PREPRINT VERSION of article published in JASA 10-02-2020, https://doi.org/10.1121/10.0000674

The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues

Laura Smorenburg[a)]

Willemijn Heeren

Leiden University Centre for Linguistics, Leiden University, Van Wijkplaats 4, 2311 BX Leiden, The Netherlands

Abbreviated title: The distribution of speaker information in fricatives

Smorenburg, L., and Heeren, W. The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. Journal of the Acoustical Society of America, 147(2), pp. 949-960. https://doi.org/10.1121/10.0000674

[a)]E-mail: b.j.l.smorenburg@hum.leidenuniv.nl

Abstract

Although previous work has shown that some speech sounds contain more speaker-dependent information than others, not much is known about the speaker information of the same segment in different linguistic contexts. The present study therefore investigated whether Dutch fricatives /s/ and /x/ from telephone dialogues contain differential speaker information as a function of syllabic position and labial co-articulation. These linguistic effects, established in earlier work on read broadband speech, were firstly investigated. Using a corpus of Dutch telephone speech, results showed that the telephone bandwidth captures the expected effects of perseverative and anticipatory labialization for back fricative /x/, for which spectral peaks fall within the telephone band, but not for front fricative /s/, for which the spectral peak falls outside the telephone band. Multinomial logistic regression shows that /s/ contains slightly more speaker information than /x/ in telephone speech and that speaker information is distributed across the speech signal in a systematic way; even though differences in classification accuracy were small, for both /s/ and /x/, codas and tokens with labial neighbours are more speaker-specific than onsets and tokens with non-labial neighbours. These findings indicate that speaker information contained by the same speech sound is not the same across contexts.

Keywords: Speech production (43.70.-h), Acoustical correlates of phonetic segments and suprasegmental properties: stress, timing, and intonation (43.70.Fq)

I. INTRODUCTION

Speakers' voices convey idiosyncratic information. In everyday communication, listeners make use of this information while interpreting what they hear and, in forensic phonetics, speech analysist use this information to acoustically characterise speakers. Although previous research has already shown that some speech sounds convey more speaker-dependent information than others (e.g. Kavanagh, 2012; Van den Heuvel, 1996), not much is known about how speaker-dependent information in the same speech sound interacts with its linguistic environment. The present study investigated the speaker-specificity, i.e. the ratio of between-speaker to within-speaker variation, of the same speech sound in different linguistic contexts. Specifically, we examined whether the speaker-specificity of Dutch fricatives varies as a function of syllabic position and labial co-articulation. Additionally, the aim was to determine which specific (combinations of) acoustic features are most successful in characterising speakers. Contrary to many previous studies that used read speech, the present study used spontaneous telephone dialogues to investigate speaker variation.

Investigating the distribution of speaker-dependent information is relevant for phonetic speech science because the role of the speaker in speech production is still largely unclear. It is known that speaker-dependent information conveys all kinds of meanings (e.g. gender identity) and that these meanings are also perceived by listeners. However, it is not clear where in the speech signal speakers have the articulatory freedom to convey speaker-dependent information, or if there are such distributional limitations. Additionally, this study may be particularly relevant for forensic speaker comparisons, where often low-quality speech samples are assessed in terms of the typicality and similarity of the speaker-dependent features they contain. The present work contributes to both fields by checking whether previously reported linguistic effects for fricatives are present in spontaneous telephone dialogues, which is a relevant speech style and channel both for everyday communication and forensic speaker comparisons, and whether these linguistic effects interact with the amount of speaker-dependent information for two highly frequent fricatives in Dutch.

A. Speaker variation

Previous work has already shown that some individual speech sounds contain more speaker-dependent information than others. For example, vowels are found to be more speaker-specific than consonants (Van den Heuvel, 1996: 145-146). Within the class of consonants, fricative /s/ – one of the speech sounds investigated in the present work – is found to be relatively speaker-specific. In Dutch read speech, /s/ was ranked below vowels and nasals, but above /r/ and plosives in terms of speaker-specificity (Van den Heuvel, 1996: 72). In English read speech, /s/ along with nasal /m/ are ranked above nasals /n/ and /ŋ/, and liquid /l/ (Kavanagh, 2012: 387-388). Studies on the speaker-specificity of fricatives that are not /s/ – such as the back fricative /x/ also examined in the present work – are rare.

As mentioned before, segments' speaker-specificity equals the ratio of between- to within-speaker variation. Theoretically, this means that factors affecting the within- and between-speaker variation have direct impact on a segment's speaker-specificity. Between speakers, variation in fricative acoustics has been observed across anatomical/physiological contexts, as well as social contexts. Regarding the former, fricative acoustics can vary as a function of the shapes and sizes of the articulators and cavities (Stevens, 2000: 411-412). In practice, this type of variation in fricative acoustics has often been observed between males and females; fricatives produced by females have higher resonance frequencies than by males, which is often explained as resulting from anatomical differences between the sexes (e.g. Jongman, Wayland, & Wong, 2000; Schwartz, 1968). This difference in production is perceivable and meaningful to listeners, as speaker sex can be perceived from isolated voiceless fricatives (Ingemann, 1968; Schwartz, 1968). Comparing speaker sex identification between fricative sounds, Ingemann (1968) found that listeners can identify speaker sex from isolated back fricatives [h, χ, x] but not from isolated front fricatives [θ, f, ɸ]. Front fricatives [s, ʃ] broke this pattern; speaker sex identification from these sounds was also above chance.

There are also social between-speaker factors that affect fricative acoustics. For example, there are well-attested effects of gender identity and sexual orientation on /s/ spectra that are not associated with anatomical/physiological differences but rather with production strategies, i.e. learned behaviour (e.g. Fuchs & Toda, 2010; Munson, McDonald, DeBoe, & White, 2006). Social class may also affect fricative spectra; Stuart-Smith (2007) found that English working-class females could be grouped with working-class males, rather than with higher-class females, on several spectral features from /s/. When looking at social identity on a larger scale, such as ethnolect, dialect, and language communities, variation in fricative spectra is also observed. For example, the so-called 'Moroccan flavoured Dutch' ethnolect is known for a retracted [s] realisation that resembles [ʃ], i.e. sibilant palatalization, in certain phonetic contexts (Mourigh, 2018). Another example is the regional variation for Dutch fricative /x/, which is produced with velar place of articulation (and thus higher resonance frequencies) in Flanders and Southern regions of the Netherlands and uvular place of articulation – often accompanied by uvular scrape, i.e. uvular trill – in Northern regions of the Netherlands (Van der Harst, Van de Velde, & Schouten, 2007).

Within speakers, it has been shown that fricative acoustics vary systematically as a function of phonetic context, speech style, and vocal effort. Regarding phonetic context, anticipatory lip-rounding has repeatedly been shown to lower resonance frequencies in fricatives (e.g. Bell-Berti & Harris, 1979; Koenig, Shadle, Preston, & Mooshammer, 2013). Anticipatory lip-rounding lowers the resonance frequencies in fricatives because the lip protrusion associated with the lip movement lengthens the anterior cavity. Notably, neighbouring labial consonants such as English bilabial /w/ and /p/ also seem to display a lowering effect on /s/ spectra (Munson, 2004), even though the lip movement for /p/ is better described as lip *closure* rather than lip-rounding. This implies that labial closure also lengthens the anterior cavity to some extent.

Speech style can also affect fricative acoustics within speakers. Maniwa, Jongman, and Wade (2009) compared clearly spoken fricatives to fricatives in a conversational speech style in American English and found that clearly spoken fricatives had longer duration, higher resonance frequencies, and – surprisingly – lower relative amplitude. Moreover, individual speakers used different strategies for producing clear speech, which were not related to speaker gender. This implies that different patterns of within- and between-speaker variation may be expected in clearly spoken speech versus conversational speech. It therefore seems important to extend research on speaker variation to include conversational speech styles.

Within speech styles, articulatory strengthening (hyperarticulation) or weakening (hypoarticulation) also affect fricative acoustics within speakers. Generally speaking, it has been shown that there are articulatory strong and weak locations in speech. Whereas the initial edges of prosodic domains such as phrases and words are generally found to be locations of articulatory strengthening (Cho & McQueen, 2005; Fougeron, 2001), the final edges of syllables, i.e. codas, are generally found to be locations of articulatory weakening compared to syllable onsets (Ohala & Kawasaki, 1984). For fricatives as a group, American English coda fricatives are found to be less identifiable (Redford & Diehl, 1999), and to have a lower intensity and a delayed and lower air pressure peak than onset fricatives (Solé, 2003). However, studies that consider different fricatives separately show inconsistent results for /s/ specifically; Redford & Diehl (1999) found coda reduction for duration in American English /s/, but not for intensity or spectral mean. Furthermore, they reported that, whereas consonant classification using linear discriminant analysis overall showed more accurately classified onsets than codas, this was not the case for /s/, where there was a reverse tendency. This lack of coda reduction for /s/ was replicated for German, where spectral mean for codas was not lower, but slightly higher than for onsets (Cunha & Reubold, 2015). Although there was no reduction effect for German /s/ in coda position, Cunha & Reubold (2015) found that codas display higher variability than onsets and that /s/ in de-accented syllables displays higher variability than /s/ in accented syllables. In other words, they reported more variability, but no reduction, in articulatory weak locations.

From the somewhat conflicting results reported above, it seems that not all fricatives reduce in the same manner or to the same extent. Rather, reduction seems to be constrained by specific production requirements (Recasens, 2004). This means that features that have high production requirements for a particular speech sound are more resistant to co-articulation and reduction than features that have low production requirements for a particular speech sound. For example, in fricatives /s/ and /x/, the resistance to anticipatory labialization might be low because there are no production requirements for the lips in /s/ and /x/. Tongue front and dorsum in the production of /s/, on the other hand, are relatively resistant to co-articulation and reduction due to the production necessity of tongue front raising for and dorsum lowering for this fricative (Recasens & Dolorspallarè, 2001). Speakers might vary in their articulatory timing, degree of co-articulation, and their reduction of specific features. This means that some speakers may be more sensitive to certain co-articulatory effects than others. As a result, the acoustic realisations of /s/

and /x/ might be more context-dependent in some speakers than others. It is therefore possible that highly context-dependent realisations, such as /s/ and /x/ in labialized context, display high between-speaker variability.

B. The distribution of speaker information

Studies on the distribution of speaker-specificity in speech are rare. Given that speaker-specificity is a ratio of between-speaker to within-speaker variation, speech samples need high between-speaker variation *and* low within-speaker variation to be speaker-specific. There are some linguistic contexts that might facilitate such environments, and thus help listeners extract speaker information. Namely, articulatory strong locations such as onsets, often argued to constitute canonical speech, may be characterised by low within-speaker variation. If these locations are not also characterised by low *between*-speaker variation, they might be relatively speaker-specific. Evidence supporting the hypothesis that articulatory strong locations are relatively speaker-specific comes from a finding that speakers were characterised more accurately using vowels receiving sentence stress, i.e. articulatory strong locations, than vowels without sentence stress (McDougall, 2006). Another example can be found in Heeren (2018), who showed that the vowel /a/ sampled from spontaneous speech contained more speaker-dependent information in content words than in function words. Content words are generally also found to be articulatory strong locations, which is evidenced by studies that found reduction in vowels sampled from function words relative to content words (Shi, Gick, Kanwischer, & Wilson, 2005; Van Bergem, 1993, pp. 38–39).

Alternatively to articulatory strong locations displaying high speaker-specificity, articulatory weak locations such as codas and highly context-dependent segments, e.g. fricatives with labial neighbours, might be characterised by high between-speaker variation and may therefore also display high speaker-specificity. Based on their work on formant and intensity dynamics, He, Dellwo, and colleagues hypothesise that speakers have more articulatory freedom in speech locations that are less constrained by articulatory targets, resulting in higher between-speaker variation in these locations. This is sometimes also referred to as variation due to target undershoot. They showed that both intensity dynamics (He & Dellwo, 2017) and formant dynamics (He, Zhang, & Dellwo, 2019) show more between-speaker variation in negative than in positive dynamics. Negative dynamics were defined as the intensity and formant slopes from the syllable's peak to the following trough, which He & Dellwo assume to be the parts of syllables associated with mouth-closing gestures. They suggest that the mouth-opening gestures (positive dynamics) are more restricted by articulatory targets.

Fricatives with labial neighbours thus display large co-articulation effects and codas often reduce compared to onsets. This suggests that fricatives in these linguistic environments are less constrained by articulatory targets and that speakers may have more articulatory freedom in these locations. In their work on coda reduction in fricatives, Cunha & Reubold (2015) indeed reported slightly more variability for fricative codas than onsets, which they attributed to target undershoot. Given that the degree and timing of labial co-articulation in fricatives might vary between speakers (Perkell & Matthies, 1992), fricatives with labialized context might also constitute relatively speaker-specific locations. However, there seems to be no quantitative evidence that fricatives in labial context show more between-speaker variability in the literature. There is, however, a study that looked at within-speaker variation for the spectral properties of /s/ as a function labial co-articulation. Replicating earlier research, Munson (2004) reported that /s/ has lower resonance frequencies when followed by rounded /u/ versus non-rounded /a/ and when followed by rounded /w/ versus vowels /a, u/, with labial – but not rounded – /p/ falling in-between. He hypothesised that variability in degree and timing of the labial co-articulation in /s/ would result in increased within-speaker variation. However, results only showed increased within-speaker variation for /s/ followed by /w/ and did not show larger within-speaker variation for /s/ followed by /u/ compared to when it is followed by /a/. It is probable that the lip-movements for /w/ versus /u/ and /p/ constitute different labial movements. Other work has shown that there are different types of labialization, e.g. different lip-area size involved in labialization for postalveolar fricatives /ʃ, ʒ/ versus /w/ (Toda, Maeda, Carlen, & Meftahi, 2003). It is therefore possible that the labial movement for /w/ is more sensitive to within-speaker variation than the labial movements for /u/ and /p/. Alternatively, /s/ followed by /w/ may display more within-speaker variation due to differences in articulatory timing between /s/ from consonant clusters versus consonant-vowel sequences.

C. Fricatives /s/ and /x/ in telephone speech

Fricative sounds are produced with a narrow constriction which results in noise generated by turbulence (Stevens, 2000: 379). The resonance frequencies of fricatives are mainly determined by the size of the cavity anterior to the narrow constriction (Stevens, 2000: 398-403). Whereas the Dutch laminal alveolar fricative /s/ has a relatively small anterior cavity and therefore high resonance frequencies, Dutch velar or uvular fricative /x/ has a medium to large anterior cavity and therefore much lower resonance frequencies. Fricative /s/ is reported to have a spectral centre of gravity of around 4.8 kHz in Standard Dutch read speech (Ditewig, Pinget, & Heeren, in press) and fricative /x/ is reported to have a spectral peak of around 1.7 kHz in Standard Dutch read speech (Van der Harst, Van de Velde, & Schouten, 2007). Most acoustic reports on /s/ and /x/ are based on studio-recorded read speech. However, this speech style is not representative of everyday communication nor of forensic speaker comparisons. It is currently unclear whether acoustic-phonetic and indexical information in /s/ and /x/ can be captured in spontaneous telephone dialogues, which is relevant for both everyday communication and forensic speech material.

Telephone signals have a limited frequency bandwidth. For example, the landline telephone dialogues worked with here have a sampling frequency of 8 kHz with an 8-bit resolution and were originally filtered at a bandwidth of 340 – 3400 Hz. Given that the spectral energy for Dutch /s/ is concentrated around 4.8 kHz (Ditewig, Pinget, & Heeren, in press), this means that the spectral energy for fricative /s/ mostly resides above the upper limit of this bandwidth (see Fig. 1). It is therefore possible that both linguistic information and speaker information from /s/ are (partly) lost in telephone speech. The spectral energy for back fricative /x/, on the other hand, falls mostly within the telephone bandwidth (see Fig. 2).

FIG. 1. Spectrograms of onset /s/ 340 – 3400 Hz telephone bandwidth from word *cd* (‘cd’, /seˈde/) spoken by a male speaker.

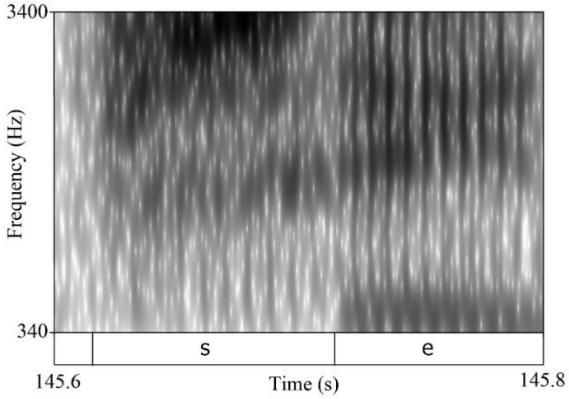
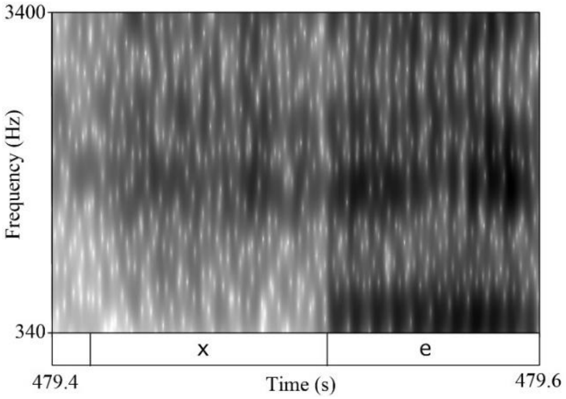


FIG. 2. Spectrogram of onset /x/ in 340 – 3400 Hz telephone bandwidth from word *geen* (‘no’, /xen/) by a male speaker.



Telephone speech also has other limitations that have to be considered in an acoustic analysis. Regarding signal-related transformative qualities, the lower formants may display an upward shift. Particularly F1 values may display a large shift of up to 14%, whereas higher formants generally remain unaffected (Byrne & Foulkes, 2004). Moreover, when this signal-related shift is paired with speaker-behaviour such as holding the phone between the cheek and shoulder, these upwards shifts are amplified (Jovičić, Jovanović, Subotić, & Grozdić, 2015). Additionally, the signal-related qualities of telephone speech are accompanied by distinct speech behaviour. For example, speakers often increase their vocal effort, possibly to adjust for increased background noises from variable environments. This effect is generally described as the Lombard effect (e.g. Junqua, 1993).

Despite the limitations of telephone speech, the use of this speech channel is preferred over higher-quality recordings when investigating speaker variation in real-world communicative contexts. Particularly in the context of forensic phonetics, telephone speech is highly relevant compared to studio-recorded (read) speech, as wiretapping telephone conversations from criminal suspects is common in police investigations in the Netherlands (Odinot, Jong, Leij, Poot, & Straalen, 2010: 82). Using higher-quality, non-telephone speech may therefore misrepresent what listeners may use in speech perception in daily conversation as well as what is possible for forensic speaker comparisons.

Finally, the availability of tokens likely to be available in spontaneous speech where the number of occurrences and the phonetic context are not controlled was also considered. Fricatives /s/ and /x/ are highly frequent in syllable onsets and to a slightly lesser extent in coda position in Dutch (Baayen, Piepenbrock, & Van Rijn, 1993), which makes them suitable speech sounds to analyse.

D. Research questions and hypotheses

Previous work has shown that fricative /s/ is a relatively speaker-specific consonant. However, reports so far have been based on studio-recorded read speech, which may display other patterns of speaker variation than conversational speech. Additionally, it might be problematic that the spectral energy for front fricative /s/ cannot be fully captured by telephone signals, which is a speech channel used in everyday communication. Spectrally-defining characteristics for back fricative /x/, on the other hand, should fall within the telephone bandwidth. The current work investigated the speaker-specificity of both of these fricative sounds in spontaneous telephone dialogues. Moreover, we investigated whether speaker-specificity varies as a function of linguistic context. Specifically, the effects of labial co-articulation and syllabic position on speaker-specificity were examined. As a secondary goal, the best acoustic measures in the speaker classification modelling were determined.

Regarding the speaker variation as a function of linguistic context, we hypothesised that articulatory strong locations (onsets and fricatives with non-labial neighbours) are characterised by low within-speaker variation and that articulatory weak locations (codas and fricatives with labial neighbours) are characterised by high between-speaker variation. However, there were no clear expectations for speaker-specificity, which equals the ratio of between- to within-speaker variation. In a second step, speaker classification analysis was performed on the data to see which (combinations of) acoustic measures and which linguistic contexts convey most speaker-dependent information. Regarding the performance of acoustic measures, previous findings report that spectral centre of gravity and standard deviation were the most speaker-discriminating features (e.g. Kavanagh, 2012). We therefore predicted that most speaker-specific information might be found in spectral as opposed to temporal or amplitudinal measures.

To answer the research questions, it was firstly ascertained whether acoustic measures extracted from /s/ and /x/ are sensitive to labial co-articulation and syllabic position in a corpus of spontaneous telephone dialogues, thus replicating and extending earlier results on read speech. Whereas previous studies strongly indicate that contextual labialization would lower fricative spectra, the literature is not clear on the effect of syllabic position, particularly for /s/. Moreover, it is unclear to what extent linguistic effects can be captured in the limited telephone bandwidth, also particularly for /s/.

II. METHODOLOGY

A. Corpus and speakers

Spontaneous telephone dialogues available in the Spoken Dutch Corpus (Oostdijk, 2000) were used to investigate the speaker-specificity in the realisation of fricatives /s/ and /x/. The telephone dialogues have a 340 – 3400 Hz band pass filter and were obtained via a switchboard. No information on the task is available, but from the recordings' content it was inferred that speakers were located in their home environment (deduced from background noises such as a crying baby or a barking dog) and were asked to converse for around ten minutes on any topic of their choosing. Variable numbers of telephone conversations – with different interlocutors – are available for each speaker in the corpus.

For 66 speakers aged 21 – 50 (M = 36.5, SD = 7.3), a total of 3,331 /s/ tokens and their adjacent contexts as well as 3,491 /x/ tokens with their adjacent contexts were first automatically segmented and provided with a broad phonetic transcription using the orthographic transcript available with the corpus. These were then manually validated by the first author. When interference such as laughter, overlapping speech from the interlocutor, or background noise showed up in the signal, tokens were excluded. Fricative tokens occurring in context with a creaky phonation were not excluded, as previous research has shown that /s/ spectra are relatively stable against creakiness (Hirson & Duckworth, 1993). Tokens were labelled as onsets (/s/: N = 1,359; /x/: N = 1,657), codas (/s/: N = 1,532; /x/: 1,453), or ambisyllabic (/s/: N = 440; /x/: N = 380). The latter category, containing tokens that cannot be categorised as either onsets or codas (e.g. *was ook* ‘was also’ [wɑsoːk]), was excluded from analysis.

As reviewed above, labialization of adjacent context affects fricative spectra. To test whether the measures extracted from telephone speech are sensitive to contextual labialization, preceding and following context was furthermore labelled as labial or non-labial. Rounded vowels /u, ɔ, o, ø, y, ʏ/, (partially) rounded diphthongs /œy, ɑu/ (see temporal patterns of lip-rounding: Bell-Berti & Harris, 1982) , and bilabial consonants /p, b, m/, were considered to be labial. Labiodental consonants /f, v, ʋ/ were not coded as labial because the teeth-to-lip movement in these sounds does not involve lip-rounding or closure, but rather eliminates the anterior cavity and can therefore not be assumed to have the same lowering effect on the spectrum. Speakers with fewer than 25 tokens per fricative sound were excluded, which excluded 23 speakers and left a total of 43 speakers with a number of sufficient tokens for both /s/ and /x/. The resulting numbers of tokens per factor level are presented in Table I.

TABLE I. Totals, and means, standard deviations, and ranges for numbers of /s/ and /x/ tokens by speaker (N = 43) and by linguistic context factor level.

		Total	Syllabic Position		Left Context		Right Context	
			Onset	Coda	Non-labial	Labial	Non-labial	Labial
/s/	Total	2,346	1,066	1,280	1,846	500	1,903	443
	M	54.56	24.79	29.77	42.93	11.63	44.26	10.30
	SD	19.29	11.32	11.16	15.70	4.95	14.52	6.61
	range	25 – 108	9 – 63	15 – 78	20 - 88	3 - 22	24 – 88	1 – 35
/x/	Total	2,820	1,460	1,360	2,336	484	2,250	570
	M	65.58	33.95	31.63	54.33	11.26	52.33	13.26
	SD	26.06	12.70	15.17	22.56	5.50	21.64	6.57
	range	27 – 124	11 – 67	9 – 73	20 – 106	3 – 29	23 – 100	3 – 31

B. Acoustic measurements

As mentioned in section I-C, the telephone dialogues available in the Spoken Dutch Corpus have a sampling frequency of 8 kHz with an 8-bit resolution and were originally filtered at a bandwidth of 340 – 3400 Hz. There are separate channels for the two speakers in each telephone conversation. A low-frequency cut-off of 500 Hz was used to reduce the influence of background noise and (partial) voicing. Hence, all measures were taken over a 0.5 – 3.4 kHz frequency range. For each fricative token, the duration, static amplitudinal and spectral measures, and dynamic spectral measures were taken in Praat version 6.0.46 (Boersma, 2001). Duration (in milliseconds, ms) was computed from fricative onset to fricative offset as characterised by the presence of aperiodic fricative noise, which was then used to establish the middle 50% of each fricative over which the static spectral measures were taken. The static spectral measures consisted of the first two spectral moments and spectral tilt. After filtering the fricative tokens to the 500 – 3400 Hz band (band pass Hann filter, smoothing = 100 Hz), the centre of gravity and the standard deviation (in Hertz, Hz) were computed from the spectrum determined over the mid-50% of the fricative, using power spectrum weighting. Spectral tilt was measured to reflect vocal effort as an alternative to absolute amplitudinal measures, and computed from the long-term average spectrum determined over the mid-50% of the fricative (bin =1 Hz) on a logarithmic frequency scale (dB/decade), using a least-squares fit. A decade is a step on the frequency scale with the power of 10, i.e. 1 Hz, 10 Hz, 100 Hz, etc. Mean amplitude (in dB) was measured over the full fricative's duration and normalised by speaker through z-transformation.

Additional to the static measures, dynamic spectral measures were computed by measuring spectral centre of gravity in non-overlapping 20%-portions of the entire fricative's duration. Coefficients from quadratic polynomial equations over the five resulting data points per fricative token constituted our dynamic measures for analysis. Both cubic and quadratic models to the data were estimated; a likelihood ratio test showed no significant difference between these two models (/s/: χ^2 (1) = 0.96, p = .33; /x/: χ^2 (1) = 0.11, p = .74). The simpler quadratic function was chosen as the fewer coefficients reduced the number of predictors in further modelling. The quadratic intercept of the dynamic CoG measure was excluded because it correlated very highly with the static CoG measure (/s/: *r* = .95, N = 2,346, p < .001; /x/: *r* = .96, N = 2,820, p < .001). The two remaining quadratic coefficients from the dynamic CoG measure will henceforth be referred to as CoG1 and CoG2.

C. Statistical analysis

The statistical analysis consisted of two parts: (1) linear mixed-effect modelling was used to check whether linguistic factors affected /s/ and /x/ acoustics in telephone speech, and (2) multinomial logistic regression was used to investigate whether the amount of speaker information in /s/ and /x/ varied as a function of syllabic position and labial co-articulation. Additionally, the relative importance of acoustic measures was estimated from the regression model. The ratio of between- to within-speaker variance, referred to as the Speaker-Specificity

Index (SSI: Van den Heuvel, 1996), was computed for all acoustic variables to assess its relationship with the regression modelling results.

1. Linear mixed-effect modelling: Linguistic effects

In the first part of the analysis, the effects of linguistic context factors on acoustic measures were investigated by means of linear mixed-effect modelling, using function *lmer()* from R package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). The dependent variable was a single acoustic measure. The fixed part of the maximal model contained binary factors for Left Context (non-labial, labial; dummy coded), Right Context (non-labial, labial; dummy coded), and Syllabic Position (coda, onset; sum coded). The random part of the maximal model contained random intercepts for Word and Speaker, as well as random slopes for Speaker over all three fixed predictors. First, a full model with maximal random structure was built by restricted maximum likelihood (REML) estimation (Barr, Levy, Scheepers, & Tily, 2013). Next, stepwise deletion was used to reduce the random structure of the model, given this lead to a better-fitting model using log-likelihood testing and this was theoretically justifiable (Bates, Kliegl, Vasishth, & Baayen, 2015). Model fit was assessed through inspection of the residuals. Duration was log-transformed (base = 10) for a better model fit.

Lastly, models were rebuilt including possibly confounding factors for Phrasal Position (initial, medial, final; sum coded) and Word Stress (non-stressed, stressed; sum coded) to see if results were maintained. For Word Stress, only tokens from content words (nouns, verbs, adjectives, and adverbs) were labelled for word stress, as function words can have stressed syllables only in special circumstances (Selkirk, 1996). This resulted in the exclusion of 16% of the data for /s/ and 12% of the data for /x/. Results from these latter models are not presented because including extended these models did not change results obtained by earlier models, although exact statistics were slightly different.

2. *Multinomial logistic regression: Speaker classification accuracies per linguistic context and acoustic measure*

Multinomial logistic regression (MLR) was used to test which linguistic context factors and measures significantly predicted the dependent variable Speaker. As a first step, function *buildmultinom()* from R Package *buildmer* (Voeten, 2019) was used to automatically build and then reduce the maximal MLR model by backward stepwise selection using likelihood-ratio tests. All acoustic measures (CoG, SD, tilt, amplitude, duration, CoG1, and CoG2) and linguistic factors (Syllabic Position, Left Context, and Right Context) were added to the maximal model as predictors for Speaker. Each acoustic measure was allowed to interact with each linguistic context factor. Highly correlating predictors ($r > .70$) were excluded, which resulted in the exclusion of spectral tilt because it correlated highly with CoG (/s/: $r = .76$, N = 2,346, p < .001; /x/: $r = .91$, N = 2,820, p < .001).

In a second step, the optimal model obtained by function *buildmultinom()* was inspected to see which combinations of acoustic measures and linguistic context factors affect speaker classification predictions. The predicted speaker classification of factor levels was compared, i.e. for Syllabic Position, speaker classification of codas is compared to onsets. This was achieved by sub setting the data on factor level and then predicting speaker classification on the resulting two datasets using the best-fitting model acquired in the previous step. This was done for factor levels from all linguistic context factors that were included in the best-fitting models. Secondly, acoustic measures and their interactions with linguistic context factors were excluded from the best-fitting model one at a time to access the relative importance of each acoustic measure.

III. RESULTS

A. Linguistic effects

Linear mixed-effect modelling results for /s/ are summarised in Table II, where it can be seen that /s/ onsets have higher CoG, higher positive spectral tilt, i.e. more high-frequency energy, higher amplitude and shorter duration than codas. In other words, all measures from /s/ except duration show coda reduction. Note also that the spectral tilt intercept in Table 2 is a positive value, i.e. there is no energy drop-off but an increase in higher frequencies. This is expected for /s/ because all the spectral energy is expected to reside in the higher frequencies of the telephone band. Regarding labial co-articulation, there were no effects for left context. For right context, /s/ tokens with right labial neighbours have lower SD, shorter duration in codas, and – opposite to what we hypothesised – higher CoG.

TABLE II. Summary of fixed effects from linear mixed-effect modelling for /s/ (N = 2,346) with Kenward-Roger degrees of freedom approximation. Reference values are ‘coda’ for Syllabic Position and ‘non-labial’ for Left and Right Context.

DV	Fixed effects	*Est.*	*SE*	*t*	*p*
CoG [Hz]	(intercept)	2537	37	68.2	<.001***
	SyllPos: onset	25	9	2.7	<.01**
	Left Context: labial	-4	29	76	.90^{NS}
	Right Context: labial	76	19	4.0	<.001***
SD [Hz]	(intercept)	603	18	32.7	<.001***
	Right Context: labial	-42	9	-4.7	<.001***
Tilt [dB/decade]	(intercept)	17.9	0.4	45.9	<.001***
	SyllPos: onset	1.3	0.4	3.3	<.001***
Amp (normalized)	(intercept)	0.04	0.03	1.5	0.12^{NS}
	SyllPos: onset	0.15	0.03	5.5	<.001***
Dur [log(ms)]	(intercept)	-1.05	0.01	-121.6	<.001***
	SyllPos: onset	-0.03	0.01	-3.9	<.001***
	Right Context: labial	-0.06	0.01	-5.5	<.001***
	Right Context * SyllPos	0.05	0.01	4.0	<.001***

Results for /x/ are summarised in Table III, which shows that /x/ onsets, like /s/ onsets, have higher amplitude than codas. Contrary to results for /s/, when left context is labialized, CoG lowers and spectral tilt decreases, i.e. there is less energy at higher frequencies. When right context is labialized, CoG lowers (although this effect is larger for onsets), tilt decreases, and amplitude decreases. The interactions between left and right context for CoG and tilt indicate that spectral lowering is attenuated by 131 Hz or 4.6 dB per decade when both left and right context are labialized. Contrasting our data for /s/, spectral tilt for /x/ shows a negative value. This shows that whereas there is no energy drop-off for high-frequency /s/, there is an average energy drop-off of 7.8 dB per decade for /x/.

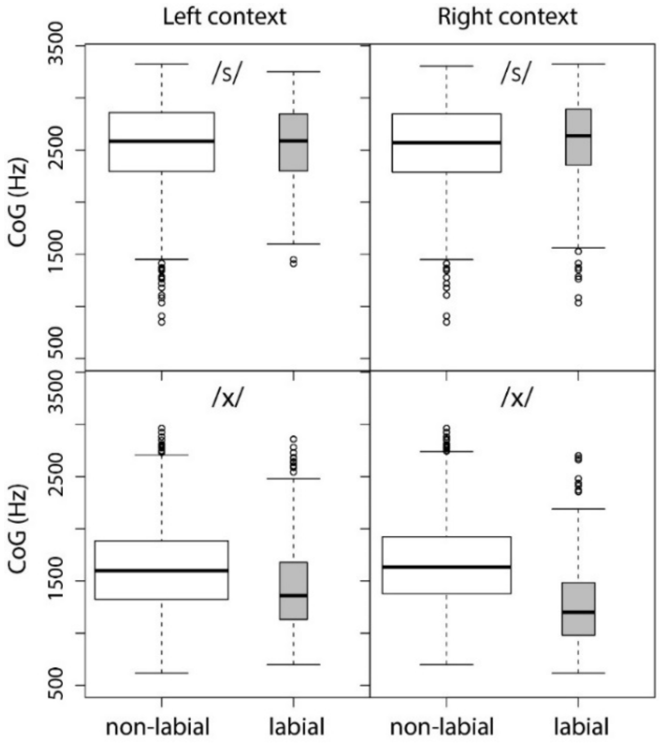
TABLE III. Summary of fixed effects from linear mixed-effect modelling for /x/ (N = 2,820) with Kenward-Roger degrees of freedom approximation. Reference values are ‘coda’ for Syllabic Position and ‘non-labial’ for Left and Right Context.

DV	Fixed effects	*Est.*	*SE*	*t*	*p*
CoG [Hz]	(intercept)	1652	34	48.6	<.001***
	SyllPos: onset	-8	13	-0.6	.52^{NS}
	Left Context: labial	-215	27	-8.0	<.001***
	Right Context: labial	-299	40	-7.6	<.001***
	SyllPos * Right Context	-99	30	-3.3	<.001***
	Left * Right Context	131	51	2.6	<.05**
SD [Hz]	(intercept)	599	14	42.5	<.001***
	SyllPos: onset	-8	4	-1.8	.08^{NS}
	Left Context: labial	19	11	1.7	.08^{NS}
	Right Context: labial	0	21	0	.98^{NS}
	SyllPos * Right Context	-57	11	-5.1	<.001***
	SyllPos * Left Context	-47	10	-4.5	<.001***
Tilt [dB/decade]	(intercept)	-7.8	1.2	-6.3	<.001***
	SyllPos: onset	-0.7	0.4	-1.6	.12^{NS}
	Left Context: labial	-7.4	0.9	-8.6	<.001***
	Right Context: labial	-8.9	1.3	-7.1	<.001***
	SyllPos * Right Context	-4.1	1.0	-4.3	<.001***
	Left * Right Context	4.6	1.7	2.7	<.01***
Amp (normalized)	(intercept)	0.03	0.03	1.0	.32^{NS}
	SyllPos: onset	0.21	0.03	7.1	<.001***
	Left Context: labial	-0.10	0.07	-1.4	.14^{NS}

	Right Context: labial	-0.27	0.07	-3.6	$<.001^{***}$
	SyllPos * Left Context	0.15	0.07	2.13	$<.05^{*}$
Dur [log(ms)]	(intercept)	1.92	0.01	212.8	$<.001^{***}$
	SyllPos: onset	0.01	0.01	1.1	$.27^{NS}$
	Right Context: labial	-0.02	0.01	-1.2	$.24^{NS}$
	Right Context * SyllPos	0.09	0.01	6.5	$<.001^{***}$

Fig. 3 illustrates the differences in linguistic context effects between the two fricatives under study, /s/ and /x/. Whereas /x/ CoG lowers when context is labial, this is clearly not the case for /s/. As hypothesised, this may be due to the telephone bandwidth. If the amount of speaker information is sensitive to linguistic context factors, the acoustic results would predict stronger effects for /x/ than /s/ in the second analysis.

FIG. 3. Boxplots for Centre of Gravity (CoG) by fricative sound, syllabic position, and left and right context labialization. The width of the box represents the number of cases.



B. Speaker classification

For /s/, the best-fitting model to predict speaker (N = 43, n = 2,346) included all acoustic measures and all linguistic context factors as significant predictors. Additionally, the following interactions between linguistic factors and acoustic measures were included: Syllabic Position with spectral CoG, mean amplitude, duration, CoG1, and CoG2; Left Context with spectral CoG, SD, and duration; and Right Context with CoG, mean amplitude, duration, CoG1, and CoG2. This model had a speaker classification accuracy of 19.5% against a chance level of 2.3%.

For /x/, the best-fitting model to predict speaker (N = 43, n = 2,820) also included all acoustic measures and all linguistic context factors as significant predictors. This model furthermore included all possible interactions between linguistic context factors and acoustic measures except for the interaction between Left Context and amplitude, duration, CoG1, and CoG2, and between Syllabic Position and CoG2. This model had a speaker classification accuracy of 18.4% (chance = 2.3%). Per linguistic context, speaker classification accuracies are similar (see Table IV), but there seems to be a small, yet systematic, advantage for articulatory weak locations, i.e. codas and tokens with labial co-articulation.

TABLE IV. Speaker classification accuracies (in %) per fricative sound and per linguistic context factor level (chance level = 2.3%).

Linguistic context		/s/	/x/
	Total	19.5	18.4
Syllabic position	Onset	19.5	18.2
	Coda	19.5	18.6
Left Context	Non-labial	18.3	18.5
	Labial	24.2	18.8
Right Context	Non-labial	18.5	17.6
	Labial	18.8	21.4

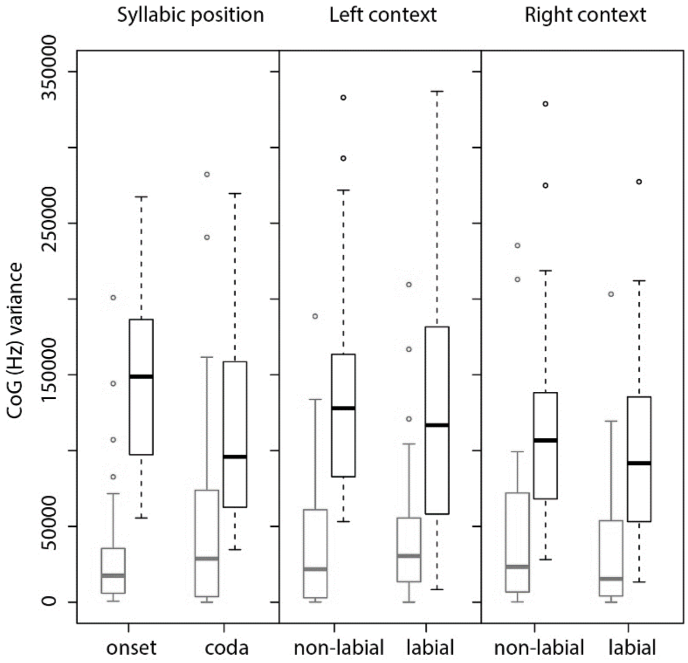
Next, the decreases in speaker classification accuracy when a single acoustic measure and its interactions with linguistic context factors were dropped from the model are presented. For example, excluding CoG and the interactions between CoG and linguistic context factors from the best-fitting model for /s/ resulted in a drop in speaker classification accuracy from 19.5% (for the optimal model) to 13.9%, which makes a decrease of 5.6%. As can be seen in Table V, CoG and SD were relatively important measures for speaker classification. Moreover, measures contributed to speaker classification in comparable ways across fricatives. The contribution of acoustic measures to the speaker classification from the MLR model is accompanied by an SSI calculated from the between- versus within-speaker variance in the data. The SSIs more or less mirror the relative ranking from the MLR model.

TABLE V. Speaker classification accuracy decreases (in %) per acoustic measure relative to the full models' speaker classification accuracy of 19.5% for /s/ and 18.4% for /x/ and speaker-specificity index (SSI) per acoustic measure for /s/ and /x/.

	/s/		/x/	
Excluded measure	Δacc	SSI	Δacc	SSI
CoG [Hz]	5.6	0.56	4.5	0.26
SD [Hz]	4.5	0.63	3.4	0.31
Dur [log(ms)]	1.9	0.07	2.1	0.10
CoG1 [Hz]	0.9	0.07	1.6	0.06
CoG2 [Hz]	1.3	0.08	1.2	0.07
Amp (norm.)	1.1	0.14	0.7	0.06

Lastly, the question whether the small advantage in speaker classification for articulatory weak locations is due to high between-speaker variation is examined. Fig. 4 shows that, for /x/ CoG, the within-speaker variance is consistently higher than the between-speaker variance (consistent with SSIs reported in Table V). Additionally, as hypothesised, the between-speaker variance seems to be increased in articulatory weak locations compared to strong locations. Against expectation, the within-speaker variation seems to be decreased in articulatory weak locations.

FIG. 4. Boxplots of between- (grey) and within-speaker (black) variances per linguistic context factor level for /x/ CoG. For each speaker, within-speaker variance was quantified as $within = (x - \mu_{speaker})^2$ and between-speaker variance as $between = (\mu_{speaker} - \mu_{total})^2$.



IV. DISCUSSION

Previous work on read speech has shown that linguistic effects such as labial co-articulation and syllabic position have effects on fricative acoustics, and that some segments, such as /s/, are more speaker-specific than other segments. The present study wished to further investigate (1) whether linguistic effects on fricative spectra are present in speech materials that were not recorded in highly-controlled circumstances, in this case, telephone dialogues, and (2) whether there is an interaction between segments' speaker-specificity and their linguistic context.

Regarding the first aim, linguistic effects were present in /x/, but were less prominent in /s/. The effect of syllabic position was present in both fricative sounds. Onsets showed higher intensity for both fricatives, which is consistent with results reported by Solé (2003) for American English fricatives. However, only for /s/ was there any indication for coda reduction in spectral measures, namely higher CoG in /s/ onsets compared to codas.

As for labialization, the results confirmed the expected linguistic effects in /x/ acoustics; both left and right labial neighbours lower the resonance frequencies in /x/ by around 200 Hz and 300 Hz respectively. This is consistent with work on /s/ from read speech where anticipatory labialization lowered spectral energy by around 300~400 Hz (Koenig et al., 2013). Two significant interaction effects for CoG and spectral tilt furthermore indicated that spectral lowering is attenuated when both left and right context are labial and that the effect of anticipatory labialization is slightly larger in onsets. Regarding the first interaction, spectral lowering in these cases might be attenuated to not undershoot the articulatory target for /x/. The second interaction could be explained by more resistance to co-articulation across word boundaries; all onsets in this dataset were word-initial and all codas were word-final. This means that right context for onsets was part of the same syllable, whereas left context for onsets was part of the previous word. Previous work, however, found only minor effects of prosodic boundaries on co-articulation of consonant cluster [kl], and then predominantly when articulation rate was slow (Hardcastle, 1985), which makes this explanation less likely. Alternatively, the second interaction may reflect a qualitative difference in the type of lip-rounding; whereas right labial context for onsets consisted of rounded vowels, right labial context for codas consisted exclusively of bilabial consonants /b, p, m/ (because codas followed by vowels were labelled as ambisyllabic). Given that Munson (2004) has shown that the labialization effect in /s/ before /p/

was smaller than before /u/, the present result that anticipatory labialization lowers /x/ spectra more in onsets is therefore likely to stem from the specific labial segments that followed /x/ in onset versus coda position.

Contrary to /x/, the /s/ acoustics did not show the expected spectral lowering in labial contexts; in fact, when right context was labial, CoG showed a small but significant increase. The lack of spectral lowering in /s/ acoustics is likely a result of the speech channel used here, as much of the spectral energy for /s/ falls above the upper limit of the telephone bandwidth. In other words, given that the effect of labial co-articulation is well-attested for /s/, it is likely that labial co-articulation effects are not captured in these data. From the literature as well as the current results on /x/, the lowering due to labialization would be on the order of 300 Hz, which – relative to 4.8 kHz for a Dutch /s/ CoG – falls outside of the telephone band.

Whereas the telephone band did not capture most energy for /s/, it did for /x/. This is supported by the mean CoG values; the mixed model's CoG intercept of 1.7 kHz for /x/ (CoG mean from the data was 1,586 Hz, SD = 421 Hz) was very similar to previously reported resonance frequencies for Dutch /x/ (Van der Harst et al., 2007). However, for /s/, the mixed model's CoG intercept of 2.5 kHz (M = 2,548 Hz, SD = 387 Hz) was around 2 kHz lower than what previous broadband studies have reported (Ditewig et al., in press). In other words, we assume that the actual spectral peaks for /s/ were far over the upper limit of the landline telephone bandwidth used here, resulting in much lower CoG values in the present analysis with a lack of linguistic effects as a result.

Regarding the dependence of the speaker information on linguistic context in spontaneous telephone speech, the speaker-specificity of fricatives /s/ and /x/ seems to be distributed across linguistic contexts in a systematic way, but differences in speaker classification accuracies were very small. In the current results, articulatory weak locations, i.e. codas and fricatives with labial neighbours, seemed slightly more speaker-specific than articulatory strong locations, i.e. onsets and fricatives with non-labial neighbours, for both /s/ and /x/. It thus seems that our data provides further evidence for the hypothesis proposed by He, Dellwo, and colleagues (2017; 2019) that speech locations that are less constrained by articulatory targets are more speaker-specific. Further examination of the between- and within-speaker variances showed that, for /x/ CoG, both between-speaker variance was increased and within-speaker variation was decreased in articulatory weak locations.

Interestingly, speech features sampled from articulatory weak locations seemed to contain more speaker-dependent information even in the absence of clear acoustic differences. Fricative /x/ acoustics were altered by linguistic context within the telephone band and simultaneously showed differences in speaker-classification per linguistic context. However, /s/ also showed higher speaker classification accuracies in articulatory weak locations, even though the expected acoustic effects for /s/ were minimal. The relative differences in speaker classification per linguistic context were very similar, and small, for /s/ and /x/. Therefore, there is a possibility that these results are dependent on the specific sampling of the current dataset, which we assume to reflect distributional patterns of conversational Dutch; there are many more /s/ and /x/ tokens with non-labial context than with labial context (see Table I in section A). We cannot exclude that the lower number of labial contexts may have resulted in an under-estimation of speaker variance in that particular context. Given the minor differences between linguistic contexts, however, the results are expected to have no major implications for either listeners' perception of speaker information or for forensic speaker comparisons.

Comparing the contribution of the different acoustic measures to the speaker-classification accuracy of the multinomial logistic regression model, our results are similar to those reported by Kavanagh (2012) for English /s/ from read speech. Namely, spectral centre of gravity and standard deviation are speaker-specific acoustic measures compared to temporal and amplitudinal measures. This might be because, whereas spectral measures reflect the size and shape of resonance cavities in the production of fricatives, this is not the case for temporal and amplitudinal measures. Notably, the contributions of acoustic measures to speaker-specificity were very similar for the two fricative sounds examined here.

Interestingly, when using the same set of measures, fricative /s/ seems to be slightly more speaker-specific than /x/ even though the spectral peak of /s/ is not captured by the telephone bandwidth. In other words, /s/ retains some speaker-specificity even in limited bandwidths. Moreover, for the Dutch situation, another highly frequent fricative, /x/, contains comparable amounts of speaker-dependent information in telephone speech. The correlation coefficient between the mean CoG values per speaker for /s/ and /x/ ($r = .46$, n = 43, p < .01) furthermore shows that the two fricative sounds carry partly complementary speaker-dependent information.

V. CONCLUSION

The present study investigated the distribution of speaker-specificity in fricatives /s/ and /x/ as a function of syllabic position and labial co-articulation. Results have firstly shown that, whereas previous findings on studio-recorded read speech can be replicated for back fricative /x/ from spontaneous telephone speech, this is less so the case for front fricative /s/. We argue that the lack of effects for labial co-articulation for /s/ is a result of the telephone bandwidth used here. Secondly, for both /s/ and /x/, results showed somewhat more speaker-specificity for codas and for tokens with labial context. However, differences in speaker-specificity per linguistic context were small. These results support the hypothesis that the role of the speaker in speech is more explicit in parts of the speech signal where speakers have more articulatory freedom.

ACKNOWLEDGEMENTS

This work was funded by the NWO Vidi grant number 276-75-010.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX Lexical Database*. Philadelphia Linguistics Data Consortium University of Pennsylvania.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” J. Mem. Lang. 68, 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). “Fitting linear mixed-effects models using lme4,” J. Stat. Softw.

Bell-Berti, F., & Harris, K. S. (1979). “Anticipatory coarticulation: Some implications from a study of lip rounding,” J. Acoust. Soc. Am. 65, 1268–1270.

Bell-Berti, F., & Harris, K. S. (1982). “Temporal patterns of coarticulation: Lip rounding,” J. Acoust. Soc. Am. 71, 449–454.

Boersma, P. (2001). “Praat, a system for doing phonetics by computer,” Glot Int. 5, 341–347.

Byrne, C., & Foulkes, P. (2004). “The “Mobile Phone Effect” on Vowel Formants,” Speech Lang. Law 11, 83–102.

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, 121-157.

Cunha, C., & Reubold, U. (2015). “The contribution of vowel coarticulation and prosodic weakening in initial and final fricatives to sound change,” In *Proc. 18th ICPhS*, Glasgow.

Ditewig, S., Pinget, A. C. H., & Heeren, W. F. L. (in press). “Regional variation in the pronunciation of /s/ in the Dutch language area,” Nederlandse Taalkunde 24.

Fougeron, C. (2001). “Articulatory properties of initial segments in several prosodic constituents in French,” J. Phonetics 29, 109–135.

Fuchs, S., & Toda, M. (2010). “Do differences in male versus female /s/ reflect biological or sociophonetic factors?” In S. Fuchs, M. Toda, & M. Żygis (Eds.), *Turbulent sounds: an interdisciplinary guide* (pp. 281–302). Berlin: De Gruyter Mouton.

Hardcastle, W. J. (1985). “Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences,” Speech Comm. 4, 247–263.

Harst, S. Van der, Velde, H. Van de, & Schouten, B. (2007). “Acoustic characteristics of standard Dutch /x/,” In *Proc. 16th ICPhS*, Saarbrücken.

He, L., & Dellwo, V. (2017). “Between-speaker variability in temporal organizations of intensity contours,” J. Acoust. Soc. Am. 141, EL488–EL494.

He, L., Zhang, Y., & Dellwo, V. (2019). “Between-speaker variability and temporal organization of the first formant,” J. Acoust. Soc. Am. 145, EL209–EL214.

Hirson, A., & Duckworth, M. (1993). “Glottal fry and voice disguise: a case study in forensic phonetics,” J. Biomedical Engin. 15, 193–200.

Ingemann, F. (1968). “Identification of the Speaker’s Sex from Voiceless Fricatives,” J. Acoust. Soc. Am. 44, 1142–1144.

Jongman, A., Wayland, R., & Wong, S. (2000). “Acoustic characteristics of English fricatives,” J. Acoust. Soc. Am. 108, 1252.

Jovičić, S. T., Jovanović, N., Subotić, M., & Grozdić, Đ. (2015). “Impact of mobile phone usage on speech spectral features: Some preliminary findings,” Int. J. Speech, Lang. Law 22, 111-124.

Junqua, J. (1993). “The Lombard reflex and its role on human listeners and automatic speech recognizers,” J. Acoust. Soc. Am. 93, 510–524.

Kavanagh, C. M. (2012). *New consonantal acoustic parameters for forensic speaker comparison*. Doctoral dissertation, University of York. Retrieved from http://etheses.whiterose.ac.uk/3980/

Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). “Toward Improved Spectral Measures of /s/: Results From Adolescents,” J. Speech Lang. Hear. Res. 56, 1175-1189.

Maniwa, K., Jongman, A., & Wade, T. (2009). “Acoustic characteristics of clearly spoken English fricatives,” J. Acoust. Soc. Am. 125, 3962–3973.

McDougall, K. (2006). “Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies,” Int. J. Speech, Lang. Law 13, 89–125.

Mourigh, K. (2018). “Stance-taking through sibilant palatalisation in Gouda Moroccan Dutch,” Nederlandse Taalkunde 22, 421–446.

Munson, B. (2004). “Variability in /s/ Production in Children and Adults,” J. Speech Lang. Hear. Res. 47, 58.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). “The acoustic and perceptual bases of judgments of women and men’s sexual orientation from read speech,” J. Phonetics 34, 202–240.

Odinot, G., Jong, D. de, Leij, J. B. J. van der, Poot, C. J. de, & Straalen, E. K. van. (2010). *Het gebruik van de telefoon- en internettap in de opsporing* (Report). Meppel: Boom Lemma uitgevers. Retrieved from https://repository.tudelft.nl/view/wodc/uuid:a4b1041c-0af4-4b30-bca2-ecc28dd79c8d

Ohala, J. J., & Kawasaki, H. (1984). “Prosodic phonology and phonetics,” Phonology 1, 113–127.

Oostdijk, N. H. J. (2000). “Corpus Gesproken Nederlands,” Nederlandse Taalkunde 5, 280–284.

Perkell, J. S., & Matthies, M. L. (1992). “Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability,” J. Acoust. Soc. Am. 91, 2911–2925.

Recasens, D. (2004). “The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters),” J. Phonetics 32, 435–453.

Recasens, D., & Dolorspallarè, M. (2001). “Coarticulation, assimilation and blending in Catalan consonant clusters,” J. Phonetics 29, 273–301.

Redford, M. A., & Diehl, R. L. (1999). “The relative perceptual distinctiveness of initial and final consonants in CVC syllables,” J. Acoust. Soc. Am. 106, 1555-1565.

Schwartz, M. F. (1968). “Identification of Speaker Sex from Isolated, Voiceless Fricatives,” J. Acoust. Soc. Am. 43, 1178–1179.

Selkirk, E. (1996). “The Prosodic Structure of Function Words,” In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping From Speech To Grammar in Early Acquisition* (pp. 187–214). Mahwah, NJ: Erlbaum.

Shi, R., Gick, B., Kanwischer, D., & Wilson, I. (2005). “Frequency and Category Factors in the Reduction and Assimilation of Function Words: EPG and Acoustic Measures,” J. Psycholinguistic Res. 34, 341–364.

Solé, M.-J. (2003). “Aerodynamic characteristics of onset and coda fricatives,” In *Proc. 15th ICPhS* (pp. 2761–2764). Barcelona.

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

Stuart-Smith, J. (2007). “Empirical evidence for gendered speech production: /s/ in Glaswegian. Change in Phonology: Papers Lab,” Phonology 9, 65–86.

Toda, M., Maeda, S., Carlen, A. J., & Meftahi, L. (2003). “Lip protrusion/rounding dissociation in French and English consonants: / w / vs. / ʃ / and / ʒ /,” In *Proc. 15th ICPhS* (pp. 1763–1766). Barcelona.

Van Bergem, D. R. (1993). “Acoustic vowel reduction as a function of sentence accent, word stress, and word class,” Speech Comm. 12, 1–23.

Van den Heuvel, H. (1996). *Speaker variability in acoustic properties of Dutch phoneme realisations.* Doctoral dissertation, Radboud Universiteit. Retrieved from http://repository.ubn.ru.nl/bitstream/handle/2066/76416/76416.pdf

Voeten, C. (2019). Buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression (R-package).

Tables

TABLE I. Totals, and means, standard deviations, and ranges for numbers of /s/ and /x/ tokens by speaker (N = 43) and by linguistic context factor level.

		Total	Syllabic Position		Left Context		Right Context	
			Onset	Coda	Non-labial	Labial	Non-labial	Labial
/s/	Total	2,346	1,066	1,280	1,846	500	1,903	443
	M	55	25	30	43	12	44	10
	SD	19	11	11	16	5	15	7
	range	25 – 108	9 – 63	15 – 78	20 - 88	3 - 22	24 – 88	1 – 35
/x/	Total	2,820	1,460	1,360	2,336	484	2,250	570
	M	66	34	32	54	11	52	13
	SD	26	13	15	23	6	22	7
	range	27 – 124	11 – 67	9 – 73	20 – 106	3 – 29	23 – 100	3 – 31

TABLE II. Summary of fixed effects from linear mixed-effect modelling for /s/ (N = 2,346) with Kenward-Roger degrees of freedom approximation. Reference values are ‘coda’ for Syllabic Position and ‘non-labial’ for Left and Right Context.

DV	Fixed effects	*Est.*	*SE*	*t*	*p*
CoG [Hz]	(intercept)	2537	37	68.2	<.001***
	SyllPos: onset	25	9	2.7	<.01**
	Left Context: labial	-4	29	76	.90^{NS}
	Right Context: labial	76	19	4.0	<.001***
SD [Hz]	(intercept)	603	18	32.7	<.001***
	Right Context: labial	-42	9	-4.7	<.001***
Tilt [dB/decade]	(intercept)	17.9	0.4	45.9	<.001***
	SyllPos: onset	1.3	0.4	3.3	<.001***
Amp (normalized)	(intercept)	0.04	0.03	1.5	0.12^{NS}
	SyllPos: onset	0.15	0.03	5.5	<.001***
Dur [log(ms)]	(intercept)	-1.05	0.01	-121.6	<.001***
	SyllPos: onset	-0.03	0.01	-3.9	<.001***
	Right Context: labial	-0.06	0.01	-5.5	<.001***
	Right Context * SyllPos	0.05	0.01	4.0	<.001***

TABLE III. Summary of fixed effects from linear mixed-effect modelling for /x/ (N = 2,820) with Kenward-Roger degrees of freedom approximation. Reference values are 'coda' for Syllabic Position and 'non-labial' for Left and Right Context.

DV	Fixed effects	*Est.*	*SE*	*t*	*p*
CoG [Hz]	(intercept)	1652	34	48.6	<.001***
	SyllPos: onset	-8	13	-0.6	.52^{NS}
	Left Context: labial	-215	27	-8.0	<.001***
	Right Context: labial	-299	40	-7.6	<.001***
	SyllPos * Right Context	-99	30	-3.3	<.001***
	Left * Right Context	131	51	2.6	<.05**
SD [Hz]	(intercept)	599	14	42.5	<.001***
	SyllPos: onset	-8	4	-1.8	.08^{NS}
	Left Context: labial	19	11	1.7	.08^{NS}
	Right Context: labial	0	21	0	.98^{NS}
	SyllPos * Right Context	-57	11	-5.1	<.001***
	SyllPos * Left Context	-47	10	-4.5	<.001***
Tilt [dB/decade]	(intercept)	-7.8	1.2	-6.3	<.001***
	SyllPos: onset	-0.7	0.4	-1.6	.12^{NS}
	Left Context: labial	-7.4	0.9	-8.6	<.001***
	Right Context: labial	-8.9	1.3	-7.1	<.001***
	SyllPos * Right Context	-4.1	1.0	-4.3	<.001***
	Left * Right Context	4.6	1.7	2.7	<.01***
Amp (normalized)	(intercept)	0.03	0.03	1.0	.32^{NS}
	SyllPos: onset	0.21	0.03	7.1	<.001***
	Left Context: labial	-0.10	0.07	-1.4	.14^{NS}
	Right Context: labial	-0.27	0.07	-3.6	<.001***
	SyllPos * Left Context	0.15	0.07	2.13	<.05*
Dur [log(ms)]	(intercept)	1.92	0.01	212.8	<.001***
	SyllPos: onset	0.01	0.01	1.1	.27^{NS}
	Right Context: labial	-0.02	0.01	-1.2	.24^{NS}
	Right Context * SyllPos	0.09	0.01	6.5	<.001***

Figure captions

FIG. 1. Spectrograms of onset /s/ 340 – 3400 Hz telephone bandwidth from word *cd* (‘cd’, /seˈde/) spoken by a male speaker.

FIG. 2. Spectrogram of onset /x/ in 340 – 3400 Hz telephone bandwidth from word *geen* (‘no’, /xen/) by a male speaker.

FIG. 3. Boxplots for Centre of Gravity (CoG) by fricative sound, syllabic position, and left and right context labialization. The width of the box represents the number of cases.