Perceptual stability of sibilants undergoing acoustic variation: Interplay between acoustic processing versus influences of articulatory and/or motor patterns

Daniel Pape[1]

[1]*McMaster University (Canada)*
paped@mcmaster.ca

One of the most interesting, long debated and classical questions in phonetics/phonology is whether phoneme perception is based on apparent (surface) acoustic forms or rather underlying representations, be it motor patterns or articulatory gestures, or a combination of both. For this project, the idea was tested that manipulating the surface forms of the two underlying sibilants /s ʃ/ into their *opposite* acoustic form and spectral shape[1] would either generate perceptual identification differences for listeners (thus pointing to the importance of underlying articulatory differences) or not (pointing to the sole importance of the presented acoustic signal without taking into account underlying articulatory information during the perception process). In other words, if an underlying /ʃ/ sibilant is acoustically changed so its new spectral shape is completely identical to the spectrum of /s/, does this acoustic manipulation generate perceptual differences in dependence of whether the original *articulatory* shape was either /s/ or /ʃ/, although the perceptually judged *acoustic* stimulus is identical?

To generate the stimuli, we recorded a number of prototypical [s] and [ʃ] items in /aCa/ context from 4 female Canadian English speakers. First, stimuli were high-pass filtered above 1000Hz to exclude possible influences of irrelevant low frequency components (with respect to sibilant perception). Then we manipulated the acoustic signals (see figure 1 left panel for example) in a stepwise manner until the [s] spectral shape was acoustically completely identical to [ʃ] (i.e. amplifying frequencies stepwise below 3000 Hz, and attenuating frequencies above 3000 Hz). We used the same procedure to change the [ʃ] spectral shape to /s/ (i.e. amplifying frequencies above 3000Hz and attenuating below 3000Hz, also in a stepwise manner). As a result of these manipulations we generated a continuum of perceptual stimuli /s ʃ/ that were (1) identical in *acoustic* spectral shape distributions (see figure 1 right panels) but (2) significantly differed in the underlying original *articulatory* configuration and thus motor configurations. If these acoustically identical, but underlyingly articulatorily different stimuli would be judged as identical by listeners, then this would be evidence of purely acoustic processing of speech stimuli in speech perception. However, if these spectrally identically stimuli would be judged differently, then this would give evidence that underlying articulatory differences (apparent before manipulation and thus supposedly still extractable after manipulation) are indeed used by listeners during the perception experiment, thus pointing to more evidence that articulatory configurations and/or gestures are processed, in addition to the acoustic surface form alone.

32 Canadian English listeners participated in the experiment for course credit, their task was a forced choice identification of the presented sibilants (/s/ or /ʃ/) for each presented audio stimulus from the four speakers. The original and manipulated stimuli were either presented in isolation (C) or embedded in their original vocalic context (/aCa/)[2]. Listeners were not allowed to repeat stimuli, and 8 repetitions of each audio file was randomly presented throughout the experiment. A 10 stimuli practice session was carried out before the main experiment. We also excluded responses above 2.5 standard deviations of each listener's mean reaction time from further analysis.

Figure 2 shows the obtained results over all listeners and all four speakers (who provided the original stimuli). It can be seen that an underlying, but acoustically manipulated /ʃ/ is completely perceived as /s/, whereas the underlying (and acoustically manipulated) sibilants /s/ are not changing sibilant perception to its opposite, but rather generate chance responses. Response changes (to the opposite sibilant) are not linear but rather follow a categorical distribution, but with only underlying /ʃ/ changing completely into the opposite sibilant perception. Significance tests performed on step 7 in figure 2 showed that the perceived differences between underlying /s/ and underlying /ʃ/ are highly significant ($t(31) = -6.211$, $p < .001$)[3], even though their acoustic spectral shape was completely identical. We did not find perceptual differences between the 4 originally recorded speakers, and there was no effect of sibilant presentation mode (i.e. whether sibilants were presented in isolation or embedded in vowels).

---

[1] Thus [s] is stepwise acoustically manipulated to have a matching [ʃ] acoustic shape, and vice versa.
[2] In order to test for the effect of neighboring vowels (versus presentation in isolation) and formant transitions
[3] Step 7 (the manipulated final stimulus result) was compared to the original (opposite) sibilant stimulus perception result.

To conclude, the two fricatives /s ʃ /examined in this study behave differently in the conducted perception experiment: only the acoustically manipulated postalveolar sibilant completely changes perception to an alveolar sibilant, whereas the underlying (but also acoustically manipulated) alveolar sound resists a perceptual class change, even though the acoustic spectral shape of the manipulated /s ʃ/ sounds is almost completely identical to their sibilant counterpart. Since the acoustic shapes to be judged are almost identical but perception results show clear differences in sibilant identification our interpretation of these results is that articulatory and/or motor patterns are indeed additionally used for phoneme processing and identification in acoustically challenging conditions, in this case to help with robust perceptual identification of sibilants.

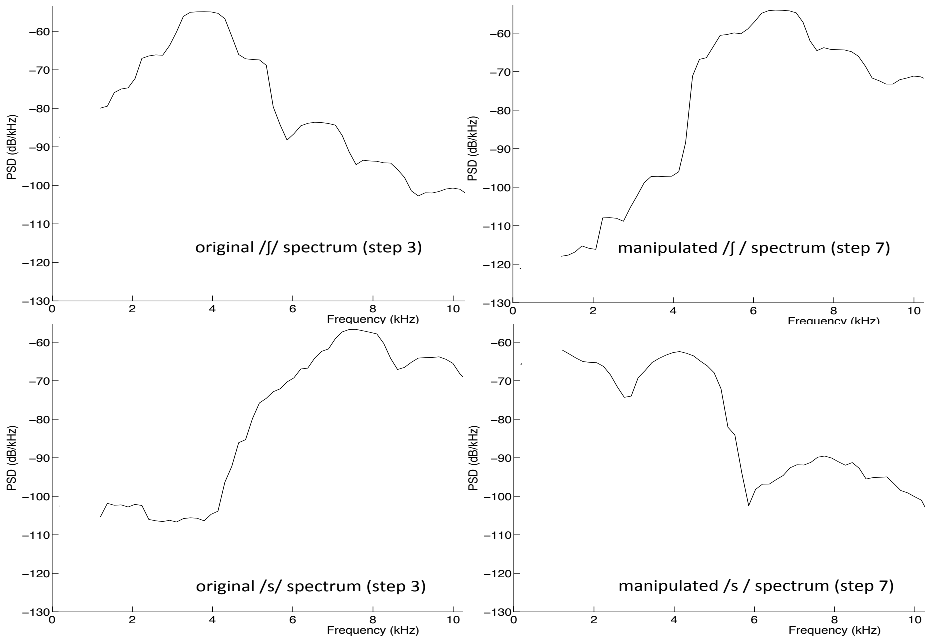

Fig. 1. Comparison between an example of the original recorded stimulus for one speaker (/ʃ/ and /s/, left panels) and the acoustically manipulated stimuli with +48dB amplification of the relevant frequencies of the opposite sibilant (right panels), i.e. *high frequency amplification for underlying / ʃ/ and low frequency amplification for underlying* /s/).

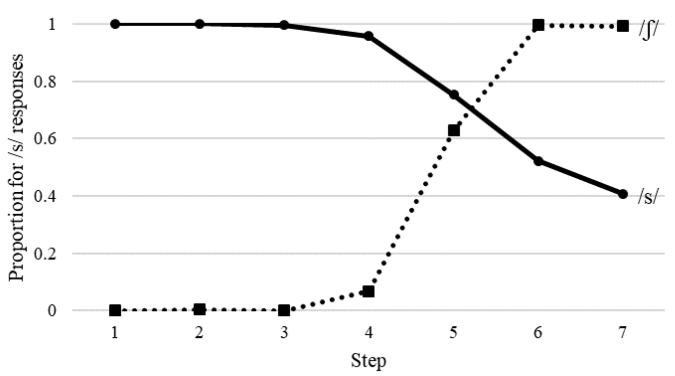

Fig. 2. Results: Probabilities of /s/ identifications (y-axis) against presented stimulus continuum (x-axis). The original recorded sibilant (/s/ or /ʃ/) is shown at step 3 (see figure 1 left panel). The stepwise increases of amplification of frequency regions relevant to the *opposite* sibilant, and thus acoustic manipulations leading to a perceptual change towards to opposite sibilant, are shown in steps 4-7. Step 7 shows the complete acoustic change where underlying /s/ would be identical in acoustic shape to /ʃ/ and vice versa (see figure 1 right panels). For completeness, steps 1-2 show amplification of *frequency regions relevant to the stimulus in question (high frequencies for /s/ and low frequencies for /ʃ/).*