The puzzle of two major spectral peaks in Polish sibilants: Acoustic analysis and articulatory synthesis

Daniel Pape (McMaster University, Hamilton, Canada) & Marzena Żygis (Centre for General Linguistics & Humboldt University, Berlin, Germany)

Sibilants with intermediate place of articulation (alveolars, alveolo-palatals and postalveolars) are generally characterized by one broad spectral peak in the mid-frequency region between 3-7kHz. This main peak is attributed to front cavity resonance (Stevens 1998, Shadle et al. 2009) whereas the back cavity (i.e. posterior to the constriction) is assumed to be acoustically inactive (Stevens 1998). The center frequency of the prominent spectral peak generally codes how fronted the sibilant is produced, with more fronted place of articulation resulting in higher frequency spectral peaks.

In order to examine how the spectral characteristics of sibilants are affected by both prosodic contrast and place of articulation we conducted a production experiment with 16 Polish speakers (8 males, 8435 items). We recorded Polish retroflex and alveolo-palatal fricatives in monosyllabic words embedded in frame sentences. We compared two prosodic conditions by producing the fricatives in either polar question or statement condition.

Figure 1 presents the mean multitaper spectra over all speakers split by prosodic condition and phoneme. For all conditions the expected major spectral peak occurs between 3kHz and 4kHz. However, all four conditions show an unexpected, strong, additional second major peak around 7kHz, with higher amplitude for the question condition. This second peak displays often the same amplitude as the first main spectral peak[1] and thus induces substantial acoustic energy in the higher frequency regions.

We pose two working hypotheses for the occurrence of the strong second peak:

1) *Front cavity hypothesis*: The second peak is assumed to be a higher resonance of the dominant front cavity, with the first major peak being the first resonance of this cavity. It has to be noted that for Polish retroflex and alveolo-palatal fricatives the sublingual cavity is actually the front cavity (see MRI images of Toda et al. 2005: 358).

2) *Additional cavities hypothesis*: The second peak is generated by resonances or additional modes in other cavities, either (1) the cavity between the teeth and the border of the lips, or (2) the back cavity behind the constriction. For example, Shadle (2009) found that both lip spreading and front cavity size varies significantly between speakers but only lip spreading (but not front cavity size) introduced spectral shape changes in postalveolar American fricatives.

To examine the probabilities of the two hypotheses outlined above, we performed a series of exploratory pilot studies. We hypothesized that if the *front cavity hypothesis* is correct, then there should be a strong correlation between the first and the second spectral peak, since the first peak is the first (main) resonance and the second peak would be a multiple of the first peak frequency. To test this hypothesis we plotted all pairs of first against second peak in a heatmap (i.e. a two-dimensional map where the number of occurrences for each pair are coded by color intensity) for all our collected Polish fricatives. Figure 2 shows these heatmaps split by phoneme identity with overlaid correlation coefficients. As can be seen, both the figure plots and the coefficient provide no evidence for a strong correlation between the first peak and the second peak; it rather seems that the frequency of the second peak is independent from the frequency of the first peak. Therefore, we conclude that the *front cavity hypothesis* should be rejected.

Prior to testing the second hypotheses we conducted a pilot study with the German postalveolar fricative (5 speakers, 164 items) to determine if a second major peak would be found for sibilants of other languages as well. As can be seen in the mean of the multitaper spectra of the German real speech data in figure 3 (left panel) there is no trace of a second peak and the spectrum only shows the prominent major (first) peak previously reported in the literature (Stevens 1998).

We then tested the *additional cavities hypothesis* by performing a series of articulatory simulations with the articulatory synthesizer VocalTractLab (version 2.1). This synthesizer has been shown to successfully model fricatives and their sophisticated noise sources (Birkholz 2014). Its articulatory targets are based on real MRI data of all German phonemes[2]. We started by using the software default for the (German) postalveolar fricative place of articulation as a baseline[3] and then successively varied the participating articulators (for fricative production) with the aim to induce a prominent second spectral peak. In figure 3 (middle panel) we show the effect of strongly protruded lips versus normal lip position on the spectra of the simulated Vocaltractlab articulations. Finally, the right panel of figure 3 shows the effect of increasing and narrowing the back cavity in the articulatory synthesizer. The comparison of middle and right panel in figure 3 shows that only the protrusion of the lips introduces an additional high-frequency spectral peak in the region around 5-6kHz. We therefore conclude that the existence of the second major peak in Polish retroflexes and alveolo-palatals can be attributed to a strong articulatory lip protrusion. In contrast, the mean spectrum of the German real speech data (see left panel of figure 3) is comparable to the neutral lip position condition in the middle panel of figure 3. In conclusion, it is claimed that Polish retroflex and alveolo-palatal sibilants are produced with extensive lip

---

[1] We calculated how many phonemes are produced with two peaks with similar amplitude (±6dB) and found that 75% of all Polish retroflexes and 84% of all Polish alveolo-palatals show this behavior.

[2] In other words, articulatory settings from other languages–in our case ideally Polish places of articulation–are not available.

[3] It has to be noted that one limitation of the software is that it only allows synthesis up to 6kHz, so we can only partly model the interesting frequency range of 5-8kHz.

protrusion, whereas the German postalveolar sibilant is produced with a rather neutral lip protrusion. Furthermore, our data and analyses show that the traditional approach of using *centre of gravity* to code the sibilant place of articulation (and thus ignoring the second major peak) is highly problematic for Polish fricatives and should be used with caution.

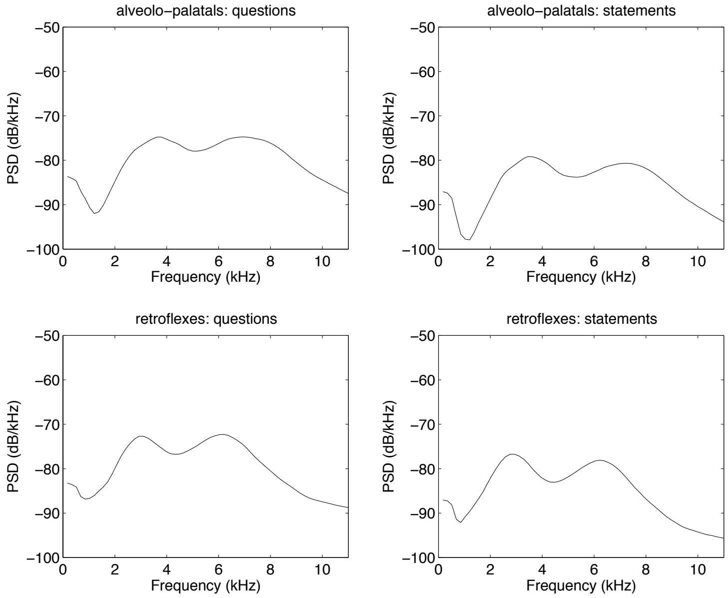

Figure1: Mean spectra of Polish alveolo-palatals (upper panels) and retroflexes (lower panels), split by prosodic condition.

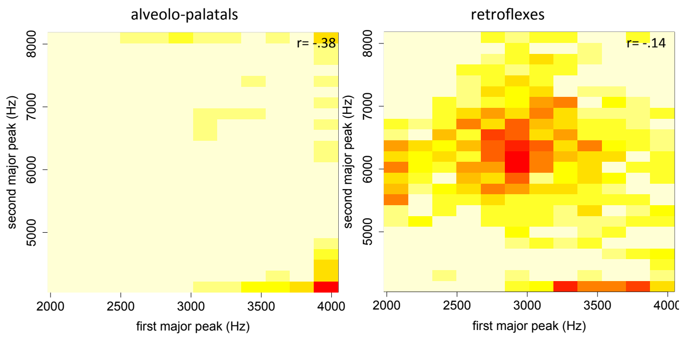

Figure 2: Heatmaps of all Polish alveolo-palatals (left) and retroflexes (right) for corresponding pairs of the first major spectral peak (x-axis) and the second major spectral peak (y-axis). The correlation coefficients between the two dimensions are printed on top of the corresponding figure. For each given x-y pair, darker colors mean higher item counts (more occurrences).

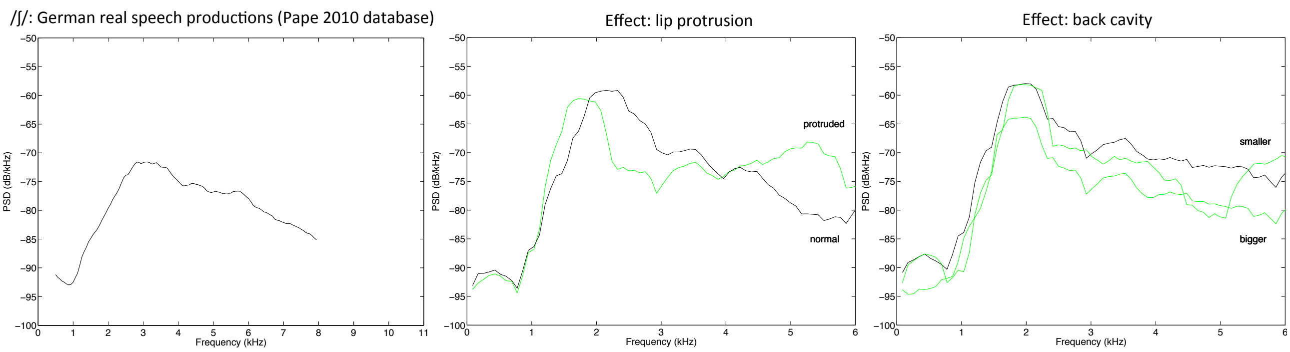

Figure 3: Spectra of German postalveolar fricative: The left panel shows the mean multitaper spectrum of real speech productions (left panel). The middle panels shows multitaper spectra of synthesized simulations (Vocaltractlab) with normal (black) and protruded (green) lips. The right panel presents multitaper spectra of synthesized simulations (Vocaltractlab) of the standard back cavity size (black) in comparison to decreased and increased back cavity size (green color).

References:

Birkholz, P. (2014). "Enhanced area functions for noise source modeling in the vocal tract" *Proc. of the 10th ISSP*, Cologne.

Shadle, C., Proctor, M., Berezina M. & Iskarous, K. (2009). "Revisiting the role of the sublingual cavity in the /s/ - /ʃ/ distinction", *157th Meeting of the ASA*, Portland, USA.

Stevens, K. N. (1998). *Acoustic phonetics*, MIT Cambridge.

Toda, M., Shinji, M., & Kiyoshi, H. (2005). "Formant-cavity affiliation in sibilant fricatives". In Fuchs, S., Toda, M. & Zygis, M. (eds.), *Turbulent sounds: An interdisciplinary guide*. Berlin: de Gruyter. 343-374.