland and J. L. Elman, "The TRACE model of speech perception," *Cognitive psychology*, vol. 18, no. 1, pp. 1–86, 1986.

[2] C. A. Fowler, "An event approach to the study of speech perception from a direct–realist perspective," *Journal of phonetics*, vol. 14, no. 1, pp. 3–28, 1986.

[3] P. S. Beddor, K. B. McGowan, J. E. Boland, A. W. Coetzee, and A. Brasher, "The time course of perception of coarticulation," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2350–2366, 2013.

[4] R. A. W. Bladon and A. Al-Bamerni, "Coarticulation resistance in English/l," *Journal of Phonetics*, vol. 4, no. 2, pp. 137–150, 1976.

[5] C. A. Fowler and L. Brancazio, "Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation," *Language and Speech*, vol. 43, no. 1, pp. 1–41, 2000.

[6] D. Recasens and A. Espinosa, "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan," *Journal of the Acoustical Society of America*, vol. 125, pp. 2288–2298, 2009.

[7] A. P. Salverda, D. Dahan, and J. M. McQueen, "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," *Cognition*, vol. 90, no. 1, pp. 51–89, 2003.

[8] D. Dahan, J. S. Magnuson, M. K. Tanenhaus, and E. M. Hogan, "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," *Language and Cognitive Processes*, vol. 16, no. 5-6, pp. 507–534, 2001.

[9] S. N. Dart and P. Nihalani, "The articulation of Malayalam coronal stops and nasals," *Journal of the International Phonetic Association*, vol. 29, pp. 129–142, 1999.

[10] 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*, 2019.

[11] P. Boersma, "Praat, a system for doing phonetics by computer." *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[12] 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.

[13] W. D. Marslen-Wilson and A. Welsh, "Processing interactions and lexical access during word recognition in continuous speech," *Cognitive Psychology*, vol. 10, no. 1, pp. 29–63, 1978. [Online]. Available: https://www.sciencedirect.com/science/article/pii/001002857890018X

[14] M. Krishnaswamy, I. Dutta, and M. Bhaumik, "Alveolar stops exhibit greater coarticulatory resistance than retroflexes and dentals in malayalam," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2582–2582, 2020.