Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

T Chiba and M Kajiyama: Pioneers in speech acoustics

Fant, G.

journal: TMH-QPSR
volume: 42
number: 1
year: 2001
pages: 059-060

KTH Computer Science and Communication

T Chiba and M Kajiyama: Pioneers in speech acoustics[1]

Gunnar Fant
Dept of Speech, Music and Hearing, Royal Institute of Technology, KTH, Stockholm, Sweden

It is my privilege to add some comments to the celebration of the 60-year-anniversary of the publication of the book, *The Vowel, its Nature and Structure* by T. Chiba and M- Kajiyama (1941).

This is a belated recognition of a monumental piece of work in speech acoustics which came to my attention only in the mid 1950s when I was preparing my book *Acoustic Theory of Speech Production*. It was of a great help in my processing of X-ray data of Russian speech sounds, e.g. in deriving area functions from midsagital projections of the vocal tract and transverse airflow sections.

The authors had a solid insight in resonator theory and were aware of the multi-formant spectral patterns of vowels, but they lacked practical tools for calculating each resonance mode of a vowel, i.e., what is now known as the F-pattern, F_1, F_2, F_3, F_4.. (Fant, 1960). Accordingly, for the vowels [**i**] and [**e**] they limited their analysis to two formants, pointing out the basic lumped element single resonator origin of the lower formant and the standing wave character of an upper formant region determined by the mouth cavity regarded as a pipe open at both ends. The latter approximation places a potential resonance in the vicinity of F3 of [**i**] and F2 of [**e**].

But their book contains much more than resonator theory. I quote the Preface:

"*In short, we trust that we have succeeded in solving almost every problem of prime importance concerning the vowel.*"

Indeed, they had covered a wide range of topics, from vocal fold physiology, glottal air flow, vocal registers, basic aspects of vocal tract configuration, the wave equation, theory of simple resonators and their equivalent networks, auditory theory and speech perception.

Here follows a few comments in the light of more recent studies. Their vocal tract area functions, including a proper larynx tube and associated Sinus Piriformis pockets, are remarkably accurate. The influence of structures at the glottal termination are now fairly well understood through the studies of Fant (1960) and by Dang & Honda (1995) and by Fant & Båvegård (1997). Thus, the larynx tube exerts a major influence on F4, and the Sinus Piriformis acts as a shunting cavity producing a main spectral minimum in the region of 4500-5500 Hz. A detailed perturbation analysis relating each formant frequency to incremental variations of all parts of the vocal tract including the larynx was performed by Fant & Båvegård (1997), who also reported dimensions of the Sinus Piriformis.

Chiba and Kijayama were the first to make systematic studies of the variation of vocal tract dimensions with sex and age. The total length, expressed as a ratio with respect to a male of age 26 as a reference, was found to be 0.87 for a female of twenty, 0.80 for a boy of nine and 0.7 for a girl of eight years of age. These all lie within the expected range of values (Fant, 1975; Goldstein, 1980). Also, they noted the larger ratio of pharynx length to mouth length in males than in females.

Chiba and Kajiyama performed experiments on the perception of vowels submitted to various combinations of low-pass and high-pass filtering and to changes in the speed of a gramophone record. As a result of these studied they formulated a quite sensible space pattern theory of vowel perception:

"A vowel, even if it has lost parts of its space pattern, can be recognised as such, as long as the remaining part retains the characteristics of the space pattern."

To illustrate this concept they referred to an analogy to vision; "We recognise an object as long as a characteristic part is visible."

By space pattern they implied frequency domain shape aspects such as the dominance of

[1] This is a contribution to a series of articles in honor of Chiba and Kajiyama to be published in the Journal of thePhonetic Society of Japan.

a single spectral region of some of the back vowels, and of two main spectral maxima in front vowels, characterised by a fixed ratio rather than absolute values. Accordingly the [i] identity of a vowel could be ascribed to a ratio 10 of the two spectral peaks, e.g, at 270 Hz and 2700 Hz or at 400 Hz and 4000 Hz.

Even if such relations nowadays are better expressed on a Bark or a Mel scale, the basic concept of a two-formant perceptual model, was approached. Among the many studies later devoted to this concept are those of Carlson et al. (1975) and Fant (1978) who demonstrated a fairly successful synthesis of two formant stimuli to match full vowels. Moreover, an additional support of the two-formant model was attained from experiments performed by Carlson and Granström with a cochlea simulating filter bank, a set of broadband bandpass filters. A spatial measure across the set of channels was derived, counting how many output taps carried approximately the same zero crossing rate. A dominant spectral component in one channel imposes its zero crossings as a forced response in neighbouring channels which masks out their free responses. A two-peak structure of front vowel and a single peak in back vowels appeared in the frequency-space distribution of zero crossings (Fant, 1978). This line of approach deserves a renewed attention in auditory modelling.

An interesting reading in Chiba & Kajiyama is the historical survey of vowel theory. From the early part of the 20th century there developed two seemingly different theories. A production oriented view appeared in the early work of Willis, later followed up by Hermann & Scripture (references missing) who pointed out the role of the glottal flow pulses, each exciting a free damped oscillation, the frequency of which need not bear a harmonic relation to the larynx tone.

Helmholtz (1912) had a correct insight in the production process but preferred a perceptually oriented, strict harmonic view of how partials of the glottis tone are reinforced by vocal tract resonances, a forced response view. So it took some time before our present time-frequency transform theory was fully established and could be incorporated as a standard tool

Anyhow, the controversy still exists in auditory modelling. Do we respond to harmonics or to formants or simply to the dynamics of regions of spectral prominence? Is it a matter of narrowband or broadband filtering or both? Surely, we can detect harmonics at a constant F0, but this is limited to lab speech experiments.

I conclude my review by expressing my deep respect for the pioneering work of Chiba and Kajiyama, for the wide range and depth of their book which will remain an inspiration for many of us as well as for coming generations in speech sciences.

References

Carlson R, Fant G & Granström B (1975). Two-formant models, pitch and vowel perception. In: Fant G and Tatham MAA (eds.). *Auditory analysis and perception of speech*. Academic Press, 55-82.

Chiba T & Kajiyama M (1941). *The Vowel, its Nature and Structure.* Tokyo-Kaiseikan Publishing Company. Ltd, Tokyo.

Dang J & Honda K (1995). Local and global effects of the piriformis fossa. *J Acoust Soc Am* 98: 2931.

Fant G (1960). *Acoustic theory of speech production.* The Hague, Netherlands: Mouton, 2nd edition. 1970 (Translated into Russian, *Nauka, Moskva,* 1964).

Fant G (1975). Non-uniform vowel normalization. *STL-QPSR* 2-3/1975: 1-19.

Fant G (1978). Vowel perception and specification. *Rivista Italiana di Acustica* II, 69-87.

Fant G & Båvegård M (1997). Parametric model of the vocal tract area function: Vowels and consonants. *TMH-QPSR* 1/1997, 1-20. Originally published in *ESPRIT/BR SPEECHMAPS (6975): Delivery 28, WP2.2*, 1-30 (1995).

Goldstein UG (1980). *An articulatory model for the vocal tracts of growing children.* Ph D. dissertation, M.I.T, Cambridge, Ma.

Helmholtz HLF v. (1912). Sensations of tone. 4th edition. 1st German edition 1862. (Citation from Chiba & Kajijama, page 48).