VARIABILITY IN SWEDISH VOICELESS FRICATIVE CONTRASTS

Carla Wikse Barrow[1], Marcin Włodarczak[1], Mattias Heldner[1] & Sofia Strömbergsson[2]

[1]Department of Linguistics, Stockholm University

[2]Department of Clinical Science, Intervention and Technology, Karolinska Institutet
{carla.w.barrow,wlodarczak,heldner}@ling.su.se, sofia.strombergsson@ki.se

ABSTRACT

Swedish voiceless fricatives vary in their realisation across dialects, as well as across word positions. In the present paper we investigate individual differences in contrast between the voiceless fricatives of twenty adult speakers of Central Standard Swedish. Visual representations and measures of contrast robustness based on spectral moments one and two (spectral centre of gravity and spectral standard deviation) were derived to explore phonetic variation across and within speakers. The described variation in contrast robustness can contribute to a broader understanding of inter- and intra-speaker variation in the realisation of Swedish fricatives.

Keywords: fricatives, acoustics, Swedish

1. INTRODUCTION

Swedish has a relatively rich inventory of voiceless fricatives with five phonemes: labio-dental /f/, alveo-dental sibilant /s/, pre-dorso-alveolar sibilant /ɕ/ (the "tje" sound), velar (or labio-velar) /ɧ/ (the "sje" sound) and glottal /h/ (see e.g. [1, 2, 3]).[1] The fricative system is even denser when allophonic variation in certain dialects is taken into account. The "sje" sound is one of few Swedish consonants that exhibits dialectal variation, with allophones such as prepalatal retroflex sibilant [ʂ̻], velar [ɧ] and velar [x] [2, 4, 3]. In Central Standard Swedish (henceforth CSS), the average speaker is expected to exhibit (at least) two allophones of the "sje" sound, [ʂ̻] and [ɧ], described as light and dark "sje" respectively, based on their perceptual qualities [2, 5]. This allophonic variation is, in part, constrained by phonological distribution, such that light "sje" occurs in postvocalic and intervocalic position and dark "sje" in prevocalic position, except in complex onsets [1]. It has been suggested that some speakers use both retroflex [ʂ̻] and non-retroflex [ʃ], that is, they apply a retroflexion rule where /rs/ sequences are pronounced [ʂ̻] whereas light "sje" is realised as [ʃ] [5]. Nevertheless, most CSS speakers neutralize this distinction, and the final fricative in "kurs" and "dusch" are indistinguishable [1, 3]. For a more comprehensive description of geographical variation, the reader is referred to [2, 4, 3], and references therein.

Static and dynamic acoustic characteristics of word initial CSS fricatives have previously been described in [6], using spectral moments analysis [7]. The results showed that /s-ɕ-ɧ/ were distinguishable through the first spectral moment (M1), on a group level. However, to the best of our knowledge, individual differences with respect to the contrast between the voiceless fricatives of CSS remains unexplored. Moreover, little is known regarding the acoustic properties and variation of light "sje" /ʂ̻/ in modern speakers of CSS. In this paper, we visualize and investigate fricative contrast in twenty speakers of CSS. We include both word initial fricatives [f, s, ɕ, ɧ] and word medial and -final [ʂ̻].

In addition, given that for the English sibilants /s-ʃ/ female speakers were reported to produce slightly larger contrast compared to male speakers [8, 9], we also investigate gender differences. Different measures of contrast robustness have been proposed (see e.g. [9]). In the current paper, we use an overall discriminability measure based on M1 that takes into account distance between fricative categories as well as dispersion within categories [8].

2. METHOD

2.1. Participants

Twenty native speakers of Swedish (10 female) between 18 and 43 years (M = 29.55, SD = 7.59) from the Svealand region took part in a speech production task. All participants provided written informed consent prior to recordings and were compensated for their participation with one movie ticket.

2.2. Task and materials

The speech production task included 28 fricative initial words (see Table 1), two fricative medial words ([kɔʂa] "to cross", [dɵʂa] "to shower") and four fricative final words ([kɔʂ] "cross", [dɵʂ] "shower", [lɑːʂ] "Lars", [garɑːʂ] "garage") that were repeated four times each, resulting in 136 target words per speaker. The fricative initial words were presented on a screen in writing as well as through headphones, whereas the fricative medial and final words were presented only in writing. The audio prompts were recorded by a female native speaker of Swedish from the Svealand region. The participants were instructed to speak clearly without hyperarticulating.

2.3. Recording

Participants were recorded in the aneachoic chamber in the phonetics lab of Stockholm University using a Brüel & Kjær Type 4189-L-001 microphone and a Motu 8M sound interface. A 44.1 kHz sampling rate and 16 bit quantization were used throughout, although the sounds were downsampled to 22 kHz prior to analysis.

2.4. Annotation and segmentation

The soundfiles were annotated in Praat [10]. The onset of the word initial fricatives was marked at a rapid increase in zero crossing and/or high frequency energy, and the offset was marked at the zero crossing prior to the onset of periodic energy of the following vowel. The segmentation procedure of the word medial and word final fricatives was less detailed. For each fricative, a fricative region of relative stability was determined. The onset and offset of this region were marked conservatively, leaving the less stable edges of the fricative segment outside the boundaries. Clear intrusions of voicing and bursts in the fricative noise were excluded from the fricative region. Therefore, fricative duration and overall intensity cannot reliably be derived from these segments and are not used in the analysis. However, as M1 has been shown to be relatively stable in the central part of voiceless fricatives [6], we believe this segmentation is sufficient to derive a representative spectral centre of gravity. All boundaries were determined through simultaneous inspection of the waveform and spectrogram. A dynamic range of 40 dB and a spectral range of 0-8 kHz were used throughout.

Recordings including disturbances, such as clearing of the throat, or background noise that affected the fricative were removed prior to analysis, leaving a total of 2671 tokens.

2.5. Feature extraction and analysis

A Praat script extracted spectral moments 1 and 2 (centre of gravity and standard deviation) from a 30 ms window Hann Window at centered at 50% of the fricative/fricative region. The spectra were filtered through a Hann pass filter (200-11000 Hz, with 100 Hz smoothing) prior to feature extraction.

The light "sje" showed significant overlap between retroflex [ʂ] and non-retroflex [ʃ] (mean M1 was 3170 Hz (SD= 515) and 3285 Hz (SD= 585), respectively), and are therefore grouped together in the plots and further analyses.

The acoustic features were imported and summarized in R [11]. The phoneme discriminability measure described in [8] was calculated for the fricatives pairs /s-ɕ/, /ɕ-ʂ/ for all speakers. The discriminability measure was calculated as the distance between fricative categories (i.e. difference between the mean M1 for each category) divided by the square root of the mean of the M1 variances of the two categories; $(\mu_{f_1} - \mu_{f_2})/\sqrt{(Var_{f_1} + Var_{f_2})/2}$, were f_1 denotes the first fricative and f_2 the second fricative.

3. RESULTS

Figure 1 shows M1-M2 plots for each speaker. Visual inspection of the plots reveals that /ɧ/, /ɕ/ and /s/ are well separated for all speakers, although some exhibit slight overlap in M1 for /ɕ/ and /s/ (see e.g. the speaker in row five; column three). Both distance between and dispersion within categories varies across speakers and gender, compare for example the speakers in the first three rows in the first column. Velar /ɧ/ is most compact in M1, while category dispersion varies across individuals with respect to the sibilants, although /s/ is often more disperse than /ɕ/.

Labio-dental /f/ is distinguishable for most participants if M2 is included, although the sound is generally most disperse in M1.

For most speakers (e.g. individuals in column two) /ʂ/ overlaps entirely or partly with /ɕ/ in the M1-M2 dimension. However, some individuals show clear separation between the two sounds (e.g. the speakers in row one; column three and row three; column five). For individuals who show separation, [ʂ] has a slightly lower spectral centre of gravity than [ɕ].

To investigate robustness of contrast, a measure of overall discriminability was calculated for the

Vowel context	/f/	/s/	/ɕ/	/ɧ/
[a]	"fall" [fal]	"juice" [saft]	"squabble" [ɕafs]	"chess" [ɧak]
[ɑː]	"plate" [fɑːt]	"auditorium" [sɑːl]	"nagging" [ɕɑːt]	"scarf" [ɧɑːl]
[iː]	"kefir" [fiːl]	"sieve" [siːl]	"wedge" [ɕiːl]	"ski" [ɧiːda]
[eː]	"wrong" [feːl]	"zebra" [seːbra]	"chain" [ɕeːdja]	"spoon" [ɧeːd]
[ɛ]	"five" [fɛm]	"cell" [sɛl]	"cane" [ɕɛp]	"ship" [ɧɛp]
[uː]	"foot" [fuːt]	"sun" [suːl]	"skirt" [ɕuːl]	"lump" [ɧuːk]
[ʉː]	"ugly" [fʉːl]	"sour" [sʉːɹ]	"bull" [ɕʉːɹ]	"sick" [ɧʉːk]

Table 1: The fricative initial targets words used in the repetition task. The words' English gloss is written in quotes and a transcription in brackets.

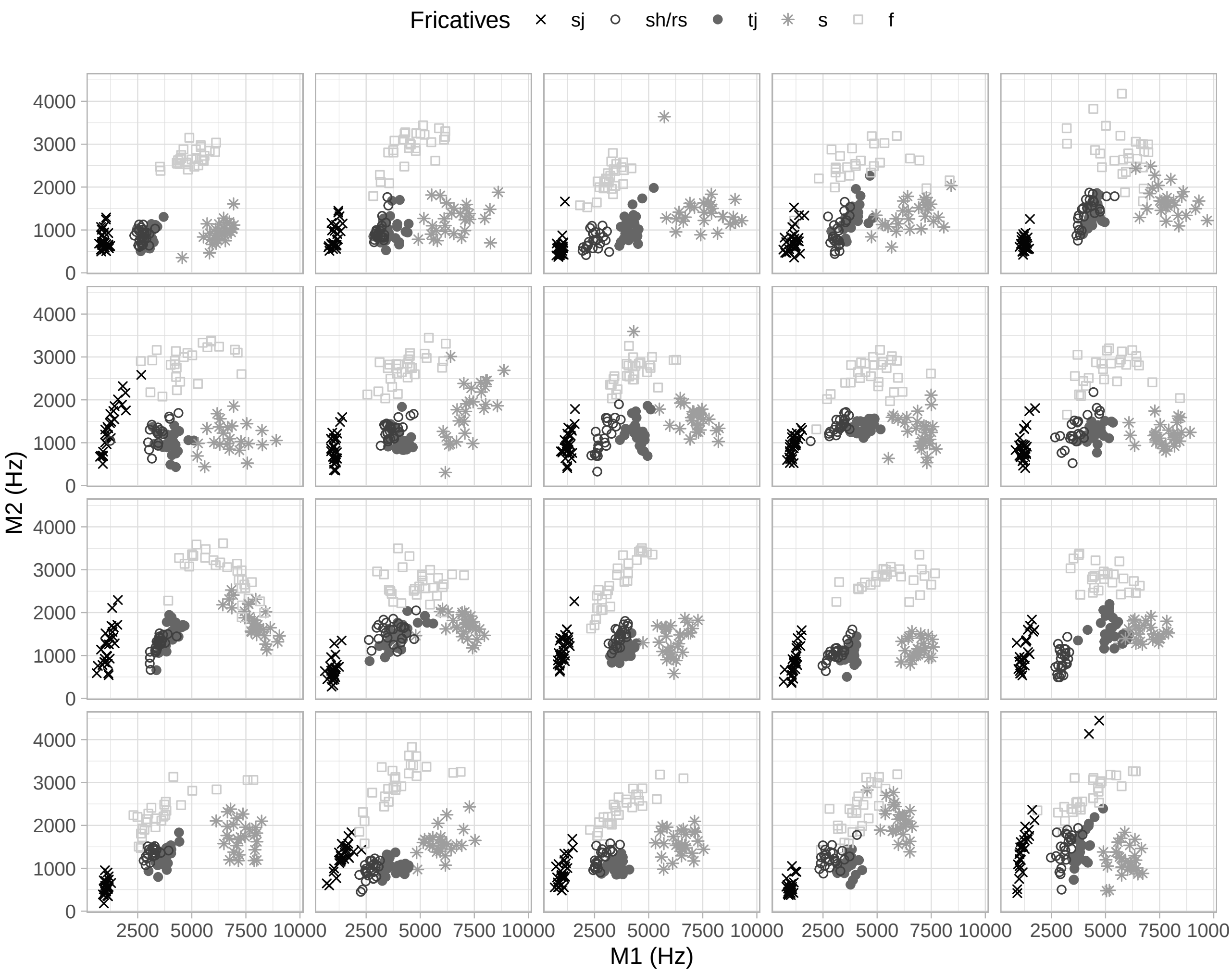

Figure 1: Scatterplot of M1 (center of gravity) and M2 (standard deviation) for Swedish /f/, /s/, /ɕ/ (tj), /ʂ/ (sh/rs) and /ɧ/ (sj), measured at the center of the fricative noise for all speakers. The top two rows show female speakers and the bottom rows show male speakers.

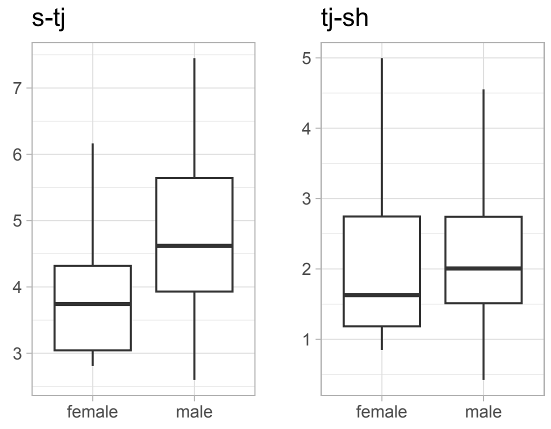

Figure 2: Discriminability in female and male speakers for the fricative pairs /s-ɕ/ (s-tj) and /ɕ-ʂ/(tj-sh)

sibilant pairs [s-ɕ] and [ɕ-ʂ]. Figure 2 shows boxplots of robustness of contrast for female and male speakers. As evident in the figure, male speakers have higher discriminability scores for both fricatives pairs, although the difference is small for [ɕ-ʂ] and the variation substantial for [s-ɕ].

4. DISCUSSION

This study is an acoustic description of Swedish voiceless fricative contrast, that highlights individual differences in variability and fricative discriminability. The results uncover differences in fricative contrast across individuals, both with regards to the dispersion and distance between the voiceless fricatives of SCC. Exploring individual differences is important as patterns which are not evident on a group level emerge. Previous descriptions (see [6]) indicate that Swedish [f] is not readily captured by M1 on a group level, and spans across the /s-ɧ/ distribution of M1 with some overlap in M2 as well. For many speakers in the present investigation, [f] is clearly distinguishable, especially if M2 is included. With respect to the light "sje", [ʂ] is separated from [ɕ] by M1 for some speakers, but most show substantial overlap with the "tje" sound in the M1/M2 dimension.

The discriminability measure shows differences across individuals with respect to contrast robustness in the sibilant fricatives. The gender-related differences described in [8], such that female speakers produced more distinctive categories than male speakers, do not seem to hold for the individuals in this study. In fact, contrary to previous findings, male speakers in the current study showed slightly higher levels of discriminability as compared to females, although the within-group variation was substantial. It should be noted that differences in filtering of the spectra affect the spectral moments calculated thereof, and although Swedish sibilants [s-ɕ] are similar to English [s-ʃ], the investigations in [9, 8] are not directly comparable to this study.

4.1. Limitations and future research

The current study is limited as it involves only two spectral parameters to describe voiceless fricatives of CSS. Including other acoustic cues related to the fricatives spectral representation as well as temporal and intensity properties, and their change over time, may provide insight into patterns that are not evident here.

Moreover, a larger sample including more words and connected speech would be beneficial. Exploring the contrast between light "sje" and the "tje" sound (/ɕ/) in dialects that primarily use the fronter [ʂ] (i.e. Northern Swedish dialects) as well as dialects with affricative realisation of tje (i.e. [tɕ] in Finland Swedish) [2, 3] would also be of interest, to capture variation in contrast and structure of voiceless fricative systems.

Finally, a developmental perspective on fricative variability and contrast would be valuable for research on acquisition of Swedish. The results presented here provide insight into the degree and type of variation that can be expected in adult productions of voiceless fricatives, which are the targets that child productions are often compared to.

5. REFERENCES

[1] T. Riad, *The phonology of Swedish*. Oxford: Oxford University Press, 2014.

[2] P. Lindblad, "Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv," Ph.D. dissertation, Lund University, 1980.

[3] G. Bruce, *Vår fonetiska geografi : om svenskans accenter, melodi och uttal*, 1st ed. Lund: Studentlitteratur, 2010.

[4] R. Shosted, "Acoustic characteristics of Swedish dorsal fricatives," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3888, 2008.

[5] C. Garlén, *Svenskans fonologi : [i kontrastiv och typologisk belysning]*, 1st ed. Lund: Studentlitteratur, 1988.

[6] C. Wikse Barrow, M. Włodarczak, L. Thörn, and M. Heldner, "Static and dynamic spectral characteristics of swedish voiceless fricatives," *The Journal of the Acoustical Society of America*, vol. 152, no. 5, pp. 2588–2600, 2022.

[7] K. Forrest, G. Weismer, P. Milenkovic, and R. N. Dougall, "Statistical analysis of word-

initial voiceless obstruents: preliminary data," *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 115–123, 1988.

[8] R. Romeo, V. Hazan, and M. Pettinato, "Developmental and gender-related trends of intra-talker variability in consonant production," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3781–3792, 2013.

[9] J. J. Holliday, P. F. Reidy, M. E. Beckman, and J. Edwards, "Quantifying the robustness of the english sibilant fricative contrast in children," *Journal of Speech, Language, and Hearing Research*, vol. 58, no. 3, pp. 622–637, 2015.

[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]." 1994-2021. [Online]. Available: https://praat.org

[11] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/

[12] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263, 2000.

---

[1] However, /h/ is sometimes described as an approximant [5] and often excluded from acoustic descriptions of fricatives (see e.g. [12]).