AFCRL - 63 - 174

AD

AF EOAR GRANT 62-82
TR

Jan 28 1963

TECHNICAL REPORT

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. Fant

ROYAL INSTITUTE OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION LABORATORY

The research reported in this document has been sponsored in part by the Cambridge Laboratories, OAR through the European Office, Aerospace Research, United States Air Force

AD

AF EOAR GRANT 62-82
TR
Jan 28 1963

TECHNICAL REPORT

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. Fant

ROYAL INSTITUTE OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION LABORATORY

The research reported in this document has been sponsored in part by the Cambridge Laboratories, OAR, through the European Office, Aerospace Research, United States Air Force

Table of Contents

Preface

This report covers accomplishments during the calender year of 1962 at the Speech Transmission Laboratory of the Division of Telegraphy-Telephony at the Royal Institute of Technology in Stockholm, Sweden. Since the purpose of the US Air Force Grant 62-82, "Acoustic Specification of Speech", covers almost all phases of the speech research program at the laboratory it has been an established policy to include all accomplishments in the reports. It should then be recognized that a part of the support originates from US Army Contract DA-91-591-EUC-2626 and from Swedish government and private funds.

Since almost all tasks have been reviewed in our quarterly progress and status reports and in scientific papers we have found it unnecessary to repeat detailed presentations of status and results. The following report is thus limited to short summaries and reviews with reference to more detailed publications.

Abstract

The scientific accomplishments cover almost all activities of the speech research group. Detailed reports on each separate task have been given in the Speech Transmission Laboratory Quarterly Progress and Status Report for 1962 and in papers presented at the Stockholm Speech Communication Seminar and the IVth International Congress on Acoustics in Copenhagen. The results may be summarized as follows:

I. Speech Analysis

Theoretical and experimental studies on the measurement of formant amplitudes have been performed. Inverse filtering and photo-electric glottography as means of indirect studying the function of the vocal cords have been performed and a theoretical study has been devoted to the relative time-delays involved. An example of the application of inverse filtering to a phathological voice with typical multiple excitation is given. Earlier voice-voiceless

identification have been reviewed and supplemented. The status of the 51-channel spectrum analyzer is reported. Among design details of general interest is an analog to digital converter with sampling stage and logarithmic converter, the SADISTOLOG. A comparative study of Swedish, English, and Polish fricatives has been undertaken and a distinctive feature interpretation of the acoustic characteristics of Polish phonemes is made.

II. Speech Synthesis

The status of the OVE II formant coding speech synthesis is reviewed and a transistorized formant circuit based on current controlled coils is described. An experiment for evaluating the effects of formant frequency tracking errors typical of a former Melpar system was made.

III. Speech Perception

A study on the perception of Swedish fricatives has included spectrum matching, synthesis experiments, and time-selective cutting and commutation of pattern segments. The object has been to evaluate the relative importance of consonant spectrum, overall noise amplitude, and formant transitions as cues for the basic phoneme distinction.

A study of the acoustic correlates of the Swedish word accent has been undertaken. A thesis work on the relative importance of successive segments in VCV syllables as determined by time-selective cuts has been completed. The potentialities of vocoder modulated vibratory transducers for cutaneous communication have been investigated with special reference to deaf subjects.

IV. List of Publications

(1) Speech Transmission Laboratory, Quarterly Progress and Status Report, No. 1/1962 (January-March)

a. Liljencrants, J.: "51-channel analyzer for spectrum sampling (The SADISTOLOG)", pp. 3-5

b. Jassem, W.: "Pitch as a correlate of Swedish word accent" pp. 4-14

c. Liljencrants, J.: "Moment weighting techniques for segmentation", pp.15-21

(5) Pickett, J.M.: "Tactual vocoder as an aid for the deaf", Paper G6 given at the Speech Communication Seminar, Stockholm 1962

(6) Fant, G., Mártony, J., Rengman, U., and Risberg, A.: "OVE II synthesis strategy", Paper F5 given at the Speech Communication Seminar, Stockholm 1962

(7) Öhman, S.: "Perceptual segments and rate of change of spectrum in connected speech", Paper D5 given at the Speech Communication Seminar, Stockholm 1962

(8) Mártony, J.: "Some experiments on perceptual cues for Swedish fricatives", Paper D4 given at the Speech Communication Seminar, Stockholm 1962

(9) Liljencrants, J.: "A few experiments on voiced-voiceless identification and time segmentation of speech", Paper C8 given at the Speech Communication Seminar, Stockholm 1962

(10) Fant, G., Fintoft, K., Liljencrants, J., Lindblom, B., and Mártony, J.: "Formant amplitude measurements", Paper C2 given at the Speech Communication Seminar, Stockholm 1962

(11) Jassem, W.: "Noise spectra of Swedish, English, and Polish fricatives", Paper B8 given at the Speech Communication Seminar, Stockholm 1962

(12) Liljencrants, J., Rengman, U., and Garpendahl, G.: "51-channel analyzer for spectrum sampling", Paper C13 given at the Speech Communication Seminar, Stockholm 1962

(14) Öhman, S.: "On the perception of Swedish consonants in intervocalic position", Thesis work, University of Uppsala; also publ. as Report No. 25, March 1, 1962, Royal Institute of Technology, Speech Transmission Laboratory, Div. of Telegraphy-Telephony

(13) Fant, G.: "Descriptive analysis of the acoustic aspects of speech", LOGOS (U.S.A.), Vol. 5, No. 1, 1962, pp. 3-17

I. Speech Analysis

1. Formant amplitude measurements

Detailed results from this investigation have been presented in two articles, in STL, QPSR 2/1962 (2.b) and in a paper for the Stockholm Speech Communication Seminar (10).

The amplitude or intensity measure of a formant differs according to the particular definition. A common characteristic of all direct measures such as initial (maximal) amplitude, mean average, root mean square average is that they fluctuate with the speaker's pitch providing formant frequencies and vocal excitation are constant. This effect originates from the superposition of the formant oscillations evoked in one vocal period on those of succeeding periods. As pitch rises the amplitude measure oscillates more and more as a function of Fn/F_0, where Fn = formant frequency and F_0 = voice fundamental frequency. On the other hand, in case of a very low pitch frequency there is the well known ripple associated with a relative long duration of a voice cycle compared with the integration time of the measuring circuitry. A detailed mathematical analysis based on Laplace transforms shows that intensity measures employed in formant vocoder should not be of the direct category. The above mentioned dependency of voice pitch is avoided only if the amplitude measure refers to the vocal tract response of a single glottal pulse. Such measures can only be derived by spectrum matching techniques.

2. Inverse filtering and voice source studies

The inverse filtering instrumentation described in Speech Transmission Laboratory, Quarterly Progress and Status Report, No. 4/1961, pp. 1-6 is still in the constructional stage and a complete set up including 4 antiformant stages and correction for higher poles will be completed during 1963. Experiments in 1962 have been made with two standard antifilter units and a low-pass filter. On one occasion we processed speech illustrating a very harsh voice with double excitation, see Briess (1.b). A recent experiment carried out in cooperation with Dr. B. Sonesson (4.a) involved the synchronous recording of an inverse filter curve and a glottograph

curve according to the photo-electric method developed by Sonesson. The degree of similarity was better than expected which adds some faith to the inverse filtering method as a means of studying laryngeal functions.

3. Speech parameter tracking

Earlier investigations on detection of speech segment boundaries and the presence of voicing in speech have been supplemented and reported to the Stockholm Speech Communication Seminar (9). Recent experiments (4.c) on the parametric representation of spectral features have involved the use of moment weighting equipment for continuous measurements of center of gravity and spread of spectral energy. This technique, at present applied to fricatives, is probably useful in automatic detection schemes as a supplement to spectral specifications and as correlates of "compactness" and "acuteness".

4. The 51-channel spectrum analyzer

A progress report on the design and construction of the 51-channel spectrum analyzer was given at the Stockholm Speech Communication Seminar (12). This instrument is intended to become our main instrument for spectral analysis. Among future applications of the instrument is the compilation of statistics of speech spectra and the spectral processing of data to be transferred to a digital computer for evaluation of automatic analysis schemes.

The 51 filters may be varied to provide 11 various combinations of filter frequencies and bandwidths. Broad-band and narrow-band analysis as well as display on linear scales, logarithmic scales, and approximations to the mel scale are included. Filters are of third order in complexity but may alternatively be switched to function as single resonance circuits. In case of narrow-band analysis means are incorporated for stepwise shifts of the 51 filters. A digitalized read out system will provide a succession of decibel amplitude versus frequency sections synchronously sampled at intervals of $12.5 \cdot 2^n$ milliseconds. In each

channel a transistorized unit for sampling, logarithmic conversion, and analog to digital conversion in 128 steps of 0.5 dB the SADISTOLOG [1.a] is incorporated.

5. Distinctive features of Polish phonemes

An attempt has been made [1.c] to describe the system of the phonematically relevant acoustic features of Polish speech according to the general principles of the Jakobson-Fant-Halle theory. A few systematic departures from the original theory involving the adoption of new features and new formulations have been attempted as a means of gaining alternative solutions which might stimulate a further development of the distinctive feature approach.

6. Swedish, English, and Polish fricatives

A paper to the Stockholm Speech Communication Seminar[11] reports the consistency in repeated measurements of the formant frequencies of fricatives. Speaker characteristics are found both in particular F_2, F_3, and F_4 frequencies and in the spectral peaks above 5000 c/s.

A comparative study of typical allophones in the three languages [3.b] has emphasized the role of formant frequencies as supplementary means of differenting classes of fricatives; e.g., palatals and retroflex sounds generally possess smaller F_4-F_2^* than sounds of other place of articulation. Overall spectral features are described.

II. Speech Synthesis

1. Status of OVE II synthesis

A status report was included in STL, QPSR 2/1962 [2.c] and in a paper to the Stockholm Speech Communication Seminar [6].

The success of these coding principles has been demonstrated in speech material presented at the Stockholm Speech

Communication Seminar. Earlier experimental work on quantization (2.c) of the parameters have shown that very small degradation is introduced when the system is digitalized at a rate of 1200 bits/sec. Future work will include the development of modifications in synthesis strategy and corresponding automatic analysis needed in formant vocoders working after the OVE II principles. Our present cooperation with the Ericsson Telephone Company might lead to positive results in this respect. Contrary to opinions expressed by US experts there are good reasons to believe that formant vocoders working after such principles will proof to be feasible.

2. Transistorized formant circuit

A prototype of a transistorized formant generator circuit based on current controlled coils has been developed (3.d). The units have a fairly linear frequency control characteristic but may not satisfy stringent demands on control current residue in the output.

3. An experiment on simulation of formant tracking errors

During a visit of S.J. Campanella and D.C. Coulter from the Melpar, Inc. we made a test of the effects of typical errors in formant coding on the OVE II type of speech (3.e). Specifically it was desired to find the role of errors in F_1, F_2, F_3 in the perception of voiced stops. The conclusion was that to produce acceptable stop consonants based on voiced transitions requires a high accuracy in tracking of each of the first three formants.

III. Speech perception

1. Evaluation of cues for Swedish fricatives

In a paper to the Stockholm Speech Communication Seminar (8) the results have been summarized from investigations published in the STL, QPSR (1.d)(2.d).

Time selective cuts and commutations of fricative segments from natural speech and a special set of experiments with variations of synthetic OVE II speech have verified the importance of fricative

noise spectrum as the primary cue for fricative recognition. The weak intensity of labials is an important characteristic. The relative role of the boundary transitions in F_2 and F_3 and F_4 is not very great except in the distinction between retroflex versus palatal fricatives where the higher loci of the latter are supplementary cues to the higher center of gravity of the noise spectrum.

2. Time segmentation as a means of evaluating the relative importance of successive segments of speech

These studies have been reported in a thesis publication [14], in ref. [1.f], and in a paper to the Stockholm Speech Communication Seminar [7].

Experiments employing successive cuts in 20 msec intervals from the beginning and from the end of vowel-consonant-vowel nonsense words reveal that the rate of growth of identification of the consonant as a function of the time location of the cuts is at a maximum at cuts very close to the conventional sound segment boundaries as observed from spectrograms. Physical and perceptual boundaries in this respect thus coincide. The relative role of an arbitrary time segment is not independent of its surroundings. Several observations of the relative role of various segments to the identification of phonetic categories were made. Thus in voiceless intervocalic stops the initial vowel together with the occlusion interval appears to provide cues for the non-voicing and the burst and the final vowel contains place and manner cues.

3. An investigation of the Swedish word accent

The distinctive tonal features of the Swedish word accent have been studied by means of direct listening and tracings from narrow-band spectrograms and from oscillograms from an automatic pitch extractor [4.b]. The general agreement between subjective and objective pitch contours was better than previously reported. In connected speech the distinctions are maintained though with overlayed sentence intonation, and with respect to the tonal group preceeding the disyllabic word.

4. Tactual vocoder as an aid for the deaf

The following summary pertains to experiments reported in ref. (1.e) and a paper to the Stockholm Speech Communication Seminar (5).

Communication through the skin offers interesting possibilities as a supplementary channel of receiving information, e.g. for speech communication with deaf in which case it may supplement lip-reading. From this point of view an existing tactual vocoder (Lucia I) was submitted to phonetical evaluation. The system at the STL comprises a 10-channel vocoder with an intensity modulation of a 300 c/s carrier supplied to each of 10 hearing aid bone conduction receivers attached to the fingers. The center frequencies of the bandpass filters in the vocoder analyzer was 210, 400, 580, 830, and 1050 c/s for the left hand units and 1800, 2250, 3320, 5800, and 7700 c/s for the right hand units. Many important phonemic distinctions that were poorly discriminated by lip-reading were received considerably better with the tactual device.

ROYAL INSTITUTE
OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION
LABORATORY
Jan 28, 1962

AF EOAR 62-82
TR
ELECTRONICS

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. FANT

ABSTRACT: Results from studies of speech analysis, speech synthesis and speech perception related to the specification of acoustical determinants of voice quality, speech sounds and other phonetic categories and speech message units are summarized with reference to work and publications from the Speech Transmission Laboratory during 1962. Design features of instrumentation for speech analysis and synthesis are reviewed.

ROYAL INSTITUTE
OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION
LABORATORY
Jan 28, 1962

AF EOAR 62-82
TR
ELECTRONICS

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. FANT

ABSTRACT: Results from studies of speech analysis, speech synthesis and speech perception related to the specification of acoustical determinants of voice quality, speech sounds and other phonetic categories and speech message units are summarized with reference to work and publications from the Speech Transmission Laboratory during 1962. Design features of instrumentation for speech analysis and synthesis are reviewed.

ROYAL INSTITUTE
OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION
LABORATORY
Jan 28, 1962

AF EOAR 62-82
TR
ELECTRONICS

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. FANT

ABSTRACT: Results from studies of speech analysis, speech synthesis and speech perception related to the specification of acoustical determinants of voice quality, speech sounds and other phonetic categories and speech message units are summarized with reference to work and publications from the Speech Transmission Laboratory during 1962. Design features of instrumentation for speech analysis and synthesis are reviewed.

ROYAL INSTITUTE
OF TECHNOLOGY
STOCKHOLM 70, SWEDEN
SPEECH TRANSMISSION
LABORATORY
Jan 28, 1962

AF EOAR 62-82
TR
ELECTRONICS

ACOUSTIC SPECIFICATION OF SPEECH

C.G.M. FANT

ABSTRACT: Results from studies of speech analysis, speech synthesis and speech perception related to the specification of acoustical determinants of voice quality, speech sounds and other phonetic categories and speech message units are summarized with reference to work and publications from the Speech Transmission Laboratory during 1962. Design features of instrumentation for speech analysis and synthesis are reviewed.

UNCLASSIFIED