INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Velum control for oral sounds

Reed Blaylock [1], Louis Goldstein [1], Shrikanth Narayanan [1]

[1] University of Southern California

reed.blaylock@gmail.com, louisgol@usc.edu, shri@sipi.usc.edu

Abstract

Velum position during speech shows systematic variability within and across speakers, but has a binary phonological contrast (nasal and oral). Velum lowering is often thought to constitute an independent phonological unit, partly because of its robust prosodically-conditioned timing during nasal stops. Velum raising, on the other hand, is usually considered to be a non-phonological consequence of other vocal tract movements. Moreover, velum raising has almost always been observed in the context of nasals, and has rarely been studied in purely oral contexts. This experiment directly contrasts velum movement in oral and nasal contexts. The results show that temporal coordination of velum raising during oral stops resembles the temporal coordination of velum lowering during nasals, suggesting that velum position and movement are controlled for both raising and lowering. The results imply that some revisions to the Articulatory Phonology model may be appropriate, specifically with regards to the treatment of velum raising as an independent phonological unit.

Index Terms: speech production, Articulatory Phonology, phonetics

1. Introduction

In speech production, the movements and postures of the vocal tract articulators are either contrastive or non-contrastive. For instance, oral and nasal sounds contrast in velum height: the velum is low for nasal sounds to allow air to pass through the nasal cavity, and high for oral sounds. But while all phonological contrasts manifest in distinctive patterns of articulatory activity, there are also regular differences in articulator activity that are not contrastive at all. For one example, the jaw is relatively low for vowels and high for obstruents, but jaw height is not itself a contrastive feature here [1]. Similarly, in oral vowels, the velum is regularly higher for high vowels and lower for low vowels [2], [3]. Though the velic port is still closed (no nasal airflow) in both contexts, it still shows a regular difference in position; despite this regularity, velum height is not a contrastive feature for vowel production.

Moll and Daniloff [4] found that the velum was not fully raised for vowels in NVC syllables, indicating either coarticulatory effects of the nasal consonant or a particular velum height for the vowels. They also found evidence of early velum lowering in CVN and CVVN syllables, which they interpreted as anticipatory lowering for the velum. However, Bell-Berti and Krakow [5] rejected this interpretation after observing velum lowering for vowels unrelated to the lowering for nasals; in addition, they showed evidence of multi-stage velum lowering in long pre-nasal vowels, interpreting these as separate lowerings for the vowel and nasal [6].

Bell-Berti et al. [7] confirmed previous findings that the highest velum positions were for obstruents, with lower positions for high vowels, then low vowels. They observed the height of the velum in non-word CVCCVC utterances like “fipmip” and “fambap”, each of which contained a nasal consonant in medial position. They reported that the overall height of the velum was lower in non-words containing low vowels and higher in non-words containing high vowels. In a subsequent study, Bell-Berti [3] used a fiberoptic camera inserted through the nasal passage to track velum motion for a single English speaker saying /itsa/ and /ista/. Similar to the previous study, it was found that velum position was lower near low vowels.

The study reported in this paper differs from much of the previous literature because it focuses on velum height during utterances that only contain oral sounds.

If velum raising during oral stops is actively controlled, then it raises the question of whether it is controlled in a fashion that is parallel to how velum lowering is controlled. Velum lowering for nasal stops has been modeled in Articulatory Phonology with an independent velum lowering gesture that can be coordinated to the oral constriction gesture [8], [9]. This can capture the relative independence of the velum gesture both in coordination in speech production and as a phonological unit.

Velum lowering does seem to be independent from oral constriction gestures it is associated with. In nasal stops, the timing of velum lowering is heavily influenced by the nasal’s prosodic location. In onset position, the velum lowers roughly synchronously with the oral closing movement; in coda position, the velum lowers well in advance of the oral closure movement. The result is variable timing that depends on the nasal’s position in the syllable, and on the strength of nearby prosodic boundaries—more lag for coda nasals in phrase-final position than in just syllable-final position [8], [10], [11], [12]. Moreover, Goldstein et al [13] showed that independent velum lowering is a possible speech error—in repeated elicitations of “bang bad” and “kim kid”, velum lowering sometimes occurred erroneously during the tongue tip gesture for /d/, without an accompanying tongue dorsum gesture.

Similarly, velum lowering appears to have an independent phonological status. For instance, [14] shows that VN sequences can become nasalized Ṽ when listeners attribute strong coarticulatory vowel nasalization to a nasal feature linked to the vowel, rather than to the following nasal stop. Historically, this has often led to a systematic sound change in which the velum lowering gesture is re-analyzed onto the vowel, and the oral gesture for the nasal coda disappears.

http://dx.doi.org/10.21437/Interspeech.2016-1408

Nasal harmony [15], [16] is another phonological phenomenon in which velum lowering works independently. In these cases, the velum lowering gesture becomes longer, nasalizing nearby segments, whereas the oral gestures originally accompanying the velum lowering remain the same length.

This behavior of the velum can be accounted for within Articulatory Phonology by treating velum lowering as its own gesture. Articulatory Phonology hypothesizes [17] a tight (close to one-to-one) relationship between phonological units and speech production tasks. The task of velum lowering creates contrastive nasality, and this nasality can be coordinated variably with tongue and lip constriction gestures.

The situation for velum raising for oral stops is quite different. Velum raising is not phonologically independent. It does not spread in harmony processes like velum lowering does. There are no sound changes in which velum raising persists when an oral constriction is lost. In Articulatory Phonology, this would lead to the prediction that velum raising is not a separate task, and this would not be expected to exhibit timing variation with respect to its oral constriction.

Instead, velum raising could be modeled as an articulator in the task of stop production—perhaps a global task for building pressure, as suggested by [18]. In this type of task, the velum would be just one of the articulators working as part of a coordinative structure to allow air to build up in the mouth in anticipation of an explosive release. Rather than working as an independent gesture like in nasals, the velum movement would result from the deployment of that coordinative structure and would not be expected to show temporal independence from oral stop constriction. Compare this with the Lip Aperture closing task for /b/: the jaw, lower lip, and upper lip all move in fixed coordination with each other. While the lips may not move at the same rate as the jaw or start/end their motion at the same time, the timing among these three articulators does not vary: the jaw always slightly precedes the lips [19], [20], [21]. The lip closure for /b/ in word-final position has the same timing as it does in coda position, and the timing in syllable-final position is the same as in phrase-final position. If the velum is recruited as an articulator for a global stop goal, then the same invariant timing relationship could hold between velum motion and tongue tip motion.

This experiment tests whether the velum raising exhibits this kind of fixed coordination or whether it functions like an independent gesture (like velum lowering), exhibiting the same sort of variable timing in oral stops that has been found in nasal stops (and in laterals [10], [11], [22]). For instance, velum raising could precede the tongue tip constriction for a /t/ in coda position, but be achieved synchronously with the tongue tip in onset position. Moreover, as the strength of the nearby prosodic boundary increases (e.g. a phrase boundary), small delays between velum and tongue tip motion should be exaggerated [11], [23]. On the other hand, if the velum is recruited as an articulator for a more global goal, one would expect to observe timing relations similar to other coordinative structures.

- Hypothesis 1: the velum has an independent gesture/goal during oral stop production
- Hypothesis 2: the velum moves as part of the global coordination for oral stop production

Hypothesis 1 predicts different timing relationships in onset and coda positions, which become stronger differences at stronger prosodic boundaries. Hypothesis 2 predicts that the timing of the velum and the tongue will not change with prosodic conditions.

2. Methods

This experiment used a real-time Magnetic Resonance (rtMR) imaging technique to capture the movements of the vocal tract articulators on video in real time. Subjects were two native English speakers from the University of Southern California community [24]. Subjects spoke lying down (face up). Airways were imaged in the midsagittal plane with a spatial resolution of 68x68 pixels (200mm x 200mm) and slice thickness of 5mm. The reconstructed frame rate for each video was 22.40 frames per second. To compensate for reduction in coil sensitivity for pixels farther away from the coil, pixel intensity in each video was normalized [25].

Two native speakers of American English produced coronal stops (/d/) and nasals (/n/) in onset and coda position. The elicitations were of a 2x2x2 design: two vowel contexts (/e/ and /o/), two stop types (oral and nasal), and two prosodic conditions (word initial and word final). Each subject repeated the eight sentences seven times for a total of 56 utterances (though not all were measurable). Each set of eight was recorded at once, into a single video, with each sentence shown to the subject one at a time. The order was always as follows:

1. Type "paid OVER" slowly.
2. Type "pay DOVER" slowly.
3. Type "pain OVER" slowly.
4. Type "pay NOVA" slowly.
5. Type "paid ALE" slowly.
6. Type "pay DALE" slowly.
7. Type "pain ALE" slowly.
8. Type "pay NAIL" slowly.

A region of interest technique was used to track articulator movement over time [26]. First, a dynamic programming algorithm established a vocal tract mid-line by selecting pixels with the highest standard deviation across time. (The mid-line was occasionally corrected manually). This method ensures that regions placed along this line will capture as much pixel intensity fluctuation as possible [25]. Pseudo-circular regions were placed over the location of tongue tip constriction and raised velum. The average pixel intensity inside the region was calculated for each frame. Tracking the pixel intensity in the region across frames yields a time series of constriction formation and release (or, equivalently, articulator motion), where higher pixel intensity indicates a greater amount of tissue in the region [25]. The pixel intensity time series were smoothed using a locally weighted linear regression technique with a relatively tight kernel width of h=.9 [26], [27]. Because of the low frame rate, the number of data points output by smoothing was three times the input number (i.e. two extra data points were linearly interpolated between each frame after smoothing).

The find_gest algorithm of [28] was used to find frames of onset and maximum constriction. The onset of motion was calculated as the frame in which constriction velocity had reached 10% of its maximum velocity (ONS) [26]. For this study, the algorithm was used to find the frames of onset and maximum constriction for the tongue tip and velum during the stops (velum lowering for nasal stops, velum raising for oral stops).

3. Results

This study measured the temporal lag between velum and tongue tip constrictions in onset and coda conditions. Three temporal landmarks were compared:

- Onset: when the vocal tract articulator began to move (ONS)
- Target attainment: when the constriction was mostly formed (NONS)
- Maximum constriction: the apex of the constriction (MAXC)

Velic landmark times were subtracted from coronal landmark times. Therefore, positive lag indicates that a coronal landmark follows a velic landmark, and negative lag indicates that a coronal landmark precedes a velar landmark. Some landmarks had zero lag, indicating that the velum and tongue tip achieved that landmark at the same time.

A simple sign test (significance: $p <= 0.05$) was performed over both subjects combined to see whether velum motion landmarks preceded, were synchronous with, or followed tongue tip motion landmarks. The results of an ANOVA indicated that the effect of speaker was not significant in any condition, so the results presented are of both speakers together (all $p > 0.05$). Values followed by asterisks are significant (Tables 1 and 2).

Onset position			
Timing	*ONS*	*NONS*	*MAXC*
Coronal precedes	17	22*	21*
Coronal follows	7	2	3
Simultaneous	3	3	3
Coda position			
Timing	*ONS*	*NONS*	*MAXC*
Coronal precedes	6	3	2
Coronal follows	17*	20*	20*
Simultaneous	0	0	1

Table 1. Temporal lag for /n/ in onset and coda

Onset position			
Timing	*ONS*	*NONS*	*MAXC*
Coronal precedes	20*	16*	15*
Coronal follows	0	4	4
Simultaneous	0	0	1
Coda position			
Timing	*ONS*	*NONS*	*MAXC*
Coronal precedes	12	4	5
Coronal follows	6	14	13
Simultaneous	3	3	3

Table 2. Temporal lag for /d/ in onset and coda

For onset nasal stops, the tongue tip preceded the velum in target attainment and maximum constriction goals, though the difference was not significant for movement onset (see Table 1). For coda nasals, the velum tended to precede the tongue tip at every landmark. This is more or less consistent with the findings from previous studies.

For oral stops in onset position, the tongue tip preceded the velum at all landmarks (see Table 2). In coda position, no difference was found between velum and tongue tip landmarks, suggesting that they moved synchronously. However, there does appear to be a trend toward early velic movement for the target attainment and maximum constriction landmarks; a larger sample size might accentuate this trend.

The sign test indicates whether one movement tended to precede or follow the other, but does not offer a comparison between lags in onset and coda position. A two-way ANOVA was run over all the oral stop data, testing for differences in vowel quality and syllable position. In each case, syllable position was a significant predictor of lag time ($p < 0.01$), but vowel quality was not ($p > .5$).

3.1. Velum ONS is later in onset than in coda

The sign test suggests that in onset position, the beginning of velum movement is later than beginning of tongue tip movement, whereas in coda position, the beginning of velum movement is not significantly different from beginning of tongue tip movement. The results of an ANOVA confirm that the beginning of velum movement with respect to the onset of the coronal constriction is later in onset position than in coda position (Figure 1), $F(1, 38) = 30.79$, $p < 0.001$.

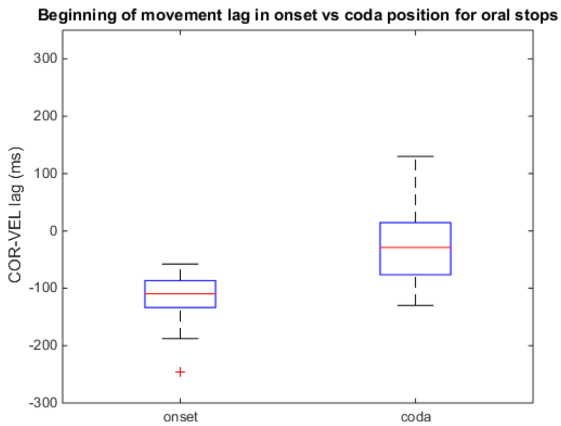

Figure 1: Oral ONS lag in onset and coda

3.2. Velum MAXC is later in onset than in coda

The sign test suggests that in onset position, the velic maximum constriction is later than tongue tip maximum constriction, whereas in coda position, the velic maximum constriction is not significantly different from tongue tip maximum constriction. The results of an ANOVA confirm that velic maximum constriction in onset position is later with respect to coronal maximum constriction than in coda position (Figure 2), $F(1, 38) = 8.21$, $p = 0.0067$.

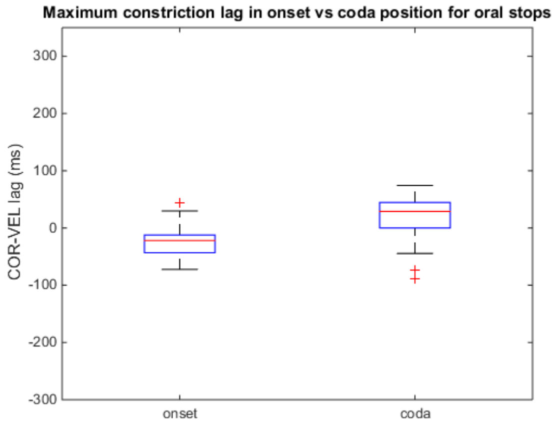

Figure 2: Oral MAXC lag in onset and coda

3.3. Velum NONS is later in onset than in coda

The sign test suggests that in onset position, velic target attainment is later than tongue tip target attainment, whereas in coda position, velic target attainment is not significantly different from tongue tip target attainment. The results of an ANOVA indicate that velic target attainment with respect to coronal target attainment in onset position is later than in coda position (Figure 3) $F(1, 38) = 9.29$, $p = 0.0042$.

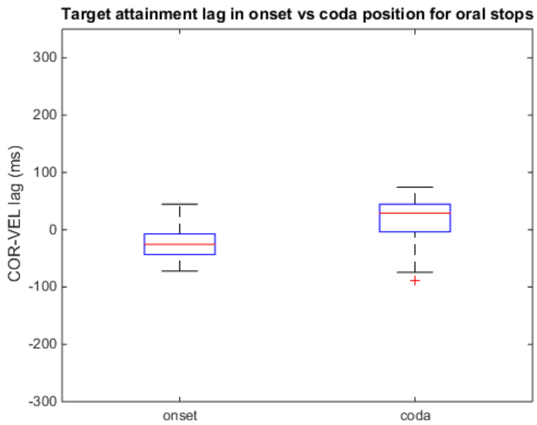

Figure 3: Oral NONS lag in onset and coda

4. Discussion

The oral stop measurements provide some evidence for Hypothesis 1—that the velum has an independent gesture during oral stops. The patterns found in oral and nasal stops are qualitatively similar: the significant delay of the velum in onset position and lack of delay in coda position (and, indeed, a trend toward early velum movement) suggests two different timing patterns based on syllable structure position; this was the prediction of Hypothesis 1.

Asserting that velum raising is an independent gesture for oral stops does not necessarily preclude a more general global goal. The median lag between tongue tip and velum target achievement and maximum constriction landmarks was closer to zero for oral stops than for nasal stops—that is, there was a trend for oral stops to have inter-gestural lag with smaller magnitude than the lag for nasals. This makes sense from the perspective of a pressure-oriented goal: if the velum reaches its goal too late, the stop could have insufficient burst; if the velum reaches its goal too early, it might also release too early and reduce the oral pressure. Even though the velum gesture is independent, it is still constrained to a fairly tight timing relationship with the tongue tip gesture. In Articulatory Phonology, this could be conceptualized as a strong coupling relationship [29], [30].

This experiment implies that some revisions may be appropriate in the Articulatory Phonology model. At this time, the velum gestures for oral stops do not have their own clock—they are always programmed to be synchronous with the oral constriction. The set of tasks for vowels does not include a velum gesture at all. Future modeling should consider incorporating velum control.

The notion of uncontrastive regularity runs counter to the one-to-one relationship between tasks and phonological units in Articulatory Phonology. If a high velum position is controlled independently, just as a low velum position is controlled for nasals, then there should be typological evidence of velum raising behaving as a phonological unit. As discussed earlier, there is no evidence of this.

The decreased magnitude of constriction landmark lag time in oral stops may offer a hint towards a compromise. The one-to-one hypothesis relies on contrasts being potentially distinguishable to the listener. With velum lowering, the relatively large lag time and consequent phonetic effect is easy for listeners to pick up. For velum raising, though, the differences—while robust and regular—are small and mostly imperceptible. It may be that velum raising is not phonologized simply because listeners never hear it change. In that case, the regular differences in timing might be related to a more ubiquitous pulling of secondary articulations toward syllable nuclei [12].

5. Conclusions

This study offers a rare look at velum movement in purely oral contexts. The control hypothesis is strengthened by evidence that the temporal coordination of velum raising during oral stops quantitatively matches the temporal coordination of velum lowering during nasals. Future work will model these findings to learn more about how the velum is and can be controlled and represented in phonological planning.

6. Acknowledgements

This work was supported by NIH grant RO1 DC007124.

7. References

[1] Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology I*, *26*, 451-470.

[2] Fritzell, B. (1969). The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. *Acta Oto-laryngologica*. Supplement 250.

[3] Bell-Berti, F. (1980). Velopharyngeal function: A spatial-temporal model. *Speech and language: Advances in basic research and practice*, *4*, 137-150.

[4] Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *The Journal of the Acoustical Society of America*,*50* (2B), 678-684.

[5] Bell-Berti, F., & Krakow, R. A. (1991). Anticipatory velar lowering: A coproduction account. *The Journal of the Acoustical Society of America*, *90*(1), 112-123.

[6] Boyce, S. E., Krakow, R. A., Bell-Berti, F., & Gelfer, C. E. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics*, *18*(2), 173-188.

[7] Bell-Berti, F., Baer, T., Harris, K. S., & Niimi, S. (1979). Coarticulatory effects of vowel quality on velar function. *Phonetica*, *36*(3), 187-193.

[8] Browman, C. P., & Goldstein, L. (1995). Gestural syllable position effects in American English. *Producing speech: Contemporary issues*, 19-33.

[9] Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-249.

[10] Krakow, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. University Microfilms.

[11] Krakow, R. (1999). Physiological organization of syllables: a review. *Journal of Phonetics*, *27*, 23–54.

[12] Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, *37*, 97–110.

[13] Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386-412.

[14] Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785-821.

[15] Cohn, A. (1990). *Phonetic and phonological rules of nasalization.* UCLA PhD dissertation. Distributed as *UCLA Working Papers in Phonetics*, 76.

[16] Walker, R. (1999). Guarani voiceless stops in oral versus nasal contexts: an acoustical study. *Journal of the International Phonetic Association* 29, 63-94.

[17] Browman, C. P., & Goldstein, L. M. (1991). Gestural structures: Distinctiveness, phonological processes, and historical change. In *Modularity and the motor theory of speech perception: proceedings of a conference to honor Alvin M. Liberman* (pp. 313–338).

[18] Mattingly, I. G. (1990). The global character of phonetic gestures. *Journal of Phonetics*, *18*, 445–452.

[19] Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, *65*(1), 156-166.

[20] Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *The Journal of Neuroscience*, *8*(12), 4628-4639.

[21] Kollia, H. B. (1994). Functional organization of velar movements following jaw perturbation. PhD Dissertation, The City University of New York.

[22] Sproat, R., & Fujimura, O. (1993). Allophonic variation in American English /l/ and its implications for phonetic implementation. *Journal of Phonetics*.

[23] Vaissière, J. (1988). Prediction of velum movement from phonological specifications. *Phonetica*, 45(2-4), 122-139.

[24] Shrikanth S. Narayanan, Krishna S. Nayak, Sungbok Lee, Abhinav Sethy, Dani Byrd. (2004). An approach to real-time magnetic resonance imaging for speech production. Journal of the Acoustical Society of America, vol. 115, no. 4, pp. 1771-1776.

[25] Lammert, A., Ramanarayanan, V., Proctor, M., & Narayanan, S. (2013). Vocal Tract Cross-Distance Estimation from Real-Time MRI using Region-of-Interest Analysis. *Interspeech*, Lyon, France, pp. 959–962.

[26] 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*, Florence, Italy, pp. 281–284.

[27] Lammert, A., Goldstein, L., Narayanan, S., & Iskarous, K. (2013). Statistical methods for estimation of direct and differential kinematics of the vocal tract, Speech Communication, *55*, 147–161.

[28] Tiede, M. (2010). "MVIEW: Multi-channel visualization application for displaying dynamic sensor movements." In development.

[29] Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209-244.

[30] 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*, (5), 25-34.