# Effects of Emotion on the Lower Lip Movements at Phrase Boundaries

*Sungbok Lee and Shrikanth Narayanan*

SAIL Lab., University of Southern California

sungbokl@usc.edu, shri@sipi.usc.edu

## Abstract

In this preliminary report we investigate the effects of emotion on the lower lip movements during the production of a monosyllabic word in two different prosodic contexts: phrase initial and phrase final. The purpose is to examine the rigidity of the phrase boundary effects under emotional perturbation of the normal, or neutral, speech production. It is found that the effect of emotion is ubiquitous in that it affects all kinematic and dynamic parameters of the lower lip movements considered in this study. The effects of emotion are not arbitrary, however, and some common characteristics of emotional speech articulation can be identified. Firstly, speakers maintain the phrase boundary effects on the duration and movement amplitude of the lower lip gesture under emotional variations. Secondly, when considering all emotion types together in a given speaker, an approximately linear relationship holds between movement amplitudes and maximum lip opening velocities when emotion varies. Therefore, the stiffness of the lower lip opening gestures seems almost invariant under emotional variations, implying that the lower lip gesture is regulated by a simple harmonic-oscillator like system with a constant stiffness. These and other findings support the hypothesis that speech production constraints imposed by the phrase boundary condition are maintained against emotional perturbations of speech articulations. In addition, the findings in the study may point to a general emotional speech production mechanism that the effects of emotion on the kinematics and dynamics of the oral speech articulators are manifested mainly in the arousal dimension of emotion.

**Index Terms**: emotion, lip movement, stiffness, phrase boundary

## 1. Introduction

It has been well recognized that articulatory realization of phonetic segments is affected by the prosodic structure or conditions imposed on the target segments. For example, a number of articulatory movement tracking studies have shown that temporal lengthening (i.e., slowed articulation) and spatial articulatory strengthening (i.e., more extreme articulation) of linguistic articulation occur at phrase boundaries [1],[2],[3].

Recent articulatory studies of affective speech have shown that attitude or emotion expressed by speakers modulates not only pitch patterns and voice qualities but also spatiotemporal articulatory movements (e.g., [4],[5]). Since affective or emotional quality of spoken utterances is encoded over a given linguistic structure of a speech utterance, it seems natural to ask the question of how emotion affects speech articulation that is already conditioned by a prosodic contrast such as a difference in phrase boundary condition. Understanding kinematic behaviors of speech articulators in such doubly-imposed prosodic conditions can be useful for modeling emotional speech production mechanism and its simulation through an articulatory speech synthesizer such as TADA [6].

Kinematics of the lip movements have been studied in the literature, for example, to illuminate the motor control and coordination strategies [7] and to compare differences between normal and speech impaired populations [8]. In this preliminary report we focus on the effects of emotion on the lower lip movement that include contribution from the jaw. Specifically, the effects of emotion on the lower lip movements during the production of monosyllabic word "FIVE" are investigated in two different prosodic contexts; phrase initial and phrase final. The main focus here is given to the investigation of how the phrase boundary effect observed in neutral or "normal" speech articulation is affected by speech emotion expression. A null hypothesis to be tested is that the speech production constraints imposed by the phrase boundary condition are maintained against emotional perturbations of speech articulations.

## 2. Method

### 2.1. EMA data collection

Speech production data from one male actor and two female actresses were collected using the Carstens AG500 electromagnetic articulograph (EMA). For each utterance, the movements of six sensors attached to the tongue tip, tongue blade, tongue dorsum, the upper and lower lips and the jaw were recorded at 200 Hz sampling rate with a simultaneous recording of speech waveforms at 16-kHz sampling rate. After the EMA data acquisition, head-movement corrections and occlusal plane rotations were performed for all utterances and the trajectory signal of each articulatory sensor was filtered with a 9th-order Butterworth filter with a 15 Hz cutoff frequency. Each sensor trajectory was then scanned for possible trajectory errors and erroneous trajectory segments were marked for exclusion during the analysis step.

### 2.2. Speech material, emotion types and perceptual evaluation of utterances

Five categorical emotions including neutral emotion were considered in the data collection. They were hot anger, cold anger, sadness and happiness. As for speech material, seven emotionally neutral sentences were designed and each sentence was produced five times in each emotion category. Each subject in the study has had more than ten years of theatrical experience and they were instructed to produce the target utterances only after they were ready to express the target emotions.

For perceptual evaluation each utterance spoken by the three subjects was presented to five native listeners of American English in randomized order using a web interface [9]. Listeners were asked to choose (1) the best-representative emotion among the six emotion categories, (2) confidence in their evaluation and (3) the strength of emotion expression. The listeners were asked to choose 'others' when they felt that none of the five given emotion categories best matched their perception. Confidence and strength were evaluated on a five-point scale of one to five.

### 2.3. Data analysis

From the seven sentences available, the sentence "*It was Nine One [Five], Two Eight Nine, [Five] Seven Six Two*" was chosen for this initial analysis. Specifically, the first (i.e., phrase final position) and the second (phrase initial position) "Five" were segmented and analyzed. Only utterances whose intended emotions by speakers were correctly identified by at least two listeners are included in the current study. A total of 34 and 38 tokens of the word "five" are analyzed for subject JN and JR, respectively, for the four emotion types of hot anger, sad, happy and neutral as reference.

#### 2.3.1. Lower lip trajectory segmentation

Using a matlab-based EMA data processing software called "mview," two instances of the word "FIVE" in the lower lip EMA trajectories of the target utterances were segmented manually as illustrated in Figure 1.
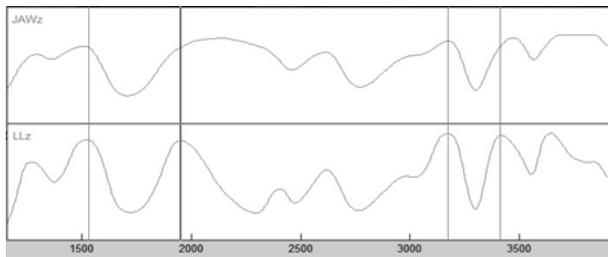


Figure 1: *Two productions of the word "FIVE" were manually segmented as illustrated by two sets of vertical lines.*

The first "FIVE" is phrase final and the second one phrase initial. Starting point of the segment is the time point where the lower lip is about to move downward and the end point is where the lower lip reaches its maximum upward position. A set of kinematic and derived dynamic parameters were measured from each segment interval and statistically analyzed for their significances as a function of emotion and phrase boundary condition.
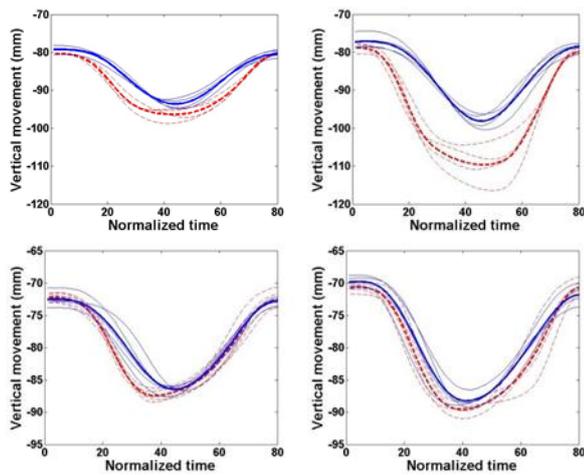


Figure 2: *Examplary lower lip trajectories of "five" are shown for subjects JN (first row) and JR (second row). In each row, the left plot is for neutral and the left plot for hot anger. For visual comparison of trajectory shapes, duration of each segment is normalized to 80 time-points (400 msec) by uniform resampling.*

Exemplary vertical lower lip trajectories of subject JN and JR are shown in Figure 2. In each plot, individual trajectroies (thin lines) are shown together with two averaged tracjectories (thick red dotted line for the first, and thick blue solid line for the second "five," respectively).

#### 2.3.2. Parameter measurement and statistical analysis

Segmented duration and movement range as well as extreme velocities during the opening and closing movements of the lower lip were measured for each segment. Division of the opening and closing interval was done by finding zero-crossing points in corresponding velocity curves. As a representative dynamic parameter that governs the lower lip movements, stiffness parameters for opening and closing lower lip movements were also estimated by dividing maximum velocities by corresponding movement ranges, respectively. Finally, measured kinematic and dynamic parameters were subject to univariate ANOVAs with emotion and phrase boundary condition as independent variables. Since what is interesting is to examine the trend of each subject's behavior in emotion expression, and not absolute values of speech parameters, each subject's data are analyzed separately.

## 3. Results

### 3.1. Duration

Segment durations are shown in Figure 1 as a function of emotion as well as phrase boundary condition for two female subjects JN and JR. In each subject, the univariate ANOVA analysis indicates that duration difference between two phrase boundaries are significant in each subject [$F(1,32) = 13.02$, p = .001 for subject JN; $F(1,36)=221.39$, p = .000 for subject JR]. Across emotions, the differences are also significant for subject JN [$F(3,30) = 23.36$, p = .000] as well as for subject JR [$F(3,34) = 21.88$, p = .064].

It is observed that durational contrasts between two phrase boundary conditions are rigidly maintained across emotions, irrespective of inter-speaker differences on duration assignment in which subject JN exhibits significantly longer segmental durations.
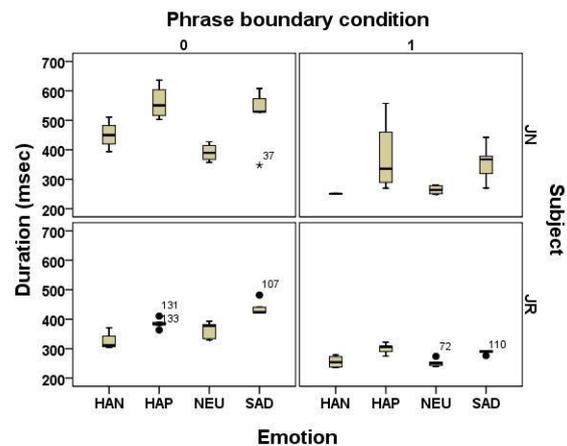


Figure 3: *Segmented durations of "five" are shown as a function of emotion and phrase boundary condition (**0**: first, phrase final. **1**: second, phrase initial). Final lengthening might have contributed to the significantly longer durations of the phrase-final "five" in each subject.*

## 3.2. Movement amplitude

A summary of all amplitude data is presented in Figure 4. It is evident for each subject that the lower lip movement range and variability are affected by emotion as well as by phrase boundary condition. A multivariate ANOVA shows that for
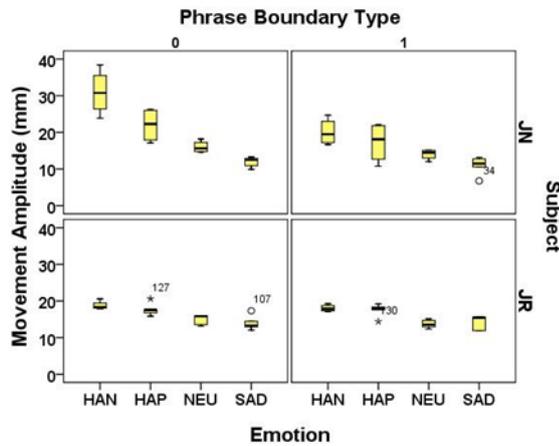
### Phrase Boundary Type



Figure 4: *Maximum movement amplitudes across emotions and phrase boundary types (0: first, phrase final five; 1: second, phrase initial) are shown in the figure. Effects of emotion and inter-speaker difference can be clearly observed.*

subject JN, both emotion and phrase boundary effects on movement amplitudes are significant [ $F(3,30) = 23.36$, p = .000 for emotion effect; $F(1,32) = 13.02$, p = .001 for phrase boundary effect]. For subject JR, however, there is no significant effect of the phrase boundary condition [$F(3,34) = 0.87$, p = .36]. These statistical significances confirm the visual impressions shown in Figure 4.

It is interesting to observe that the strength (i.e., magnitude) of movement amplitude is closely linked with the emotion quality assessment in the arousal dimension [9]. It is noted that the dynamic ranges of amplitude maneuvers are somewhat less for subject JR, when compared to subject JN, across emotions as well as across phrase boundary conditions.

## 3.3. Movement amplitude and maximum lip opening velocity

Movement amplitudes and maximum opening velocities of the lower lip movements are represented in Figure 5 as scatter plots. It visualizes the relationships between movement amplitudes and maximum lip opening velocities for each subject as a function of emotion in two phrase boundary conditions. A multivariate ANOVA analysis indicates that for both subjects the effects of both emotion and the phrase boundary condition on maximum opening velocity are significant [$F(3,30)=7.57$, p=.001 for emotion; $F(1,32) = 49.18$, p = .000 for the phrase boundary condition for subject JN, for instance. JR also shows similar results.]

It is evident that when all the emotion categories are considered together, approximately linear relationships hold between movement ranges and maximum velocities in each subject. For subject JN, such linear relationships imply that the stiffness of the lower lip movement is approximately invariant across different emotion expressions, although both absolute magnitudes of amplitudes and maximum movement velocities of the lower lip movements vary as a function emotion.
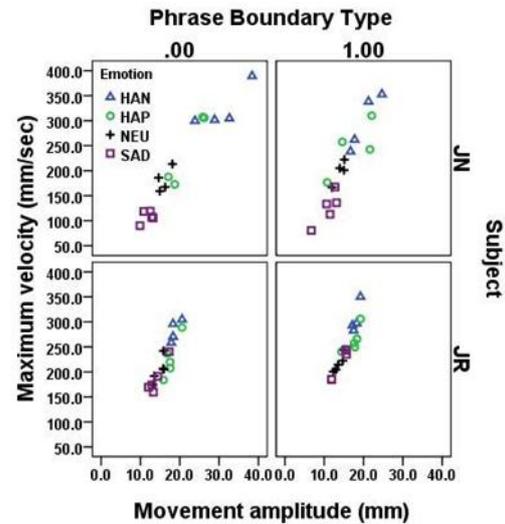


Figure 5: *Movement ranges and maximum opening velocities (i.e., speed) of the lower lip movements are shown for subjects JN and JR in two phrase boundary conditions (0: first, phrase final "five"; 1: second, phrase initial "five"). These linear relationships imply a simple harmonic-oscilltor like behavior of the lower lip opening gestures with a constant stiffness, irrespective of emotional variation.*

## 3.4. Stiffness parameter

Slopes which represent linear relationships between movement amplitudes and corresponding maximum velocities portray stiffnesses which govern the lower lip movement dynamics (cf., [10]). Stiffness parameters are explicitly computed from the measured values of the lip movement amplitudes and extreme velocities and the results are summarized in Figure 6.
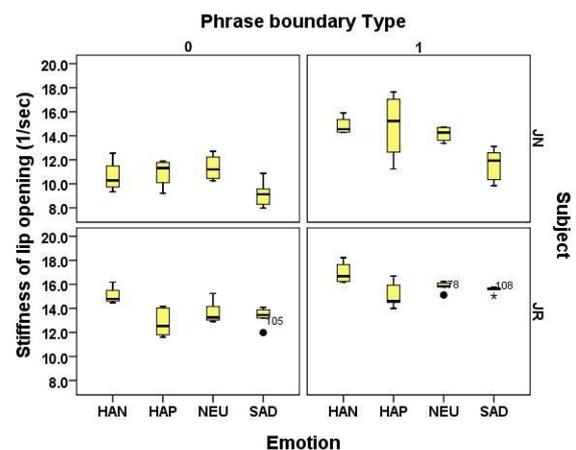


Figure 6: *Stiffness parameters of the lower opening in two phrase boundary conditions (0: phrase final, 1: phrase initial) are summarized as a function of emotion for each subject.*

Both emotion and phrase boundary effects are significant [$F(3,30) = 6.64$, p=.002 and $F(1,36)=45.04$, p=.000 for JN; $F(3,30)=8.46$, p=.000 and $F(1,36)=55.49$, p=.000 for JR]. For both subjects, the phrase boundary effects are much stronger than the emotion effect.

It is observed that while stiffness varies significantly as the phrase boundary condition changes, its inter-relationship among emotions are maintained in each speaker, which implies the rigidity of the phrase boundary effects under emotional perturbation. Speaker-dependency in stiffness control is also observed, with relatively larger stiffness values for subject JR. This should be the case in fact because JR shows much smaller ranges of amplitude variation across emotions (cf. Figure 4), while exhibiting comparable maximum velocities.

## 4. Discussion

Although a relatively short segmental unit (i.e., a monosyllabic word) was analyzed in this study, it was able to identify some interesting characteristics of emotional speech production. The effect of emotion seems ubiquitous in that it affects all kinematic and dynamic parameters of the lower lip movements measured in the current study. However, the effects of emotion do not appear to be arbitrary, and some common characteristics of emotional speech articulation can be identified irrespective of inter-speaker differences. Firstly, both speakers maintain the phrase boundary effects on the duration and movement amplitude of the lower lip opening across emotional perturbations, which imply the rigidity, or resistance, of linguistic prosodic effects against emotional perturbations of the normal or neutral speech articulations. Secondly, it has been observed that approximately linear relationships hold between movement amplitudes and maximum lip opening velocities irrespective of emotion types and phrase boundary conditions, although inter-speaker differences in stiffness control are significant. Therefore, for a given speaker, the stiffness of the lower lip opening gestures seems almost invariant under emotional variations, although the behaviors of individual data points of amplitudes and maximum movement velocities are significantly different under emotional perturbations of their normal speech articulations.

Another interesting observation is that the strength, or magnitude, of the lower lip movement amplitudes (i.e., displacements) and movement velocities are closely linked to the arousal, or activation, dimension of emotion [11]. For instance, the strength is strongest for hot anger and weakest for sadness (see Figure 4). This implies that the speech motor control system is somehow affected by the arousal state of the autonomic nervous system [12]. Therefore, the results in this study may point to a general underlying mechanism of the articulatory movements for emotion encoding: the effects of emotion on the kinematics and dynamics of the oral speech articulators such as the tongue, jaw and lips are manifested mainly in the arousal dimension of emotion.

## 5. Summary

While considerable inter-speaker differences exist in the kinematic and dynamic parameters of the lower lip movements under emotional perturbation, some common patterns of the lower lip behaviors, irrespective of differences in subject and emotion type, have been observed: (1) the phrase boundary effects on the durations and movement amplitudes, or displacements, of the lower lip gestures are maintained under emotional variations in spoken utterances, and (2) the vertical lower lip gesture is regulated by a simple harmonic-oscillator like system with a stiffness that is almost invariant across emotions and individuals. In summary, the results of the study provide evidence that supports the hypothesis that the speech production constraints imposed by phrase boundary condition are preserved against emotional perturbations of speech articulations. Investigations of other speech articulators such as the tongue and the jaw as well as the effects of emotions in various prosodic contexts will follow based on the further analysis of the current database.

## 6. Acknowledgements

## 7. References

[1] Edwards, J., Beckman, M.E., Fletcher, J. "The articulatory kinematics of final lengthening," J. Acoust. Soc. Am. 89, 369-382, 1991.

[2] Fougeron, C. and P. Keating. "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. 101, 3728-3740, 1997.

[3] Byrd, D., Saltzman, E., "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. Phonetics 31, 149–180. 2003.

[4] D. Erickson, O. Fujimura, B. Pardo, "Articulatory correlates of prosodic control: Emotion and Emphasis," Language and Speech, 41, 395-413, 1998.

[5] Lee, S., Yildirim, S., Kazemzadeh, A. and Narayanan, S., "An Articulatory study of emotional speech production", Interspeech, Lisbon, Portugal, 497-500, 2005.

[6] http://www.haskins.yale.edu/tada_download/index.php

[7] P. Van Lieshout, C. Rutjens, P. Spauwen, "The Dynamics of Interlip Coupling in Speakers With a Repaired Unilateral Cleft-Lip History," J. Speech, Lang. and Hear. Res., 45, 5-19, 2002.

[8] V. Gracco, A. Löfqvist., "Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements," J. Neuroscience, 14, 6585-6597, 1994.

[9] J. Kim, S. Lee, S. Narayanan, "An Exploratory Study of the Relations between Perceived Emotion Strength and Articulatory Kinematics," Interspeech, Florence, Italy, 2011.

[10] K. Roon, A. Gafos, Ph. Hoole, Ch, Zeroul., "Influence of articulator an manner of stiffness," Proceedings 16th ICPhS, Saarbrücken, 409-412, 2007.

[11] R. Cowie, R. Cornelius., "Describing the emotional states that are expressed in speech," Speech Communication 40, 5–32, 2003.

[12] LeDoux, J., "Emotional networks and motor control: a fearful view," Progress in Brain Research, 107, 437-446, 1996.