INVESTIGATING OVERLAPPING GESTURES IN ACOUSTIC SIGNALS: THE CASE OF FRICATIVES AND SONORANTS IN CLUSTERS.

Jérémy Genette

Universiteit Antwerpen
jeremy.genette@uantwerpen.be

ABSTRACT

It is investigated whether the difference between sonorants and fricatives in terms of overlapping articulatory gestures can be tracked in the acoustic signal. To this aim, a new method using spectral comparisons is introduced to make possible the consistent analysis of speech sounds characterised by markedly different acoustic features. The efficiency of this method is tested by evaluating overlapping articulatory gestures in French initial CCV syllables with either a fricative or a sonorant as C2 (N=1904). The stimuli were produced by 20 native speakers of Belgian French who participated in a reading task. The main findings confirm the results of previous articulatory studies, indicating that sonorants exhibit longer articulatory overlap with the following vowels than fricatives. By obtaining consistent results, this paper validates the suggested method for evaluating the dynamics of overlapping articulatory gestures.

Keywords: methodology, coarticulation, acoustics, epenthetic vowels, sound change.

1. INTRODUCTION

Phonetic explanations for recurrent sound change patterns have triggered much interest throughout the history of linguistics [1, 2]. According to Ohala [3, 4, 5], sound changes are the result of phonetic synchronic variations that are phonologized over time. In this framework, physical constraints on speech production and perception may favour the misinterpretation of the acoustic signal by listeners who do not compensate for coarticulation and consequently do not retrieve the speaker's intended phonological target. Therefore, such confusions can trigger the emergence of a sound change by incorporating the fortuitous – even if somehow constrained – phonetic variation into the phonology of the language. Articulatory gestures which overlap to different extents are typical examples of physical constraints that potentially lead listeners to reinterpret the signal due to coarticulation.

Research on such overlapping articulatory gestures has been implemented by analysing directly articulatory data [6] or by observing how overlapping gestures can affect perception [7]. If differences in overlap timing can be observed in speech production and if they are consequently perceived by listeners, it is assumed that cues for such overlapping gestures can be found in the acoustic signal. Such coarticulatory effects have been acoustically studied via dynamic analyses of formants for vowels (e.g. [8]) and sonorants (e.g. [9]), of the centre of gravity for fricatives (e.g. [10]), etc. However, to the best of our knowledge, there is no technique that can be consistently applied to speech sounds characterised by markedly different acoustic features, such as sonorants and fricatives.

The main objective of this paper is to introduce a general acoustic method for assessing the amount of coarticulatory influence between flanking phones, by using a Spectral Similarity Index. The efficiency of the method will be tested by observing whether it grasps the expected differences in overlapping articulatory gestures in CCV syllables whose C2 is either a sonorant or a fricative.

2. OVERLAPPING GESTURES IN CCV

One of the most recurrent sound change patterns is the simplification of complex syllabic structures into CV syllables. As for CCV syllables, the simplification can consist of either the deletion of one of the two consonantal segments (i.e., CCV > CV) or the insertion of a vowel between the consonants (i.e., CCV > CVCV).

In *Articulatory Phonology* (cfr. [11, 12]), the insertion of epenthetic vowels can be attributed to a mistiming of the different articulatory gestures. More specifically, an epenthetic vowel can result from a reduced articulatory overlap between C1 and C2 or from the articulatory gesture of the V being anticipated up to the C1-C2 transition [13]. The former account suggests the presence of a short gap between the articulatory implementation of C1 and C2, leaving the vocal tract open. The latter account suggests that the articulatory gesture for the vowel might be anticipated as early as the C1-C2 transition. Both gestural timings are thought to

favour the reinterpretation of CCV as CVCV by the listener. The gestural scores of both accounts are shown in Fig. 1.

It has long been noted, however, that some phonetic contexts favour vowel epenthesis more than others. For example, it is commonly assumed that the insertion of epenthetic vowels between consonants is favoured by plosive + sonorant rather than by plosive + fricative clusters [7]. This can reasonably be attributed to the precise articulatory configuration in terms of constriction width and pressure required for fricatives [6], while sonorants are characterized by a reduced build-up of oral pressure [14]. In Ohala's [3, 4, 5] terms, the specific articulatory configuration needed for the production of fricatives, as opposed to that of sonorants, could be the physical constraint that make the stop + fricative clusters less subject to reinterpretation than stop + sonorants clusters, hence less subject to sound change.

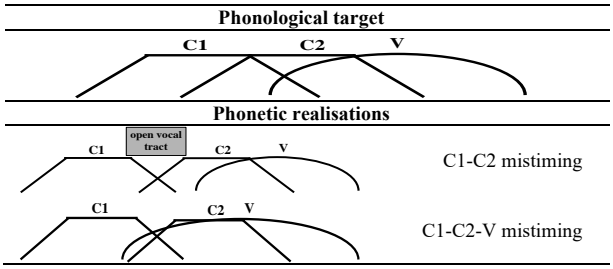


Figure 1: Gestural scores of the phonological target of CCV syllables and potential phonetic realisations (adapted from [13]).

In other words, the longer gestural overlap between a sonorant and the following V causes the listener to reinterpret a CCV syllable more easily as CVCV. Both articulatory and perception studies previously confirmed the propensity of plosive + sonorant clusters to favour the insertion of epenthetic vowels [6, 7]. On the one hand, the current evidence indicates that the vocalic articulatory overlap is longer in plosive + sonorant than in plosive + fricative clusters [6]. On the other hand, perceptual experiments confirm that CCV syllables are more often reinterpreted as CVCV if C2 is a sonorant [7]. Given the consistency of articulatory and perceptual evidence, it can reasonably be assumed that such differences in articulatory timing are reflected in the acoustic signal. The acoustic lens sheds light both on what is produced and perceived and enables a better integration of speech production and perception in the study of sound change.

In the following, a method is presented to study overlapping gestures acoustically. It is then observed whether it detects the difference in articulatory timing between sonorants and fricatives.

3. EXPERIMENT

To assess the extent of articulatory overlap between a speech sound (the target phone) and its flanking phones, its spectrum is compared to a theoretically non-coarticulated spectrum. The non-coarticulated spectrum is computed by instructing speakers to produce several repetitions of words containing the target phone with the preceding and following phones chosen to maximise the range of articulatory dimensions that might modify its spectral composition through coarticulation. From those stimuli, the spectrally stable portion of the target-phone is manually extracted, and its Long-Term Average Spectrum (LTAS) is computed. The gathered LTAS of all repetitions are averaged per-target phone and per-speaker to obtain a per-phone and per-speaker reference LTAS.

Then, the speech sounds to be analysed are divided into time-normalized frames, and the LTAS of each frame is computed. The spectral comparison is then performed between the LTAS of the extracted frames and the per-phone and per-speaker reference LTAS via a Spectral Similarity Index (SSI). By doing so, the amount of similarity between both spectra can be evaluated throughout the duration of the sound.

In this study, speech samples were collected from native French speakers via a self-paced reading task to investigate the dynamics of articulatory overlap between sonorants and fricatives in CCV syllables.

3.1. Participants and procedure

Data were collected from 20 native speakers of Standard Belgian French (11 women and 9 men, age: $M = 47.6$ years, $SD = 18.9$). They participated in a randomized self-paced reading task of French (pseudo-)words. All stimuli were presented visually using a *PsychoPy* experiment [15] on a computer screen placed at a comfortable reading level for each participant. The productions were recorded on a TONOR TC 30 in a quiet room. Before the start of the task, subjects were invited to read practice stimuli. The stimuli comprised both (pseudo-)words with initial CCV syllables containing the test stimuli and the reference stimuli used to elicit the reference LTAS.

3.2. Materials

3.2.1. Test stimuli

The materials for this study consisted of (pseudo-) words with a CCV syllable in initial position. The C1 of the CCV syllable was one of the plosives /p,

b, t, d, k, g/. The C2 was either one of the sonorants /ʀ, l, m/ or one of the fricatives /s, z/. The V was one of the three cardinal vowel /i, a, u/. In total, the stimuli set amounts to 90 pseudo-words (6 C1 * 5 C2 * 3 V) and 40 existing French words. The selected existing words were the most frequent lexical items with a given C1-C2-V combination, according to the *Lexique 3.83* database [16]. Some stimuli were excluded due to technical issues and omissions in the responses of the participants, leaving 1904 tokens.

3.2.2. Reference stimuli

The selected sonorants and fricatives were produced in 4 repetitions of 15 words in which they were followed by one of the 3 cardinal vowels /i, a, u/. For instance, /z/ was produced in 4 repetitions of the words *hasard* /azaʀ/ "chance", *visite* /vizit/ "tour" and *bouzouki* /buzuki/ "bouzouki". The reference stimuli were randomly inserted within the test stimuli during the reading task.

3.3. Acoustic analysis

3.3.1. Test stimuli

The C2 of the CCV syllables were segmented manually after the release burst of the C1 and before the beginning of the vowel. Then, each C2 was divided into 10 time-normalized frames. For each frame, the LTAS was computed using a frequency band of 100 Hz across the sampling frequency (48 kHz) via PRAAT [17] and the *PraatR* package [18].

3.3.2. Reference stimuli

The spectrally stable portion of the sonorants and fricatives – i.e., the portion of the sound that is least affected by coarticulation – in the reference stimuli were segmented manually through visual inspection. The LTAS was then computed for each segment, using a frequency band of 100 Hz throughout the sampling frequency (48 kHz) via PRAAT [17]. The LTAS of each C2 were averaged to obtain a per-speaker and per-C2 reference LTAS.

3.4. Spectral Similarity Index

By considering each LTAS as a vector of n dimensions, the 10 frames of the target-C2 were compared with the reference LTAS of the corresponding C2 via the Spectral Similarity Index (SSI) as in Eq. 1, where x stands for the reference spectrum, y is the test spectrum, i is the frequency bin and n is the number of frequency bins.

The resulting SSI values indicate the similarity between the reference C2 and each frame of the test stimuli. A high SSI value indicates a high similarity between both LTAS.

$$(1)\ SSI = \frac{\sum_{i=1}^{n}(x_i-\bar{x}).(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}.\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

3.5. Statistical analysis

Multi-level modeling was used as the statistical technique. Models of increasing complexity were built step by step by including random and fixed effects one after the other [19, 20]. The statistical analysis was carried out in R [21] using the *lme4* package [19] and the *lmerTest* package [22] to obtain *p*-values. The final model included time (i.e., the frame number), a quadratic effect of time, C2 type (sonorant vs. fricative), lexicality (word vs. pseudo-word), and their interactions, as well as subjects and items nested within subjects as random effects.

4. RESULTS

Table 1 shows the results of the final model, and a graphical representation of the SSI dynamics throughout C2 is presented in Fig. 2. In Table 1, only the main fixed effects are presented, as well as the interaction effects that are significant. The results show that SSI increases with time, as indicated by a significant effect of time (p<.001). The effect of time is not linear, as shown by the significant quadratic effect of time (p<.001). The main effect of C2 type (sonorant vs fricative) is also significant (p<.001), but the nonlinear effect of time is different depending on the type of C2, as shown by the significant interaction between both factors (p<.001) and by the significant three-way interaction with time (p<.001). The main effect of lexicality is, however, not significant (p=.68).

	Estimate	SE	*t*-value	*p*-value
Intercept	809.30	0.005	149.94	<.001***
Time	0.18	0.03	6.02	<.001***
Quadratic time	0.24	0.06	3.89	<.001***
C2 type [son.]	0.06	0.006	10.28	<.001***
Lexicality [word]	0.004	0.009	0.42	0.68
C2 type [son.] * Quadratic time	-3.69	0.09	-4.24	<.001***
Time * Quadratic time	-0.42	0.03	-11.18	<.001***
C2 type [son.] * Time * Quadratic time	0.45	0.005	8.55	<.001***

Table 1: Fixed main effects and significant interaction effects on SSI; C2 type [fri.] & Lexicality [pseudo-word] = ref. category.

As can be seen in Fig. 2, the SSI dynamics of fricatives exhibit a bell-shaped trajectory, while sonorants show a shallow SSI increase at the beginning followed by a relatively flat SSI trajectory. The sharp and short decreases in SSI at the end of fricatives indicate that the coarticulary influence of the vowel begins late and does not reach the first half of the fricative. The steep

increase at the beginning can be attributed to the coarticulatory influence of C1. In contrast, in sonorants, the maximum level of dissimilarity due to the vowel is quite stable over the last two-thirds of C2 and gently decreases towards its beginning. Yet, after visual inspection of the spectrograms, the low SSI at the beginning of sonorants, must be attributed to a period of relative silence between the release of the C1 occlusion and the occlusion needed to produce /m/ (see Fig. 3.a.), rather than to coarticulatory transition patterns as in /ʀ/ and /l/ (see Fig. 3.b.). Nevertheless, the more stable SSI curve in sonorants indicates that the effect of the vowel, which is expected to be maximal at the C2-V boundary, remains constant for a longer period of time in sonorants than fricatives.

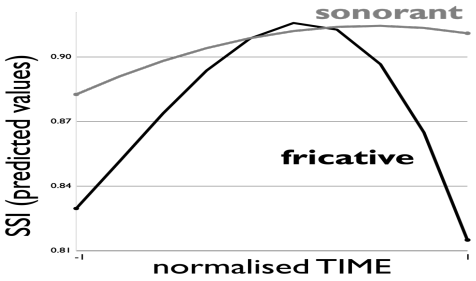
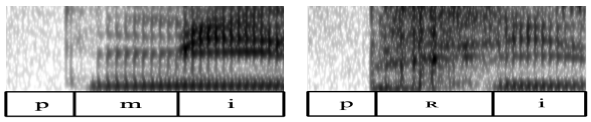


Figure 2: SSI dynamics (predicted values) as a function of C2 type (fricative vs sonorant).



(a) (b)

Figure 3: Example spectrograms of CCV syllables with /m/ (a) and /ʀ/ (b) as C2.

5. DISCUSSION

This study aimed at observing whether differences in the timing of coarticulatory gestures in CCV syllables can be observed in the acoustic signal. For this purpose, the SSI metrics was designed to assess the extent to which a given frame of C2 is spectrally different from a theoretically non-coarticulated spectrum. The C2 of 1904 clusters were analysed with the suggested method. The main finding is that sonorants tend to overlap more with the successive vowel than fricatives, which is consistent with previous articulatory and perceptual studies. In fact, the results show a clear difference in behaviour between fricatives and sonorants.

The evolution of the SSI in fricatives shows a bell-shaped curve, indicating that the coarticulatory influence of the vowel covers the second part of the C2, but not earlier. This suggests that the articulatory gestures of the vowel do not extend to the beginning of the C2. Therefore, it is assumed that the phonetic realization of the CCV syllable is relatively close to the phonological target, as shown in Fig. 1. This makes it unlikely that the signal of the CCV syllable is misinterpreted by the listener as CVCV. The large decrease in gestural overlap could be attributed to the precise articulatory configuration in terms of airflow and constriction width necessary to produce frication noise [6]. In contrast, the SSI trajectory of the sonorants is much more stable and follows a rising trend. It potentially indicates that the coarticulatory influence of V covers a larger portion of C2 and merges with the influence of C1 at the beginning of C2. The observed major articulatory overlap between sonorants and vowels, attributed to the less constrained articulatory specifications of sonorants, might lead listeners to reinterpret the signal as CVCV, rather than as a CCV syllable.

It must be noted tough that the absolute SSI values do not provide information about the amount of the coarticulatory influence *per se* because the SSI heavily depends on the spectral variability between reference stimuli, hence the average high SSI values observed with sonorants. However, the shape of the trajectories provides insights into how the preceding and following speech sounds affect the SSI for each C2 type. An analysis considering the different types of sonorants as well as the effect of potential allophonic realisations of French /ʀ/ would provide more detailed insights, but this is beyond the scope of the present paper.

In short, this technique could provide an alternative to more expensive and invasive articulatory methods, albeit with limitations. By complementing articulatory and perceptual studies, acoustic analyses such as this one could also lead to a deeper understanding of sound change processes.

6. CONCLUSIONS

The present study aimed at observing whether differences in the timing of coarticulatory gestures in CCV syllables can be tracked in the acoustic signal. A new method is suggested by computing a per-speaker and per-C2 reference LTAS with which to compare the LTAS of the test C2.

The results show that the coarticulatory influence of V on C2 was longer for sonorants than fricatives, confirming the results of previous studies. The results indicate the feasibility of tracking overlapping articulatory gestures in the acoustic signal. The suggested technique cannot identify specific articulatory gestures, but it permits the assessment of coarticulatory dynamics in speech sounds differing in acoustic features, such as sonorants and fricatives.

7. ACKNOWLEDGEMENTS

Thanks are due to the participants of this study. The research reported in this paper was partly supported by a research grant from the Research Foundation – Flanders (FWO) [grant G004321N].

8. REFERENCES

[1] Pinget, A.-F. 2015. *The actuation of sound change.* LOT.

[2] Garrett, A., Johnson, K. 2013. Phonetic bias in sound change. In: Yu, A. C. L. (ed.), *Origins of sound change: Approaches to phonologization*. Oxford University Press, 51–97.

[3] Ohala, J. J. 1989. Discussion of Lindblom's 'Phonetic invariance and the adaptive nature of speech'. In: Elsendoorn, B. A. G., Bouma, H. (eds.), *Working models of human perception*. Academic Press, 175–183.

[4] Ohala, J. J. 1989. Sound change is drawn from a pool of synchronic variation. In: Breivik, L. E., Jahr, E. H., (eds.), *Language change: Contributions to the study of its causes*. Mouton de Gruyter, 173–198.

[5] Ohala, J. J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13, 155–161.

[6] Recasens, D. 2018. *The production of consonant clusters: Implications for phonology and sound change*. De Gruyter Mouton.

[7] Fleischhacker, H. A. 2005. *Similarity in phonology: Evidence from reduplication and loan adaptation*. University of California.

[8] Krull, D. 1989. Consonant-vowel coarticulation in spontaneous speech and in reference words. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1(5), 101–105.

[9] Themistocleous, C., Fyndanis, V., Tsapkini, K. 2022. Sonorant spectra and coarticulation distinguish speakers with different dialects. *Speech Communication* 142, 1–14.

[10] Lulaci, T., Tronnier, M., Söderström, P., Roll, M. 2022. The time course of onset CV coarticulation. *Fonetik 2022 - the XXXIIIrd Swedish Phonetics Conference*.

[11] Browman, C. P., Goldstein, L. M. 1986. Towards an articulatory phonology. *Phonology* 3, 219–252.

[12] Browman, C. P., Goldstein, L. M. 1992. Articulatory phonology: An overview. *Phonetica* 49, 155–180.

[13] Buchwald, A. B., Rapp, B., Stone, M. 2007. Insertion of discrete phonological units: An articulatory and acoustic investigation of aphasic speech. *Language and Cognitive Processes* 22(6), 910–948.

[14] Ladefoged, P. 1997. Linguistic phonetic descriptions. In: Hardcastle, W. J., Laver, J. (eds.), *The handbook of phonetic sciences*. Wiley, 589–618.

[15] Peirce, J.W. 2007. PsychoPy–psychophysics software in Python. *Journal of neuroscience methods* 162(1-2), 8–13.

[16] New, B., Pallier, C. 2020. Lexique 3.83. http://www.lexique.org

[17] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot. International* 5(9), 341–345.

[18] Albin, A. 2014. PraatR: An architecture for controlling the phonetics software "Praat" with the R programming language. *Journal of the Acoustical Society of America* 135(4), 2198.

[19] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1).

[20] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4), 390–412.

[21] R Development Core Team 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[22] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2015. Package 'lmertest'. *R Package Version* 2(0), 734.