The effects of linguistic contexts on the acoustics and strength-of-evidence of /s/

Smorenburg, B.J.L.; Heeren, W.F.L.

The effects of linguistic contexts on the acoustics and strength-of-evidence of /s/

Laura Smorenburg and Willemijn Heeren
Leiden University Centre for Linguistics, Leiden, The Netherlands
{b.j.l.smorenburg|w.f.l.heeren}@hum.leidenuniv.nl

Previous research has shown that linguistic structure and phonetic contexts can affect the acoustics and consequently the strength-of-evidence in speaker comparisons. For example, stressed vowels seem to perform better than unstressed vowels [1] and vowels from content words seem to perform slightly better than vowels from function words [although only in multinomial regression, not in likelihood-ratio analysis: 2].

Fricative /s/ is a relatively speaker-specific consonant, but is reported to be strongly affected by coarticulatory labialization, which lengthens the anterior cavity and lowers the resonance frequencies in /s/ [e.g. 3]. Data from Dutch spontaneous telephone speech has shown that slightly more speaker information is available when fricatives /s/ and /x/ occur in these labial contexts [4], which was attributed to between-speaker variation in the degree and timing of the co-articulatory movement. We now investigate the effects of phonetic context and syllabic position on British English /s/, also considering speech channel effects.

Method

Materials consisted of mock telephone conversations with an accomplice taken from Task 2 in WYRED [5]. One 15-min conversation per speaker (*N*=60, all adult males from Wakefield, Yorkshire) was analysed. Per speaker, ~100 /s/ tokens along with their immediate phonetic neighbours were manually segmented and labelled on syllabic position. Spectral moments (M1: centre of gravity, M2: variance, L3: skewness, L4: kurtosis), duration, and spectral tilt were measured for each /s/ in the simultaneously recorded studio and landline telephone channel. M1 was also measured dynamically in 5 non-overlapping windows and captured with a quadratic polynomial fit. Effects of contextual labialization (non-labial, labial) and syllabic position (onset, coda) were assessed with linear mixed-effects (LME) modelling for the acoustics and with multinomial logistic regression (MLR) and MVKD [6] likelihood ratio analysis (LR) for the speaker discrimination. Only speakers with at least 10 tokens per factor level were included in the analysis (*N*=55).

Results

For M1 measured in the studio channel, it can be seen in Table 1 that acoustic results are mostly congruent with the literature [e.g. 3, 4]. There are significant effects of labialization, although anticipatory (i.e. right context) effects are relatively small compared to carry-over (i.e. left context) effects. Coda reduction is also observed. Generally, these effects are not maintained in the telephone channel, rather, they sometimes go in the opposite direction and seem random.

	Studio (550-8000 Hz)			Telephone (550-3400 Hz)		
Effects	*Est.*	*SE*	*t*	*Est.*	*SE*	*t*
(intercept)	5190	77	67.3	2075	32	64.2
Left context = LABIAL	–365	20	–18.7	112	10	10.6
Right context = LABIAL	–94	22	–4.3	–31	12	–2.6
Syll. Position = CODA	–200	15	–13.2	–1	8	–0.1
Left x Syll. Position						
Right x Syll. Position	–118	37	–3.2	68	20	3.4

Table 1. Best-fitting LME model for M1 in the studio and telephone channels (*N*=55, *n*=6634). Hertz ranges refer to the measurement ranges per channel, not the available signal.

Regarding the speaker discrimination, both the MLR and LR analyses showed small contextual sampling effects in the studio, but not (in MLR), or to a lesser extent (in LR), in the telephone channel. However, even in the studio channel, the effect of syllabic position is negligeable and the effect of context labialization not consistent for preceding versus following context (see Figure 1 and Table 2). The effect of speech channel, on the other hand, is much larger for both the acoustics and speaker discrimination. To conclude, contextual sampling effects are present in broadband, but not so much in narrowband signals. It is rather the speech channel that has the largest effects on the acoustics and strength-of-evidence of English /s/.

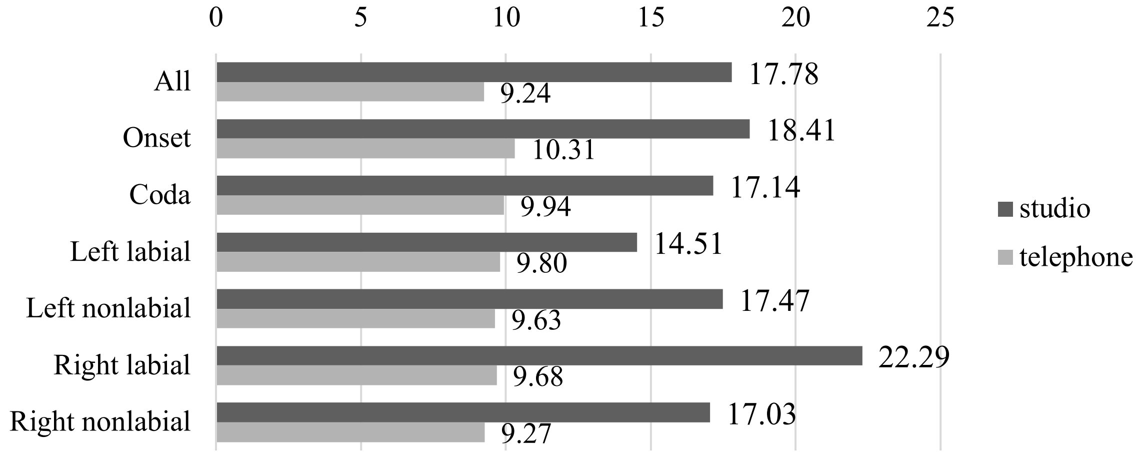

Figure 1. MLR speaker-classification accuracies (in %) using duration, M2, L3, L4, spectral tilt, and linear and quadratic M1 coefficients as predictors. Chance level = 1.82%.

	Studio (550-8000 Hz)					Telephone (550-3400 Hz)				
	LLR^{SS}	LLR^{DS}	Cllr	$Cllr^{min}$	EER	LLR^{SS}	LLR^{DS}	Cllr	$Cllr^{min}$	EER
All	1.77	–2.73	0.52	0.46	16.04	0.76	–0.52	0.82	0.71	24.02
Onset	1.83	–2.86	0.51	0.45	14.09	0.85	–1.02	0.72	0.63	23.12
Coda	1.96	–3.00	0.49	0.45	14.08	1.05	–0.84	0.72	0.64	23.50
Left labial	1.20	–1.16	0.69	0.61	19.31	0.64	–0.30	0.85	0.77	28.43
Left nonlabial	1.75	–2.58	0.53	0.48	16.17	0.67	–0.48	0.80	0.74	24.97
Right labial	1.88	–3.82	0.50	0.36	10.68	0.90	–0.87	0.72	0.66	25.05
Right nonlabial	1.52	–2.27	0.58	0.49	15.47	0.74	–0.51	0.81	0.73	25.36

Table 2. Calibrated LLRs, Cllr, $Cllr^{min}$ and EER. Max. sample size (n=18) per speaker per condition.

References

[1] McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Towards a new approach using formant frequencies. *Int. J. of Speech, Lang. and the Law, 13*(1), 89–125.

[2] Heeren, W. F. L. (2020). The effect of word class on speaker-dependent information in the Standard Dutch vowel /aː/. *J. of the Acoust. Soc. of Am., 148*(4), 2028–2039.

[3] Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward Improved Spectral Measures of /s/: Results from Adolescents. *J. of Speech Lang. and Hearing Res., 56*(4), 1175.

[4] Smorenburg, L., & Heeren, W. (2020). The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. *J. of the Acoust. Soc. of Am.*, *147*(2), 949–960.

[5] Gold, E., Ross, S., & Earnshaw, K. (2018). The “West Yorkshire Regional English Database”: Investigations into the generalizability of reference populations for forensic speaker comparison casework. In Proceedings of *INTERSPEECH* (Vol. 2018–Sept., pp. 2748–2752).

[6] Morrison, G.S. (2007). Matlab implementation of Aitken & Lucy’s (2004) forensic likelihood-ratio software using multivariate-kernel-density estimation.