Technical Note 1969-55

A. V. Oppenheim

Computer Generated Speech Spectrograms

28 November 1969

Prepared under Electronic Systems Division Contract AF 19(628)-5167 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Lexington, Massachusetts

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

COMPUTER GENERATED SPEECH SPECTROGRAMS

A. V. OPPENHEIM

Group 62

TECHNICAL NOTE 1969-55

28 NOVEMBER 1969

LEXINGTON MASSACHUSETTS

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract AF 19(628)-5167.

ABSTRACT

One of the early applications of the fast Fourier transform was in spectral analysis of speech waveforms, and it was immediately evident that, at least in principle, speech spectrograms could be generated by computer using the FFT. In this note the procedure and results of such an experiment are presented.

Accepted for the Air Force
Franklin C. Hudson
Chief, Lincoln Laboratory Office

COMPUTER GENERATED SPEECH SPECTROGRAMS

One of the early applications of the fast Fourier transform (FFT) algorithm was in spectral analysis of speech waveforms for bandwidth compression, automatic speech and speaker recognition, speech displays, etc. It was immediately evident, and suggested both formally and informally,[1,2] that, at least in principle speech spectrograms could be generated on a computer using the FFT. The purpose of this note is to present the procedure and results for one such attempt.

The computational procedure consists of first determining the Fourier transform of the speech to be analyzed by weighting the speech with a Hanning window, then transforming using the FFT and computing the spectral magnitude. The number of points transformed, and the time increments between spectral cross-sections, depend in each case on the tradeoff desired between time and spectral resolution. An attempt was made to keep the analysis time approximately the same for wide and narrow band spectrograms. As a consequence, when the window duration, and consequently the size of the Fourier transform, is doubled, thereby increasing spectral resolution, the time increment between spectral cross-sections is doubled.

The spectrograms are displayed on a conventional Hewlett-Packard 1200 AR oscilloscope. The intensity to be displayed is used to modulate the duration of an unblanking pulse applied to the z-axis. The resulting hard copy is obtained from a time exposure, using Polaroid type 52 film. In the present system, the spectrograms are displayed on a grid with 256 points on the vertical (frequency) axis and 512 points on the horizontal (time) axis. Values on this grid are obtained from the computed values, which lie on a less dense grid, by linear interpolation.

The contrast and high frequency shaping desired in spectrogram displays is generally dependent on the specific application for which they are to be used. Consequently, a provision for varying both has been incorporated in the present system. Variable contrast is incorporated by applying a variable exponent to the spectral magnitude prior to D-A conversion. Frequency shaping is then applied by multiplying this modified spectral magnitude by a frequency dependent gain. The overall system is illustrated in Fig. 1. In Fig. 2 and Fig. 3 are displayed some typical spectrograms obtained in this way, together with the corresponding voiceprint spectrograms for reference. The sentence is "He took a walk every morning." as spoken by a male speaker. The input speech has been pre-emphasized at 6 db/octave starting at 350 Hz and sampled at a 10 kHz rate. Figure 2a shows a narrowband voiceprint spectrogram of the pre-emphasized speech and with high frequency shaping. Figures 2b, c, d correspond to a narrowband spectrogram, computed using a window length (and Fourier transform size) of 512 points. A new spectral cross-section is computed each 9.6 msec. In Fig. 2b, the parameter γ as defined in Fig. 1 is unity and no frequency shaping has been applied. Figure 2c corresponds to no frequency shaping but a γ of (2/3). Figure 2d represents a γ of unity and linear shaping starting at 1.25 kHz with a slope of 1.6 (1/kHz). Figure 3a shows a wideband voiceprint spectrogram of the pre-emphasized speech and with high frequency shaping. Figures 3b, c, d corresponds to a wideband spectrogram, computed using a Fourier transform size of 128 points. A new spectral cross-section is computed each 2.4 msec. In Figures 3b to 3d, the contrast and frequency shaping are the same as in Figs. 2b to 2d respectively. Figure 3e is identical to Fig. 3d except for an expanded time scale.

One of the potential advantages of computer generated spectrograms is the tremendous flexibility available. Bandwidths, for example, can be chosen to more nearly match the kind of speech being analyzed and in fact can be time variable, controlled for example by the spectral derivative or the fundamental frequency. Furthermore, the spectrogram display can be more precisely correlated with other speech displays such as the waveform and spectral cross-section. It seems reasonable to anticipate either special purpose digital hardware or computer attachments that can provide both hard copy and on-line spectrographic displays of speech, with a quality comparable to or perhaps better than other presently available spectrographic displays.

REFERENCES

1. Ernst Rothauser and Dietrich Maiwald, "Digitalized Sound Spectrograph Using FFT and Multi-print Techniques", Journal of the Acoustical Society of America (Abstract) Vol. 45, No. 1, p. 308, 1969.

2. G.D. Bergland, "A Guided Tour of the Fast Fourier Transform", IEEE Spectrum, July 1969, pp. 41-52.

ACKNOWLEDGMENT

I would like to express my appreciation to Mr. Don Johnson and Mrs. Ann Fried for the assistance that they provided.

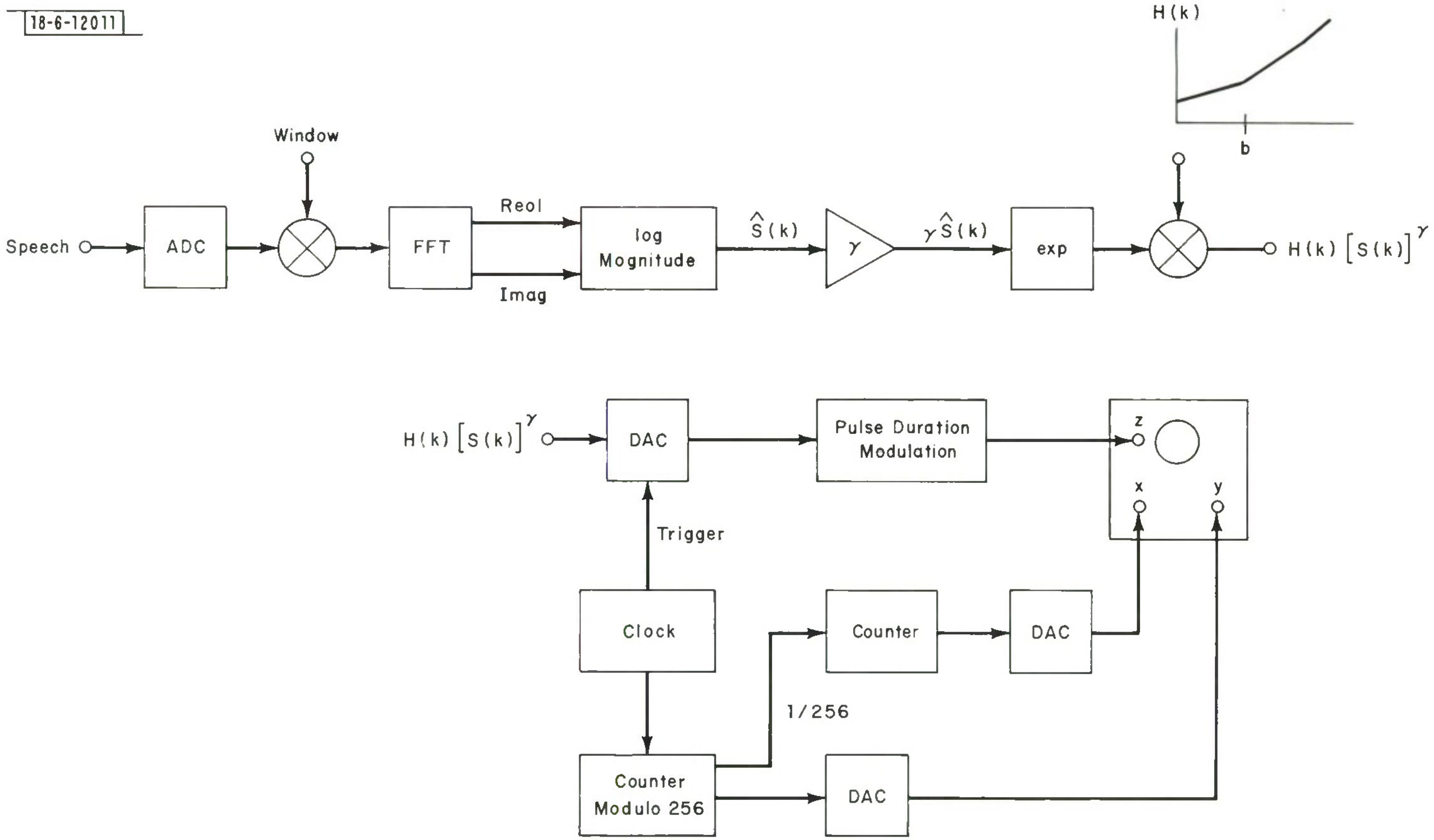


Fig. 1. Block diagram for obtaining and displaying spectrograms.

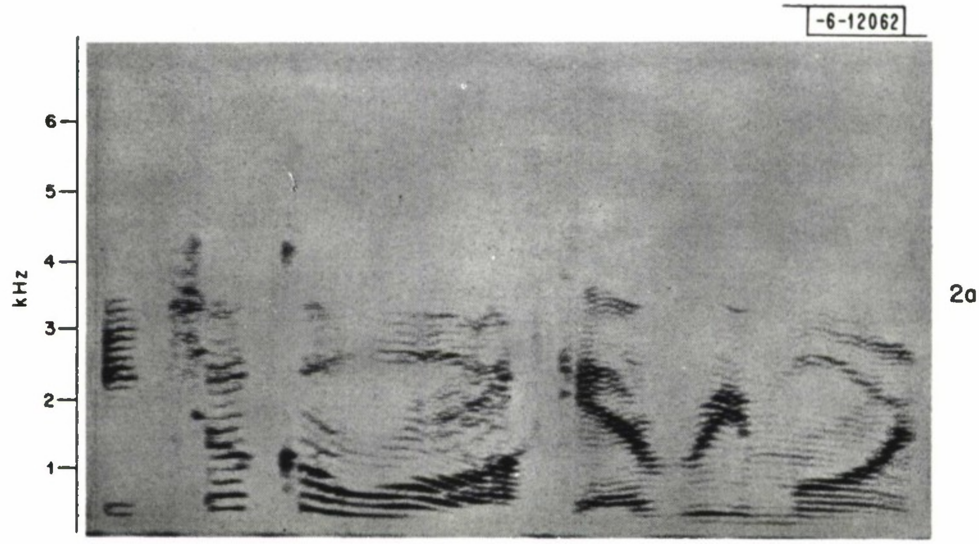


Fig. 2. Narrowband spectrograms.

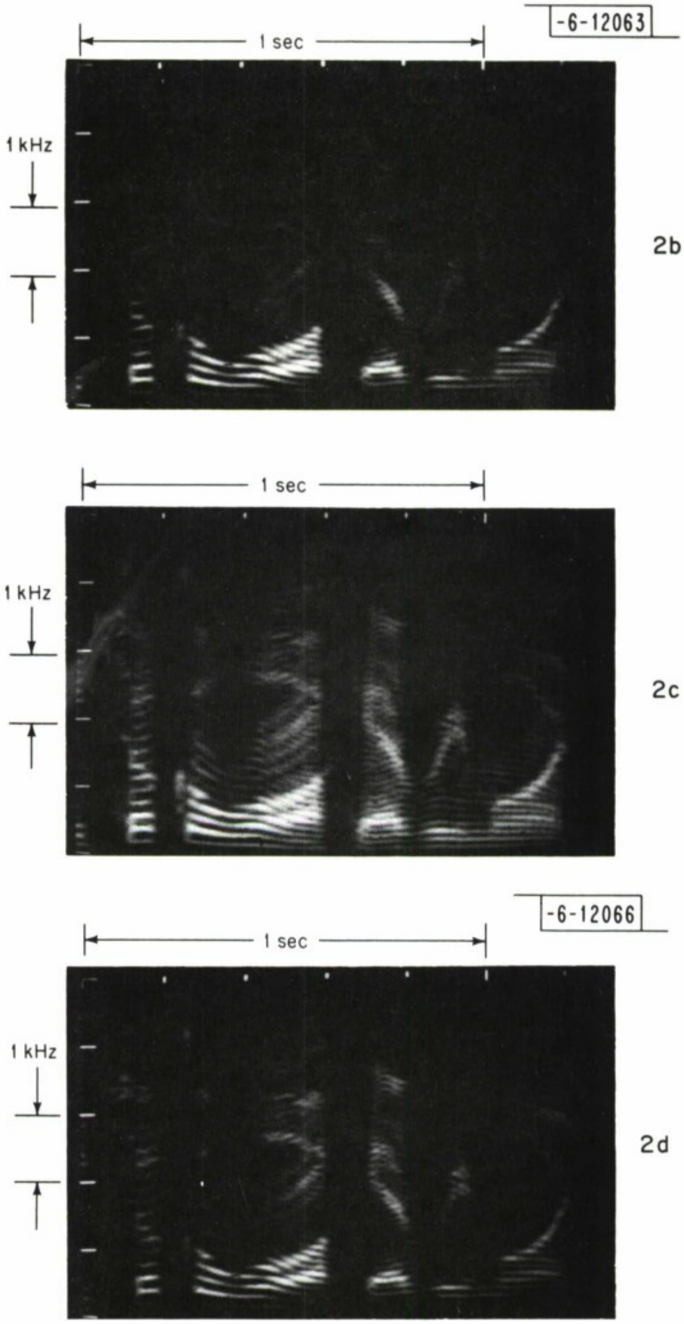
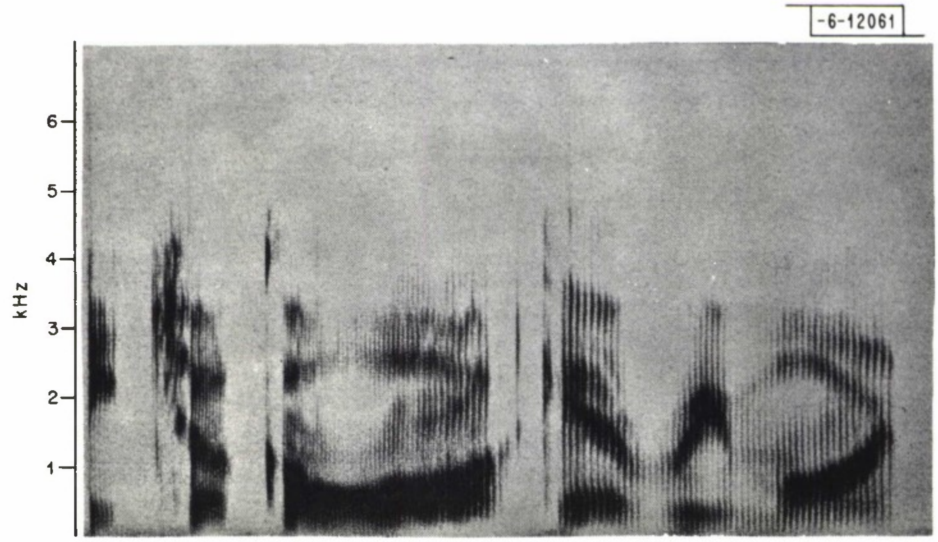


Fig. 2. Continued.



3a

Fig. 3. Wideband spectrograms.

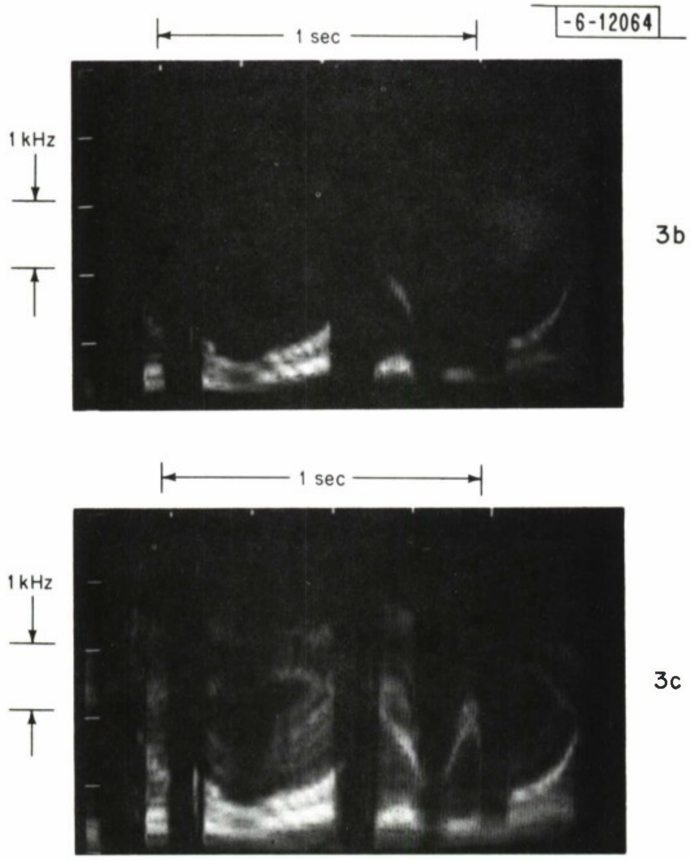


Fig. 3. Continued.

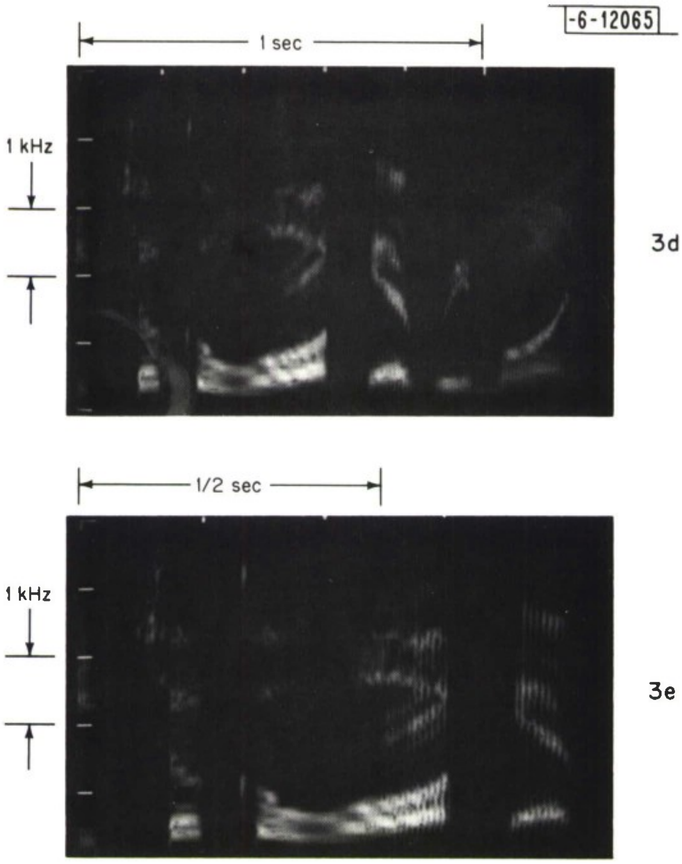


Fig. 3. Continued.

DOCUMENT CONTROL DATA – R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Lincoln Laboratory, M.I.T.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP None

3. REPORT TITLE

Computer Generated Speech Spectrograms

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Note

5. AUTHOR(S) *(Last name, first name, initial)*

Oppenheim, Alan V.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
28 November 1969	14	2

8a. CONTRACT OR GRANT NO. AF 19(628)-5167 b. PROJECT NO. 649L c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) TN 1969-55 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* ESD-TR-69-352

10. AVAILABILITY/LIMITATION NOTICES

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Air Force Systems Command, USAF

13. ABSTRACT

One of the early applications of the fast Fourier transform was in spectral analysis of speech waveforms, and it was immediately evident that, at least in principle, speech spectrograms could be generated by computer using the FFT. In this note the procedure and results of such an experiment are presented.

14. KEY WORDS

speech	spectral analysis
speech recognition	spectrograms
speech compression	