Kinematic evidence of centering during vowel production

Benjamin Parrell[1], Mark Tiede[2], Vince Gracco[2], Doug Shiller[3]
[1]University of Wisconsin–Madison, [2]Haskins Labs, [3]University of Montreal

INTRODUCTION: Typically, the role of the auditory system in online control of speech has been probed by delaying auditory feedback [1] or by altering its spectral characteristics [2-3]. Many studies have shown through these methods that the neural control of speech movements is indeed sensitive to these external auditory perturbations. Recently, an alternative method to masking or external perturbations of sensory feedback has been proposed [4-6]. Rather than imposing sensory perturbations, this method leverages the natural variability found in speech production to examine how speakers alter their productions online. These studies have shown that vowel productions which initially fall near the edge of the sound's distribution in F1/F2 space (for a given talker) exhibit movement towards the middle of the distribution over time, a phenomenon known as *centering*. While the sensorimotor mechanisms underlying this behavior are not clear, it has been suggested that auditory feedback may play a role. First, masking noise has been shown to attenuate the magnitude of the centering behavior [5]. Second, trials which fall near the edge of the acoustic vowel distribution generate neural signals that are similar to those generated when auditory feedback is externally perturbed, suggesting that the auditory system is able to distinguish these peripheral trials from typical productions [4]. Here we test whether centering can be observed at the level of speech kinematics in addition to speech acoustics. The examination of speech kinematics also allows us to test whether centering behavior occurs prior to the vowel midpoint, when auditory feedback would not be available. If such centering occurs, it would suggest that this behavior may rely at least partially on sources other than auditory feedback, such as internal predictions [7], somatosensation, or increasing restrictions on permitted variability at the planning level [8].

METHODS: This pilot study involved two adult native speakers of American English (one male, one female), with no reported speech, language or hearing deficits. All procedures were approved by the Yale University IRB. Electromagnetic articulography (*EMA*; Wave, Northern Digital Inc.) was used to measure the 3D position of sensors attached to the tongue (midsagittal dorsum, blade and tip), jaw (upper and lower incisors, left premolar), and lips (upper/lower) relative to the head. Participants were instructed to produce individual words (visually presented on a computer display) drawn from the set *Ed*, *add*, *ebb*, *ab*, *shed*. 60 repetitions of each word were produced in pseudo-randomized order.

For each token, F1/F2 traces (in mels) spanning the vowel were averaged over a 50 ms window at vowel onset (*Window-1*), as well as a 50 ms window centred in the middle of the vowel (*Window-2*). A measure of *vowel distance* was calculated within each time window as the Euclidean distance between the trial's F1/F2 values and the median F1/F2 values within that time window. Vowel centering was then calculated as the change in vowel distance between Window-1 and Window-2. Using the same two time windows (Window-1 and Window-2) identified on the basis of the acoustic signal for each utterance, the mean x- and z-position was computed for each of the three tongue markers. A measure of *kinematic vowel centering* was computed using the same approach described above for acoustics, only in this case using the x and z positions (midsagittal plane) of the tongue in place of F1 and F2.

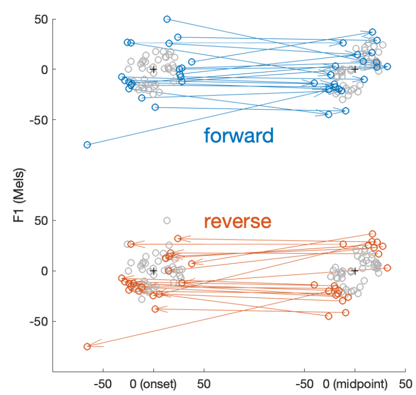

Figure 1: Illustration of centering (blue) and reverse centering (red) analyses.

An additional analysis used the same two time-windows, but *reversing* their temporal order in the analysis (i.e., treating the data from Window-2 as if it was the onset of the movement, and Window-1 as if it were the vowel midpoint, see Fig. 1). By examining whether the original centering measure exceeds this new measure of *reverse centering* (both of which include the same regression to the mean effects due to measurement noise or random physical variation), we can test the robustness of the centering effect as a phenomenon beyond simply a statistical artifact [6]. Statistical analyses of acoustic and kinematic measures was conducted using linear mixed-effects models in MATLAB with random intercepts and slopes for each participant.

RESULTS: Mean acoustic centering from vowel onset to vowel midpoint was 12.5 mels and the mean reverse centering was 7.4 mels (Fig. 2). There was a significant effect of direction ($p = 0.01$), and word ($p = 0.03$), but no interaction ($p = 0.07$). This suggests that there was more centering than reverse centering, consistent with the hypothesis that centering is not solely caused by regression to the mean. However, the amount of centering in the current data is substantially lower than that which has been previously reported [4-6]. This may indicate that the two speakers we recorded happen to fall on the lower end of the normal range of centering behavior, or potentially that the EMA protocol (including affixation of sensors) affects centering by altering normal speech patterns.

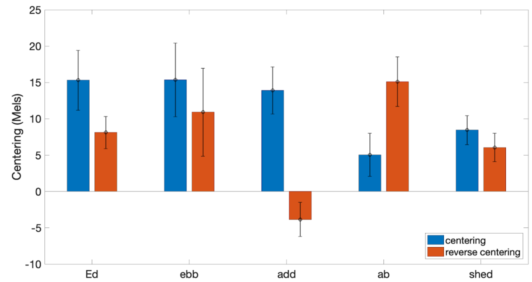

Figure 2: Acoustic centering (blue) and reverse centering (red), shown by word, pooled across both participants. Means and standard errors shown.

Kinematically, the Tongue Dorsum (TD) sensor showed a significant effect of direction ($p = 0.003$), with a mean centering of 0.37 mm and a mean reverse centering of 0.07 mm. There was no significant effect of word or interaction between word and measurement (both $p = 0.09$). For the Tongue Tip (TT), there was no main effect of direction, but there was a main effect of word ($p = 0.002$) as well as an interaction between measurement and word ($p = 0.001$). *Ed* showed more reverse centering than centering, while *add, ab,* and *shed* showed more centering than reverse centering. The measurements were roughly equal for *ebb*. No other sensor showed a significant effect of direction.

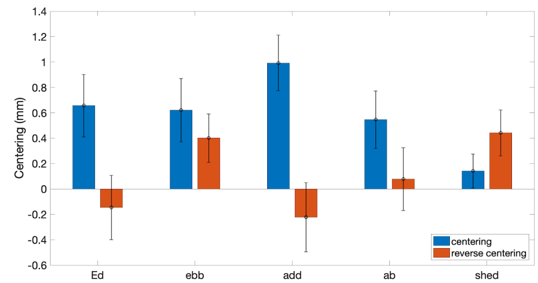

Figure 3: Kinematic centering for the tongue dorsum (blue) and reverse centering (red), shown by word, pooled across both participants. Means and standard errors shown.

In order to assess the extent to which centering may be driven by factors other than auditory feedback, we conducted an additional analysis examining articulatory centering around the time of vowel onset. Here, we compared the articulatory position before vowel onset (-100 to -50 ms prior to onset) to the articulatory position soon after vowel onset (25-75 ms after onset). Critically, this second time window is prior to the time when corrective alterations to ongoing speech based on auditory feedback could be expected (~150 ms after onset). Both the Jaw ($p = 0.014$) and the TD ($p = 8e\text{-}6$) showed more centering than anti-centering. A marginal effect was found for the TB ($p = 0.051$). No other effects were found for the jaw, but the TD also showed significant effects of word ($p = 0.004$) as well as an interaction between word and measurement ($p = 0.0005$).

DISCUSSION: We have shown that centering is visible in speech kinematics. Centering is most consistently seen in the tongue dorsum, though it may also be seen in other articulators such as the tongue tip and jaw. Importantly, centering is present not only after the acoustic onset of the vowel, but also from before to immediately after vowel onset. This suggests that centering is driven, at least partly, by factors other than auditory feedback. These potential influences include somatosensory feedback, internal predictions (of auditory feedback, somatosensory feedback, and/or articulator positions), and increasing restrictions on the permitted variability at vowel midpoint compared to vowel onset [8].

[1] Yates, A. 1963. Delayed Auditory Feedback. *Psychological Bulletin* 60, 213–32.

[2] Purcell, D., Munhall, K. 2006. Compensation Following Real-Time Manipulation of Formants in Isolated Vowels. *J Acoust Soc Am* 119, 2288–97.

[3] Burnett, T., Freedland, M., Larson, C., Hain, T. 1998. Voice F0 Responses to Manipulations in Pitch Feedback. *J Acoust Soc Am* 103, 3153–61.

[4] Niziolek, C., Nagarajan, S., Houde, J. 2013. What Does Motor Efference Copy Represent? Evidence from Speech Production. *J Neurosci* 33, 16110–16.

[5] Niziolek, C., Nagarajan, S., Houde, J. 2015. The Contribution of Auditory Feedback to Corrective Movements in Vowel Formant Trajectories. *Proc, 18th ICPhS*, Glasgow, UK.

[6] Niziolek, C., Kiran, S. 2018 Assessing Speech Correction Abilities with Acoustic Analyses: Evidence of Preserved Online Correction in Persons with Aphasia. *International Journal of Speech-Language Pathology*.

[7] Parrell, B., Ramanarayanan, V., Nagarajan, S., Houde, J. 2019 The FACTS Model of Speech Motor Control: Fusing State Estimation and Task-Based Control. PLOS Computational Biology 15, e1007321.

[8] Keating, P. 1990. The Window Model of Coarticulation: Articulatory Evidence. In: Kingston, J., Beckman, M. *Papers in Laboratory Phonology I*. Cambridge: Cambridge University Press, 451-470.