Perceptual cue weighting of intervocalic velar plosives in Canadian English

Daniel Pape, Olivier Mercier

paped@mcmaster.ca; McMaster University (Hamilton, Canada)

When producing speech, acoustic signals contain a number of extractable parameters, or acoustic cues, which then, on the perceptual side, can be used to categorize and classify different phonemes. The actual perceptual relevance of the different available acoustic cues and their interaction is known as perceptual cue-weighting. We investigated how Canadian English listeners (n=38) from the Greater Toronto Area identify the voicing category of intervocalic velar plosives (i.e. /k/ versus /g/) for biomechanically generated stimuli differing systematically and orthogonally in a number of acoustic parameters known to influence stop voicing. For this reason, we embedded the velar stops in intervocalic /a/ contexts to create an experimental stimuli set that varied the following acoustic cues to stop voicing along four continua: (1) voicing maintenance throughout consonantal closure (VM), (2) voice onset time (VOT), (3) duration of consonantal closure (CL) and (4) duration of the previous vowel segment (VL). We were interested in the importance, weighting and interaction of all four acoustic cues when tasking listeners to classify the VCV sequence in a forced-choice test (choice: /aka/ or /aga/). We used biomechanical modelling, a synthesis approach known to generate natural, highly controllable and very accurate stop consonants, as shown e.g. by accurate reproductions of important fine phonetic details like articulatory loops (see e.g. Perrier et al. 1998).

In the following, we present findings for three experiments: **Experiment 1** investigates the cue weighting of three concurrent acoustic cues (VM, VL, CL) in the absence of an *acoustic* VOT cue[1]. **Experiment 2** then determines an ambiguous VOT value (i.e. the perceptual categorical boundary between English /k/ and /g/ perceptions) which was specifically measured for our biomechanical stimuli. Additionally, the experiment examined how varying VOT values[2] from 0ms to 100ms would interact with the VM parameter to influence voicing perception when duration cues were held ambiguous (i.e. intermediate VL and CL durations). This experiment thus explored not only the effect of VOT on voicing perception, but also how VOT interacts hierarchically with other cues. **Experiment 3** then adds the ambiguous VOT boundary (measured in experiment 2) to the previous other three cues already examined in experiment 1 (VM, VL, CL). Therefore, experiment 3 compares how the perceptual cue weighting of the three acoustic cues (VM, VL, CL) is affected when a perceptually ambiguous VOT value is either present (exp. 3) or absent (exp. 1) in the acoustic signal, in other words how the perceptual effect of absence vs. presence of ambiguous VOT on the voicing distinction manifests on English listeners when faced with the task of judging the effects of the other acoustic cues (VM, VL, CL).

The results from the three experiments reveal that, for our native Canadian English listeners, VM has a significantly higher influence on voicing perception than all other cues (VM > VL > CL), and that only at low VM levels other cues increase their influence, as can be seen in figure 1. The absence vs. presence of ambiguous VOT values does not change perception significantly for Canadian English listeners (see figure 1, blue lines): they still mostly rely on VM, to some extent on VL but not on CL. The presence of ambiguous VOTs pulls the listener responses more towards voiceless identifications, but this interaction strongly depends on the presented voicing maintenance value: The interaction is strongest for low VM values but

---

[1] For this condition, no acoustic stop burst is present in the signals. The temporal relationship between articulatory vocal tract opening and onset of the voicing of the following vowel is identical for all stimuli.

[2] The parameter VOT was introduced by adding identical acoustic burst signals with varying temporal distances with respect to the voicing onset of the following vowel. Thus, VOT here is defined as an acoustic cue consisting of varying temporal distances between spliced burst (identical for all stimuli) and vowel voicing onset (also identical for all stimuli).

disappears for higher values (above 25 ms voicing during closure). With respect to listener variability, one of the main results for experiment 2 is that individual listeners utilized very different strategies in establishing their individual cue-weighting patterns for their intervocalic voicing perception in Canadian English, as can be seen in figure 2.

Our results add additional evidence that the perceptual cue weighting process for intervocalic plosives is highly complex and cannot be argued to rely dominantly on the classic main cue(s). Finally, the dominance of VM in our results is surprising since it is often argued that voicing maintenance during stop closure is not an important parameter for English stop voicing classification.

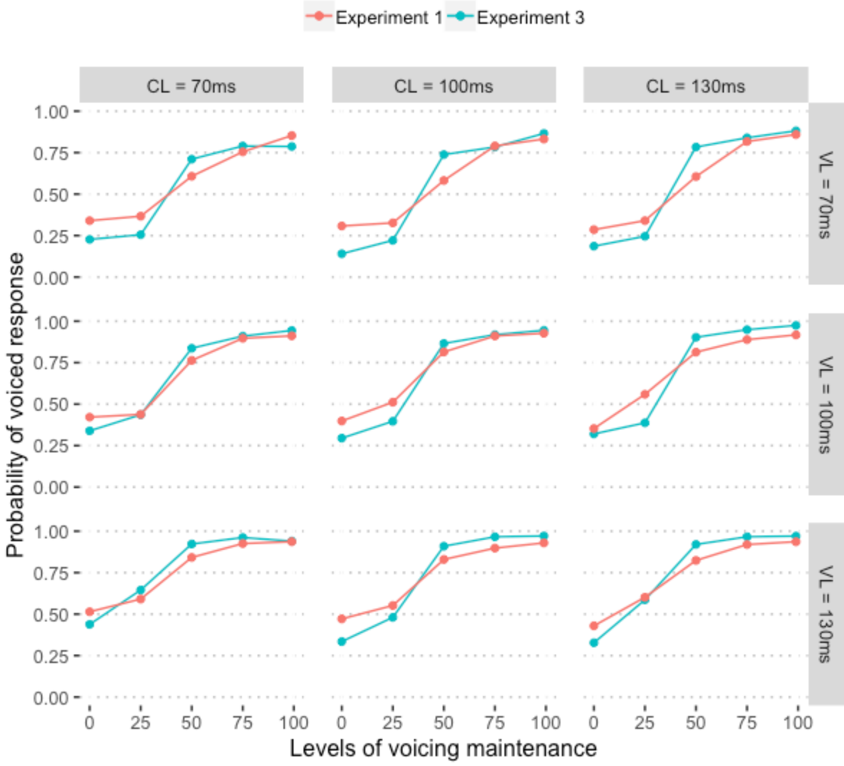

Figure 1: Perception results for experiment 1(red lines) versus experiment 3 (blue lines). Shown are the presented VM levels during the consonant duration (ranging from 0% to 100%) on the x-axis versus the probability that the listeners chose a voiced response (i.e. /aga/ instead of /aka/). The columns show the variation of consonant duration and the lines show the variation of the preceding vowel duration. As can be seen, the inclusion of an ambiguous VOT cue (red lines) compared to the absence of a VOT cue (blue lines) did not change listener responses in a relevant manner. Listeners were most sensitive to VM variation, especially in the lower VM regions (0% - 25%), were less sensitive to vowel duration differences (more in the lower VM region) and barely sensitive to stop consonant duration.

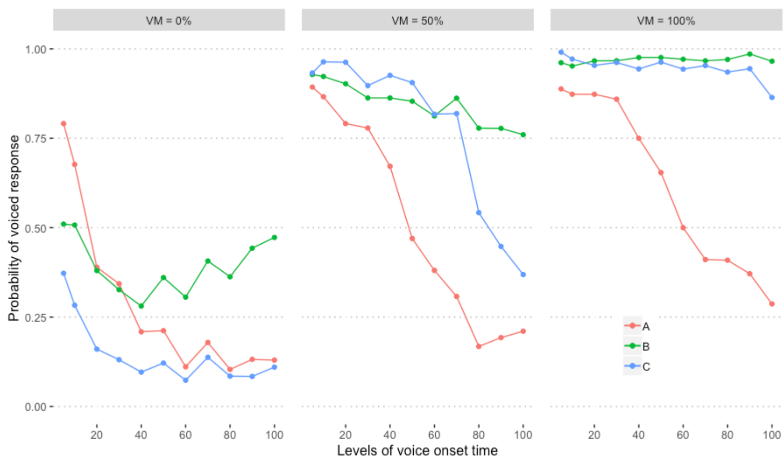

Figure 2: Results for experiment 2 for the comparison of VOT (x-axis) versus voicing maintenance levels (panels from left to right) for different listener response patterns, i.e. different group responses. Group A (in red) is clearly more sensitive to the

VOT perceptual cue and rather insensitive to the simultaneously presented VM difference (ranging from completely unvoiced (VM=0%) to fully voiced stimuli (VM=100%). In contrast, group C is rather sensitive to the presented VM differences but appears to be insensitive to the presented VOT differences.

Perrier, P., Payan, Y., Perkell, J., Jolly, F., Zandipour, M., Matthies, M. (1998): "On loops and articulatory biomechanics", In *ICSLP-1998*, paper 0112.