SPEAKER VARIATION IN DUTCH /x/ AND /s/ AS A FUNCTION OF SYLLABIC POSITION AND LIP-ROUNDING

Laura Smorenburg & Willemijn Heeren

Leiden University Centre for Linguistics
B.J.L.Smorenburg@hum.leidenuniv.nl; W.F.L.Heeren@hum.leidenuniv.nl

ABSTRACT

This study aimed to investigate the interaction between intra- and inter-speaker variation, i.e. speaker-specificity, and linguistic information in fricatives. Previous research has shown that linguistic information such as syllabic position and contextual lip-rounding may cause variation in fricative acoustics. Acoustic measures from Dutch fricatives /x/ and /s/ were extracted from spontaneous telephone speech for 57 male speakers. The speaker-specificity of these measures was examined as a function of syllabic position and anticipatory and perseverative lip-rounding.

Linear mixed-effect models showed no, or small, effects of syllabic position on spectral, amplitudinal, and temporal measures and showed effects of contextual lip-rounding predominantly for /x/. Linear discriminant analysis showed that fricative acoustics contain speaker-specific information. Syllabic positions differed somewhat in degree of speaker-specificity.

These results show that, in fricatives, speaker variation is slightly affected by linguistic information.

Keywords: speaker-specificity, fricatives, syllabic position, contextual lip-rounding, speech production

1. INTRODUCTION

Although it has been shown that both speaker-dependent and linguistic information cause variation in speech sounds, it is not clear if and how these interact. This is relevant from both a theoretical perspective to study the contribution of the speaker in fricative productions, as well as a practical perspective (with implementations in forensic voice comparisons) to find locations with more speaker-dependent information. The present study investigated speaker variation as a function of linguistic information.

Linguistic information has been shown to affect speech segment acoustics. Namely, it has been shown that there are articulatory strong and weak locations in speech. For example, the edges of prosodic domains such as phrases and words are generally found to be locations of articulatory strengthening [6, 8, 9]. Another example, and a main focus of the present study, is coda consonant reduction, which poses that codas are articulatory weak locations compared to onsets [15]. For fricatives, aerodynamic and acoustic data on American English /f, v, s, z, ʃ, ʒ/ seems to support this; codas (defined as prepausal and preconsonantal) were found to have a slower pressure build-up, a lower pressure peak, a delayed onset of audible frication, and a lower amplitude than onset fricatives [22]. However, when looking at American English /s/ specifically, [19] show that, while /s/ durations are shorter in coda than in onset position, amplitude and spectral centre of gravity do not show reduction. Interestingly, when a discriminant analysis was performed on the acoustic data, consonant classification performance was better for onsets than codas for all consonants except /s/. Similar results are reported for German fricatives /s, ʃ/; slightly higher spectral centre of gravities were found for codas than onsets [7]. The authors also reported higher variability for codas than onsets and more variability for deaccented than accented positions.

Together, these findings indicate that not all fricatives reduce in the same manner or to the same extent, which might be explained by studies showing that consonant reduction seems to be constrained by production requirements [17]. As a result, some consonantal features are more resistant to coarticulation and reduction than others. For example, the well-reported effect of lip-rounding on fricatives [4, 12, 13, 14, 20, 21] may be explained by the fact that the lips are often not actively engaged in fricative production, so the coarticulatory resistance to lip-rounding in fricatives is very low. The tongue blade and dorsum are more resistant to coarticulation and reduction in fricatives because of the production necessity of constrictions formed with the blade and dorsum in fricatives [18].

Speaker-related variation has also been shown to affect speech segment acoustics, resulting in speaker-specificity of speech sounds. Studies show that some segments are more speaker-specific than others. For example, in Dutch, fricative /s/ was ranked below vowels and nasals, but above /r/ and plosives in terms of speaker-specificity [23]. Fricative /s/ has also been shown to contain speaker-

specific information in both English [10] and Dutch [24] read speech. To our knowledge, no reports of speaker variability in acoustic data for Dutch /x/ exist in the literature. Fricatives /s/ and /x/ were selected for the present study because they are highly frequent fricatives in Dutch [11].

In sum, although it has been shown that some locations in speech are susceptible to articulatory strengthening or weakening, these effects are not uniform, as some consonantal features are more resistant to reduction and coarticulation. Moreover, it is not clear how these articulatory strong or weak locations interact with speaker-specificity. Additionally, most studies have examined these effects in read speech, which is not representative of speech material used in forensic voice comparisons. The present study investigated if speaker variation is affected by syllabic position and contextual lip-rounding in spontaneous telephone speech, and if so, which acoustic measures in which contexts are relatively speaker-specific. Based on previous findings [4, 12, 13, 20, 21], we hypothesised that spectral measures show significant effects of contextual lip-rounding. In light of conflicting findings on coda reduction in fricatives (particularly /s/) [7, 19, 22], we had no strong prediction for the effect of syllabic position on spectral, amplitudinal, and temporal measures.

2. METHODOLOGY

2.1. Corpus data

A total number of 3,492 /x/ tokens and 3,073 /s/ tokens from 57 male speakers of Standard Dutch aged 18-50 were automatically segmented and manually validated from spontaneous telephone speech available in the Spoken Dutch Corpus [16]. Word-initial onsets and word-final codas were automatically coded based on lexical form, but codas followed by vowels were recoded as ambi-syllabic (/x/: N = 378, /s/: N = 412) and excluded from the present analysis. Additionally, speakers with fewer than 25 tokens were excluded. This resulted in 3,067 /x/ tokens and 2,661 /s/ tokens. Adjacent segments to the left and right of each fricative were coded as rounded or non-rounded. Vowels, /u, ɔ, o, ø, y, ʏ/, diphthongs /œy, ɑu/, and bilabial consonants /p, b, m/ were considered to be rounded.

2.2. Acoustic measures

For each fricative token, the duration, spectral centre of gravity (CoG), spectral standard deviation (SD), and spectral tilt (Praat's spectral tilt function with robust fit method on a logarithmic frequency scale) were taken over the middle 50% of each fricative's duration, over a 0.5–4.0 kHz band in Praat [5]. Additionally, polynomial cubic fits derived from the spectral mean over five non-overlapping 7-ms windows that were evenly spaced over the fricative's full duration were computed (see [12]). This dynamic measure was computed for tokens with durations of minimally 35 ms.

2.3. Analysis

2.3.1. Linear mixed-effect modelling

Linear mixed-effect models with random intercepts for Word and Speaker, random slopes for Speaker, and fixed predictors for Left Context (0 = non-rounded, 1 = rounded), Right Context (0 = non-rounded, 1 = rounded), and Syllabic Position (0 = coda, 1 = onset) were run separately per measure for /x/ and /s/. Models were fitted using function `lmer` from R package `lme4` [3]. The initial step was to build a full model with a maximal random structure by restricted maximum likelihood (REML) estimation [1]. Next, stepwise deletion of random structure was used to reduce the random structure of the model, given this was theoretically justifiable [2]. Random-effect correlations were excluded. In a last step, the fixed factors were estimated by stepwise deletion. Models were compared using the likelihood ratio test. Speaker variation was inspected using caterpillar plots, which visualised random intercept and slope coefficients by speaker.

2.3.2. Linear discriminant analysis

The data set was not balanced enough across contextual lip-rounding conditions to run separate linear discriminant analyses for rounded versus unrounded context conditions. Therefore, separate linear discriminant analyses were run for onset (/x/: N = 1,580, /s/: N = 1,435) and coda tokens (/x/: N = 1,436, /s/: N = 1,225) to determine contributions of individual acoustic measures to speaker-classification performance per syllabic position for /x/ and /s/. Outliers, defined as being more than three standard deviations removed from the mean, were excluded. The highest correlating measures (all r > .55) were excluded, which excluded the cubic intercept from the dynamic CoG measure. Per syllabic position, speakers with fewer than 10 tokens were excluded, for /x/ resulting in 1,491 onset and 1,376 coda tokens from 50 and 48 speakers respectively and for /s/ resulting in 1,186 onset and 1,375 coda tokens from 48 and 50 speakers respectively. Only the first discriminant functions that, together, explained at least 75% of the variance were considered.

Table 1. Fixed effects in best-fitting linear mixed-effect models for /x/ and for /s/

	/x/									/s/					
	CoG (Hz)			SD (Hz)			duration (ms)			CoG (Hz)			duration (ms)		
Effect	*β*	SE	t	*β*	SE	t	*β*	SE	t	*β*	SE	t	*β*	SE	t
(intercept)	1728	33	51.8	679	16	43.3	91	2	42.9	2698	40	70.0	105	2	44.7
Left Context	−153	33	−4.6	70	16	4.4				−59	26	−2.2			
Right Context	−229	33	−6.6	49	23	2.2	−21	4	−5.5				−26	4	−7.1
Syll. Position				−9	8	−1.1	−2	3	−0.9	49	17	2.8	−19	3	−7.4
Left × Right	110	49	2.3												
Left × Syll.P				−73	18	−4.1									
Right × Syll.P				−108	21	−5.1	35	5	6.7				29	5	5.8

Note. Blank cells indicate that these predictors were not included in the best-fitting linear mixed-effect model.

3. RESULTS

3.1. Linear mixed-effect models

3.3.1. Fixed effects

For both /x/ and /s/, the estimates, standard errors, and t-values for all fixed effects in the best-fitting models are displayed in Table 1. Not displayed are best-fitting models that contained an intercept only and thus showed no significant effects, i.e. models for /x/ spectral tilt (–28 Hz, SE = 1 Hz), /s/ spectral tilt (–25 Hz, SE = 1 Hz), and /s/ spectral SD (626 Hz, SE = 19 Hz).

As can be seen in Table 1, for /x/, CoG shows a decrease when Left Context (–153 Hz) or Right Context (–229 Hz) are rounded. However, when both Left and Right context are rounded, these lowering effects are attenuated (110 Hz). For /s/, Left Context rounding decreased CoG (–59 Hz) and onsets had higher CoGs (49 Hz) than codas.

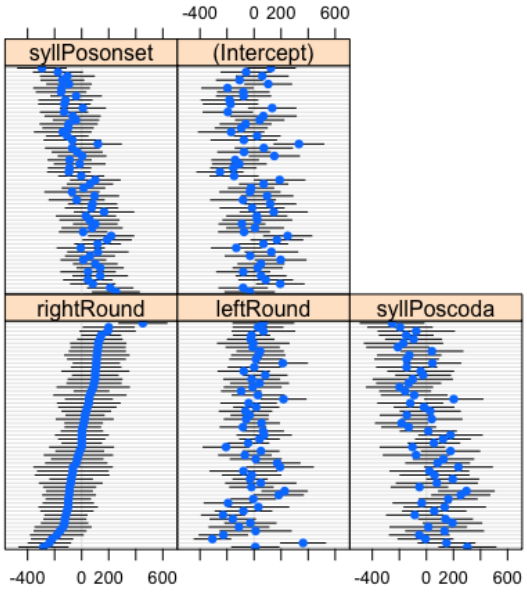
Whereas there were no significant effects for /s/ SD, /x/ SD shows an increase when Left Context (70 Hz) and Right Context (49 Hz) are rounded. However, the significant interactions between Left Context and Right Context with Syllabic Position indicate that the effects of Left and Right Context rounding are different for onsets and codas; contextual rounding increases SD only in codas.

For duration, /x/ shows a decrease when Right Context is rounded (–21 ms). The interaction indicates that the effect of Right Context rounding is different for onsets and codas; onset duration increases and coda duration decreases as a function of Right Context rounding. For /s/, Right Context rounding also decreases duration (–26 ms) and onsets are shorter (–19 ms) than codas. The interaction between Right Context and Syllabic Position indicates that the effect of Right Context rounding is different for onsets and codas; onset duration increases and coda duration decreases as a function of Right Context rounding.

3.3.2. Random effects

All random by-speaker intercept and slope coefficients show speaker variation, with some showing large differences between speakers. For example, random by-speaker slopes for the effect of Right Context on /x/ CoG, for which the fixed-effect intercept was −229 Hz, show a wide range (see Figure 1). This indicates that the effect of Right Context rounding is speaker-specific.

Figure 1: Caterpillar plots of random by-speaker intercepts and slopes of the /x/ CoG model



3.2. Linear discriminant analysis

Speaker classification performances per fricative and per Syllabic Position are displayed in Table 2.

Table 2: Cross-validated speaker classification performance (%) and chance-level (%) per fricative and per Syllabic Position

	onset		coda	
	class.	chance	class.	chance
/x/	12.5	2.0	15.7	2.1
/s/	18.5	2.1	15.4	2.0

After excluding measures that correlated highly with other measures (which excluded only the cubic intercept coefficient from our dynamic CoG measure), all linear discriminant models included the following set of acoustic-phonetic measures: spectral CoG, SD, and tilt, duration, and three dynamic CoG coefficients. Correlations between acoustic-phonetic measures and discriminant functions express measures' individual contributions to classification performance, i.e. speaker-specificity. Across discriminant models per fricative and per syllabic position, spectral tilt, followed by spectral SD and CoG were the best predictors.

4. DISCUSSION

The present study investigated speaker variation as a function of linguistic information in fricatives. To do so, effects of syllabic position and contextual lip-rounding were firstly examined with linear mixed-effect models.

Previous studies have demonstrated an effect of context lip-rounding on fricative spectra [4, 12, 14, 20, 21]. The present study, using spontaneous speech data, was only able to confirm this effect for /x/ and less clearly for /s/. For /x/, the effect of anticipatory lip-rounding was larger than that of perseverative lip-rounding, however an interaction also shows that when both left and right context are rounded, this lowering effect is attenuated.

Regarding the effect of syllabic position, previous studies have reported somewhat conflicting results. The present study finds some evidence that fricative onsets constitute stronger articulatory locations than codas. Spectral SD in /x/ increased when context is rounded only in codas. This indicates that, for /x/, codas seem less resistant to contextual lip-rounding than onsets. For /s/, spectral CoG in onsets was 49 Hz higher than in codas, which indicates that onsets are stronger articulatory locations than codas.

Although /s/ onset durations were shorter than coda durations, the interaction between right context rounding and syllabic position indicates that onsets in right rounded context were longer in duration than codas. Given that our coda tokens were sometimes phrase final, the lack of longer durations for onset /s/ and /x/ might be confounded with phrasal position. Future inclusion of a predictor variable for phrasal position may confirm this.

After confirming small effects of syllabic position on /x/ and /s/ spectra, the question remained whether these differences interacted with speaker-specificity. Linear discriminant analysis showed no substantial differences in speaker-classification performance per fricative and per syllabic position. For /x/, codas contained slightly more speaker-specific information, whereas for /s/, onsets contained slightly more speaker-specific information. Future research will look at possible confounds such as word stress and morphosyntactic status of the word the fricative occurs in to see whether these small differences remain.

Looking at the specific acoustic measures that contributed to speaker discrimination, a more consistent picture emerges for /x/ and /s/; spectral tilt, SD and CoG performed best, whereas duration and dynamic CoG coefficients performed worst. SD and CoG have before been shown to be relatively well-performing speaker discriminants in English /s/ [10]. Spectral tilt, which showed no effects of contextual lip-rounding or syllabic position for either /x/ or /s/, performed best as a discriminant for both fricatives.

Importantly, linear discriminant analysis shows that both /x/ and /s/ contain speaker-specific information, despite the limited frequency band of our data (0.5 – 4.0 kHz telephone speech). This has implications for forensic phonetics, where analysed speech material is often similar to the data set analysed here and the goal is to compare speaker measures across recordings. The differences in speaker classification across linguistic contexts, however, were so small that they can have no practical consequences for forensic speech comparisons at this time. Especially given the often very limited speech material in forensic casework, this a useful result.

5. CONCLUSION

The present study has found that Dutch fricatives /x/ and /s/ from spontaneous telephone speech contain speaker-specific information, confirming findings from non-spontaneous speech data. For both /x/ and /s/, spectral tilt, CoG, and SD were the most speaker-specific acoustic measures. Moreover, it seems that speaker-specificity interacts with linguistic information. However, differences in speaker classification between syllabic positions were very small and it is currently unclear whether the different speaker-classification performances per fricative and per syllabic position are solely due to differences in syllabic position. Future research will include possible confounds to test this.

6. ACKNOWLEDGEMENTS

This research was supported by a Netherlands Organization for Scientific Research VIDI grant (276-75-010).

7. REFERENCES

[1] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Memory and Language* 68(3), 255-278.

[2] Bates, D., Kliegl, R., Vasishth, S., Baayen, H. 2015. Parsimonious Mixed Models.

[3] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Statistical Software* 67, 1-48.

[4] Bell-Berti, F., Harris, K. S. 1979. Anticipatory coarticulation: Some implications from a study of lip rounding. *J. Acoust. Soc. Am.* 65(5), 1268–1270.

[5] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341–347.

[6] Cho, T., McQueen, J. M. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *J. Phonetics* 33(2), 121-157.

[7] Cunha, C., Reubold, U. 2015. The contribution of vowel coarticulation and prosodic weakening in initial and final fricatives to sound change. *Proc. 18th ICPhS* Glasgow, 27-31.

[8] Fougeron, C., Keating, P. A. 1998. Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* 101(6), 3728-3740.

[9] Fougeron, C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *J. Phonetics* 29(2), 109–135.

[10] Kavanagh, C. M. 2012. New consonantal acoustic parameters for forensic speaker comparison. Diss. University of York.

[11] Luyckx, K., Kloots, H., Coussé, E., Gillis, S. 2007. Klankfrequenties in het Nederlands. In: *Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems.* Academia Press, 141-154.

[12] Munson, B. 2001. A method for studying variability in fricatives using dynamic measures of spectral mean. *J. Acoust. Soc. Am.* 110(2), 1203–1206.

[13] Munson, B. 2004. Variability in /s/ production in children and adults. *J. Speech Lang. and Hearing Research* 47, 58-69.

[14] Nittrouer, S., Whalen, D. H. 1989. The perceptual effects of child–adult differences in fricative-vowel coarticulation. *J. Acoust. Soc. Am.* 86(4), 1266–1276.

[15] Ohala, J. J., Kawasaki, H. 1984. Prosodic phonology and phonetics. *Phonology* 1, 113–127.

[16] Oostdijk, N. H. J. 2000. Corpus Gesproken Nederlands. *Nederlandse Taalkunde* 5, 280–284.

[17] Recasens, D. 2004. The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). *J. Phonetics* 32(3), 435–453.

[18] Recasens, D., Dolorspallarè, M. 2001. Coarticulation, assimilation and blending in Catalan consonant clusters. *J. Phonetics* 29, 273–301.

[19] Redford, M. A., Diehl, R. L. 1999. The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *J. Acoust. Soc. Am.* 106(3), 1555-1565.

[20] Shadle, C. H. 1986. The acoustics of fricative consonants. *J. Acoust. Soc. Am.* 79(2), 574.

[21] Shadle, C. H., Scully, C. 1995. An articulatory-acoustic-aerodynamic analysisof [s] in VCV sequences. *J. Phonetics* 23, 53–66.

[22] Solé, M. 2003. Aerodynamic characteristics of onset and coda fricatives. *Proc. 15th ICPhS* Barcelona, 2761-2764.

[23] van den Heuvel, H., Rietveld, T. 1992. Speaker related variability in cepstral representations of Dutch speech segments. *Proc. ICSLP 92* Banff, 1581-1584.

[24] van den Heuvel, H. 1996. Speaker variability in acoustic properties of Dutch phoneme realisations. Diss. Radboud Universiteit.