HISTORY OF CHIBA AND KAJIYAMA AND THEIR INFLUENCE IN MODERN SPEECH SCIENCE

Takayuki Arai

Dept. of Electrical and Electronics Eng., Sophia Univ., Tokyo, Japan
Speech Communication Group, RLE, MIT, Cambridge, MA, USA
arai@sophia.ac.jp

ABSTRACT

More than 60 years ago, Chiba and Kajiyama published "The Vowel: Its Nature and Structure" in 1942, and it was fundamental to the establishment of the modern acoustic theory of speech by Stevens, Fant, and other eminent scientists. This book approached the mechanism of vowel production and perception from the viewpoints of physiology, physics and psychology, and importantly, it integrated them together for the first time. They showed that the waveform of a vowel is treatable by Fourier analysis, introduced the concept of the electric-circuit analog to simulate a resonance of the vocal tract, and succeeded in calculating vowel spectra from data of the vocal tract shape. In the present study, we first review the topics of this historical book and reconfirm that it established the basis of currently accepted theories on vowels, such as source-filter theory and perturbation theory. Furthermore, we confirm that their accomplishments were extremely influential for many researchers in the history of modern speech science. Finally, the usefulness of "Chiba and Kajiyama" from the pedagogical point of view is discussed. Arai (J. Phonetic Soc. Jpn., 2001) replicated Chiba and Kajiyama's physical models of human vocal tract and showed that they are extremely effective in the classroom. We further extend these physical models, as educational tools, to consonants, such as nasals, stridents and liquids (/r/ and /l/) based on modern literature, particularly "Acoustic Phonetics" (Stevens, 1998).

INTRODUCTION

Chiba and Kajiyama's book was published more than 60 years ago, and both the Phonetic Society of Japan and the Acoustical Society of Japan made special issues of its 60th anniversary. Once you read the articles from these special issues in addition to the original book, you will understand more about the depth of their work. The importance of this book lies in the fact that all related areas were merged into a single science; especially, Chiba seemed to keep to his steady policy to introduce natural science, namely physics, into the study of phonetics (Maekawa, 2002).

The features of Chiba and Kajiyama's study can be summarized as follows (Maekawa and Honda, 2001, Kasuya, 2001; Maekawa, 2002; Honda 2002): 1) they collected the physiological data and measured the three-dimensional vocal tract shape (area function) by using the most advanced technologies at the time including the X-ray imaging device; 2) they calculated vowel spectra / resonance frequencies from the data for the first time; 3) they introduced electrical circuit theory and established the acoustic theory of vowel production; and 4) they concluded that the acoustic nature of vowels is determined by vocal tract shape. In the present paper, we thus review the topics of this historical book and reconfirm that it established the basis of currently accepted theories on vowels. Furthermore, we confirm that their accomplishments were extremely influential and are still useful from the pedagogical point of view.

FOUR PARTS IN "CHIBA AND KAJIYAMA"

"The Vowel" consists of the following four parts, and many of the results were obtained by using the most advanced technologies at the time including an oscillograph, a stroboscope, and an X-ray apparatus.

Part 1 "The Action of the Larynx": The voice source was analyzed, i.e., physiology of larynx, glottal air flow, and the dynamic aspects of the vibration of the vocal folds under various voice registers.

Part 2 "The Mechanism of Vowel Production": One of the main topics of Part 2 is a historical dispute of so-called "Harmonic (Steady State) vs. Inharmonic (Transient) Theories" (Maekawa, 2002; Honda, 2002). Chiba and Kajiyama took the position that the Harmonic Theory (vowel sounds are considered as a forced response) and the Inharmonic Theory (vowel sounds are considered as a free damped oscillation) are intrinsically the same (Maekawa and Honda, 2001), and they applied Fourier analysis to obtain vowel spectra. Then, Part 2 follows discussions of the theory of simple resonators and their equivalent networks, and basic aspects of vocal tract shape. In general, the vowels /i/ and /e/ are explained by a Helmholtz resonator, while the vowel /a/, /o/, and /u/ are represented by a double resonator (Honda et al., 2004). The natural frequencies calculated from the approximated vocal tract coincided fairly accurately with the values obtained by Fourier analysis (Maekawa, 2002).

Part 3 "The Measurement of the Vocal Cavity and the Calculation of Natural Frequencies": The vocal tract shape was measured using a combination of X-ray photography, palatography, and laryngoscopic observation of the pharynx (Maekawa, 2002). The cross-sectional area function for each vowel was then used to calculate the spectra of the sounds (Maekawa and Honda, 2001). They successfully approximated the first two formant frequencies from the vocal tract shapes, and the frequencies matched well to the ones calculated from natural speech sounds (Motoki, 2002). The acoustic theory of resonators involved in vowel production provided significant new insights into the relation between vocal-tract shape, distribution of pressure and velocity amplitude, and formant frequencies (Stevens, 2001). The book also demonstrates the very wide range of cross-sectional areas that exist across vowels in the pharyngeal region of the vocal tract (Stevens, 2001). In Chapter XI, the distribution of volume and particle velocities in each vocal tract was computed. The discussion of the effect of the characteristics of vocal tract shape on its natural frequency is a generalization of the approximation that was done for each individual vowel in the previous chapter (Maekawa and Honda, 2001).

Part 4 "A Subjective Study of the Nature of a Vowel": They discussed human perception of vowels in contrast to the production theory (Maekawa and Honda, 2001). It involves systematic studies of the variation of vocal tract dimensions with sex and age (Fant, 2001). They further examined the problem of vowel normalization by developing a space-pattern account of vowel perception in opposition to formant-based accounts (Honda, 2002). By space pattern they implied frequency domain shape aspects such as the dominance of a single spectral region of some of the back vowels, and of two main spectral maxima in front vowels, characterized by a fixed ratio rather than absolute values (Fant, 2001). They also explained that resonance determines vowel quality by applying the concept that the cochlea has a frequency-analysis mechanism with low resolution (Kasuya, 2001).

PEOPLE INFLUENCED BY "CHIBA AND KAJIYAMA"

Any researchers engaged in speech science may benefit from Chiba and Kajiyama directly or indirectly. Gunnar Fant and Kenneth N. Stevens might be the two who are most influenced by them. The first time that Stevens came across "The Vowel" (see Fig. 1) was around 1950, and the book showed him how it was possible to combine his interests in speech and in electrical engineering (Stevens, 2001). Stevens' own research on articulatory/acoustic relations for vowels, beginning in the early 1950's, was stimulated by Chiba and Kajiyama and Fant's source-filter theory. In the mid 1950s, Fant was preparing his book "Acoustic Theory of Speech Production" (Fant, 1960), and "The Vowel" was of a great help in his processing of X-ray data of Russian speech sounds (Fant, 2001).

Table 1 shows the events occurred around that time period. In 1950, the following people were in the Acoustics Lab. at MIT: Prof. R. H. Bolt (director), Prof. L. L. Beranek (technical director), K. N. Stevens (doctoral student), J. L. Flanagan (master's student), and Gunnar Fant (visitor). In 1952, Stevens got his doctoral degree. After that, he interviewed at Bell Labs but ended up with as half-time research staff at MIT and half-time consultant at BBN. In 1954, Stevens became an assistant professor at MIT. At the same time, Beranek left MIT and moved to BBN (Stevens, 2004). (Stevens was the first doctoral student of Beranek, and Flanagan was the first doctoral student of Stevens. All three were the president of the Acoustical Society of America. They were also all awarded the National Medal of Science, but interestingly, in the reversed order.)

By the way, Morris Halle recalls that Roman Jakobson had a copy of Chiba and Kajiyama around 1950 (Halle, 2004). (Halle became Jakobson's student in 1948 at Columbia University, and they moved to Harvard University in 1949.) A draft of the thank-you letter written by Jakobson for Chiba is stored in the MIT Archives. The letter is from Cambridge, Feb. 4, 1951, and reports that a copy was sent to Jakobson from Chiba around that time. In 1951, Halle bought Chiba and Kajiyama's book secondhand (Halle, 2004). The back cover has the handwritten date of March, 1942, with the compliment of the author. Another document stored in the MIT Archives was a letter written by Chiba for Jakobson. The letter is from Tokyo, Aug. 8, 1956, and says "With regard to the republication of *The Vowel*, I am going to apply to the Ministry of Education for a subsidy which will, I hope, be granted, though not to my desired extent, because our Ministry is now willing to do its best for the advancement of international cultural exchange."

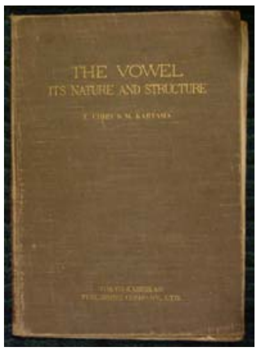

Figure 1. "The Vowel" possessed by Stevens.

WHAT "CHIBA AND KAJIYAMA" TAUGHT US

Source-Filter Theory

Human beings are able to independently control phonation (source) at the larynx and articulation (filter) at the vocal tract, and Chiba and Kajiyama solved the mechanisms of speech production based on the concept of phonation and articulation scientifically and systematically (Kasuya, 2001). Fant was trained in electrical circuit theory in 1944 and 1945 from his teacher who was an expert on filter theory. Then, Fant encountered "Chiba and Kajiyama," perhaps when he visited MIT (Fant, 2004). Their view of phonation and articulation merged with Fant's filter theory. It lead to the so-called "source-filter theory of vowel production" in the modern acoustic theory of speech production (Fant, 1960), and this is one of the reasons that Chiba and Kajiyama is counted as a classic in a history of science (Maekawa and Honda, 2001).

Perturbation Theory

A general approach to perturbation theory is based on a theorem by Ehrenfest (1916). Perturbation theory tells us the relations between vocal tract configurations and formant frequencies by examining the changes in the formant frequencies that occur as a result of small perturbations of the area function in some region along the length of the vocal tract (Stevens, 1998). Chapter XI of "Chiba and Kajiyama" shows a number of figures giving the calculated distribution of sound-pressure amplitude and velocity amplitude for the first two formants for different vowel configurations based on their measurement. It is considered that Chiba and Kajiyama showed the physical phenomenon of wave propagation in the vocal tract for the first time (Motoki, 2002). It was, however, only some decades later that the relevance of such plots in predicting the acoustic effects of perturbations in vocal-tract shape was recognized (Stevens, 2001). Fig. 93 in "The Vowel" was cited by Fant and adopted by many subsequent monographs of acoustic phonetics, including very recent ones (Maekawa, 2002).

Petagogical applications

In the section "Artificial Vowels" (pp. 128-131) Chiba and Kajiyama synthesized vowel sounds, using physical models based on sectional measurements made of vocal tracts, and compared the synthetic outputs to those of natural vowels. Arai (2001) replicated Chiba and Kajiyama's physical models of the human vocal tract (Fig. 2) and showed that they are extremely effective in the classroom when demonstrating vowel production, especially, in a demonstration on what determines the quality of a vowel, source-filter theory, and perturbation theory, by combining a sound source, such as an artificial larynx. We extend these physical models, as educational tools, to consonants, such as nasals (Fig. 2), stridents and liquids (/r/ and /l/) based on modern literature, particularly "Acoustic Phonetics" (Stevens, 1998). Recently, Arai's models were used in Stevens' class on Speech Communication at MIT (Fig. 2). The students showed a lot of interest in seeing and hearing real models of the vowels with an excitation source. They were mainly interested in the different shapes for the vowels, and the capability to simulate a variety of other shapes (Stevens, 2004).

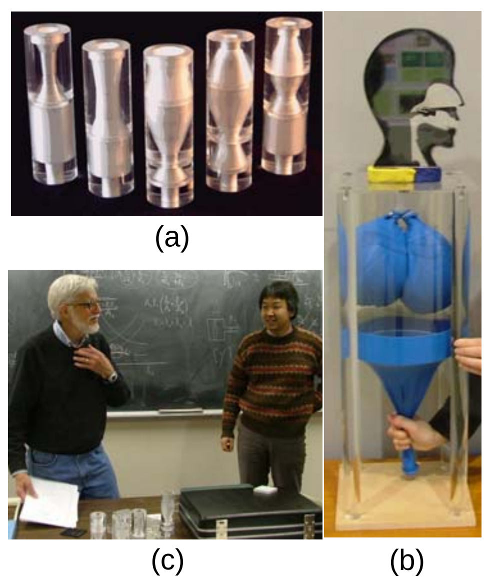

Figure 2. (a) Arai's models of human vocal tract as educational tools. (b) A model for nasalized vowel /a/ with a lung model. (c) Arai's models used in a class by Stevens.

ACKNOWLEDGEMENTS

I would like to thank all of the people who helped me in various ways, especially Ken Stevens, Joe Perkell, Stefanie Shattuck-Hufnagel, Sharon Manuel, Janet Slifka, other members of the Speech Communication Group at MIT, Morris Halle of MIT, Ben Gold of MIT Lincoln Lab., and Gunnar Fant of KTH.

REFERENCES

Arai, T. (2001) The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity, *J. Phonetic Soc. Jpn., 5(2)*, 31-38.

Ehrenfest, P. (1916) *Proc. Amsterdam Acad., 19*, 576-597. (Citation from Schroeder, M. R. (1976) *J. Acoust. Soc. Am., 41*, 1002-1010.)

Fant, G. (1960) *Acoustic Theory of Speech Production*, The Hague, Netherlands: Mouton.

Fant, G. (2001) T. Chiba and M. Kajiyama, pioneers in speech acoustics, *J. Phonetic Soc. Jpn., 5(2)*, 4-5.

Fant, G. (2004) *Personal communication*.

Gold, B. (2004) *Personal communication*.

Halle, M. (2004) *Personal communication*.

Honda, K. (2002) Evolution of vowel production studies and observation techniques, *Acoust. Sci. & Tech., 23(4)*, 189-194.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S. & Takano, S. (2004) Exploring human speech production mechanisms by MRI, *IEICE Trans. Inf. & Syst., E87-D(5)*.

Jakobson, R. MC72. Institute Archives and Special Collections, MIT Libraries, Cambridge, Massachusetts.

Kasuya, H. *et al.* (2001) Overview in each research field: Speech, *J. Acoust. Soc. Jpn., 57(1)*, 11-20.

Maekawa, K. & Honda, K. (2001) On the Vowel, Its Nature and Structure and related works by Chiba and Kajiyama, *J. Phonetic Soc. Jpn., 5(2)*, 15-30.

Maekawa, K. (2002) From articulatory phonetics to the physics of speech: Contribution of Chiba and Kajiyama, *Acoust. Sci. & Tech., 23(4)*, 185-188.

Motoki, K. (2002) Three-dimensional acoustic field in vocal-tract, *Acoust. Sci. & Tech., 23(4)*, 207-212.

Stevens, K. N. (1998) *Acoustic Phonetics*, Cambridge, MA: MIT Press.

Stevens, K. N. (2001) The Chiba and Kajiyama book as a precursor to the acoustic theory of speech production, *J. Phonetic Soc. Jpn., 5(2)*, 6-7.

Stevens, K. N. (2004) *Personal communication*.

APPENDIX: REPRODUCING SPECTRA OF THE VOWELS IN CHIBA AND KAJIYAMA

Chiba and Kajiyama showed that Fourier analysis is applicable to vowel sounds. However, they executed the Fourier transform manually. The following figure shows the reproduced waveforms from the Chiba and Kajiyama's book and their spectra by the modern computer software XKL (a revision of the software package developed by Dennis Klatt).

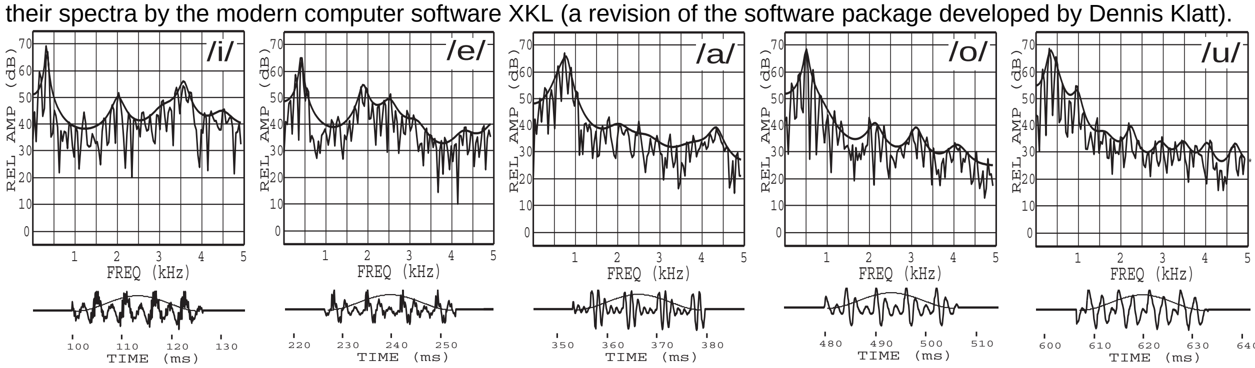

Table 1. Chronology associated with “Chiba and Kajiyama.”

1929	Chiba started Phonetics Lab. at Tokyo School of Foreign Languages
	"Speech and Hearing" by Fletcher (New York: van Nostrand)
	Acoustical Society of America (ASA) was founded
1933	Kajiyama joined the Phonetics Lab at Tokyo School of Foreign Languages
1934-39	Period for the research project on "The Vowel"
1939	“The vocoder” by Dudley (Bell Labs Record, 17, 122-126)
	World War II started
1941	Pacific War started
1942	"The Vowel" was published (Tokyo-Kaiseikan)
1945	The end of the war
1947	"Visible Speech" by Potter, Kopp, and Green (New York: van Nostrand)
1948	Stevens became a doctoral student at Acoustics Lab., MIT
1950	The first citation of "The Vowel" by Fant (MIT Acoustics Lab. Quarterly Progress Report, Jul.-Sep.)
	Chiba became a professor of English at Sophia University
1952	Stevens received Sc.D. (The dessertation: “The perception of sounds shaped by resonant circuit”)
	Jakobson, Fant and Halle (1952) cited “The Vowel” (Acoustics Laboratory Technical Report 13, MIT)
1953	Stevens, Kasowski, and Fant (1953) cited "The Vowel" (JASA, 25, 734-742)
1954	Stevens became an assistant professor at MIT
1955	Stevens and House (1955) cited "The Vowel" (JASA, 27, 484-493)
1958	The 2nd Edition of "The Vowel" was published from the Phonetic Society of Japan
	Fujimura joined the Speech Communication Group at MIT as Visiting Researcher (until 1961)
	Fujisaki joined the Speech Communication Group at MIT as Fulbright Scholar (until 1959)
1959	Chiba passed away (in December in Tokyo, 76 years old)
1960	"Acoustic Theory of Speech Production" by Fant (The Hague, Netherlands: Mouton)
1965	"Speech Analysis, Synthesis, and Perception" by Flanagan (New York: Springer-Verlag)
	Cooley & Tukey published a paper on FFT (Mathematics of Computation, 19, 297-301)
1968	LPC by Atal & Schroeder and Itakura & Saito at International Congress on Acoustics (Tokyo, Japan)
1969	"Digital Processing of Signals" by Gold and Rader (New York: McGraw-Hill) *

*Note: In 1966, Stevens asked Ben Gold to spend one year on MIT campus in the Speech Communication Group (because of Gold’s speech research). At that time, Charles Rader and Gold had developed a fair amount of material on DSP (which originally began as a book on vocoders). The DSP work by Gold and Rader was motivated almost completely by speech. In Stevens’ Group, Gold decided that there was enough DSP material for a graduate seminar, but not quite enough for a text book. Near the end of Gold’s class, his graduate assistant, Tom Crystal, who also was a fellow at Bell Labs, showed him a paper which was just being circulated at Bell Labs, by Cooley and Tukey. Soon after, Gold’s old boss (who hired him in 1953) introduced him to Alan Oppenheim, then a young Assistant Professor in the E.E. Department at MIT. Oppenheim had done his Ph.D. thesis on Homomorphic Deconvolution (using analog techniques). Oppenheim became very interested in the DSP approach to his problem. Meanwhile, Tom Stockham, who was a friend of Oppenheim and worked at the Lincoln Lab., had learned about the FFT and had developed the idea of high speed convolution. Given these developments, Gold now felt that a book on DSP was warranted. In 1966 and 1967 Gold became friends with Larry Rabiner, who was then Stevens' Ph.D. candidate. Rabiner did take Gold’s DSP class. About the time of publication, Ronald Schafer became Oppenheim's graduate student and Oppenheim had begun to teach his DSP graduate course. These friendships (Oppenheim & Schafer, and Rabiner & Gold) led to the publication of two more books. It was just a coincidence that this was just the moment that the ideas of DSP erupted (Gold, 2004).