12th International Seminar on Speech Production
Providence, RI, USA, June 1-4, 2020

Gestural coordination in non-native onset clusters: An electromagnetic articulography study

*Hung-Shao Cheng[1], *Matthew Masapollo[2], Christina Hagedorn[3], & Adam Buchwald[1]

[1]*Department of Communicative Sciences and Disorders, New York University*
[2]*Department of Speech, Language, & Hearing Sciences, University of Florida*
[3]*Department of English, CUNY College of Staten Island*

Introduction. A major goal of research in speech production has been to explicate how speakers control the duration, timing, and sequencing of articulatory gestures. Over the years, a number of studies have investigated this issue by examining how speakers articulate novel speech sequences, and, in particular, consonant clusters, that are not found in their native language (see, e.g., Davidson, 2005, 2006, 2010; Wilson, Davidson, & Martin, 2014; Buchwald, Calhoun, Rimikis, Steinberg Lowe, Wellner, & Edwards, 2019; Segawa, Masapollo, Tong, Smith & Guenther, 2019). Most of these studies have explored adults' initial attempts at producing non-native clusters using a combination of perception-based segmental transcription and simple acoustic measures (but see Davidson, 2005). Relatively less, however, is known about the articulatory aspects of non-native production. Examining the articulatory movement characteristics of non-native speech sequences provides an opportunity to not only increase our understanding of early aspects of speech motor sequence learning but also speech organization and production planning more generally. The current research addresses this gap in the literature by examining the coordination of the speech articulators (tongue, lips and jaw) when speakers attempt to produce non-native consonant sequences in syllable-initial (i.e., onset) contexts.

Acoustic analyses and perceptual data on the production of phonotactically-illegal onset clusters (e.g., Davidson, 2006, 2010; Wilson, Davidson, & Martin, 2014; Buchwald, Calhoun, Rimikis, Steinberg Lowe, Wellner, & Edwards, 2019) have revealed that some non-native clusters are produced less accurately than others are. For example, Buchwald *et al.* (2019) reported that native English speakers tend to produce /f/-nasal clusters (e.g., /fn/) more accurately than stop-stop clusters (e.g., /gd/), and Davidson (2010) reported that voiceless stop-stop clusters are produced more accurately than voiced stop-stop clusters. Moreover, and of particular relevance to the current study, recent laboratory-based speech training studies (Buchwald *et al.*, in preparation) have demonstrated that English speakers refine their productions of /fn/ clusters even once they are perceptually acceptable. Specifically, the duration of the portion of the acoustic signal associated with the nasal segment systematically decreased following motor training, presumably due to increased spatial and temporal overlap among the gestures corresponding to neighboring fricative and nasal segments. Collectively, these findings suggest that speakers (implicitly) continually update their implementation of gestural scores (e.g., by temporally optimizing consonant-to-consonant transitions). Critically, however, the aforementioned studies assessed non-native production using acoustic analysis alone, which provides an indirect measure of articulatory movement characteristics. In the present study, we use electromagnetic articulography (EMA) to quantify the articulatory phasing between successive /f/ and /n/ gestures in non-native onset clusters.

A second set of findings motivating the current research comes from studies examining the relative timing and timing stability of gestures in native speech sequences. Specifically, kinematic

---

* joint 1st author

studies of onset clusters in English (Browman & Goldstein, 1988; Byrd, 1996; Marin & Pouplier, 2010; Pastatter & Pouplier, 2017), have reported that the timing of the onset as a whole with respect to the vowel remains relatively stable, regardless of the number of consonants and the composition of a given onset cluster (i.e., "C-center" effect). On the basis of these findings, some theoretical accounts have proposed that onsets and vowels are fused together as cohesive action units, or "gestural molecules," at least at some level of the speech production process (for discussion, see Goldstein, Byrd, & Saltzman, 2006; Pastatter & Pouplier, 2017). That is, each consonant is in-phase coupled with the following vowel while anti-phase coupled with each other, resulting a stable "C-center" coordination pattern (Browman & Goldstein, 1988; Marin & Pouplier, 2010; Pastatter & Pouplier, 2017). Here, we examine whether auditorily acceptable /fn/ onset clusters exhibit onset-vowel C-center timing.

Methods. Data collection is currently underway. Ten native English speakers will be recruited and tested. We will record and analyze articulatory movement data from productions of read disyllabic (CCVC.C[C]VC) non-words, comprised of /fn/ onset clusters in the first syllable (e.g., /fnɑb.zud/). We also included words with the native /sm/ onset cluster and relevant singletons (/f/ and /n/) to serve as controls (e.g., /fɑb.zud/, /nɑb.zud/). All cluster targets, both native and non-native alike, were selected such that each consonantal constriction was made with a different primary articulator. In this way, it will be possible to identify the maximum vocal tract constriction for each consonant in the position trajectories. The target words will be produced in a simple carrier sentence in order to control vocal tract configuration before and after the segments of interest. Speech movements will be recorded using an NDI Wave EMA system in a sound-treated booth. Six sensors will be attached to the midsagittal surface of the articulators: tongue tip, tongue blade, tongue dorsum, upper lip, lower lip, and jaw. Reference sensors will be used to correct for head movement. Participants' productions will be recorded with a microphone. The text stimuli will be presented on a lap top screen. Each target nonword will be produced three times pseudo-randomly, with no words occurring consecutively. The EMA data will be low-pass filtered, corrected for head motion, and re-referenced to each individual speaker's occlusal plane in MATLAB. Using the Mview algorithms, developed by Mark Tiede at Haskins Laboratories, the kinematic data will then be computed, visualized and analyzed. Target gestural landmarks (e.g., max constriction [MaxC]) will be semi-automatically labelled using the *findgest* procedure.

Planned analyses and predictions. Following Browman and Goldstein (1988), we will calculate the means and standard deviations for each of the following dependent measures of duration: (1) the interval between the MaxC of the leftmost (/f/ or /s/) consonantal gesture and the MaxC of the coda consonant of the first syllable; (2) the interval between the MaxC of the rightmost (/n/ or /m/) gesture and the MaxC of the coda of the first syllable; (3) the interval between the temporal midpoint of MaxC for C_1 (/f/ or /s/) and C_2 (/n/ or /m/) (i.e., the C-center) and the MaxC of the coda. If there is a relative constant phasing between initial consonant sequences and the following vowel, then we predict that the third duration measure will be the most stable (i.e., exhibit the lowest standard deviation). The planned analyses will provide fundamental information on gestural coordination in non-native production. In so doing it will shed light on how speakers begin to learn novel sound sequences within extant temporal frames.

Selected References. Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45, 140-155. / Buchwald, A., *et al*. (2019). Using tDCS to facilitate motor learning in speech production: The role of timing. *Cortex, 111,* 274-285. / Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan, *J. Phonetics*, 38, 272-288.