Systematic Variation in the Articulation of the Korean Liquid across Prosodic Positions

Yoonjeong Lee[1], Louis Goldstein[1], Shrikanth Narayanan[2]

[1]Department of Linguistics, University of Southern California, USA;
[2]Department of Electrical Engineering, University of Southern California, USA
yoonjeol@usc.edu, louisgol@usc.edu, shri@sipi.usc.edu

ABSTRACT

This study examined articulatory composition of the Korean liquid in the phonological contexts that condition its allophony using real-time MRI. During the liquid, tongue tip constriction, tongue body raising and tongue root fronting motions were observed. Inter-vocalic liquids (flap percept) were produced with shorter tongue tip movement duration and smaller displacement than onset or coda liquids (lateral percept). Moreover, tongue root fronting during the liquid was associated with the smaller displacement in the flap contexts (inter-vocalic) than the lateral contexts (onset and coda). Interestingly, there was *no* tongue body raising observed in the flap context, suggesting that the allophony between lateral (onset and coda) and flap (inter-vocalic) is not only attributable to the overall gestural reduction, but also to a categorical distinction in gestural composition. This compositional difference challenges phonetic/phonological models that view position-sensitive allophony as resulting exclusively from dynamic variation of the abstract component gestures.

Keywords: Articulation, Korean liquid, Syllable positions, Gestural composition, Real-time MRI

1. INTRODUCTION

Speech is made up of discrete articulatory events unfolding over time. Phonetic/phonological models that take this dynamic characteristic of speech into account view continuous changes in the speech output as lawful consequences of the dynamical systems controlling its production [2]. These models invite questions about what kind of invariant phonological gestures compose an utterance and how reduction and overlap with adjacent gestures in time and space result in context-dependent variations in acoustic or perceptual properties of utterances [3,4,6].

This study examines position-sensitive realization of Korean liquid, which is less well understood in the articulatory domain. Korean purportedly exhibits a single liquid phoneme that shows multiple allophones alternating between lateral and rhotic articulations. Traditionally, this liquid allophony has been described as a single phoneme with an underspecified feature at a phonological level, which becomes a flap [ɾ] syllable initially, but a lateral [l] syllable finally [8]. To date, there has been only one articulatory study on the Korean liquid [14] and it still remains unclear what articulatory characteristics drive the lateral vs. flap percepts of this single liquid phoneme.

To shed light on articulatory composition of the Korean liquid in different phonological contexts, this study sets out a real-time MRI experiment. Based on previous work on laterals in other languages and from the previous articulatory report by Oh and Gick [14], this study hypothesizes that 1) there will be at least two underlying phonological gestures comprising the single liquid phoneme—tongue tip and tongue body gestures in every context; 2) The underlying phonological gestures will be influenced by phonological contexts (i.e., syllable positions) in such a way that the articulatory variability in time and space leading to the lateral [l] and rhotic [ɾ] percepts is the outcome of gestural reduction. From previous work on English laterals [4,17], we know that although the tongue tip gesture of a liquid can be quite reduced in phonologically weak positions (e.g., inter-vocalically), the liquid with the reduced tongue tip gesture still produces a lateral percept. Therefore, what drives the flap percept instead of a lateral might have to do with the *tongue body* gesture of the liquid that is reduced due to context. We expect that *both* tongue tip and tongue body gestures for the liquid will be produced with shorter duration in inter-vocalic position (flap percept) than in syllable onset or coda positions (lateral percept). In connection with this, inter-vocalically, we also expect reduced magnitude of the tongue tip and tongue body gestures attributable to the temporal reduction (i.e., gestural overlap and reduced activation interval).

2. METHODS

2.1. Subjects and stimuli

Two male Seoul Korean speakers (JK1, JK2) participated in this study.

Disyllabic pseudo-words were designed to elicit a range of liquid (**L**) articulation in different syllable positions (see Table 1). These included onset position ([V#**L**V], where # indicates an accentual phrase boundary), inter-vocalic position ([V**L**V]), and (pre-consonantal) coda position ([V**L**pV]). The liquid was always embedded in V**L**(p)V sequences, in which the flanking vowels were one of the 5 peripheral vowels in Korean (/a/, /ɛ/, /i/, /o/, /u/) and were identical to each other (V=V). Each V**L**(p)V sequence was placed in the middle of the carrier sentence [tɕa, ___ hɛpoa] (Let's ___ say-imperative; presented visually in Korean). Note that for the onset condition (V#**L**V) the carrier sentence was modified by substituting "tɕa" with one of the 5 vowels so that the vowel put before the accentual phrase matched the other flanking vowel in the following syllable (e.g., [**o**, l**o**po hɛpoa] for /o/). Without any specific instructions on phrasing, the speakers always put a major phrase boundary immediately preceding the test word (Intonational Phrase boundary auditorily), and each test word formed an accentual phrase. In total, 180 tokens were collected and analysed (3 syllable position conditions x 5 vowels x 2 speakers x 6 repetitions).

Table 1: A list of test words with V**L**(p)V sequences in IPA.

vowel	V#**L**V [#onset]	V**L**V [inter-vocalic]	V**L**pV [coda]
a	(a)#**l**apa	a**ɾ**a	a**l**pa
ɛ	(ɛ)#**l**ɛpɛ	ɛ**ɾ**ɛ	ɛ**l**pɛ
i	(i)#**l**ipi	i**ɾ**i	i**l**pi
o	(o)#**l**opo	o**ɾ**o	o**l**po
u	(u)#**l**upu	u**ɾ**u	u**l**pu

2.2. MRI data acquisition

MR image and audio data were acquired using an MRI protocol developed for research on speech production [13]. During scans, speakers lying supine read the test sentences that were presented on a back-projection screen, which they could read from within the scanner through a mirror. Each sentence was presented one at a time. Subjects were imaged in the mid-sagittal plane with a 9-interleaf spiral sequence (TR = 6.028 ms, Field of view = 200 x 200 mm, flip angle = 15^{o}), a resolution of 68 x 68 pixels with 2.9 mm width, and a reconstruction rate of 33.18 frames/s. Audio was simultaneously recorded at a sampling frequency of 20 kHz inside the MRI scanner and was noise corrected [1].

2.3. MRI data analysis

Tongue movements were estimated using changes in the average pixel intensity in speaker-specific regions-of-interest of the vocal tract (ROIs: tongue tip [TT], tongue body [TB] and tongue root [TR]), located along an automatically derived midline of the airspace (see Fig. 1) [11]. To minimize noise or random intensity fluctuations, all resulting signals were smoothed by a locally weighted linear regression [10]. The weighting function used was a Gaussian kernel K with a standard deviation of *h* samples, where *h* = .8, giving a smoothing window width of roughly 90 ms given the sampling period of 30.14 ms.

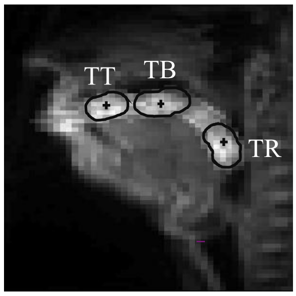

Figure 1: Mean image with correlated ROI locations overlaid (JK1).

Temporal landmarks for TT were algorithmically identified using the velocity of a manually located measurement window (algorithm by Mark Tiede at Haskins Laboratories) on the mean-pixel-intensity-based contours. These landmarks included the time points of movement onset (referred to as *V1* in Fig. 2), constriction onset (target achievement), constriction maximum (*L* in Fig. 2), constriction offset, and movement offset (*V2* in Fig. 2). Movement onset is defined as the point where the velocity first crosses the +/–10 percent threshold of the first peak velocity. Target achievement is defined by locating the point where the velocity falls below the same percent threshold. Constriction maximum is defined by identifying the zero-crossing point in the velocity signals. Constriction offset is defined as the first threshold-crossing point before the second peak velocity. Movement offset is defined as the point where the velocity falls below the same +/–10 percent threshold at the right edge of the window.

TT measurements included movement duration (time between movement onset and constriction offset; referred to as (a) in Fig. 2), and movement displacement (mean pixel intensity difference between the points of onset and constriction maximum; (b) in Fig. 2). Unlike TT, temporal landmarks for TB and TR during the liquid were hard to identify because their movements were often too small. To test whether there is any tongue body raising or tongue root fronting during the liquid in the various syllable contexts, the TB and TR intensity values were measured at the points in time corresponding to the three TT landmarks—i.e., *V1* (at TT movement onset), *L* (at TT constriction maximum) and *V2* (at TT movement offset). Based on the obtained values, TB raising displacement and TR advancement displacement were calculated as the difference between intensity values at *V1* and *L*.

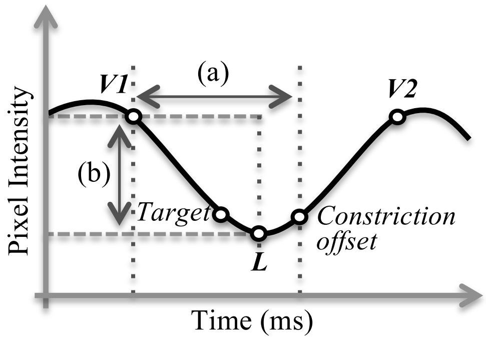

Figure 2: Temporal landmark locations.

3. RESULTS

Statistical evaluation was made based on an analysis of variance (ANOVA) for each subject independently. When significant variation among conditions was detected, post-hoc comparisons using the Tukey's Honest Significant Difference were conducted on all possible pairwise contrasts.

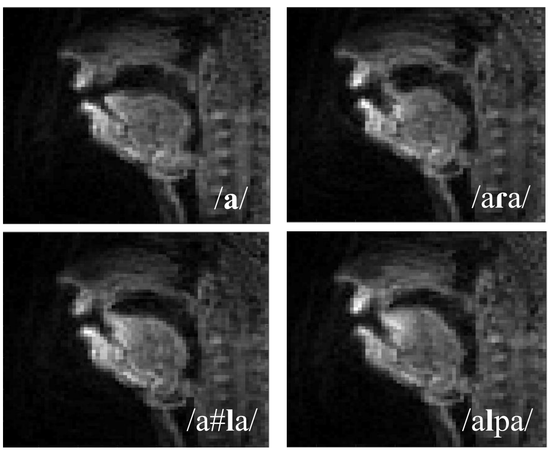

Figure 3: An example set of images taken from the frame of TT constriction maximum during /**a**/, /a**r**a/, /a#**l**a/ and /a**l**pa/ (JK1).

For both speakers, inter-vocalic liquids were produced with shorter TT movement duration compared to onset or coda liquids (all at *p*<.001). In terms of TT displacement, the mean pixel intensity difference was largest in coda position, intermediate in onset position and smallest inter-vocalically (all at *p*<.001). Thus, inter-vocalic liquids (flap percept) were produced with the reduced tongue tip gesture compared to onset or coda liquids (lateral percept) (see Fig. 3), and more tongue tip displacement was associated with coda laterals than onset laterals (compare /a**l**pa/ with /a#**l**a/ in Fig. 3).

As illustrated in Fig. 3, while a tongue tip constriction gesture is always present for the liquid, the tongue body/root shape for the liquid is largely similar to the vowel. Nonetheless, for the lateral context (/a#**l**a/ and /a**l**pa/), some tongue body raising or tongue body/root fronting can be observed, suggesting some tongue body contribution to lateral production, but much less, or perhaps none, is observed in the flap context (/a**r**a/).

In order to determine whether any TB raising at all was observed for the liquids, the pixel intensity values of TB were compared at time points *V1*, *L* and *V2* (Fig. 4) in an ANOVA. In the onset (V#**L**V) or coda (V**L**pV) contexts, TB raising intensity values significantly differed at the three measurement points (all at *p*<.05; dashed/dotted lines in Fig. 4). However, inter-vocalically (V**L**V), there was no effect of different measurement points on TB raising intensity values (presented as solid lines in Fig. 4). This indicates that there is *no* TB raising associated with the flap articulation. The reader is referred back to Fig. 3 for the tongue body height comparison between /**a**/ and /a**r**a/.

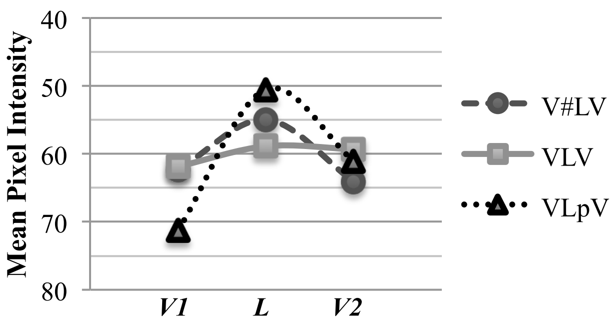

Figure 4: Mean pixel intensity values of TB raising during *V1*, *L*, *V2* (JK1).

The pixel intensity values of TR compared during *V1*, *L* and *V2* revealed that some tongue root fronting motions were always present during the liquid in every context (figure not given, all at *p*<.05). As was the case with the tongue tip, the tongue root movement during the liquid was significantly reduced in the inter-vocalic liquid context as reflected in the reduced difference between TR values at *V1* and *L* for the inter-vocalic context (all at $p<.05$). In addition, the gestural magnitude, measured as the difference between *V1* and *L*, systematically differed between onset and coda positions for TB and TR (as it did for TT, discussed above). Larger displacement of all the relevant gestures was associated with liquids in coda positions compared to those in onset positions (all at $p<.05$). This finding is consistent with the previous acoustic reports on the Korean liquid: a more stable lateral percept is found in coda position than in onset position [9,16]. Taken together, our results showed that liquid production varied dynamically with different phonological contexts, which was attributable to gestural reduction.

4. DISCUSSION

Overall, our results confirmed that Korean liquid production varied dynamically with different syllable position contexts. The lateral (onset and coda contexts) and the flap (inter-vocalic contexts) articulations were different in that there was a substantial gestural reduction of tongue tip (as reflected in both movement duration and magnitude measures) as well as tongue root (displacement measure). We attribute the inter-vocalic flap articulation in part to gestural reduction conditioned by phonological contexts of the overlapping gestures with conflicting production goals [7,15].

However, the most dramatic difference between lateral and flap was that there was *no* tongue body raising at all in the flap context. Thus, our study suggests that the allophony between lateral (onset and coda) and flap (inter-vocalic) is not only attributable to the overall tongue tip and tongue root gestural reduction, but also to a categorical distinction in gestural composition. This compositional difference challenges phonetic/ phonological models that view position-sensitive allophony as resulting exclusively from dynamic variation of the abstract component gestures [4]. Of course, the lack of observed movement could be the limiting case of reduction to 0. However, what would remain unexplained on that view is why the TB movement reduces so much more than the TT or TR movement.

Exactly what the compositional difference in gestures is between the flap and lateral contexts remains unclear. Usually, tongue body raising and tongue root fronting would be consequences of a single palatal or velar constriction task. Here the TR fronting is observed without TB raising in the inter-vocalic context. Another thing that remains unexplained for now is how this gestural composition difference results in the tongue side-lowering goal that is responsible for the lateral percept [12]. In the case of English, the main articulatory distinction between the lateral and flap is that the flap is produced with tongue side contact as reported in previous work with a palatogram [5,18]. It appears that onset and coda liquids in Korean are perceived as laterals, but unlike English, their production mainly involves the tongue body raising, not retracting. In order to better understand the role of the tongue body support with respect to the lateral articulation and the resulting perceptual effect, further evidence with different methods (e.g., electropalatography or tongue shaping analysis) is needed.

5. ACKNOWLEDGEMENTS

Work reported here was supported by NIH Grant R01 DC007124.

6. REFERENCES

[1] Bresch, E., Nielsen, J., Nayak, K., Narayanan, S. 2006. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *JASA*, 120(4), 1791-1794.

[2] Browman, C., Goldstein, L. 1986. Towards an articulatory phonology. *Phon. Yrbook*. 3, 219-252.

[3] Browman, C., Goldstein, L. 1992. Articulatory Phonology: An Overview. *Phonetica*, 49, 155-180.

[4] Browman C., Goldstein, L. 1995. Gestural syllable position effects in American English. *AIP*.

[5] Byrd, D. 1994. Palatogram reading as a phonetic skill: a short tutorial. *JIPA*, 24, 21-34.

[6] Fowler, C. 1980. Coarticulation and theories of extrinsic timing. *J. Phon*. 8, 113-133.

[7] Fukaya, T., Byrd, D. 2005. An articulatory examination of word-final flapping at phrase edges and interiors. *JIPA*, 35(1), 45-58.

[8] Iverson, G., Sohn, H-S. 1994. Liquid representation in Korean, *Theoretical issues in Korean ling*. 77-100.

[9] Jun, J. 2000. Preliquid nasalization. *Korean J. Ling*. 25(2), 191-208.

[10] Lammert, A., Goldstein, L., Iskarous, K. 2010. Locally-weighted regression for estimating the forward kinematics of a geometric vocal tract model. *Proc. Interspeech* Makuhari, 1604-1607.

[11] Lammert, A., Ramanarayanan, V., Proctor, M., Narayanan, S. 2013. Vocal tract cross-distance estimation from real-time MRI using region-of interest analysis. *Proc. Interspeech* Lyon, 959-962.

[12] Ladefoged, P. 1980. What are linguistic sounds made of? *Language*, 56, 485-502.

[13] Narayanan, S., Nayak, K., Lee, S., Sethy, A., Byrd, D. 2004. An approach to real-time magnetic resonance imaging for speech production. *JASA*, 115, 1771-1776.

[14] Oh, S., Gick, B. 2002. An ultrasound study of articulatory gestures of Korean /l/. *Proc. 1st ICSS* Seoul.
[15] Parrell, B. Narayanan, S. 2014. Interaction between general prosodic factors and language-specific articulatory patterns underlies divergent outcomes of coronal stop reduction. *Proc*. *10th ISSP* Cologne.
[16] Seo, M-S. 2004. A production-based study of the Korean liquid in onset position. *SPPM*, 10(2). 257-276.
[17] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *J. Phon*. 21, 291-311.
[18] Stone, M., Faber, A., Raphael, L. J., Shawker, T. H. 1992. Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ] and [l]. *J. Phon*. 20, 253-270.