First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

SEQUENTIAL CONTROL MODEL OF SPEECH ARTICULATION IN PRODUCING WORD UTTERANCE

Naoki Kusakawa, Kiyoshi Honda, and Yuki Kakita

Kanazawa Institute of Technology
Kanazawa-Minami, Ishikawa, 921 Japan

ABSTRACT

Speech articulation is executed by motor commands to various speech muscles. The subsystem of motor programming autonomically outputs a sequence of commands according to intended signals for articulatory target. The present study examined such a function of motor program for coordinated speech gestures by speculating motor commands from electromyographic (EMG) data. A waveform separation technique was applied to extract component commands from ensemble average EMG waveforms. The data used in this study contained six tongue muscles and two other muscles, which were recorded during production of /əpVp/ word utterances. The motor score was obtained from the results, which demonstrated temporal and hierarchical organization of sequential control on motor commands.

1. INTRODUCTION

Speech articulation is a unique motor behavior, since precise and smooth movements are executed without largely depending on visual information. While executions of other body movements rely on visual monitoring with the aid of internal model of the world, speech articulation is executed without using visual function at all. Acoustic and somatosensory feedback signals may be used, but they play roles mainly during speech acquisition process, or they are used after executing movements only for detecting errors between predicted and recieved feedback signals.

Another unique characteristics of speech articulation is that intensions to produce speech sounds are evident. While intentions for usual body movements are vaguely defined, acoustic target (or intended speech signals) formed in the brain can be described by acoustic outputs, because speech sounds are produced as a copy of internal representation of speech sounds. The stream of signals from acoustic targets to acoustic outputs consists of several steps which occur in the brain and the periphery: (1) conversion from acoustic target to articulatory target, (2) conversion from articulatory target to motor commands, (3) conversion from motor commands to articulatory movements, and (4) final realization of speech sounds via articulatory-to-acoustic conversion process. (see the figure [1].)

The main topic of the present study is to seek for a method to examine the motor program of coordination control. In the signal flow described above, speech articulation is also realized as a

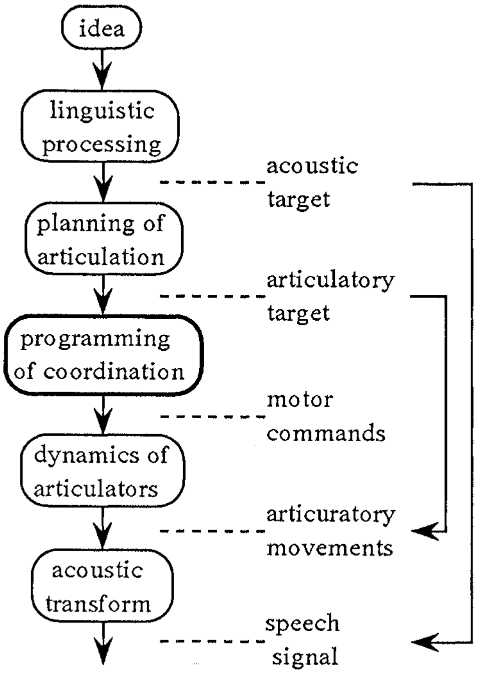

Figure [1] A functional model of speech production process showing component units with the flow signals among them. (The arrows on the right side of the figure indicate the correspondence between internal and external activities.)

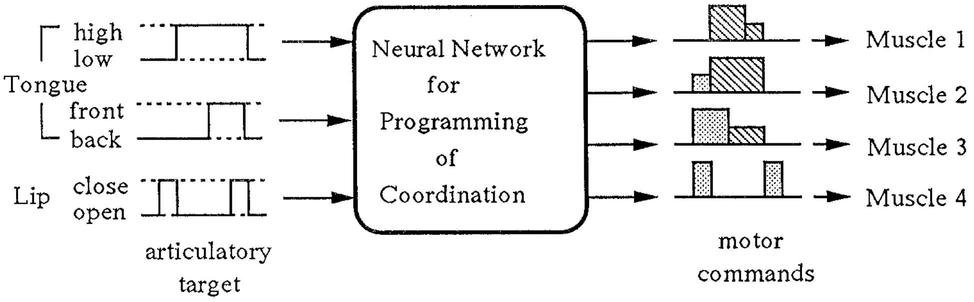

Figure [2] The unit of coordination control and the schematized drawing of the input and output signals.

copy of intended articulatory target, just like speech sounds are produced as a copy of acoustic target. The conversion from articulatory target to motor commands takes place in the neural network between the association motor cortex and the speech motor cortex via the cerebellum (Allen and Tsukahara, 1974). This neural network functions as a subsystem of coordination control, which autonomically generates proper motor commands to speech muscles when the target signals are input (see the figure [2]). We can speculate articulatory target as an input for the subsystem by analyzing observed physiological events. Motor commands, or the outputs from the subsystem, can be examined by analyzing electromyographic signals obtained from speech muscles during speech. Once inputs and outputs of the subsystem are determined, we can identify the internal function of the subsystem, e.g., neural coding from articulatory intention to motor commands.

In this study, integrated EMG signals from each muscle were analyzed to show the orchestration of muscles as a "motor score", which was composed of a sequence of discrete command signals. The analysis method was computational waveform separation technique employed on ensemble average EMG data (Honda, Kusakawa, and Kakita, 1990). Extracted component commands were examined to speculate on temporal and hierarchical organization of muscles.

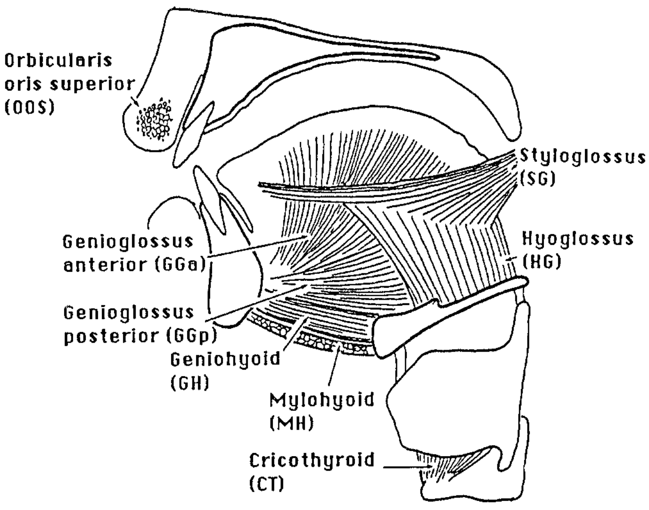

Figure [3] The anatomy of the muscles used in the experiment.

2. EMG Data for the Analysis

The EMG data used in this study were collected at Haskins Laboratories from one male American subject. The original data are already reported by Baer, Alfonso, and Honda (1988). The EMG recording was done simultaneously with audio recording and jaw movement trace during production of isolated /əpVp/ word utterances which contain 11 English vowels. The subject produced each word with listening pre-recorded stimulus sounds and repeated each word 10 times for ensemble averaging analysis. The figure [3] shows the muscles used in the experiment. The EMG data were integrated and digitized at a rate of 200 samples per second. Then ensemble averaging was performed to show temporal changes of muscular activities.

3. Method for Reconstructing the Motor Commands from the EMG Signals

The waveform separation method used in this study decomposes the ensemble average EMG waveform into several component waveforms. The Lorentz curve, which represents a shape of the 2nd order resonance characteristics, is used as a component waveform, since this curve is defined by a simple function having only three parameters of the peak height, the bandwidth and the position. The optimized values of these three parameters were computed by the simplex method (Nelder and Mead, 1963). In this study, this method was applied by minimizing the squared errors between the ensemble average EMG waveform and the waveform synthesized by each of Lorentz curves. In the next step, a square pulse was fitted onto each Lorentz curve in order to schematize a sequence of motor commands in a discrete form. The pulse height and the bandwidth of a square pulse were directly derived from the height and the width of the Lorentz curve. The figure [4] shows an example of the data manipulations applied to the EMG waveform from the orbicularis oris superior (OOS) muscle.

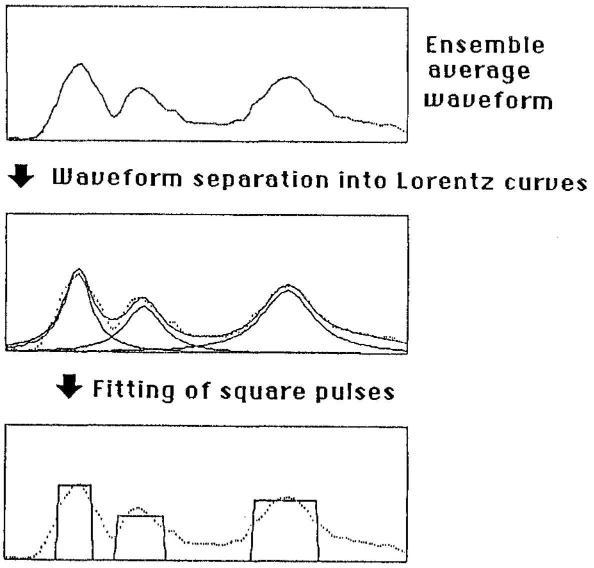

Figure [4] An example of the waveform separation analysis and square pulse fitting.

4. Sequential Control Cycles in /əpVp/ Utterances

The above results indicate that the subsystem of coordination control in the functional model noted before outputs the optimized magnitude and timing of motor commands to speech muscles. Considering articulatory gestures for word utterance as a chain of movements caused by sequential control on motor commands, an attempt is made to speculate sequential control cycles which construct the gestures for a word utterance.

The figure [5] shows the examples of "motor score" which are schematized from the analyzed results. The vertical lines indicate the segmentation for each cycle of sequential control. The cycles for /əpip/ utterance are initiated by the commands to a few muscles to produce the word initial /ə/ vowel (Cycle#1). Next occurs the command to the OOS to produce bilabial closure associated with the commands to a few other muscles, the MH and the GGp, to support the increase of the intraoral pressure for /p/ closure (Cycle#2). The following vowel segments seem to consist of a few cycles where the magnitude of the command is different from each other. Two cycles are fitted in this example for simplicity. The earlier part has large pulses on the GGs for the transition from the preceding consonant to the vowel /i/ (Cycle#3), and the later part shows reduction of pulse height, corresponding to sustained articulation for the vowel (Cycle#4). The entire sequence ends with the commands to the OOS and other muscles (Cycle#5).

As for /əpap/ utterance, the commands for /p/ closure occur mainly in the OOS (Cycle #2). The command in the HG, which was observed in /əpip/, disappears since the tongue position is already low. The following vowel segment are

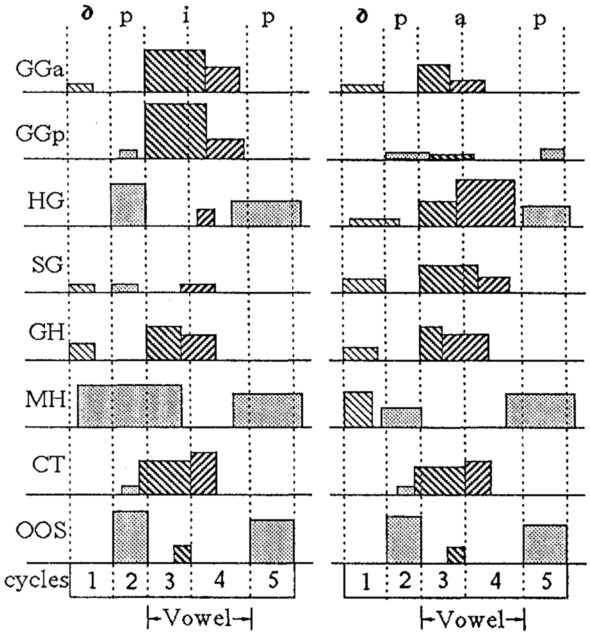

Figure [5] The muscle scores showing schematized results of the analysis for /əpip/ and /əpap/ utterances, segmented by the sequential control cycles.

also divided into two cycles, however, the earlier transitional part has short or weak commands to the HG and SG (Cycle #3), followed by large commands to these muscles (Cycle #4).

The figure [6] is a hypothetical representation of the above results, which shows temporal and hierarchical organization of motor commands. In the figure, each action unit has a hierarchically structured command set to execute a unit movement. These units are sequentially activated when the start signal is input.

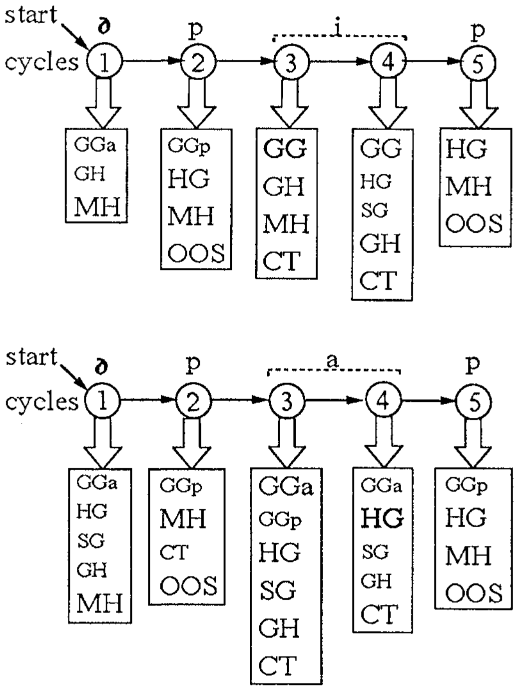

Figure [6] A model of sequential control on motor commands to show temporal organization of action units and hierarchically structured command set.

5. Discussions

As a first step to examine motor program for speech articulation, we made an attempt to describe motor score of speech muscles from EMG data. In the motor scores, we notice that a sequence of command pulses on muscles vary across utterances in terms of the timing and the magnitude. The command pulses on a few muscles for producing the initial /p/ closure appear to have a good synchronization, while, in the following vowel segments, a number of pulses occur in different timing, suggesting that temporal organization of coordination control may not always require a rigid synchronization across muscles. Our model of sequential control to observe motor commands as components of action units, shown in the figure [6], is only a working hypothesis to explain such a coordination control for speech articulation. Further experiments combined with kinematic measurements will provide a comprehensive picture of the coordination control.

Through the present study, we have learned that articulatory intentions for a particular gesture are composed not only of the intentions to change vocal tract shape by tongue and jaw movements, but also of the intentions to produce intraoral pressure rises during lip closure for /p/. Other physiological events, such as contact patterns of the tongue and the palate or glottal tension, also may compose intended signals for articulatory targets. Experimental studies employing various measurements seem to be necessary to describe articulatory intentions in terms of such physiological events.

Acknowledgement

This study is supported in part by the NIH grant (NS-13617) to Haskins Laboratories.

References

Allen, G. I. and Tsukahara, N. (1974). “Cerebrocerebellar communication systems.” *Physiological Reviews*, 54, 957-1006.

Baer, T., Alfonso, P. J., and Honda, K. (1988). “Electromyography of the tongue muscles during vowels in /əpVp/ environment.” *Ann. Bull. R.I.L.P., Univ. Tokyo*, 7, 7-18.

Honda, K., Kusakawa, N., and Kakita, Y. (1990). "An EMG analysis on sequential control cycles of articulatory activity during /əpVp/ utterances." Paper presented at 2nd Seminar on Speech Production: Models and Data, Leeds, May 13-15.

Nelder, J. A. and Mead, R. (1963). “A simplex method for function minimization.” *The Computer Journal*, 7, 308-313.