Perception of Canadian English intervocalic and isolated sibilants: Processing of acoustic information or underlying (articulatory) vocal tract configurations?

San-Hei Kenny Luk[1] and Daniel Pape[1]
[1]*McMaster University, Hamilton, Canada*

Acoustic and articulatory speech perception theories aim to explain how listeners map the incoming acoustic signals to phonetic categories. In contrast to acoustic theories, articulatory theories (e.g. Motor Theory [1] or Direct Realist Theory [2]) hypothesize that listeners would recover the underlying articulatory information or motor commands, whereas acoustic theories only rely on the spectral content of the acoustic signal for phoneme identification. To test the hypothesis whether articulatory information can be recovered in isolated and intervocalic sibilant perception, we examined how manipulating different frequencies ranges will affect identification of Canadian English sibilants /s ʃ/. Specifically, we either (a) manipulated the amplitude of the frequency range where the main spectral peaks are found (i.e. around 3kHz for /ʃ/ and 7kHz for /s/) and, therefore, including front cavity and thus place of articulation information (from here on defined as *relevant* frequencies) or (b) the amplitude of frequency ranges where the sibilant place of articulation is *not* coded, i.e. articulatory irrelevant frequency ranges for the perception of a given sibilant (from here on defined as *irrelevant* frequencies with respect to front cavity resonance and thus place of articulation). The more interesting manipulation is altering *irrelevant* frequencies in the acoustic signal: While maintaining the underlying articulatory information, the manipulated sibilants would still be *articulatorily* cued as the original sibilants (i.e. defined by front cavity properties), however they would be *acoustically* cued as the alternative sibilants (i.e. /s/ as /ʃ/, and /ʃ/ as /s/), with completely changed acoustic spectral shape due to the strong acoustic manipulation (see figure 1).

We conducted an identification (speech perception) experiment with 31 listeners to examine how strong manipulation of acoustic information would switch Canadian English listeners' identification scores. Manipulated fricatives were either presented in isolation, re-embedded (after manipulation) in their original /a_a/ context or cross-spliced into their alternative sibilant's vowel context (e.g. fricative /ʃ/ embedded in /a_a/ context where /s/ was removed and replaced with /ʃ/). We found that the listeners' switch in identification categories strongly depends on the phoneme category of the *underlying* sibilant (see figure 2 left): Listeners identified acoustically /s/-like /ʃ/ completely as the alternative sibilant /s/ (see (1)), however the acoustically /ʃ/-like /s/ was perceived only at chance level as the alternative sibilant /ʃ/ (see (2)). In other words, manipulations of the *irrelevant* parts of the spectra with respect to the front cavity frequency range lead to a categorical switch for the *underlying* /ʃ/ stimuli but *not* for the *underlying* /s/ stimuli. The presentation of manipulated fricatives in original vowel context did not change these results. However, the cross-splicing in alternative sibilant vowel context ensured the previously lacking phoneme boundary shift for underlying /s/ stimuli (see figure 2 right), so the inclusion of the original /s/ formant transitions shifted listener responses towards 100% /ʃ/ perception. It seems that perceptual cue-trading between fricative noise and vowel context takes place here.

In conclusion, we here provide additional evidence that articulatory information can in fact be recovered for sibilants, but recovery strongly depends on the *underlying* phonetic category (i.e. in this case restricted to alveolars), possibly due to rivaling articulatory gestures like the presence (/ʃ/) or absence (/s/) of lip rounding gestures facilitating or hindering articulatory parsing. However, we also found that acoustic information clearly dominates articulatory information in the identification process, at least for the tested Canadian English voiceless sibilants.

(1) Manipulation of underlying /ʃ/ stimuli: Strong acoustic manipulation of the *irrelevant* frequency regions (no manipulation for the vocalic part) lead to a complete switch to the /s/ category, with stronger amplitude manipulations leading to more switches. Manipulating *relevant* frequency regions (steps 1 and 2) resulted in the expected ceiling effect of /ʃ/ judgments. The original recorded /ʃ/ stimulus is step 3.

(2) Manipulation of underlying /s/ stimuli: Strong acoustic manipulation of the irrelevant frequency regions did *not* lead to a complete switch towards the /ʃ/ category, with the highest manipulation of step 7 resulting in a mere chance judgment (around 50% /s/ responses and 50% /ʃ/ responses). Manipulating relevant frequency regions (steps 1 and 2) resulted in an expected ceiling effect of /s/ judgments. The original recorded /s/ stimulus is step 3.

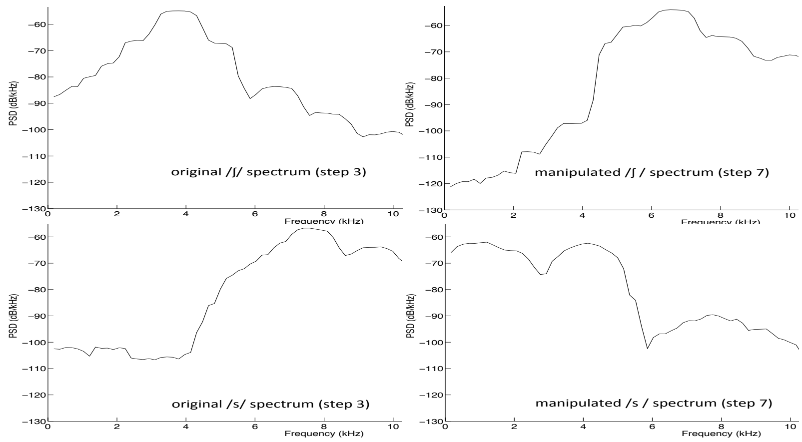

Figure 1. *Comparison between the acoustic spectra of the recorded stimuli (/s/ and /ʃ/, left panels) and acoustic spectra of the manipulated stimuli with strongly amplified irrelevant frequency range (right panels, step7 in figure 2 with +48dB amplification). As a result, the acoustically manipulated stimuli were acoustically very similar in overall spectral shape to the other examined sibilant (/s/ as /ʃ/ and /ʃ/ as /s/).*

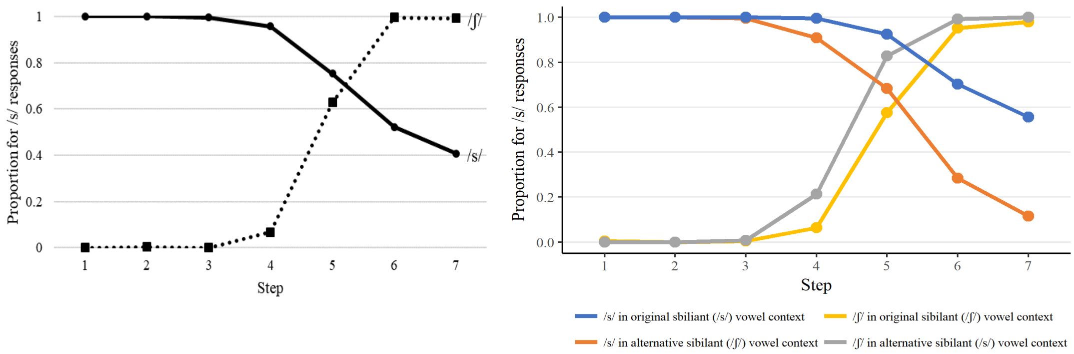

Figure 2. *Identification results for Canadian English sibilant perception (left: isolated fricative noises; right: fricatives in original vowel context or cross-spliced alternative sibilant vowel context. The original recorded sibilant (/s/ or /ʃ/) is step 3, amplified relevant frequency stimuli are steps 1+2, and amplification of irrelevant frequency regions are steps 4-7. The magnitude of the acoustic manipulation increases linearly from step 4 to step 7 (step4 = 12dB, step7 = 48dB). See text for definition of relevant and irrelevant frequency regions.*

[1] Galuntucci, B., Fowler C.A. & Turvey M.T. 2006 The motor theory of speech perception reviewed, *Psychonomic Bulletin & Review* 13(3), 361-377.

[2] Fowler, C.A. (1986). An event approach of the study of speech perception from a direct- realist perspective. *Journal of Phonetics*, 14, 3-28.

[3] Stevens, K. 1998. *Acoustic Phonetics.* MIT Press.

[4] Jongman A, Wayland R. & Wong S. (2000). Acoustic characteristics of English fricatives, *Journal of the Acoustical Society of America* 108, 1252-1263.