Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Historical notes

Fant, G.

journal: TMH-QPSR
volume: 47
number: 1
year: 2005
pages: 009-019

Historical notes*

Gunnar Fant

Response to interview questions posed by Louis-Jean Boe and Pierre Badin in 1996. The headings pertain to their questions. Minor editorial revisions 2005.

1. Education and background, first studies

I started my studies at the Royal Institute of Technology, KTH, in 1938 and graduated with a degree in Electrical Engineering in 1945. My studies were interrupted by three years of military service of which two as a reserve officer in the Swedish Coast Artillery.

An intermediate degree of licentiate was received in 1952 with a thesis on "The heterodyne filter" which allowed continuously variable band-pass filtering with very steep filter skirts.

Following Swedish tradition my main doctoral degree was received as late as in 1958 when I was already well established in the field. The thesis defense is of some historical interest. It has been reviewed by James Flanagan (1979) who served as the main faculty opponent. My second opponent was Eli Fischer-Jörgensen and the third opponent was the British pioneer in speech synthesis, Walter Lawrence, who during the ceremony demonstrated his PAT, (Lawrence, 1953) in a dialogue with my OVE I.

My mentor at the KTH was the professor of Telegraphy and Telephony, Torbern Laurent, a specialist in filter and transmission line theory. As a master thesis topic for me in 1945 he suggested studies of the relation between intelligibility and speech bandwidth reduction, theoretically and with special application to various types of hearing loss. My main references were the early works of Fletcher (1929) and Collard (1930).

The thesis work paved the way to an employment at the Ericsson Telephone Company 1945-1949 for investigating the frequency domain characteristics of Swedish speech sounds, which was motivated by the need to understand the consequences of reserving a part of the telephone network audio band to tone signaling.

I was given free hands to develop my own instrumentation. I constructed a wave-analyzer for frequency analysis of sustained sounds, recorded on a galvanometric oscillograph. I also produced time-frequency-intensity spectrograms, by a manual assembly of amplitude-time curves from a number of successive band-pass regions. For this work, I had access to one of the first magnetic tape recorders of that period. My broadband spectrograms with black-white binary coding of spectral intensity (Fig. 1), were quite similar to those I had just learned about from the early publications of Visible Speech at Bell Laboratories (Potter et al., 1947). I had access to an an-echoic chamber and a high quality condenser microphone. Absolute amplitude calibrations of sound pressure levels at a specified distance were made in all sessions. The detailed data of formant structure of vowels as well as of consonants and the influence of voice effort were documented in internal reports (Fant, 1948, 1949), but the main results were not published until 1959 together with the theory of cascade formant synthesis (Fant, 1959). The vowel data compared well with those of Peterson & Barney (1952).

During 1947-1949 I had part time employment at my old department at KTH, planning for future studies of speech and language structure and systems evaluation. At that time one of my projects was the development of a speech audiometer. My speech data from Ericsson was introduced in the frame of an audiogram to visualize how much of the

* The later period from 1997 is not covered in the overview. More recent accounts are given in:
Fant G (2004). *Speech Acoustics and Phonetics. Selected Writings*. Kluwer Academic Publisher/Springer
Fant G (2004). Speech research in a historical perspective. In: Slifka J & Manuel S (eds) *From Sound to Sense: Fifty Years of Discoveries is Speech Communication*. MIT conference June 11-13, 2004

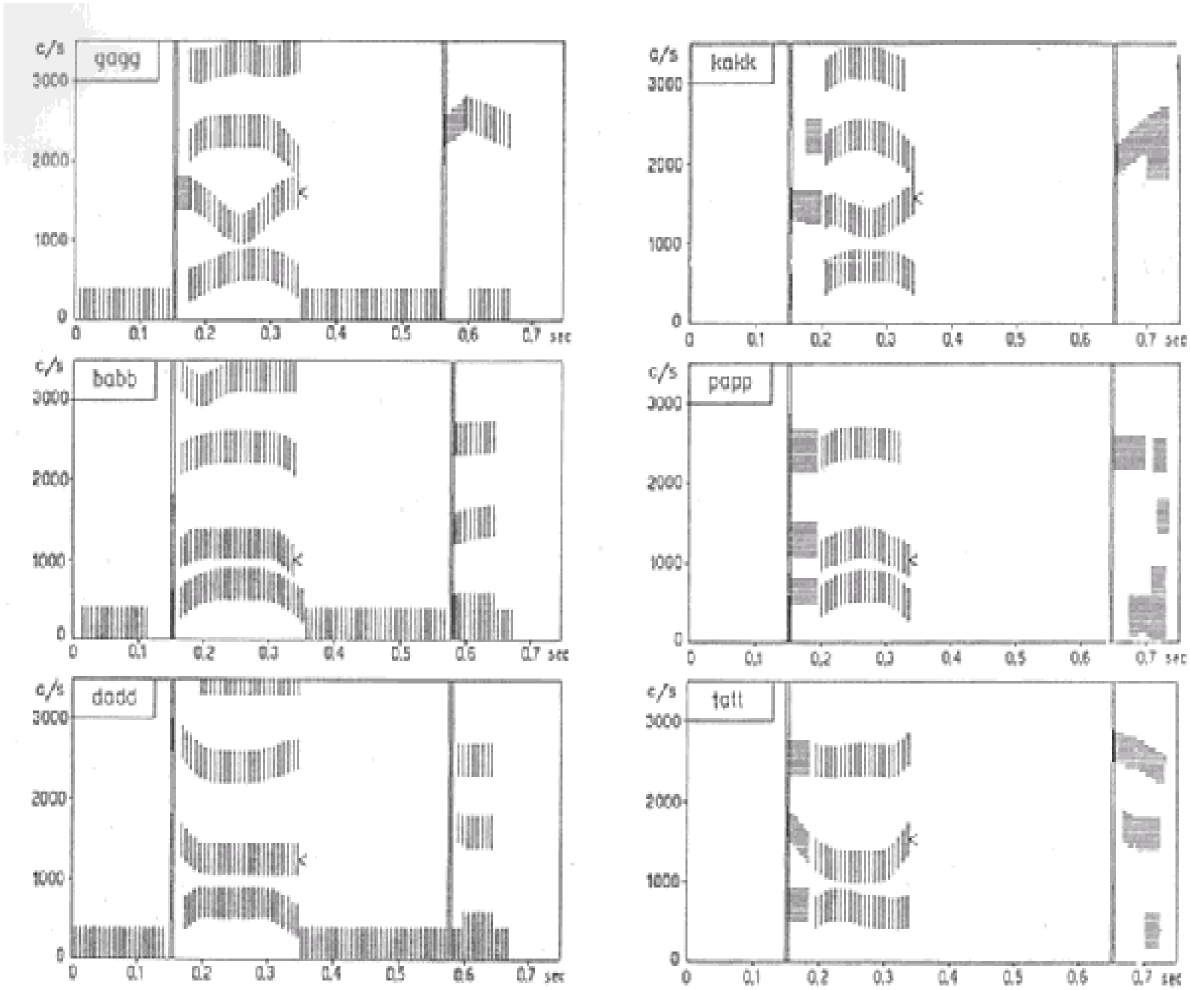


Figure 1. Stylized spectrograms. From Fant (1949, 1959).

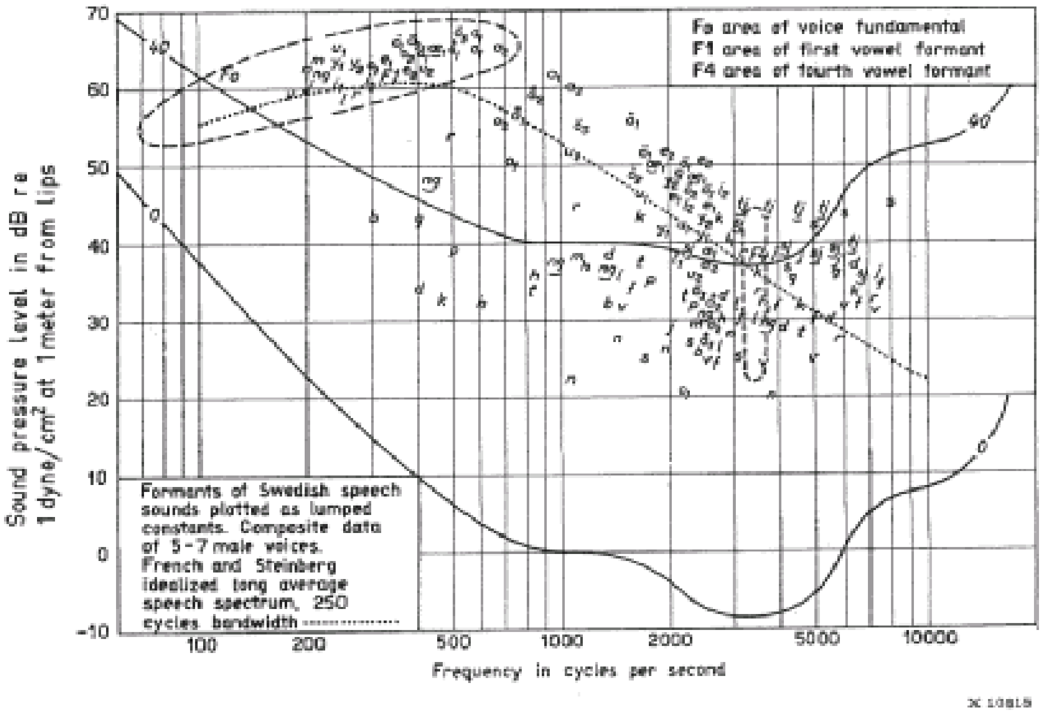
speech spectrum could be perceived considering a certain hearing loss and speaking distance. The overall spectral area of formant regions (Fig. 2), was eventually adopted as the "formant banana" in audiology practice (Fant, 1995b). I was also involved in the auditory training of deaf children.

Experiments on vowel synthesis with formant circuits in cascade were initiated at the KTH in 1949. In 1952, when I came back from the USA, the one-hand control system of OVE I was added (Fant, 1953). It enables a continuous variation of vocalic sounds in an F2 versus F1 field, with F0 varied by a rotation of a hand control. A possible application conceived of was as a speech prothesis, but with the advent of text-to-speech synthesis this idea became outdated. However, the system has had an appreciable pedagogical value. A later version developed by Johan Liljencrants is shown in Chistovich et al. (1966b).

In Stockholm, in 1948 I met with the famous acoustician Leo Beranek, who arranged for my stay at the acoustics laboratory of MIT 1949-1951. During the first months I was part time attached to the Psycho-Acoustics Laboratory of Harvard University where I followed ongoing work on auditory functions and speech perception. Here I met S. S. Stevens and other prominent psychologists such as J. C. R. Licklider and George Miller.

2. Preliminaries to speech analysis

It was here at a seminar I gave, that I first met Roman Jakobson. I reviewed my speech analysis work at Ericsson 1946-1949 and made some comments on the relation between vowel and consonant spectra that fitted into his ideas about distinctive features. I could thus add some substance to his concept of the features compact



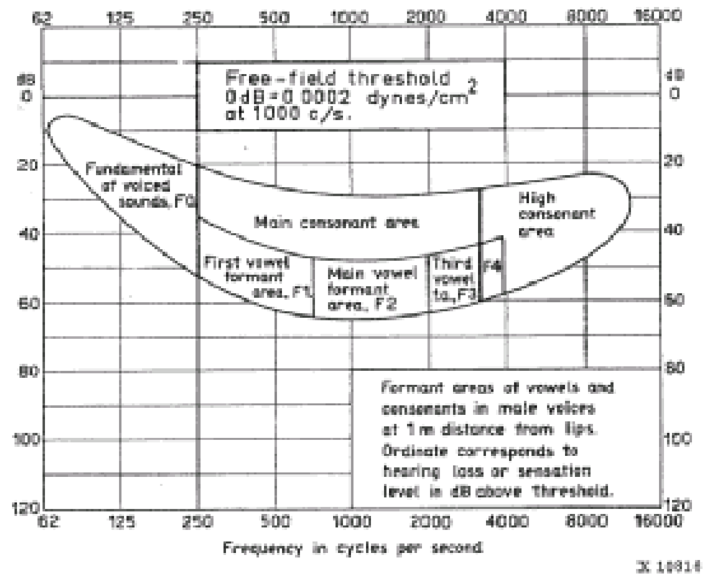


Figure 2. Summary of formant regions within the frame of an audiogram (in audiology referred to as the "speech banana"). From Fant (1959).

versus diffuse and grave versus acute, as realized in both vowels and stop consonants. The common denominator of compactness versus diffuseness is the central concentration versus spread of spectral energy. The burst spectra* of unvoiced stops in final release that I showed him are reproduced in Figure 3.

Morris Halle joined us in the work on an integrated theory of distinctive features, (Jakobson et al., 1952). It developed into a joint venture where Roman posed the questions, I acted as a scientific medium suggesting feature correlates, and Morris was the secretary. Discussions were lively.

Before meeting with Roman Jakobson I had been exposed to only a moderate amount of traditional phonetics, and I had not attended any courses in linguistics. As a result of our cooperation, I now gained a fundamental insight in the relation of phonology to phonetics and how to structure my own thinking about speech research. The basic concept is that of the speech code and its realization in successive levels of the speech chain. In a narrow sense, the speech code is the relation of phonology to phonetics. In a broader sense, it is the relation of any unit or category on the message level to observable aspects in production, in the speech wave and in perception.

Preliminaries aimed at a minimal redundancy encoding of phonological contrasts into binary features. Following Jakobson's basic approach, our system grouped together traditional manner and place features in consonants with acoustically and perceptually similar features in vowels, which called for a definition of production level categories in terms of cavity shapes rather than by articulatory patterns as was later adopted by Chomsky & Halle (1968). The latter system, which is production oriented, has some advantage in generative grammar and phonetics but lacks the consistent handling of perceptual and acoustical dimensions attempted in Preliminaries.

The choice of features can always be disputed and both systems have their specific strength and weakness, see Fant (1973) for a review. The Jakobson-Fant-Halle features are not sufficient for describing all the languages of the world and some feature combinations may seem rather abstract. The same can be said about the Chomsky-Halle features. To create a language universal feature system requires a more detailed analysis of speech production including exotic languages. But this could be a feasible task. Our speech organs do not have unlimited resources.

The basic issue of invariance is well founded in Roman Jakobson's philosophy. Invariance of feature correlates can, but need not be absolute. The great variability of phonetic manifestations, which has plagued speech analysis and has caused so much discussion is overcome by accepting the notion of *relational invariance*. The way out of the dilemma is to structure the sources of variability with respect to all contextual, dialectal and speaker specific factors as well as paralinguistic features. This is the ultimate object of the search for the *speech code* as it has developed in my thinking. Within a defined context, disregarding deletions and substitutions, the phonetic essence of a distinction remains, *ceteris paribus*, the same.

The main importance of Preliminaries has been to add a structure to general phonetics, decomposing phoneme inventories into features as information bearing elements. Details of the system may be criticized but it has promoted the concept of the speech code and its nature. The system has influenced methodology in assessing hearing loss and speech defects and it has been of theoretical support for automatic speech recognition. The Chomsky-Halle feature system has been more suited for text-to-speech synthesis.

3. Acoustic theory of speech production

At MIT I gained much from ongoing work on acoustics, circuit theory, transform theory and information theory. Here I started my work on the acoustics of speech production. I had close contacts with Ken Stevens, then a graduate student, a relation that has remained intact

---

* Figure 3 produced at the Ericsson Telephone Company in Stockholm 1949 is unique in several aspects. The observed male/female differences in F2 and F3 peak locations in the aspiration parts of the burst spectra of [p] and [t] reflect scale factors of the order of 10-20% which agree with those later established for vowels (Fant, 1975). The male and female [k] spectra are quite similar and show the same location of the main peak. The acoustical determinant is the length of the mouth cavity in front of the constriction, which is about the same for males and females. Moreover, the absolute intensity calibration preserves relative spectral levels. Comparable data from later studies do not exist.

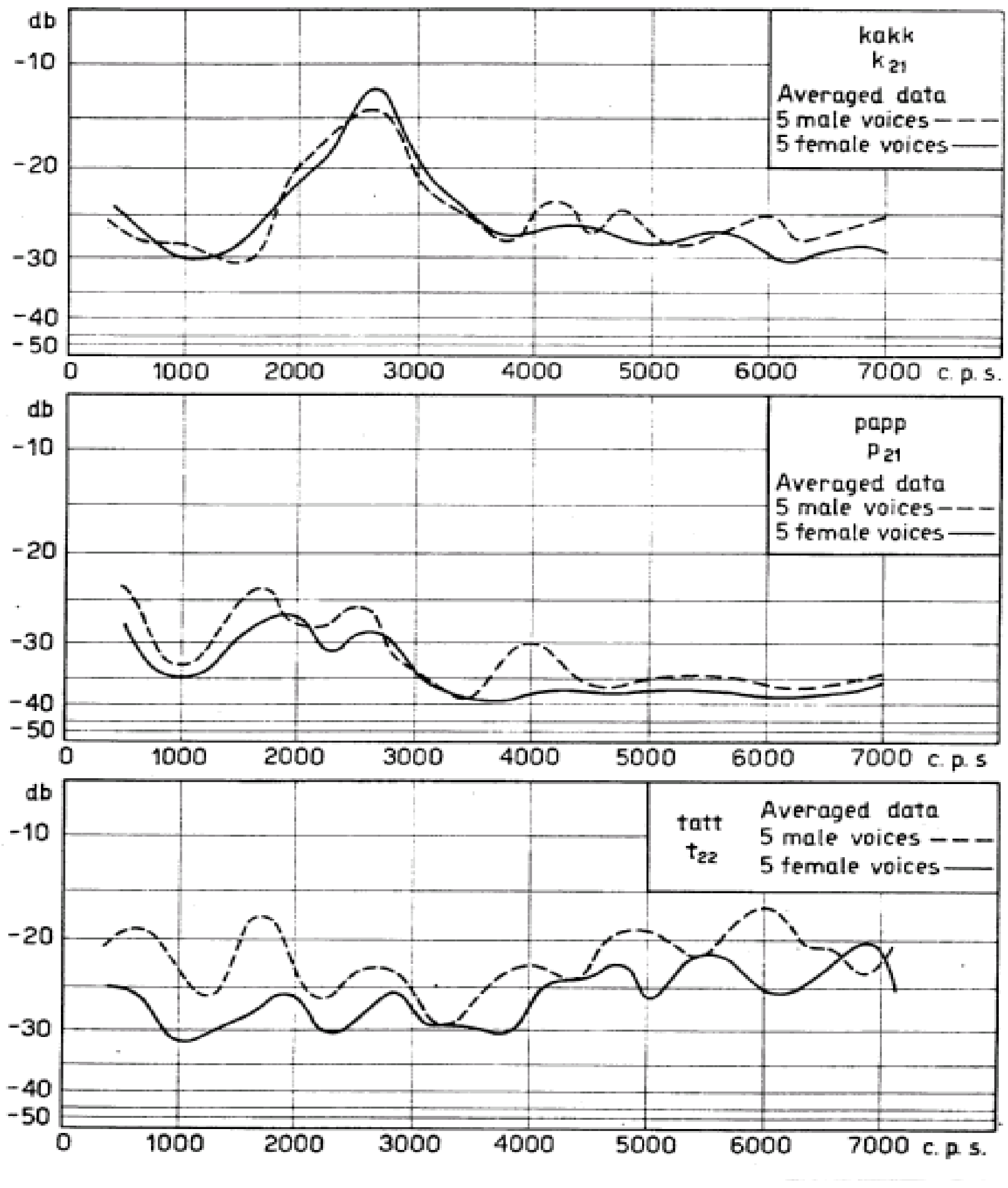


Figure 3. Burst spectra of final [k] [p] [t]. Average data for 5 males and 5 females. From Fant (1949, 1959)

throughout our lives, in virtue of common interests and scientific approach.

In cooperation with Roman Jakobson and Morris Halle I pursued X-ray studies of sustained articulations of a Russian subject which gave me material for my book (Fant, 1960). The incitement was obvious. Only by a profound insight in the articulatory and phonatary processes and their acoustic consequences can we gain an understanding of the speech code. Articulatory interpretation of spectrograms is the key for the language of Visible Speech.

The most important development in the history of speech research is undoubted the sound spectrograph developed at Bell Laboratories (Potter et al., 1947). The time-frequency-intensity patterns of Visible Speech provided a new insight in human speech which has remained our most valuable resource for objective studies. The contrast between the discrete units of phonology and the continuity of speech became evident. An American linguist, Charles Hockett (1955) phrased his view of spectrograms patterns as "what you could expect from smashing eggs onto a moving belt". An early and excellent work on acoustic phonetics was that of Martin Joos (1948).

To this period also belongs the Pattern Play Back of Haskins Laboratories (Cooper et al., 1952). It was a fascinating experience for me to test my internalized knowledge of speech patterns, acquired in Sweden, by simply painting a stylized conceived spectrogram of an arbitrary sentence on a plastic sheet which was played back, monotone but intelligible. The simplicity and pedagogical virtue of this device has not been matched by modern computer technology. The Pattern Play Back made possible important studies of speech perception. A critical remark, however, is that the relevance of some of the findings were limited by constraints such as the limited number of formants.

The work on Acoustic Theory of Speech Production (Fant, 1960) was started at MIT. My theoretical background was a combination of the transmission line theory I had learned in Sweden and general acoustics and circuit theory I had studied at MIT. I had not been influenced by the pioneering work of Chiba & Kajiyama (1941) which I came across much later. Back in Sweden 1952 my work continued and became a part of my doctoral thesis in 1958, which also included two additional monographs, one containing results from my speech analysis work at Ericsson 1945-1948 together with the theory of cascaded formant synthesis (Fant, 1959), and one describing speech analysis techniques (Fant, 1958).

The most quoted part of my book is the section on nomograms relating formant frequencies to a vocal tract parameterized tube model. The novelty of this model and that of Stevens & House (1955) developed at the same time, is that it demonstrates the importance of the place and cross-sectional area of the main VT constriction, which has a greater acoustical relevance than the traditional front-back, high-low vowel plane. We now conceive of the vowel [a] as a back vowel with a narrow constriction rather than as a maximally open front vowel.

4. Speech acoustics and simulation means

LEA, the electrical transmission line analog of the vocal tract, described in Fant (1960, 1958) was developed to serve as a simulation tool for studying the relations between articulatory and acoustic patterns. Prior to its construction, in the early 1950's, I also had experience from calculations of VT transfer functions carried out with the first electronic computer in Sweden, BESK, as well as with its predecessor, the binary relay calculation machine BARK. To this earlier period also belongs the work of Stevens et al. (1953). The LEA system (Fant, 1958, 1960) made possible predictions of formant frequency and bandwidth patterns which are not easily observable in speech analysis.

We now have more elaborate vocal tract simulation systems (Badin & Fant, 1984; Lin, 1990; Lin & Fant, 1990) and an extended vocal tract model (Fant & Båvegård, 1995) to gain an understanding of speech dynamics and to pave the way towards articulatory synthesis.

Formant coded synthesis has come to use in perceptual experiments, e.g. for the matching of two-formant approximations to complete vowels in the frame of the F2-concept (Carlson et al., 1970) and also for the determination of perceptual boundaries (Chistovich et al., 1966a; Chistovich et al., 1966b).

5. Your relation with phonetics and linguistics

At the linguistics conference in Oslo 1958 I gave two presentations. One on new methods in experimental phonetics, (Fant, 1958) and one together with Marianne Richter on the relative occurrence of words, letters and phonemes in telephone conversations and in news paper materials and novels (Fant & Richter, 1958). The latter contribution also included a study on the relation between word frequency and rate of occurrence in Swedish, inspired by the work of French et al. (1930) at Bell Telephone Laboratories. Our study also contained data on the relative complexity of words in terms of number of syllables and phonemes per word in various corpora (Fant, 1967). One of the practical applications was in the design of word lists for "articulation" tests, i.e. for assessing hearing loss and speech reception in various conditions of room acoustics.

The fusion of engineering and linguistics interests in speech is a logical consequence of the multidisciplinary character of speech research and growing applications in speech and language technology. An example is that Sven Öhman and Björn Lindblom, both with a background in linguistics, came to work with me for a period and returned to humanities as professors, Öhman at Uppsala University and Lindblom at Stockholm University bringing widened insights in research objectives and methods.

6. Relations between phonetics and engineering

Instrumental resources for basic and applied speech research and also aids for the handicapped, have been greatly advanced by developments in engineering and computer science laboratories. Efficient analysis and re-synthesis software systems are now available.

The influx of new people from engineering and computer science in speech research is increasing rapidly, as judged by the growth of ESCA meetings and the International Conferences on Spoken Language Processing. As a part of the fusion of interests, linguistics departments may compete with engineering departments for funds in both basic and applied research. Cooperative efforts in Sweden and in the EU have been fruitful, but there still remains a divergence of interests and approaches.

In an international perspective, there are two opposite trends. One is that of fractionalization, to choose very narrow objects of studies, e.g. details of speech processing, speech production or perception or to formulate, with or without much experimental evidence, some basic philosophic aspect of the overall speech communication process. The other trend is towards integration of knowledge as a foundation for general phonetics or for the design of text-to-speech synthesis system. This is the code approach.

A difficulty is that the entire system is not merely the sum of the fractional parts. The overall code requires a systematic generative approach with building blocks that fit together with respect to choice of units, terminology and linguistic frame.

A profound knowledge of the speech code would undoubtedly promote advance in speech recognition but we are not there yet. Our present insight is far from complete and not easily accessible. No wonder that speech recognition people favor the statistical approach with HMM and neural networks which have performed remarkably well.

7. Participation to congresses and journals

In the early period of my carrier the meetings of the Acoustical Society of America were of considerable importance to me. Today, my scientific interests are best matched by the International Congress of Phonetic Sciences (ICPhS), the International Conference on Spoken Language Processing (ICSLP, and the ESCA meetings. Major journals of interest are Speech Communication, JASA, Journal of Phonetics, and Phonetica. Today a valuable source of information derives from institutional reports, e.g. the working papers of UCLA and of University of Lund and our KTH reports, the STL-QPSR now named TMH-QPSR.

8. The speech laboratory at KTH

The Speech Transmission Laboratory was formed after my return to Sweden in 1951. In addition to speech analysis and synthesis, a major project during the first years involved language statistics and system evaluations as already mentioned. A substantial growth was encountered in the 1960's on the basis of American grants and later by increased Swedish funding. In 1966, when I received my professorship, we were already a group of 20 people. In 1997, we were the department of Speech, Music and Hearing with a personnel of more than 50 people and five professors; Björn Granström in speech communication, our head of department, Johan Sundberg in music acoustics, Arne Leijon in hearing technology (after Arne Risberg), Rolf Carlson in speech technology, and Johan Liljencrants in electroacoustics and speech. I myself am retired since 10 years but still active. [*]

In the earlier period, the growth of the department has on the whole been that of a self organizing system without any conscious plan to meet certain needs. We have attracted people by virtue of our resources, tradition and scientific breadth. A substantial part of our senior staff has been recruited from former graduate students. We have had an active role in the training of people from linguistics, audiology and speech pathology. Sven Öhman and Björn Lindblom have already been mentioned. We now have a

[*] Johan Sundberg and Johan Liljencrants are now retired and Arne Leijon has moved to a different department of the KTH.

nation wide cooperation with linguistics departments and medical centers in Sweden and with research centers in other countries. Our international contacts, including foreign visitors, have also been promoted by our STL-QPSR reports which were distributed in 800 copies to more than 50 countries, which equals that of Speech Communication. Now they are avaailable on-line only. Several important contributions reside in the STL-QPSR and have not been published elsewhere, see the directory of our published articles in STL-QPSR 1/1995.

The interaction with music acoustics has been very profitable. We have a common interest in the function of the human voice. The music acoustics group under Johan Sundberg has attained an internationally leading position in its field and has recently incorporated multimedia studies of music performance.

Our engagement in the handicap area dates back to an early period of work in speech audiometry and my lecturing at the school for the deaf in Stockholm in the 1950's. During this period, we started research and development of visual aids for speech correction and tactile speech reception systems. Our present engagement in this area involves more complex visual aids for speech training of the deaf and for spoken second language training. We have also participated in a EU project on special speech coding in aids for severely hard of hearing people.

The main application for the text-to-speech system developed by Rolf Carlson and Björn Granström and marketed by Infovox, has been in aids for the blind and speech handicapped (Carlson et al., 1990; Galyas et al., 1992). More than 1000 blind people in Sweden use the Infovox system and the multi-language modules have had a good market. The terminal synthesizer is essentially that of OVE II (Fant & Martony, 1962). Johan Liljencrants, (1969) has taken an active part in the implementation as well as of the OVE I, and he has developed a special parallel system, OVE III. His small company Fonema is now active in supplying complete systems and software for spelling and speech training with the Infovox synthesis module.

During the last 10 years our speech communication group has directed a greater part of its efforts to interactive man-computer dialog systems involving recognition, synthesis, language engineering and studies of speaker variability which implies a change in research profile towards a more system oriented but highly interdisciplinary area.

Our contacts with the Pavlov Institute of Physiology belong to an earlier period of our history. I met Ludmilla Chistovich the first time at the International Congress of Acoustics in Stuttgart 1959. In 1966, she was our guest at KTH. The same year and in 1973 she organized international symposia in Leningrad. Her work, especially the publication of Kozhevnikov & Chistovich (1965) on speech production and perception and control mechanisms, introduced a novel paradigm of approach, and her work on vowel perception gave incitements to studies of gross spectral attributes in relation to individual formants and a search for systematic trends in perceptual boundaries (Chistovich et al., 1966). A special issue of Speech Communication (No. 4, 1985), was devoted to Ludmilla Chistovich.

9. Relations between basic research and applied research

There is a world wide politically induced and increasing trend of priority given to funding of applied development work of immediate industrial interest, which reduces sources for basic research. This is not a new trend. From an early stage I have been used to selling basic research under a cover of potential applications, a situation which prevails today everywhere.

The transfer of knowledge from research to applications deserves some comments of historical interest. One issue is the need of the applicant to possess a basic background of understanding. I have a few examples. One is the false expectation evoked in a small USA company to succeed in speech recognition by developing detectors for the distinctive features of Jakobson et al. (1952). A small number of features were supposed to suffice for all languages of the world. The brave investment came to an end after a year - a naive approach to a potential strategy that has attained a renewed interest, e.g. at M.I.T.

A working prototype is an attractive object for industrial engagement but shortcuts may occur. One such is that the complete design of our OVE II synthesizer, including the function generator for handling parametric control parameters, was copied by a representative of the Melpar company during the 1962 speech communication seminar in Stockholm. Photographs were taken at night without our consent. A year later the Melpar company had its synthesizer EVA on sale. We managed to reach an agreement whereby Arne Risberg and myself where offered royalties. However, only a few copies were sold.

There has always been a tension between forensic science and speech research. In early 1970, the head of the FBI visited Stockholm and gave an interview on the success of Voice Prints for speaker identification. I was asked by a newspaper to comment on the scientific feasibility. I gave my frank opinion about the premature state of the art and the risks involved. The next day on the front page of the newspaper there appeared a résumé with two photographs, one with the text; "Edgar J. Hoover, Chief of the FBI" and the other one; "Gunnar Fant, a possible FBI-Enemy Number 1".

How do I grade the relative importance of my work in different areas? This is a difficult question to answer. From an earlier period, the most well known objects are the already cited studies of the acoustics of speech production and the development of synthesis techniques. Over the years a substantial number of studies in speech production, speech analysis, the human voice source and prosody have been carried out. Most of them have been published in our quarterly reports (see the index in STL-QPSR 1/1995). I have also been engaged in fundamental studies of divers speech under hyperbaric conditions and with various gasmixtures (Fant & Lindqvist-Gauffin, 1968). A more peripheral project was the development of a phonetically structured Cued Speech sign alphabet for deaf subjects (Fant, 1972).

During the last 10 years a substantial time has been spent in studies of prosody (Fant & Kruckenberg, 1989, 1994, 1996; Fant et al., 1991a, b and c). We have also performed studies of metrical structure in poetry reading (Kruckenberg & Fant, 1993). Another main area of interest has been properties of the voice source (Ananthapadmanabha, 1984; Fant et al., 1985; Fant, 1993). A recent publication is devoted to the voice source in connected speech (Fant, 1997a).

The main motivation for my work has been to increase the understanding of fundaments and to combine such insights for the benefit of general theory and development work.. An integrated view of segmentals and prosody is highly needed.

There exists a *knowledge gap* for future developments but also an *information gap* between present status of the art in science and potential applications. Our multidisciplinary speech conferences have an important mission to facilitate the connections. Today, the information gap within a research group is often greater than that between specialists from different groups and countries. You may have to go to a conference to learn about a next door neighbour's work, that you could benefit from.

The challenge for future research is thus to force the speech code, i.e. to predict the articulatory, acoustic and perceptual manifestation of any utterance given the message transcript and the particular language, dialect, speaker and situational context. One can conceive of this task as a very advanced project of deriving rules for speech synthesis. In simple words, it is the ultimate aim of general phonetics and linguistics.

The question is now whether we shall be able to handle all this developing knowledge in explicit rules, or if we have to rely on computers to learn the code. Some kind of compromise emerges already. Much statistically relevant information can be stored in neural networks, but we should not give up the challenge of securing a profound insight.

My forecast for the future is that a more solid and integrated view of speech and language structure will develop and find its way also into speech recognition work. Present systems suffer from a substantial information loss in the front end, where prosody and articulatory continuity is more or less ignored.

The need for a more complete recognition, e.g. including emotions and speaker type, is anticipated in the planning of automatic telephone translation projects. More primitive experimental systems exist, but the demand for an effective dialogue performance will put pressure on basic research.

A developing application of speech research is to improve methods of language teaching. Much could be gained by intensified cooperation between phonetics, speech technology and pedagogics. So, why put your hopes to future availability of fancy speech and language translating systems when you could learn a foreign language? A problem of great social relevance is the need to improve methods and aids for spoken second language acquisition. This is a suitable area for EU projects, as well as for international cooperation.

Sweden's entry into the EU has stimulated our research contacts in Europe. This is my personal experience from the Esprit SPEECH-

MAPS project on speech production and perception, the successful outcome of which to a great extent is due the excellent coordination from the Grenoble group. EU grants have given us new scientific contacts and financial sources. However, a problem is that the access to EU funding has reduced the availability of national resources for more basic research.

Today I enjoy having the freedom of time, and an intact curiosity to remain in speech research, the most multidisciplinary of all scientific fields.

References

Ananthapadmanabha TV (1984). Acoustic analysis of voice source dynamics. *STL-QPSR* 2-3/1984: 1-24.

Badin P & Fant G (1984). Notes on vocal tract computations, *STL-QPSR* 2-3/1984: 53-108.

Carlson R, Granström B & Fant G (1970). Some studies concerning perception of isolated vowels. *STL-QPSR* 2-3/1970: 19-35.

Carlson R, Granström B & Hunnicutt S (1990). Multilingual text-to-speech development and applications. *Advances in Speech, Hearing and Language Processing*, Ainsworth AW, editor (JAI Press).

Chiba T & Kajiyama M (1941). *The Vowel its Nature and Structure*, Tokyo.

Chistovich L, Fant G, de Serpa-Leitao A & Tjernlund AP (1966a). Mimicking of synthetic vowels. *STL-QPSR* 2/1966: 1-18.

Chistovich L, Fant G & de Serpa-Leitao A (1966b). Mimicking and perception of synthetic vowels, part II. *STL-QPSR* 3/1966: 1-3.

Chomsky N & Halle M (1968). *The Sound Pattern of English*. New York: Harper and Row.

Collard J (1930). Calculation of the articulation of a telephone circuit from the circuit constants. *Electrical Communication*, Vol 8 no 3, 141.

Cooper FS, Delattre PC, Liberman AM, Borst JM & Gerstman, LJ (1952). Some experiments on the Perception of Synthetic Speech Sounds. *J Acoust Soc Am*, 24, 597-677.

Fant G (1948). *Analys av de svenska vokalljuden*. LM Ericsson report H/P-1035.

Fant G (1949). *Analys av de svenska konsonantljuden. Talets allmänna svängningsstruktur*. LM Ericsson report H/P-1064.

Fant G (1953). Speech communication research. *IVA* (Swedish Academy of Engineering Sciences) No. 8: 331-337.

Fant G (1958). Modern instruments and methods for acoustic studies of speech. *Proc VIIIth Int Congress of Linguists*, *Oslo Univ Press*, 282-358.

Fant G (1959). *Acoustic Analysis and Synthesis of Speech with Applications to Swedish*. Ericsson Technics No. 1, 1959.

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

Fant G (1967). *Kompendium i talöverföring.* KTH-TMH.

Fant G (1972). Q-codes. Int Symp on Speech Communication Ability and Profound Deafness, Stockholm 1970 (AG Bell Ass for the Deaf, Washington, DC, eds), 261-268.

Fant G (1973). *Speech Sounds and Features*. MIT Press.

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

Fant G (1993). Some problems in voice source analysis. *Speech Communication* 13: 7-22.

Fant G (1994). Lectures in China 1985. Translated into Chinese. *Academia Sinica* (172 pages).

Fant G (1995a). The LF-model revisited. Transformations and frequency domain analysis. *STL-QPSR* 2-3/1995: 119-155.

Fant G (1995b). Speech related to pure tone audiograms. *Profound deafness and speech communication.* Plant G & Spens K-E (eds), Whurr Publ Ltd, London, 299-305.

Fant G (1997a). The voice source in connected speech. *Speech Communication* 22: 125-139.

Fant G (1997b). Acoustical analysis of speech In: Crocker MJ (ed), *Encyclopedia of Acoustics, John Wiley,* Vol .4*:* 1589-1597.

Fant G & Båvegård M (1995). Parametric model of VT area functions: Vowels and Consonants. *ESPRIT/BR Speech Maps* (6975), Delivery 28 WP2.2, 1-30.

Fant G, Liljencrants J & Lin Q (1985). A four-parameter model of glottal flow. *STL-QPSR* 4/1985: 1-13.

Fant G & Lindqvist-Gauffin J (1968). Pressure and gas mixture effects on divers speech. *STL-QPSR* 1/1968: 7-17.

Fant G & Kruckenberg A (1994). Notes on stress and word accent in Swedish, *STL-QPSR* 2-3/1994: 125-144. Also published in *Proc Int Symp on Prosody, 18 Sept 1994, Yokohama,* 19-36.

Fant G & Kruckenberg A (1996). On the quantal nature of speech timing. *Proc ICSLP-96,* Vol. 4: 2044-2047.

Fant G & Kruckenberg A (1989). Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 2/1989: 1-83.

Fant G & Martony J (1962). Speech synthesis instrumentation for parametric synthesis (OVE II). *STL-QPSR 2/1962*: 18-19.

Fant G & Richter M (1958). Some notes on the relative occurrence of letters, phonemes, and words in Swedish. *Pro. VIIIth Inl Congress of Linguists,* Oslo Univ Press, 815-816.

Fant G, Kruckenberg A & Nord L (1991a). Durational correlates of stress in Swedish, French and English. *J Phonetics* 19: 351-365.

Fant G, Kruckenberg A & Nord L (1991b). Prosodic and segmental speaker variations. *Speech Communication* 10, 521-531.

Fant G, Kruckenberg A & Nord L (1991c). Language specific patterns of prosodic and segmental structures in Swedish, French and English. *Proc of the XIIth ICPhS,* Aix-en-Provence, 1991, 118-121.

Flanagan JL (1979). From the doctorial thesis defense of C Gunnar M Fant. Verbatim Remarks of the Principal Opponent. In: Lindblom B & Öhman S (eds). *Frontiers of Speech Communication Research,* London: Academic Press.

Fletcher H (1929). *Speech and Hearing*. New York.

French NR, Carter CV & Koenig S (1930). The words and sounds of telephone conversations. *Bell System Technical Journal* 9:290-324.

Galyas K, Fant G & Hunnicutt S (1992). *Voice output communication aids*. A study sponsored by the International Project on Communication Aids for the Speech Impaired, IPCAS. The Swedish Handicap Institute, Stockholm (86 pages).

Hockett CF (1955). *Manual of Phonology*. Indiana Univ Publ in Anthropology and Linguistics, No. 11, Bloomington.

Jakobson R, Fant G & Halle M (1952). *Preliminaries to speech analysis. The distinctive features and their correlates.* Acoustics Laboratory, Massachusetts Inst of Technology, Technical Report No. 13 (58 pages). Published by MIT press, seventh edition, 1967.

Joos M (1948). Acoustic Phonetics. *Language* 24: 1-136.

Kozhevnikov VA & Chistovich LA (1965), *Speech Articulation and Perception.* US Dept of Commerce Publ. JPRS:30 543; TT: 65-31233. Translated from Russian. (See also Vol. 4 (1985) of Speech Communication devoted to Ludmilla Chistovich.)

Kruckenberg A & Fant G (1993). Iambic versus trochaic patterns in poetry reading. *Nordic Prosody VI,* Stockholm, 123-135.

Lawrence W (1953). *The synthesis of Speech from Signals which have a Low Information Rate*. Communication Theory. Jackson W (ed), London.

Liljencrants J (1969). Speech synthesizer control by smoothed step functions, *STL-QPSR* 4/1969: 43-50.

Lin Q (1990). *Speech Production and Articulatory Synthesis*. Ph D thesis, Dept of Speech Communication and Music Acoustics, KTH.

Lin Q & Fant G (1990). A new algorithm for speech synthesis based on vocal tract modeling. *STL-QPSR* 2-3/1990: 45-52.

Peterson GE & Barney GE (1952). Control methods used in the studies of vowels. *Acoust Soc Am* 24: 629-637.

Potter RK, Kopp AG & Green HC (1947). *Visible Speech*. New York.

Stevens KN & House AS (1955). Development of a quantitative description of vowel articulations. *J Acoust Soc Am*, 27: 484-493.

Stevens KN, Kasowsky S & Fant G (1953). An electrical analog of the vocal tract. *J Acoust Soc Am* 25: 734-742.

(STL-QPSR = Speech Transmission Laboratory Quarterly Progress and Status Reports, KTH, Stockholm, now (TMH-QPSR, Speech, Music and Hearing, Quarterly Progress and Status Reports).