**The effect of realtime visual feedback on vocalic targets**

Elizabeth Stelle\(^1\) (stell@alumni.ubc.ca), Eric Vatikiotis-Bateson\(^1\), and Caroline L. Smith\(^2\)

\(^1\)The University of British Columbia, \(^2\)The University of New Mexico

**Introduction:** The dynamic nature of phonological contrasts is seen in both speech perception (e.g. Bertelson et al., 2003) and speech production (e.g. Houde & Jordan, 1998). In speech production, realtime perturbations to a speaker's auditory feedback cause compensatory shifts in the production of the perturbed phonological category. Speech production feedback thus plays a role in the maintenance of phonological contrasts (e.g. Lane et al., 2007). While removing or changing feedback has been explored in a number of experimental paradigms, adding a novel source of feedback has been somewhat less studied. Visual speech information has long been known to play an important role in speech perception (Sumby & Pollack, 1954; McGurk & MacDonald, 1976), but there is growing evidence that it can also play a role in speech production. Recent research has shown that visual feedback—by means of a mirror or realtime video—has an effect on the production of whole utterances produced under delayed auditory feedback conditions (Jones & Striemer, 2007; Chesters et al., 2015; Stelle et al., 2015). The present research extends this work by testing whether realtime visual speech feedback can also affect the production of vowel targets.

**Hypothesis:** We hypothesized that visual feedback would be used during speech production to maintain accurate speech targets, especially in difficult speaking conditions, when the speech motor control system is more likely to rely on feedback mechanisms. We tested this by comparing the effect of visual feedback on speech produced with and without a bite block. Following Lane et al., (2005), we looked at within and between category vowel differences, measured in terms of acoustics. Between category differences, or vowel contrast, refer to the degree of separation between vowel categories, while within category differences, or vowel dispersion, refer to the variability of the various instances of a given category. We predicted that vowel contrast would be diminished during bite block production and increased with visual feedback, and that vowel dispersion would be increased during bite block production and diminished with visual feedback. As a complement to the predicted acoustic differences, we predicted that articulatory differences would also be found during speech production with visual feedback; specifically, that lower face motion would be greater with visual feedback than without it.

**Methods:** A within-participants, 2x2 factorial design was used. The four conditions combined the factors, contrasting whether the oral perturbation and visual feedback were absent or present. Condition order was pseudo-randomized across participants; participants always completed the conditions without the bite block before the conditions with the bite block, but the order of the visual feedback conditions was randomized within each of these subgroups. Stimuli were seven American English monophthongs produced in an /hVd/ context. The stimuli were presented multiple times in each condition. Participants were instructed to listen to a word and then repeat it while looking at the monitor, which displayed either a cross or realtime visual feedback of the participant. Participants were audio and video recorded throughout the experiment.

**Results:** The effects of the bite block and visual feedback manipulations were assessed in terms of vowel dispersion, operationalized as Euclidean distance, and vowel contrast, operationalized as Mahalanobis distance. The Mahalanobis distances were used to calculate the average vowel space (AVS) distance; this was the mean of the distances between all ordered vowel pairs. The two measures were made at three points during the vowel: 25%, 50%, and 75%. Effects were assessed by model comparison of mixed effects models. A summary of the AVS results is displayed in Figure 1. Model comparisons revealed that visual feedback significantly increased vowel contrast at the 25% time point (p<0.05), while at the 50% time point the interaction between oral perturbation and visual feedback was close to significant (p=0.08489); there was a tendency for visual feedback to increase vowel contrast, but more so during normal production than bite block production, the latter being counter to our prediction. There were no significant effects at the 75% time point. A summary of the Euclidean distance results is displayed in Figure 2. Model comparisons revealed that visual feedback significantly decreased vowel dispersion at the 25% time point (p<0.05), with a tendency to also do so at the 50% time point (p=0.07056). The oral perturbation significantly increased dispersion at the 50% time point (p<0.05) and had a tendency to do so at the 25% time point. There was a significant interaction between oral perturbation, visual feedback, and...
phoneme, at the 75% time point (p<0.05). Post-hoc tests suggest that this interaction is primarily driven by the /u/ vowel; vowel dispersion increased during bite block production, especially for /u/ when visual feedback was absent. Analysis of the video data was performed using optical flow analysis (Barbosa et al., 2008) to infer magnitudes of motion (MM) of the lower face from inter-frame changes in pixel brightness. Initial results suggest that, for a subset of participants, mean MM increases with visual feedback for non-high vowels, and that overall there is a significant positive correlation between mean MM and AVS for non-high vowels (p<0.05). Overall, our results provide further support for the dynamic nature of phonological targets; not only can they be refined by a novel source of speech production feedback—visual feedback—but in this context the productions change to enhance visually salient aspects of articulation.

Figure 1: Vowel contrast, measured as Average Vowel Space, across conditions at three points in the vowels.

Figure 2: Vowel dispersion, measured as Euclidean distance, across conditions at three points in the vowels.

References