The time course of onset CV coarticulation

Tugba Lulaci[1], Mechtild Tronnier[1], Pelle Söderström[1,2], Mikael Roll[1]
[1] *Centre for Languages and Literature, Lund University, Sweden*
[2] *MARCS institute, Australia*
tugba.lulaci@ling.lu.se, mechtild.tronnier@ling.lu.se, pelle.soderstrom@ling.lu.se, mikael.roll@ling.lu.se

Abstract

The study investigates the center of gravity in onset fricatives as a main acoustic feature to assess the relation between vowel pronunciation and coarticulatory spectral characteristics of the onset consonant. /s/- and /f/-initial CV sequences were analyzed with backness, roundedness and height of the vowel as predictors of fricative center of gravity. Results showed that the first 15 ms of an onset fricative could carry predictive cues to the upcoming vowel.

Introduction

Speech production consists of different steps. Models of speech production commonly include planning and execution as two main levels. As a speech sound category, fricatives occur at the end of various complex production steps (Pützer et al., 2019; Themistocleous, 2017). When we listen to spoken language, the auditory system extracts behaviorally pertinent information from acoustic stimuli, and rapid analysis of spectro-temporal patterns and components in the auditory stream is crucial for successful speech perception (Ardila & Bernal, 2016; Checkik & Nelken, 2012).

Coarticulation – the phonetic influence on a specific speech sound from surrounding sounds – plays an important role in spoken word recognition (McQueen, Norris & Cutler, 1999), along with factors such as prosody and formant states and transitions (Hardcastle & Hewlett, 1999). Speech processing occurs rapidly, and spectro-temporal cues to the identity of a specific phoneme usually can develop from over tens of milliseconds to hundreds (e.g., in the case of vowels). Considering the fact that normal speech has a rate of 8-15 phonemes per second, fast processing is required (Schreiber & McMurray, 2019). The effective analysis of upcoming speech sounds plays a role in making this process faster and more accurate. Furthermore, the continuous prediction of speech sound has been proposed to be central to successful spoken word recognition (Gagnepain, Henson & Davis, 2012).

Pre word-onset coarticulation in English with definite and indefinite articles is helpful for processing continuous speech, and coarticulation cues can be helpful for identification of novel words and real words (Salverda et al., 2014). We investigated the time course of acoustic correlates (fricative center of gravity) of CV coarticulation at word onset to determine at which point in time acoustic information about subsequent vowel articulation becomes available, with an aim to test the potential perceptual availability of these anticipatory acoustic features in the future. Previously, most studies investigating coarticulatory effects in fricatives on subsequent speech sounds have used online experimental techniques, such as eye-tracking (McMurray et al., 2018). In the present study, we used offline acoustic measurements to test the effects of coarticulation as possible cues for predicting upcoming vowels using linear regression models. We recorded words beginning with /s/- (Figure 1, Figure 2) and /f/-initial CV sequences (Figure 3, Figure 4), using vowel backness, roundedness and height as predictors, and the center of gravity of the onset fricative as the dependent variable to assess the relation between the spectral characteristics of the onset consonant and subsequent vowel pronunciation. Center of gravity (CoG) is the most used parameter for analyzing the spectrotemporal properties of fricatives (Zygis et al, 2015). Spectral features of fricatives have been researched widely, with a focus on sibilants. We operationalized the center of gravity (CoG) as the average frequency of a spectrum, weighted by the power spectrum. CoG is generally higher in fricatives and plosives compared to other consonants since they have relatively increased amplitude at higher frequencies (van Son & Pols, 1996).

Material and method

Twenty monosyllabic words beginning with /s/ and twenty monosyllabic words beginning with /f/ were recorded in isolation in citation form by a Central Swedish speaker. The recordings contained 17 different sets with 34 stimuli in total. A Praat script was used to identify the word onset and to calculate CoG for different time slices of the consonants. The CoG Praat function was used. It is defined as

$$\frac{\int_0^\infty f|S(f)|^p\, df}{\int_0^\infty |S(f)|^p\, df}$$

for a spectrum given by S(f), where f is the frequency and a power of p = 1 was used (Boersma & Weenink, 2021). Word onsets were identified as the point where sound intensity levels exceeded 35 dB, using another Praat script.

The CoG of /f/ and /s/ were measured in Praat. The CoG data was extracted in 4 slices for each word onset fricative. The slices were 15 ms, 35 ms, 75 ms, and 135 ms long, all beginning at word onset. The obtained data were analyzed using IBM SPSS Statistics (Version 28.0). Linear regression models assessed the relation between CoG as a dependent variable and the degree of vowel roundedness, backness, and height as independent variables. In several languages, backness and roundedness are related to each other. Front vowels are usually unrounded, while back vowels are usually rounded. A relationship between roundedness and height is also visible, high vowels being generally more rounded than low ones (Ladefoged & Maddieson, 1996). Six degrees of height were defined: low (0), mid (1), mid-high (2), high (3), near-high (4) and near-low (5). The predictor backness had three degrees: front (0), central (1), and back (2). Roundedness had two degrees: unrounded (0) and rounded (1). Roundedness, backness, and height degrees were derived from the phonological representation in Riad (2014). Our lists contain Swedish long and short vowel allophones. [ɑː] [ɔ] [uː] [oː] [ɵ] [œ] [œː] [øː] [ø̞] [ʏ] [ʉː] [yː] [iː] [ɪ] [eː] [a] [æ] [æː] [ɛ̝] [ɛː] were used for both the /f/ and /s/ list. Analyses were performed separately for /f/ and /s/.

As shown in Figures 1-4, we did not include the formant transition part in the analysis. The last slice ended just before the onset of the formant transition. Note that word onset times differed from each other since we used a script to define the word onset. We used the same vowels for the /f/ and /s/ words.

Results

Regression model for /s/

The regression for /s/ was statistically significant for the 15-ms slice ($R^2 = 0.287$, Adjusted $R^2 = 0.281$, $F_{(3, 336)} = 45.107$, $p < 0.001$). Roundedness ($p < 0.001$) and height of the subsequent vowel ($p < 0.001$) predicted fricative CoG while backness ($p = 0.216$), did not. The regression model for the 35-ms slice of /s/ was also statistically significant ($R^2 = 0.391$, Adjusted $R^2 = 0.385$, $F_{(3, 336)} = 71.848$, $p < 0.001$). Roundedness was the strongest predictor for 35-ms slice ($p < 0.001$) and height ($p = 0.016$) also predicted CoG while backness ($p = 0.195$) did not. Regression for 75-ms /s/ was found statistically significant ($R^2 = 0.587$, Adjusted $R^2 = 0.583$, $F_{(3, 336)} = 159.196$, $p < 0.001$). To formulate a regression equation with only significant terms, we re-ran the regression model for the first slice, using only roundedness ($p < 0.001$) and height ($p < 0.001$) as predictors $F_{(2, 337)} = 66.784$, $p < 0.001$ (equation 1). Examples of the /s/ words are shown in Figures 1 and 2.

$$CoG = 7061 - 1359\, roundedness - 134\, height \tag{1}$$

Roundedness ($p < 0.001$) and backness predicted CoG in the 75-ms slice ($p = 0.002$) while height did not ($p = 0.591$). We calculated the regression equation for the third slice to reflect the significant components, backness ($p = 0.002$) and roundedness ($p < 0.001$) as predictors, $F_{(2, 337)} = 239.154$, $p < 0.001$ (equation 2).

$$CoG = 7508 - 116\, roundedness - 122\, backness \tag{2}$$

The regression model for /s/ was significant also in the 135-ms slice ($R^2 = 0.600$, Adjusted $R^2 = 0.596$, $F_{(3, 336)} = 167.808$, $p < 0.001$). Roundedness ($p < 0.001$) and backness ($p < 0.001$) predicted CoG in this slice while height ($p = 0.773$) did not.

Regression model for /f/

For the 15-ms slice of /f/, the regression model was significant ($R^2 = 0.036$, adjusted $R^2 = 0.028$, $F_{(3, 336)} = 4.231$, $p = 0.006$). Results showed that roundedness of the vowel significantly predicted CoG in the initial fricative ($p = 0.001$), but backness ($p = 0.110$) and height ($p = 0.255$) did not. The regression model for the 35-ms slice was also significant ($R^2 = 0.030$ adjusted $R^2 = 0.022$, $F_{(3, 336)} = 3.489$, $p = 0.016$). Again, roundedness ($p = 0.006$) was a significant predictor, whereas backness ($p = 0.095$) and height ($p = 0.175$) were not. No significant regressions were found for the 75-ms slice ($R^2 = 0.020$, adjusted $R^2 = 0.011$, $F_{(3, 336)} = 2.307$, $p = 0.076$) or the 135-ms slice ($R^2 = 0.019$, adjusted $R^2 = 0.010$, $F_{(3, 336)} = 2.139$ $p = 0.095$). To obtain a regression equation with only significant terms, we re-ran the regression model for the first slice with only roundedness ($p < 0.001$) as a predictor, $F_{(1, 338)} = 8.765$, ($p = 0.003$) (equation 3). Examples of the /f/ words are shown in Figures 3 and 4.

$$CoG = 6722 - 808\, roundedness \tag{3}$$

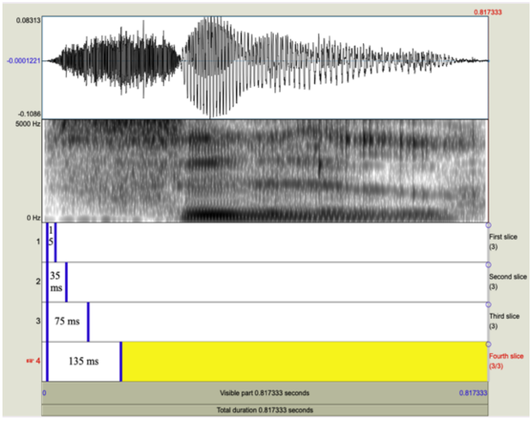

Figure 1. Sound waveform, spectrogram, and slices of [siːl]

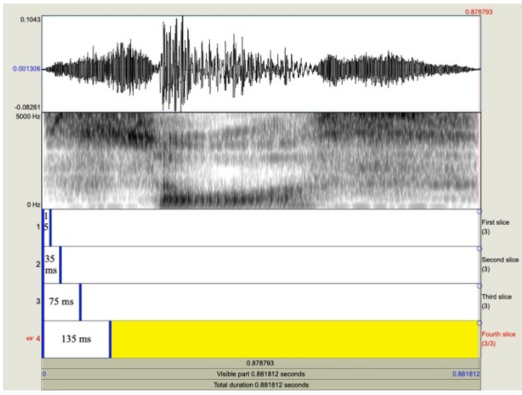

Figure 2. Sound waveform, spectrogram, and slices of [soːs]

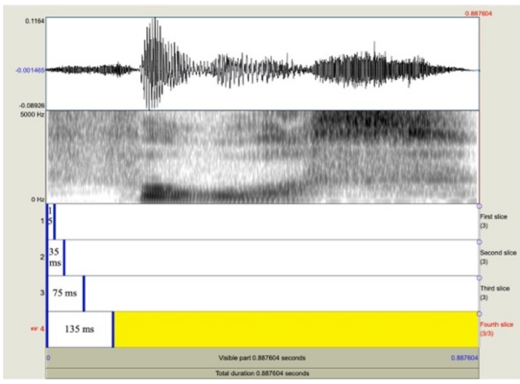

Figure 3. Sound waveform, spectrogram, and slices of [foːs]

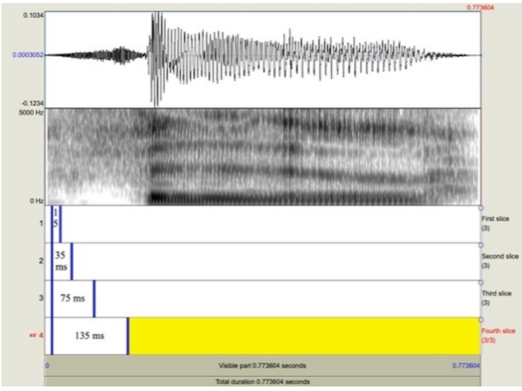

Figure 4. Sound waveform, spectrogram and slices of [fi:l]

Discussion

Results showed that coarticulation provides acoustic cues to the features of the upcoming vowel as early as during the first 15 ms after word onset. The acoustic feature of the upcoming vowel that was found to be most strongly reflected in the center of gravity measure of the initial fricative was roundedness. The regression equations show that the backness, height, and roundedness of the vowel reduced the average frequency of the spectral signal in the onset fricative. A future hypothesis is that the reduction in this frequency might be useful for perception.

Roundedness was the main predictor in the first two slices of /f/, while for /s/, roundedness was a significant predictor across all four slices. Furthermore, the roundedness effect of /s/ was larger compared to that of /f/. The reason might be that in the labiodental articulation of /f/, the lip movement at the beginning of the consonant could give rise to acoustic features which are similar to those associated with roundedness. For /s/, backness predicted CoG in the last two slices (75 ms and 135 ms). It thus seems that backness becomes more detectable as the consonant gets closer to the vowel. Vowel height predicted CoG in the first two slices (15 ms and 35 ms), and could thus be a more important cue for the vowel at the beginning of the consonant.

Perception studies have shown that fine-grained phonetic cues are used in lexical selection (McMurray et al., 2008; McQueen et al., 1999). Eye tracking has indicated that it is possible to use cues from fricatives /s/ and /ʃ/ to predict the upcoming phoneme, mainly by using roundedness as a cue. Fricatives carry slightly stronger cues than other consonants since the place of articulation provides clear traces in the frequency spectrum and by the formant transitions in the vocoid (Schreiber & McMurray, 2019). In the present study, the roundedness of the upcoming vowel predicted the CoG of the word onset as early as 15 ms, both for /f/ and /s/.

Conclusions

Center of gravity measurements show a weighted spectral average. This study used CoG as a measure to trace the acoustic correlates of the articulation of an upcoming vowel during a word onset consonant. The results showed that CoG values reflected information derived from the upcoming vowel, mainly in terms of roundedness, but for /s/, also height and – at a later stage – backness. There were generally stronger acoustic cues to the upcoming vowel in /s/ than in /f/. Future behavioral studies will investigate the perceptual availability of these early coarticulatory effects.

Acknowledgements

This study was supported by the Swedish Research Council (Grants No. 2018.00632 and 2019-03063), Knut and Alice Wallenberg Foundation (Grant No. 2018.0454), Crafoord Foundation (Grant No. 2017.0006) and Marcus and Alice Wallenberg Foundation (Grant No. 2018.0021).

References

Ardila, A., & Bernal, B. (2016). From Hearing Sounds to Recognizing Phonemes: Primary Auditory Cortex is A Truly Perceptual Language Area. *AIMS Neuroscience*, *3*, 454-473. doi: 10.3934/Neuroscience.2016.4.454

Boersma, P., Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.2, retrieved 7 April 2021 from http://www.praat.org/

Chechik, G., & Nelken, I. (2012). Auditory abstraction from spectro-temporal features to coding auditory entities. *Proceedings of the National Academy of Sciences*, *109*(46), 18968-18973. doi:10.1073/pnas.1111242109

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current biology*, *22*(7), 615–621. doi: 10.1016/j.cub.2012.02.015

Hardcastle, W. J., & Hewlett, N. (1999). *Coarticulation : theory, data, and techniques*. Cambridge University Press.

IBM Corp. Released (2021). *IBM SPSS Statistics for Macintosh, Version 28.0*. Armonk, NY: IBM

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychon Bull Rev*, *15*(6), 1064-1071. doi: 10.3758/pbr.15.6.1064

McMurray, B., Galle, M., Farris-Trimble, A., & Seedorff, M. (2018). What the /f/? We're not done with fricatives yet. Integrating across time and frequency bands. *The Journal of the Acoustical Society of America*, *144*, 1716-1716. doi: 10.1121/1.5067610

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25(5)*, 1363–1389. doi:10.1037/0096 1523.25.5.1363

Pützer, M., Moringlane, J. R., Sikos, L., Reith, W., & Krick, C. M. (2019). fMRI and acoustic analyses reveal neural correlates of gestural complexity and articulatory effort within bilateral inferior frontal gyrus during speech production. *Neuropsychologia*, *132*, 107129. doi: 10.1016/j.neuropsychologia.2019.107129

Riad, T. (2014). *The Phonology of Swedish*. OUP Oxford.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of memory and language*, *71*(1), 145-163. doi: 10.1016/j.jml.2013.11.002

Schreiber, K. E., & McMurray, B. (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics*, *81*(4), 1147-1166. doi: 10.3758/s13414-019-01712-9

Themistocleous, C. (2017). Effects of Two Linguistically Proximal Varieties on the Spectral and Coarticulatory Properties of Fricatives: Evidence from Athenian Greek and Cypriot Greek [Original Research]. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.01945

van Son, R., & Pols, L. (1996). *An Acoustic Profile Of Consonant Reduction*. doi: 10.1109/ICSLP.1996.607908

Zygis, M., Pape, D., Jesus, L., & Jaskuła, M. (2015). The effects of intonation on acoustic properties of fricatives, *Proceedings of the 18th ICPhS (International Congress of Phonetic Sciences), Glasgow, UK, August 2015*