Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 120–141
(S. Karger, Basel/New York 1965).

From the Speech Transmission Laboratory, Royal Institute of Technology (KTH)
Stockholm

Formants and Cavities

By Gunnar Fant

1. Introduction

The scene is a phonetics laboratory. The investigator has just turned off his Sona-Graph, taken off the paper, and attempts to read the spectrogram. What does he see and what does he make out of it? Obviously this depends on how well he is acquainted with the language of Visible Speech[17,9,10] and all these details in electronics, acoustics, physiology, linguistics, and psychology which provide the theoretical substance of general phonetics.

In order to establish a language of Visible Speech it would be sufficient to correlate linguistic elements with spectrographic patterns. In this task one inevitably returns to questions like, what is the origin of all these details that we have made visible and what is their auditory significance?

What I am especially interested in is the theory of speech production. With some knowledge of the underlying acoustics processes the spectrogram relates a story of the articulatory and phonatory history of the utterance. Those looking for an efficient rationale for decoding speech messages out of the acoustic structure will find useful analogies and concepts in speech production. This level of specification, although not as easily accessible, is less obscured by details than is the spectrographic pattern.

Speech production once the major object of classical phonetics now enjoys a healthy renaissance and there are many sub-levels to be considered, such as the motor command or neurological level, the dynamics of muscular movements, and the dynamics of cavity size variations. With those ambitions in mind my topic here today

is rather restricted, since I am dealing with the static aspects alone. What are the formant-cavity relations in speech?

2. The Concept of Formant

First of all let us see what is meant by the terms formants and cavities. A negativistic starting point would be to declare that the whole topic is antiquated and of academic interest only since people differ so much in their concepts of formants and experts declare that any part of the spectrum of a sound is dependent on all parts of the air filled interior of the vocal tract. What is a cavity and what is not in the continuous structures involved?

The real trouble might start already when taking spectrograms since under unfavorable conditions of high fundamental pitch the formants will be poorly defined in a harmonic spectrum and the so-called broad-band filter of the Sona-Graph will portray harmonics and not formants in these instances, see Figure 1.

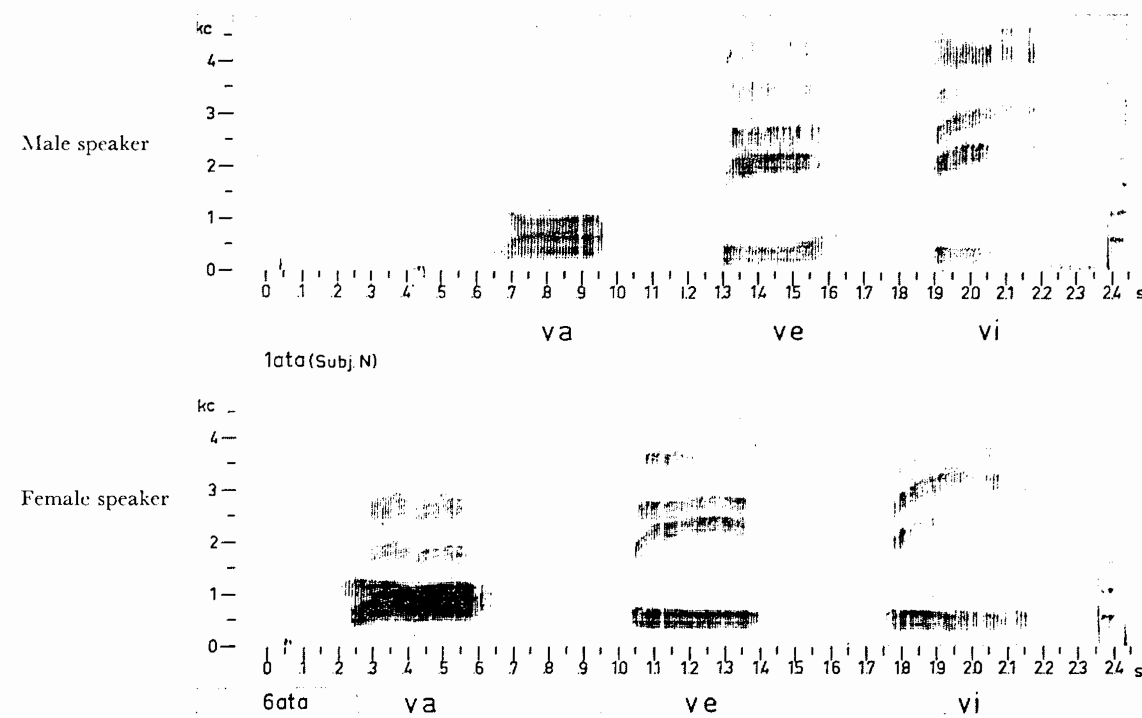

Fig. 1. Broad filter (B = 300 c/s). Sona-Graphic analysis.
A. Male voice
B. Female voice

A vowel formant of number n has the attributes of frequency F_n in c/s, bandwidth B_n in c/s[7], and amplitude level[12] L_n in dB. Of these the frequency and the bandwidth is defined as the corresponding quantities of the associated vocal tract resonance in mathematical language pole. A formant is not identical with any specific harmonic and the formant frequency is therefore not identical with the frequency of a specific harmonic. For practical purposes, dealing with male voices, we can simply and accurately measure the formant frequency as the center of the formant band of a wide-band spectrogram.

Sometimes the spectrum is obscured by two or more formants constituting a single cluster with a single peak, and a spectrum matching procedure is needed in order to decide how many vocal resonances were present and their characteristics. This view may be criticized by those preferring auditory criteria since they can claim that a formant should be defined as an auditory unit.

A single spectrum maximum should accordingly be described as a single stimulus. This point of view is indeed fruitful to follow in theories of speech perception. When studying speech structure on the acoustic level with reference to articulation, however, we need to identify formants as separate individuals even if two resonances incidently merge in a single peak. It is the demand for continuity of physical parameters intended as correlates of continuous articulatory motions and continuous perceived qualities that calls for this consistency.

To a first approximation, neglecting spectrum and voice fundamental frequency F_0, vowel qualities may be expressed in terms of formant frequencies alone, a specification in terms of three formants providing a reasonable accuracy. This type of rationale [8,16] is primarily developed for the study of non-nasalized vowels. With nasalization present it is generally possible to find out from continuity considerations what is a formant of the true vowel system and what is not. The frequencies of these formants, the so-called F-pattern, may also be traced without too great trouble in liquids and in semivowels and in voiced fricatives.

When dealing with nasal murmur sounds there are obvious difficulties and a new set of formants (constituting an N-pattern) is adopted. The established view[7,13] on the number of poles and zeros in nasal murmur segments is fit for a revision. Recent experiments performed by *O. Fujimura*[14] at the Speech Transmission Laboratory have revealed several more poles and zeros than anticipated in our previous studies.

In stops and in fricatives the F-pattern is often identifiable at least in vocalic boundary regions. The formant structure of the higher part of the spectrum, e.g. the region above 4000 c/s in fricatives could of course also be described in terms of the F-pattern. In this frequency region it is more practical to make use of larger units in terms of prominent parts of the spectrum without reference to the detailed peak structure. Spectrum matching techniques and synthesis experiments are helpful in arriving at valid approximations.

3. Vocal Cavity Structures

The procedure for investigating cavity formant relations should start with a study of what these cavities of the speech organs look like. Secondly the description must be simplified in order to arrive at a model that can be subjected to calculations. Finally certain systematic variations in the geometry of the model should be introduced in the calculations in order to find out which aspects of the structures determine a specific pattern of formant frequencies. Out of this general study there can be derived statements such as to what extent a specific formant frequency under specified articulatory or acoustic conditions will be dependent on a specific cavity or structure or any other characteristic of the compound resonator system.

What I have to say here is essentially a summary of some of the findings reported in ref.[7]*. In addition, I have samples of more recent physiological and acoustical data, especially of nasals and quite recently gained evidence of the significance of cavity wall vibrations as a factor determining formant frequencies.

The first step in any calculation is to estimate the cross-sectional area perpendicular to a central place coordinate in the vocal tract, running from the lips to the glottis. Typical examples of a tracing from a sideview of the subject and a corresponding "area function" in continuous and quantized form is shown in Figure 2. In this example the calculations were made from an area function divided in 21 successive parts of varying length and cross-sectional area.

The calculations performed on the X-ray material of Russian speech analyzed in ref.[7] were more successful than anticipated in

* Some of the material of ref.[7] is reviewed in detail in ref.[19].

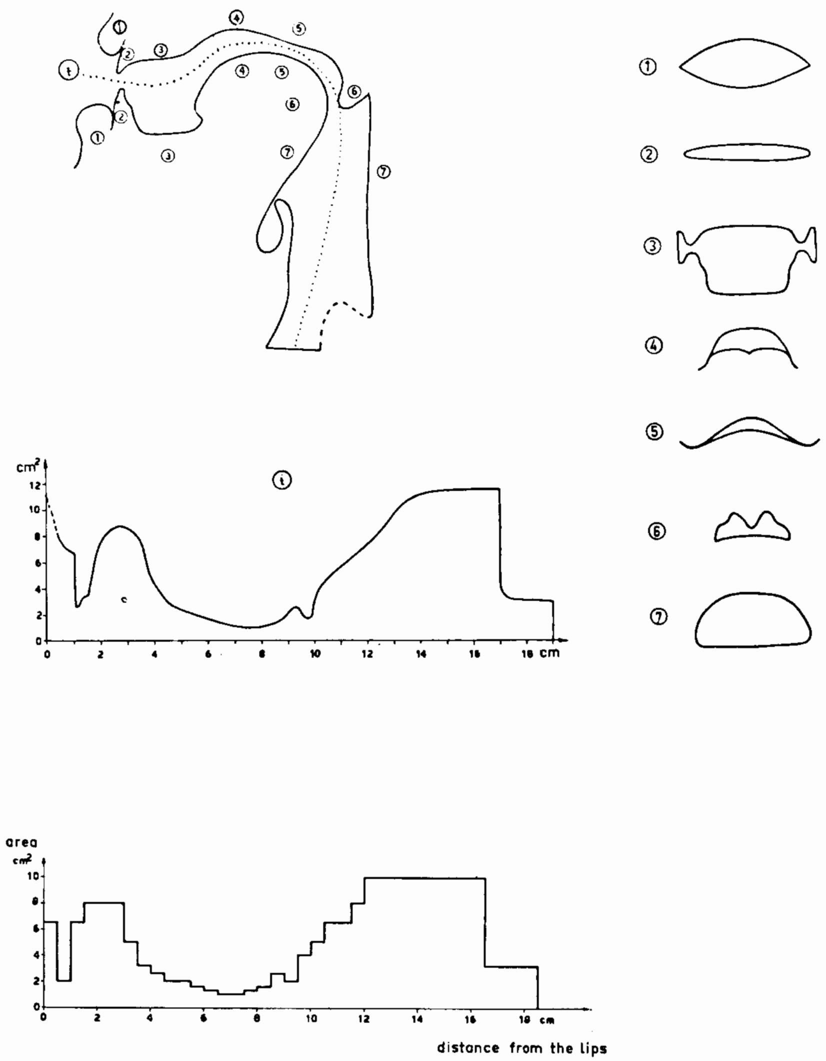

Fig. 2. The procedure for arriving at a stepwise approximation of the vocal tract area function from the X-ray data. Cross-section figures are made for a number of planes perpendicular to the central line from the glottis to the lips. A continuous outline of the area function is drawn and finally the area data within successive 0.5 cm sections are quantized according to the requirements of either the numerical calculations or those performed with LEA. (See ref.[7].)

view of the limited insight available of the lateral dimensions, i.e. from the right to left side of the subject.

The fact that the calculations came out very neatly does not guarantee that the physiological interpretations were correct in all

details although the major assumptions with respect to lateral dimensions appear to have been valid.

One obvious difficulty was to gain any insight in the structure of the nasal pathways. The calculations in ref.[7] were carried out on a nasal area function, Figure 3, derived from anatomical data* fitted to observable overall dimensions of the Russian subject.

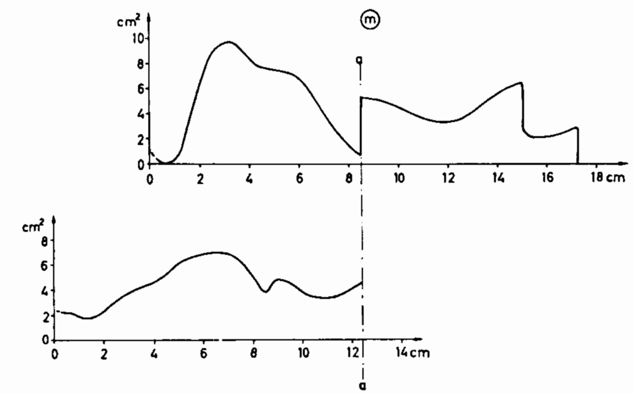

Fig. 3. Area function of complete vocal tract including nasal system (above) for the nasal consonant [m]. After *Fant*, ref.[7].

The representation of a nasal cavity system in terms of an area function alone as in Figure 3 is an oversimplification and disguises several interesting features which have acoustic relevance, as will be discussed in section 4. The primary data from perpendicular cuts of a casting of the nasal pathways, as determined by *Bjuggren*, are shown in Figure 4. The two surfaces on each side of a cut are displayed with an order scale numbered in cm along a central coordinate from the nostrils to the connection with the epipharynx. The left and right passages meet at about 8 cm from the nostrils. The symmetry is not perfect and much larger differences between the left and the right passage occur frequently. The complicated outline with an upper medium and lower channel and a very large circumference which is up to 4 times as large as that of a circular section of the same area is typical of the central parts of the nasal cavity system.

One way of visualizing cross-sections of the vocal tract is by means of so-called "laminographic" or "tomographic" X-ray

* Post-mortem anatomical study performed by Dr. *Gunnar Bjuggren*, Sabbatsbergs Sjukhus, Stockholm.

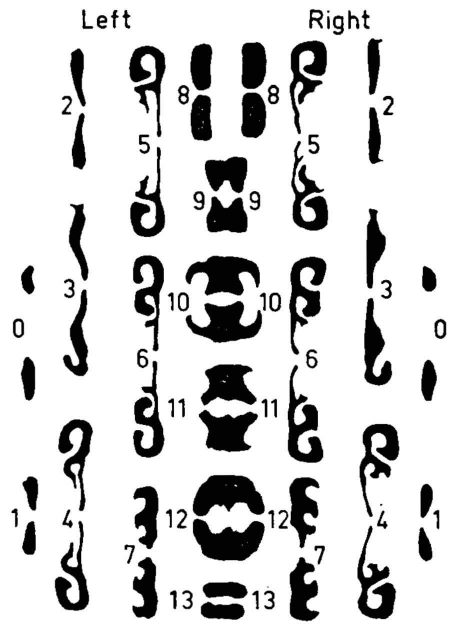

Fig. 4. Cuts through a casting of nasal cavities. Numbers pertain to distance in cm from the nostril end. Data by *G. Bjuggren.*

photos. Figure 5 is an ordinary sideview of a Swedish male subject articulating the vowel [a]. Figure 6 is a frontal tomographic view through the larynx and pharynx of the same subject articulating the vowel [i]. Figure 7 shows tomographic cuts through the middle of the mouth cavity of the same subject articulating [a] and [u]. Figure 8 pertains to cuts in a horizontal plane through the middle of the pharynx of the vowels [*a*], [u], and [i].

In these pictures made by the Swedish radiologist *P. Edholm** are several interesting facts to observe. The extreme differences in pharynx cross-sectional area comparing [i] and [a] are not only a matter of the distance between the back of the tongue and the posterior pharynx wall. The narrow pharynx passage of the vowel [*a*] is also accentuated by the short distance in the left-right direction. This general trend quoted by *Chiba* and *Kajiyama*[4] was anticipated in my work, ref.[7].

Here I would like to insert the comment that I consider not only the Swedish long vowel [a:] but also the short [a] to be a back

* Of Karolinska Sjukhuset, Stockholm. The tomograph equipment was designed by Professor *Lindblom.*

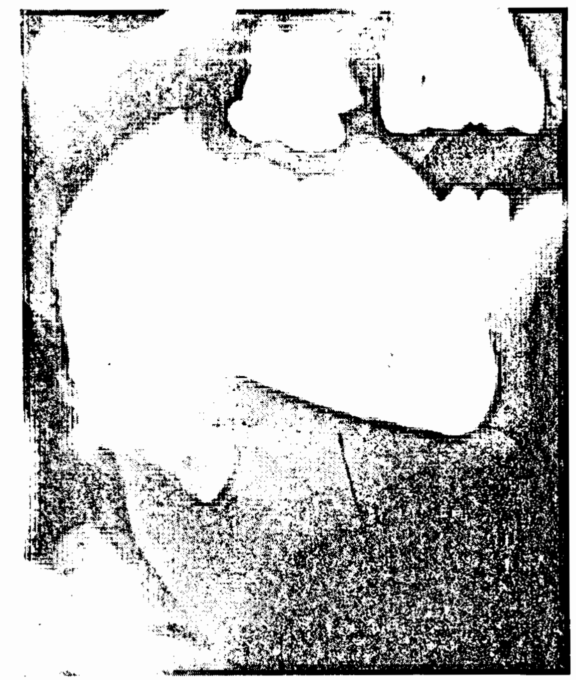

Fig. 5. Tomogram of male subject articulating the vowel [*a*]. (Photo by *P. Edholm.*)

Fig. 6. Frontal tomographic cut through the larynx and pharynx of the same subject as in Fig. 5 articulating the vowel [i].

vowel and that I find several reasons for extending this articulatory classification to other languages since the main distinctive feature of the vowel [a] compared with front vowels is the relative narrow pharynx. The highest point of the tongue, being the holy reference of classical articulatory phonetics, has no acoustic relevance in this system and is fit for revision itself. The narrowing of the pharynx on the other hand is a necessary requirement for making an [a].

The vertical cuts through the mouth of Figure 7 reveal the plastic narrowing of the tongue when elevated as in the vowel [u] leaving appreciable columns of air on both sides.

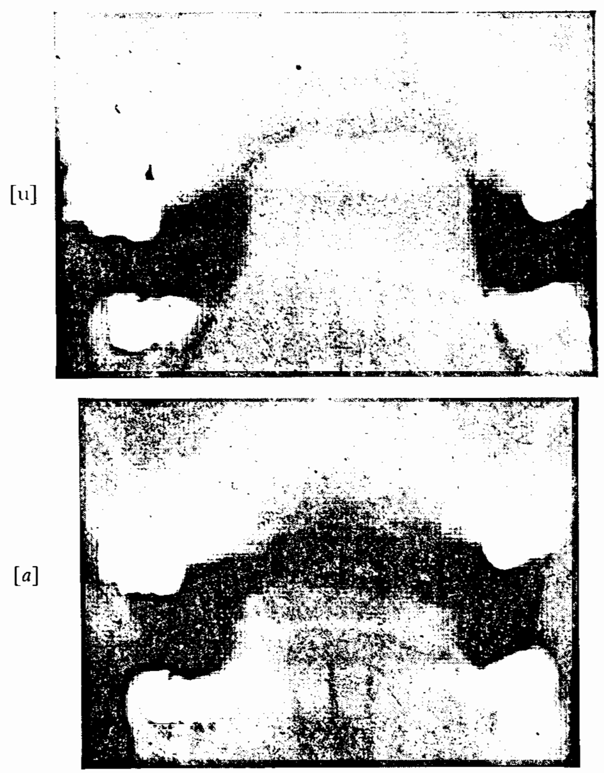

Fig. 7. Tomographic cuts through the middle of the mouth cavity of vowels [ɑ] and [u].

The frontal view of the pharynx in Figure 6 cuts through the larynx tube displaying the small Sinus Morgagni just above the vocal cords and the Sinus Piriformis side pockets on both sides of the larynx tube. These have been incorporated in some of the models submitted to calculations in ref.[7].

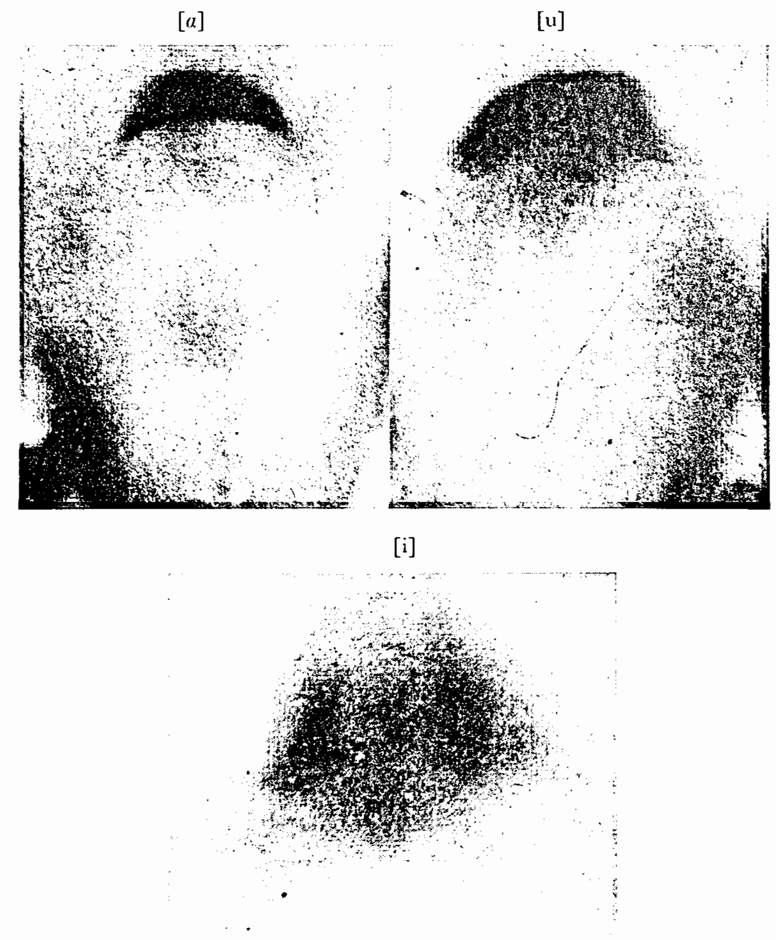

Fig. 8. Tomographic cuts horizontally through the middle of the pharynx of [ɑ], [u], and [i].

Figure 9 finally shows tracings of sideviews of articulations of Swedish vowels. Complete spectrum sections of standardized vowels selected in specific quanta of F_1 and F_2 are shown in Figure 10. These two figures have been published before[9] but are quoted here since they effectively correlate articulatory and acoustic data in terms of both formant frequency patterns and typical formant amplitude levels.

4. Vocal Tract Models

Chiba and *Kajiyama*[4] should be credited for being the first to effectively show the transmission line structure of the vocal tract.

Fig. 9. X-ray tracings of Swedish vowels arranged to conform with an F_2 versus F_1 vowel diagram as in Fig. 10.

Figure 11 taken from their book[4] illustrates the single tube representation of a neutral sound. Pressure and velocity of the sound wave inside the tube have a phase difference of 90 degrees. At the frequency of a formant there is always a pressure minimum and velocity maximum at the lips and a pressure maximum and velocity minimum at the glottis. At the frequency of the first formant the standing wave pattern occupies $^1/_4$ of a full wavelength and $^3/_4$ and $^5/_4$ wave-

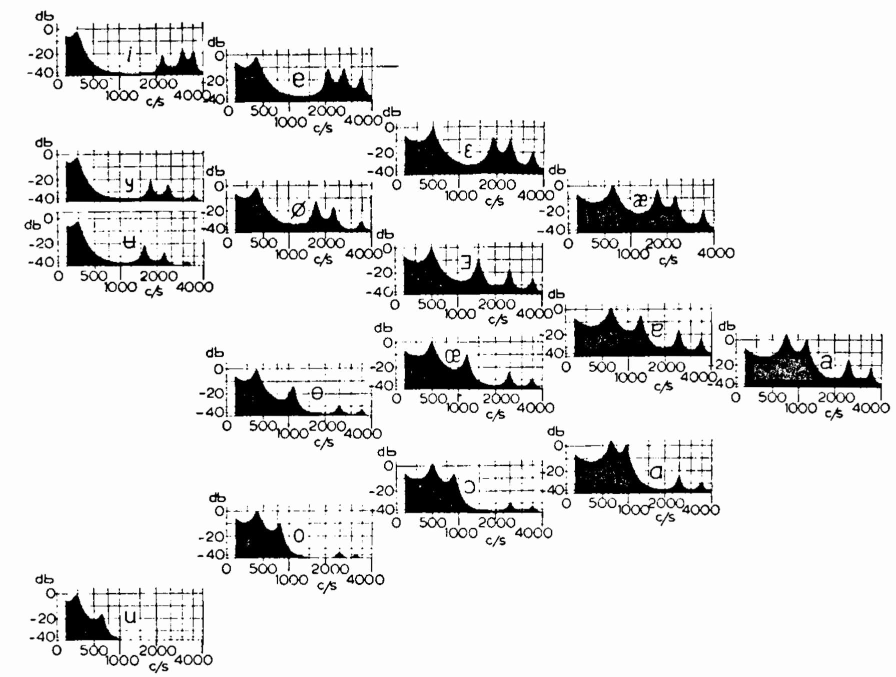

Fig. 10. Spectra of synthetic vowels assuming a standard voice source. The vowels are selected in terms of combinations of quantal values of F_1 and F_2 and F_3.

lengths at F_2 and F_3 respectively. The length of the standard tube, 17.5 cm end-corrections included, corresponds to a F-pattern of 500, 1500, 2500 c/s, etc. The most simple rule governing the relations between vocal tract configurations and F-patterns is that if a homogeneous tube is constricted at a place where one of its formants has a velocity minimum, there will follow an increase of the formant frequency whereas a constriction at a place of a velocity maximum results in a decrease of the formant frequency. These rules are related to the fact that the standing wave carries largely kinetic energy at a velocity maximum and largely potential energy at a velocity minimum, i.e. at a pressure maximum. In terms of circuit elements the region of velocity maximum may be replaced by an inductance and the region of pressure maximum by a capacitance as far as the particular formant is concerned*.

A simple but very effective approximation of the vocal tract is by means of two connected tubes of different lengths and cross-

* Calculations based on the Webster horn equation have been successfully carried out by *Ungeheuer*[25].

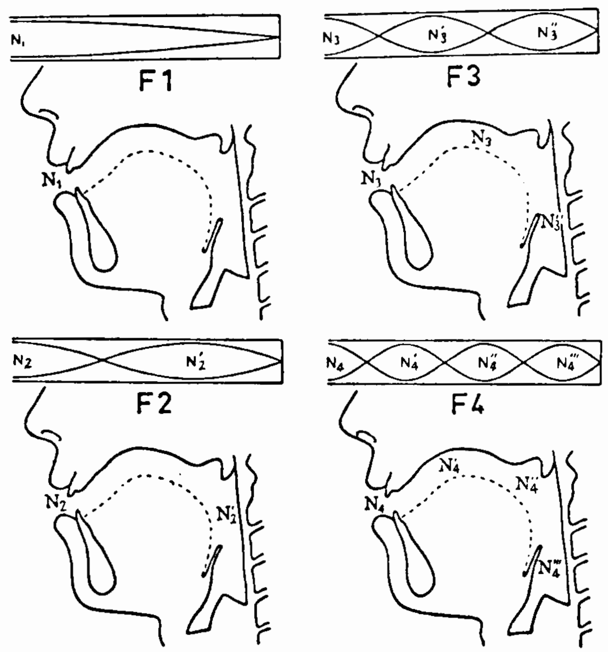

Fig. 11. Distribution of volume velocity at the frequencies of each of the first four resonances of an ideal neutral articulation in which the vocal tract simulates a tube of constant cross-sectional area (after *Chiba* and *Kajiyama*, ref.[4]).

sectional area as exemplified by Figure 12 which originates from the very early years of my work on the acoustics of speech. Of special significance is the two-tube approximation of [a] and [i] which are polar opposites in terms of articulatory features.

The double Helmholtz resonator shown in Figure 13 has played a very important role in the history of acoustic phonetics and I still use it for some approximate calculations. When the double resonator model is represented by lumped circuit elements, i.e. by the two volumes and the two necks, circuit theory will account for two formants only. As a matter of fact F_1 and F_2 of the Russian vowel [ï*] shown in Figure 2 may be calculated with a reasonable accuracy from this model. Formant frequency F_1 of most front vowels can be calculated with a reasonable degree of accuracy from a single Helmholtz resonator model of the entire vocal tract, the front part of which is the neck and the back part of which acts as the volume. In general the vocal tract behaves like a continuously inhomo-

* Nonpalatalized allophone of [i].

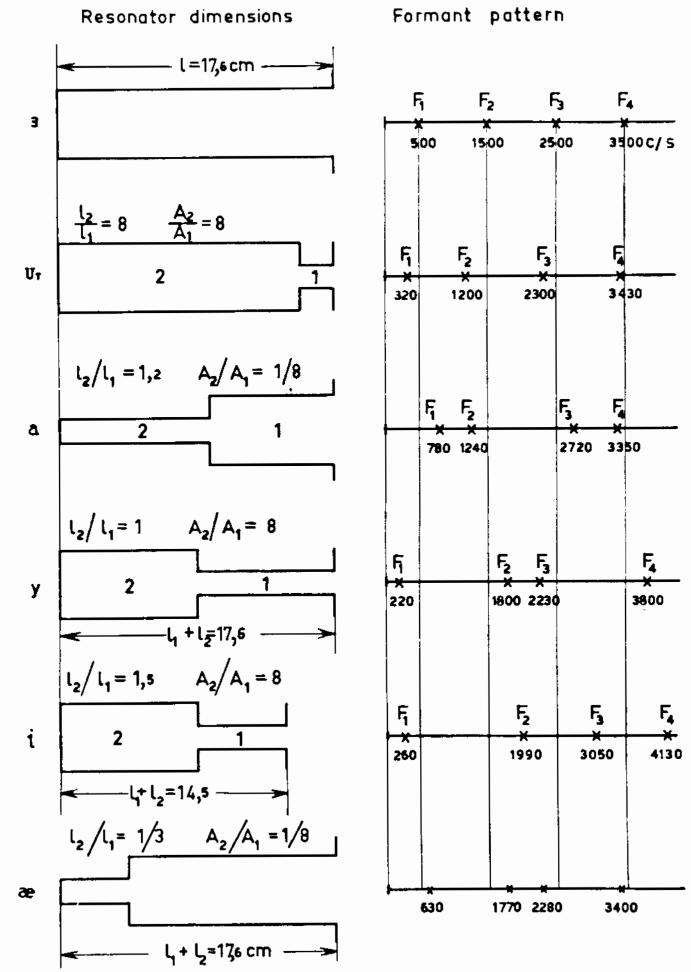

Fig. 12. Twin-tube resonators and associated F-patterns approximating a few vowels, see ref.[7]).

geneous transmission line. In order to simplify calculations *Stevens* and *House* designed a three-parameter model, whereby the complete articulation is specified by (1): the degree of opening at the lips, (2): center coordinate of the major tongue constriction, and (3): the degree of opening of the constriction.

The nomograms of Figure 15 show the relation between the place of tongue constriction and formant frequencies under conditions of a constant and small tongue constriction area and a set of varying degrees of lip-opening. This kind of nomogram is very useful for

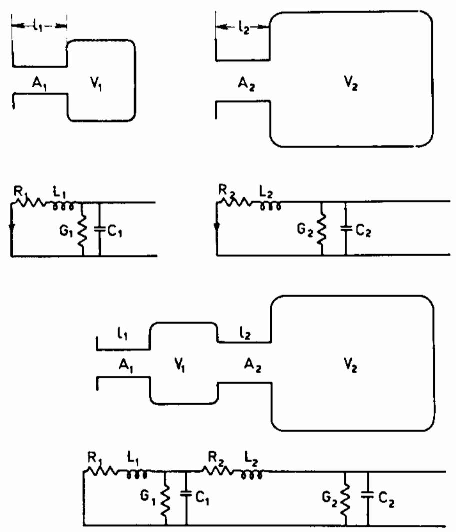

Fig. 13. Single and double Helmholtz resonators and equivalent electrical networks, see ref.[7].

translating various articulatory positions to an F-pattern. The low position of F_2 close to F_1 typical of back vowels is apparent as is the high F_2 closer to F_3 in front vowels. The place of maximum F_2 is more posterior than that of maximum F_3, the latter position representing prepalatal articulations.

5. *Formant-Cavity Affiliations*

There are several means of studying cavity formant relations. One is with reference to Figure 15. As the tongue constriction is moved from an extreme position at the glottis end of the model to the other extreme at the lips, it is to be seen how F_1 first shows a slight

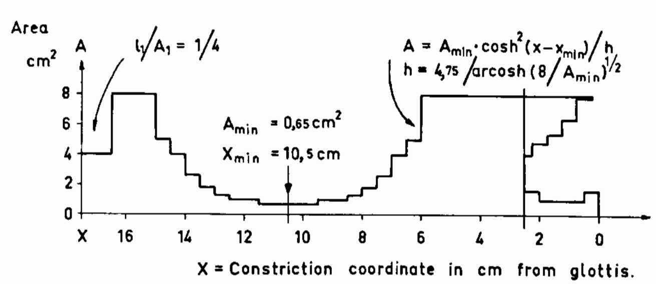

Fig. 14. Three-parameter vocal tract model based on a horn-shaped tongue section. A larynx tube, as well as the sinus piriformis cavities surrounding the larynx tube, have been incorporated as fixed cavities, see ref.[7].

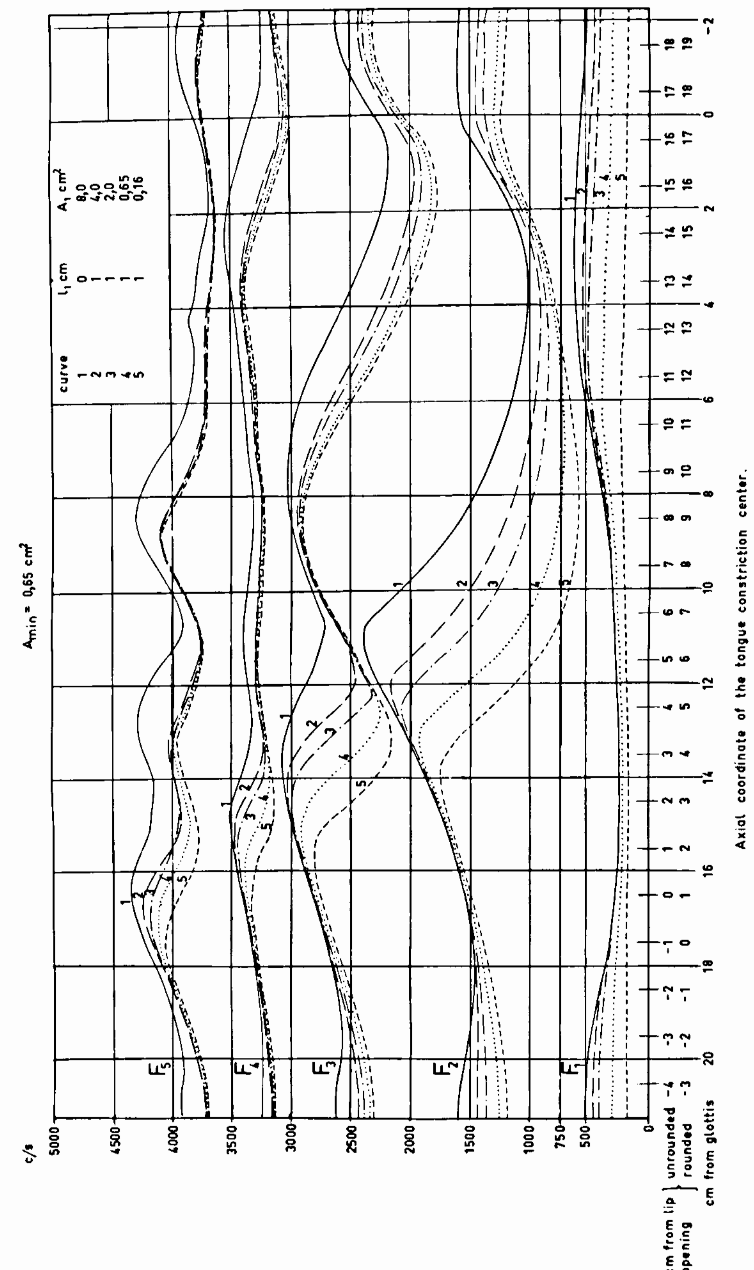

Fig. 15. Nomograms of the first five formant frequencies as a function of varying place of the major tongue constriction assuming a very narrow tongue passage and with various degrees of superimposed lip-rounding in the model of Fig. 14. (See *Fant*, ref.[7].)

tendency of rise and then falls. Here F_1 shifts from a mixed front and back cavity affiliation to being more affected by the back cavity*. In this forward movement F_2 starts with a fall and then rises as it gets more affiliated with the front cavity than with the back cavity.

However, at an advanced place of articulation in front of the place of maximum F_2 the second formant will be more dependent on the back cavity, now as a half-wavelength resonance. In this advanced location F_3 is the fundamental resonance of the front cavity. Further back at a medio-palatal to velar region F_3 is associated with a standing wave resonance of the back cavity. In back-vowels F_3 is a standing wave resonance of the front cavity.

The effect of superimposed lip-rounding on formant cavity affiliations is to shift the critical coordinates in an anterior direction, as can be seen from Figure 15. On the other hand if a lip-rounding affects one formant more than others this is a sign of an especially prominent affiliation of this formant with the front cavity. This situation is typical of the prepalatal [i] the F_3 of which is very sensitive to lip-rounding.

Quantitative measures of the degree of dependency of any formant frequency on any small variation of the vocal tract area function were given in ref.[7]. One of the several variations introduced was to make a small increase in the cross-sectional area at the coordinates of maximum area in the front and the back cavity. The tabulation is quated here, * denotes negative values.

Vowel	$\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_2}{F_2} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_1}{F_1} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_2}{F_2} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_3}{F_3} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_3}{F_3} \frac{C_2}{\Delta C_2}$
[ɑ]	0.07	0.19	0.23	0.11	0.18	0
[o]	0.05	0.33	0.37	0.22	0.25	0
[u]	0.18	0.28	0.20	0.15	0.06	0
[ï]	0.02*	0.39	0.49	0.02*	0.04	0.45
[i]	0.01*	0	0.53	0.39*	0.04	0
[e]	0.08*	0.01*	0.42	0.26*	0.01	0.23

Observe e.g. the lack of association of F_2 of [i] with the front cavity volume C_1. The same effect is observed when merely shortening the front cavity of [i] by removing one 0.5 cm section, see ref.[7], p. 120,

* This discussion pertains to the role of cavities alone. In half-open and close front vowels the tongue constriction is of equal importance as one of the basic elements of a Helmholtz resonator.

Table 2.33–4. Whilst discussing the vowel [i] it should be appreciated that what is said above pertains to the Russian and the Scandinavian [i]-vowels which are prepalatal whereas the [i] is articulated more towards the medio-palatal region in English. Also the age and sex of the speaker should be considered. Women generally have shorter necks than men, i.e. their ratio of mouth length to pharynx length is greater than in males which determines an increased affiliation of their F_2 with the front cavity.

The fourth or fifth formants of a male voice have a tendency to be strongly affected by the larynx tube. Even without such a tube there would fall a vocal tract resonance not too far away but the essential feature is that the addition of a larynx tube to a model, everything else being constant, is to place one more formant in the 3500 c/s region thus boosting this part of the spectrum. Those who have a short larynx tube or insufficient closure at the arytenoids will lack this formant. This is typical of female voices.

The effect of the Sinus Morgagni pockets is to cause a slight shift down in some of the formant frequencies and to set a sharp upper frequency limited at about 4500 c/s in the vowel spectrum, see ref.[7], p. 102, and ref.[3].

The search for formant cavity associations once approached with unsufficient theoretical tools and an overoptimistic attitude[20, 21, 15] was miscredited by the impact of transmission line theory in vocal tract analysis. This subject may now be treated with better theoretical tools, although our knowledge of the physiological facts has not advanced to the same extent. In fact, we still may have something to learn of the "old school". *Sovijärvi's*[20, 21] ensemble of seven variable and eleven fixed formants of the speech spectrum was in part a fiction but some of these speculations deserve a renewed interest now when *Fujimura* has shown[14] that there are at least twice as many formants in a nasal murmur sound than we had anticipated from spectrum matching work[13] and the averaged area function of the nasal tract I used in my calculations[7]. A sample of *Fujimura's* observations[14] pertaining to transmission measurements with sound injected from a vibrator externally at the throat and a pick up of the sound at the nostrils is shown in Figure 16. Observe, for instance, the split first formant as in nasalized sounds which together with the several other spectral peaks can be ascribed to the asymmetry of the particular nose, the left and right passages differing in their dimensions.

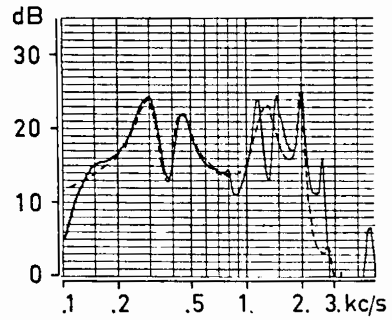

Fig. 16. Sine wave sweep frequency characteristics of the cavity system of a palatal nasal consonant co-articulated with a neutral vowel (solid line). Vibrator externally at the throat and pickup at the nostrils. A spectrum matching comprising some of the lower resonances is included (broken line). In an accurate match something like 6 poles and 3 zeros would be needed in the frequency range below 2000 c/s. (After *O. Fujimura.*)

Finally a statement on the role of the cavity walls. It was initially suggested by *van den Berg*[1,2] and I elaborated the theme in my book[7], that in sound with a very low F_1 there is a substantial energy loss from sound transmission through the walls of the vocal tract, wherever these walls are thin enough to allow vibration to be set up. In ref.[7] I calculated a limiting frequency of 150 c/s for the air inside the vocal tract resonating with the mass element of the walls all outlets being closed. Recently I discovered that this shunting mechanism is responsible for the typical nasal quality of the speech of a diver submitted to a high overpressure.

Recent experiments performed together with Dr. *B. Sonesson** have shown that the most typical characteristics of the speech of a person submitted to a high overpressure is the rise in F_1, the immobilization of F_1, and the inability of the subject to produce sounds with an F_1 lower than a certain limit set by the overpressure. What happens is that an increase in overpressure does not affect the velocity of sound, c, but it affects the density ρ of the air proportionally and thus the impedance level $\rho c/A$ whereas the impedance of the cavity walls is not affected by the pressure increase and its relative load will thus increase.

The lowest possible frequency of F_1 is simply the square root of the atmospheric pressure in ATA times a constant which is the limiting F_1 of normal speech and of the order of 200 c/s. With a pressure of 6 ATA the low limit of F_1 is thus of the order of 500 c/s.

* Will be reported on briefly in STL-QPSR 2/1964.

The typical effect is that close and half-open vowels, as well as voiced occlusives, fricatives, and semivowels and nasals, all fall within the same very narrow range of F_1 around 500–600 c/s. To the first approximation

$$F_1 = \sqrt{F_{10}^2 + F_{1k}^2}$$

where F_{10} is the F_1 without consideration to the cavity wall load and F_{1k} the limiting frequency at complete vocal tract closure.

Examples of speech at 1 and 6 ATA pressure is shown in Figure 17. This effect facilitates the calculation of the limiting F_1 in normal speech and enforces the existence of this limit to be considered when calculating the filtering properties of the vocal tract. If this effect is not taken into account calculations of F_1 will thus tend to give too low values, a fact that is supported by experience. But for this effect it would be difficult to explain the very stable base-band formant frequency at about 200 c/s of voiced consonants. It remains to study to what extent the low F_1 limit varies with speaker category.

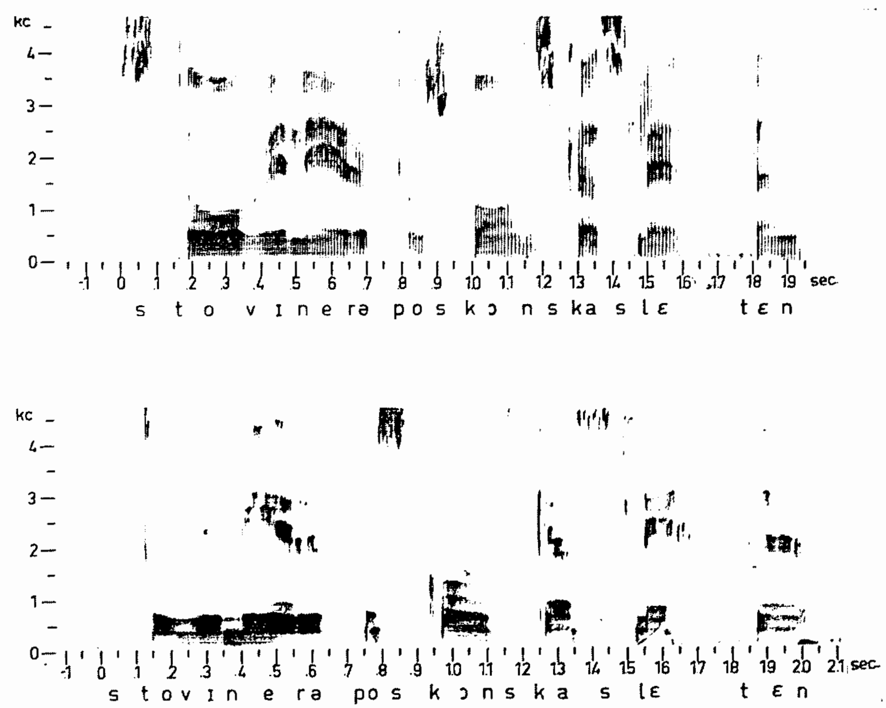

Fig. 17. The main effect on speech of increasing the air pressure in the room, where the subject is seated, is to increase the lowest limit of F_1.

References

1. *van den Berg, Jw.:* Physica van de stemvorming met toepassingen; Diss. Groningen ('s-Gravenhage 1953).
2. *van den Berg, Jw.:* Calculations on a model of the vocal tract for vowel /i/ (meat) and on the larynx. J. acoust. Soc. Amer. *27:* 332–338 (1955).
3. *van den Berg, Jw.:* On the role of the laryngeal ventricle in voice production. Folia phoniat. *7:* 57–69 (1955).
4. *Chiba, T.* and *Kajiyama, M.:* The vowel – its nature and structure (Tokyo 1941).
5. *Delattre, P.:* The physiological interpretation of sound spectrograms. PMLA, *LXVI:* 864–875 (1951).
6. *Dunn, H. K.:* The calculation of vowel resonances, and an electrical vocal tract, J. acoust. Soc. Amer. *22:* 740–753 (1950).
7. *Fant, G.:* Acoustic theory of speech production ('s-Gravenhage 1960).
8. *Fant, G.:* Acoustic analysis and synthesis of speech with applications to Swedish. Ericsson Technics *15:* 3–108 (1959).
9. *Fant, G.:* The acoustic of speech. Proc. Third int. Congr. Acoustics, Stuttgart 1959; pp. 188–201 (Amsterdam 1961).
10. *Fant, G.:* Descriptive analysis of the acoustic aspects of speech. Logos *5:* 3–17 (1962).
11. *Fant, G.:* Sound spectrography. Proc. Fourth int. Congr. Phonetic Sciences, Helsinki 1961; pp. 14–33 ('s-Gravenhage 1962).
12. *Fant, G.; Fintoft, K.; Liljencrants, J.; Lindblom, B.* and *Mártony, J.:* Formant-amplitude measurements. J. acoust. Soc. Amer. *35:* 1753–1761 (1963).
13. *Fujimura, O.:* Analysis of nasal consonants. J. acoust. Soc. Amer. *34:* 1865–1875 (1962).
14. *Fujimura, O.* and *Lindqvist, J.:* The sinewave response on the vocal tract. Speech Transmission Lab., quart. Progr. and Status Report No. 1, pp. 5–10 (KTH, Stockholm 1964).
15. *Hála, B.:* Nature acoustique des voyelles. Acta Univ. Carol. Med. (Praha 1956).
16. *Peterson, G. E.* and *Barney, H. L.:* Control methods used in a study of vowels. J. acoust. Soc. Amer. *24:* 175–184 (1952).
17. *Potter, R. K.; Kopp, A. G.* and *Green, H. C.:* Visible speech (New York 1947).
18. *Rayleigh, L.:* The theory of sound. Vols. I and II (New York 1945). (First Amer. Ed.)
19. *Sapozhkov, M. A.:* The voice signal in cybernetics and communications (Moscow 1963). (This book provides an inventory of speech research with a large number of references.)
20. *Sovijärvi, A.:* Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache. Ann. Acad. Sci. fenn. B XLIV (Helsinki 1938).
21. *Sovijärvi, A.:* Die wechselnden und festen Formanten der Vokale, erklärt durch Spektrogramme und Röntgenogramme der finnischen Vokale. Proc Third int. Congr. Phonetic Sciences, Ghent 1938; pp. 407–420 (Helsinki 1938).
22. *Stevens, K. N.* and *House, A. S.:* Development of a quantitative description of vowel articulation. J. acoust. Soc. Amer. *27:* 484–493 (1955).
23. *Stevens, K. N.* and *House, A. S.:* Studies of formant transitions using a vocal tract analog. J. acoust. Soc. Amer. *28:* 578–585 (1956).
24. *Trendelenburg, F.:* Einführung in die Akustik (Berlin 1950). (Zweite Auflage.)
25. *Ungeheuer, G.:* Elemente einer akustischen Theorie der Vokalartikulation (Berlin 1962).

Author's address: Dr. G. Fant, Speech Transmission Laboratory, Royal Institute of Technology (KTH), *Stockholm 70* (Sweden).

Discussion

H. M. Truby (San Jose, Calif.) (*Truby's* summary of his own critique): In spite of the overwhelming effects of the past 55 min of slides, I feel I must utilize some part of the few remaining discussion moments to point up certain pertinencies. The title selected by Docent *Fant, Formants and Cavities*, would lead members of the Congress to assume that something of the *relationship* of formants and cavities was to be discussed, as might well be the case after all the years of research applied to the establishment of such implied correlates. However, we have not been so blessed.

I hold before you a picture of a little girl. This picture tells a complex story to me which no one else in the hall could possibly be in a position to appreciate. This picture is a single instant in time, yet when *I* look at this picture an entire dynamic tale unfolds. The little girl of the picture is my daughter, and the picture acts as a multiple stimulus, the nature of which you are all aware of by this point in my analogy, for I've seen her happy and sad and at work and at play and asleep, and so on. Now, when we see slide after slide of such single-instants-in-time, it is only as products of our own individual experiences that we are able to evaluate their static natures in dynamic terms. Certainly *formants* do not reveal themselves entirely – and ofttimes not even identifiably – on sound spectrograms or similar displays, from the standpoint of their per*ception*, and *cavities* cannot be defined in terms of cross-sectional areas unless they happen to be highly idealized cavities, such as are *never* found in the human or any other animal form... and rarely in nature at all, except as a man-made artifact. How then can it be proposed that a discussion of a few arbitrarily selected instants in time extracted from visual displays of acoustic continua will reveal anything of the dynamic nature of the speech signal? We know as little now, apparently, of the relationship of formants to cavities and vice versa as was reported during the Oslo Congress of Linguists in 1957, which situation I cticicized expressly as *A First Feature of Indivisibility* in my own Oslo paper, *Visible and Indivisible Speech.* I say, 'apparently', since in the twelve minutes actually allotted by the Committee for the pre-discussion phase of this paper, there *would* have been time to report on the dynamic analysis aspects *now* available to speech research. Cineradiography, spectral timesampling, overall intensity dynamics, fundamental frequency flow – all these and other media for analysis are discussable at this point of the history of the phonetic sciences. Having been exposed to the implications of 'hocus pocus linguistics', in the face of what we are asked to infer from this array of time and space *cross-sections* just presented us, I should like to propose the adjunctive term 'hocus pocus phonetics'. In short, calculations in static terms of dynamic – yea, kinematic – activities have little significance for the phonetician. The high-light of my long-time colleague's presentation was the single slide of the F_1 through F_5 nomogram which, if not dynamic, was at least spatial, and which is featured in his doctoral dissertation.

I cannot hope that my remarks will not be misconstrued in some quarters and my point overlooked, but my principle intention is to point out that the present state of the art *does* permit a dramatic account of the acoustic and physiologic correlates of speech. It's there for the telling...

Epilogue

The above Discussion was, of course, motivated by the *delivered* version of *Fant's* paper. The *written* version, carefully edited by the author after the Congress, has been altered in several places significantly relevant to certain points of my verbal critique, e.g., last line, p. 120 – first line, p. 121; lines 3–8, p. 121; etc. Even though, in this instance, my own remarks have lost none of their pertinence, presented papers should not, in my opinion, be mere rough drafts of *Proceedings* reports.