6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

RELATIONS BETWEEN UTTERANCE SPEED AND ARTICULATORY MOVEMENTS

OKADOME, Takesi *Tokihiko Kaburagi* *Masaaki Honda*

NTT Laboratories/CREST, JST

3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan

houmi@idea.brl.ntt.co.jp, kabu@idea.brl.ntt.co.jp, hon@idea.brl.ntt.co.jp

ABSTRACT

A relation between utterance speed and the amount articulatory behavior for each phoneme reduces is investigated on the basis of the articulatory data taken by using a magnetic sensory system. The relation between the utterance speed and the change of the utterance timing for each phoneme is also focused on. Two linear-regression models are proposed: one predicts reduced articulatory movements and the other adjusts the utterance timing of each phoneme according to the utterance speed.

10.21437/Eurospeech.1999-38

1. INTRODUCTION

Articulatory-based speech-synthesis requires high-fidelity generation of articulatory behavior. In particular, the control of the speech rate requires knowing the relation between utterance speed and articulatory movement. Among the most prominent properties of fast speech is reduction (Levelt, 1989). Previous work, phonological or acoustical, focused on vowel reduction and tried to clarify the rules of vowel reduction, that is, when does a vowel reduction occur (for example, Levelt, 1989) or how does vowel reduction occur (for example, Lindblom, 1983). Reduction may, however, also occur for consonants. Also, previous work did not predict how much articulatory movements reduce. for each phoneme. On the basis of the articulatory data taken by using a magnetic sensory system, this paper presents the relation between utterance speed and the amount articulatory behavior for each phoneme reduces, and proposes a linear regression model which predicts the amount articulatory behavior will reduce.

The control of speech timing also plays an important role in speech. We, therefore, also focus on the relation between utterance speed and the change of the utterance timing for each phoneme. Work derived from Harris et al. (1986), that is based on acoustic data, found some kinds of the temporal invariance. We describe another kind of temporal invariance with regard to the articulatory data. Kohler (1986) presented a linear model of the timing production that is based on acoustic data. We also propose a linear model which adjusts the utterance time of each phoneme according to the utterance speed.

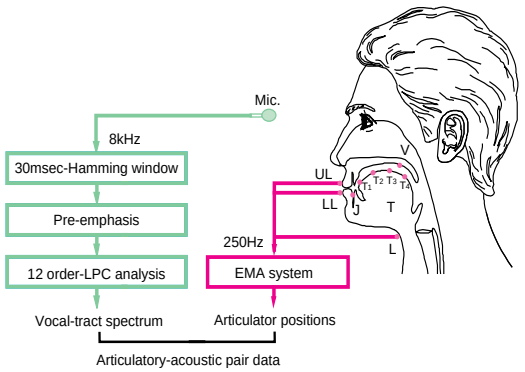

Figure 1. The experimental system.

2. EXPERIMENT

The material used in the experiment consists of 144 sentences in which three male subjects read under three utterance-speed conditions: fast, normal, and slow. With 250 Hz sampling in both the vertical and horizontal orientations, we observed the following 9 points on the articulator: the jaw (J), the upper lip (UL), the lower lip (LL), the tongue points (T1, T2, T3, and T4), the velum (V), and the larynx (L) (See Figure 1). For the observed data, we first assigned the articulatory timing for each phoneme. The time assignment was done by marking manually the time at which the kinematic feature of each phoneme was most remarkably appeared. For example, we put the marker for /b/ to the time at which the lips are closed. We call the time assigned for a phoneme the *articulation timing for the phoneme*. The mean utterance velocities of subject HM in normal, fast, and slow speeds are 13.38, 17.18, and 11.78 phonemes/sec., respectively; Those of subject OT are 15.18, 17.26, and 13.04 phonemes/sec.; Those of subject MT are 12.76,

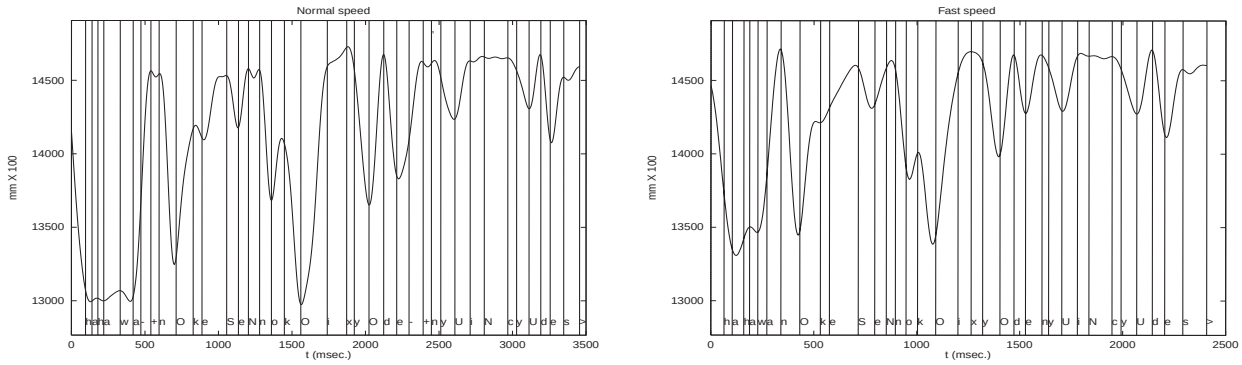

Figure 2. Tongue-tip trajectories in the vertical orientation of the normal (left) and fast (right) speech.

15.88, and 10.88 phonemes/sec.

3. ARTICULATORY REDUCTION

3.1. Reduction phenomena

Figure 2 shows the tongue-tip trajectories of subject HM in the vertical orientation when he read a sentence at fast and normal speeds. Although the trajectory in the fast speech is similar to that in the normal speech, a comparison of the trajectories shows that articulatory behavior for some phonemes in the fast speed reduces. For example, the vertical positions of the tongue tip at the time for the long vowel /O/ in the fast speech are remarkably higher than those in the normal speech.

Pairwise-comparison tests for differences between the fast and normal speed utterances at the articulation timing for each phoneme reveal that the ratio of significant differences to non-significant diffirences is 58.2% on average, even though the ratios are different among the subjects (Subject HM: 49.4%; TO: 36.2%; MT: 89.0%). Those between the slow and normal speed utterances also show the ratio is 62.3% on average (Subject HM: 58.3%; TO: 37.3%; MT: 91.3%). Although the ratio for the vowel is higher than that for consonant, we can see that reduction for each type of consonant also occurs. (vowel: 97.5%; consonant: 69.3%).

Figure 3 shows that position differences in the vertical orientation between the fast and normal speed utterances at each articulatory organ for each type of phonemes. For the vowel, velar, and alveolar, the differences at the low lip and tongue body are relatively larger than those at the tongue tip. In particular, the differences at the tongue tip for the velar and alveolar are extremely small. On the other hand, for the labial, the differences at the lip and front tongue are smaller than the back tongue. These results are phonologically plausible.

3.2. Predicting reductions

On the basis of the experimental data, we construct a linear-regression model for predicting the position and velocity of the articulatory organ for each phoneme in a speech at an arbitrary speed V. The linear predicting model is expressed by:

$$\begin{aligned} y &= a_1 x + a_2 x' + a_3 x'' + a_4 t + a_5 V_n + a_6 V, \\ w &= b_1 v + b_2 v' + b_3 v'' + b_4 t + b_5 V_n + b_6 V, \end{aligned}$$

where y is a position of the articulatory organ for a phoneme in the speech at an arbitrary utterance speed V, x is the position of the articulatory organ for the phoneme in the normal speech, x' is that for the previous phoneme in the normal speech, x'' is that for the successive phoneme in the normal speech, t is the interval times between the previous and the successive phonemes in the normal speech, and V_n is the average velocity of the articulatory organ for the phoneme in the normal speech; w is a velocity of the articulatory organ for a phoneme in the speech at an arbitrary utterance speed, v is the velocity of the articulatory organ for the phoneme in the normal speech, v' is that for the previous phoneme in the normal speech, v'' is that for the successive phoneme in the normal speech. The regression coefficient for positions is 0.80 and that for velocities is 0.76 on average, which are rather high.

To evaluate the prediction model, we used 6 sentences which are not in the set of the 144 sentences used for constructing the model. The average error of predicted positions for fast speed utterances is 1.35 mm and that for slow speed utterances is 1.53 mm, whereas the average error without prediction for for fast speed utterances is 1.74 mm and that for slow speed utterances is 1.82 mm. Figure 4 shows the average errors of positions of the tongue tip for each type of phonemes with and without prediction by the model. It also shows those of the low lip and jaw.

For a given sequence of phonemes, using the position and velocity for each phoneme in the sequence as a constraint, we can formulate the trajectory by calculating the *minimum-acceleration trajectory* for each point on the articulator that coincides with the extremum of the following cost function:

$$\frac{1}{2}\int_0^{t_f} \left(\left(\frac{d^2x}{dt^2}\right)^2 + \left(\frac{d^2y}{dt^2}\right)^2 \right) dt, \qquad (1)$$

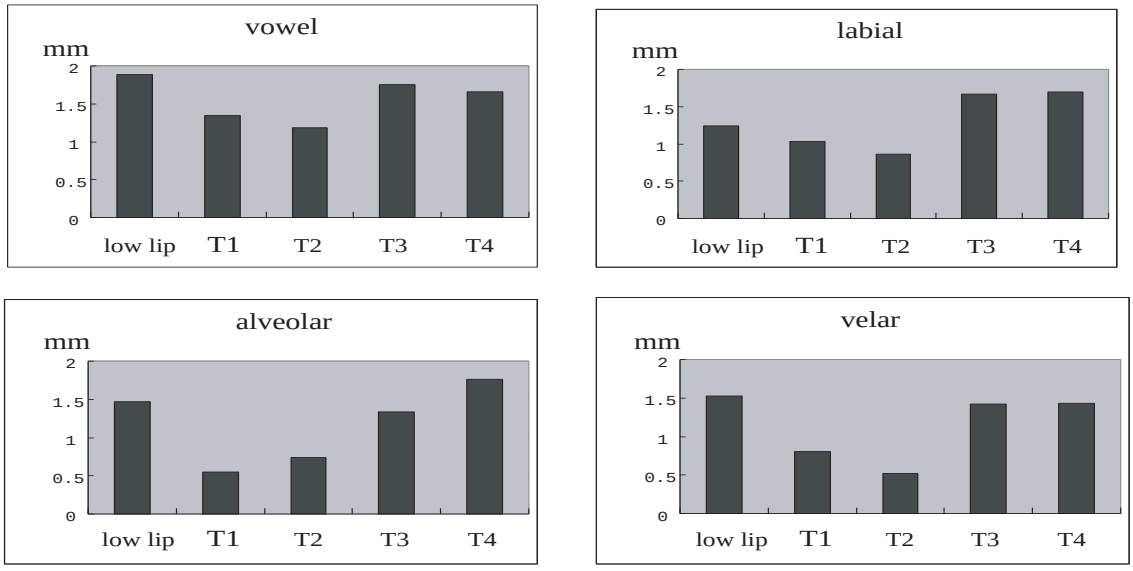

Figure 3. Position differences in the vertical orientation between the fast and normal speed utterances at each articulatory organ for each type of phonemes.

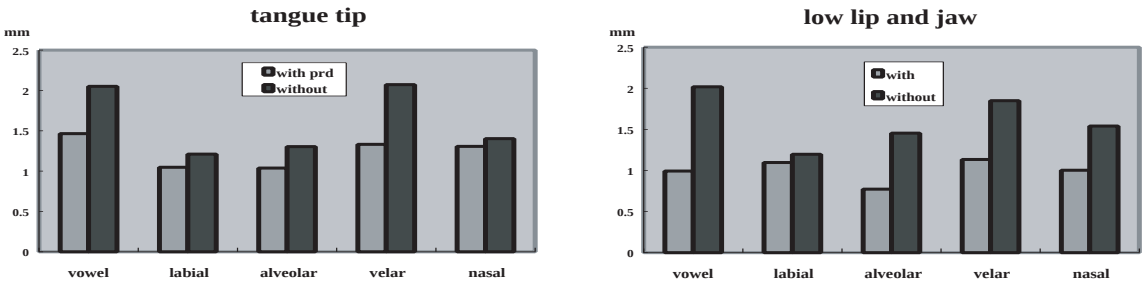

Figure 4. The average errors of positions of the tongue tip (low lip and jaw) for each type of phoneme with and without prediction by the linear-regression model.

where (x, y) are the time-varying Cartesian coordinates on the sagittal plane of the point on the articulator (c.f. Okadome et al., 1998). The average distance between an observed trajectory and a minimum-acceleration trajectory calculated using the position and velocity predicted by the linear-regression model is 1.53 mm for the fast speed utterance and 1.73 mm for the slow speed utterance. On the othe hand, The average distance between an observed trajectory and a minimum-acceleration trajectory calculated using the position and velocity without prediction is 2.12 mm for the fast speed utterance and 2.33 mm for the slow speed utterance (see Figure 5).

These results show that the linear-regression model permits us to effectively predict the articulatory reduction.

4. Articulation timing

4.1. An invariance

Let v_1 and v_2 be utterance speeds and let $x = p_1 p_2 \cdots p_n$ be a phonemic sequence. We define the *elasticity rate* of an articulation interval of the speech at v_1 to that at v_2 for x by

$$\frac{t_{v_1}(x) - t_{v_2}(x)}{t_{v_2}(x)},$$

where $t_v(x)$ denotes an interval time from p_1 to p_n of x in the speech at utterance speed v. Table 1 lists the average elasticity rates of the fast speech to the normal speech for the CV syllables and that for the VC syllables. It also lists that of the slow speech to the normal speech. This table shows that the average elasticity rate for all CV syllables is almost equal to that for all VC syllables.

4.2. Linear-regression model for timing

We also construct a linear-regression model for predicting the interval time between successive phonemes in a speech at an arbitrary speed. The

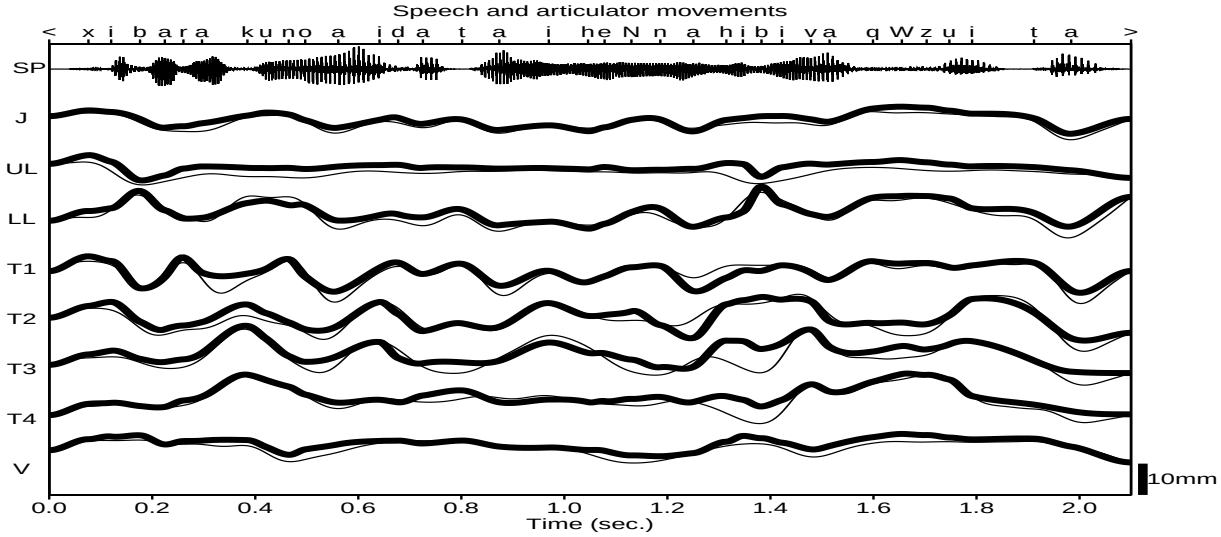

Figure 5. Observed (thin) and predicted (thick) trajectories based on the minimum-accelaration model with the linear-regression model for reduction.

Table 1. The average elasticity rates of the fast (slow) speech to the normal speech for all CV syllables and that for all VC syllables.

fast vs. normal	MH	OT	MT
CV	0.195	0.094	0.158
VC	0.192	0.094	0.167

slow vs. normal	MH	OT	MT
CV	0.090	0.110	0.121
VC	0.090	0.090	0.115

linear predicting model constructed based on the experimental data is expressed by:

$$u = c_1 t + c_2 t' + c_3 t'' + c_4 x + c_5 x' + c_6 x'' + c_7 V_n + c_8 V,$$

where u is the interval time between the articulation timings of two successive phonemes p_0, p_1 in a speech at an arbitrary utterance speed V, t is the interval between the two phonemes p_0, p_1 in the normal speech, t' is that between p_{-1}, p_0 in the normal speech, t'' is that between p_1, p_2 in the normal speech, x is the position (the value of a channel) of p_0 in the normal speech, x' is that of p_{-1} in the normal speech, and x'' is that of p_1 in the normal speech, and V_n is the average velocity of the articulatory organ for the phoneme in the normal speech. The regression coefficient for the intervals is 0.79 which is rather high.

For each of the two successive phonemes in the 6 sentences that we tested, we calculated the interval time between them using the linear-regression model. The results show that 70.1 percent of the predicted interval times were inside the permitted ranges, where a permitted range is defined to be the range from the observed interval minus 20 msec. to plus 20 msec.

5. CONCLUSION

This paper first presents the relation between utterance speed and articulatory reduction for each phoneme. It also focuses on the relation between the utterance speed and the change of the utterance timing for each phoneme. The paper also proposes a linear-regression model which predicts reduced articulatory movements and presents a linear model which adjusts the utterance time of each phoneme according to the utterance speed.

REFERENCES

1. Harris, S., B. Tuller, and J. A. S. Kelso (1986). Temporal invariance in the production of speech. In: *Invariance and Variability in Speech Processes*, LEA, Hillsdale, NJ, pp. 217-245.

2. Kohler, K. J. (1986). Invariance and variability in speech timing. In: *Invariance and Variability in Speech Processes*, LEA, Hillsdale, NJ, pp. 268-299.

3. Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

4. Lindblom, B. (1983). Economy of speech gestures. In: *The Production of Speech*, Springer-Verlag, New York, pp. 217-245.

5. Okadome, T., T. Kaburagi, and M. Honda (1998). Trajectory formation of articulatory movements for a given sequence of phonemes. *Proceedings of the 5th International Conference on Spoken Language Processing*, **7**, 3131-3134.