Articulatory Effort in Different Speaking Rates

Chen-huei Wu, Chilin Shih

Department of Linguistics, University of Illinois, Urbana-Champaign, USA

wu9@illinois.edu, cls@illinois.edu

Abstract

Articulatory effort has been widely discussed under different speech production models. A good way to explore this issue is to test the hypothesis that speakers naturally desire to minimize articulatory effort during fast speech rates, such as undershoot of the target sound. This study demonstrates articulatory effort at different speaking rates by examining articulatory trajectory using the Electromagnetic Articulograph AG500. The results suggest that the articulator undershoots a target and the valley of the target might go deeper while the velocity increases.

Index Terms: articulatory effort, speech production

1. Introduction

Articulatory effort has been widely discussed with various speech production models. One aspect views speech production as a compromise between ease of articulatory effort and communication accuracy. A good way to understand the mechanism of speech production and related issues is to examine articulatory movement during different speech rates. This paper investigates production data related to different speech rates, using an Electromagnetic Articulograph AG500 (hereafter EMA). The goal of this study is to illustrate articulatory effort in terms of the trajectory of the tongue movement.

The speech production model as a compromise between economy of articulation and speech intelligibility has been promoted in the segmental domain to account for phonetic variations of speech sounds [1, 2, 3, 4, 5, 6]. The speaker would minimize the articlutory effort for ease of production, while the speakers would still need to maintain production accuracy. The Hypo- & Hyper-articulation (H&H) model of speech production [7] posits that speech as a motor activity is constrained by economy of articulatory effort according to the communication situation. That is, when intelligibility needs not to be improved, speech production will tend toward a reduced form, i.e. hypo-articulation. On the other hand, when intelligibility could be decreased, the speaker tends to increase articulatory effort in a more hyperarticulated mode of production. As a result, overarticulation may lead to enhancement of the acoustic distance or salience of sound contrasts. Therefore, depending on different speaking conditions, it has been shown that clear speech is characterized with greater clarity, longer duration of sound segments, and expanded vowel space [8].

Further, the vowel undershoot model [9] hypothesizes that undershooting of the target position is an effect of speaking styles and conversational demands. Shih [6] has shown that the phonetic reduction of tone 2 sandhi in Mandarin at different prosodic positions in the surface forms meet the demands in speech communication by simulating an articulatory-based prosody model the Soft *Template Markup Language* [6, 10, 11] . Figure 1 (cited from [6]) shows three schematic cases that explain the discrepancies between an intended target (dashed lines) and their surface realization (solid lines with arrow) which exert less articulatory effort.

In example (a), a moving articulator continues moving in the same direction. In example (b), the smooth trajectory represented by the solid arrow is an approximation that takes less effort and less time. The solid line with the arrow in example (c) is one of many possible approximations to realize two targets.

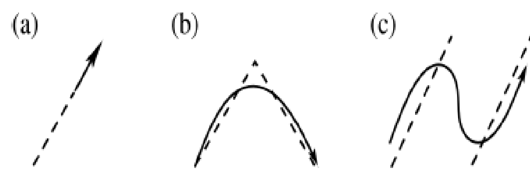

Figure 1: *Hypothetical targets (dashed lines) and examples of target approximation (solid lines) that follow a trajectory that requires less articulatory effort.*

2. Hypotheses

According to the H & H model, compared to clear speech, normal speech is produced with less articulatory effort and is characterized by, but not limited to reduced vowel space, shorter durations and articulatory undershoot. Kuehn and Moll [12] reported that if the speaking rate increases, some speakers tend to increase articulatory velocities with less articulatory undershoot, while others produce more articulatory undershoot and decrease velocities.

In the present study, speakers produced a test sentence with multiple repetitions in normal speaking conditions with different speech rates. It is expected that in order to increase the speaking rate, the articulatory trajectory will be understoot to some degree.

3. Methodology

3.1. Subject and speech materials

Four young adult native English speakers without speech or hearing deficits participated in the experiment, one male and three females. The test sentence 'The quick brown fox jumps over the lazy dog' was repeated 13 times by each speaker. The first repetition was not used for analysis since subjects were getting used to the EMA.

3.2. Procedure

The data were obtained from a Carstens 3D Eletromagnetic Articulograph AG500. This apparatus consists of the EMA cube with six transmitter coils generating magnetic fields at different frequencies at defined orientations, 12 sensors and channels, and a computer with an automatic calibration unit, real time display, and head movement correction systems, and a receiver. Six transmitter coils generate alternating magnetic fields at different frequencies. Sensors, which are built of small coils, are fixed onto the articulators of the subject. The alternating currents induced by the alternating magnetic fields have different strengths as a function of the distance and the angle of the sensor to the respective transmitter coil. A microphone mounted on a microphone stand and extended on a boom stick is connected to the audio box, transferring the speech signal to the synchronizer. All articulatory and acoustic data are acquired and synchronized simultaneously through the EMA system and stored on the hard disk of the computer.

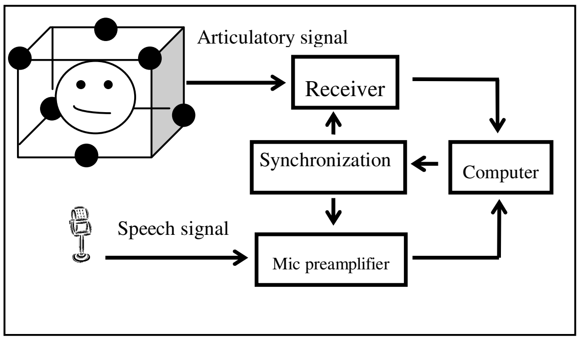

Figure 2: *Schematic diagram of EMA setup*

The participants were seated on a wooden chair and their heads were positioned in the center of the EMA cube. The filled circles in Figure 2 schematize the approximate locations of the sensors used in this experiment. In order to track the movements of the tongue, lips and the jaw, sensors were mounted on the tongue tip (TT, 1cm behind apex), the tongue body (TM, 2cm behind the tongue tip sensor) and the tongue back (TB, 2cm behind the tongue body sensor), as well as on the lower incisor (LI), the upper lip (UL), the lower lip (LL), and the right and left corners of the lips. Other sensors were attached to the bridge of the nose and the left and right tragi as reference points to normalize head movement.

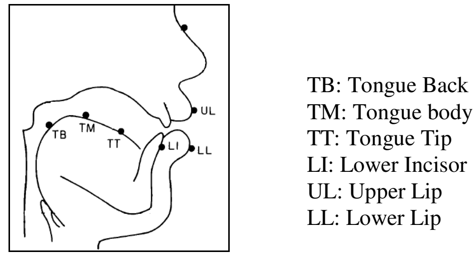

Figure 3: The schematic view with sensors on the tongue and lips as well as a reference point on the bridge of the nose

After obtaining the data, head movement corrections were carried out. The speech was segmented into phonetic and word levels using an acoustic aligner [14] and manually checked by the author.

4. Results and Discussion

Figure 4 demonstrates the articulatory trajectory of the tongue body movement in Z-position, the inferior and superior dimension, produced by subject 1. The upper panel shows the trajectory of the fastest speaking rate and the lower panel shows the slowest speaking rate in the production of subject 1. The solid black circle in the fast speaking rate shows the smooth trajectory compared to the same target in the slow speaking rate. The valley of the word ‘brown’ was deleted in the fast speaking rate.

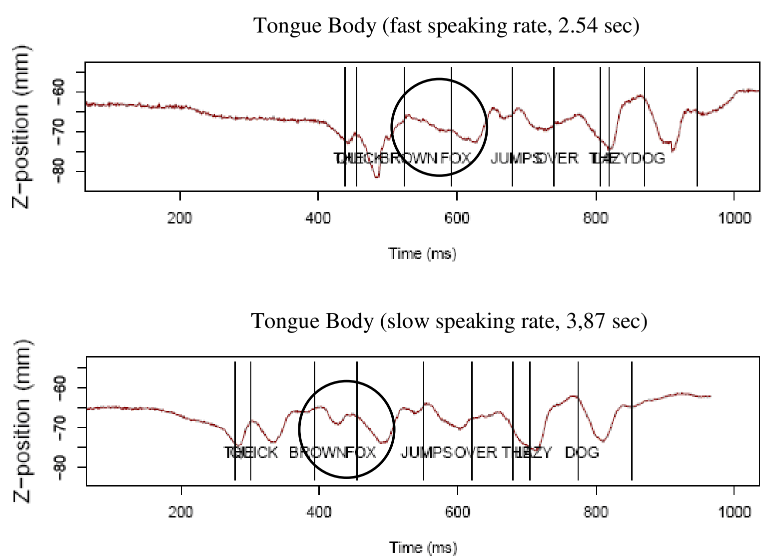

Figure 4: Articulatory trajectory of the tongue body movement in Z-position produced by subject 1.

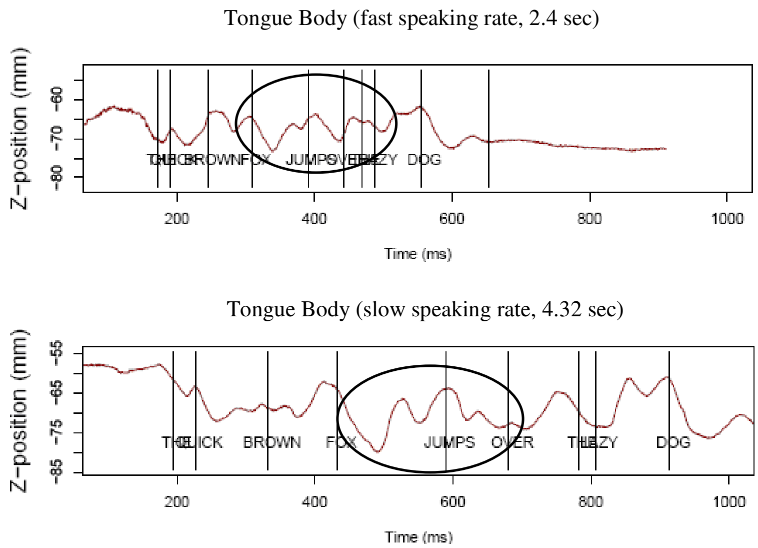

Figure 5: Articulatory trajectory of the tongue body movement in Z-position produced by subject 2.

Figure 5 demonstrates the aritculatory trajectory of the tongue body movement in Z-position produced by subject 2. When the trajectory goes up, it shows that the tongue body moves upward and vise versa. The circle in the fast speaking rate shows the undershoot of the valley in the words 'fox' compared with that in the slow speaking rate. Also, the smooth trajectory of the word 'jumps' suggests less articulatory effort at the fast speaking rate than that at the slow speaking rate. As we can see, there are two valleys of the word 'jumps', while they are deleted in the fast speaking rate.

Figure 6 shows articulatory trajectory of the tongue body movement in Z-position produced by subject 3. The first circle shows the smooth trajectory of the word 'quick' during fast speech, compared to that in the slow speech. The second circle in the fast speaking rate illustrates the undershoot of the second valley in the word 'fox', while the target is clearly realized in the slow speaking rate. Again, the third and forth circles present the undershoot of the valley in the word 'lazy' and 'dog', respectively, during the fast speaking rate, while the valley is much deeper in the slow speaking rate.

Figure 7 shows the articulatory trajectory of the tongue body movement in Z-position produced by subject 4. The circle shows the undershoot of the target in the fast speaking rate, compared to that in the slow speaking rate. The trajectory of other words are approximately the same at both fast and slow speaking rates.

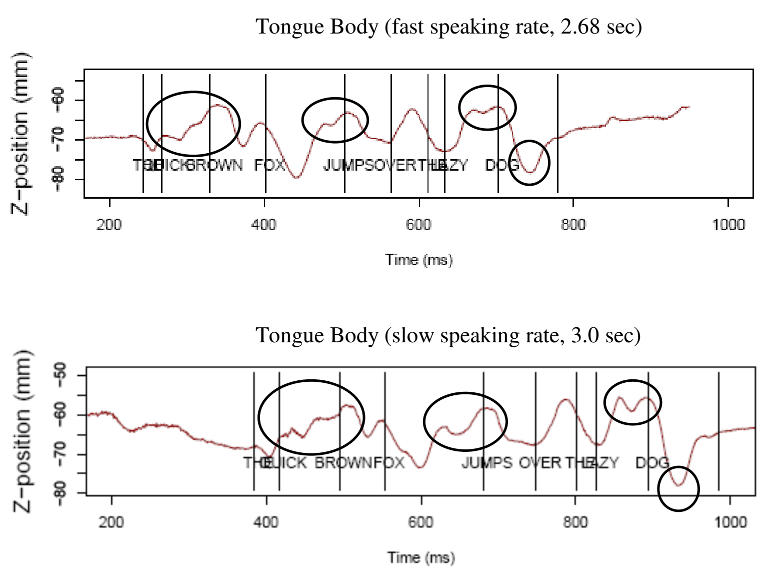

Figure 6: Articulatory trajectory of the tongue body movement in Z-position produced by subject 3.

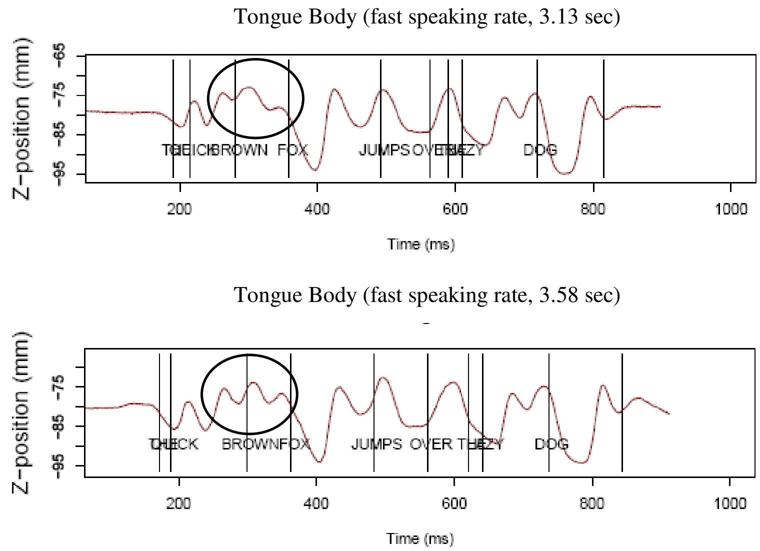

Figure 7: Articulatory trajectory of the tongue body movement in Z-position produced by subject 4.

In previous studies [5, 12, 13], it has been found that acoustic characteristics and articulatory efforts will be different in different speaking styles. Clear and slow speech requires more articulatory effort, such as greater intensity, expanded vowel space, longer duration of speech segments and greater articulatory trajectories. In order to increase speaking rates, some speakers will increase articulatory velocities with little articulatory undershoot or little change in movement distance; others may not increase articulatory velocities and have more articulatory undershoot.

In the current study, all four speakers show articulatory undershoot when they increase speaking rates. The result is along the line with the H & H model.

5. Conclusions

This study explores articulatory effort during slow and normal speech by examining articulatory data gathered using the Electromagnetic Articulograph AG500. The finding provides the articulatory evidence that the trajectory of the tongue movements are undershot to satisfy the constraint, ease of articulation, when the velocity increases.

6. Acknowledgements

This work is supported in part by NSF IIS-0623805, NSF IIS-0534133, and a Critical Research Initiatives Grant from UIUC. Any opinions, findings and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the University of Illinois. Also, I thank the participants and Hsin-Yi Lu and Shawn Chang for their help preparing the experiment.

7. References

[1] Nelson, W. L., "Physical principles for economics of skilled movements", Biological Cybermetics 46(2): 135-1147, 1983.

[2] Ohala, J., "The segment: primitive or derived? Gesture, segment, prosody", in G. J. Docherty and D. R. Ladd [Eds], Papers in Laboratory Phonology 2, 166-183, Cambridge, 1992.

[3] Guenther, F. H., "Speech sound acquisistion, coarticulation, and rate effects in a neural network model of speech production", Psychological Review 102(3): 594-621, 1995.

[4] Guenther, F. H., Hampson, M. and Johnson, D., "A theoretical investigation of reference frames for the planning of speech movements", Psychological Review 105(4): 611-633, 1998.

[5] Perkell, J. S., Zandipour, M., Matthies, M. L. and Lane, H., "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues", Journal of the Acoustical Society of America, 112(4): 1627-1641, 2002.

[6] Shih, C., "Linking phonology and phonetics: An implementation model of tones", Interfaces in Chinese Phonology, 99-120, 2008.

[7] Lindblom, B., "Explaining phonetic variation: A sketch of the H & H theory", in W. J. Hardcastle and A. Marchal [Eds], Speech production and speech modeling, 403-439, Kluwer Academic publishers, 1990.

[8] Moon, S.-J., "An acoustic and perceptual study of undershoot in clear and citation-form speech", Ph.D dissertation, University of Texas at Austin, 1991.

[9] Lindblom, B. "Spectroraphic study of vowel reduction", Journal of the Acoustical Society of America 35:1773-1781, 1963.

[10] Kochanski, G., and Shih, C. "Prosody modeling with soft templates", Speech Communication 39.3-4:311-352.

[11] Kochanski, G., Shih, C., and Jing, H. "Quantitative measurement of prosodic strength in Mandarin", Communication 41.4: 625-645.

[12] Moon, S. –J., and Lindblom, B. "Interaction between duration, context, and speaking style in English stressed vowels", Journal of the Acoustical Society of American, 96(1): 40-55, 1994.

[13] Kuehn, D. P., and Moll. K. L., "A cineradiographic study of ve and ev articulatory velocities," Journal of Phonetics 4: 303-320, 1976.

[14] Yuan J. and Liberman M., The Penn Phonetics Lab Forced Aligner, 2008. http://wms-609.sas.upenn.edu/research/alignment/align.htm