

PHONETICS AND PHONOLOGY IN THE LAST 50 YEARS

Gunnar Fant

Dept. of Speech, Music and Hearing, KTH, Sweden

gunnar@speech.kth.se

ABSTRACT

My overview has four parts. One is concerned with the two years 1949-1951 I spent at MIT, a pioneering era of speech research of fundamental importance in shaping my research interests, and the start of a lifetime co-operation with Ken Stevens and with Roman Jakobson and Morris Halle. An outcome of our joint work on feature theory was the X-ray studies we performed on a Russian immigrant, which also provided the foundation for my book Acoustic Theory of Speech Production. The second part is an attempt to structure the half century of developments into four successive periods. The third part is concerned with brief surveys of each of a number of activities in acoustics, phonetics and phonology, and applications in speech technology, handicap systems, music acoustics and a number of medical specialities. The fourth part is devoted to my subjective overview of the entire field, how research topics are selected, and basic issues such as the nature of invariance and variability, and the concept of the speech code. From 1950 to present days there has been a gradual shift from basic research towards more application-oriented activities, promoted by the rapidly expanding information technology. As a result, statistical methods operating on large data banks, backed up by increasing computer power, now dominate system developments. However, there is now a growing insight that we can not realize far reaching goals without incorporating more fundamental knowledge about human speech communication and information bearing elements of speech.

MY EARLY YEARS AT MIT

First of all I want to express my thanks to the organizers for inviting me to give this introductory talk. From a scientific point of view, MIT has been my second home. It is now a challenge and a pleasure to share with you some recollections from early years. I will tell you how it started. It was through an invitation from Leo Beranek, whom I met when he visited the Ericsson Telephone Company in Stockholm in 1948. I was then engaged in a speech analysis project. He suggested that I continue research and studies at the MIT Acoustics Laboratory. This was how I came to MIT in November 1949. I stayed till May 1951. Several return trips were made in the following years.

Leo Beranek arranged for me to spend a few months at the Harvard University Psycho-acoustic Laboratories, where I could follow ongoing work on auditory functions and speech perception. I met S.S. Stevens and other prominent psychologists, such as J.C.R. Licklider, Walter Rosenblith, George Miller and also Ira Hirsh, then a graduate student.

It was here at a seminar I gave that I first met Roman Jakobson. I reviewed my speech analysis work at Ericsson 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 versus diffuse and grave versus acute, in both vowels and consonants. The common denominator of compactness versus diffuseness is the central and

concentrated, versus spread of spectral energy. I showed him the following examples of burst spectra in the release phase of final unvoiced Swedish stops, Figure 1. The [k] has a single, centrally located peak, which accounts for the compactness. There is a high frequency emphasis of the [t], qualifying for acuteness and a low frequency emphasis for [p] which is grave. These spectra originate from my Ericsson studies in 1949 (Fant, 1959).

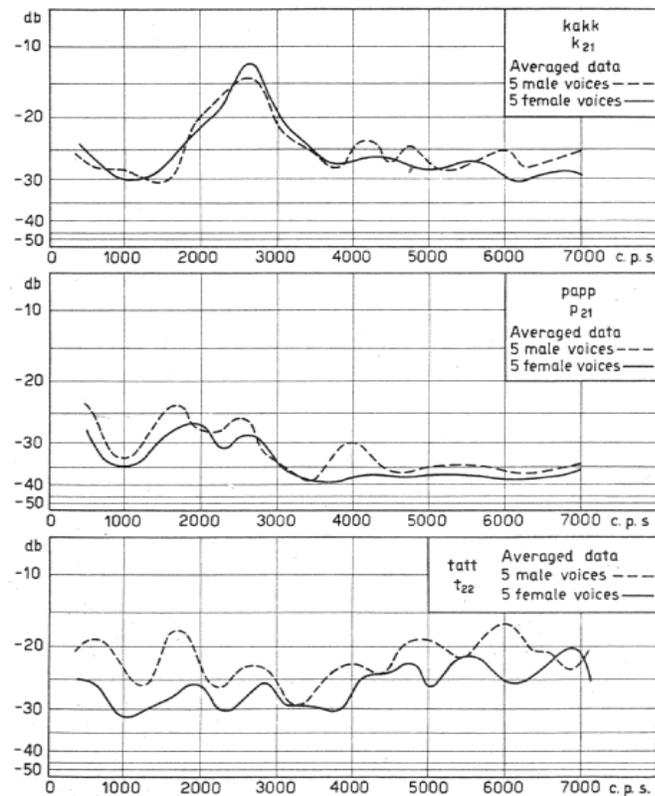


Figure 1. Burst spectra of Swedish aspirated [k], [p] and [t], preceded by a short vowel [a]. Average of 5 male speakers broken lines, average of 5 female speakers solid lines. From Fant (1949, 1959).

Morris Halle joined us in the development of distinctive feature theory, (Jakobson et al., 1952). This became 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. As a part of our project we performed X-ray studies of a Russian immigrant, which provided data for my book *Acoustic Theory of Speech Production* (Fant, 1960). At that time I started my theoretical work. Ken Stevens was then a graduate student. In 1952 he defended his thesis on "The perception of speech sounds shaped by resonance circuits". During my stay at MIT we started a cooperation on transmission line theory applied to the vocal tract (Stevens, et al., 1953).

My stay at MIT was in a creative period of interdisciplinary contacts, a true pioneering era. Information theory had recently been introduced and caused considerable excitement. However, apart from its great importance in telecommunications (Shannon & Weaver, 1949), information theory did not open new insights in human behaviour, but it gave us a frame and has contributed with one important theoretical tool, the concept of redundancy. Redundancy operates on two levels. One is in classifying structural units in terms of their relations. In the Jakobson, Fant, Halle feature system a small number of features defined a large sound inventory, and redundancy was accordingly reduced to a minimum. The other aspect of redundancy is the contribution of details to ensure correct identification.

CHRONOLOGICAL OVERVIEW

In attempting to summarize more than half a century of accomplishments I am aware of the limitations. My review is bound to be personal, limited by my own experiences and evaluations, see also Fant (1996, 2004a). Also, because of the restricted frame, many important contributions had to be omitted. More detailed overviews are to be found in the individual presentations to our conference. A worthwhile reading is the History of Phonetics in the United States (Ohala et al., 1999) with presentations of separate institutions and research groups where Ken Stevens contributed with an account for activities at MIT.

Before 1950

Of major importance for our field was the development of the sound spectrograph at Bell Laboratories (Potter et al., 1947). An early and initiated study of acoustic phonetics was that of Martin Joos (1948). My early work at Ericsson was reported in Fant (1948, 1949) and later in Fant (1959). A forerunner of vocal tract theory was that of Chiba and Kajiyama (1941). My first contact with the field of speech and hearing was through Fletcher (1928).

1950 – 1965

- Pioneering era. Inter-disciplinary trends
- Acoustic phonetics. Feature theory
- The speech communication chain
- Intelligibility, system evaluation
- Hearing loss and hearing aids
- Speech compression. Bandwidth reduction
- Early period of speech synthesis

1965 – 1980

- Computers in general use
- Digital speech processing
- Text-to-speech systems under way
- Speech recognition
- Auditory theory. Models of speech perception
- The search for invariance

1980 – 1995

- Models of the voice source
- Articulatory modelling, MRI
- Speech recognition and text-to-speech systems
- HMM and neural networks
- Man computer dialog systems
- Handicap aids

1995 – 2004

- Large data bases, also for text-to-speech synthesis
- Demands for prosody modelling
- Automatic telephone translation
- Talking-head speech synthesis
- Insert hearing aids gaining market
- Advances in neuro-physiological studies

Comments

I shall now make a brief attempt to comment speech research and applications in the four successive periods. This is a difficult task. Many projects such as speech synthesis extend over the entire 54-year period, or they have had an early start and have been resumed with full activity later. Examples are hearing loss and hearing aids and models of speech perception.

1950-1965

The first period, arbitrarily assigned to the years 1950-1965, was a pioneering era of interdisciplinary contacts. The sound spectrograph, developed already in the 1940s at Bell Laboratories (Potter et al.,1947) gave us an insight in the nature of speech signals, and prompted questions of how speech patterns are related to the linguistic frame and the nature of information bearing elements. A classical study of vowels is that of Peterson & Barney (1952). An early acoustic phonetic study of spectrographic data in terms of manner and place features and of the temporal distribution of information bearing elements appeared in Fant (1962).

The apparent complexity of spectrographic patterns called for studies of the entire speech chain, production, speech wave patterns and perception. Reading a spectrogram proved to be quite difficult even for experts. However, it was found that knowledge of speech production provided an important key. This insight has prevailed in the entire history of speech research and has influenced production-oriented theories of speech perception. Basic studies of vocal tract acoustics gave a support to the expanding domain of speech synthesis.

In this early period much work was devoted to assessments of intelligibility as a function of properties of a communication system, including room acoustics and hearing loss. Major technical applications were systems for data reduction and bandwidth saving in telephony, e.g. the vocoder, see Fant & Stevens (1960). Most of present days activities were initiated in this early period.

An incitement for data reduction projects was the apparent mismatch between the linguistically defined information rate in speech and the bit-rate required for effective transmission, differing

by something like a factor of 1/1000 or more. Today a standard rate in telephony is 12 Kbit/s. For special applications it can be compressed to the order of 4 Kbit/s (Spanias, 1994).

1965-1980

In the second period, 1965 to 1980, computers had become a standard laboratory tool, which marked the transition from analog to digital processing in all aspects of speech research, and was the foundation for text-to-speech synthesis and speech recognition. Auditory theory and models of speech perception were expanded. The topic of invariance and variability created a large interest as manifested by an MIT symposium (Perkell & Klatt, 1986). Complete text-to-speech systems were developed at this time.

1980-1995

In the third period, 1980–1995, statistically oriented methods of speech processing, HMM and analog neural networks became standard tools for speech recognition and applications in man-computer dialog systems. Text-to-speech technology provided valuable aids for the blind as well as for speech and language handicapped people.

In this period articulatory modelling of speech became a major object of research. It was promoted by the developing MRI and other more direct techniques. Studies of the human voice source in connected speech were initiated. Musical acoustics expanded, especially in connection with studies of the singing voice (Sundberg, 1987).

1995-2004

Ken Stevens' book *Acoustic Phonetics* was published in 1998. It has contributed unique insights of monumental value, especially of speech production, collected during a major part of his half a century of research.

A main trend in the fourth period has been a shift of emphasis towards speech technology applications with an influx of people and groups from computer science, applied linguistics and psychology. A recent development in speech synthesis is the addition of a talking head for public information systems and auditory handicap applications (Beskow, 2003).

The concept of automatic telephone translation has become a challenge. At one end of a telephone line it would involve the combination of language and speech recognition and analysis of speaker type, speaking style and emotions. On the other end it assumes a language translation and a re-synthesis of speaker characteristics and speaking style, adjusted to match demands of the target language. This would be the far-reaching objectives of speech communication research, adopted by the ATR laboratories in Japan and in part also in Germany and other places. Obviously, these ambitions require a much broadened knowledge base.

OVERVIEW OF SELECTED AREAS

I shall now attempt a more detailed survey of the status of the art in a number of selected areas. The first is our basic concept of the speech chain. We are concerned with mechanisms and data on the production end, the conversion of production data to speech wave data, and the mechanisms of speech perception and cognition. The underlying assumption is that the speaker and the listener each possess a language competence, which could be the same or different.

THE SPEECH COMMUNICATION CHAIN

SPEAKER

LANGUAGE COMPETENCE

PRODUCTION

MECHANISMS

DATA

SPEECH WAVE

DATA

LISTENER

PERCEPTION

DATA

COGNITION

LANGUAGE COMPETENCE

Many of you have contributed to knowledge within this frame. In the first place I would like to mention the early work of James Flanagan (1965); Peter Ladefoged (1967, 1975) and of course Ken Stevens (1998). A knowledge of the encoding between successive links is implicit, but the speech chain is merely a skeleton for adding insights in language units and of the nature of information bearing elements. Such studies are badly needed to ensure further progress in speech synthesis, speech recognition and other applications.

Speech production

Speech production has been a major object of research during the entire half century. It has now gained increased importance, especially for applications in articulatory synthesis. Our major reference is that of Stevens (1998). Several methods of tracking articulatory movements have been developed. One was the Micro-beam X-ray system initiated by Osamu Fujimura. Now the main tools are MRI scanning and the use of magnetic sensors and ultrasound. Important contributions have been made here at MIT by Joe Perkell and his group (Perkell & Matthies, 1992). One object has been to study compensatory modes of articulation. Among earlier work can be mentioned the processing of a cineradiographic film (Perkell, 1969). A detailed study of vocal tract computations with special reference to losses and gas mixtures appear in Badin & Fant (1984). Fricative production has been studied by Badin (1989).

Today the main interest is devoted to three-dimensional mapping of the entire articulatory system as pursued in Stockholm and Grenoble (Engwall, 2003). A new approach to the design of an articulatory synthesizer was suggested by Lin and Fant (1992).

The voice source

The voice source is a major component of the production mechanism. An influential mechanical model was that of Ishizaka and Flanagan (1972). Studies of the human voice source by means of inverse filtering techniques have provided a base for acoustical models. One is the LF model (Fant et al., 1985; Gobl, 2003), which has been applied to descriptions of voice qualities, but more importantly to voice source dynamics in connected speech (Fant, 1997; Fant &

Kruckenbergh, 2004ab) with applications in speech synthesis. The voice source in singing has been studied by Sundberg (1987); Sundberg et al. (1999).

Properties of the voice source add to perceived stress. One is a spectral slope parameter referred to as "spectral tilt" (Sluejter & van Heuven, 1996). An increase of vocal effort is usually accompanied by a relative increase of formant amplitudes, especially of F2 and F3, in relation to the amplitude of the voice fundamental (Hanson & Chuang, 1999; Fant et al., 2000).

Aerodynamics

Early studies of Ladefoged (1967) on the role of the subglottal pressure in speech have been followed up by Rothenberg (1968); Ohala (1990); Stevens (1998) and by Liljencrants et al. (2000). More recent studies of Fant et al. (2004ab) on the covariation of F0, voice intensity and sub-glottal pressure in prose reading have revealed close functional ties, which allow a prediction of intensity from F0 and sub-glottal pressure.

Intonation control

Our insights in laryngeal control mechanisms and the modelling of F0 in connected speech with applications in speech synthesis are largely due to Hiroya Fujisaki (2004). His model has its roots in the early work of Öhman (1967). Basic physiological studies of laryngeal mechanisms have been performed by Titze (1989).

Vowel perception

I shall now turn to perception. A most influential study of speech perception and the organization of speech production was published by Kozhevnikov & Chistovich (1965). Their findings concerning two-formant approximations of vowels was followed up by several research groups, at KTH by Carlson et al. (1970); Fant (1978).

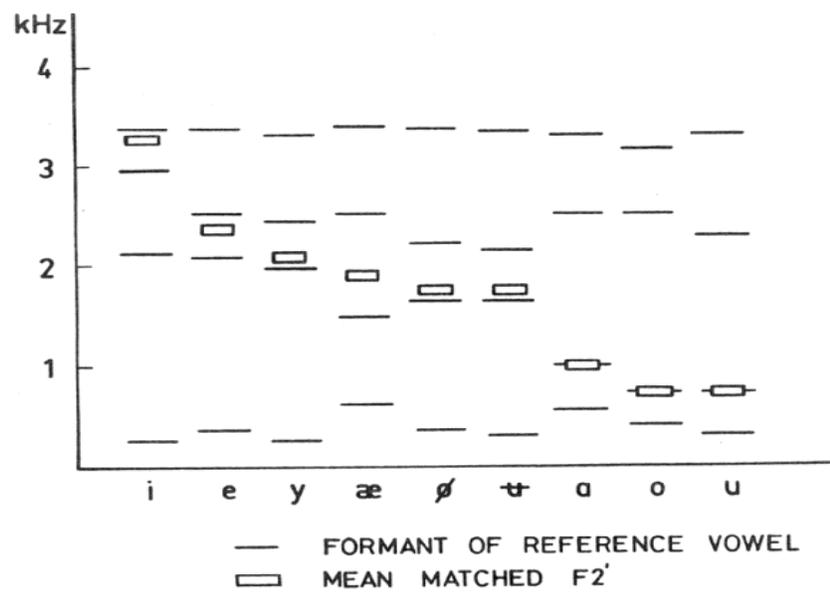


Figure 2. F2' (F2 prime) of Swedish vowels (Carlson et al., 1970).

Figure 2 shows the formant patterns of a sequence of Swedish vowels. The two-formant versions have the same F1 as the full vowel and an upper formant labelled F2' (F2 prime). F2 prime comes close to F2 in back vowels and in the vicinity of F3 and even F4 for a long tense vowel [i:] of a male speaker.

Carlson et al. (1970) found that a cochlea model operating on the distribution of zero crossings within a set of band-pass filters gave a support to the hypothesis of frequency selectivity, which could span larger domains than a single vowel resonance. This time-domain processing was inspired by the findings of Sachs & Young (1979).

Motor theory of speech perception

A hot topic in theories of speech perception has been to what extent a listener employs his capacity as a speaker for the identification of speech sounds. In an extreme version of the motor theory of speech perception the listener consults an internalized model of the vocal tract as a reference. This would be in analogy with our need for a production model when attempting to read spectrograms. It is tempting to project our models of speech processing on brain functions.

A less specific version of the theory was claimed by Alvin Liberman (Liberman & Mattingly, 1985). Present days neuro-physiological studies provide some support. Brain scanning experiments have revealed that speech motor functions are to some extent activated simultaneously with the auditory cognitive process. But we are still in an early stage of exploring brain functions in speech and hearing.

How do we listen to speech? Liberman (1986) introduced the notion of listening in the speech mode as different from a pure auditory process, in which degraded speech may be sensed as strange sounds only.

Speech synthesis

I shall now turn to speech synthesis. The main area of development started around 1950 with applications in speech perception. A most remarkable system of great influence as a research tool was the Haskins Laboratory Playback system (Liberman et al., 1952), Figure 3.

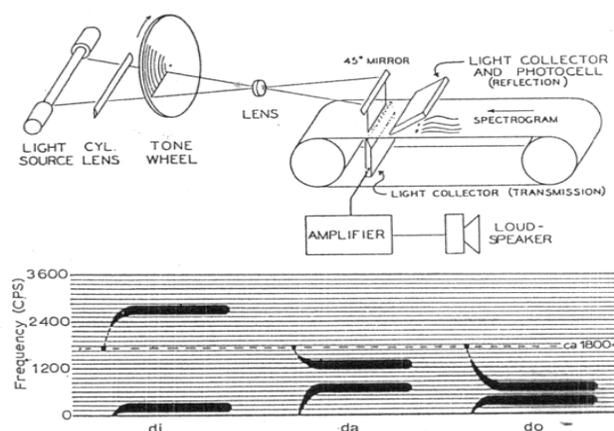


Figure 3. The Haskins Laboratory Play Back synthesizer.
From Liberman et al. (1952).

It produced a monotone replica of a spectrogram or of a stylized hand painted formant pattern. When I visited them in early 1950, my experience from speech analysis enabled me to paint a sentence pattern on a plastic sheet, which when played back produced an acceptable replica. Studies of prototype formant patterns and speech perception at Haskins gave a first insight in the language of Visible speech.

At MIT a combination of resonance circuits, POVO, was used for perception experiments. A more sophisticated system was my Swedish OVE 1 vowel synthesizer, a complete 5 resonance cascade analog of the vocal tract (Fant, 1953), Figure 4. F1 and F2 were continuously varied by hand with a simultaneous dynamic control of F0. It could produce sentences in which voiced consonants were approximated by vocalic glides. After some practice I could produce quite satisfactory sentences.

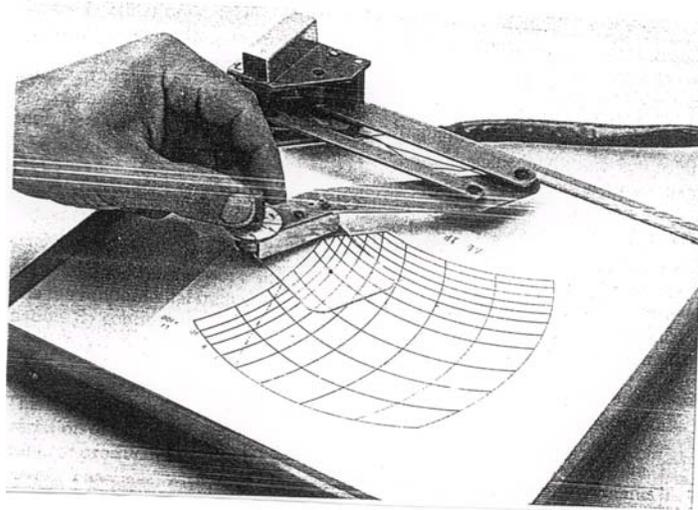


Figure 4. The OVE 1 synthesizer operating in an F1 versus F2 plane with separate control of F0.

At that time the first programmable synthesizer was demonstrated by the Englishman Walter Lawrence (1953). He came to MIT with two huge boxes of equipment. His Parametric Talker, PAT, employed optically scanned function lines.

A more versatile system for synthesis from parametric function lines was developed at our lab in Sweden around 1958 and was demonstrated at the speech communication seminar in Stockholm in 1962, (Fant & Martony, 1962). It enabled high quality copy synthesis from information derived from speech analysis. The main outline of the synthesizer was the same as we still use in Stockholm. It contained a vowel branch, a nasal branch and a fricative branch in parallel, each with cascaded formant circuits fed from a voice source and a noise source. A well known sentence from that time was "I enjoy the simple life" produced by John Holmes (1962),

an English pioneer in speech research and synthesis. He was the first to demonstrate that copy synthesis could produce sentences closely matching originals.

A major advance in speech technology was the development of complete text-to-speech systems, initiated around 1970-1980 at Bell laboratories (Cooker et al., 1973) at KTH in Stockholm (Carlson & Granström, 1975, 1990), at MIT (Klatt, 1980, 1987), and in an earlier period also at Haskins Laboratories (Mattingly, 1968). The formant coding could at times produce fairly natural sounds, but the intelligibility was not very high. The introduction of di-phone coding, first introduced at Bell Laboratories (Olive, 1977) overcame this problem at the expense of some concatenation artefacts.

Present outlooks on text-to-speech synthesis may be found in Dutoit, T. (1997) and in Sagisaka et al. (1997). Today we have two competing technologies. One is an extension of di-phone coding, e.g. in the PSOLA and the MBROLA systems. The latter has been adopted for implementing our prosody rules (Fant et al., 2002). An alternative system now gaining ground is the so called UNIT SELECTION, which allows concatenation of larger units such as complete syllables, words and short phrases selected from a very large recording of a single speaker. It sounds quite human and is especially useful for limited vocabulary applications. However, a limitation is that it usually fails in providing a continued natural prosody across prosodic boundaries, and is thus less suited for the reading of unrestricted texts.

Much hope has been directed to articulatory synthesis as an ultimate solution, but the progress has been slow. As already mentioned, advanced work on vocal tract mapping and aerodynamics are under way. Most of this work has had an emphasis on static properties or on isolated dynamic gestures, as studied at Haskins laboratories (Kelso et al., 1986). However, without integration into a proper acoustic phonetic and linguistic frame, our present knowledge will not suffice to meet the demands for text-to-speech applications. We have some tools but not a complete system of rules. These may in part be derived by inverse transformations, i. e. a prediction of articulatory configurations and gestures from existing spectrographic data, Lin & Fant (1989); McGowan (1994). As proposed by Stevens & Bickley (1991) a hybrid version would be to employ a formant based system, which is controlled by higher level, articulatory oriented rules.

During the last 15 years a substantial part of my work together with Anita Kruckenberg has been directed towards studies of prosody with application in text-to-speech synthesis. (Fant et al., 2002; Fant & Kruckenberg, 2004ab). The novelty of our approach lies in the normalization of intonation contours on a semitone scale, which allows average data for males and females to be calculated. Our Swedish tonal accents 1 and 2 are superimposed on basic prosodic modules for F0 onset, rise and decay within syntactically selected parts of speech. The amplitude of F0 modulations as well as phoneme duration are controlled by a continuously varied prominence parameter.

Some aspects of our system are language universal and some are language specific, which has allowed us to transfer rules from Swedish prosody to English and French, using existing sound inventories from Babel-Infovox (now Acapela Group) standard voices. As a consequence, we have been able to demonstrate the switching of language codes, e.g. simulating an English talker reading a French text retaining his English pronunciation and prosody. An even more

striking example is the simulation of a French talker attempting to read an English text. They suggest applications in language teaching.

Speech recognition

The relative success of speech synthesis has created an illusion that we have a profound insight in the speech code (Fant 2004b). This illusion becomes especially apparent when operating in the reverse direction, that is, given a record of the speech wave we attempt to decipher what was said. This is a reminder when attempting spectrogram reading, and explains the rather slow progress in speech recognition.

Speech recognition is still in a developing stage, but important applications are now emerging, e.g. in limited vocabulary, man computer interface systems. However, it has taken us the better part of the half century. As long as we rely on statistical methods only, the general problem of unlimited vocabulary, speaker independent recognition remains out of reach. A profound general knowledge of the speech code must develop and penetrate into development programs. A neglected dimension is speech prosody and its interaction with segmental units.

Now in the back mirror, I recollect an extreme example of early optimism. Some time around 1970 a small company was created in New England with the prospect of applying a distinctive feature approach to speech recognition. Their outlook was as follows: "Jakobson, Fant and Halle claim that a limited number of features is sufficient to describe all the languages of the world. Let's attempt to define them in more detail, and we will have solved the problem of speech recognition." After a couple of years the firm went bankrupt.

It has taken considerable patience for funding agencies to support research and development over long periods of time, and we in the branch have accordingly learned to sell basic research under the cover of applications. An example of impatience comes to my mind. It was from the 1960s. MIT speech research was then in part supported by the Cambridge Air Force Research Centre. Their contract officer, Weiant Wathen-Dunn, a most friendly and insightful man, made the following impertinent remark at a banquet: "Speech research is like a huge pit. You can throw any amount of money into it and nothing comes out." Well, it took some time, but here we are and much has come out. I have him in good memory. It was through Weiant Wathen-Dunn that we received US Air Force support for work in Sweden. Additional support from the National Institute of Health and from the US Army in the 1960s were of considerable importance in developing our laboratory.

Forensic applications

Speech recognition has close ties to speaker recognition and speaker verifications. I am not very impressed by the present status of the art. In 1970 Stockholm was visited by the head of FBI. He was interviewed by Dagens Nyheter, our leading daily newspaper, where he expressed quite optimistic views on lie detectors and especially about voice print methods for speaker identification. This created some interest and the newspaper turned to me for comments before the interview was published. My response was rather negative, which resulted in a first page article with two photographs, one of Edgar J. Hoover, head of the FBI, and one of Gunnar Fant in the role of an FBI enemy number one.

Handicap aids and medical applications

Speech technology has had many applications in handicap aids, e.g. in text-to-speech for the blind and speech output aids for non-speaking individuals with neurological handicaps. (Galyas et al., 1992; Carlson et al., 1990). In Stockholm we have experience of aids for more extreme cases of language impairment, requiring the use of symbol encoding. One application has been speech output facilities for a Bliss symbol chart. It generates synthetic sentences predicted from a sequence of selected input symbols (Hunnicut, 1984). A related topic is aids for predicting computer text (Hunnicut, 1989). The user is given a list of possible continuations of his initial text.

There are many applications in the use of text-to-speech for language learning. One is to let a dyslectic person manipulate a synthesizer to train letter to sound rules, which will increase his phonological awareness.

I anticipate advanced applications to be reached in aids for language teaching. But it requires a more solid knowledge base of speech prosody than what is generally available today. After all, the basic difficulty in understanding a second language speaker usually lies in incorrect prosody. The same argument applies to our reactions to less advanced text-to-speech synthesis in our own language, which may sound as an immigrant speaking.

Technical methods and results from speech research have penetrated into several medical fields. Close at hand are activities in speech and hearing departments. My own experience dates back to my first active years of research in Stockholm around 1948. I was part time involved in speech audiometry and performed a study to predict how much of the speech spectrum could be available for a person, given his pure tone audiogram and a specific speaking distance.

At Ericsson in 1947-1949 I had collected formant data of vowels and consonants with absolute calibration of sound pressure levels. This is shown in Figure 5 together with an overview of formant regions. This graph has been adopted in audiology under the name of the "speech banana" (Fant, 1995).

Hearing aids have not developed much during the half century. In order to suit a specific hearing loss and the needs of a specific situation, there now exist means for individual control of gain and frequency response and of the sensitivity to dynamic variations. However, technical facilities can only in part compensate impaired auditory and cognitive functions, especially in situations with a group of people talking and under unfavourable room acoustics. The best solution is to have a portable microphone to point at the partner.

Cochlear implant aids employ a specific speech processing for frequency space encoding in the inner ear (Wilson, 1993). This has been one object of clinically oriented research. Although auditory patterns conveyed by the aid preserve only gross features of speech, a language code develops after some training (Risberg & Agelfors, 1984; Agelfors & Risberg, 1991). Successful applications for very young deaf children have now resulted in recommendations for implant aids in both ears.

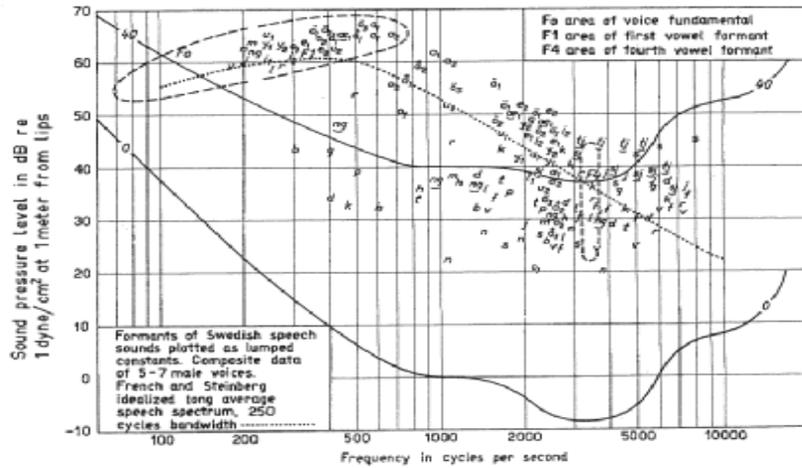


Fig. 45. Sound pressure level versus frequency plot of the vowel and consonant formant data.

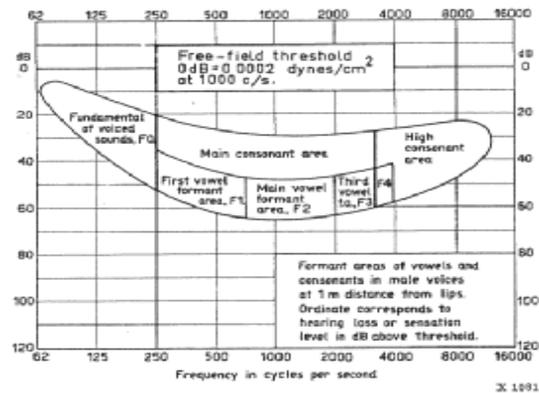


Fig. 46. Schematized presentation of the average speech spectrum within an audiogram in terms of main formant areas. Sensation levels (down) are relative to the standardized free-field threshold.

Figure 5. Above, scatter plot of Swedish vowel and consonant formant data. Below, summary view of formant domains in terms of sensation level in a pure tone audiogram. From Fant (1949, 1959, 1995).

Visual aids for the hard of hearing and deaf have a tradition from tests performed in the 1940s at Bell Laboratories when a group of deaf subjects were trained to read a moving spectrum display. A limited recognition was reported. Similar studies at KTH in Stockholm in the 1960s were not very successful.

However, visual displays have proved to be more helpful in speech training. In more recent years advanced computer displays of speech parameters have been successfully used in second language speech training (Öster, 1989).

Much work is nowadays devoted to the development of talking heads as a support for text-to-speech systems for public use. They are also expected to find use in aids for communication with hard of hearing people and for speech training.



Figure 6. Example of talking head. From Beskow (2003).

Another possibility is communication through the tactual sense. In 1961 Mac Pickett, then a guest in our lab in Stockholm, made an evaluation of a system employing a channel vocoder modulating vibrators attached to each of the ten fingers (Pickett, 1962). A reasonable vowel discrimination was reported, but the system was judged to be unpractical. Simpler systems employing a single or a few vibrators in a one-hand control have come in use. A limitation is the rather large time constant in tactual reception.

A handicap application we have been involved in was a manually controlled intonation device for vibrators to be used by larynx-ectomized subjects (Branderud et al., 1974). It was inspired by the success of the manual F0 control in the OVE 1 synthesizer.

Intonation training may be carried out with a personal feedback aid employing an F0 extractor attached to the neck (Fourcin, 1981).

Sign language for the deaf is one possible area of application. One category is systems intended as a support to lip reading. I have in mind the Cued Speech system developed around 1965. It excluded all phonemes that could be inferred visually. With some support from Arne Risberg I developed a system without this restriction, structured in distinctive feature categories (Fant, 1972). It has been tested in a Japanese school for the deaf.

Divers' speech

Acoustic theory of speech production has been adopted for assessments of cleft palate surgery and other clinical operations. This reminds me of a related study I have been involved in. A special case of speech distortion occurs in divers' speech when operating at a great depth,

requiring a specific gas mixture for breathing, usually based on helium which accounts for the well known Donald Duck effect. This is a linear frequency transposition induced by the increased velocity of sound. I have contributed with an understanding of the effects of the specific density of the gas in combination with the velocity of sound.

The incitement for the study was a report from a Swedish Navy physiologist Bertil Sonesson, that divers when inserted in a decompression chamber, breathing air, sounded nasal. He had installed a cineradiographic equipment in the chamber, but the X-ray pictures showed normal naso-velar functions. A spectrographic analysis revealed an apparent increase of the first formant frequency, which I could relate to a greater than normal participation of the vocal cavity walls, induced by the higher than normal density of air. The closed vocal tract resonance frequency is raised due to the reduction of the impedance contrast between cavity walls and the enclosed air column.

To suit specific diving conditions it is now possible to choose a gas mixture for a certain diving depth, with the view of optimising the performance of "speech unscramblers" for spectral restoration of the transmitted speech (Fant & Lindqvist-Gauffin, 1968). My colleague Johan Liljencrants has been active in the development of such systems.

TRENDS AND BASIC ISSUES

At his stage I would like to make an overview of trends and basic issues. We have experienced an impressive growth of research and development within an expanding number of areas, but it has taken us half a century or more, and this is a rather long time. Here, I would like to make some critical remarks.

The symbiosis between technology and basic research that made possible the advance, now shows a tendency to turn into polarization. Speech technology has become highly dependent on statistical tools and large data bases, whilst phonetics tends to become fractionalized by narrowly defined problems or by abstract issues with at times little or no relevance for the overall code of spoken language.

Also the transfer of available knowledge from phonetics to technology has not been the best. As a result it has been up to those developing text to speech systems to perform their own studies and data collections. Dennis Klatt accordingly managed to collect an impressive basic knowledge about speech and of speaker performance. The same can be said about Carlson and Granström in Sweden. The concept of the speech code is closely connected to the requirements for text-to-speech conversion.

I shall now turn to some general aspects of theory and data, and more specifically about structure and data. Here follow some condensed statements.

Structure with data

Speech code

Data without structure

Noise

Structure without data

Abstraction

Structured data qualifies as a part of the speech code. But data without structure or any connecting reference is analogous to noise. An example would be when attempting to decipher a spectrogram without any supporting knowledge or when listening to speech with the speech mode of perception disconnected.

Structures without data remain abstractions, but they could be supported by intuitive guesses to be verified later. This might be said about Noam Chomsky's theme that the child is born with some degree of linguistic competence. Phonology should be a guide to phonetics, contributing with structure alone. As a reaction, the sub-field of laboratory phonology has developed with special meetings to be held. Can we now expect a happy marriage between phonetics and phonology, or should we maintain a respectful divorce? (I am quoting Peter Ladefoged arguing with Morris Halle).

Anyhow, today after more than half a century of research, we have not been able to document anything like a basic version of the speech code for any language. We have gross views of frames needed for structuring, but our data is incomplete or antiquated.

Phonetic studies are not only concerned with structures and data. We are aided by principles of more global nature. Ken Stevens (1972) has contributed with a theory of the quantal nature of speech, based on stable articulatory regions, which do not require precise positioning of the articulators. Together with Anita Kruckenberg I have developed a theory of quantal timing in speech (Fant & Kruckenberg, 1996) and in poetry reading (Kruckenberg & Fant, 1993). A more recent contribution of Ken Stevens (2003) is a concept of universal phonological features based on landmarks from structured spectrographic data.

In my view prosody has been a neglected area in speech technology but is now gaining importance (Dutoit, 1997; Sagisaka et al., 1997). Expressive features have recently been treated by Campell, 2004). The major effort has been in studies of intonation and stress (Gårding, 1989, 1983; Fant et al., 2000; Fant & Kruckenberg, 2004ab; Hirst, 2004) but there remains much to be learned about the interaction between syntax and prosody (Bailly, 1989) and between prosody and the segmental frame. Lindblom (1989) has coined the terms hyper versus hypo speech, in other words clear versus reduced forms. A related aspect is the role of the syllable as an organizing unit, (Fujimura, 1992).

In closing up this overview I intend to focus on a basic theme. In quest of the speech code, we are faced with issues concerning invariance and variability. This was the subject of a conference arranged here at MIT (Perkell & Klatt, 1986). My view on this subject is inspired by Roman Jakobson. Invariance exists in a relational sense only, to be tested "*ceteris paribus*", that is in the same context. Absolute invariance is a property of the perceptual-cognitive process induced by linguistic competence rather than a property of the physical form.

Invariance ceases to be a problem if we systematically develop rules for structuring variability of all kinds, not only language, dialectal and contextual variations, but also variations specific to speaker, speaking style and emotions. Much more effort is needed to develop the code and make it available for practical applications as well as for the advance of general phonetics and linguistics. It is only with a profound knowledge of the speech code and human behaviour that we can realize ultimate goals of advanced and reliable systems.

This is an opinion that I share with many of you and especially with Ken. His large and productive work will be a guide for the future.

References

- Agelfors, E. & Risberg, A. (1991) Speech perception abilities of patients using cochlear implants, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 3/1991, 29-40.
- Badin, P. & Fant, G. (1984) Notes on vocal tract computation, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 2-3/1984, 53-108.
- Badin, P. (1989) Acoustics of voiceless fricatives, Production theory and data, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 3/1989, 33-55.
- Badin, P. & Fant, G. (1989) Fricative modelling: Some essentials, In *Proc. of Eurospeech 89*, Paris (edited by J. Tubach and J.J. Mariani), Vol. II, 23-26.
- Bailly, G. (1989) Integration of rhythmic and syntactic constraints in a model of generation of French prosody, *Speech Communication*, 8, 137-146.
- Beskow, J. (2003) *Talking Heads - Models and applications for multimodal speech synthesis*, Doctoral Thesis, Department of Speech, Music and Hearing, KTH, Stockholm.
- Campell, N. (2004) Expressive speech: simultaneous indication of information and affect, In *From Traditional Phonology to Modern Speech Processing. Festschrift for prof Wu Zongji* (edited by G. Fant, H. Fujisaki, J. Cao & Y. Xu) Foreign Language Teaching and Research Press, Beijing, 49-58.
- Chiba, T. & Kajiyama, M. (1941) *The Vowel its Nature and Structure*, Tokyo, Tokyo-Kaseidan.
- Carlson, R. Granström, B. & Fant, G. (1970) Some studies concerning perception of isolated vowels, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 2-3/1970, 19-35.
- Carlson, R. & Granström, B. (1975) A text-to-speech system based on a phonetically oriented programming language, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 1/1975, 1-4.
- Carlson, R. Granström, B. & Hunnicut, S. (1990) Multilingual Text-to-Speech Development and applications, In *Advances in Speech, Hearing and Language Processing* (edited. by A. W. Ainsworth) JAI Press.
- Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English*, Harper and Row, New York.
- Cooker, C. H., Umeda, N. & Browman, C. P. (1973) Automatic synthesis from ordinary English text, *IEEE Trans, Audio and Electroacoustics*, 21, 293-297.
- Dutoit, T. (1997) *An introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, London.
- Engwall, O. (2003) Combining MRI, EMA & EPG measurements in a three-dimensional tongue model, *Speech Communication*, 4, 303-329.
- Fant, G. (1948) *Analys av de svenska vokalljuden*, L.M. Ericsson protokoll H/P 1035 (52 pages).
- Fant, G. (1949) *Analys av de svenska konsonantljuden*, L.M. Ericsson protokoll H/P 1064 (139 pages).
- Fant, G. (1953) Speech Communication Research, *IVA, Swedish Academy of Engineering Sciences*, No. 8, 331-337.

- 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*, Mouton, The Hague.
- Fant, G. (1962) Descriptive analysis of the acoustic aspects of speech, *LOGOS*, Vol 5, No. 1, 3-17.
- Fant, G. (1964) Auditory patterns of speech, In *Models for the Perception of Speech and Visual Form*, Boston, Mass., Nov. 11-14, 1964.
- Fant, G. (1968) Analysis and synthesis of speech processes, In *Manual of Phonetics* (edited by B. Malmberg) Amsterdam, North-Holland Publ. Co. Chapt. 8, 173-276.
- Fant, G. (1972) Q-codes, *International Symposium on Speech Communication Ability and Profound Deafness, Stockholm 1970* (edited by A.G. Bell Ass. for the Deaf) Washington DC, 261-268.
- Fant, G. (1978) Vowel perception and specification, *Rivista Italiana di Acustica II*, 69-87.
- Fant, G. (1995) Speech related to pure tone audiograms, In *Profound deafness and speech communication* (edited by G. Plant and K.E. Spens), Whurr Publ. Ltd, London, 299-305.
- Fant, G. (1996) Historical notes. Response to interview questions posed by Louis-Jean Boe and Pierre Badin, *KTH-TMH manuscript*, 15 pages.
- Fant, G. (1997) The voice source in connected speech, *Speech Communication* **22**, 125-139.
- Fant, G. (2004a) More than half a century in phonetics and speech research. (To be published in *Gunnar Fant Selected Writings*, Kluwer Academic Publisher).
- Fant, G. (2004b) On the Speech Code, (To be published in *Gunnar Fant Selected Writings*, Kluwer Academic Publisher).
- Fant, G. & Kruckenberg, A. (1996) On the quantal nature of speech timing, *Proceedings of the International Conference on Spoken Language Processing*. 1996, 2044-2047.
- Fant, G. & Kruckenberg, A. (2004a) Analysis and Synthesis of Swedish Prosody with Outlooks on Production and Perception, In *From Traditional Phonology to Modern Speech Processing. Festschrift for prof Wu Zongji* (edited by G. Fant, H. Fujisaki, J. Cao & Y. Xu) Foreign Language Teaching and Research Press, Beijing, 73-95.
- Fant, G. & Kruckenberg, A. (2004b) An integrated view of Swedish prosody, Voice production, perception and synthesis (To be published in *Gunnar Fant Selected Writings*, Kluwer Academic Publisher).
- Fant, G., Kruckenberg, A. & Liljencrants, J. (2000) The Source-Filter Frame of Prominence, *Phonetica* **57**, 113-127.
- Fant, G., Kruckenberg, A., Gustafson K. & Liljencrants, J. (2002) A new approach to intonation analysis and synthesis of Swedish, *Speech Prosody 2002, Aix en Provence*. Also in *Fonetik 2002, TMH-QPSR*, 2002.
- Fant, G., Liljencrants, J. & Lin, Q. (1985) A four-parameter model of glottal flow, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 4/1985, 1-13.
- Fant, G. & Lindqvist-Gauffin, J. (1968) Pressure and gas mixture effects on divers speech, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 1/1968, 7-17.

- Fant, G. & Martony, J. (1962) Speech synthesis instrumentation for parametric synthesis (OVE II), *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 2/1962, 18-24.
- Fant, G. & Stevens, K. N. (1960) Systems for speech compression, *Fortschritte der Hochfrequenztechnik*, Vol. 5. Akademische Verlagsgesellschaft M.B.H., Frankfurt am Main, 229-262.
- Faulkner, A., Fourcin, A. J. & Moore, B. C. J. (1990) Psychoacoustic Aspects of Speech Pattern Coding for The Deaf. *Acta Oto-Laryngologica, Suppl.* 469, 172-180.
- Flanagan, J. L. (1965) *Speech Analysis Synthesis and Perception*, New York, Springer.
- Fletcher, H. (1929) *Speech and Hearing*, New York.
- Fourcin, A. J. (1981) Laryngographic assessment of phonatory function, *ASHA Reports* 11, 116-127.
- Fujimura, O. (1992) Phonology and Phonetics. A syllable Based Model of Articulatory Organization. *Acoustical Society of Japan (e)* 13, 39-48.
- Fujisaki, H. (2004) Information, prosody, and modelling with emphasis on the tonal features of speech. *From Traditional Phonology to Modern Speech Processing, Festschrift for prof. Wu Zongji* (edited by G. Fant, H. Fujisaki, J. Cao & Y. Xu) Foreign Language Teaching and Research Press, Beijing, 111-128.
- Joos, M. (1948) Acoustic Phonetics, *Language* 24,1-136.
- Branderud, P., Galyas, K. & Svensson, S. (1974) Experimental device for control of intonation with an artificial larynx. A status report, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 4/1974, 38-41.
- 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).
- Gobl, C. (2003) *The voice source in speech communication*, Thesis at the Department of Speech, Music and Hearing, KTH.
- Gårding, E. (1989) Intonation in Swedish, *Working papers, Lund University Linguistics Department* 35, 63-88. Also in *Intonation Systems* (edited by. D. Hirst and A. Di Cristo), Cambridge University Press 1998, 112-130.
- Gårding, E. (1993) On parameters and principles in intonation analysis, *Working Papers Lund University, Dept. of Linguistics*, 40, 25-47.
- Hanson, H. & Chuang, E. (1999) Glottal characteristics of male speakers. Acoustic correlates and comparison with female data, *Journal of the Acoustic Society of America* 106/2, 1064-1077.
- Hirst, D. (2004) Speech prosody: from acoustics to interpretation, In *From Traditional Phonology to Modern Speech Processing, Festschrift for prof. Wu Zongji* (edited by G. Fant, H. Fujisaki, J. Cao & Y. Xu) Foreign Language Teaching and Research Press, Beijing, 177-188.
- Holmes, J. (1961) Notes on synthesis work, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH* 1/1961, 10-12.
- Hunnicutt, S. (1984) Bliss symbol-to-speech conversion: Blisstalk, *Journal of the American Voice I/O Society*, Vol. 3, Also in *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 1/ 1984, 58-77.

- Hunnicutt, S. (1989) Using syntactic and semantic information in a word prediction aid, In *Eurospeech 89, Paris, Vol 2* (edited by J. Tubach & J. J. Mariani) CPC Consultants Ltd., Edinburgh, U.K. 191-193.
- Ishizaka, K. & Flanagan, J. L. (1972) Synthesis of voiced sounds from a two-mass model, *Bell System Technical Journal*, 51, 1233-1268.
- 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, MIT press, seventh edition, 1967.
- Kelso, J. A., Saltzman, E. L. & Tuller, B. (1986) The dynamical perspective on speech production; Data and theory, *Journal of Phonetics*, 14, 29-59.
- Klatt, D. H. (1980) Software for a cascade/parallel formant synthesizer, *Journal of the Acoustic Society of America*, 67, 971-995.
- Klatt, D. (1987) Review of text to speech conversion for English, *Journal of the Acoustical Society of America*, 82, 737-793.
- Kozhevnikov, V. A. & Chistovich, L. A. (1965) *Speech Articulation and Perception*, US Dept. of Commerce Publ. JPRS:30 543;TT: 65-31233 (translated from Russian). See also *Speech Communication 1985*, Vol. 4, devoted to Ludmilla Chistovich.
- Kruckenber, A. & Fant, G. (1993). Iambic versus trochaic patterns in poetry reading, *Nordic Prosody VI*, Stockholm, 1993, 123-135.
- Ladefoged, P. (1967) *Three areas of experimental phonetics*, Oxford University Press.
- Ladefoged, P. (1975) *A course in phonetics*, 2nd ed. 1985, 3rd ed. 1993, Orlando; Harcourt Brace.
- Lawrence, W. (1953) The synthesis of Speech from Signals which have a Low Information Rate, In *Communication Theory*, (edited by W. Jackson), London.
- Lieberman, A. M., Delattre, P.C. & Cooper, F. S. (1952) The role of selected stimulus variables in the perception of the unvoiced consonants, *American Journal of Psychology*, 65, 497-516.
- Lieberman, A. M. & Mattingly, I. G. (1985) The Motor Theory of Speech Perception, *Cognition*, 21, 1-36.
- Lieberman, A. M. (1986) Comments to the presentation of G. Fant in *Invariance and variability of speech processes* (edited by J.S. Perkell and D. Klatt) Lawrence Erlbaum Ass. Publ., 490-492.
- Liljencrants, J. Fant, G. & Kruckenber, A. (2000). Subglottal pressure and prosody in Swedish. *Proc. International Conference on Spoken Language Processing, 2000*, Beijing.
- Lin, Q. & Fant, G. (1989). Vocal tract area function parameters from formant frequencies, In *Proc. of Eurospeech 89, Paris* (edited by J. Tubach and J. J. Mariani), Vol. 2, 673-676.
- Lin, Q. & Fant, G. (1992) An articulatory speech synthesizer based on a frequency domain simulation of the vocal tract, *IEEE-ICASSP 1992*, San Francisco, paper 173, 1992.
- Lindblom, B. (1989) Explaining phonetic variation; A sketch of the H&H theory, In *Speech Production and Modelling*, (edited by W.J. Hardcastle & A. Marchal) Kluwer Academic Publishers, Netherland, 403-439.
- Mattingly, I. G. (1968) Experimental methods for speech synthesis by rule, *IEEE Transactions on Audio and Electroacoustics*, AU-16, 198-202.

- McGowan, R. S. (1994) Recovering articulatory movements from formant frequency trajectories using task dynamics and a generic algorithm; Preliminary model tests, *Speech Communication*, 14, 19-48.
- Ohala, J. J. (1990) Respiratory activity in speech. In *Speech Production and Speech Modelling*, (edited by W.J. Hardcastle and A. Marchal) Kluwer Academic Publishers, 23-53.
- Ohala, J. J., Bronstein, A. J., Busà, M. G., Lewis, J. A. & Weigel, W. F. (editors) (1999) *A Guide to the History of the Phonetic Sciences in the United States*, University of California, Berkeley.
- Öhman, S. (1967) Word and sentence intonation: a quantitative model, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 2-3/1967.
- Olive, J. P. (1977) Rule synthesis of speech from diadic units, *ICASSP*, Vol. 77, 568-570.
- Öster, A-M. (1989) Applications and experiences of computer based speech training, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 4/1989, 37-44.
- Perkell, J. S. (1969) Physiology of speech production; Results and implications of a quantitative cineradiographic study, Research monograph No. 53, Cambridge MA, MIT Press.
- Perkell, J. S. & Klatt, D. (1986) (editors) *Invariance and variability of speech processes*, Lawrence Erlbaum Ass. Publ.
- Perkell, J. S. & Matthies, M. L. (1992) Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within and cross-subject variability, *Journal of the Acoustical Society of America*, 91, 2911-2925.
- Peterson, G. E. & Barney, H. L. (1952) Control methods used in the study of vowels, *Journal of the Acoustical Society of America*, 24, 2, 175-184.
- Pickett, J. M. (1962) Tactual vocoder as an aid in speech transmission to the deaf, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH* 1/1962, 23-29.
- Potter, R. K., Kopp, A. G. & Green, H. C. (1947) *Visible Speech*, New York.
- Risberg, A. & Agelfors, E. (1984) On the relation between frequency discrimination ability and the degree of hearing loss, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 4/1984, 59-71.
- Rothenberg, M. (1968) *The breath stream dynamics of simple-released plosive production*, Bibliotheca Phonetica No 6, Basel, S. Karger.
- Sachs, M. B. & Young, E. D. (1979) Encoding of steady state vowels in the auditory nerve; Representation in terms of discharge rate, *Journal of the Acoustic Society of America*, 66, 470-479.
- Sagisaka, Y., Campell, N. & Higuchi, N. (1997) *Computing Prosody*, Springer Verlag.
- Shannon, C. E. & Weaver, W. (1949) *The mathematical theory of communication*, Urbana.
- Sluijter, A. M. C. & van Heuven, V. J. (1996) Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America*, 100/4, 2471-2484.
- Spanias, A. (1994) Speech coding: A tutorial review, *Proceedings of the IEEE*, 82, 1541-1582,
- Stevens, K. N. (1952) *The perception of speech sounds shaped by resonance circuits*, ScD. dissertation, Massachusetts Institute of Technology, Cambridge.

- Stevens, K. N. (1972) The quantal nature of speech. Evidence from articulatory-acoustic data, In *Human communication: a unified view* (edited by P. B. Denis & E. E. David Jr), New York, McGraw Hill, 51-66.
- Stevens, K. N. (1998) *Acoustic Phonetics*, Cambridge MA, MIT Press.
- Stevens, K. N. (2003) Acoustic and perceptual evidence for universal phonological features. *Proc. XVth International Congress of Phonetic Sciences*, Barcelona, 33-38.
- Stevens, K. N. & Bickley, C. A. (1991) Constraints among parameters simplify control of the Klatt formant synthesizer, *J. of Phonetics*, Vol 19, No.1, 161-174.
- Stevens, K. N., Kasowski, S. & Fant, G. (1953) An electrical analog of the vocal tract, *Journal of the Acoustical Society of America* 25, 734-742.
- Sundberg, J. (1987) *The Science of the Singing Voice* (author's translation of Röstlära), DeKalb: N. Ill. Univ. Press.
- Sundberg, J. Andersson, M. & Hultqvist, C. (1999) Effects of subglottal pressure variations on professional baritone singers' voice sources. *Journal of the Acoustical Society of America*, 105/3, 1999.
- Titze, I. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *Journal of the Acoustical Society of America* 85/2, 901-906.
- Wilson, B. S. (1993) Cochlear implants, audiological foundations, In *Signal processing* (edited by R.S. Tyler), Singular Publishing Group, Inc, San Diego, California.