Modeling Vocal Tract Organs Based on MRI and EMG Observations and Its Implication on Brain Function.

Kiyoshi Honda*, Hiroyuki Hirai* and Naoki Kusakawa**

*ATR Human Information Processing Research Laboratories
**ATR Institute International
2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-02, Japan

1. Introduction

This paper describes our recent studies on the structure and the function of speech organs with some implications on brain organization in speech production. Our experimental studies have been carried out to derive a vocal tract model from morphological and physiological data. Recently emerged technology of magnetic resonance imaging (MRI) of human body provides new and abundant information of the structure of vocal tract organs. Electromyographic (EMG) data help us understand the motor organization of these organs. Such morphological and myofunctional information also gives some insights into brain function in speech production and perception. The conceptual framework for inferring the organization of speech is based on the process of the interaction between the brain and the peripheral organs in human development. The brain perceives peripheral organs in the same way as perceiving the outside world. In other words, the central nervous system recognizes the characteristics of the structure and the function of the peripheral speech organs. Then, the system acquires new strategies or modifies old mechanisms to perform purposeful actions for producing speech sounds.

In this paper, the morphological and functional aspects of F0 and vowel production are discussed based on our recent findings from observations and analyses of MRI and EMG data. Midsagittal MRI data of the larynx have revealed overall organization of F0 control. The analysis of tongue EMG data have suggested an image of internal representation of vowel system which is equivalent in perception and production. A hypothesis is proposed to account for the origin of these speech phenomena from a view point of the relationship between the brain and the periphery.

2. Laryngeal structure and its implication to F0 control

This section describes our recent findings from MRI study on the larynx and the interpretation on sensory-motor organization of F0 control. Laryngeal control of F0 has self-evident mechanism of rotation of the cricoid and the thyroid cartilages (Katsuki, 1950; Sonninen, 1954). The cricothyroid muscle is attached to the relevant cartilages. Due to anatomical location of this small muscle, F0 control mechanism has been attributed to the cricothyroid function to raise F0. On the contrary, it is not very clear whether F0 lowering is caused by relaxation of F0 raising muscle(s) or by specialized F0 lowering muscles. It has been known that the strap muscles show F0 lowering activity (Atkinson, 1973; Erickson, 1976; Sugito, 1982; Shimada, 1991). It is our empirical knowledge that larynx lowering tends to be associated with F0 lowering. However, no explanation has successfully led us to understand the chain of mechanisms to cause F0 lowering. Our MRI study on the larynx and the surrounding structure (Hirai, Honda, Fujimoto, & Shimada, 1993) gives us a better understanding about overall mechanisms and organization of F0 control.

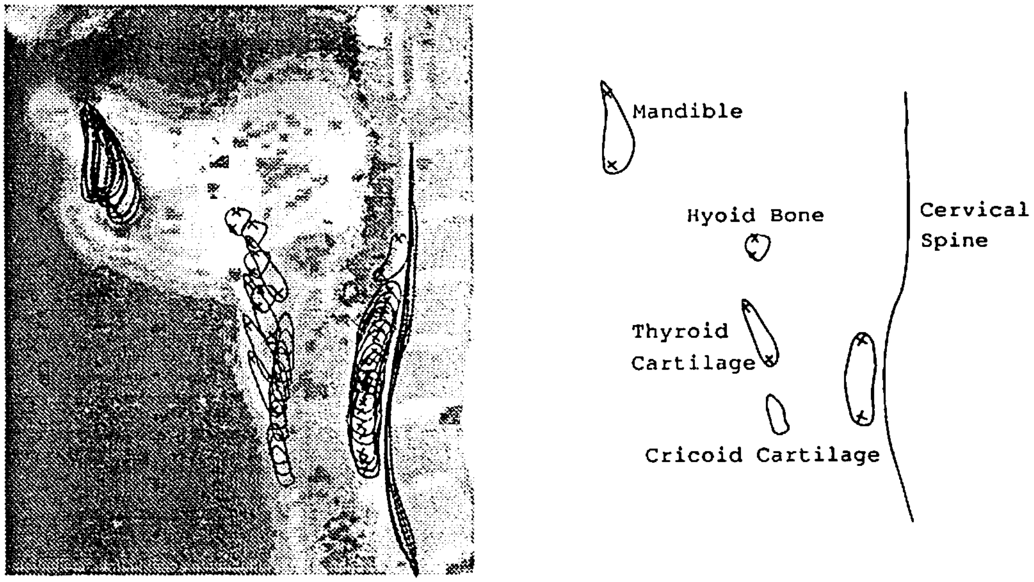
2.1 MRI observation of perilaryngeal morphology at different F0 levels

Midsagittal magnetic resonance images of perilaryngeal structure were taken from three male subjects during phonation at different F0 levels. The devise is 1.0 Tesla clinical purpose imager (Shimadzu SMT-100). Table 1 shows the imaging method and parameters used for our study. The field-echo method is one of fast-scan methods, and a single slice image can be taken in four seconds. This method also has an advantage for the purpose of imaging the laryngeal cartilage, because brighter images are obtained for the cartilaginous tissue. The subjects produced sustained vowel /a/ in the range of 1.0 - 1.5 octaves in a down scale to subjects' lowest possible F0.

Table 1. Imaging Parameters for the MRI experiment

Devise: SMT-100GUX (Shimadzu)
Magnetic Field: 1 Tesla
Slice: 10 mm at the midsagittal plane
Methods: Field Echo
    TE: 12 msec
    TR: 30 msec
Time: 4 sec / frame

Fig. 1. shows MRI pictures and the traces of laryngeal and perilaryngeal structures. The midsagittal images of the mandible, the hyoid bone, the thyroid cartilage, the cricoid cartilage, and anterior contour of the cervical spine are traced at all of the F0 levels and superimposed on an MRI picture for the lowest F0. Each trace of rigid structures has two landmarks (measurement points) for movement analysis. In each picture, vertical movement of the larynx is observed as a major factor of positional changes of these structures. Movements of the jaw and the hyoid bone are also observed in the data. These pictures also demonstrate that these movements occur in order to allow the larynx to move along the anterior wall of the cervical spine. The finding which is not expected from our previous knowledge is that vertical movement of the larynx can induce a rotation of the cricoid cartilage via the curvature of the cervical spine. This finding is evident for the data from the subjects A and C. The cervical spine shows a curvature, called "lordosis" in an anatomical term, and the posterior plate of the cricoid cartilage appears to tilt along this curvature, as it moves vertically. This tilt of the cricoid cartilage, associated with F0 lowering, causes a rotation of the cricothyroid joint toward the direction of shortening the vocal fold. When the curvature of the cervical spine is not available, as shown on the data from the subject B, the cervical spine actively changes its tilt angle so that the cricoid cartilage can rotate.



(a) Subject A

(b) Subject B

(c) Subject C

Fig. 1 MRI pictures of the three male subjects and trace of perilaryngeal structure.

Measurements of positional changes of these structures were performed to confirm the above observations on the images. Fig. 2. shows vertical movements of the larynx at different F0 levels. The vertical positions of the cricoid cartilage relative to the position of the second cervical vertebra (C2) were plotted on the figure. The tendency of larynx lowering with F0 lowering is observed in all three subjects. The average rate of larynx lowering is roughly 20 mm per octave. Fig. 3 shows relative angles of the posterior plate of the cricoid and the thyroid cartilages and the tilt of the contour of the cervical spine at the position of the cricoid cartilage. These measures are standardized at the values for the highest F0. The data show that the measures of the cricoid cartilage and the cervical spine change in parallel, supporting the previous visual observation on the mechanism of cricoid rotation due to the lordosis of the cervical spine.

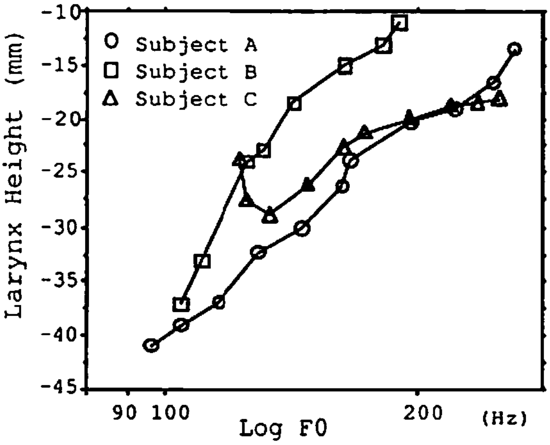
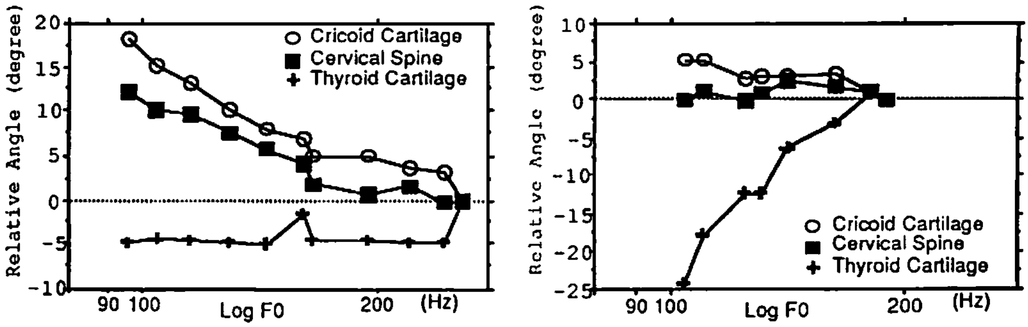


Fig. 2 Vertical movement of the larynx in different F0 levels.

The involvement of cervical structure in F0 control is a strategy which is unique to the structure of human body. The lordosis of the cervical spine at the level of the larynx composes a part of the curvature of the entire spinal structure. Evolutionary morphology explains that the spinal curvature, lordosis and kyphosis in cycle, is the result of morphological development towards human bipedalism (Hayama, 1991). The descent of the larynx which occurred associated with the enlargement of the skull in the phylogeny of the vertebrates also contributes to enhancing the extent of F0 control by the mechanism described above. It is interesting to find that F0 change, an element of human speech production, have been developed with particular evolutionary alternation of the vertebrate body which is called "hominization".

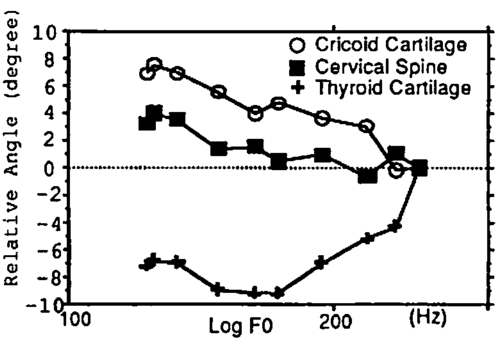
2.2 F0 related events and their implication on neural process

It seems that we have a spectrum of knowledge on peripheral events associated with F0 changes. Fig. 4 illustrates peripheral mechanisms of F0 control which have been proposed in the



(a) SUBJECT A

(b) SUBJECT B



(c) SUBJECT C

Fig. 3 Rotation of the cricoid and the thyroid cartilages in different F0 levels.

literature. This figure also explains that the brain possesses the knowledge of these available mechanisms. The integration of F0-related laryngeal events in the following discussion is useful to understand some aspect of neural process of speech production. Motor commands, which are generated in the motor cortex, cause various events in the periphery of the body to execute intended gesture for F0 control. These laryngeal events associated with F0 changes are the stretch of the vocal folds by cricothyroid rotation, vertical movement of the larynx by the strap and other muscles, the change of pulmonary air pressure by the respiratory organs, and so on. These events are monitored by the brain via sensory organs to form sensory representation of F0 changes. Fig. 5 shows a box-and-arrow representation of the integration of such neural information. In the figure, motor representation of F0 control is shown as a rise-and-fall pattern. The motor pattern causes various events in peripheral structures, and sensory organs in the corresponding parts send afferent information to the brain according to each event. The figure illustrates that these sensory inputs resemble with each other. In other words, phenomenological changes of the events are similar in representative forms. The sensory cortex receives these sensory inputs, and detects the meaning of the events by integrating the information. Since those patterns are correlated with each other, the manner of integrating F0 related sensory information can be a simple neural computation of correlation in the association cortex.

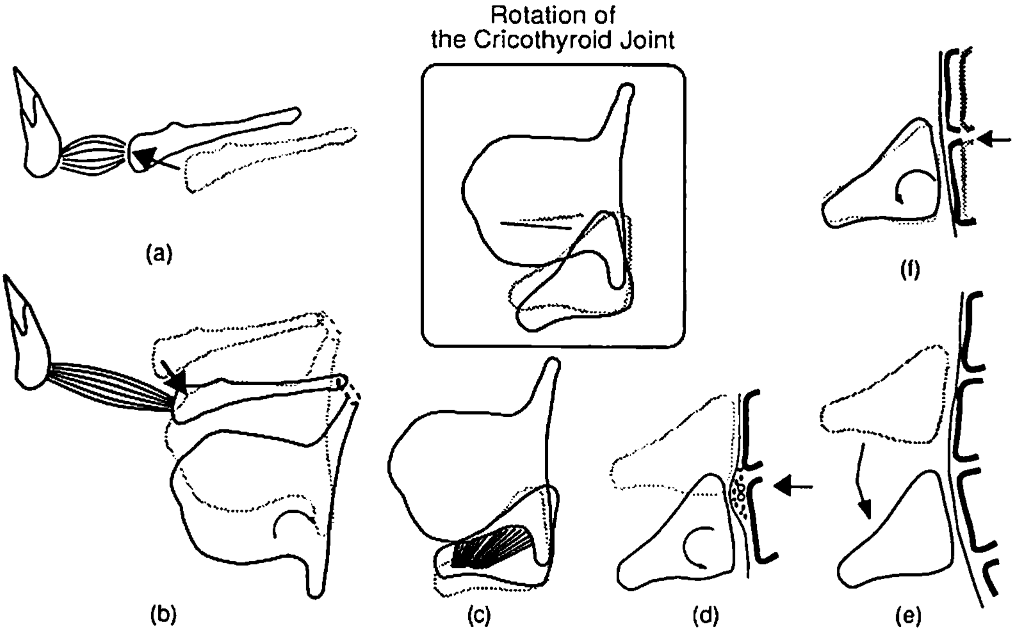


Fig. 4 Peripheral mechanisms of F0 control. (a) hyoid bone movement for F0 raising, (b) hyoid bone movement for F0 lowering, (c) the action of the cricothyroid muscle for F0 raising, (d) the action of the cricopharyngeus muscle for F0 lowering (Honda & Fujimura, 1991), (e) the rotation of the cricoid cartilage for F0 raising and lowering due to larynx lowering along the curvature of the cervical spine, (f) active tilting of the cervical spine for F0 raising.

The concept of informational integration by calculation of correlation is further demonstrated to account for the relationship between sensory and motor association areas in the brain, as shown in Fig. 6. The internal representation of the motor pattern for F0 control is the tension of the vocal folds. The internal representation of pitch corresponds to the fundamental frequency (F0) of voiced sounds. The association cortex can process two kinds of information to be identical when they are correlated. Once such a transparent relationship is established, motor pattern may automatically induce sensory pattern, and vice versa. The motor pattern formed in the anterior part of the brain is fed back to the posterior part of the brain to elucidate the sensory pattern which is similar to the original information. This concept of association of motor and sensory information accounts for neuronal background of close relationship between production and perception in speech. It can be noted that such sensory-motor relationship has been established by a resonance of morphological diversion of human speech organs and the development of cognitive function of the brain.

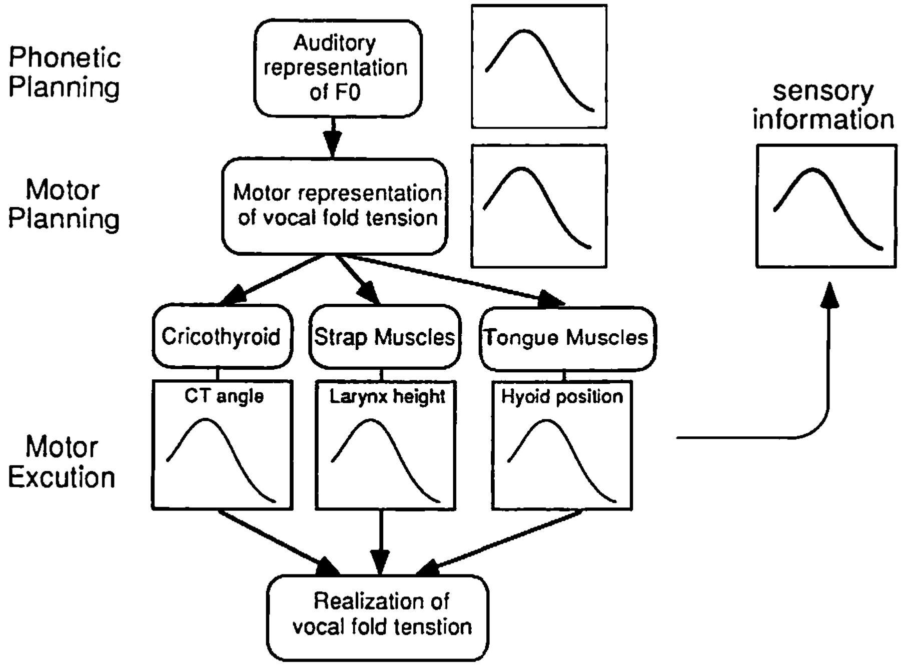


Fig. 5 The process of F0 control from internal tonal representation of F0 to its realization. The resulting proprioceptive information from sensory afferent information formulates sensation of pitch in conjunction with auditory information.

3. EMG analysis of tongue muscle activity for vowel production

Another example of illustrating the relationship between morphological alternation and sensory-motor patterning in speech function is observed in vowel gesture of the tongue. This section describes our recent analysis on electromyographic (EMG) data of tongue muscles in vowel production (Kusakawa, 1993). The deformation of the tongue is a unique phenomenon in human motor system. While the most of motor organs employ joint movements, the organ of vowel production is composed of muscle bundles alone. This unique anatomical composition of the tongue induces an atmosphere of complexity. However, the tongue musculature does not have such a complicated mechanical functions as the larynx does, as far as vowel gestures are concerned. The EMG data of the extrinsic muscles of the tongue have indicated that the contraction of the muscles produced a deformation which can be predicted by their anatomical distribution (Baer, Alfonso, & Honda, 1988).

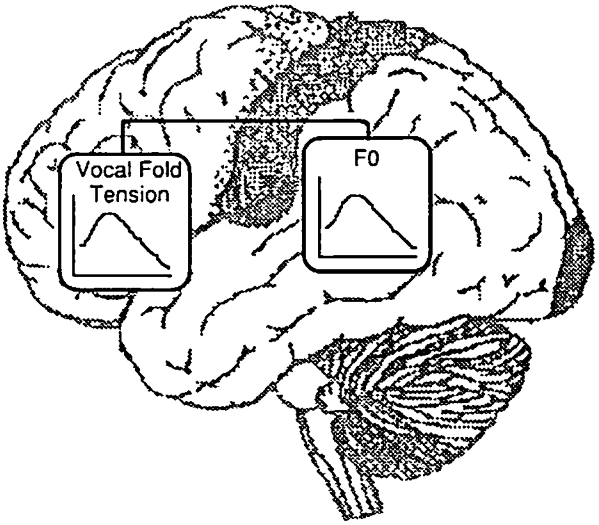


Fig. 6 The relationship of motor and sensory patterns of F0 in the brain. The module of F0 in the brain has a transparent relationship between motor and sensory representations.

3.1 The articulatory trajectory of the tongue in the muscle force space

Anatomical background of vowel production is the composition of the extrinsic tongue muscles and the right-angled cavity of the oropharyngeal tube in human body. The variation of vowel articulation corresponds to the variation of contractile pattern, or group selection, of these muscles. The functional components of the tongue muscles for vowel production is illustrated by a simple computer model in Fig. 7. It has been understood that the genioglossus muscle, the largest muscle in the tongue along the midline, has functional subdivisions; empirically, the anterior fibers (GGA) and the posterior fibers (GGP). As a counteracting fashion to the genioglossus, two muscles are attached to the both sides of the tongue; the styloglossus (SG) and the hyoglossus (HG). The figure shows deformational effects of each muscle contraction, and also illustrates a functional relationship among those muscles as two pairs of antagonistic muscles. An antagonistic pair of GGP and HG determines tongue position along the primary (high-front to low back) oblique axis. Another pair of GGA and SG defines the secondary (low-front to high-back) axis. This observation on component tongue movements is supported by Maeda's X-ray cinematographical analysis for articulatory modeling (Maeda, 1990).

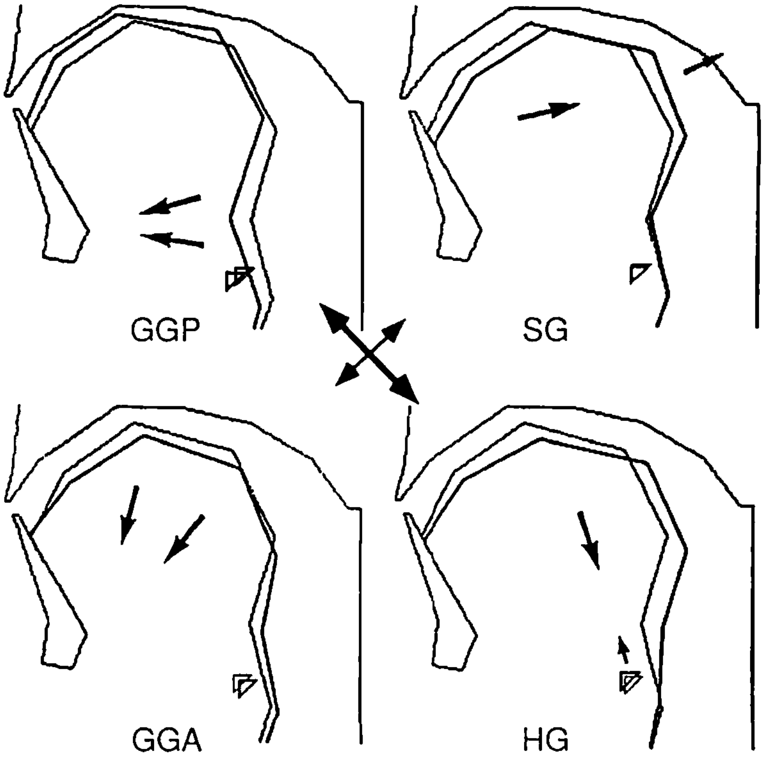
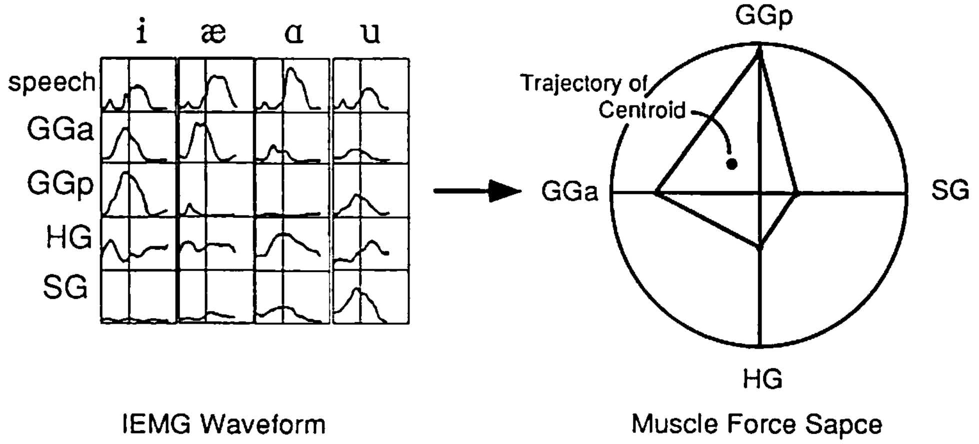


Fig. 7 Antagonistic relationship among the extrinsic tongue muscles. Two pairs of antagonist muscles control the position of the tongue in the vocal tract cavity.

The concept of antagonist pairs of the tongue muscles has also been supported by our recent study (Kusakawa , Honda, & Kakita, 1993) for reconstructing internal representation of motor pattern for vowel gestures. The EMG data which were recorded at Haskins Laboratories (Baer, Alfonso, & Honda, 1988) have a complete set of the extrinsic tongue muscles, and the utterance has 11 English vowels in /əpVp/ word frame. The ensemble average integrated EMG signals from four extrinsic muscles were transformed into "the articulatory trajectory in the muscle force space," which was defined by two orthogonal axes of the antagonist muscle pairs. The trajectory represents overall force and direction applied to the tongue by combined contraction of muscles. Fig. 8 shows the method for the transformation from EMG to the articulatory trajectory in the muscle force space. Fig. 9 demonstrates computed trajectories for the word utterances with four extreme vowels and the acoustic data for the words from the same experiment. The trajectories of the utterances start from initial /əp/ cluster to the extreme points, or the targets, of the vowels /i/, /æ/, /ɑ/, and /u/. The distribution of the vowels resembles with vowel distribution in classical phonetic space. The distribution of acoustic data, in the F1-F2 diagram in the figure, also has a similarity with the vowel pattern in the muscle force space. It is speculated that the coincidence of vowel distributions in the acoustic and articulatory domains should provide considerable influence on the organization of speech function in the brain.


IEMG Waveform

Muscle Force Sapce

Fig. 8 The method for composing the articulatory trajectory for English vowels in the muscle tension space (Kusakawa, 1993).

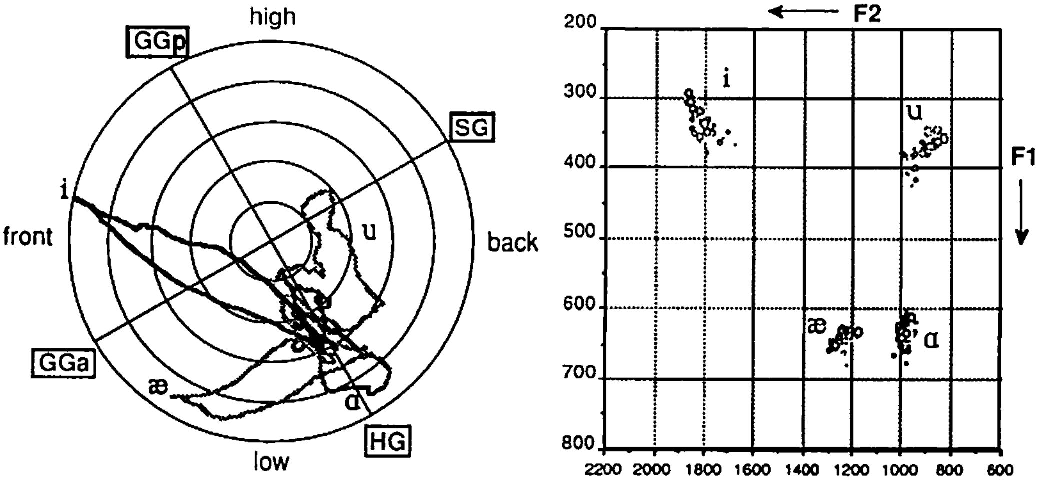


Fig. 9 The comparison of vowel's articulatory trajectory and formant pattern in F1-F2 diagram.

3.2. Vowel distribution in sensory and motor representations

Fig. 10 is the analogy of the association of sensory and motor patterns in the internal representations applied for the case of the organization of vowels. The internal representation of the motor pattern of vowels is the distribution of vowel in the articulatory space, i.e., the position of the tongue relative to the vocal tract wall. The sensory representation of vowel pattern is the formation of vowel distribution in sensory space, i.e., the acoustic characteristics of vowels. These sensory and motor representations of vowels can be hypothetically illustrated by measurable parameters. In the figure, the distributions of vowels are shown by tongue positions for motor pattern and F1-F2 diagram for sensory pattern. The resemblance between them suggests another evidence of close relationship in motor and sensory organization. Similar to the case of the organization of F0 control, the integration of motor and sensory pattern for vowels can be performed by a simple calculation of correlation.

The resemblance between these two vowel representations also supports the evidence of the interaction between morphological development of human vocal tract and brain function. The vowel distribution in F1-F2 diagram only depends on vocal tract area function, i.e., positions of a stricture in the vocal tract. A straight vocal tract could demonstrate the same vowel distribution in acoustic space. However, vowel distribution in articulatory space strongly depends on the formation of human vocal tract of a right-angled cavity. Morphological formulation of two-pairs of antagonist muscles of the tongue also contributes to vowel distribution in the articulatory space. The organization of vowel system can be recognized as the result of the development of human vocal tract associated with cerebral enlargement.

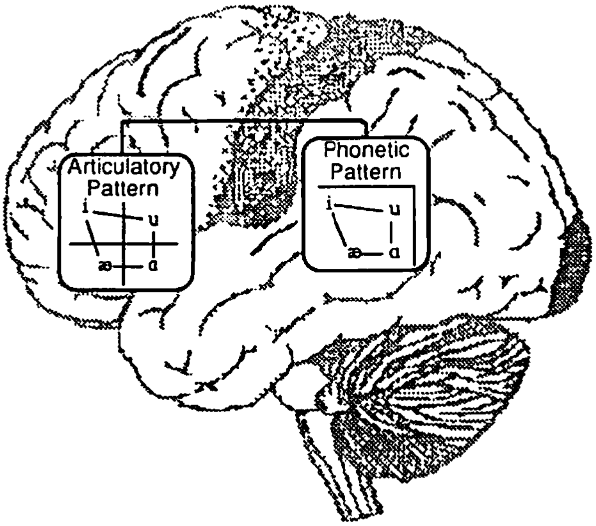


Fig. 10 The relationship of motor and sensory patterns of vowels in the brain. The module of vowel system in the brain also has a transparent relationship between motor and sensory representations.

4. Conclusion

The observations of human phonatory and articulatory organs by means of MRI and EMG experiments suggest that an element of speech is formulated as a module which has a transparent relationship between motor and sensory representations. The F0 module is composed of the representations of vocal fold tension and perceived pitch. The vowel module is formed by the representations of tongue positions and acoustic characteristics of vowels. The overall image of speech pattern seems to be produced by the integration of these component modules. The motor and sensory patterns in each module resemble with each other. This coincidence is a result of morphological development of human speech organs. The development of brain function to recognize peripheral structures is another factor to establish the association of motor and sensory patterns. To summarize, the organization of speech depends on the interaction between peripheral structures and brain function. The interaction is based on the characteristics of peripheral morphology which is innate to human body. The concept proposed in this paper also accounts for the background of innate human speech.

Acknowledgment

The EMG data in this paper were recorded and pre-analyzed at Haskins Laboratories in New Haven. This part of the study was supported by NIH Grant to Haskins Laboratories (PI: Professor Katherine S. Harris).

References

Atkinson, J. E. (1973). Correlation analysis of the physiological factors controlling voice frequency. *J. Acoust. Soc. Am.*, 63 (1), 211-222.

Baer, T., Alfonso, P. J., & Honda, K. (1988). Electromyography of the tongue muscles during vowels in /əpVp/ environment. *Ann. Bull. RILP*, 22, 7-19.

Erickson, D. (1976). *A Physiological Analysis of the Tones of Thai.* Unpublished Ph.D. dissertation, Univ. of Connecticut.

Hayama, S. (1991). The origin of human vocal organs. In A. Shibatani, K. Nagano, T. Yoro, (eds.) *Evolution, 4,* University of Tokyo Press, 173-198. (in Japanese)

Hirai, H., Honda, K., Fujimoto, I. & Shimada, Y. (1993). Analysis of magnetic resonance images on the physiological mechanism of fundamental frequency control. *ATR Technical Report,* TR-A-0166. (in Japanese)

Honda, K. & Fujimura, O. (1991). Intrinsic vowel F0 and phrase-final F0 lowering: phonological vs. biological explanations. In J. Gauffin and B. Hammerberg (eds.) *Vocal Fold Physiology,* San Diego: Singular Publishing Group, 149-158.

Katsuki, Y. (1950). The function of the phonatory muscles", *Japan J. Physiol.,* 1, 29-36.

Kusakawa, N., Honda, K. & Kakita, Y. (1993). Construction of articulatory trajectories in the space of tongue muscle contraction force. *ATR Technical Report,* TR-A-0171,1993. (in Japanese)

Maeda, S. (1990). Compensatory articulation during speech; evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*, Dortrecht: Kluwer Academic Publishers, 1990, 131-150.

Simada, Z. B., Niimi, S. & H. Hirose. (1991) On the timing of the sternohyoid muscle activity associated with accent in the Kinki dialect. *Ann. Bull. RILP*, 25, 39-45.

Sonninen, A. (1954). Is the length of the vocal cords the same at all different levels of singing? *Acta Otolaryngol. Suppl.* 118, 219-231.

Sugito, M. (1982). *Studies of Japanese Accent*, Tokyo: Sanseido, (in Japanese).