# Relation between Geometry and Kinematics of Articulatory Trajectory Associated with Emotional Speech Production

*Sungbok Lee, Tsuneo Kato and Shrikanth Narayanan*

Speech Analysis and Interpretation Laboratory (SAIL),
USC Viterbi School of Engineering,
University of Southern California, Los Angeles, California 90089, USA
sungbokl@usc.edu, tsuneo.kato@sipi.usc.edu, shri@sipi.usc.edu

## Abstract

We investigate whether articulatory movement trajectories follow the nonlinear invariant relationship between the tangential velocity and the curvature, i.e. the one-third power law. The power law holds for articulatory trajectories of phonetic rendering, although the exponent is in the range 0.35–0.42 when averaged across speakers and emotions but is relatively invariant under speaker/emotion differences. However, the velocity gain factor, a proportional constant in the power law, is sensitive to such variations. While the power law reflects some common articulatory movement characteristics for phonetic rendering across speakers and emotions, the prosodic variation/speaker idiosyncrasy are reflected in the velocity gain factor.

**Index Terms**: articulatory trajectory, one-third power law, velocity gain factor, emotional speech production

## 1. Introduction

Movement trajectory of the endpoint (e.g., pen tip) associated with a hand-related motor tasks such as writing or drawing are curvilinear in nature whose shape or geometry is related to the current task goal. A number of studies [1,2,3] have shown that a systematic relationship exists between the tangential velocity and the curvature along the trajectory. The relationship is known as the one-third power law which dictates that the tangential velocity along the trajectory is proportional to the one-third power of the radius of curvature (i.e., the reciprocal of the curvature) and the proportional constant is called the velocity gain factor [see Eq. (1)]. The power law represents an invariance that the movement velocity is being regulated by the motor control system in conjunction with the shape or the geometry of the trajectory. Interestingly, it has been found that the power law is a preferred mechanism not only in the task-oriented motor control but also in the visual perception of uniform (i.e., constant-velocity) movements [2]. This finding has been regarded as a supporting evidence of a coupling between motor and perceptual processes. What is an important point in these findings is that any consistent relationship between quantities related to kinematic and geometrical properties of a movement trajectory can provide an important clue for understanding the nature of the underlying motor control regime and related perceptual process.

Speech-related articulatory movement data acquired by a point-tracking device such as Electromagnetic Articulography (EMA) show that they also exhibit complex trajectory shapes during speech production as illustrated in Figure 1. It is thus natural to ask whether such point-based articulatory movements for phonetic rendering follow the power law. Tasko and Westbury [4] were the first to investigate the power law of speech-related movements using X-ray microbeam trajectory data. The study found that the power law also holds for articulatory movement trajectories, although the exponents are not exactly one-third. The range of the exponent value was 0.34 - 0.44 across all speakers and articulators examined in the study.

In the current study we also investigate the power law associated with speech articulations based on the analysis of an EMA articulatory trajectory database. Our investigation, however, considers not only normal speech but also emotional speech production. Since the knowledge related to articulatory movement control is an essential ingredient to model or to simulate human-like speech production, it is desirable to know the behavior of the power law under various speech production conditions. For instance, it has been found that the tongue tip and the jaw movement ranges and velocities increase when speakers are in angry emotion or in annoyed speaking state [5,6]. It is interesting to know how these prosodic articulatory variations affect the power law. Therefore, two main foci of this study are to verify the previous finding in [4] and to examine how the power law behaves when speech production deviates from the normal or neutral condition. To our knowledge, this is the first investigation of the power law associated with emotional speech production which is in a sense more extreme articulation and thus it is expected that the results would provide some insights on the behaviors and generalizability of the power law to various speaking conditions.

## 2. Methods

### 2.1. Speech material

EMA sensor trajectory data analyzed in this study were collected from one male (AB) and two female speakers

(JN, LS) using the Carstens' Ag200 EMA system [5]. The EMA system samples articulatory data at 200Hz and acoustic data at 16kHz. Two reference sensors for head-movement correction were put on the nose and the center of upper incisors. Three articulatory sensors were attached on the tongue tip, the lower lip and the lower incisor (i.e., the jaw). Four different emotions—neutral, angry, sad and happy—were simulated by the subjects. They produced a set of 10 sentences, generally neutral in semantic content, and produced each sentence five times in a random order for a given emotion category, yielding 200 utterances (10 sentences x 5 repetitions x 4 emotions) per subject. While such simulated emotion productions are known to be different from spontaneous unscripted productions, they are useful in providing a controlled approach (similar to the wide use of read speech in phonetic experiments).

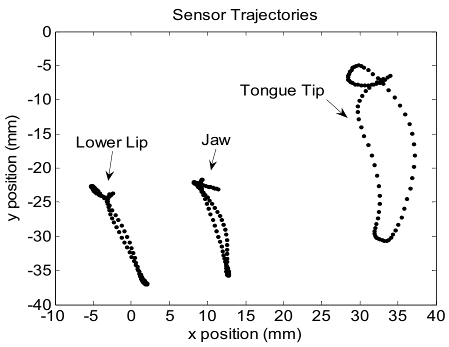


***Figure 1**: Three sensor trajectories for word "doctor." It is clear that the articulatory trajectories are curvilinear in general and that the movement speed (i.e., distances between dots) varies in conjunction with the curvature of trajectory.*

The raw articulatory data including audio obtained by the EMA system were assembled into matlab data files. During the post-processing, each trajectory data was smoothed after correction for head movement and rotation to the occlusal plane so that the x-axis is parallel to the subject's occlusal plane. Each sensor trajectory signal was then differentiated in order to obtain velocity as well as acceleration components which were smoothed with a 9-th order Butterworth filter of cutoff frequency 15-Hz.

Each utterance was evaluated for the target emotion quality by the three native American English speakers on a one (lowest emotion quality) to five (highest) point scale. Among 600 utterances, 503 utterances were scored 3 or more points by at least one evaluator and were analyzed in the current study.

### 2.2. Measurements of parameters associated with the power law.

The one-third power law can take several different forms (c.f., [1]). In this study, it is defined by the equation

$$V(t) = KC(t)^{-\alpha}, \quad (1)$$

where V(t) and C(t) are the tangential velocity and the curvature of trajectory at time t, respectively, and α is the exponent and K is a proportion constant called velocity gain factor (VGF) [1]. The tangential velocity is computed by

$$V(t) = \sqrt{(x'(t)^2+y'(t)^2)}, \quad (2)$$

and the curvature is computed by the formula

$$C(t) = |x''(t)y'(t)-y''(t)x'(t)|/(x'(t)^2+y'(t)^2)^{3/2}, \quad (3)$$

where $x'(t)$, $y'(t)$ and $x''(t)$, $y''(t)$ are velocity and acceleration terms of x and y movements, respectively, along the trajectory. By taking a natural logarithm, Equation (1) can be converted into the equation of line,

$$\ln(V(t)) = -\alpha \ln(C(t)) + \ln K. \quad (4)$$

Then, the exponent α and the velocity gain factor K can be estimated from the slope and the intercept of the regression-line fitted to a set of data points (lnC(t), lnV(t)). Example plots of the data distribution and the regression-line fit are shown in Fig. 2.

It is noted that the exponent and velocity gain factor are estimated at the utterance level in the current study. That is, the end-pointed whole utterance interval is the segment unit in which the two trajectory-related parameters are estimated for the tongue tip, the lower lip and the jaw movements. For statistical treatments, the utterance-level parameter values are arranged into a table as a function of subject and emotion, and multivariate ANOVAs are performed for significance tests of the speaker and emotion factors.

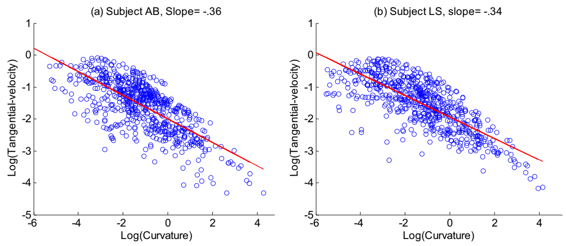


***Figure 2**: Regression-lines fitted to two scatter plots of the tangential velocity and curvature of the tongue tip trajectory for subjects AB and LS. The data points are obtained from one instance of utterance #7 in neutral emotion. Strong negative correlation between the tangential velocity and the curvature is clear. The exponent α (i.e., the absolute value of slope) is near 1/3, although slightly higher in general.*

## 3. Results

Our major concern in this study is to examine the behaviors of the power exponent α and the velocity gain factor (VGF) as functions of speaker and emotion. In Table 1, the averaged values of the exponent and the velocity gain factor across all speakers and utterances are listed as functions of articulator and emotion. In Figs. 3, 4 and 5, the effects of speaker and emotion on the parameters are graphically shown in box plots for the tongue tip (TT), the jaw (JAW) and the lower lip (LL), respectively.

***Table 1****: The values of the power exponent α and the velocity gain factor (vgf) as a function of articulatory movement point and emotion. Averaged values across subjects and utterances are shown and the numbers in parentheses are standard deviations.*

| | | Neutral | Angry | Sad | Happy |
|---|---|---|---|---|---|
| TT | α | .366 (.042) | .371 (.039) | .369 (.036) | .375 (.041) |
| | VGF | .143 (.017) | .176 (.033) | .121 (.017) | .151 (.029) |
| JAW | α | .353 (.032) | .368 (.036) | .364 (.033) | .371 (.033) |
| | VGF | .092 (.023) | .099 (.034) | .072 (.016) | .093 (.031) |
| LL | α | .411 (.043) | .407 (.033) | .420 (.032) | .413 (.034) |
| | VGF | .088 (.019) | .107 (.011) | .074 (.016) | .084 (.016) |

Some general trends can be observed from the table and the figures. The range of the averaged exponent values is 0.35 – 0.42 across articulators and emotions, although the exponent itself is not exactly one-third but a little bit higher in general. This is especially true for the lower lip movement. However, the tongue tip and the jaw show a similar tendency on average. It is also observed that the averaged exponent values across emotion for a given articulator are quite similar. This indicates that the exponent is relatively stable against articulatory and prosodic modifications of speech production associated with emotion expression [5].

However, that is not the case for the velocity gain factor. It is quite variable across emotions, and the figures also suggest that it varies much across speakers. The velocity gain factor seems quite affected by the personal articulation characteristics as well as by emotion-dependent variations in speech production. Multivariate ANOVAs reveal that in the case of the tongue tip (TT), the exponent α vary significantly with speaker ($F(2) = 14.347$, $p =.000$) but not with emotion ($F(3) = 1.181$, $p = .316$) and emotion and subject interaction is also not significant ($F(6) = 1.534$, $p = 0.165$). However, the velocity gain factor K significantly varies with emotion ($F(3)= 172.384$, $p=.000$) and subject and emotion interaction is also significant ($F(6)=19.745$, $p=.000$). This confirms that the exponent α of the tongue tip movement is relatively stable under emotion-related modification of speech articulation.

For the jaw and lower lip movements, the exponent α also significantly varies not only with speaker but also with emotion. Therefore, both the exponent α and the velocity gain factor can be affected by the personal differences in speech articulation but the former is more stable than the latter under articulatory and prosodic variations in speech production. This is also supported by the statistical observations that the F-values of the main effect of emotion and the interaction effect of speaker and emotion in the multivariate ANOVAs are always much smaller for α, which is a supporting evidence that speaker- and emotion-dependencies are always less in the case of α. Therefore, it is reasonable to state that the exponent α reveals some invariance characteristics under speaker and prosodic articulatory variations. Examination in a smaller segmental level such as word-or syllable-level may reveal the further details of the behavior of the power law in terms of the invariance.

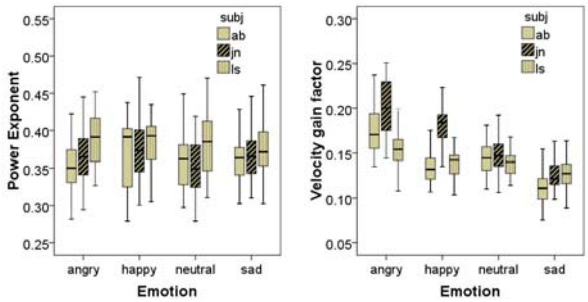


***Figure 3****. Box plots of the power exponent value (left) and the velocity gain factor (right) are shown for the tongue tip (TT) as a function of speaker and emotion. The variability associated with the power exponent across emotion is much less than that of the velocity gain factor, implying relatively stable α values across emotions.*

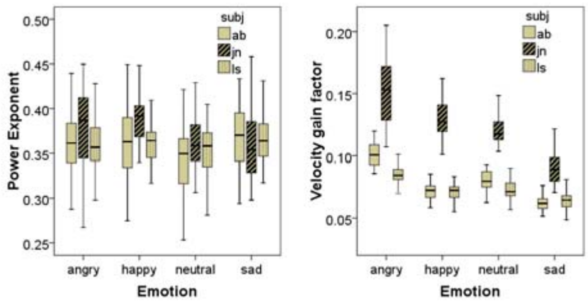


***Figure 4****. Box plots of the power exponent (left) and the velocity gain factor (right) are shown for the jaw (JAW) track. The tendency is similar to the case of the tongue tip.*

The lower lip velocity gain factor is relatively large for angry emotion and small for sad emotion. This trend is clearer for the tongue tip and the jaw than for the lower lip. The tendency can be attributed to the findings in previous studies that the tongue tip and the jaw movement range and the velocity become larger when speakers are in angry emotion or in annoyed state [5,6]. One underlying reason can be inferred from the fact that the tangential velocity increases only when the trajectory is relatively straight. That is, the curvature itself does not vary much during such movement intervals and thus the velocity gain factor should increase in order to match the increased tangential velocity. This is supported by the statistical observation that the correlation between the velocity gain factor and the median value of the tangential velocity is very high (r = .88 for the tongue tip and r = .89 for the jaw) across all speakers and utterances.

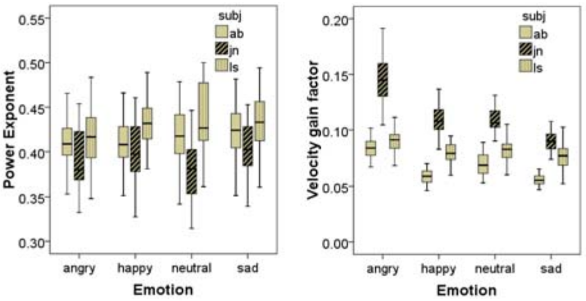


***Figure 5****. Box plots of the power exponent (left) and the velocity gain factor (right) are shown for the lower lip (LL). Inter-speaker variability in both the exponent α and the velocity gain factor is conspicuous for the lip movement. Also, the exponent value is largest on average among the three articulators considered.*

## 4. Discussion

There is no a priori reason that the shape or the geometry of a trajectory and the movement velocity along the trajectory should follow a certain rule such as the one-third power law. Trajectory shape and movement velocity are mutually independent physical quantities that can vary arbitrarily with respect to each other. However, this study verifies that such an invariant relation (at least in approximation) exists in articulatory movement trajectories, indicating a similar control regime which constrains the tangential velocity according to the geometry or the shape of the trajectory exist in the speech motor control system as in hand-related motor tasks such as writing and drawing.

However, when compared to the hand motor tasks [1,3], the parameters associated with the power law in speech articulation are more susceptible to perturbations in measurement environment such as speaker and emotional variations. Specifically, this study indicates that prosodic articulatory variation and speaker idiosyncrasy are mainly reflected in the velocity gain factor. Whereas, the relative invariance of the exponent α implies some common articulatory movement characteristics for phonetic rendering across speakers and emotions. Therefore, separated and meaningful interpretations of the exponent and the velocity gain factor are possible in the power law associated with speech production.

There exist articulator-specific differences in the realization of the power law. The low lip trajectory exhibits a significantly higher value of the power exponent α when compared to that of the tongue tip or the jaw. In the case of the velocity gain factor, it is the tongue tip that takes a significantly higher value. The former seems related to the difference in physical and physiological properties of the muscles which are responsible for altering the lip configuration. The difference in stiffness and viscosity should affect the trajectory movement characteristics in the realization of power law by the lower lip following the current motor command. The latter should be related to the fact that the tongue tip velocity is usually much higher than those of the jaw or the lower lip.

In the current study the power law is examined at the utterance level. Therefore, only averaged behaviors have been studied. However, future investigations should be conducted at a finer level such as syllable in order to examine the phonemic or gestural level behaviors of the power law, to uncover the underlying dynamics of the phonemic gestures. For instance, a constriction-forming articulatory gesture is frequently modeled using the second-order damped harmonic oscillator and therefore the articulatory motion governed by the equation should conform to the power law. This and other related topics are our research directions in the near future.

## 5. Conclusions

We investigated whether articulatory movements for speech production follow the invariant relationship between the tangential velocity and the curvature, i.e., the one-third power law. It is confirmed that the articulatory movement trajectories follow the power law as in hand-related motor tasks, although the exponent itself is not exactly one-third but a little bit higher and more variable in general for speech articulators. The range of the averaged power exponent values across speakers and emotion categories is 0.35 – 0.42. Both the exponent and the velocity gain factor are affected by the personal idiosyncrasies in speech articulation but the former is much more stable than the latter under prosodic variations in speech production, especially for the tongue-tip movement. Examination at the level of smaller segmental units such as word- and syllable can reveal further details and insights of the behavior of the power law.

## 6. Acknowledgements

Work supported in part by NIH, Army and ONR.

## 7. References


[1] Viviani, P., & Cenzato, M., “Segmentation and coupling in complex movements,” J. Exp. Psychol.: Hum Percept. Perform. 11, 828–845, 1985.

[2] Viviani, P., & Stucchi, N., “Biological movements look uniform: evidence of motor-perceptual interactions,” J. Exp. Psychol.: Hum Percept. Perform. 18, 603–623, 1992.

[3] Viviani, P. Flash, T., “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning,” J. Exp. Psychol.: Hum Percept. Perform., 21, 32-53, 1995.

[4] Tasko, S.M. and Westbury, J. R., “Speech-curvature relations for speech-related articulatory movement”, Journal of Phonetics, 32, 65-80, 2004.

[5] Sungbok Lee, Serdar Yildirim, Abe Kazemzadeh, and Shrikanth Narayanan, “An articulatory study of emotional speech production,” In Proceedings of Eurospeech, Lisbon, Portugal, October 2005.

[6] D. Erickson, O. Fujimura, B. Pardo, “Articulatory correlates of prosodic control: Emotion and emphasis,” Language and Speech, 41, 395-413, 1998.