Ann. Bull. RILP, No. 13, 1-11 (1979)

TEMPORAL PATTERNS OF ARTICULATORY AND PHONATORY CONTROLS*

Masayuki Sawashima

Abstract Several problems on temporal patterns of articulatory and phonatory controls are discussed from the viewpoint of the physiology of voice and speech production. In Japanese word accent where articulatory control is closely related in timing with phonatory control, EMG data as well as acoustic data reveal that the articulatory control consistently precedes the laryngeal control of the vocal pitch. Recent data on the velocity of movements of various parts of the articulatory organs in syllable repetition are presented. Flexibility of the articulatory target should be taken into account for a better understanding of effects of coarticulation and speaking rate on the dynamic control of articulatory movements. The central mechanism of skilled movements should be referred to as the physiological base of adaptive control of articulatory movements.

1. Introduction

There have been a wide variety of discussion on topics concerning the problem of temporal organization and prosody in speech production. In this paper, the author does not intend to cover all the relevant topics but will discuss some of them from the viewpoint of the physiology of voice and speech production.

2. Relative Timing of F_0 Control and Articulatory Control and Related Problems.

2.1 F0 Control by the Larynx

The contour of vocal pitch (F_0) in speech is an indispensable element of the prosodic features. There have been numerous works on the mechanism of F_0 control during speech. It is generally agreed that the cricothyroid is the only muscle which is uniquely related to F_0 control. The contraction and relaxation of this muscle cause the F_0 to rise and fall respectively, especially in the higher pitch range. As is well known, this muscle actively elongates the vocal folds, which results in an increase in the longitudinal tension and a decrease in the effective mass. The vocalis muscle also, in combination with the cricothyroid, contributes to changes in the longitudinal tension. Up and down movements of the larynx correlated with F_0 change have also been observed by many researchers. But the question of how these movements can affect condition of the vocal fold and the resulting vocal pitch is still to be answered. An old and new

---

*This is a revised version of a paper presented at the US-JAPAN Joint Meeting on Acoustics, November 1978. The work was partly supported by a Grant-in-Aid for Scientific Research, Ministry of Education, No. 310707 and No. 410207.

problem for us is that of the F_0 lowering mechanism. It is well known that in some languages, such as tone languages and pitch accent languages such as Japanese, some of the linguistic information is manifested by a rapid fall in F_0. Experimental observations have indicated the activity of some of the extrinsic laryngeal muscles correlated to F_0 lowering 1) 2) 3) 4) or low F_0 5). However, these results are not as consistent as those obtained for cricothyroid activity in relation to F_0 changes. Furthermore, the timing of the activity of the extrinsic laryngeal muscles relative to the fall in F_0 suggests that these muscles may not be causal mechanisms, but rather supplementary or maintaining mechanisms for F_0 lowering following the relaxation of the cricothyroid. In this regard, it is interesting to note that in Japanese accent patterns, and probably in some tonal patterns in other languages, a rise in F_0 or a high F_0 state precedes the rapid F_0 fall. A further problem regarding the contribution of these muscles is how their activity can affect the vocal fold condition. This is actually the same question as for the up and down movements of the larynx. Actually these up and down movements are closely related to the activities of the extrinsic muscles. Some researchers claim that the up and down movements, as well as the activities of the extrinsic laryngeal muscles, may be transferred to the longitudinal tension. 6) 7) Another new claim is that the up and down movements may alter the "vertical tension" of the vocal folds. 8) 9) However, there appears to be no anatomical base where the up and down movements may directly result in a vertical compression or tensening of the vocal folds.

2-2 Relative Timing of F_0 Control and Articulatory Control

In Japanese, the accent pattern as represented by the F_0 contour is closely linked with a time segment called the "mora". Thus, laryngeal F_0 control should be expected to have some appropriate timing relative to supraglottic articulatory control. Fujisaki 10) 11) has made a quantitative analysis of this problem using his functional model of articulatory and phonatory control based on the acoustic analysis of speech waves. We are now studying the same problem using a physiological parameter, the EMG's of the speech muscles in combination with acoustic analysis. The study is now underway. Here I will present some preliminary data from our study and point out several questions arising from this kind of experiment.

Figure 1 shows the acoustic data of speech samples of two 2-mora words with a rising accent pattern, /ai⌝/ and /ia⌝/. The F_0 contour and formant shift data obtained by Fujisaki's method are displayed on the same time axis for each word. The vertical line T_p is the estimated onset time of pitch transition and T_f that for formant transition. From the acoustic data, it is apparent that the onset of formant transition, whcih corresponds to the onset of the articulatory gesture, takes place earlier than that for the F_0 transition which corresponds to the phonatory control. This result is consistent with Fujisaki's previous data.

Figure 2 shows integrated and smoothed EMG curves for each of the tokens used for the acoustic analysis. Averaging repeated utterance samples is not useful for this kind of timing analysis because of time smearing. CT indicate the cricothyroid muscle time curve for F_0 control and GG the genioglossus muscle time curve for tongue movement, which

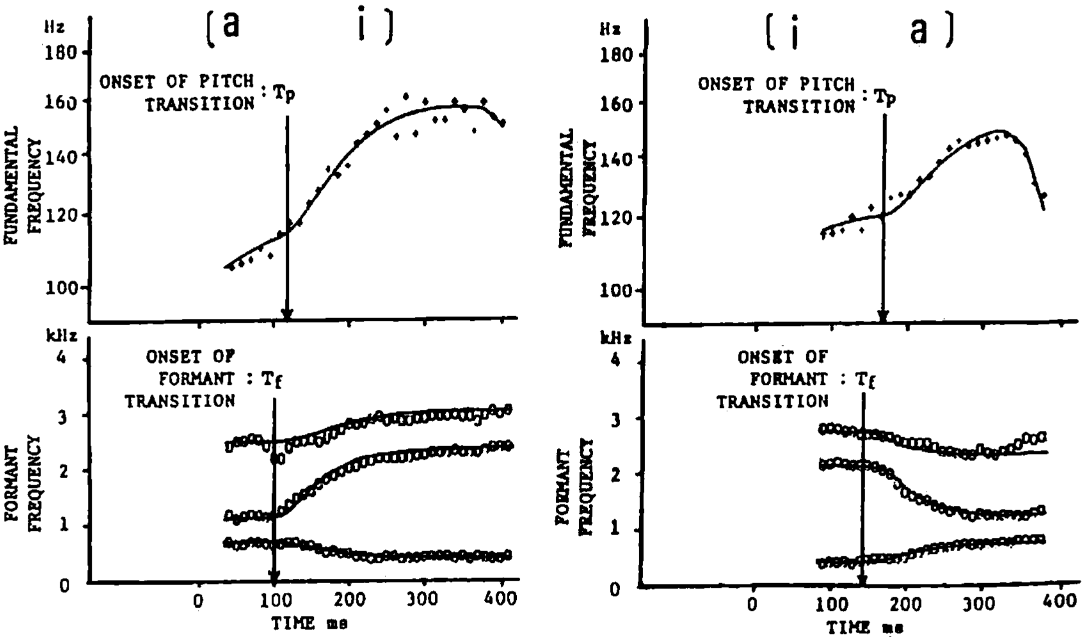

Fig. 1: Comparison of F_0 and formant transitions for /ai˥/ and /ia˥/.

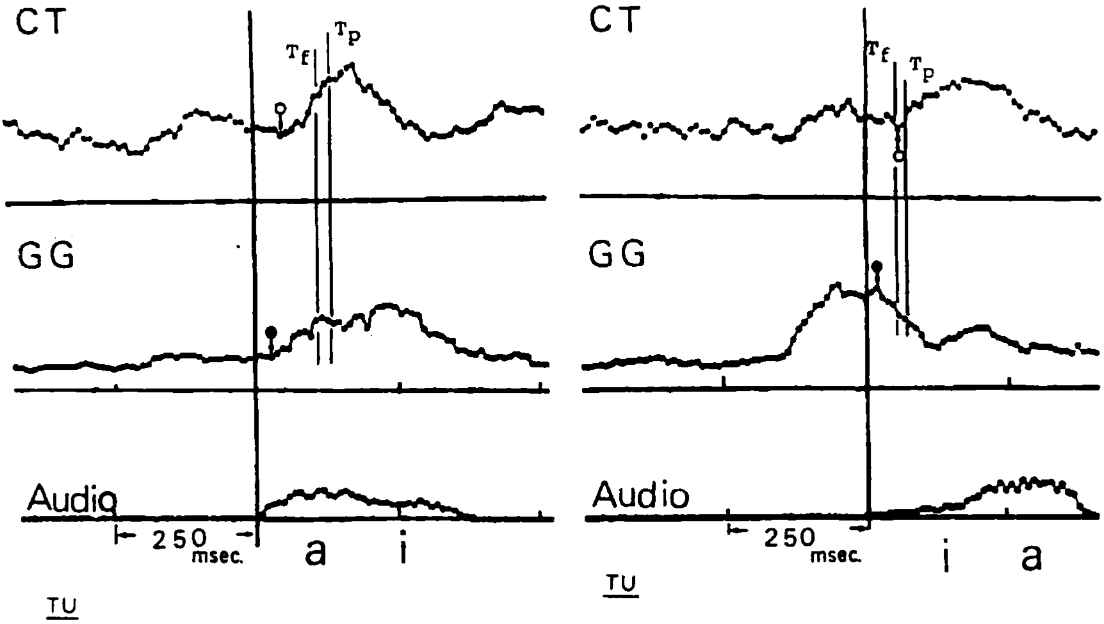

Fig. 2: Comparison of EMG time curves of cricothyroid (CT) and genioglossus (GG) muscles for the same samples as Fig. 1.

corresponds to the formant shift. The thick vertical bar indicates the onset of the vowel of the first mora. Time points of T_p and T_f obtained by acoustic analysis are also indicated by thin vertical bars on the EMG curves. The open circles on the CT curves indicate the onset of the increase in cricothyroid activity for the pitch rise. The solid circle on the GG curve of /ai / indicates the onset of the increase in GG activity corresponding to the tongue fronting from /a/ to /i/. The solid circle on the GG curve for /ia / indicates the onset of the decrease in GG activity, which presumably corresponds to the tongue backing from /i/ to /a/. These onset time points were placed by visual observation. For these samples, the relative timing between phonatory and articulatory control observed from the EMG data is in good agreement with that observed from acoustic data.

Two problems remain to be solved for this physiological experiment. The first concerns the definition of the onset of the change in the EMG curve. We should develop a more objective method than simple visual observation for locating the onset time, as Fujisaki has for his acoustic data. Second, we took the decrease in GG activity as an indication of tongue backing. It would be better to select some other muscle which is activated for tongue backing. This question is related to the problem of the "synchronism" of the different muscle activities for a given basic articulatory gesture. The problem may, in some part, be attribued to the dynamic properties of the different parts of the speech organs.

In the electromyographic study of American English, Smith 12) has observed that the onset of activity of the mylohyoid muscle was approximately 20 msec earlier than that of the genioglossus for a given consonant articulation. Smith concludes from this data that the motor commands to different muscles which participate in a specific speech gesture can be issued nonsynchronously by the cortical control areas, and that this lack of synchrony can have functional significance in speech production.

3. Speed of Articulatory Movements in Syllable Repetition

In studying the temporal organization of articulatory movements, various dynamic characteristics such as the velocity of the movement of different parts of the speech organs should be taken into account. Formerly, data relevant to this question were obtainable by measuring the maximum frequency in repetition of a given syllable. According to Hudging and Stetson 13), the maximum rate of syllable repetition (mean value per second) is: 6.7 for the lip in /pu/, 8.2 for the tip of the tongue in /tat/, 7.1 for the back of the tongue in /ka/, and 6.7 for the velum in /tun/. Recently, with the development of the x-ray microbeam system, 14) we have made detailed analyses of the articulatory movements in syllable repetition by tracking the movement of metal pellets placed on various parts of the articulatory organs. 15) 16) Figure 3 shows the displacement and the velocity of lip movement in repetitions of /pa/ obtained by x-ray microbeam tracking. It should be noted that in the normal subject, the repetition of the movement is carried out quite regularly in terms of the amplitude, velocity, interval and also the direction of the movement. In the normal subject it is also noted that attempts to make the syllable repetition at a slower rate does not result in a decrease in the velocity but in an increase in the closure period of the given consonant as compared to

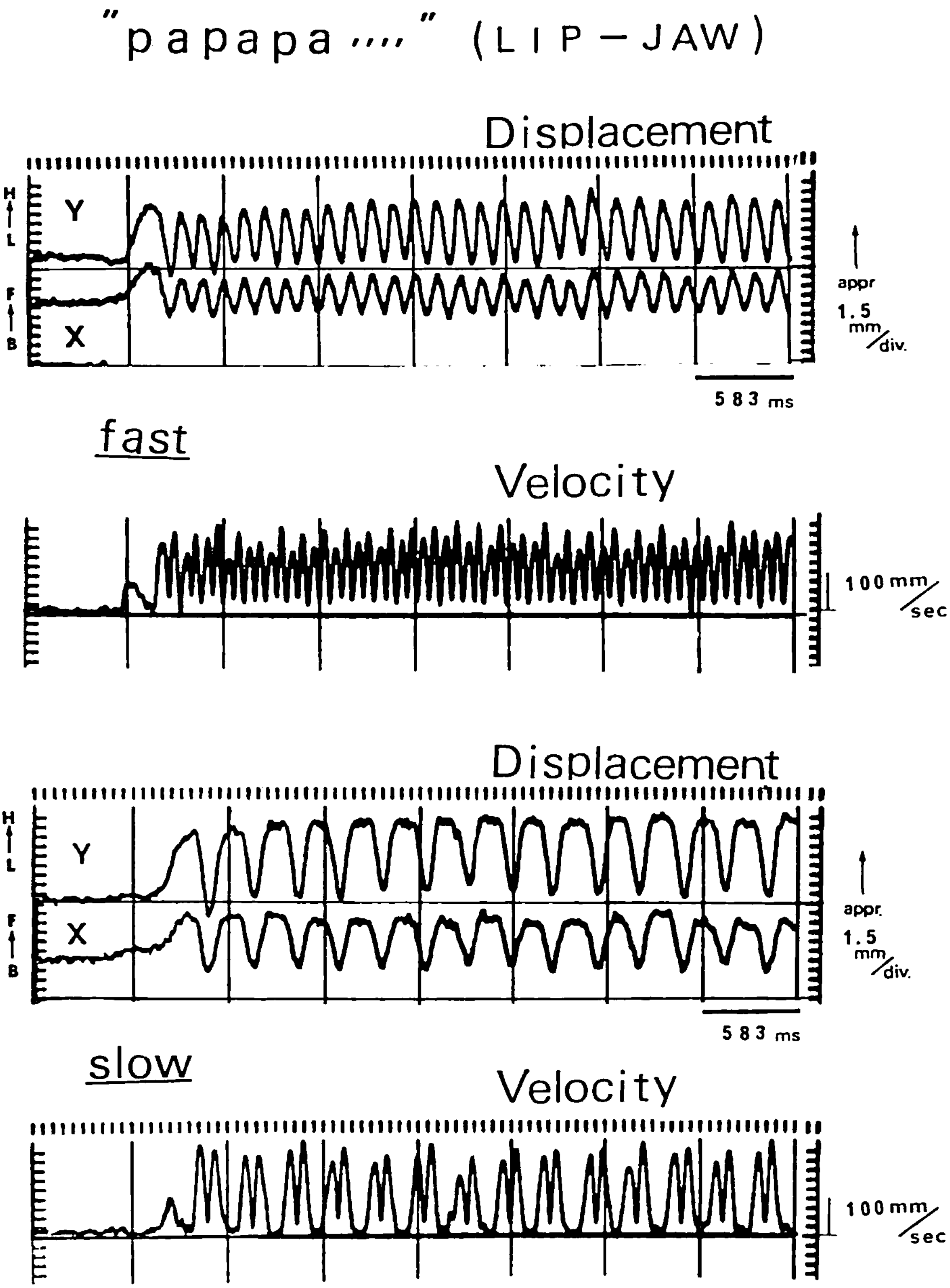

Fig. 3: Displacement and velocity of lip movement for repetitions of /pa/.

a faster rate of repetition. EMG's of the pertinent articulatory muscles show a fairly rhythmic pattern of activation-suppression corresponding to the movement with a clear reciprocal activity pattern between the antagonistic muscle pairs.

The maximum velocity (mm/sec) in syllable repetition is: 190-250 for the lip in /pa/, 220 for the tongue tip in the /t/ of /pataka/, 200-220 for the tongue back in /ka/ and 105 for the velum in /teN/. It is noted that the velocity of velar movement is definitely slower than that of the others.

4. Dynamic Control of Articulatory Movements.

4-1 Adaptive Motor Control and the Flexibility of Articulatory Targets

It is well known that in connected speech articulations, motor control in terms of the peripheral neural input to the muscles for a given articulatory gesture or target varies with different phonetic environments. An example of this context dependent control manifested in velar movement 17) is shown in Fig. 4. The left part of the figure displays the time course for velar height obtained by fiberscopic observation, for three types of utterances. Velar lowering takes place for the nasal sound /N/. However, the velar height for the /d/ of the frame sentence is quite the same irrespective of the presence or absence of the nasal sound, or its position. In the right part of the figure, EMG patterns of the levator muscle are displayed for three types of utterances. Here, there are great differences in the muscle activities for the /d/ of the frame sentence depending on the presence or absence and the position of the nasal sound.

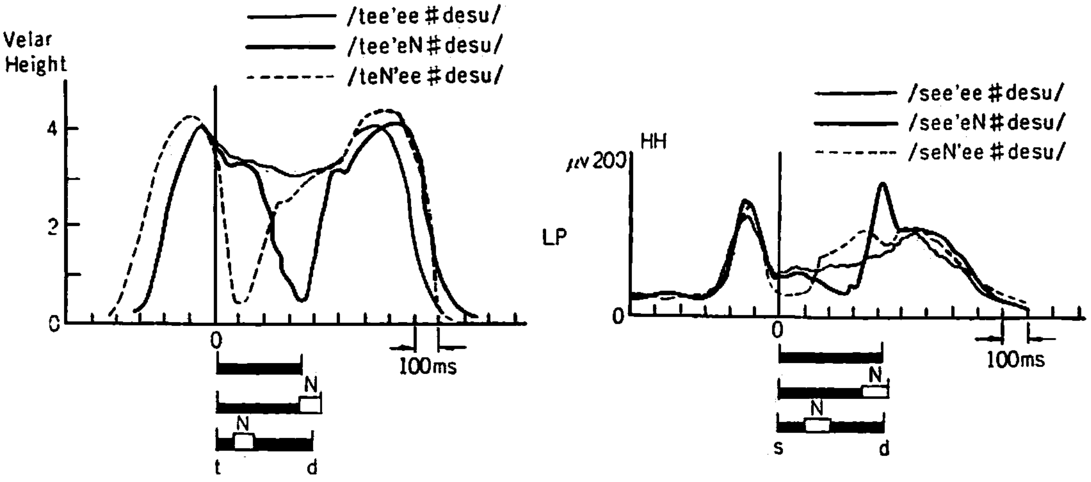

Fig. 4: Time curves of velar height and EMG of the levator muscle for utterances of words containing nasal and non-nasal sounds.

Stress and speaking rate also affect the dynamic control of articulatory movements. The physiological nature of stress, whether lexical or emphatic, may be characterized in general by an increase in the peripheral neural input to the muscles for the articulatory manifestation of a given phonetic segment. Thus, an increase in articulatory constraints, segment duration, acoustic power, the fundamental frequency of the voice, etc. results.

The effect of speaking rate is somewhat more complex. Experimental results reveal that an increase in speaking rate results in an increase in articulatory effort as reflected in EMG levels for some consonants, for example, lip gestures for /w/ and /p/,18) and velar elevation for obstruents.19) On the other hand, there is a decrease in the activity of the tongue muscle for vowel articulation.18) This difference between consonants and vowels may be attributed to a difference in the flexibility of the articulatory target. Vowels are assumed to be more flexible than consonants. The amount of the flexibility may also depend on tense-lax oppositions, if they exist in the phonetic segment. Environmental conditions, such as stress may also affect flexibility. Actualization of coarticulatory effects may also be reflected in the flexibility of the articulatory target. Fig. 5 shows the trajectory of the tongue pellets for /k/ in the utterance of /makae/ and /makie/. In /makie/, as compared with /makae/, the tongue gesture for /k/ closure shifts forward due to the coarticulatory effect of the following front vowel /i/. The articulatory target for /k/ is much more flexible along the front-back axis than along the high-low axis of the tongue position. The flexibility for /t/ appears to be far less than that for /k/, especially for the tongue tip.20)

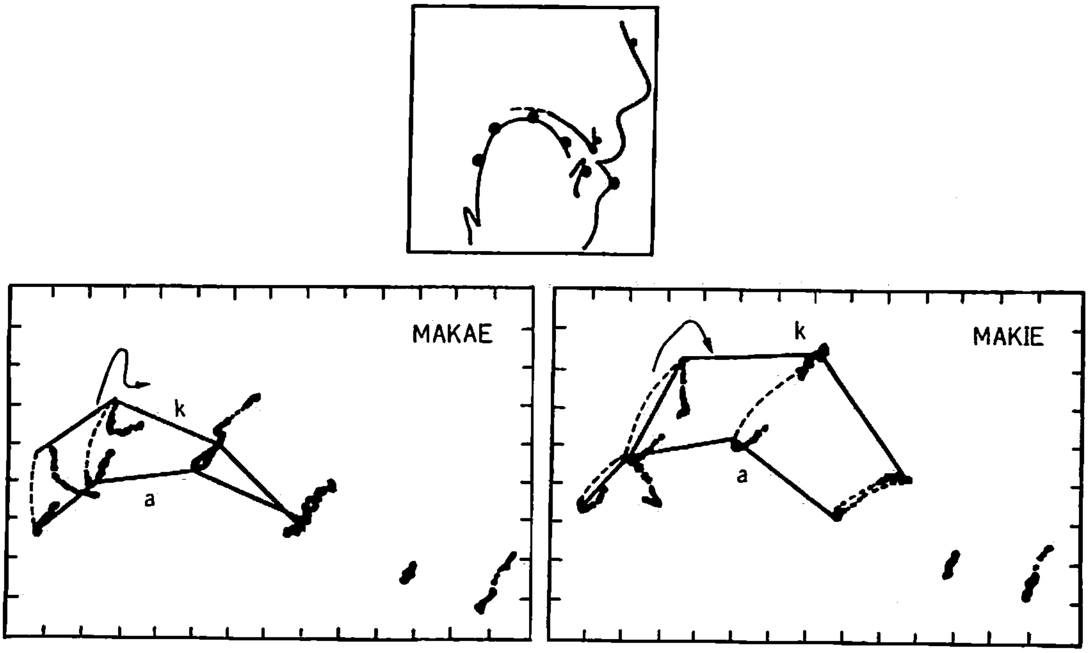

Fig. 5 : Trajectory of the tongue pellets for /k/ in /makae/ and /makie/.

Some flexibility still remains for the tongue root, manifesting the same type of coarticulation as with /k/, although the extent of the shift is small.

For velar coarticulation, the high position for obstruents should allow very little flexibility, and the low position for the nasals may show a considerable degree of flexibility. Velar height for vowels should not be unspecified. Each vowel should be associated with its own target value and should also be characterized as having a great deal of flexibility. It follows that this kind of multi-valued system should be placed at some level in the characterization of the process of speech production.

Kent and Minifie 21) proposed a hierarchical model of speech production. The model consists of the following steps: 1) the rhythmic grouping of syllables (multi-syllable unit or phrase), 2) the syllable forms (individual syllable shapes and associated intrasyllabic rules) 3) the phonemes or segments (basic phonetic constitutents of syllables) 4) the set of phonetic features (goals or targets defined for each phoneme or segment), 4) the pattern of articulatory transitions (articulatory realizations of phonetic features), and 6) the sequence of neuro-motor commands (neural instructions to the muscles, derived from transitional requirements). If we follow this model, the multi-valued system mentioned above may be incorporated at the level of the "set of phonetic features". Also, dynamic adaptive motor control may be organized at the level of the "pattern of articulatory transitions."

4-2 Central Mechanism of Skilled Movements.

The central mechanism of dynamic adaptive control of articulatory movements, inclusive of the problem of "open loop" or "closed loop" control, has been discussed by several authors. 22) 23) It should be kept in mind here that the articulatory movements of the native speaker in his own language, although speech specific, are learned, skilled voluntary movements. In this sense, it would be useful to consider the central mechanisms of other skilled movements as was suggested by Allen and Tsukahara 24). They discussed the participation of the cerebellum in the planning and carrying out of the voluntary movement of limbs in their extensive review of cerebro-cerebellar communication systems. According to these authors, the lateral cerebellum (cerebellar hemisphere) may be considered to participate in the pre-programming or long-range planning of voluntary limb movement. The intermediate cerebellum (pars intermedia), on the other hand, works as a feedback system to the motor cortex in execution of the movement. They state:

"Once the movement has been planned within the association cortex, with the help of the cerebellar hemisphere and basal ganglia, the motor cortex issues the command for movement. At this point the pars intermedia makes an important contribution by updating the movement based on the sensory description of the limb position and velocity on which the intended movement is to be superimposed. This is a kind of short-range planning as opposed to the long-range planning of the association cortex and lateral cerebellum.......

In learning a movement, we first execute the movement very slowly because it cannot be adequately preprogrammed. Instead, it is performed largely by cerebral-intervention as well as the constant updating of the intermediate cerebellum. With practice, a greater amount of the movement can be preprogrammed and the movement can be executed more rapidly(without reference to peripheral sensory input)".

Thus, for learned movements the cerebellum provides an internal substitute for the external world. The lateral cerebellum is considered to perform this operation.

The cerebellum participates in two types of motor control: closed loop and open loop control. In closed loop control the pars intermedia integrates information from the motor cortex and peripheral sensory inputs for updating the ongoing movements. In open loop control the lateral cerebellum participates in organizing preprogrammed movements which eliminate the need for peripheral neural input and allow one to increase the speed of execution. The transition from closed loop control to open loop control takes place by learning and practicing skilled movements.

Although Allen and Tsukahara discuss only the control of limb movement, their primary observations may be applied to speech production. In studying speech dynamics, we should consider the neurophysiological basis for the development and organization of skilled movements as well as exploring the various speech specific problems.

5. Summary

In this review, several topics related to the temporal patterns of articulatory and phonatory controls were discussed from the viewpoint of the physiology of voice and speech production. An electromyographic study on the relative timing between vocal pitch (F_0) control at the larynx and the articulatory tongue gestures in Japanese word accent was presented. The problems to be solved in this kind of experiment are the definition of the onset of the muscle activity and the possible non-synchronism of the different muscle activities for a given articulatory movement. The latter problem may, in some part, be attributed to the dynamic properties of the different parts of the speech organs.

The analysis of articulatory movements in syllable repetition is considered to be a promising approach in studying some basic dynamic characteristics of the articulatory organs. Recent data obtained on the velocity of the movements of various parts of the articulatory organs were presented.

Dynamic patterns of articulatory movements are characterized by a fluctuation in the actualization of the articulatory target according to various conditions such as coarticulation and speaking rate. This fluctuation is considered to be attributed to the flexibility of the articulatory target for a given segment. The amount of flexibility may vary with different segments, and also according to environmental conditions. Consequnetly a multi-valued system reflecting these properties should be incorporated at some level in the characterization of the process of speech production.

Dynamic patterns of articulatory movements are also characterized by the context dependent adaptive control of muscle activities. The account of the central mechanism of the skilled movements of limbs as proposed by Allen and Tsukahara is considered to be a good basis for a better understanding of articulatory motor control including open loop vs. closed loop problems.

References

1. Shimada, Z. and H. Hirose, (1971). "Physiological correlates of Japanese accent patterns, " Ann. Bull. RILP, 5, 41-49.
2. Sawashima, M., Y. Kakita and S. Hiki (1973). "Activity of the extrinsic laryngeal muscles in relation to Japanese word accent, " Ann. Bull. RILP, 7, 19-25.
3. Erickson, D. M. (1976). "A physiological analysis of the tones of Thai", Univ. Connecticut Dissertation.
4. Sugito, M. and H. Hirose (1978). "An electromyographic study of the Kinki accent. " Ann. Bull. RILP, 12, 35-51.
5. Atkinson, J. E. (1978). "Correlation analysis of the physiological factors controlling fundamental voice frequency, " J. Acoust. Soc. Am. 63, 211-222.
6. Sonninen, A. (1956). "The role of the extrinsic laryngeal muscles in length adjustment of the vocal cords in singing, " Acta Otolaryng. Suppl. 130.
7. Zenker, W. and A. Zenker (1960). "Über die Regelung der Stimmlippenspannung durch von aussen eingreifende Mechanism, " Folia phoniat. 12, 1-36.
8. Ohala, J.H. (1972). "How is pitch lowered?" J. Acoust. Soc. Am. 52, 124.
9. Stevens, K.S. (1977). "Physics of laryngeal behavior and larynx modes, " Phonetica 34, 264-279.
10. Fujisaki, H. and H. Sudo, (1971). "A model for the generation of fundamental frequency contours of Japanese word accent, " J. Acoust. Soc. Japan 27, 445-453.
11. Fujisaki, H. (1977). "Functional models of articulatory and phonatory dynamics. Temporal organization of articulatory and phonatory controls in realization of Japanese word accent, " in Dynamic Aspects of Speech Production, edited by M. Sawashima and E S. Cooper, 346-366, Univ. Tokyo Press.
12. Smith, T.S. (1972). "Some temporal aspects of speech motor commands, " Language Center Occasional Papers, 13 , Univ. Essex. 151-162.
13. Hudging, C. V, and R.H. Stetson (1937). cited from Lehiste (1970) Suprasegmentals, MIT Press.
14. Kiritani, S., K. Itoh and O. Fujimura (1975). "Tongue-pellet tracking by a computer-controlled x-ray microbeam system, " J. Acoust. Soc. Am., 57, 1516-1520.

15. Hirose, H., S. Kiritani, T. Ushijima and M. Sawashima (1978). "Analysis of abnormal articulatory dynamics in two dysarthric patients, J. Speech and Hearing Disorders, 43, 96-105.

16. Hirose, H, S. Kiritani, T. Ushijima, H. Yoshioka and M. Sawashima (1978), "Patterns of dysarthric movements in patients with Parkinsonism, " Ann. Bull. RILP, 12, 73-85.

17. Ushijima, T. and H. Hirose (1975). "Velar movement and its motor command, " Haskins Laboratories Status Report on Speech Research, SR-41, 207-216.

18. Gay, T., T. Ushijima, H. Hirose, and F.S. Cooper (1974). "Effect of speaking rate on labial consonant-vowel articulation, " J. Phonetics, 2, 47-63.

19. Ushijima, T., H. Hirose, and M. Sawashima (1977). "Effects of accent and speaking rate on the control of velar articulation, " Ann. Bull, RILP, 11, 23-30.

20. Kiritani, S. (1978). "Perturbations of the consonant and vowel articulation by the adjacent segments, " J. Acoust. Soc. Japan, 34, 132-139.

21. Kent, R. D and F. D. Minifie, (1977). "Coarticulation in recent speech production models, " J. Phonetics, 5, 115-133.

22. Mac Neilage, P (1970). "Motor control of serial ordering of speech. Psych. Rev., 77 , 182-196.

23. Moll, K. L., G.N. Zimmermann, and A. Smith (1977). "The study of speech production as a human neuromotor system, " in Dynamic Aspects of Speech Production, edited by M. Sawashima and F.S. Cooper, 107-127, Univ. Tokyo Press.

24. Allen, G. I. and N. Tsukahara (1974). "Cerebrocerebellar communication systems, " Physiol. Rev. 54, 957-1006.