Acoustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Pitfalls of spectrogram readings of flaps

Minjung Son

Yale University and Haskins Labs, 300 George St. Suite 900, New Haven, CT 06511, USA
son@haskins.yale.edu

Several acoustic studies have revealed stable characteristics of the speech signal for rhotics, whereas articulatory studies have found different articulatory strategies across speakers for producing these segments. However, no study has showed that the phonologically reduced flap gesture, which is argued to be a lateral in the underlying representation, can sometimes fail to be reflected in the acoustic correlates of an articulatory event. Using electromagnetic midsagittal articulometry, a case study has been conducted for Korean. Productions of two Seoul-Korean female speakers were collected using real words in the within-word condition. The subjects read target words within a single accentual phrase and repeated eight times. Each target word (/ili/ and /ala/) appeared in its own distinct natural sentence at two speech rates (fast and comfortable). A total of 32 tokens were available for analysis. Tongue tip gestures were examined simultaneously with spectrograms. Results from one speaker indicated that sometimes no acoustic correlates of the tongue tip gesture for a flap were present (6.3% of productions) although articulation was obviously produced [Work supported by NIH DC00403].

1 Introduction

Flaps are allophones of word-medially unstressed alveolar voiceless/voiced stops /t, d/ for American English (e.g., /lædəɹ/ → [læɾəɹ], 'ladder' [9]) of either pre-stressed or post-stressed alveolar voiceless/voiced stops /t^h, d/ for Xiangxiang Chinese (e.g., /Vt^{h}V/ or /VdV/ → [VɾV] [24]), while a flap is known to be an allophone of a lateral /l/ for Korean (e.g., /puli/ → [puɾi], 'beak' [22]). Although flaps/taps in some languages have been studied using different methodologies with or without simultaneous acoustic analysis (cf., X-ray studies on alveolar flaps in dental taps in Spanish [14]; EPG studies on alveolar taps in Catalan [18, 19]; acoustic-aerodynamic studies on alveolar flap in Xiangxiang Chinese [24]), articulatory behavior of flaps has not studied on Korean. The goal of the present acoustic/EMMA study is to identify whether articulatory properties are correlated to acoustic properties and to determine which domain shows invariance.

2 Background

2.1 Korean flapping rule

(1)

a. /pi+li/ → [piɾi] 'corruption'

b. /kalu/ → [kaɾu] 'powder'

c. /p'alli/→ [p'alli] 'hurry'

In (1.a and 1.b) a lateral becomes a flap when it occurs intervocalically. However, as shown in (1.c), /ll/ does not undergo flapping when two laterals occur in the intervocalic position. This allophonic rule is conditioned by a small prosodic boundary in Korean, e.g., accentual phrase, being understood as a lenition process [26]. For Korean, an acoustic study [22] showed a short closure duration (20 ms.) and a high frequency of voicing (95%). Bursts occurred less than half of productions (42%).

2.2 Flaps: phonetic description

A flap is a sound when an active articulator contacts briefly either the dental, alveolar, or post-alveolar area of the roof of the mouth [13]. According to Ladefoged [12], this apical sound is made 'moving the active articulator tangentially to the site of the contact, so that it strikes the upper surface of the vocal tract in passing'. In a simultaneous acoustic/aerodynamic study, there was intra-speaker variation in the phonetic realization of a flap derived from /d/ or /t^h/, which rendered several variant types [24]. In this acoustic/aerodynamic study, different oral airflow had its corresponding acoustic correlates—increase of oral air flow at the onset of release is correlated with a clear release burst, decrease of oral airflow with a short occlusion, and no change in oral airflow with no formant amplitude change.

2.3 Articulatory Phonology

Gestures, which is a task accomplished by a set of articulators, are the units of phonemic contrast, phonological patterning, and phonetic description [2, 3]. These primitive units are coordinated, which is achieved by a dynamical coupling oscillators associated with the gestures [15, 16], so as to achieve a desired phase relation between pairs of gestures [6, 7]. By hypothesis, this information (e.g., nodes specifying gestures such as tongue tip critical for constriction degree and phase relations such as in-phase (stable mode) and anti-phase (less stable mode)) is present in phonological representation of an utterance, which is encoded by using a gestural coupling graph [6, 7, 15, 16]. To quote Browman and Goldstein [3], by hypothesis, 'phonetic variation can be captured either by quantitative variation in the input parameter specification of a given gesture, or as a direct output consequence of overlap of invariant gestural units'. A flap, which is traditionally known to be derived from a lateral in the intervocalic position, can be defined using two constriction variables—three gestural landmarks of the tongue tip (gesture onset, target attainment, and release) for constriction location and constriction minima for constriction degree. Within the framework of AP [2, 3], a flap is assumed to have a reduced tongue tip constriction gesture in time and space, involving a reduction process of the tongue tip gesture, which in turn results in 'a short (single timing slot) open-closed-open contour in an autosegmental representation (Banner-Inouye in 1989 from [3]). That is, short constriction duration and less constriction for a flap can be understood as a weakening process [21, 26].

3 Methods

3.1 Data collection and subjects

The Perkell-system electromagnetic midsagittal articulometer system (EMMA) at Haskins Laboratories was used to collect kinematic data of the tongue [17]. Four transducers were attached on the tongue at roughly equal distances (i.e., on the tongue tip, anterior tongue body, posterior tongue body, and tongue dorsum), and two on the upper lip and lower lip. They were aligned with each subject's vocal tract in the midsagittal plane.

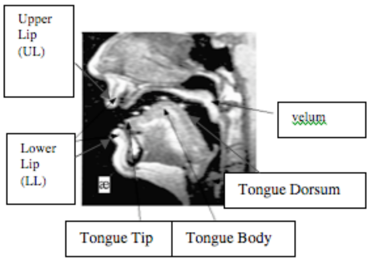


Figure 1. Typical receiver locations imposed on an MRI image.

Stimuli containing target words with tongue tip constriction gestures were part of another experiment [21]. Two vocalic contexts (i.e., /ili/ and /ala/) were used. We used short natural sentences (/cənip**pili**ka simhakun/ 'there is too much transference corruption'; /cəna pakacilɨl p**ala**/ 'Ceona sells gourds'). The presentation of stimuli containing target words was blocked by speech rate condition (fast and comfortable). Speakers' productions occurred with interruption from the four repetitions of a different sentence, rendering a total of eight repetitions of each target word. The speech signal was sampled at 20 kHz. Acoustic data were collected with a Sennheiser shotgun microphone at the time of acquiring articulatory data. Stimuli were presented in Korean on a computer screen.

A total of two female native speakers of Seoul Korean participated. Both subjects, none with any speech impairment, were naive to the purpose of the experiments. They spent their first twenty-three years mostly in Seoul, and identified themselves as speaking Seoul-Korean dialect.

3.2 Measurements

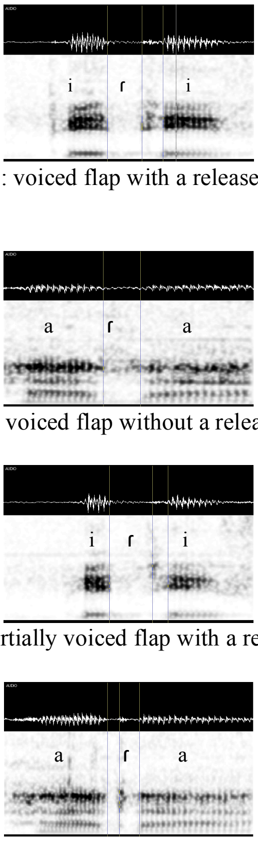
For this study, spatial and temporal properties of gestures are measured with respect to the constriction degree of the tongue tip (TTCD). An estimation of palate position is obtained by locating the maximum vertical positions attained by the tongue transducers during the entire course of the experiment (the vertical positions are presumably limited by the palate) and finding the convex hull that encloses these vertical tongue maxima. The hull is re-sampled at a 0.5 mm horizontal resolution.

For measurements of the tongue tip gesture, tongue tip constriction degree is measured as the Euclidean distance between the sensor on the tongue tip and the pseudo-palate. In measuring temporal and spatial values, two functions in MVIEW are used [23]. To find the maximally constricted values for the tongue tip constriction degree (during /ɾ/), a function (Snapex) is employed that finds the nearest position extremum (velocity zero) to the time location at which the user clicks. To find the times of gestural onset, target attainment, and release a second function (Findgest) is employed. The gestural onset is defined when the velocity of the corresponding constrictor exceeds a threshold, defined as a percent of the difference between the local speed minimum and speed maximum. Time of target attainment and release are measured in similar ways. In analyzing data, a 20% of threshold is used. For spectrogram analysis closure duration and intensity were measured using Praat [1].

4 Results: Acoustic data

4.1 Intra-speaker variability

An intervocalic flap [ɾ] showed several different patterns in its phonetic realization. Although there was frequency difference between the two speakers, two phonation types occurred, showing a higher frequency in voiced productions than voiceless (including partially voiced). Intra-speaker variation in the realization of an intervocalic flap is presented using sound waveforms and spectrograms in Fig. 2. Voiced productions occurred with or without a release burst (Fig. 2.i, 2.ii). Voiceless or partially voiced productions occurred with a release burst (Fig. 2.iii, 2.iv).



i. [iɾi]: voiced flap with a release burst

ii. [aɾa]: voiced flap without a release burst

iii. [iɾi]: partially voiced flap with a release burst

iv. [aɾa]: voiceless flap with a release burst

Figure 2. Flaps with an occlusion. The top channel is a waveform. The bottom channel is a sound spectrogram. Each time window is approximately 2.5 milliseconds.

In addition to different patterns of an intervocalic flap illustrated in Fig. 2, we also observed that several flaps were not distinguished at all from adjacent vowels in both waveforms and spectrograms. We did not see acoustic changes typical of a flap. For convenience's sake, we call it extreme [24]. Shown in Fig. 3, underlying lateral was reduced to the extent of deletion, rather than being phonetically realized as a typical flap (e.g., a short occlusion with or without a release burst).

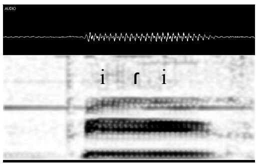


Figure 3. [iɾi]: A flap without closure duration. The top channel is a waveform. The bottom channel is a sound spectrogram. The time window is approximately 2 milliseconds.

In Table 1, the frequency of each phonation type is shown in percentages. Some tokens (one token for S1 and seven for S2) were eliminated from a further analysis due to uncertainty about whether it was a release burst or noise caused by the friction of the transducer attached on the tongue tip against the roof of the mouth.

		Speaker 1 N=31	Speaker 2 N=25
Voiced	Fully voiced	84%	76%
	Partially voiced	6%	16%
Voiceless		10%	8%

Table 1. Frequency of two phonation types.

The duration of a flap was measured as followed. If there was a burst in the sound spectrogram and the waveform, we measured the closure duration and the burst. If there was no burst, the closure duration was measured exclusively. If there was no indication of closure in the sound spectrogram, we measured the portion, which exhibited decrease in amplitude in the waveform; if this was not available, we skipped measuring closure duration. Along with these measurements, we also acquired intensity measured at the mid point of the closure duration. If none of these durational measurements was available, the intensity was measured at the time location of a tongue tip minimum (see Fig. 8.ii of section 5.3). The means are shown in Table 2. Standard deviations are in italics in a parenthesis.

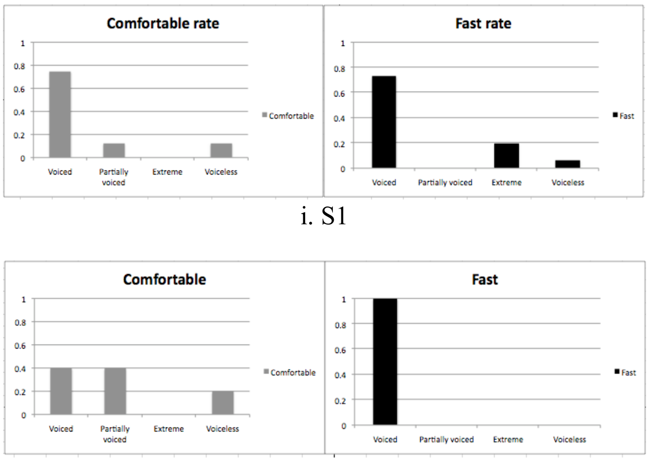
Acoustic		Speaker 1		Speaker 2	
		Duration in ms.	Intensity in dB	Duration in ms.	Intensity in dB
vd	Occlusion & release burst	36.9 (*8.24*) N=14	63.1 (*2.72*)	46 (*8.4*) N=6	59 (*1.1*)
	Occlusion	23.5 (*2.49*) N=9	63.2 (*4.24*)	29.9 (*11*) N=13	62 (*2.64*)
	Extreme	U/A N=3	68.4 (*6.39*)		
part vd	Occlusion & release burst	49.6 (*1.77*) N=2	58.4 (*.35*)	80.6 (*5.48*) N=4	55.4 (*.46*)
vl	Occlusion & release burst	45.8 (*6.54*) N=3	58.8 (*2.23*)	61.5 (*19.3*) N=2	58.9 (*.41*)

Table 2. The means and standard deviations of the total duration and of intensity by subject.

The mean duration of flaps was the largest for the partially voiced flaps, which was in turn followed by the voiceless ones, which was in turn followed by the voiced ones. Confined to voiced flaps, the mean duration of flaps was longer when there was an occlusion and a release burst co-occurred. The intensity values were larger for tokens without acoustic demarcation for flaps.

4.2 Rate effect

The two speakers coherently produced partially voiced flaps confined in comfortable rate in 13% of productions in each rate for S1 (Fig. 4.i) and in 40% for S2 (Fig. 4.ii). Fast rate exhibited three extreme cases in 20%, rendering 3 out of fifteen tokens at fast rate for S1.



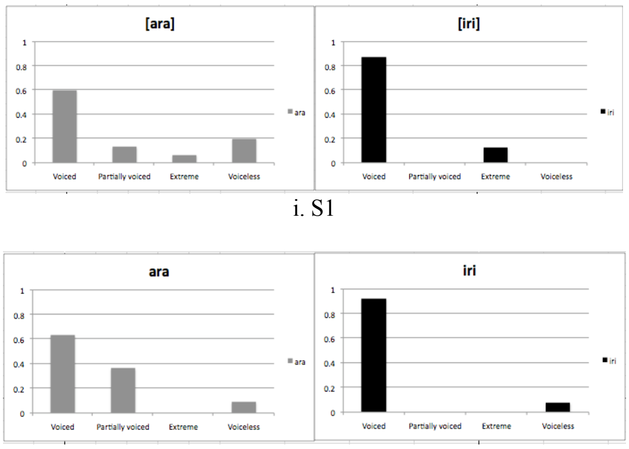
i. S1

ii. S2

Figure 4. Frequency of each flap type in percentages. Percentages are calculated based on the number of each flap type over the total number of tokens in each condition (rate). Bars in gray represents comfortable rate, bars in black fast rate.

4.3 Vocalic context effect

For S1, four different glottal patterns were observed in [aɾa] sequences, while fewer patterns occurred in [iɾi] sequences (Fig. 5.i). A similar pattern was also observed for S2 (Fig. 5.ii). For both speakers, high frequency in the emergence of fully voicing with an occlusion was obtained across different vocalic contexts.



i. S1

ii. S2

Figure 5. Frequency of each flap type in percentages. Percentages are calculated based on the number of each flap type over the total number of tokens in each condition (vocalic contexts). Bars in gray represents the /a/ context, bars in black the /i/ context.

5 Results: Tongue tip gesture

5.1 Tongue tip formation duration

Two speakers exhibited tongue tip constriction gestures in 100% of productions. The means for the formation duration of the tongue tip gesture, measured from gestural onset to target attainment, were shown by subject and by speech rate.

	Speaker 1 (N=32)		Speaker 2 (N=32)	
	Fast	Comfort	Fast	Comfort
[iɾi]	53 *(6.5)*	57.5 *(9.11)*	73.25(*9.67*)	91.5(*7.61*)
[aɾa]	54.25(*3.77*)	62.75 (*4.77*)	60.5 (*4.5*)	61.25(*2.81*)

Table 3. Means and standard deviations of formation duration of the tongue tip constriction gesture in milliseconds.

In a pairwise comparison of the means using paired sample t-tests, S1 had rate effects on the formation duration of a flap in the /a/ context (t(7) = -4.348, p <.01). The result of a flap in the /i/ context for S2 was significant, showing shorter duration for fast rate (t(7) = -4.311, p <.01).

5.2 Tongue tip constriction duration

There was longer constriction duration for a flap [ɾ] in the /a/ context for S1 (t(7) = -1.552, p < .05), while other comparisons did not reach statistical significance. The average values of constriction duration, measured from target achievement of the tongue tip gesture to release, are shown by subject and by speech rate in Table 4.

	Speaker 1 (N=32)		Speaker 2 (N=32)	
	Fast	Comfort	Fast	Comfort
[iɾi]	10.3(*.71*)	10.3(*.71*)	30.5(*12.1*)	44.25 (*20.2*)
[aɾa]	9.25(*1*)	10.8(*1*)	15.3 (*4.7*)	13 (*1.5*)

Table 4. Means and standard deviations of constriction duration of the tongue tip constriction gesture in milliseconds.

5.3 Tongue tip constriction degree

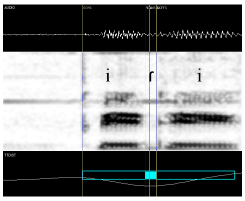
For S1, fast rate consistently exhibited less constriction in fast rate of each vocalic context (t (7) = 3.140, p<.05 for [iɾi]; t (7) = 19.745, p <.0001 for [aɾa]). However, this did not hold true for S2 (p>.05). The average values of constriction degree using Snapex [23] are shown by subject and by speech rate in Table 5.

	Speaker 1 N=32		Speaker 2 N=32	
	Fast	Comfort	Fast	Comfort
[iɾi]	5.51 (*.4*)	4.7(*.57*)	7.54(*.81*)	8.16 (*.85*)
[aɾa]	10.7 (*.16*)	8.21(*.36*)	12.9(*1.22*)	12.6 (*.78*)

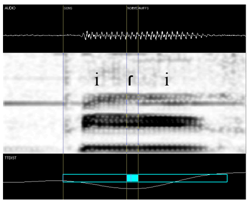
Table 5. Means and standard deviations of constriction degree of the tongue tip constriction gesture in millimeters. Smaller numbers indicate more constriction.

5.4 No acoustic correlates of an articulatory event

In order to determine whether gestural reduction occurs in some productions of flaps, we simultaneously take into account acoustic and articulatory data. In Fig. 3 of section 4.1, we show two tongue tip constriction gestures: one is for a typical flap with an occlusion and a release burst, and the other for an acoustically deleted flap, namely an extreme case. As shown in Fig. 6, no articulatory deletion occurred although it failed to be reflected in acoustic correlates of an articulatory event.



i. [iɾi]: voiced flap with a short occlusion and a release burst



ii. [iɾi]: no acoustic correlates of the tongue tip constriction gesture

Figure 6. The top channel is a waveform. The next channel is a sound spectrogram. The bottom channel is for constriction time functions for the tongue tip constriction gesture. Each time window is approximately 2 milliseconds. The first vertical line is a timing point for the gesture onset, the second for the target attainment, the third for the tongue tip minimum, and the forth for the release.

In order to find whether no acoustic correlates of the tongue tip gesture was due to a reduced tongue tip displacement, we compared the two patterns. Since extreme cases were only confined to fast speech for S1, means and standard deviations for the subset of this speaker were calculated (N=12 for flaps with an occlusion; N=3 for extreme cases). In Table 6, the numerical values for the constriction duration and for the constriction degree are shown.

Articulatory measures	Acoustic pattern	Gestural duration & displacement
Formation duration	Acoustic occlusion	53.4 (*5.38*) ms.
	Extreme	54.7 (*5.03*) ms.
Constriction duration	Acoustic occlusion	9.85 (*.99*) ms.
	Extreme	9.33 (*1.15*) ms.
Constriction degree	Acoustic occlusion	8.21 (*2.78*) mm.
	Extreme	7.63 (*2.75*) mm.

Table 6. Means and standard deviations for constriction duration in milliseconds and for constriction degree in millimeters. For formation duration and constriction duration, larger numbers indicate longer duration. For constriction degree, smaller numbers indicate narrower constriction.

As shown in Table 6, the mean value for the formation duration for the tongue tip gestures was slightly longer for extreme cases than flaps with an acoustic occlusion, but constriction duration was shorter. Narrower constriction was observed in extreme cases. However, none of comparisons was significant using a pairwise comparison of the means using paired sample t-tests (p>.05). Therefore, we conclude that the tongue tip gestures which did not have acoustic correlates was as large as those with an acoustic occlusion.

6 Discussion

This paper examines the phonetic nature of a flapping lateral in /VlV/ sequences, analyzing acoustic and articulatory data. Although stable tongue tip gestures for a flap are consistently executed in the phonetic level across speakers, acoustic data show higher variability, showing several cases of deletion of a flap. The present study does not confirm that an acoustic extreme case *should* have correlates of categorically reduced tongue tip. We would like to examine why this articulatory invariance is not represented in acoustics. One possible reason is that kinematic data for the tongue tip gesture is only obtained from one transducer attached in the midsagittal plane of subject's vocal tract. Hence, the absence of acoustic correlates of the tongue tip gesture, which are further correlated with an air pressure buildup in the oral cavity, presumably indicates no simultaneous contact of the sides of the tongue against the roof of the mouth. However it is still unclear why the invariant tongue tip gesture is not reflected in acoustics at all for the three tokens. Future articulatory studies should include data on the sides of the tongue. Future perceptual studies should also investigate whether or not a tongue tip gesture that does not have acoustic correlates has a possible impact on the listener's perception of a flap.

Regarding glottal gestures of a consonant in the intervocalic position, glottal abduction of a consonantal gesture is considered as a reduction process (e.g., unstressed word-medial /t/ → [ɾ] in American English [9]). However, intervocalic voiceless flaps in the phonetic form obtained in the current study (10% for S1 and 8% for S2; see also [22]) indicate no reduction. This glottal spreading in some productions of flaps is not is phonetically grounded (cf., segmental context effects). For example, voiceless flaps are induced by /h/ coalesced with /t/ in Tümpisa Shoshone [11] and voiceless flaps occurs in the word-final position for Spanish [25]. One possible reason of the glottal spreading gesture may be due to 'enhancing consonantality' of a flap against flanked vowels in /VɾV/ sequences [4].

Acknowledgments

This work was supported by NIH DC 00493 to Haskins Laboratories. The author wishes to thank Hosung Nam for his comments.

References

[1] Boersma, Paul, David Weenink, Praat: doing phonetics by computer (version 4.3.14) [computer program]. http://www.praat.org (2005)

[2] C. Browman and L. Goldstein, "Gestural specification using dynamically-defined articulatory structures', *Journal of Phonetics* 18, 299–320 (1990)

[3] C. Browman and L. Goldstein, "Articulatory phonology: An overview", *Phonetica* 49, 155–180 (1992)

[4] T. Cho and S. Jun, "Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean", *Chicago Linguistics Society* 36, 31-44 (2000)

[5] J. Dayley, *Tümpisa (Panamint) Shoshone Grammar*, University of California Publications in Linguistics 115, UC Press. (1989)

[6] L. Goldstein, D. Byrd, and E. Saltzman "The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to Language via the Mirror Neuron System*. M. Arbib (ed.), Cambridge University Press, 215–249 (2006)

[7] L. Goldstein, I. Chitoran, and E. Selkirk "Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber", *Proceedings of the 16th International Congress of Phonetic Sciences*, 241–244 (2007)

[8] S. Jun, "*The Phonetics and Phonology of Korean Prosody"*, Ph.D. dissertation. The Ohio State University, Columbus, Ohio. (1993)

[9] D. Kahn, *Syllable-based generalizations in English phonology*. Indiana University, Linguistics Club, Bloomington. (1976)

[10] Y. Kim-Renaud, *Korean Consonantal Phonology*, PhD Dissertation, University of Hawaii. (1974)

[11] R. Kirchner, "Phonological contrast and articulatory effect", *Segmental Phonology and Optimality Theory*. In L. Lombardi (ed.), Cambridge: Cambridge University Press. (2001)

[12] P. Ladefoged, *A Phonetic Study of West African Languages*. Cambridge University Press, Cambridge. (1968)

[13] P. Ladefoged and I. Maddieson, *The Sounds of World's Languages*, Blackwell Publishers. (1998)

[14] M. Monnot and M. Freeman, "A comparison of Spanish single –tap /ɾ/ of velar activity during speech", *Cleft Palate Journal* 4, 58–69 (1972)

[15] H. Nam "Articulatory modeling of consonant release gesture", *Proceedings of the 16th International Congress of Phonetic Sciences* 8, 625–628 (2003)

[16] H. Nam and E. Saltzman "A competitive, coupled oscillator of syllable structure", *Proceedings of the 15th International Congress of Phonetic Sciences*, 2253–2256 (2003)

[17] J. Perkell, M. Cohen, M. Svirsky, M. Matthies, I. Garabieta, and M. Jackson, "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements", *Journal of the Acoustical Society of America* 92, 3078–3096 (1992)

[18] D. Recasens and A. Espinosa " Phonetic typology and positional allophones for alveolar rhotics in Catalan", Phonetica 63, 1–28 (2007)

[19] D. Recasens and M. D. Pallarès "A study of /□/ and /r/ in the light of the "DAC" coarticulation model", Journal of Phonetics 27, 143–169 (1999)

[20] E. Saltzman and K. Munhall, 'A dynamical approach to gestural patterning in speech production', *Ecological Psychology* 1, 333–382 (1989)

[21] M. Son, *The Nature of Korean Place Assimilation: Gestural Overlap and Gestural Reduction*. PhD Dissertation, Yale University. (2008)

[22] E. Sung "The relationship between phonetics and phonology: Problematic cases for a unified module", 음성.음운.형태론 연구 13 (2), 269–287 (2007

[23] M. Tiede "MVIEW: software for visualization and analysis of concurrently recorded movement data", (2005)

[24] Z. Ting "Understanding flapping in Xiangxiang Chinese: Acoustic and aerodynamic evidence', *Proceedings of the 16th International Congress of Phonetic Sciences*, 393–396 (2007)

[25] M. S. Witley, *Spanish/English contrasts: A course in Spanish Linguistics*, Georgetown University Press, (2002)

[26] K. Yoon, C. Brew, and M. Beckman, "Letter-to-sound for Korean", Proceedings of 2002 IEEEE Workshop on Speech Synthesis (2002)