The effects of intonation on acoustic properties of fricatives

Marzena Żygis[1], Daniel Pape[2], Luis M.T. Jesus[2,3], Marek Jaskuła[4]

[1] Centre for General Linguistics (ZAS) & Humboldt University, Berlin, Germany
[2]IEETA, University of Aveiro, Portugal
[3]ESSUA, University of Aveiro, Portugal
[4] West Pomeranian University of Technology, Szczecin, Poland
zygis@zas.gwz-berlin.de,danielpape@ua.pt,lmtj@ua.pt,Marek.Jaskula@zut.edu.pl

ABSTRACT

This study investigates acoustic properties of the Polish retroflex and alveolo-palatal fricatives and affricates /ʂ/, /t͡ʂ/, /ɕ/ and /t͡ɕ/. The sibilants were produced in (i) yes/no questions with rising intonation and (ii) answers with falling intonation. Multitaper spectra were used to compare the two places of articulation. Results show that the Centre of Gravity (COG) did not differentiate retroflexes from alveolo-palatals in the question condition. The same consonants, however, display significantly different COG values if they are produced in the statement condition. Other spectral measurements (highest spectral peak, standard deviation, kurtosis) were also significantly different with regard to the two places of articulation and intonation types.

The results indicate that special attention should be given to intonational differences when acoustic properties of sibilants are investigated.

Keywords: segment-prosody interaction, sibilants, intonation

1. INTRODUCTION

Much attention has been devoted to the investigation of spectral properties of fricatives in general, and sibilants in particular. Among various parameters, the spectral centre of gravity (COG) is probably the most frequent parameter used in the acoustic analysis of fricatives [1], [2], and successfully distinguishes sibilants produced at different places of articulation [3].

At the same time, studies which focus on spectral properties of fricatives have not taken into consideration intonational properties of the words/sentences in which the segments were embedded, following a still common tendency to investigate segments and intonation separately.

More recently, however, some evidence was provided from the perspective of intonational research, according to which intonation patterns are not only encoded in vocalic but also in consonantal elements of speech. In particular, it was shown that sibilants contribute to differences in intonation by changing their spectral properties, e.g. higher COG was found for rising and lower COG for falling intonation [4], [5], [6].

The present paper seeks to provide an answer to the question of how spectral properties of sibilants are affected by intonational properties. Although linked in a sense to previous studies, the present investigation has a different aim. It does not analyse the contribution of sibilants to intonation patterns, but rather makes a point about the role of intonation on the spectral properties of sibilants. The main aim of the study is, thus, to show that spectral properties of sibilants – classically seen as invariant properties – might be seriously affected, and to some extent even blurred if intonation is not controlled for.

2.METHODS

2.1. Corpus

In order to test the influence of intonation on spectral properties of sibilants, we investigated Polish phonemic retroflex and alveolo-palatal sibilant clusters consisting of (i) retroflex fricatives and affricates /ʂt͡ʂ/ and (ii) alveolo-palatal fricatives and affricates /ɕt͡ɕ/. The clusters always appeared in word-final position in sixteen words (eight words for each cluster). The words were monosyllabic and occurred in the sentence-final position which in turn differed in intonation contour: (a) rising intonation in yes/no questions and (b) falling intonation in statements:

a) Widzi tę kość [kɔɕt͡ɕ]?
Does he see this bone?
b) Widzi tę kość [kɔɕt͡ɕ].
He sees this bone.

Eight male native speakers of Standard Polish read all sentences three times. Sentences with both cluster types (divided into two sets) were randomised together with sentences containing bisyllabic words. In total we obtained 1536 items to analyse (16 words × 2 sentence types [statement, question] × 2 sibilant types [fricative, affricate] × 3 repetitions x 8 speakers).

All recordings were conducted in a sound-proof room, using a TLM103 Neumann microphone connected to a ProTools system with a Digi 003 interface (sample rate 44100 Hz). The items were analysed with PRAAT (version 5.3.57) and MATLAB (version R2007b).

2.2. Measurements

The following measurements were taken from the multitaper spectra for both fricatives and affricates:

- frequency of the highest spectral amplitude peak of the frication noise in the range from 20 Hz to 5 kHz (we excluded frequencies above 5kHz to concentrate on the main spectral peak around 3kHz, see figure 1)
- spectral Centre of Gravity (COG, first spectral moment),
- standard deviation (SD, second spectral moment)
- skewness (third spectral moment),
- kurtosis (fourth spectral moment),
- duration of the fricative and the fricative part of the affricate.

Apart from duration, all parameters were measured at the acoustical midpoint of each fricative/affricate which was defined as the equal temporal distance between the onset and the offset of the frication noise (which also defined the target phoneme duration). At this acoustical midpoint multitaper spectra with a 12 ms window (512 point Hamming window) were computed. The power spectral density (PSD) was estimated via the Thomson multitaper method (linear combination with unity weights of individual spectral estimates and the default FFT length) available in the MathWorks Signal Processing Toolbox Version 6 [7].

2.3. Statistics

Linear mixed effect models were employed for the investigated variables (highest spectral peak, COG, SD, skewness, kurtosis, duration) which were studied as effects of INTONATION (rising vs. falling), PLACE (retroflex, alveolo-palatal) and MANNER (fricative, affricate) as well as the interaction of INTONATION and PLACE. ITEM and SPEAKER were taken as random effects. This model was chosen due to the best fit (lowest AIC values); it was significantly different from models with other interactions. All p-values reported in the paper are pMCMC values.

The statistical analyses were conducted in the R environment software (version 3.0.2) [8].

3. RESULTS

First of all, the results confirm that questions were indeed produced with rising intonation: the mean F0 values at the beginning and at the end of the vowel preceding the clusters were 138 Hz (sd: 26Hz) and 225 Hz (sd: 50 Hz), respectively. In statements, the corresponding mean values were 122 Hz (sd: 22 Hz) and 108 Hz (sd: 20 Hz) respectively, thus proving the falling intonation in statements.

Figure 1 and 2 show multitaper spectra for all retroflex and alveolo-palatal fricatives split by intonation condition (question/statement). In addition, the overlaid mean spectrum is shown and the spectral regression lines with the endpoint/startpoint $\overline{F}$ at 3 kHz are displayed [9].

Figure 1: Multitaper spectra for all ***retroflex fricatives*** (light colour) with superimposed mean spectrum (black lines) measured at the acoustic midpoint of the fricative. Dashed lines show the spectral regression lines.

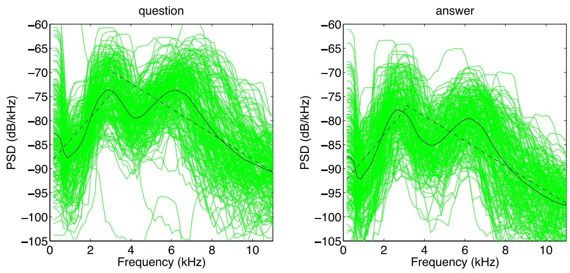
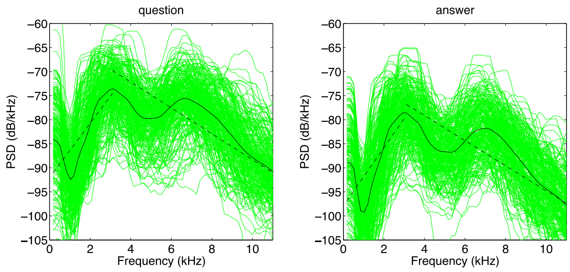


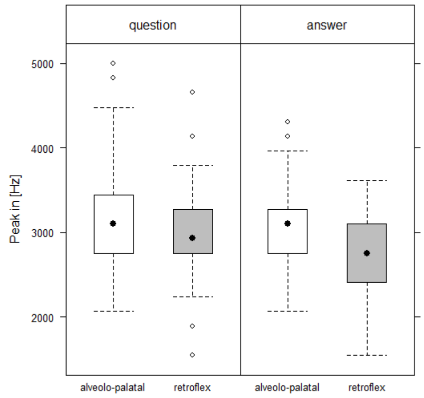
Figure 2: Multitaper spectra for all ***alveolo-palatal fricatives*** (light colour) with superimposed mean spectrum (black lines) measured at the acoustic midpoint of the fricative. Dashed lines show the regression lines.



It can be seen that in addition to the typical broad spectral peak corresponding to retroflex and alveolo-palatal place of articulation (around 3kHz) there is an additional prominent spectral peak around 7kHz, which in the retroflex question condition has even higher amplitudes as the 3kHz peak.

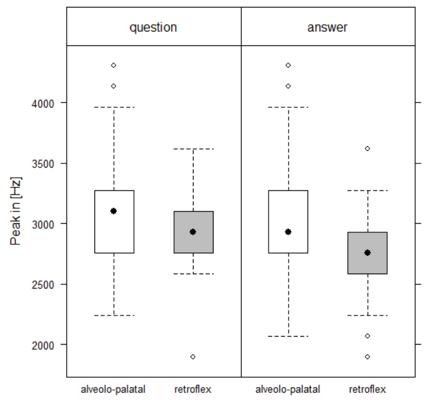
Statistical comparison of the highest spectral peak values points to significantly lower peaks in retroflexes than in alveolo-palatals in both questions (t= -2.075 p<.05) and statements (t= -2.872 p<.01). This is shown in Figure 3.

Figure 3: The highest spectral peak of fricatives



In affricates, the highest spectral peak was significantly lower in retroflexes than in alveolo-palatals when segments were produced in both question (t=-2.963 p<.001) and statement (t=-3.809 p<.001) condition. Figure 4 illustrates the results.

Figure 4: The highest spectral peak of affricates



With respect to COG in fricatives, our results indicate that there was no statistically significant difference between retroflex and alveolo-palatal fricatives when they were produced with rising intonation in questions (t=1.31). The significant difference was, however, observed in statements where the COG values were higher in alveolo-palatals than in retroflexes (t=-2.73 p<.05), as shown in Figure 5.

As far as affricates are concerned neither in questions nor in statements a significant difference in COG values was found (questions: t=1.31, statements t=-1.35). The results are shown in Figure 6.

Figure 5: Centre of Gravity of fricatives

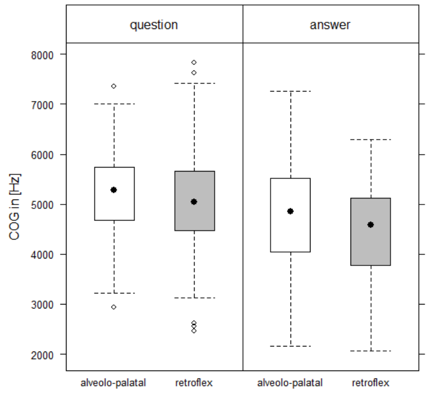
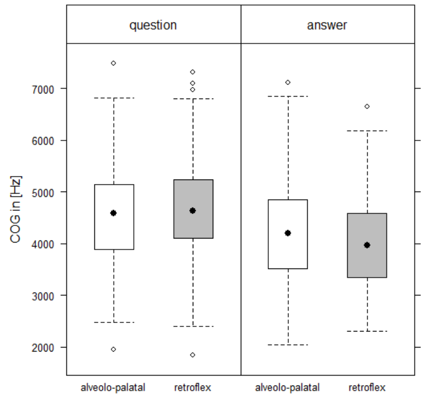
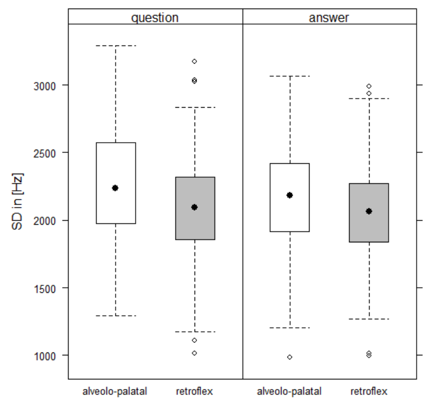


Figure 6: Centre of Gravity of affricates



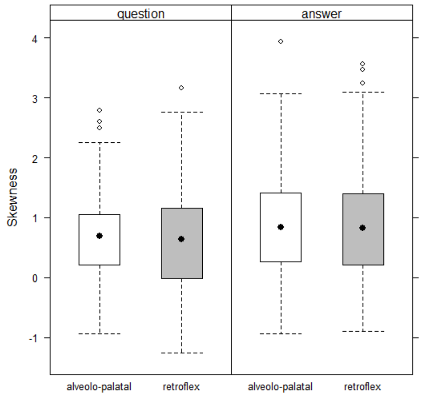
Standard deviation, the second spectral moment, was significantly higher in alveolo-palatals than in retroflexes in questions and in statements for both fricatives (questions: t= 4.745 p<.001; statements: t=3.814 p<.001) and affricates (questions: t=3.71 p<.001; statements: t= 8.412 p<.001) (see Figure 7 for fricatives).

Figure 7: Standard deviation of fricatives.



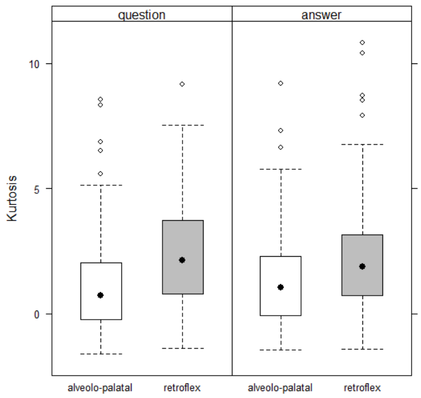
Skewness, the third spectral moment, differentiated retroflex from alveolo-palatal fricatives neither in questions (t=-0.694) nor in answers (t= -0.065). This is shown in Figure 8. For the affricates the skewness was significantly lower in retroflexes than in alveolo-palatals in questions (t=-5.607 p<.001) but not in answers (t=-1.893).

Figure 8: Skewness of fricatives.



Kurtosis, the fourth spectral moment, which is an indicator of the overall broadness of the spectral peak, was higher in retroflex fricatives than in alveolo-palatal ones in both questions (t=5.640 p<.001) and answers (t= 4.947 p<.001), as shown in Figure 9. Similarly, in affricates, kurtosis was higher in retroflexes than in alveolo-palatals in questions (t= 6.234 p<.001) but not in answers (t=1.897).

Figure 9: Kurtosis of fricatives



Comparison of the phoneme durations showed that retroflex fricatives were significantly longer than alveolo-palatal ones in questions (t= 3.396 p<.01) but not in answers (t=1.706). Frication duration of affricates was shorter in retroflexes than in alveolo-palatals when both answers (t=-12.40 p<.001) and questions were produced, (t=-13.01 p<.001) were produced.

4. DISCUSSION AND CONCLUSIONS

Previous research has shown that different types of intonation contours are encoded in consonants, leading to the hypothesis that they help to convey the intended intonation [10].

For fricative (and other noise) spectra, the main low-frequency spectral prominence ($\bar{F}$) directly codes the fricative place of articulation. For our data, values around 3kHz are found for both alveolo-palatals and retroflexes. However, it is not clear what process generates the clearly visible prominent spectral peak around 7kHz (see Figure 1 and Figure 2), as above 5kHz nonplanar modes and complex nonlinear effects are dominant in the noise spectra ([9]:445). It is due to this second prominent peak, however, that the COG values are drawn towards higher frequencies, resulting in higher COG values for the retroflex question condition than for the alveolo-palatal question condition. In other words, the COG values in *questions* do not code place of articulation of the two Polish phonemes because the prominent high-frequency peak manipulates the COG values using different magnitudes for retroflexes compared to alveolo-palatals. For the *statement* condition this effect of the high-frequency peak is less pronounced, thus resulting in the observed and expected significance in COG values between alveolo-palatal and retroflex *fricatives*. With respect to the non-significance comparing the place of articulation in *affricates*, our assumption is that the expected COG differences are cancelled due to their sentence-final position known for contrast neutralisation. Therefore, for Polish sibilants the COG measurement clearly is not appropriate to code the difference in place of articulation (at least not in question condition).

The present acoustic investigation shows that different intonation patterns can significantly influence the acoustic properties of sibilants. What the results clearly suggest is that special attention should be given to intonational differences when acoustic properties of fricatives, especially COG, are under investigation. Different intonation contour types might lead to different results, in our case the non-significance of the COG values between two different places of articulation in Polish.

Finally, the results point to a tight link between segments and prosody and their mutual interaction, a point which undoubtedly deserves more attention.

Acknowledgments

This work was partially funded by National Funds through FCT - Foundation for Science and Technology, in the context of the projects UID/CEC/ 00127/2013, Incentivo/EEI/UI0127/2014 and .the post-doctoral fellowship SFRH/BPD/ 48002/2008. This research has also been supported by Bundesministerium für Bildung und Forschung (BMBF, Germany) Grant Nr. 01UG1411 to Marzena Żygis.

7. REFERENCES

[1] Jassem, W. 1979. Classification of fricative spectra using statistical discriminant functions. In: Lindblom, B., Öhman, S. (eds.), *Frontiers of Speech Communication Research*. New York: Academic Press, 77-91.

[2] Stevens, K. N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

[3] Gordon, M., Barthmaier, P., Sands, K. 2002. A cross-linguistic acoustic study of voiceless fricatives. *JIPA* 32, 141-171.

[4] Niebuhr, O. 2012. At the edge of intonation: The interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica* 69, 7-27.

[5] Ritter, S., Röttger, T. 2014. Speakers modulate noise-induced pitch according to intonational context. *Proceedings of 7th International Conference on Speech Prosody*, 890-894.

[6] Żygis, M., Pape, D., Jesus, Luis M.T, Jaskula, M. 2014. Intended intonation of statement and polar questions in Polish whispered, semi-whipered and normal speech modes. *Proceedings of the 7th International Conference on Speech Prosody*, 678-682.

[7] MathWorks (ed.) 2007. Signal Processing Toolbox 6 User's Guide. Natick: MathWorks.

[8] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, http://www.R-project.org/, version 3.0.2.

[9] Jesus, L. M. T., Shadle, C. 2002. A parametric study of the spectral characteristics of fricatives. *Journal of Phonetics* 30, 437-464.

[10] Niebuhr, O. 2008. Coding of intonational meaning beyond F0: Evidence from utterance-final /t/ aspiration in German. In: *Journal of the Acoustical Society of America* 124, 1252-1263.