ANTICIPATORY LABIALIZATION IN C[u] CLUSTERS: A CROSS-LINGUISTIC STUDY

Veno Volenec

Concordia University
veno.volenec@concordia.ca

ABSTRACT

This paper presents a cross-linguistic investigation of anticipatory labialization (AL) in C[u] clusters. It addresses two underexplored research questions: (1) Is AL always present in a consonant (C) that is followed by [u] in all languages? (2) Is there a cross-linguistic difference in the temporal unfolding of AL in C[u] clusters?

Lip movements of 10 native speakers of 10 languages were video-recorded during the production of all the C[u] clusters that exist in each language. Participants read isolated words beginning with C[u] (e.g., *rule* [ɹuɫ]) from a screen; this trial was repeated three times. The duration of AL before [u] was determined using OpenFace2.2.

The results showed that: (1) all 10 languages displayed AL in every token of every C; (2) there is a statistically significant difference in the duration of AL in C from C[u] between the 10 languages, ranging from 83.0 ms (Japanese) to 109.7 ms (Brazilian Portuguese).

Keywords: anticipatory labialization, V-to-C coarticulation, cross-linguistic study

1. INTRODUCTION

Anticipatory labialization (AL) is a coarticulatory effect where the lip rounding gesture from a rounded sound, such as [u], is temporally extended in the regressive direction ('right-to-left' if the sounds are transcribed), influencing one or more of the preceding consonants [1, 2]. For example, in the English word *rule* [ɹʷuɫ], the lip rounding gesture originating from the vowel [u] encapsulates the entire production of the consonant [ɹ], so that this consonant, which otherwise does not entail lip rounding (e.g., *read* [ɹid]), becomes rounded or labialized.

Despite the vast experimental and theoretical literature on labial articulation and coarticulation in various languages [10, 11, 12, 13, 14, 15, 16], only a few studies have undertaken a cross-linguistic investigation of AL, and none of them have directly compared AL across more than two languages [3].

Lubker and Gay [4] used electromyography to compare AL in Swedish and American English. The results showed that while AL was present in both languages in all the syllables in which consonants appeared before rounded vowels, AL in Swedish was systematically produced with more extensive and more precise lip-protrusion movements than AL in English. Also, the timing of AL was found to differ in the two languages: for the same number of consonants, the onset of AL occurred earlier for the Swedish than for the English speakers.

Boyce [5] compared patterns of lip rounding in VCV sequences for speakers of English and Turkish. She measured EMG activity in the *orbicularis oris* muscle and found that Turkish speakers produced "plateau" patterns of movement, where lip rounding in a C is approximately the same as in both Vs, whereas in English they produced "trough" patterns, where lip rounding is greater in Vs than in a C. Because the study was done on VCV sequences, it cannot be determined whether the intervocalic C was influenced by the AL of the following vowel or the carryover labialization from the preceding vowel, nor if there are any temporal differences in the labialization of C between the two languages.

Noiray and colleagues [6] found no significant differences in AL in terms of lip protrusion and constriction between American English and Quebec French, while data on its scope and duration were not provided.

Thus, while previous research sheds some light on the similarities and differences in AL between languages, various aspects of it remain insufficiently explored, in particular whether it systematically occurs in all languages with rounded vowels and whether its spatio-temporal properties vary cross-linguistically. With the goal of contributing to this research area, this paper addresses two questions. First, is AL always present in a consonant that is followed by [u] or is it systematically absent in that context in some languages? Second, is there a cross-linguistic difference in the temporal unfolding of AL in C[u] clusters? In order to facilitate direct cross-linguistic comparison, this study only concentrates on AL originating from [u] (because [u] is typologically the most frequent rounded vowel [7]), and it only investigates AL in word-initial C[u] clusters (because

CV syllables are typologically most common [8], and because word-initial C[u] clusters ensure that only anticipatory and not carryover coarticulation is present).

2. METHOD

2.1 Participants

This study included 10 participants, seven of whom were female and three were male. Each person is a native speaker of one of the 10 studied languages: Brazilian Portuguese, Croatian, American English, Quebec French, Italian, Japanese, Jordanian Arabic, Lithuanian, Persian, and Telugu. The crucial criterion for the inclusion of a language in this study was for a language to contain the phoneme /u/ in its vowel inventory. Since the phonetic literature does not indicate that biological and social factors such as gender, age and level of education play a role in the realization of AL [3, 4], these parameters were not tracked.

2.2 Materials

For each of the 10 languages, a list of words was created. The purpose of these lists was to elicit pronunciations that feature AL. The lists contained only words that begin with C[u] clusters. Every particular language's word list covered all of the consonants that may appear word-initially before [u] in that language. Since the consonant inventories differ between languages, the word lists containing the C[u] clusters were also partially different. For example, the Japanese word list contained 10 different words, each beginning with a particular C[u] cluster (such as [nuka], [ɾusu], [t͡suika], etc.), while the Lithuanian word list contained 16 C[u] clusters. Six consonants appeared in all of the languages: [n, t, d, s, ʃ, g]. All labial consonants were deliberately excluded from the study in order to avoid any possible clash between the gestures associated with the labial place of articulation and labial coarticulation.

The word lists were compiled in collaboration with the participants. The investigator consulted the extant phonological descriptions of the 10 languages and compiled a list of all possible C[u] configurations, given the languages' consonant inventories. The participants were then asked to think of a single word in their language that featured each of those clusters. Thus, the word lists contained all of the possible word-initial C[u] clusters, with a single example for every such cluster. Importantly, the topic, the context and the goal of the study were not revealed to the participants during this process. For every language separately, the words were placed into an automated PowerPoint presentation in which a single word appeared on a slide every 6 seconds. The entire word list was presented three times in a row, so that ultimately three tokens per C[u] cluster were obtained. Such presentations were used to elicit pronunciations that feature AL.

2.3 Procedure and data analysis

During the experiment, the participants were seated in front of a computer with a video camera (a 2021 16-inch MacBook Pro). The camera was slightly below their eye level at an approximate distance of 75 cm. They were instructed to read the words of their native language as they appeared on a screen in a natural and spontaneous way. As they were pronouncing the words from an automated PowerPoint presentation, a camera was recording their entire face with a 1920x1080 pixel resolution at 30 frames per second.

In order to determine the duration of AL, the applications OpenFace2.2 and Wondershare Filmora X were used. OpenFace [9] is an automated facial behavior analysis toolkit driven by artificial intelligence. For every frame of a video, it places landmarks (dots) on a person's face and keeps track of the location of those landmarks in a 3D space (Figure 1). Out of 130 facial landmarks, 20 are placed over the lips. All of the video recordings gathered in the experiments were processed in OpenFace, which generated an Excel spreadsheet with 3D coordinates of lip movements during the production of all the C[u] clusters. These data showed the exact moment of the beginning of lip rounding, i.e., the onset on AL.

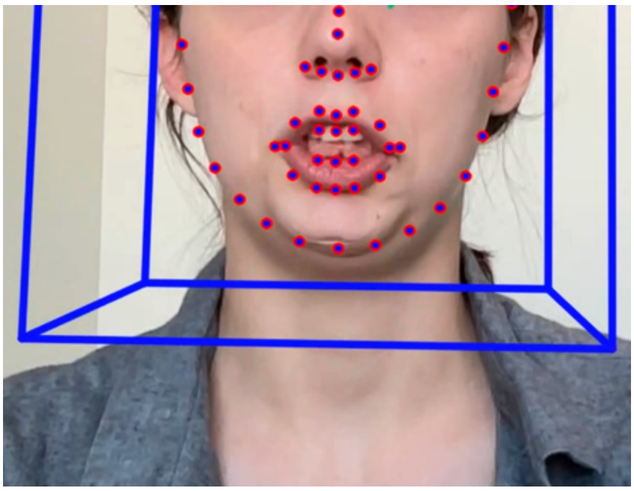

Figure 1: Automated facial landmark detection in OpenFace2.2. The software uses artificial intelligence to keep track of landmarks (blue dots in red circles) as they move in 3D space over time.

The video processing software Wondershare Filmora X was used to determine the ending of AL during the production of the consonants. Audio-visual inspection of the recorded videos, conducted frame by frame, was used to ascertain the beginning of the vowel [u]. Specifically, the beginning of [u] was established on the basis of two criteria: the onset of the auditory perception of [u] and the onset of the formant structure characteristic for the high back rounded vowel. The temporal distance between the onset of AL as determined by OpenFace and the onset of [u] as determined by Wondershare Filmora X was taken as the duration of AL. Thus, these two applications were used in order to determine both the presence versus absence of AL and the duration of AL in C[u] clusters.

Subsequently, the mean duration of AL was calculated for every language separately by summing the durations of AL in all C[u] tokens and dividing that by the number of tokens. One-way ANOVA was used to test if there are statistically significant differences in the duration of AL between the 10 tested languages.

3. RESULTS

All 10 languages displayed anticipatory labialization (AL) in every token of every word-initial consonant that was followed by [u]. There was not a single instance of an absence of AL in any of the tested words. Furthermore, in all of the tested words in all languages, the anticipatory lip rounding gesture preceded all consonantal acoustic cues. In other words, AL could always be observed as the first overt articulatory gesture associated with a C in a C[u] cluster, before any audible signal was emitted by the speaker. For example, in the production of the word *succo* [suk:o] 'juice', the Italian speaker started to round her lips 40 milliseconds before the audible friction associated with [s] began. The instant before [s] became audible is captured in Figure 2, showing that AL preceded the consonant's acoustic cue.

The average duration of AL (expressed in milliseconds) for every language is presented in Table 1. The total average duration of AL in all C[u] clusters calculated across all 10 language is 96.7 ms. The third column shows the respective standard deviations, which are relatively high, but strikingly similar cross-linguistically. Thus, the results show that different languages display very similar variation in the duration of AL.

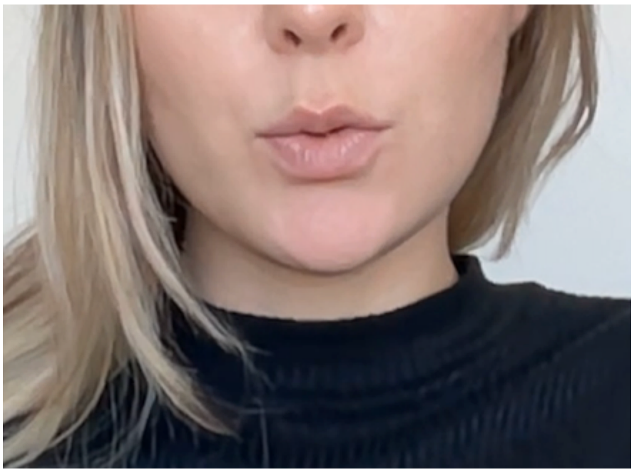

Figure 2: Anticipatory labialization as it is produced 10 ms before the onset of audible friction of [s] in a [su] cluster (in the word *succo* [suk:o] 'juice') by an Italian speaker.

Language	Mean duration of AL (ms)	Standard deviation
Brazilian Portuguese	109.7	18.6
Croatian	89.3	18.5
English	92.3	18.7
French	92.7	19.5
Italian	102.0	19.6
Japanese	86.3	18.7
Jordanian Arabic	104.2	20.1
Lithuanian	96.1	19.1
Persian	97.6	18.7
Telugu	96.7	19.7

Table 1: Mean duration of anticipatory labialization in C[u] clusters in 10 languages.

A one-way ANOVA showed that there is a statistically significant difference in the duration of AL in C from C[u] between the 10 languages: $F = 4.773$, $p = 0.01$. While there are no significant differences in AL between some languages, for example between English and French, some of them do display significant differences, the most prominent difference being between Japanese with the shortest mean AL of 86.3 ms and Brazilian Portuguese with the longest mean AL of 109.8 ms. The shortest observed instance of AL lasted for 40 ms and it appeared during [n] in the Persian word [nur] 'light'. The longest observed AL lasted for 150 ms during [ʃ] in the Brazilian Portuguese word [ʃuvə] 'rain'.

4. DISCUSSION

The results of this study showed that anticipatory labialization (AL) was always present during the production of a consonant that was followed by a high back rounded vowel [u]: it was consistently observed in all 10 of the tested languages, in all classes of consonants, in all tokens of all the words that were uttered during the experiment. The wide cross-linguistic distribution of AL cannot be attributed to historical factors because AL is ubiquitous in genetically unrelated languages; Telugu is a Dravidian language; Jordanian Arabic is an Afro-Asiatic language, Japanese is a Japonic language, and the rest belong to various sub-families of Indo-European. While previous research showed that it is likely that *particular parameters* of AL are determined by language-specific factors [4, 5, 10, 11, 12, 13, 14, 15, 16], the fact that it systematically appears in genetically unrelated languages with vastly different phonologies suggests that AL *as a general phenomenon* stems from universal (i.e., non-language-specific) phonetic principles. Following [17], [18] and [19], it can be assumed that the main motivation for AL is perceptual: as a result of anticipatory coarticulation, information about an upcoming sound, in this case [u], is available to the listener before that sound is fully articulated, and this prior information may facilitate more accurate perception than would be the case if all acoustic and articulatory cues were confined within the temporal boundaries of that sound.

The results related to the temporal unfolding of AL showed that it cannot be said that AL has a fixed universal duration cross-linguistically, because there is a statistically significant difference in the duration of AL between at least some of the examined languages. These results are consistent with the two previous cross-linguistic studies of AL [4, 5] that also found between-language differences in the spatio-temporal parameters of AL. It is likely that the temporal differences in AL result from the differences in the underlying phonological systems of the languages in question. It has previously been shown that syllable structure, prosody, and the contrastive vs. redundant nature of the feature [±ROUND] can all play a role in the realization of coarticulatory effects [3]. Thus, AL has a different duration in different languages because during speech production the effect of AL is superimposed onto different, language-specific phonological representations. The interaction between the language-specific phonology and the non-language-specific nature of AL explains why on the one hand languages can differ in the average duration of AL (because AL is manifested on different phonological systems) and why on the other hand all of the tested languages display strikingly similar AL variability as indicated by the standard deviations in Table 1 (because the variability of AL is due to idiosyncratic biomechanical factors that are the same irrespective of a speaker's native language).

The main significance of this study is that it provides the first direct cross-linguistic comparison of AL on a relatively large sample of languages, whereas previous studies have focused either on AL in individual languages [10, 11, 12, 13, 14, 15, 16] or on comparing two languages [4, 5, 6]. This work has several limitations, which indicate avenues for future research. First, it would be beneficial to expand the array of compared languages as much as possible in order to increase the reliability of cross-linguistic generalizations about AL. Second, this study has only focused on the temporal aspect of AL, leaving aside its spatial properties. Future cross-linguistic investigations of AL should also aim to determine how languages differ in the shape and extent of lip-rounding during AL. Third, the influence of phonological factors on the realization of AL merits further exploration in order to clarify what exactly it is that, for example, yields a short mean AL in Japanese (86.3 ms) and a long mean AL in Brazilian Portuguese (109.8 ms).

In summary, this paper explored anticipatory labialization in C[u] clusters from a cross-linguistic perspective. It addressed two research questions: is AL always present in a consonant that is followed by [u] in all languages; and is there a cross-linguistic difference in the temporal unfolding of AL in C[u] clusters? The results showed that all 10 of the studied languages displayed AL in every token of every consonant from a C[u] cluster, and that while AL can be seen as a universal phonetic phenomenon, the temporal properties of AL can vary significantly between languages due to their different phonologies.

5. ACKNOWLEDGEMENTS

The writing of this paper was made possible by the selfless support and encouragement of Elizabeth Tobyn, to whom I extend my heartfelt gratitude.

6. REFERENCES

[1] Farnetani, E., Recasens, D. (2013). Coarticulation and connected speech processes. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds.), *The Handbook of Phonetic Sciences*. 316–351. Oxford: Wiley-Blackwell.

[2] Volenec, V. (2015). Coarticulation. In: Davis, J. (ed.) *Phonetics: Fundamentals, Potential Applications and Role in Communicative Disorders*. 47–86. New York: Nova Science Publishers.

[3] Farnetani, E. (1999). Labial coarticulation. In: Hardcastle, W. J., Hewlett, N. (eds.), *Coarticulation: theory, data and techniques*. 144–163. Cambridge: Cambridge University Press.

[4] Lubker, J. F., Gay, T. (1982). Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America* 71, 437–448.

[5] Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America* 88, 2584–2595.

[6] Noiray, A., Cathiard, M. A., Abry, C., Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. In: Maassen, B., Van Lieshout, P. (eds.), *Speech Motor Control: New Developments in Basic and Applied Research*. 153–171. Oxford: Oxford University Press.

[7] Ladefoged, P, Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

[8] Gordon, M. (2016). *Phonological Typology*. Oxford: Oxford University Press.

[9] Baltrušaitis, T., Zadeh, A., Lim, Y. C., Morency, L. P. (2018). OpenFace 2.2: Facial Behavior Analysis Toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition*. URL: [https://github.com/TadasBaltrusaitis/OpenFace].

[10] Benguerel, A. P., Cowan, H. (1974). Coarticulation of upper lip protrusion in French. *Phonetica* 30, 41–55.

[11] Lubker, J. F., McAllister, R., Carlson, P. (1975). Labial co-articulation in Swedish: a preliminary report. In: Fant, C. G. M. (ed.), *Proceedings of the Speech Communication Seminar.* 55–64. Stockholm: Almqvist and Wiksell.

[12] Descout, R., Boë, L. J., Abry, C. (1980). Labialité vocalique et labialité consonantique. 111–126. *Labialité et Phonétique, Publications de l'Université des Langues et Lettres de Grenoble*.

[13] Bell-Berti, F., Harris, K. S. (1979). Anticipatory coarticulation: some implications from a study of lip rounding. *Journal of the Acoustical Society of America* 65, 1268–1270.

[14] Magno-Caldognetto, E., Vagges, K., Ferrigno, G., Busà, M. G. (1992). Lip rounding coarticulation in Italian. *Proceedings of the International Conference on Spoken Language Processing* 92/1, 41–64.

[15] Sussman, H. M., Westbury, J. (1981). The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. *Journal of Speech and Hearing Research* 24, 16–24.

[16] Perkell, J. S., Matthies, M., (1992). Temporal measures of anticipatory labial coarticulation for the vowel /u/: within- and cross-subject variability. *Journal of the Acoustical Society of America* 91, 2911–2925.

[17] Kühnert, B., Nolan, F. (1999). The origin of coarticulation. In: Hardcastle, W. J., Hewlett, N. (eds.), *Coarticulation: theory, data and techniques*. 7–30. Cambridge: Cambridge University Press.

[18] Fowler, C. A. (2005). Parsing coarticulated speech in perception: effects of coarticulation resistance. *Journal of Phonetics* 33, 199–213.

[19] Harrington, J., Hoole, P., Pouplier, M. (2013). New Directions in Speech Production. In: Jones, M. J., Knight, R. (eds.), *The Bloomsbury Companion to Phonetics*. 242–259. London: Bloomsbury.