3-3 Study of Speech Communication Mechanisms based on the Understanding of Human Biological Functions

Kiyoshi HONDA
Advanced Telecommunications Research Institute International

Human speech acts as vehicles of thought, communication, emotional state, and physical identity; these functions together contribute to our life and social organization. The value of speech is given by its versatility, which is derived basically from innate biological abilities equipped to humans. If the machinery emerged that handled all the speech functions, it would deserve human-equivalent significance. Towards evolution of spoken language automata as a prospective goal, we launch a research project to look into biological functions underlying human speech in the three areas; form and function of the speech organs, biophysical modeling of speech production mechanisms, and functional imaging of speech-related pathways.

Keywords
Speech, Biological functions, Individuality, Speaker normalization, Magnetic resonance imaging

1 Introduction

Human speech plays a significant role in our daily life and social organization as a vehicle of information exchange. When we consider the functional aspects of speech, we can first recognize it as a medium of thought, which is often referred to as “internal speech.” This symbolic speech, not expressed in sound, is used for reasoning and mental writing, and is fundamental to intellectual activity and cultural transmission. Speech is widely recognized as a tool of communication and its value is inestimable. If the interruption of telecommunication were to somehow take place, we can readily guess that this would give rise to extreme confusion in our society. Speech acts reflect emotional consciousness through tonal changes, and play an important role in creating human relationships. In addition, speech functions as a means of identification of talkers and a source of information concerning physical conditions. More specifically, we can guess who and how a person is even from coughing or other non-speech sounds.

Such a variety of speech functions are the subject of extensive investigations; if machinery were to emerge to technologically comprehend all of human speech functions, such machinery would have value equivalent to humans as an artificial organism. Speech research must aim at understanding basic biological functions of humans and aim towards creating machinery capable of performing a variety of speech functions.

2 Issues in speech-communication research

2.1 Brief history of speech research

Speech research has spanned a period of 300 years, originating in an experiment on vocal cord vibration in 1700 in France. In the late 18th century, von Kempelen analyzed mechanisms of producing vowels and consonants, and manufactured a mechanical speech synthesizer. In the 19th century, Wheatstone and Helmholtz developed vowel-production theories, and they found vocal cavity resonance, or formant, to be a characteristic that determines vowel color[1]. At the time, J. Muller was conducting research in phonation mechanisms using a model of the vocal folds and is now generally credited as being the first person to advocate the source-filter theory. His theory has led, in the 20th century, to Chiba and Kajiyama's vowel research based on X-ray images and to Fant's research for the acoustic theory of speech production[2]. This acoustic theory was gradually simplified, and the vocal tract is now commonly regarded as a single tube. Although such simplification leads to a general understanding, it is unrealistic to assume such simple mechanisms in the speech production process. It may be time to reexamine the biological mechanism by which speech is produced.

2.2 Issues to be investigated

While human speech mechanisms have in many ways been clarified in the past 300 years, there remain many issues to be investigated. We will discuss the following two issues as tasks to be solved in the 21st century.

The first topic is the talker characteristics, or individual information conveyed by speech sound. Tone color varies from person to person, just as the appearance of individual faces varies. The individuality in face and speech is an aspects of one's physical identity that cannot be modified by intention. In order to develop future biometric technologies, knowledge of physical factors causing talker characteristics must be established. Since we have little evidence on this matter, it may be reasonable to begin with the most likely sources such as the gross shape and side-branches of the vocal tract. As shown in Fig. 1, the vocal tract consists of two side-branches, i.e., the nasal cavity and the periform fossa, and their characteristics are reflected by the detail of the spectral envelope[3]. Furthermore, the length and shape of the vocal tract have been hypothesized to be factors determining talker characteristics, and it is further speculated that the length of the pharyngeal cavity influences the formants[4].

Second is the "exchangeability" of vowels. There is a marked difference in vowel color between adults and children, due to differences in body size. However, we know that speech comprehension is maintained in parent-child communication, without consciousness of this difference. Such vowel exchangeability is referred to as "vowel normalization" and has long been debated; however, it

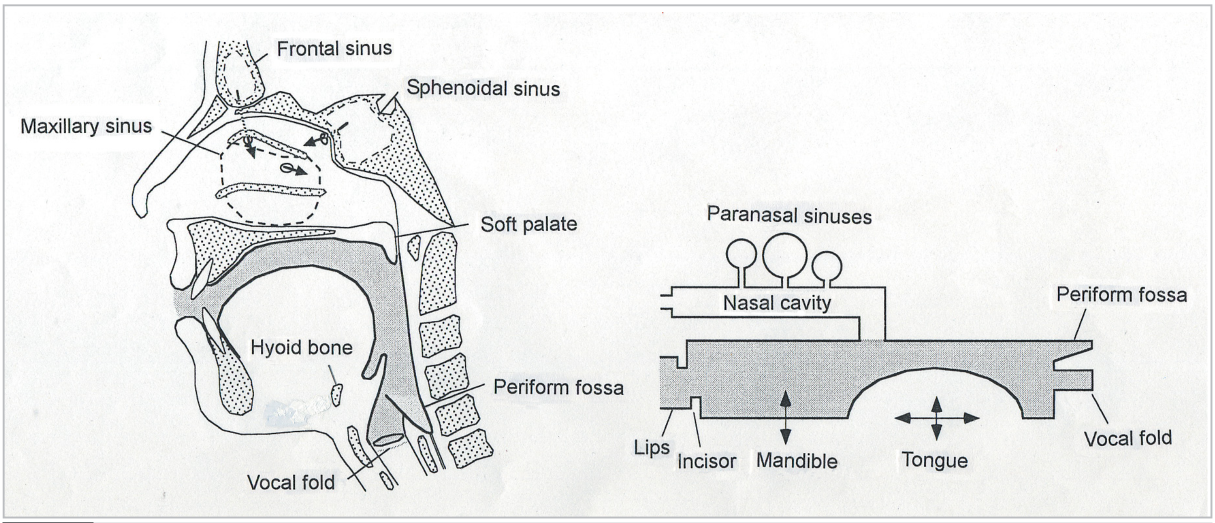

Fig.1 *Fine structure of the speech organ (left) and an acoustic tube model (right)*

remains unresolved in speech-perception theories to date. The vowel formants vary widely, in inverse proportion to the length of the vocal tract. The vowels /a/ in adults and /o/ in children are distributed within nearly the same area, whereas they are easy to identify. How we categorize vowels from among speakers is an important issue to be investigated in speech recognition research. However, so far we have no physiological evidence of vowel normalization mechanisms.

2.3 Biological foundations of speech

The above two issues explain that speech demonstrates exchangeability as a communication tool, while it sufficiently distinguishes individuality. Such individual identification appears to have been generated through the evolutionary process, from the protomorphic stage onward. In animals, vocal sounds act as means of individual identification and message communication. In particular, vocal identification between parents and children has been investigated in birds and mammals. In the case of birds, auditory function is known to be more dominantly involved in individual identification than is visual function. Studies of animal speech-behavior could provide useful knowledge in understanding the biological functions of speech in humans.

3 Speech research using magnetic resonance imaging (MRI)

The Biological Speech Science project at ATR is conducting a morphological measurement of the speech organs, a modeling of the speech-production mechanisms, and functional brain imaging during speech production, with a focus on the biological aspects of speech communication. A few findings obtained by MRI are introduced below.

3.1 Morphological measurement of the speech organ

The shape of the vocal tract has been studied by X-ray imaging. MRI has enabled the first three-dimensional imaging of vocal tract, and the accuracy of measurement has been insured by dental imaging technique. Fig. 2 (a) shows an image of the vocal tract extracted from the MRI data when the vowel /e/ is produced. As illustrated by this figure, the vocal tract, long considered a simple tube, consists of large and small branches. Recent technical advancement has also enabled three-dimensional motion imaging[5]. Fig. 2 (b) shows the time course of vocal tract area function extracted from three-dimensional motion imaging when the Japanese vowels /a/, /i/, /u/, /e/, and /o/ are produced. Thus, MRI presents a significant contribution to the reevaluation of the acoustic theory of speech production

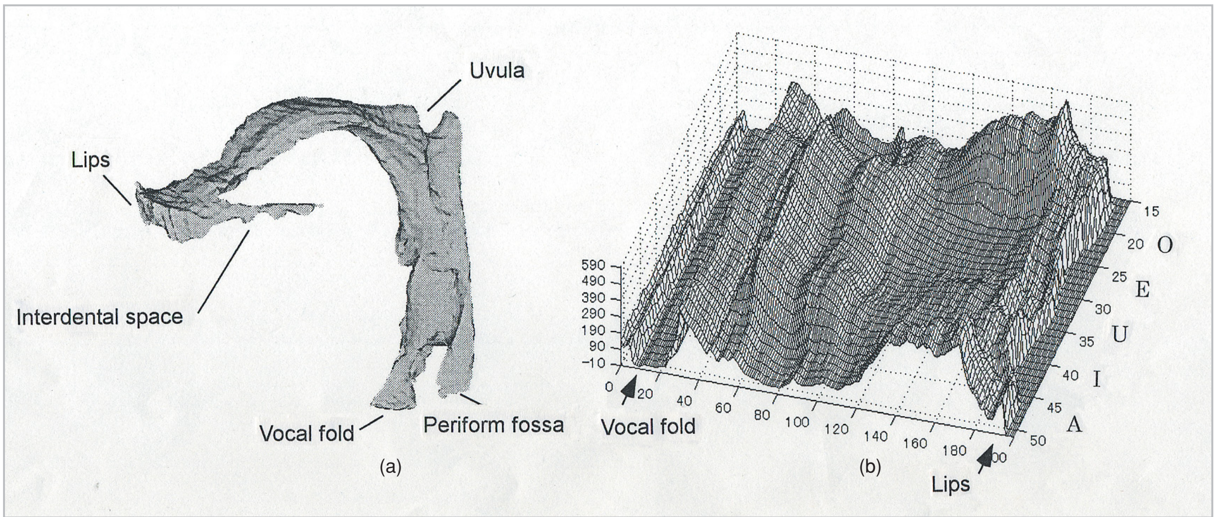

Fig.2 *Three-dimensional features of the vocal tract (left) and temporal sequence of the area function associated with the voiced vowels /a/, /i/, /u/, /e/, and /o/*

and the elucidation of factors involved in speech individualities.

3.2 Brain activity during speech production

The cerebral cortex has been investigated in detail in relation to brain activity during speech production and hearing. It is necessary, from the standpoint of biological functions, to extend measurement areas to include the brain stem. Functional MRI (fMRI) faces several technical problems in the imaging of the brain stem. We investigated orientation of the imaging plane to be suitable for recording brainstem activity during speech production. Fig.3 shows a result during spontaneous speech production. As shown in the figure, activity of the general motor circuitry, such as the thalamus, pons, and cerebellum, is visible. We are also planning to investigate a method for reducing noises that are generated from the machinery in anticipation of speech perception research, with the aim of using MRI for visualizing the neuronal pathways involved in the production and perception of speech.

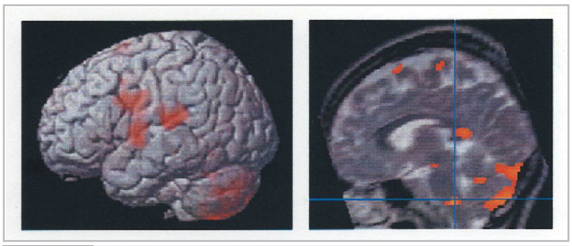

Fig.3 Active areas in the cerebral cortex (left) and activities of the motor pathway in the brain stem (right) during speech production

References

1 T. Chiba and M. Kajiyama, "The Vowel: Its Nature and Structure", Tokyo: Tokyo-Kaiseikan, 1941.

2 G. Fant, "Acoustic Theory of Speech Production", The Hague, Mouton, 1960.

3 K. Honda, "Biological foundations of speech", Gengo no Kagaku, Vol. 2. Tokyo, Iwanami, 1998.

4 K. Honda, "Facial appearance and vocal quality in man", J. Acoust. Soc. Jpn, 57, 308-313, 2001.

5 H. Takemoto, K. Honda, S. Masaki, Y. Shimada, I. Fujimoto, S. Takano, and K. Takeo, "Three-dimensional motion imaging of speech organs using a synchronized MRI sampling method", Proceedings of Acoust. Soc, Jpn, pp.251-252, Mar. 2001.

***Kiyoshi HONDA**, M.D., D. Ms.*

Project Leader, Biological Speech Science Project, Information Sciences Division, ATR International

Speech Science