SOUND SPECTROGRAPHY

GUNNAR FANT

1. ACOUSTIC PHONETICS

The scientific study of speech has developed considerably during the last 24 years since the last international congress of phonetic sciences.

The traditional articulatory or rather physiological phonetics is being supplemented more and more by acoustic phonetics. In part this is due to the increasing interest in speech research from communication engineering quarters (11). Progress in methods of speech transmission and speech data processing presupposes advanced knowledge of human speech on all levels of specification including speech production, acoustic speech structure, and speech perception. The basic notion of speech as signals which are transmitted through successive stages within both the listener and speaker part of the system and transformed into different physical forms by a coding process in each of these stages is a commonly used model (28), which stems from information theory.

Acoustic phonetics centers around the speech wave as defined by the sound pressure fluctuations affecting a microphone in front of the speaker. Our most important means of studying the physical structure of the speech wave is through spectrographic records but there exist many supplementary means of analysis, e.g. through oscillograms and records of speech intensity and voice fundamental frequency, etc.

The acoustic speech wave is more open for insight than any other physical manifestation of a human utterance. This potentiality is of a considerable advantage once the language of Visible Speech is known (32, 33). Beginners will certainly have difficulties – those who are more sophisticated will have to admit that there still remains very much to learn about this language. A maximally complete description can be achieved at the price of an overdetailed representation only.

Spectrographic records may possess undeniable artistic qualities but they are not studied for the sake of their own beauty. Acoustic phonetics aims at relating speech wave data to any other observable aspect of the speech act. Of primary interest is to relate to the speech wave data to linguistically defined signs and categories belonging to the message aspect (15). The success of such a task is much dependent on our insight in the codes whereby the sound patterns may be related to the function of the human speaking organs (3, 10, 41) and to the capacities of the hearing mechanism.

These general constraints of what can be said and what can be heard are of basic importance for the development of specificational schemes. The rules for physiological interpretation of spectrograms provide the key for learning to read Visible Speech.

With access to this code the reading of spectrographic records provides a very interesting and stimulating insight in the speaker's particular oral behavior in producing a specific utterance. In this respect the spectrographic records supplement the auditory transcription of an informant's speech. It is true that the predictability of articulation from speech wave patterns is not as single valued as translating from speech production to speech waves. However, the possibilities of compensatory modes of articulation providing the same speech wave patterns are not very serious if sufficient evidence is taken into account. Here the interests of traditional articulatory phonetics (24) and modern acoustic phonetics (6, 16) meet. The more experienced a phonetician is in articulatory phonetics, the easier it will be for him to learn the elements of Visible Speech.

A knowledge of the general constraints of the auditory system as a sound analyzer and its response to speech-like stimuli is much needed for a proper evaluation of the relative importance of various sound pattern aspects. Synthetic speech has been of considerable value for studying these criteria (12, 17, 18) but there are many other important techniques for speech pattern evaluations of e.g., by means of small differential eliminations or substitutions of the sound within specific time (43, 44) and frequency intervals.

The raw data from a spectrogram is never subjected to an auditory evaluation in all its details. A single harmonic does not have an independent auditory significance. The spectrographic patterns must first be expressed in a simplified form pertaining to the essentials of the energy distribution in frequency and time. The specific constraints on such energy distributions imposed by the properties of the speech generating mechanism, for instance the relations between formant intensities and the pattern of formant frequencies (5, 7, 10) allow for valuable simplifications in the specifications and provide a guide for a natural choice of parameter for a systematic variation of the composition of speech-like stimuli in speech perception experiments.

However, the procedure sketched above is more of a theoretical strategy than an established procedure in speech research. Our knowledge of speech wave structure and of speech production and speech perception is still rather incomplete. Although there has been considerable advance in speech research the last 15 years we are still in a rather early stage of development. It was stated by Joos in 1948 (16) and I expressed the same opinion in Oslo 4 years ago (6) that acoustic phonetics is still in its infancy. In view of the fact that speech spectrography has existed for a long time it is astonishing that so few large scale experimental studies have been accomplished. The reason for this slow development is in part due to instrumental difficulties, in part due to a lack of a rationale for specification and interpretation of experimental results. The investigator too easily drowns in a sea of details of unknown significance if he attempts to make use of all observable data.

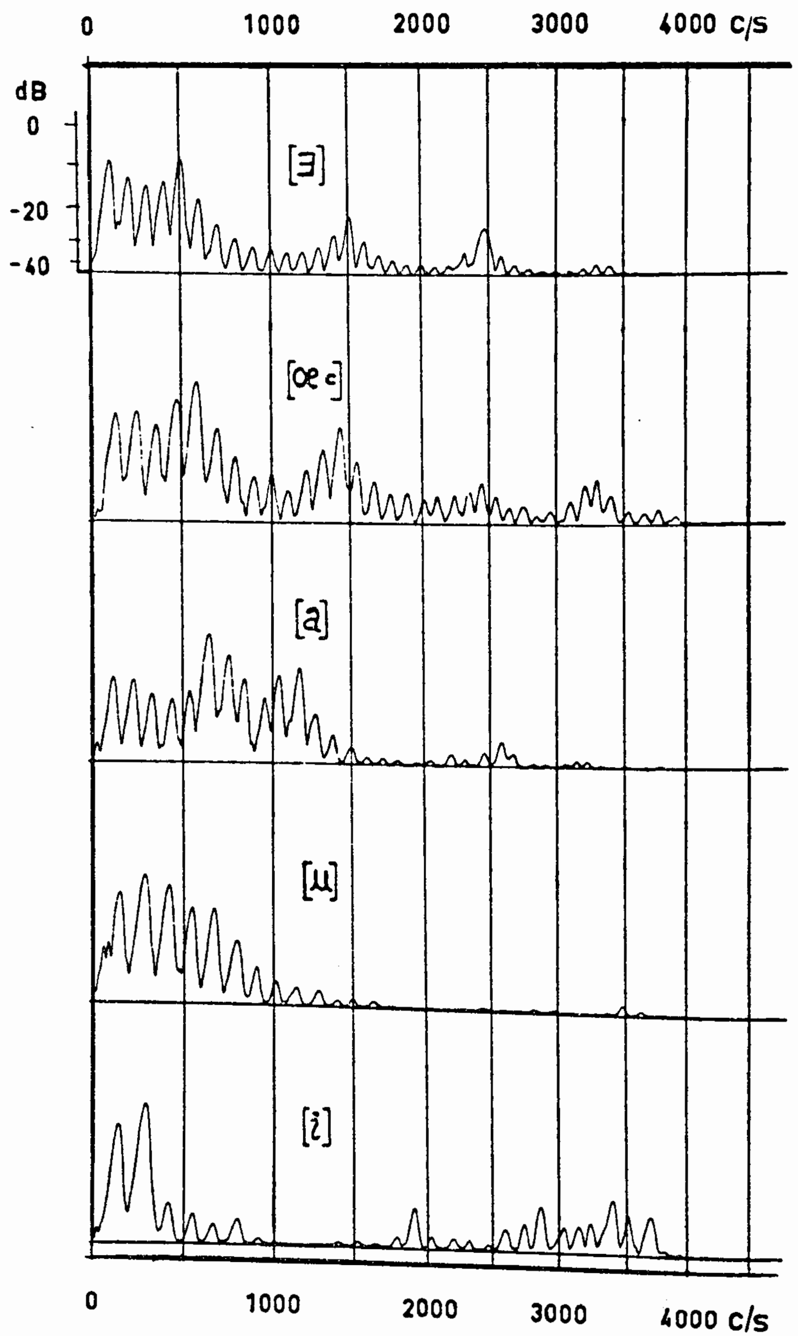

Fig. 1. Sweep frequency analysis of sustained speech sounds. The sign [ɜ] denotes a neutral reference vowel produced from OVE I. The 4000 c/s range has been run through in 3 seconds with a 31 c/s-wide filter. Mingograph recorder.

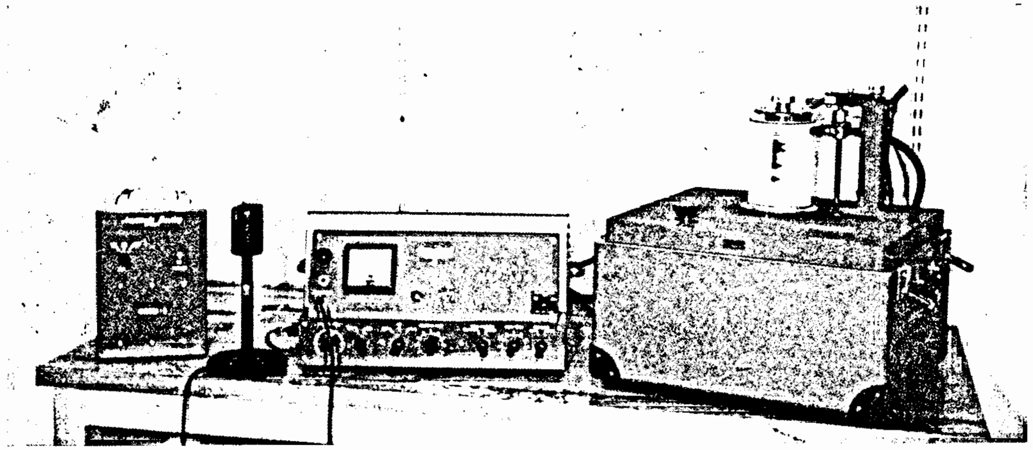

Fig. 2. The Sona-Graph spectrograph.

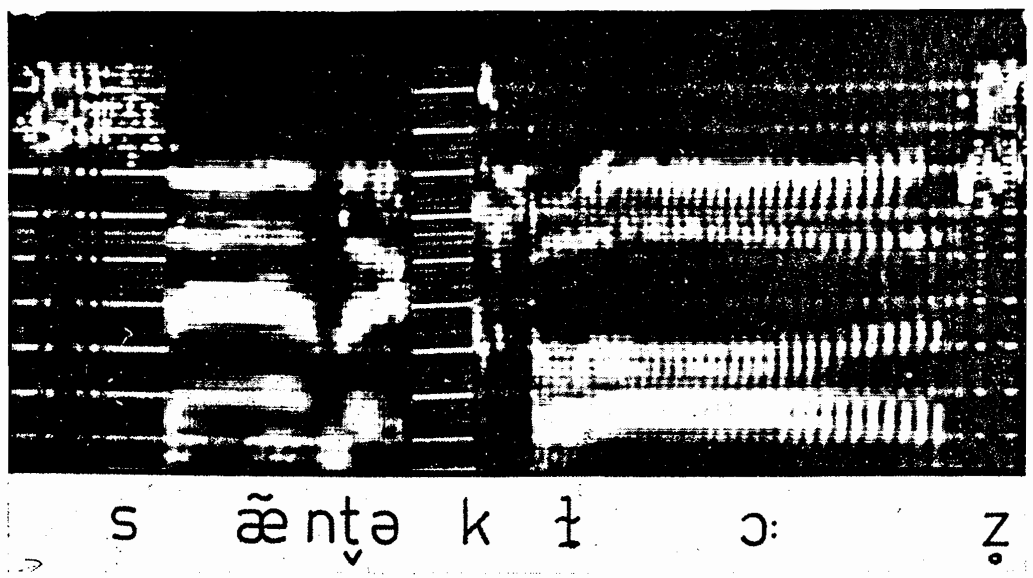

Fig. 3. Spectrogram taken with the 48-channel spectrograh. Text is "Santa Claus". Compare with Fig. 10.

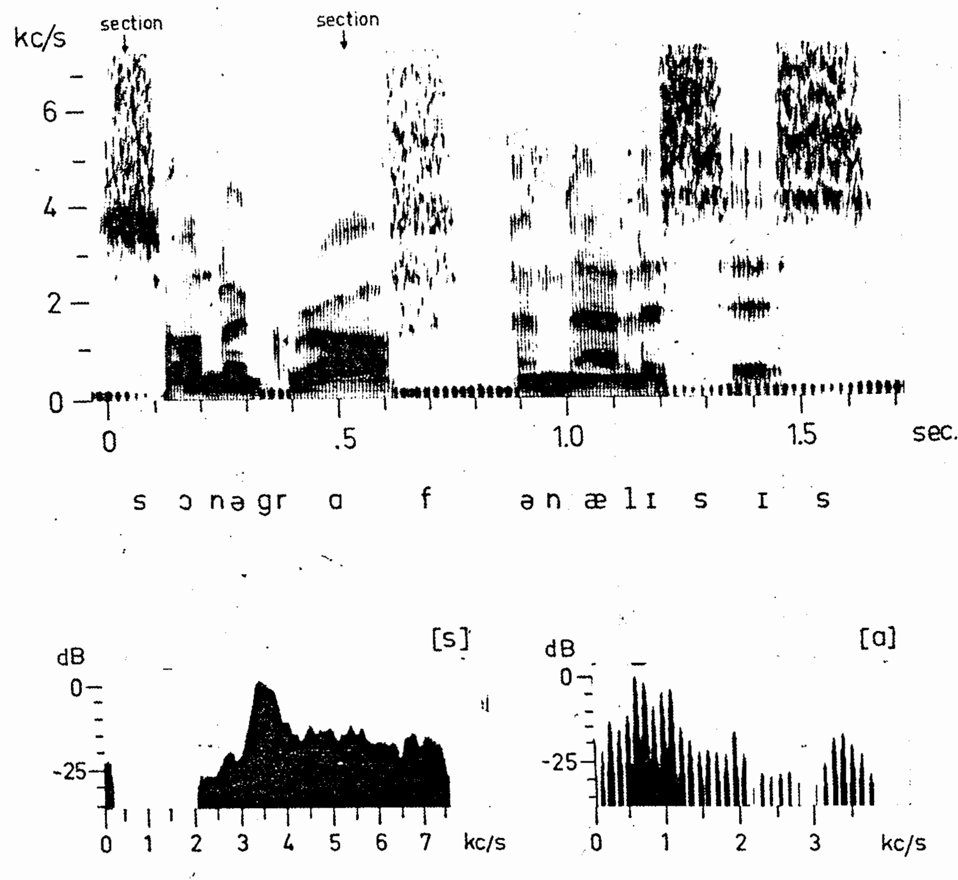

Fig. 4. Broad-filter spectrogram and narrow-filter sections taken with the Sona-Graph.

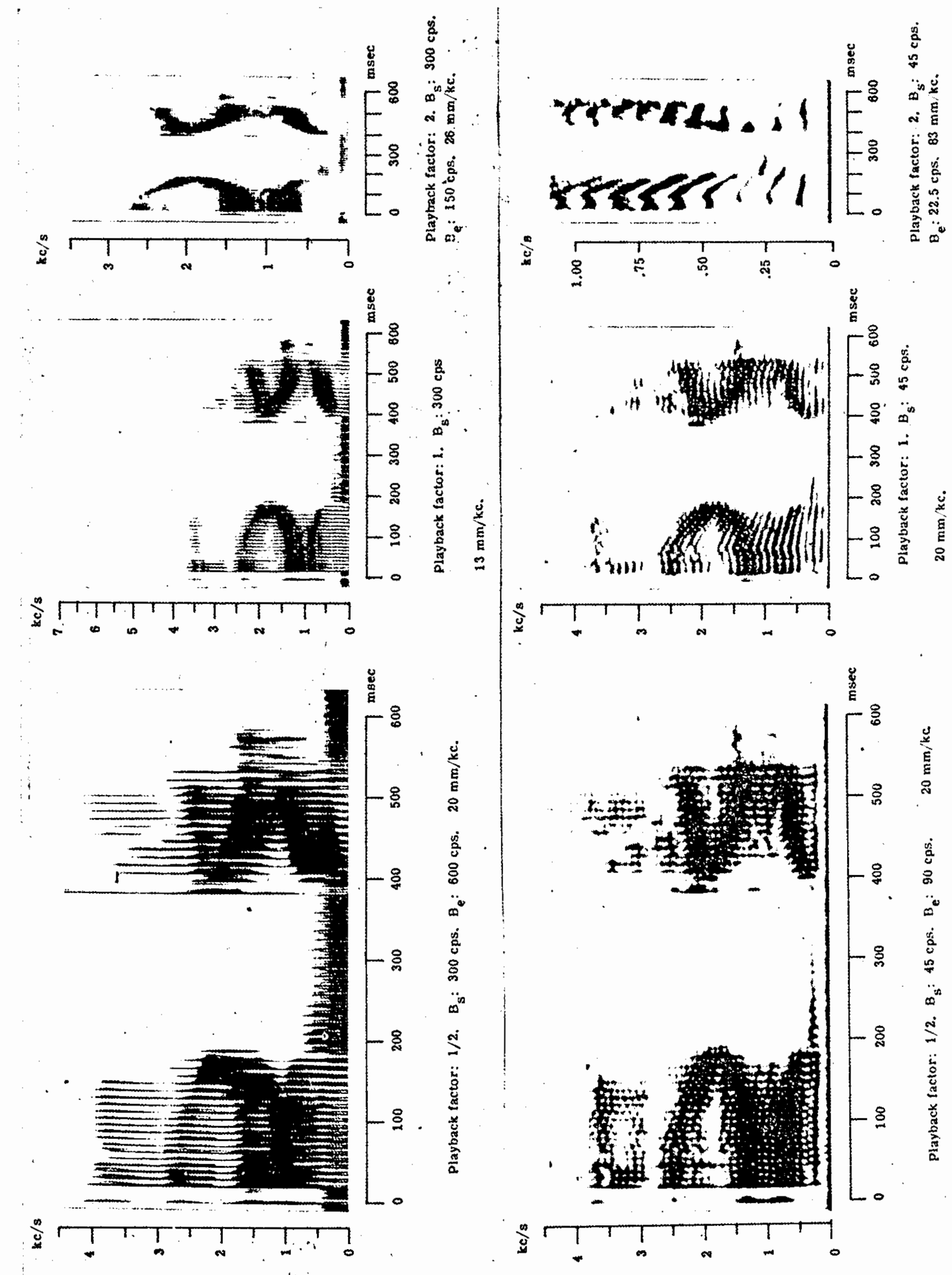

Fig. 5. Spectrograms illustrating the effects of varying the speed of the tape-recorder from which the speech material is transferred to the Sona-Graph. Broad-filter analysis above, narrow-filter analysis below. Half speed to the left, normal speed in the middle, and double speed to the right. The utterance is the VCV syllable "agga".

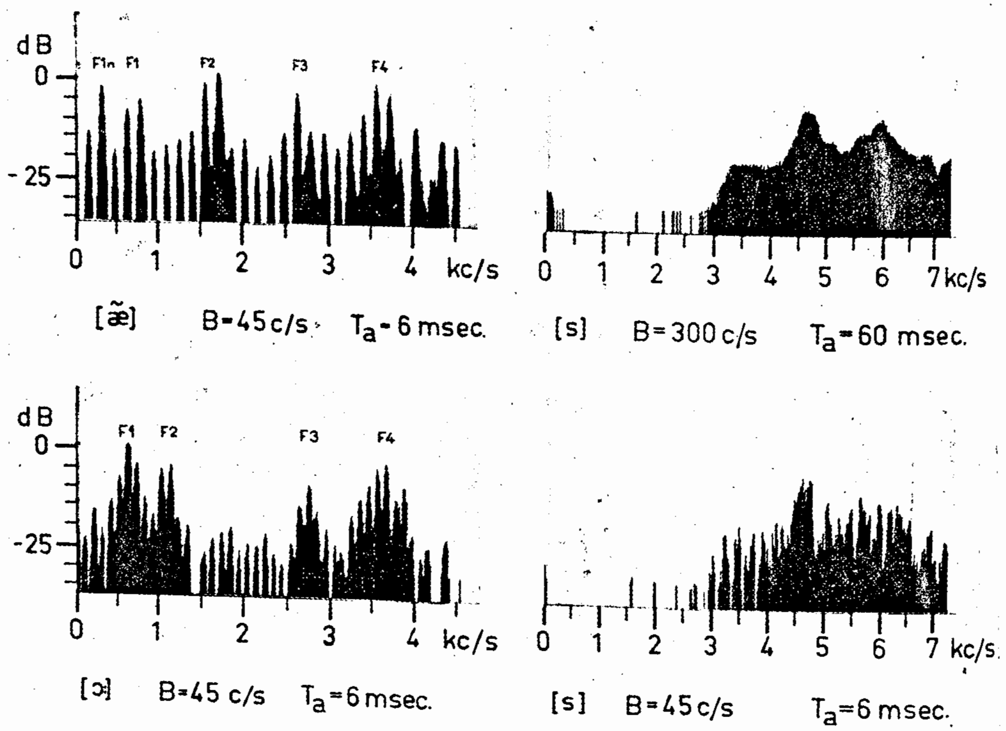

Fig. 6. To the left normal narrow-filter Sona-Graph sections of the vowels [æ̃] and [ɔ:]. Observe the split first formant of the nasalized [æ̃]. – Upper-right section pertains to the fricative [s] analyzed with the broad filter and an extra large integration time constant. Lower-right section pertains to the same [s] sample but analyzed with the narrow filter and the normal integration time constant of the instrument. – The rugged fine structure of the lower section is an instrumental artefact related to the low product of analysis filter bandwith B and integration time constant T_a.

2. SPECTRUM ANALYSIS TECHNIQUES

The sound spectrograph is perhaps the most important research tool in present-day speech research both in phonetic laboratories and at communication engineering institutions. This statement may not be agreed upon by all those present here. Some may speak up for speech synthesizers, others for high speed digital computers, especially in view of the fact that large, versatile digital computers can be made to simulate both analysis and synthesis processes.

However, the common feature of analysis and synthesis methods lies in the spectral representation of speech waves. It is a well established fact that the speech spectrum provides a much more useful reference than oscillographic displays.

The human ear is an effective sound analyzer that we would like to be able to duplicate instrumentally in all its major functions. The major trouble is the brain part of the system that determines important aspects of the auditory functions that have as yet not been sufficiently investigated in psychoacoustic experiments. The mere fact that the selectivity and discriminability of time events are not constant but appear to be functions of the overall type of stimulus indicates the level of the difficulties we meet with when attempting to predict the auditory functions in response to the rather complicated speech wave structure.

Aside from the difficulties of reaching a complete insight in the black box we label hearing it is appearent that simplified spectrum analysis by hearing has been of considerable importance in the early history of experimental phonetics, e.g. in the works of Helmholtz (14), Paget (26), Stumpf (42). I have the greatest respect for the accurate formant frequency analysis of Danish vowels performed by Smith (37) by hearing alone.

Fourier analysis of oscillographically determined wave forms was one of the earliest methods of objective analysis. I would like to mention the pioneering work of the Finnish phonetician Pipping (31) who derived the complete amplitude and phase spectra of vowels by this method as early as 80 years ago. Fourier analysis of speech wave forms now incidently enjoys a renaissance thanks to high speed computer techniques (23).

Sweep frequency analysis of sustained sounds has been an important technique but has the obvious drawback that its use is limited to prolonged isolated sounds and thus excludes the study of connected speech. The sweep frequency analyzers currently in use 30 years ago required a subject that could maintain a steady sung vowel for a period of one or two minutes. I have in mind the analysis performed by Barczinski and Thienhaus (1) and by Sovijärvi (38). In modern design the time of analysis is brought down to the order of 4 seconds for the analysis of a 4000 c/s frequency range with a 32 c/s-wide filter (6) and the method may supplement standard sound spectrography for detail studies of well reproducible stationary sounds. Amplitude versus frequency spectra recorded by the sweep frequency method are exemplified in Figure 1. Such recordings generally display a clear and well defined harmonic structure.

In the Key Electric Sona-Graph and other modern sound spectrographs the frequency location of the analyzing filter is shifted through the frequency range of interest just as in the sweep frequency method. The essential difference is that the speech to be analyzed is played through the spectrograph a large number of times during the course of analysis. The memory for storage of the speech material is a magnetic recording drum which has a time span of 2.4 seconds. The time needed for the analysis process is of the order of 5 minutes in the Sona-Graph.

The origin of the Sona-Graph is the sound spectrograph developed for the Visible Speech Project at the Bell Telephone Laboratories (32, 33, 29). Several laboratory constructions of spectrographs have been made according to these principles. The most exact and versatile apparatus of this type I have seen is that at Gordon E. Peterson's Laboratory, University of Michigan, Ann Arbor, Michigan.

The common drawback of these spectrographs is the relative long time required for the processing but they have the benefit of a very detailed spectral portrayal and an optimally fine resolution in time allowing a detailed insight in the time varying spectrum of connected speech. The restrictions in the speed of analysis have been overcome in a recent design of Gill (13) employing a circulating memory which the speech continuously enters and leaves after having been rotated round the memory loop many times at a very high speed. This causes a translation of the speech material by a very large factor in frequency at a retained time scale allowing a very large reduction of the time required for analysis. A related method making use of phase coincidence (2) filters has been tried in an American spectrograph, the Simaramic, designed for studies of under-water sounds.

One important class of high speed sound analyzers makes use of a large set of band-pass filters working in parallel the outputs of which are scanned by a rotating switch. Our 48-channel analyzer (6) designed by H. Sund is one of the few analyzers of this type in use for phonetic research. A recent design according to similar principles is that of P. Denes for University College, London. The time-frequency-intensity sound spectrum is recorded on continuous photographic film in these two analyzers and that of Gill (13). The advantage is that very long pieces of speech may be analyzed in a very short time. This is an advantage when one merely intends to use the spectrogram as a reference record for a large speech material. The frequency resolution of these high speed analyzers, with the possible exception of that of Gill is not quite as good as in the Sona-Graph, compare for instance Figure 3 with that of Figure 7. It should be appreciated that these high speed analyzers can in a short time produce a spectrographical material that may keep a phonetician busy for a life time.

The sound spectrograph of the Sona-Graph type and its use in phonetic research has been extensively described in the literature (29, 32, 33). A few practical hints on the use of the Sona-Graph will be given here in addition to those discussed by Lindblom in his paper for this congress (21). The normal type of spectrograms and sections are illustrated by Figure 4. It is possible to expand the time scale and compress simultaneously the frequency scale. This is achieved by a play-back speed reduction of a factor of 2 when playing speech material into the Sona-Graph. This process will increase the effective bandwidth of the analysis filter to twice the normal value which in case of the broad-band filter implies a $2 \times 300 = 600$ c/s effective width. This feature is sometimes of value for avoiding the appearence of a harmonic fine structure in "broad-band analysis" of high pitched female voices. Another apparent advantage is the increased time resolution which emphasizes the periodicity aspect of low pitched male voices. Similarly, a play-back speed increase will compress the time scale and expand the frequency scale which is of some benefit in voice fundamental frequency analysis. The effective reduction of the filter width will provide an increased resolution in the harmonic analysis.

There are several practical precautions which should be made in using the Sona-Graph. One is to prevent overloading by careful check of the input signal level to the instrument (see Lindblom's paper (21)). Spectral sections of fricatives should not be taken with the narrow-band filter since the overlaid statistical fluctuation will be considerable. Instead the broad-band filter should be used and the integration time of the intensity processing part of the analyzer should be increased by a factor of 2–3.

Other suggestions for instrumental improvements would include a reshaping of the pre-emphasis filter HS to provide a smooth 6-dB/octave rise from 200 c/s to 5000 c/s instead of the present fairly constant response at frequencies below 800 c/s followed by a rapid rise in the response. This change will contribute to clean up the F1 and bass-band region of the spectrogram thus avoiding excessive marking in this region. Another suggestion is to replace the present mechanical sampling switch for sectioning by an electronically gated switch. This would improve the accuracy in sampling. Most spectrographs are nowadays supplied with a frequency scale expander, which is very helpful for intonation studies based on narrow-band harmonic analysis.

3. INTERPRETATION OF SPECTROGRAPHIC DATA

It was stated in the first chapter that we cannot make full use of all observable data from a spectrogram, at least not on a quantitative basis. The purpose of analysis is rather to obtain answers to specific questions of a phonetic nature. Quantitative measurements have generally been limited to formant frequencies of vowel-like sounds (7, 30) but this is by no means an inherent restriction of the instrument. Sound spectrographs deserve to be used more than they have been but they need all the engineering care they can get to function satisfactorily.

Some of the general problems encountered in taking formant frequency measures are ventilated in the paper by Lindblom. The accepted general definition of a formant is a maximum in the sound spectrum. Alternative definitions could be adopted for the formant frequency but it is nowadays commonly accepted to refer to the frequency

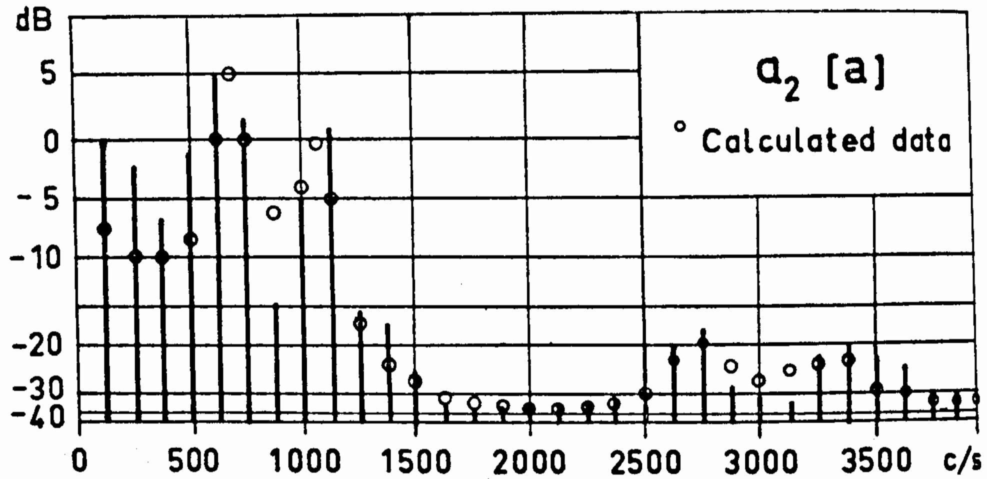

Fig. 7. Spectrum of a sustained vowel [a]. The calculated points have been derived from a set of preselected formant frequencies. (From *Ericsson Technics*, Vol. 15, No. 1, 1959).

of the corresponding vocal resonance (pole). In a well isolated formant the difference between the peak frequency of the spectrum envelope and the corresponding resonance frequency is of academic interest alone since it is much less than the uncertainty set by the harmonic structure.

The formant frequencies are not merely the basic correlates of vowel quality as judged from several studies. It can be shown that the essential shape features of a vowel spectrum are predictable from a knowledge of the formant frequencies (5), see also the paper by Mártony et al. (22). One of these basic rules interrelating formant frequencies and the intensity level of any part of the spectrum is that a decrease in formant frequency F_1 will cause a shift down in the spectrum level of the entire spectrum above F_1 at a rate of 12 dB per octave shift in F_1. The implications of this rule for phonetic theory, and syllable division in particular, is apparent. Articulatory narrowing is the cause of decrease in F_1 which is automatically followed by the above mentioned intensity reduction. At constant phonatory power, i.e. vocal source strength, the articulatory movements alone thus have the power of determining the time-variable intensity changes from a vowel to a following voiced consonant.

One important aspect of these relations is the converse. The intensity level at one part of the vowel spectrum compared with the levels at other parts of the spectrum carries some information on the frequencies of all formants. The accuracy in formant frequency measurements may thus be improved by taking into account data from the spectrum outside the region of the formant peak. This is practically accomplished by the spectrum matching technique, also known under the name of "analysis-by-synthesis", from the work at M.I.T. (40). The technique is examplified in Figure 7 which shows to what extent the measured harmonic line spectrum of a vowel [a] is predictable from the analytical model based (7) on the formant frequencies. The

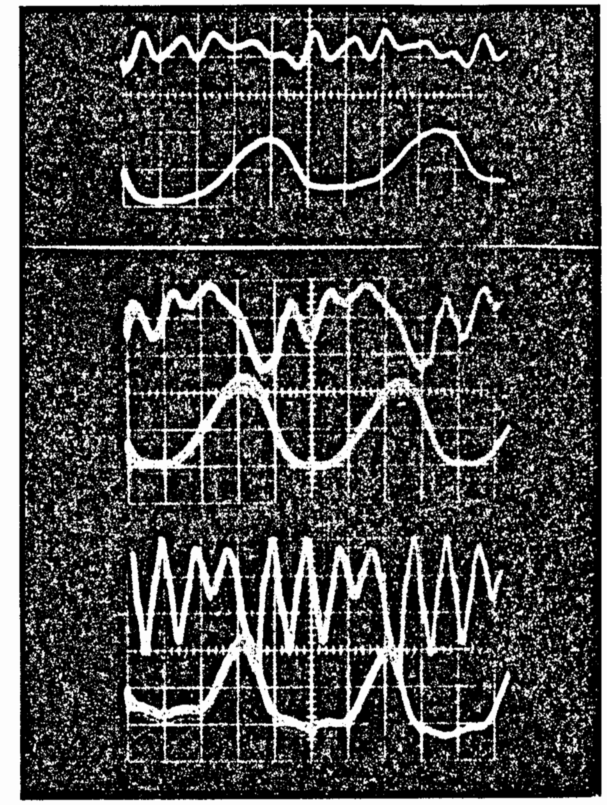

Fig. 8. Oscillograms of the vowel [æ] and of the regenerated glottal flow function. Low, medium, and high voice efforts are illustrated from the top to the bottom of the figure.

figure exemplifies the first stage of the matching based on the first guess of formant frequencies. The matching proceeds with a systematic variation of formant frequencies until the difference between the synthetic and natural spectrum is minimal.

The pronounced dip at 800 c/s which marks a deviation from the ideal model is a typical property of the vocal source spectrum of the particular speaker.

Spectrum matching according to these principles is outside the technical resources of most phonetic laboratories. The gross relations between spectrum shape and the pattern of formant frequencies should, however, be known by any phonetician attempting to collect data on formant frequencies of the separate vowels of a particular language or dialect. This knowledge constitutes an insurance against errors in the identification of a formant number. I know of several vowel studies which are invalidated by inconsistent formant specifications, e.g. F_3 labeled as F_2 etc. Simple continuity considerations can be relied upon for the practical evaluations.

The time domain correspondence of spectrum matching is the inverse filtering technique whereby each formant is compensated until the oscillographic display of the vowel shows a residue which is merely a pulse train at the rate of the voice fundamental. This method, if carefully planned, provides valuable information on the glottal functions (25, 8), see Figure 8, but the primary aim is generally to measure the formant frequencies. We do not yet have sufficient experience of this speech microscope[1] method to fully evaluate it but one can anticipate trouble in case of nasalized vowels containing two separate first formants, one nasal and one oral as in the [æ̃] of Figures 3 and 10.

The most powerful method in investigating the acoustic correlates of phonetic entities is to construct contrastive sentences differing only in terms of the category to be studied. The importance of the concept of distinctive features in phonemic theory is undeniable; no less is its importance in practical experimental work. Without the elimination of the conditioning effects due to the speaker's individuality and all contextual factors it is often difficult to get a clear view of the information bearing elements of the speech wave. An example of a minimum distinction, that of the continuant /v/ versus the interrupted /b/, is shown in Figure 9. The difference in the degree of articulatory closure accounts for the less attenuated second formant of /v/ as compared with /b/. This acoustic difference is paralleled by the higher frequency of F_1 in /v/ compared with /b/ in comformity with the general relations between articulatory narrowing, F_1, and the level of higher formants discussed above. This is a typical example of how a knowledge of the articulation-speech wave relations and the interrelations of the acoustic variables of specification contributes to the interpretation of spectrographic records.

A systematic study of the general principles of taking spectrographic measures in connected speech is at present being undertaken at our laboratory. The basic problem is how to divide the visible pattern in a succession of natural units. Are there any such units? The answer is positively yes. The sound spectrogram provides an excellent basis for the study of the durations of phonetic units. There are sharp breaks in the pattern associated with the major discontinuities in speech production, such as the onset of a fricative following a vowel or the step to or from complete closure in the vocal tract. Less unambiguous speech segment boundaries may be defined by variations in the formant pattern or by the appearance or disappearance of nasalization cues, i.e. the split versus single first formant.

The present specificational system[2] is based on a primary division of sound features in two basic categories; those determining *segment boundaries* which will be called *segment type features* and those specifying the contents of speech segments beyond the categorization inherent in the type features which we call *segment pattern features*. Segment type features thus concentrate on the discontinuous aspects of speech, in phonetic terminology referred to as manner of production. The segment pattern features, on the other hand, correspond more to the place of articulation.

[1] So called by W. Lawrence who has developed this method.

[2] A more detailed presentation of this tentative system is given in the Royal Institute of Technology, Speech Transmission Laboratory, *Quarterly Progress and Status Report* No. 2/1961, pp. 1–11. The system is related to that of Fant-Jakobson-Halle (15) but has a broader purpose.

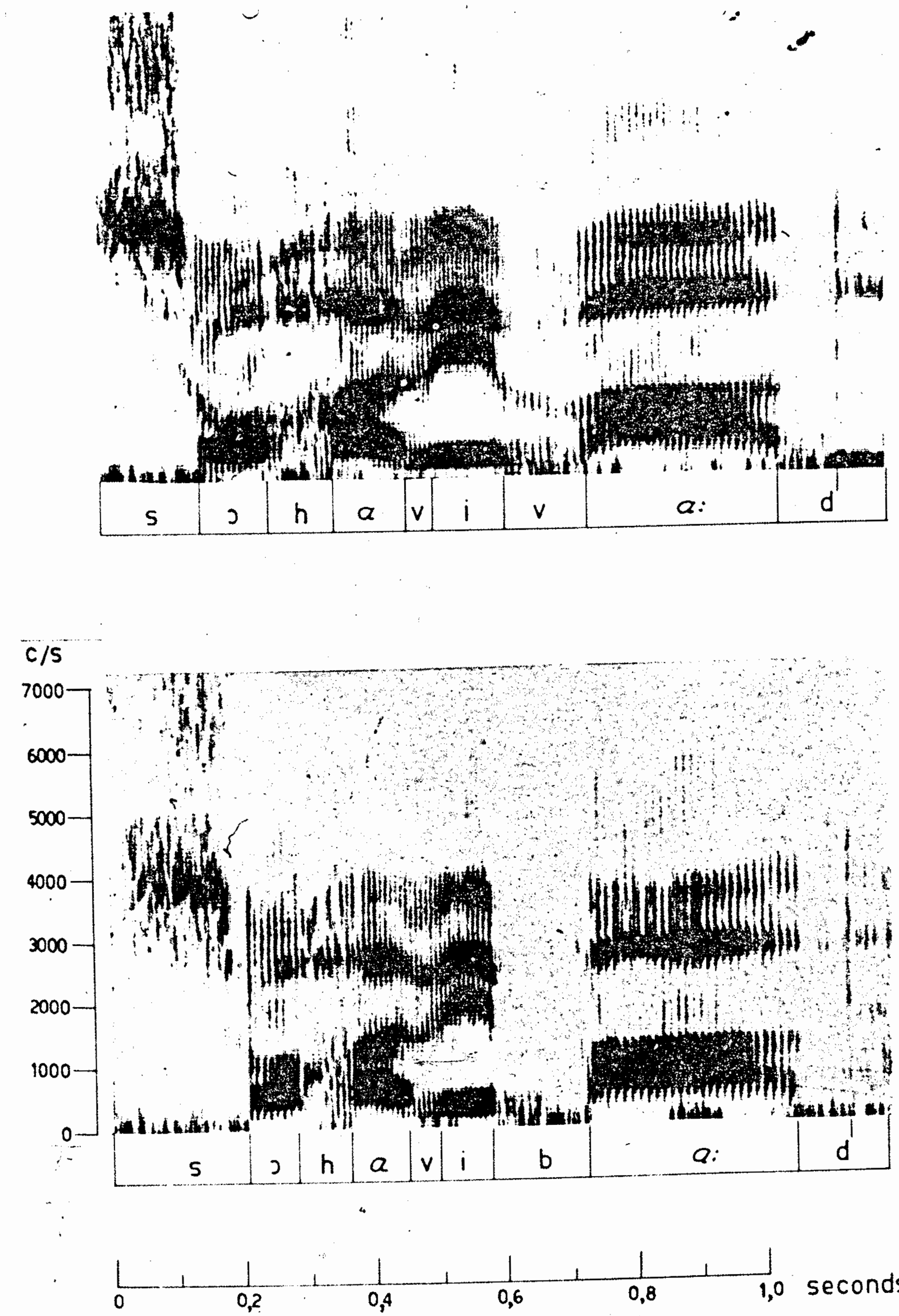

Fig. 9. Spectrograms illustrating the minimal distinction between [v] and [b]. The speaker is the Norwegian phonetician Ernst Selmer.

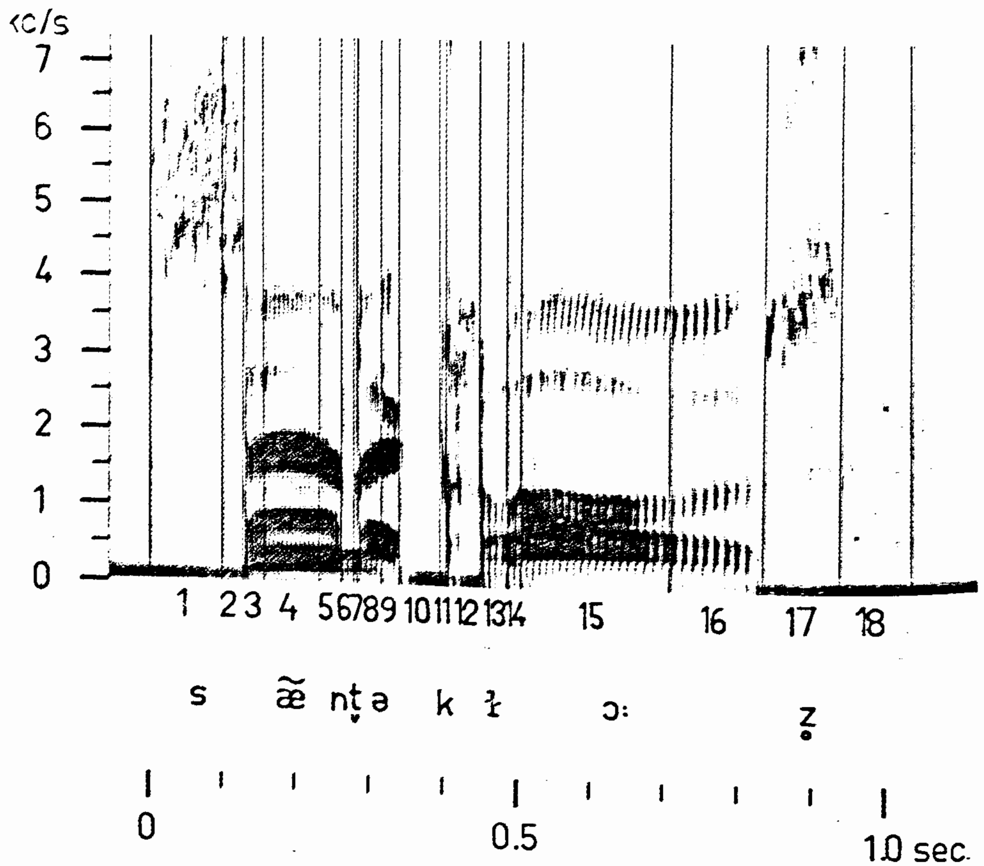

Fig. 10. Segmentation of a broad-band spectrographic record. Text: "Santa Claus."

SEGMENT TYPE FEATURES

Sources features	1	voice
	2	noise
	3	transient
Resonator features	4	occlusive
	5	fricative
	6	lateral
	7	nasal
	8	vowellike
	9	transitional[3]

The specification of any segment thus starts out with a number of binary decisions as to the presence or absence of each of the type features. The analysis then proceeds with the segment pattern features which are measured in terms of more or less continuous functions the particular choice of which depends upon the segment type classification. These data may finally be interpreted in terms of the most probable place of articulation. In automatic speech recognition the data from a number of successive segments should enable a phonemic or allophonic identification.

Most of the segment pattern features may be quantized in terms of an arbitrary number of standard articulatory positions or in terms of the acoustic parameter values allowing such identifications. A final identification of the phonemes of a piece of speech sometimes requires a few additional acoustic data such as intensity, duration, and voice fundamental frequency within segments supplementing the place of articulation data. In other words, not all of the segment pattern features are related to the place of articulation. The articulatory patterns may be described acoustically on the basis of two basic categories of measures, the *F-pattern* defined by the formant frequencies F_1 F_2 F_3 F_4 and *spectral energy distribution* generally defined by a spectrum envelope (10). In vowellike sounds the essential shape of the spectrum envelope is predictable from the F-pattern[4] and the F-pattern is then a sufficient basis for the specification.

In the case of fricatives and stops, on the other hand, both the F-pattern and the spectral energy distribution aid in signaling the specific place of articulation. The greater the energy content in the consonantal sound segment the greater perceptional importance its spectral energy distribution will have. A typical example is the difference between the relative intense burst of unvoiced stops compared with the relatively low intensity burst of voiced stops. Similarly, the nasal consonants differ more in terms of formant transitions as defined by the F-pattern, than by their inherent spectrum.

[3] An intermediate degree of feature 9 labeled glide has been adopted in some of the preliminary studies (Figure 11). Although phonemic motivations exist for this extra distinction it does not seem practical to maintain it in the acoustic analysis.

[4] Specific deviations from this rule define the lateral and nasal type features.

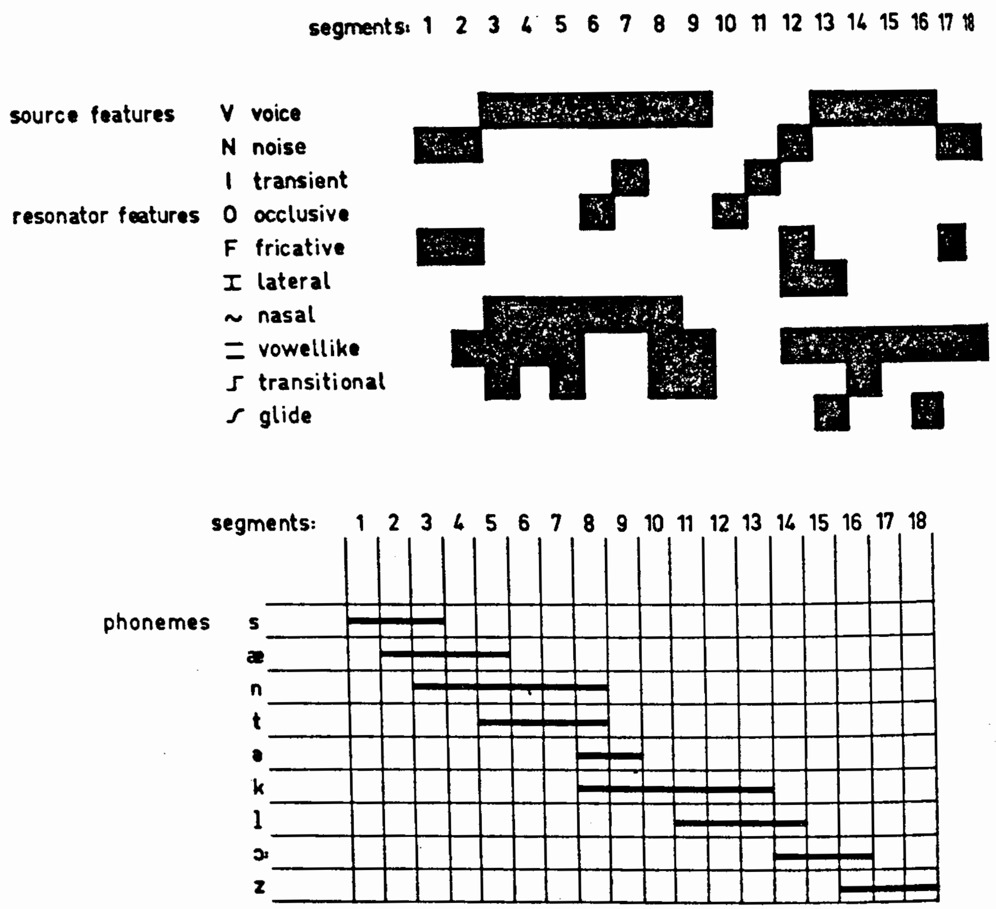

Fig. 11. Classification of the acoustic segments of the utterance Santa Claus, Fig. 10, in terms of segment type features. Estimated phoneme-segment dependency is indicated below.

The view of speech as composed of one sound segment for each phoneme is an oversimplification which is allowable on a phonemic level only. A purely phonetic segmentation according to the principles above of a specific sentence, see e.g. Figure 10, results in a larger number of sound segments than phonemes. Each sound segment generally carries cues for the identification of several successive phonemes and each phoneme is generally associated with several successive sound segments, see Figure 11. Well-known examples of these interrelations are the dual role of vocalic segments in signaling a vowel and an adjacent consonant and the dependency of a stop sound on the burst and a following vocalic element.

There is a wealth of information collected on the perceptional importance of the various cues in speech from the synthesis work at Haskins Laboratories (17, 18). Figure 12 exemplifies the nature of the systematic variations of transitional cues and spectral energy distribution cues.

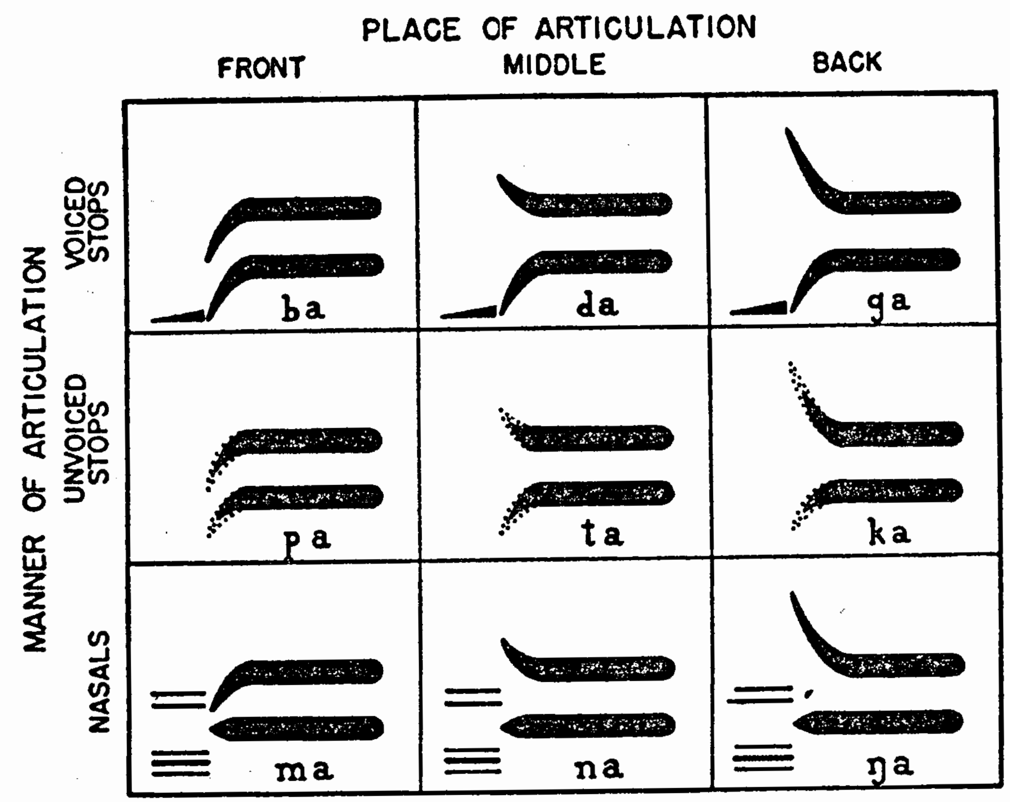

Fig. 12. Spectrographic patterns that illustrate the transition cues for the stop and nasal consonants in initial position with the vowel /a/. The dotted portions in the second row indicate the presence of noise (aspiration) in the place of harmonics. (After Alvin M. Liberman, 1957, *J. Acoust. Soc. Am.*, 29 (1957, pp. 117–123).

It would be interesting to follow up these classical experiments with synthesizers that retain the natural constraints of the human speaking mechanism. I am considering not only articulatory analogs such as DAVO (36) of M.I.T. but in particular synthesizers of the OVE II type (34, 35, 8, 4) which are coded in terms of acoustic parameters but retain the natural relations between formant frequencies and spectrum shapes, as discussed in connection with Figure 10. But now I am approaching the subject of the next paper for this congress.

Speech Transmission Laboratory
Royal Institute of Technology
Stockholm

REFERENCES

(1) Barczinski, L., Thienhaus, E., "Klangspektren und Lautstärke deutscher Sprachlaute", *Arch. Néerland. Phon. Exp.*, 11 (1935), pp. 47–58.

(2) Bickel, H. J., "Spectrum analysis with delay-line filters", *Ire Wescon Convention Record* (1959), Part 8, pp. 59–67.

(3) Delattre, P., Liberman, A. M., Cooper, F. S., "The physiological interpretation of sound spectrograms", *PMLA*, LXVI (1951), pp. 864–875.

(4) Fant, G., "Speech communication research", *IVA*, 24 (1953), pp. 331–337.

(5) Fant, G., "On the predictability of formant levels and spectrum envelopes from formant frequencies", *For Roman Jakobson* ('s-Gravenhage, 1956), pp. 109–120.

(6) Fant, G., "Modern instruments and methods for acoustic studies of speech", *Acta Polytechnica Scandinavica*, 246/1958, 84 pp., also publ. in *Proc. of VIII Internat. Congress of Linguists, Oslo 1958*, pp. 282–358 (1958).

(7) Fant, G., "Acoustic analysis and synthesis of speech with applications to Swedish", *Ericsson Technics*, 15, No. 1 (1959), pp. 1–106.

(8) Fant, G., "The acoustics of speech", *Proc. of the 3rd Internat. Congress on Acoustics, Stuttgart 1959*, ed. by L. Cremer, Vol. I (Amsterdam, 1961), pp. 188–201.

(9) Fant, G., "Descriptive analysis of the acoustic aspects of speech", paper given at the Wenner-Gren Foundation for Anthropological Research Symposium on Comparative Aspects of Human Communication, Burg Wartenstein, Austria 1960; to be publ. in *LOGOS*, the Bulletin of the National Hospital for Speech Research (1962).

(10) Fant, G., *Acoustic Theory of Speech Production* ('s-Gravenhage, 1960).

(11) Fischer-Jørgensen, E., "What can the new techniques of acoustic phonetics contribute to linguistics?", *Proc. of VIII Internat. Congress of Linguists, Oslo 1958*, pp. 433–478.

(12) Flanagan, J. L., "Difference limen for vowel formant frequency", *J. Acoust. Soc. Am.*, 27 (1955), pp. 613–617.

(13) Gill, J. S., "A versatile method for short-time spectrum analysis in 'real-time'", *Nature*, 189 (1961), pp. 117–119.

(14) Helmholtz, H., *On the Sensations of Tone* (New York, 1954) (Second English Edition).

(15) Jakobson, R., Fant, G., Halle, M., "Preliminaries to speech analysis", Massachusetts Institute of Technology, Acoustics Laboratory, *Techn. Report* No. 13 (1952), 3rd printing.

(16) Joos, M., "Acoustics phonetics", *Language*, 24 (1948), pp. 1–136.

(17) Liberman, A. M., "Some results of research on speech perceptions", *J. Acoust. Soc. Am.*, 29 (1957), pp. 117–123.

(18) Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P., Cooper, F. S., "Minimal rules for synthesizing speech", *J. Acoust. Soc. Am.*, 31 (1955), pp. 1490–1499.

(19) Lindblom, B., et al., "Evaluation of spectrographic data sampling techniques", Speech Transmission Laboratory, R.I.T., Stockholm, *QPSR*, 1/1960, pp. 11–13.

(20) Lindblom, B., "Formant frequency measurement", Speech Transmission Laboratory, R.I.T., Stockholm, *QPSR*, 2/1960, pp. 5–6.

(21) Lindblom, B., "Accuracy and limitations of Sona-Graph measurements", paper presented at the IV Internat. Congress of Phonetic Sciences, Helsinki 1961; see this volume. pp.

(22) Mártony, J., et al., "Zur Analyse und Synthese von Vokalen und Geräuschlauten", paper presented at the IV Internat. Congress of Phonetic Sciences, Helsinki 1961; see this volume, pp.

(23) Mathews, M. V., Miller, J. E., David, E. E., Jr., "Pitch synchronous analysis of voiced sounds", *J. Acoust. Soc. Am.*, 33 (1961), pp. 179–186.

(24) Menzerath, P., de Lacerda, A., *Koartikulation, Steuerung und Lautabgrenzung* (Berlin-Bonn, 1933).

(25) Miller, R. L., "Nature of the vocal cord wave", *J. Acoust. Soc. Am.*, 31 (1959), pp. 667–677.

(26) Paget, R., "The production of artificial vowel sounds", *Proc. Royal Soc.*, A 102 (1923), p. 75, and *Human Speech* (London, 1930).

(27) Peterson, G. E., "The information bearing elements of speech", *J. Acoust. Soc. Am.*, 24 (1952), pp. 629–637.

(28) Peterson, G. E., "Fundamental problems in speech analysis and synthesis", *Proc. of the VIII Internat. Congress of Linguists, Oslo 1958*, pp. 267–281.

(29) Peterson, G. E., "Parameters of vowel quality", *J. of Speech and Hearing Research*, 4 (1961), pp. 10–29.

(30) Peterson, G. E., Barney, H. L., "Control methods used in a study of the vowels", *J. Acoust. Soc. Am.*, 24 (1952), pp. 175–184.

(31) Pipping, H., *Om klangfärgen hos sjungna vokaler* (Helsinki, 1890).

(32) Potter, R. K., et al., "Technical aspects of visible speech", Bell Telephone System, *Monograph* B-1415, 1946; *J. Acoust. Soc. Am.*, 17 (1946), 89 pp.

(33) Potter, R. K., Kopp, A. G., Green, H. C., *Visible Speech* (New York, 1947).

(34) *Quarterly Progress and Status Report*, Speech Transmission Laboratory, R.I.T., Stockholm, STL–QPSR 1/1961 (January–March), April 15, 1961.

(35) *Quarterly Progress and Status Report*, Speech Transmission Laboratory, R.I.T., Stockholm, STL–QPSR 2/1961 (April–June), July 15, 1961.

(36) Rosen, G., "Dynamic analog speech synthesizer", *J. Acoust. Soc. Am.*, 30 (1958), pp. 201–219.

(37) Smith, S., "Analysis of vowel sounds by ear", *Arch. Néerland. Phon. Exp.*, XX (1947), pp. 78–96.

(38) Sovijärvi, A., "Die wechselnden und festen Formanten der Vokale erklärt durch Spektrogramme und Röntgengramme der finnischen Vokale", *Proc. III Internat. Phonet. Conf., Ghent 1938*, pp. 407–420.

(39) Sovijärvi, A., *Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache* (Helsinki, 1938).

(40) Stevens, K. N., "Toward a model for speech recognition", *J. Acoust. Soc. Am.*, 32 (1960), pp. 47–55.

(41) Stevens, K. N., House, A. S., "Studies of formant transitions using a vocal tract analog", *J. Acoust. Soc. Am.*, 28 (1956), pp. 578–585.

(42) Stumpf, C., *Die Sprachlaute* (Berlin, 1926).

(43) Truby, H. M., *Acoustico-cineradiographic analysis considerations with especial reference to certain consonantal complexes.* Thesis work, *Acta Radiologica*, Suppl. 182 (Stockholm, 1959).

(44) Öhman, S., "On the contribution of speech segments to the identification of Swedish consonant phonemes", Speech Transmission Laboratory, R.I.T., Stockholm, *QPSR*, 2/1961, pp. 12–15.

(45) Öhman, S., "Relative importance of sound segments for the identification of Swedish stops in VC and CV syllables", Speech Transmission Laboratory, R.I.T., Stockholm, *QPSR*, 3/1961, pp. 6–14.