TEMPORAL ORGANIZATION OF OFF-GLIDES IN AMERICAN ENGLISH

Fang-Ying Hsieh and Louis Goldstein

Department of Linguistics, University of Southern California
fangyinh@usc.edu

ABSTRACT

This paper tests the hypothesis that offglides /j/ and /w/ are best analyzed as a syllable coda through examining the temporal organization of offglide gestures within a syllable. Kinematic data were collected from four American English speakers using realtime MRI. The gesture timing of offglides was compared in singleton and complex coda contexts to determine whether offglides exhibit local gestural coordination with the preceding vowel, as would have been predicted by the non-competitive coordination model hypothesized for complex coda. Results show a temporal shift of the offglides when other coda consonants are present, suggesting a different coordination relationship for offglides in English. Accordingly, we propose a unique gestural organization that involves competitive coupling to account for the current findings, which captures both the findings and their dual phonological nature.

**Keywords**: offglides, articulatory timing, syllable organization.

1. INTRODUCTION

Glides, or "semi-vowels," often pose a problem for phonology due to their dual nature: they are similar to high vowels in terms of articulatory correlates, but occupy syllable margins like consonants. In English, controversy arises as to whether offglides (in diphthongs like [aj] and [aw]) should be affiliated to the syllable nucleus or coda position. Arguments for [aj] and [aw] as complex nuclei were mainly based on the fact that they behave like long vowels in terms of phonotactics [1]. Such a behaviour is expected as these diphthongs were historically derived from long vowels [i:] and [u:]. In prosody, however, diphthongs are indistinguishable from a nucleus-coda (VC) sequence. It therefore raises the questions of whether it is necessary to analyze offglides as a part of the syllable nucleus in synchronic English grammar and how they are distinguished from consonant codas.

In this paper, the syllable role of English offglides is addressed from the perspective of gestural coordination. Coordination of articulatory movements is governed by linguistic structure and in the domain of syllable, previous articulatory studies have found that temporal coordination of gestures reflects syllable constituency: successive onset consonants overlap relatively more with the vowel compared with the coda consonants. In the framework of Articulatory Phonology (AP) [4] and task-dynamics [5], this difference has been modelled as differences in speech planning: Onset consonants are hypothesized to be coupled in-phase ($0°$, synchronous) to the vowel and anti-phase ($180°$, sequential) to each other, giving rise to the competition among onset consonants and substantial temporal overlap between the entire cluster and the vowel. On the other hand, coda consonants are hypothesized to only involve local organization, with the first coda consonant (C1) anti-phase coupled to the vowel and the following consonants anti-phase or eccentrically (with arbitrary relative phases) coupled to the preceding consonant [6]. This non-competitive coupling predicts no increase in temporal overlap between V and C1 when the number of gestures increases in coda position. These two types of gestural organization are illustrated in the coupling graphs below (Fig 1).

Figure 1: Gestural organization for complex onset (left) and for complex coda (right).
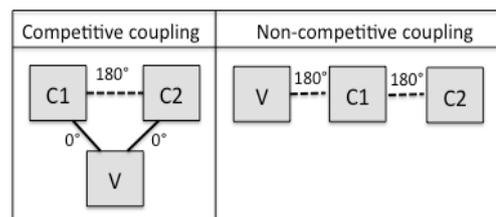
In this paper, the temporal behavior of offglide gestures in [j] and [w] in English was compared with previous findings about coda consonants in English [3] to diagnosis the syllable role of offglides. Specifically, we compared the temporal organization of these offglides in syllable final position (singleton G) versus in a syllable coda cluster (complex GC) to see if temporal overlap between V and G remains unchanged with increase in coda complexity. Based on the temporal organization observed in this paper, we propose a unique coordination relationship for English offglides to capture the current findings. The proposed model treats offglides as a "special coda" that shares properties with the coordination relation of coda consonants, but also differs in a crucial way,

which in turn captures the unitary status of VG as complex nucleus.

2. METHODOLOGY

2.1. Data acquisition

Midsagittal speech articulation from four adult native speakers of American English (one male and three female) was captured using a custom MR protocol developed especially for research on speech production [7]. Synchronous noise-cancelled audio was also collected at a sampling rate of 20 kHz during MRI acquisition. During the scan, participants lie supine with the head restrained in a fixed position to facilitate comparisons across acquisition. Data were reconstructed into a sequence of video frames with spatial resolution of 68-by-68 pixels, with pixels 3 mm in width, at a temporal rate ranging from 23.18 to 32.45 frames per second.

All target words are monosyllabic words embedded in a final (*pre-IP boundary*: I didn't think I'd see ____.) or non-final (*pre-ip or smaller boundary*: I didn't think I'd see ____ anymore.) position of a carrier phrase. The stimuli used for the current study (18 target words) were presented in Table 1 below.

Table 1: List of target words.

Offglide	G	C	GC
/**j**/	pie	po**t**	figh**t**
		po**d**	pie'**d**
	tie	to**p**	ty**p**e
		To**m**	ti**m**e
/**w**/	pow	po**t**	pou**t**
		po**d**	pow'e**d**
	tau	to**t**	tou**t**
		To**dd**	tow**n**

These stimuli were designed to allow for a comparison of the gesture timing between complex coda (GC) and corresponding simple coda (G or C). Each target word begins and ends with either a labial (/p/, /f/) or an alveolar (/t/, /d/) gesture except for the /w/-series because offglide /w/ cannot be followed by noncoronals in English phonotactics [8]. All target words occur in both final and non-final positions; a total number of 4 to 8 tokens were collected for each target sentence from each subject.

2.2. Data analysis

To automatically quantify the constriction location, constriction degree, and gestural kinematics of the participating gestures from the MR images, we adopt a method that directly uses the mean intensities within localized regions-of-interest (ROIs) of the vocal tract [9, 10, 11]. This technique is based on the fact that pixel intensity values are a function of soft tissue density and that localized changes in tissue density signify a vocal tract constriction, e.g. higher intensity in a region implies the presence of more tissue and thus narrower constriction. The ROI technique has been shown to be an efficient and robust way to extract linguistically meaningful constriction degree information.

A set of three grid lines perpendicular to the vocal tract midline (derived automatically) was used as the ROI for each non-pharyngeal gesture, while a horizontal rectangle was used as the analysis region for the vowel gesture based on the assumption that pharyngeal constriction (that produces vowel [a]) may possibly involve a less precise location of constriction than oral constrictions, requiring only a narrowing in the upper or lower pharynx (see Fig. 2).

Figure 2: Representative ROI locations for measuring labial (Lab), alveolar (Alv), palatal (Pal), dorsal (Dors), and pharyngeal (Phar) gestures. Figure shows locations for subject JN.
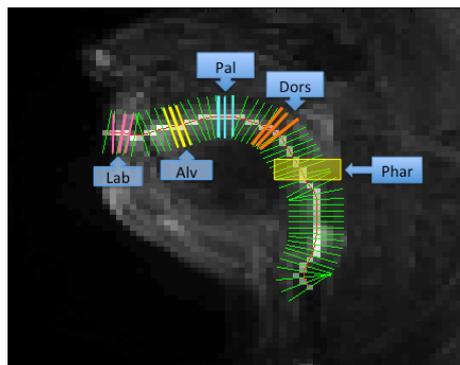

The analysis region for each gesture was chosen from a set of candidates in the corresponding vocal tract region (e.g. the grid lines below the hard palate for palatal /j/ gesture) using specific algorithms. The method for selecting the analysis region for pharyngeal gesture can be found in [12]. For other gestures, following [13], the optimal analysis region was defined dynamically as the site of maximal constriction (intensity) change in the corresponding image frames because this is the region over which the articulator moves most dynamically towards and away from its constriction target.

2.3. Data measurement

Averaging pixel intensity of the ROIs (i.e. all pixels through which each set of three gridlines passes for non-pharyngeals and pixels within the rectangle for pharyngeal gesture) for the image sequences during the production of a target word yields intensity time

functions for each participating gesture, exemplified in Fig 3. These time functions were smoothed using a locally weighted linear regression [14] in order to eliminate the random fluctuations across frames. The data were upsampled by a factor of 5 (between frames) and the smoothing window width was 0.9 frame. They reflect change in constriction degree over time for the gestures and thus allow us to define their kinematic events of and to investigate the temporal coordination among them.

Figure 3: Intensity time functions of the highlighted regions in Fig 2 during the production of "pout" (x- axis: frame number, fps = 32.45; y-axis: pixel intensity)
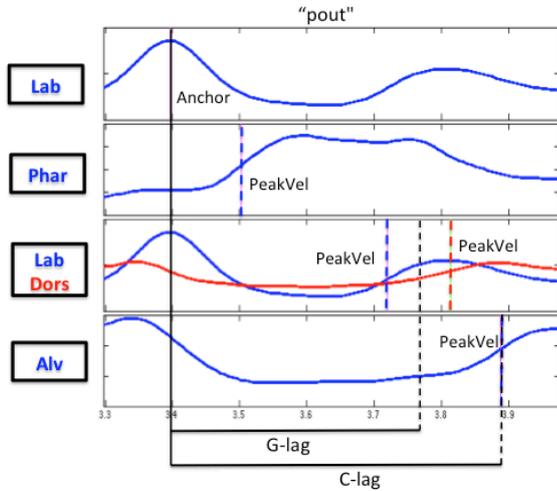
In the current study, we are particularly interested in investigating whether the temporal coordination of an offglide (/j/ or /w/) patterns with coda consonants, which have shown no temporal shift when another coda consonant was added to the syllable [3]. Following [3], the maximum constriction (peak intensity) in the onset consonant gesture was used as a constant anchor (Anchor) with respect to which the relative timing of the glide/coda gestures of interest was measured. The relative timing of target gesture (C or G) to the anchor was calculated as the lag between the peak velocity (PeakVel) of the closing phase and Anchor (G_Lag: the timing of glide gesture; C_Lag: the timing of consonant coda gesture) (see Fig 3). In the case of glide /j/, which consists of a single palatal gesture (tongue raising), G_Lag was measured as the temporal interval between Anchor and PeakVel of the palatal gesture. For glide /w/, which is composed of a labial (lip approximation) and a dorsal (tongue dorsum backing and raising) gesture, the midpoint between the PeakVel's of the two gestures was used as the reference of the glide gesture, and its G_Lag was measured from Anchor to this midpoint, shown in the third panel of Fig 3.

3. RESULTS

The relative gesture shift was determined by comparing the C_Lag and G_Lag of singletons with corresponding lags in the clusters. To reduce effects from inter- and intra-speaker variability in speaking rate, these temporal lag measures were normalized before subjecting to statistical analysis. Assuming variability in speech rate is encoded in the temporal interval between onset consonant (Anchor) and vowel (Phar PeakVel), which is shared by all tokens, C_Lag's and G_Lag's were separately regressed over the CV timing and transformed as residuals from the predicted value.

Three-Way ANOVAs were separately carried out for each offglide (/j/ and /w/) and for each lag measure (C_Lag and G_Lag). The three factors are: Coda Type (singleton or complex), Position (sentence-final or non-final), and Onset (alveolar or labial). Results suggest a temporal shift of both offglides (G_Lag), as main effects was found for Coda Type in both cases (/w/: $F_{(1,104)} = 40.35$, $p < 0.001$; /j/: $F_{(1,277)} = 74.4$, $p < 0.001$). As can be seen from the means in Table 2, the G_Lags are substantially longer in singleton (Sing) than in complex CG cluster (Comp). This implies that the glide gesture shifts *leftward* towards the vowel in a cluster. Unlike glide gestures, the consonant gesture either shifts rightward (C_Lag in /w/-data: $F_{(1,263)} = 20.54$, $p < 0.001$), or remains unchanged (C_Lag in /j/-data: $F_{(1,372)} = 1.98$, $p = 0.16$).

Table 2: Average G_Lag and C_Lag (raw data) for each condition.

| (ms)		/j/		/w/	
Ons	*Pos*	*Comp*	*Sing*	*Comp*	*Sing*
		C_Lag			
Lab	fin	373	416	488	416
	nf	297	306	390	306
Alv	fin	397	402	476	428
	nf	308	299	385	324
		G_Lag			
Lab	fin	309	414	382	500
	nf	263	297	318	342
Alv	fin	326	425	285	480
	nf	276	323	274	383

4. DISCUSSION

The timing of offglides /j/ and /w/ in English was examined in coda cluster to determine if it is locally timed to the preceding vowel in the presence of a following consonant, as would have been predicted

by the hypotheses of VC-coupling in syllable coda position. Results of the current study diverged from this prediction as we found a significant leftward shift of the glide gestures towards the vowel in a coda cluster, while the following consonant may either shift rightward or maintain similar relative timing to the anchor even in the presence of a glide gesture ahead of it in the same syllable. Accordingly, the VC-coupling hypothesis in syllable coda position is not sufficient in accounting for our finding of VG timing, which suggests the glide gesture bears a closer relationship to the preceding vowel when followed by another coda consonant.

However, this may not necessarily imply that VG sequence is a linguistically meaningful unit and V-G holds an intrinsically different relation from a vowel-coda relation. Arguments for diphthongs as "complex nuclei" tend to come from the phonotactic restrictions held between diphthongs and its following consonant. When a VG sequence is not followed by another coda consonant, there does not seem to be any argument against the glide gesture being structurally analyzed as a syllable coda. Besides, through examining the articulatory correlates of ambisyllabicity in English offglides /w, j/ and liquid /l/, Gick [15] found that all gestures of syllable-final [j] and [w] exhibited final reduction compared with word-initial onglides—a property usually found in coda consonantal gestures. Therefore, it seems a singleton glide gesture can be modelled in the same way as a coda consonant—it is anti-phase coupled to the vowel, and the leftward shift occurs due to a competition with other coda consonants. Based on the observed temporal patterns, we propose to account for this consonant shift in terms of a competitive anti-phase coupling structure in GC clusters, shown in Fig 4.

Figure 4: Proposed gesture organization for post-vocalic GC sequence.
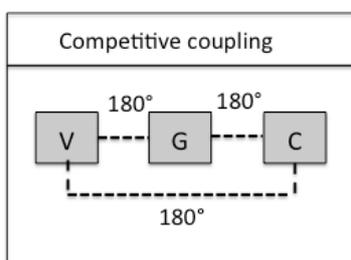
In this hypothesized coupling graph, the glide gesture is coupled anti-phase to the vowel, just like a consonant coda. Nevertheless, when an additional consonant coda enters this coupling relationship, it is coupled anti-phase (180°) to the preceding glide gesture as well as to the vowel gesture. This specification of the additional consonant in a sense captures the unitary phonotactic properties of VG—the coda C is coupled to both V and G, consistent with the fact that possible coda Cs are a function of both gestures in the so-called "complex nucleus."

The coupling specifications between GC and VC are incompatible since VG also demands anti-phase coupling, resulting in competition among the three coupling relations. The glide gesture may therefore shift leftward to compromise these competing demands, as found in the current study. Such a coupling graph also predicts a rightward shift of the consonant gesture, which was found in C_Lag results for /w/-data.

Unlike the case of /w/, the VC timing was preserved whether offglide /j/ was inserted or not. This discrepancy is not unexpected since one crucial difference between /j/ and /w/ is the number of gestures involved. More gestures participating in the proposed competitive coupling relationship entails stronger competition, and therefore significant rightward shift of C in /wC/ was predicted. One the other hand, the stable VC timing regardless of the insertion of offglide /j/ may be modelled as different quantitative specification of coupling strength associated with the graph's edges (following [16]): if the VC coupling is specified with stronger coupling strength compared with the VG and GC coupling, the consonant C will likely to maintain a stable timing relation with respect to the vowel.

4. CONCLUSION

The temporal organization of the offglides [j] and [w] was investigated in this paper. Results showed that a GC coda cluster behaves differently from a CC coda cluster. We propose to account for these temporal patterns as a competition among multiple anti-phase coupling relations specified in GC coda clusters. This proposed model is advantageous since it is typologically plausible—it does not rely on stipulated relations among gestures as anti-phase coupling more has been shown to be one of the most stable coupling modes (anti-phase and in-phase) and available within AP. Finally, the proposal captures both the experimental findings in this paper and the phonotactic properties of VG.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Selkirk, E. 1982. The syllable. In van der Hulst, H. & Smith, N. (eds.), *The structure of phonological representations, part II*. Dordrecht: Foris. 337-83.

[2] Browman, C. P., Goldstein, L. 1988. Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45, 140-155.

[3] Marin, S., Pouplier, M. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14, 380-407.

[4] Browman, C. P., Goldstein, L. 1992. Articulatory phonology: an overview. *Phonetic*a 49, 155-180.

[5] Saltzman, E., Munhall, K. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382.

[6] Goldstein, L. 2011. Back to the past tense in English. In Bravo, R., Mikkelsen, L., & Potsdam, E. (eds.), *Representing language: Essays in honor of Judith Aissen.*

[7] Narayanan, S., Nayak, K., Lee, S., Sethy, A., Byrd, D. 2004. "An approach to real-time magnetic resonance imaging for speech production." *JASA* 109, 2446.

[8] Hammond, M. 1999. *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford University Press.

[9] Bresch, E., Katsamanis, A., Goldstein, L., Narayanan, S. 2010. Statistical multi-stream modeling of real-time MRI articulatory speech data. *Interspeech* Makuhari, Japan, 1584–1587.

[10] Lammert, A., Proctor, M. and Narayanan, S. 2010. Data-Driven Analysis of Realtime Vocal Tract MRI using Correlated Image Regions. *Interspeech* Makuhari, Japan, 1572–1575.

[11] Proctor, M., Lammert, A., Katsamanis, A., Goldstein, L., Hagedorn, C., Narayanan, S. 2011. Direct Estimation of Articulatory Kinematics from Real-time Magnetic Resonance Image Sequences. *Interspeech*, 281-284.

[12] Hsieh, F., Goldstein, L., Byrd, D., Narayanan, S. 2013. Truncation of Pharyngeal Gesture in English Diphthong [aI]. *Interspeech* Lyon, France.

[13] Hagedorn, C., Proctor, M., Goldstein, L. 2011. Automatic analysis of singleton and geminate consonant articulation using real-time Magnetic Resonance Imaging. Interspeech Florence, Italy, 409-412.

[14] Atkeson, C., Moore, A., and Schaal, S. (1997). Locally-weighted learning. AI Review 11: 11-73.

[15] Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In Local, J., Ogden, R., Temple, R. (eds), *Phonetic Interpretation: Papers in Laboratory Phonology IV*. Cambridge University Press. 222-236.

[16] Goldstein, L., Nam, H., Saltzman, E., Chitoran, I. 2008. Coupled oscillator planning model of speech timing and syllable structure. *Proc. Phonetic Conference of China and the International Symposium on Phonetic Frontier 8th*.