ARTICULATORY MODELING OF CONSONANT RELEASE GESTURE

Hosung Nam

Yale University & Haskins Laboratories
hosung.nam@haskins.yale.edu

ABSTRACT

This study examines the dynamics of consonant release motion. The release in syllable onset is shown to exhibit consonant-like speed to reach the rest position but the rest position corresponds to that of a concomitant vowel. We attempt to model such dual status of the release motion in the task-dynamic speech production model by splitting dynamic blending parameter into stiffness and target.

Keywords: articulatory phonology, task dynamics, consonant release

1. INTRODUCTION

In articulatory phonology, speech gestures are atomic units and can be appropriately organized to build up higher level structures such as syllables, words, phrases, etc. Gestures are defined in terms of vocal-tract constriction at articulatory organs (LIPS, TT(tongue tip), TB(tongue body), VEL(velum), GLO(glottis)) and each vocal tract variable involves relevant model articulators. Co-articulation is naturally expressed in time functions of gestural activation, *gestural scores*.

Recent studies have shown that constriction for consonants in syllable onset begins synchronously to tongue body opening for vowels, and proposed structural hypotheses on gestural temporal coordination in which simple phasing dynamics underlies [1][3][4]. In addition, it has been shown that consonants (especially obstruents) could be understood as a sequential combination of closure (CLO below) and release gestures (REL below) [1][3][7]. This study attempts to investigate the dynamic behavior of REL[1] and discuss how to model it to fit to empirical articulatory movements.

2. RELEASE GESTURE OF CONSONANTS

How can the release motion for consonants be understood and modeled in speech production? The task-dynamic model provides a way of understanding and implementing of various gestural patterns [6]. In the model, a gesture is a constricting motion parametrically defined on the tract variable (organ), and distributes its driving influence to associated articulators.

One possible answer is that release motion could be understood as articulators' return to the neutral position. In the model, a neutral shape of vocal tract is defined in terms of the articulators. Each articulator has its own neutral state and returns to it at a defined speed when any tract variable associated with the articulator is not active. Browman [1], however, demonstrated through computational simulations the necessity of explicit REL in addition to CLO for obstruent consonants. The return to neutral configuration for consonantal release without an explicit REL might not be able to ensure sufficient acoustic radiation for the following vowel[2]. In addition, Nam [3] proposed that positing an underlying REL as well as CLO for a single consonant can help account for its asymmetrical phonetic variations between onset and coda.

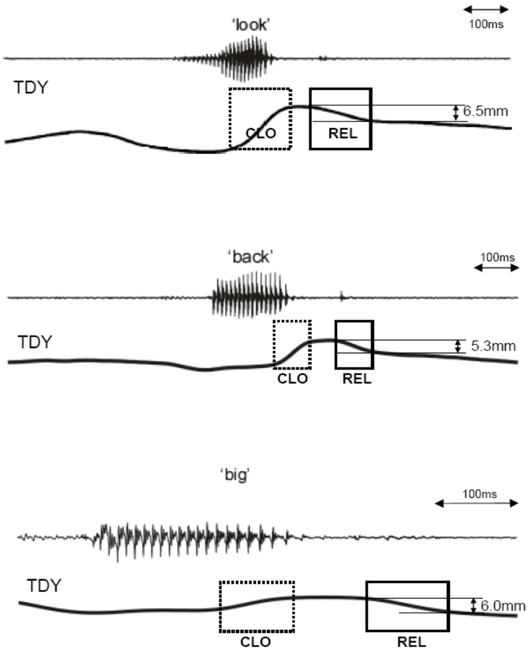
2.1. Constriction target of release gestures

In the task-dynamic model, a constriction gesture[3] is expressed by a simple mass-spring system below:

(1) $$m\ddot{x} + b\dot{x} + k(x - x_0) = 0$$

where x is *a tract variable,* x_0 is the constriction *target* as rest position, m is set equal to 1, b is currently set to critical damping, and k is the *stiffness* term controlling gestural speed to target. We only focus on frequency and target. CLO should no doubt have invariant values for constriction target (e.g. -2 mm for stops and 1 mm for fricatives). Then what is the constriction target for consonant release gesture? Following the intergestural phasing hypothesis [1][3][4], REL in onset is temporally overlapped with vowels whereas that in coda is not. Thus, constriction target for REL can be correctly estimated only in a boundary final position, where no vowel gesture is hypothesized to be active while consonants being articulatorily released. The achievement of release motion is hypothesized to be invariant as not influenced by vowel height at coda position. Here we need to focus on velar consonant release because non-velar (coronal, labial) consonants do not involve TB and thus never overlap in time with vowels involving TB within a tract variable. As illustrated in Fig. 1, the constriction targets of REL for velar stops, /k/ and /g/, are estimated in different vowel contexts (/u/, /ae/, /i/).

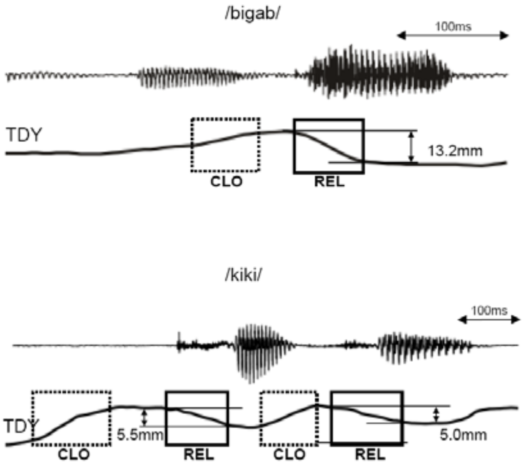
Figure 1: Waveforms and vertical displacements of tongue dorsum (TDY) for 'look', 'back' and 'big'. Dotted boxes are gestural activation interval of CLO and solid boxes are that of REL. (from Wisconsin X-ray microbeam database [8])



Activation interval for each gesture is delimited by measuring velocity below a threshold and represented by a rectangular box. The left end of each box corresponds to the beginning of an action and the right end to completion of reaching a target position. Velar consonants are shown to exhibit consistent release target (6.5, 5.3, 6.0 mm) regardless of the types of tautosyllabic vowels.

As for velar consonants in onset, tongue dorsum is quickly lowered for the release motion after their closure but unlike in the boundary-final position, the release motion exhibits different displacement targets from complete constriction position depending on the types of tautosyllabic vowels. Fig. 2 illustrates waveforms and vertical displacements of tongue dorsum over time for /bigab/ and /kiki/.

Figure 2: Waveforms and vertical displacement of tongue dorsum (TDY) for /bigab/, /kiki/.

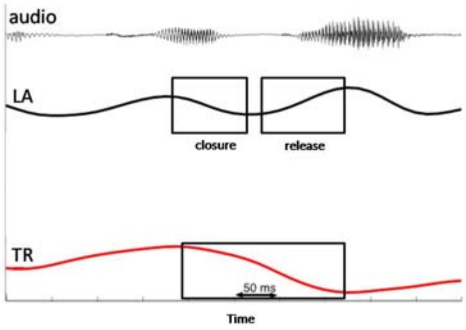


Tongue dorsum is lowered 13.2 mm for /g/ release in /bigab/ but 5.5 mm and 5.0 mm for /k/s in /kiki/. Unlike in coda, the realization of constriction target for REL in onset is variant depending on overlapping vowel gestures. Further, the minima of tongue dorsum shown at the end of REL corresponds to the target position that should be realized for each vowel. However, importantly, the displacement of tongue dorsum for REL delimited as solid boxes is as quick as CLO.

2.2. Stiffness of release gestures

Articulatory data show that a vowel gesture is intrinsically slower than a consonant in achieving its target so it should have a lower frequency. Since the labial consonant /p/ and the unrounded vowels /i, a/ do not share a tract variable, their activation span can be easily delimited on their vocal-tract trajectories of constriction as exemplified in Fig. 3. Here both closure and release actions for /p/ in LA are approximately twice faster than the vowel gesture in TR.

Figure 3: Waveforms and tract variables for /pi**pa**p/ over time by an American English speaker. Gestural activations are delimited for /p/ and /a/ in LA (lip aperture) and TR[4] (tongue root), respectively.



This implies that the release motion should be controlled by the dynamics of the consonant, not the vowel, even though its target is realized identically to the vowel's.

3. MODELING OF CONSONANT RELEASE

In this section, we will discuss how to model velar consonants in onset position, with the same temporal coordination as schematized in Fig. 3 using TADA (task-dynamic speech production model, [5]). It is tested whether the dynamics of syllable-initial velar consonants shown in articulatory data could be correctly modeled in the current version of the task-dynamic model.

3.1. Dynamic blending

The task-dynamic model allows temporal overlap of gestures within a tract variable, which creates competition in their dynamics. Dynamic influences of individual gestures with temporal overlap are weight-averaged at a given time, which is called *blending*. Each gesture can be specified by different blending weights (parameters). Velar consonants in syllable onset are a good example for this issue because the velar consonant and vowel employ the same tract variable, TB. The closure gesture for velar consonants is concomitant with the opening gesture for vowels. TB constriction target for the closure should be less than 0 while that for vowels should be much greater than 0. In the current model, CLO has greater blending parameters than vowels, which implies that CLO is dominant over vowels in realizing the dynamics when it overlaps with vowels. Thus, the closure for velar consonants can be realized at its desired target regardless of the presence of a vowel gesture with a different target.

However, the constriction target for consonant release in onset is more complicated. As described in section 2, the unique dynamics of REL can be summarized twofold: (1) it behaves like consonant as it moves as quickly as CLO whereas (2) when it is overlapped with a vowel within a tract variable, its constriction target corresponds to the vowel's target.

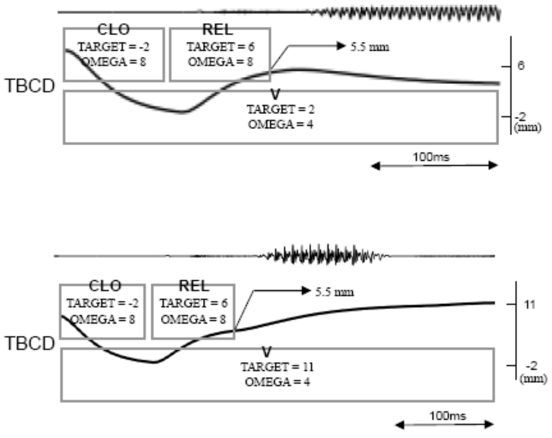
3.1.1. Simulations

Gestural scores for /ki/ in 'key' and /ka/ in 'cop' are modeled in two different blending parameters of consonant release. In simulation 1, REL is dominant over vowels, and in simulation 2, vowels are dominant over REL. For both simulations, TB constriction target of CLO for /k/ is set to -2 mm, which is a virtual target beyond a palate wall and that of REL is set to 6 mm, which approximately corresponds to what is observed in Fig. 1. TB constriction targets desired for the vowels /i/ and /a/ are set to 2 mm and 11 mm, which are estimated target values based on articulatory kinematic data. To reflect the asymmetry (roughly 1:2 ratio) between vowels and consonants in movement velocity as shown in Fig. 3, their frequencies are set to 4 Hz and 8 Hz respectively. To implement the difference in dynamic dominance of REL over vowels, blending parameters are differently specified. In the current task-dynamic model, the blending parameter of vowel is 0.01^5. In simulation 1, the blending parameter of REL is set to 1 to make REL dominant over the vowel. In simulation 2, the blending parameter of REL is set to 0.0001 to make vowels dominant over REL. The tract variables and acoustic outputs are generated using TADA.

3.1.2. Results

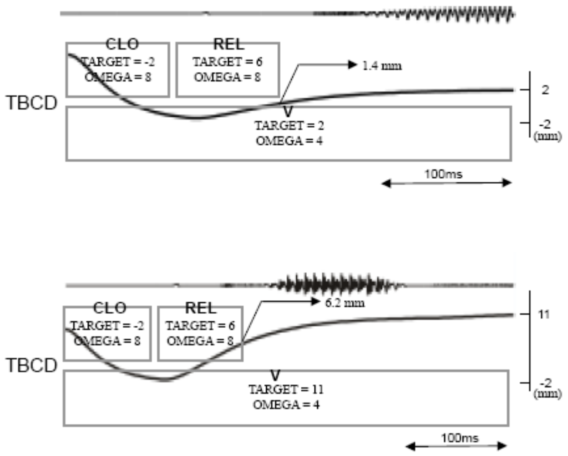
Fig. 4 and 5 illustrate the outputs of simulation 1 and 2, respectively. Each figure contains waveforms and TBCD (tongue body constriction degree) over time for 'key' and 'cop' respectively.

Figure 4: Simulation 1 (REL dominant over V in blending): Input gestural scores (grey boxes) with target and stiffness specified in them, output tract variable TBCD and waveforms for 'key' (top) and 'cop' (bottom).



The gestural scores are input to task-dynamic model, which in turn generates tract variable trajectories, which are superimposed over the input gestural scores. Tract variable trajectory of TBCD is measured at the end of REL activation to see how close TBCD is realized to the desired targets of vowels during the release action. In simulation 1 (Fig. 4), REL does not exhibit variation in TBCD regardless of its vowel context: 5.5 mm for /i/ and /a/ both, which is not empirically the case. Also, as for ‘key’, tongue body moves downward for REL passing the vowel target and upward again for the vowel target, which is rarely observed in empirical data.

Figure 5: Simulation 2 (V dominant over REL in blending): ‘key’ (top) and ‘cop’ (bottom). Other parameter settings are identical to simulation 1.



In simulation 2 (Fig. 5), release motions are realized depending on the ambient vowels: 1.4 mm for /i/ and 6.2 mm for /a/ but they are not enough for the desired vowel targets, 2 mm for /i/ and 11 mm for /a/. This is because of low stiffness of V, which is dominant over REL.

3.2. Split blending parameters

To summarize, REL should realize both its intrinsic stiffness and the ambient vowel’s target when it is overlapped with a vowel within a tract variable. In the current model, the dynamic parameters (target, stiffness, and damping) for each gesture are all coupled and behave together in terms of their blending influences. This study proposes specifying blending parameters separately between target and stiffness as a solution to such dual status of REL. I.e., while REL is dominant over V in stiffness blending, V is dominant over REL in target blending. For a new simulation, the blending parameters are set to 1, 0.01, 100 for V, REL target, and REL stiffness, respectively. Simulation outputs show that the release motions for /k/ are realized at 1.8 mm and 10.2 mm in constriction degree for /i/ and /a/, respectively, which are very close to the desired constriction targets (2 mm and 11 mm) for the vowels.

4. CONCLUSION

This study describes the dual status of the release motion of consonants especially when they are temporally overlapped with vowels in syllable onset position. Further, it is shown that splitting dynamic blending parameter into target and stiffness in the task-dynamic model allows us to correctly model the unique behaviors of the release.

5. REFERENCES

[1] Browman, C. 1994. Lip aperture and consonant releases. *Papers in laboratory phonology III: Phonological structure and phonetic Form:* Patricia Keating (eds.), 331-353. Cambridge: Cambridge University Press.

[2] Browman, C. P. & Goldstein, L. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures, *Les Cahiers de l'ICP, Bulletin de la Communication Parlée,* vol. 5, pp.25-34.

[3] Nam, H. in press. A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. *Laboratory phonology VIIII*

[4] Nam, H. & Saltzman, E. 2003. A competitive, coupled oscillator of syllable structure. *Proceedings of the XIIth International Congress of Phonetic Sciences* (3): 2253-2256.

[5] Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. 2004. TADA: An enhanced, portable Task Dynamics model in MATLAB. Acoustical Society of America, New York, *Journal of the Acoustical Society of America*, 115(5,2): 2430

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

[7] Steriade, D. 1993. Closure, release, and nasal contours. Nasals, nasalization, and the velum, *The phonetics of nasal phonology: theorems and data.* Marie K. Huffman, and Rena A. Krakow (eds.) (Phonetics and Phonology Series 5), 401-470.

[8] Westbury, J. 1994. *X-ray microbeam speech production database user’s handbook Version 1.0* Madison: University of Wisconsin

---

[1] We focus on the dynamics of velar consonants exhibiting a gestural overlap with a vowel. The velar consonants, which share an organ (tongue body: TB) with vowels, could generate greater complexity.

[2] This is because the dynamic stiffness of passive motion such as the return to neutral position is in general considered lower than that of active motion.

[3] *x* can be either constriction degree or location. This study only focuses on constriction degree.

[4] Tongue root (TR) can be employed to find activation interval for some vowels such as /i/.

[5] The values used for blending are relative to each other.