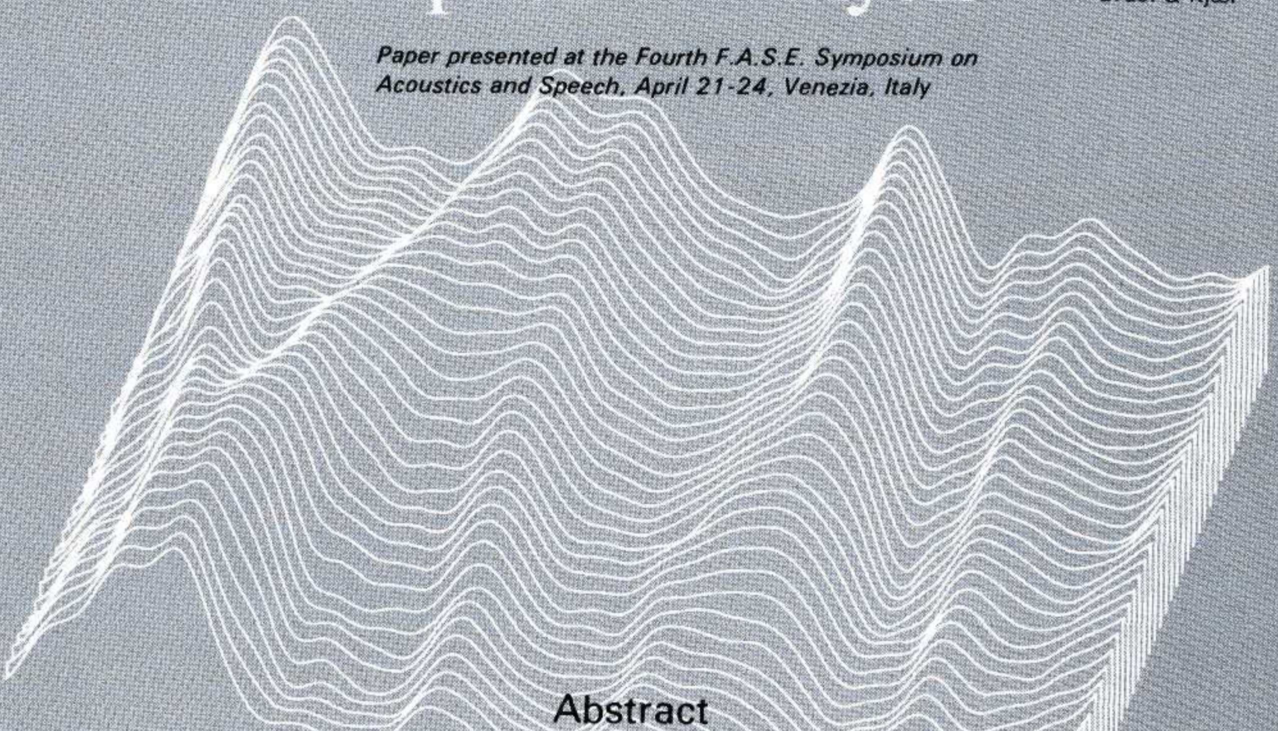
Application of a long Memory FFT Analyzer in Speech Analysis

by *N. Thrane*, Brüel & Kjær

Paper presented at the Fourth F.A.S.E. Symposium on Acoustics and Speech, April 21-24, Venezia, Italy



Abstract

The introduction of dedicated FFT analyzers having a long input memory — i.e. 10K samples — has opened up new possibilities for simpler forms of speech analysis. Scanning a 1K sample long time window through the 10K sample input memory permits seeing a running spectrum and studying its time variations. This can be repeated in its entirety or in part at will and at different speeds. Addition of a small desktop calculator allows shorter time windows to be used. Also a "scan" cepstrum can be performed for pitch extraction and formant analysis.

Introduction

The analysis of speech as a nonstationary signal is normally implemented on mini-computers coupled to array processors. In contrast, this paper describes the use of a dedicated, self-contained instrument for the simpler forms of analysis. The instrument contains an ordinary FFT processor which operates on a 1K sample time signal, thereby producing a 400 line spectrum ranging from DC up to a chosen maximum frequency. However, the input memory of the instrument is ten times longer than the length of the analysis window, i.e. 10K samples. This allows the analysis window to be "scanned" through the total input memory and hence displaying the spectrum as it changes as a function of time. A more detailed discussion of the instrument can be found in the reference.

066—81

Scan Analysis

When the instrument is used in the 0 — 5 kHz range the time length of the input memory is 800 ms, which is sufficient to contain one or more syllables of speech. In this case the analysis window is 80 ms long, but since a Hanning weighting function is normally used the effective length of the analysis window is only 30 ms. A scan of this analysis window through the complete input memory can be performed either manually or automatically. In the automatic scan the step size between the individual analyses can be varied from 80 ms (i.e. 10 spectra per scan) down to 0,7 ms (1153 spectra per scan). Each analysis takes approx. 100 ms so the duration of a complete scan can be varied from 1s up to 2 minutes. In this way a "slow motion" presentation of the spectral changes can be observed directly on the instrument screen. Since the scan analysis does not change the original content of the input memory, the scan can be repeated at will, analysing any part of the signal at any convenient speed.

Fig.1 shows the result of such a scan analysis. The word "Montreal" was recorded in the 800 ms input memory: however, only the part containing the "ea" sound was analysed. The step size used was 3,1 ms. The three-dimensional plot used here to present the results does not reflect the visual impression of the live "slow motion" analysis, but only the complexity of the results.

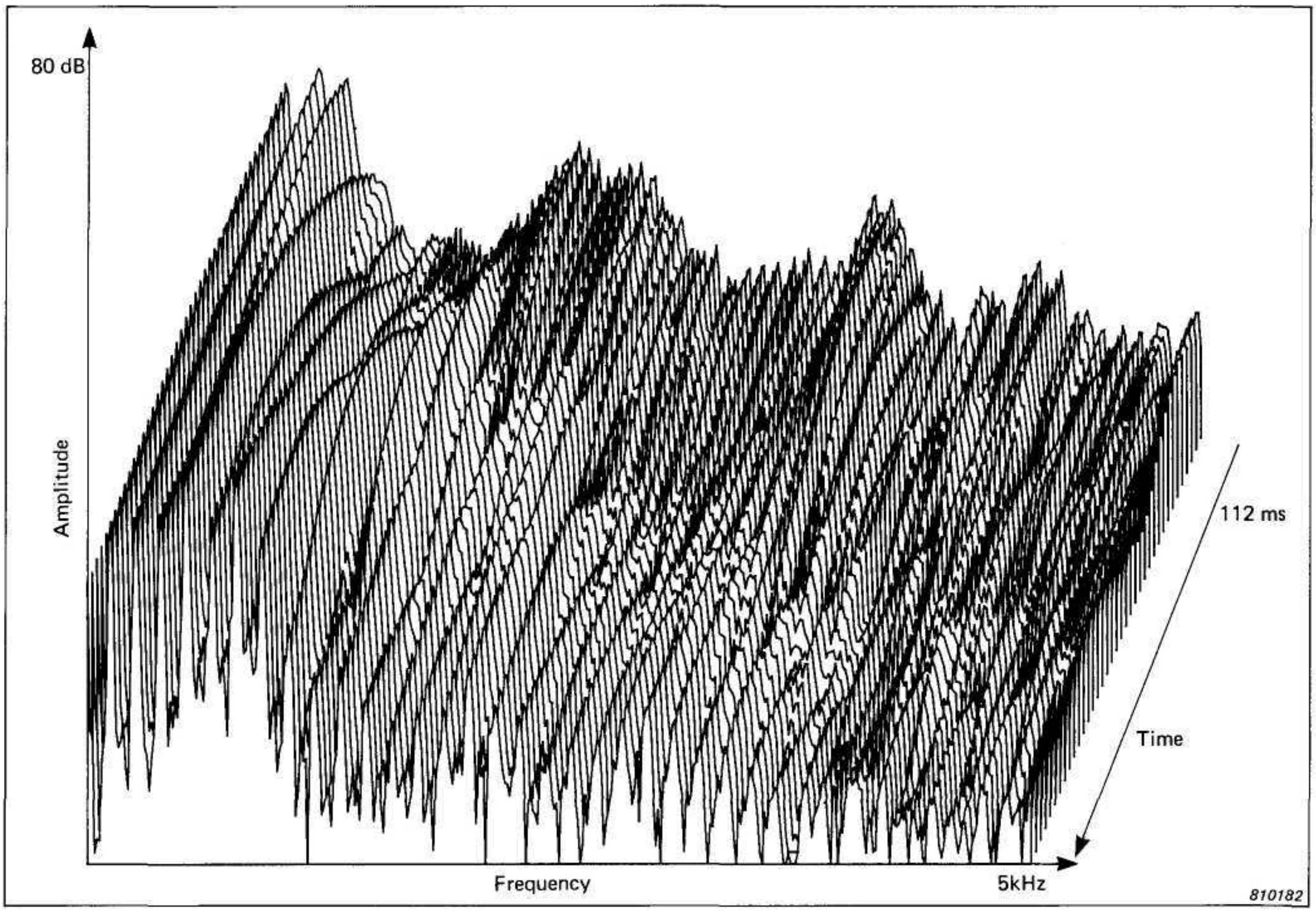


Fig.1. Scan spectrum of "ea" in "Montreal"

Cepstrum Liftering

In order to simplify the information contained in the 3-D landscape of Fig.1 a "cepstrum liftering" was performed of each spectrum. An example is shown in Fig.2. The log power spectrum of a vowel sound is seen in Fig.2.a and its cepstrum in Fig.2.b. (The cepstrum is defined as the inverse Fourier transform of the log power spectrum).

The rapid fluctuations in the spectrum caused by the voice frequency (150 Hz) and its harmonics are reflected in the cepstrum as a sharp peak at a "quefrency" of 6,6 ms plus the second and third "rahmonics". Conversely, the slowly varying parts of the spectrum, i.e. the formants, are found at low quenfrencies. Since the two effects are additive and separated in quefrency, application of the "short pass lifter" characteristics shown in Fig.2.c will remove the influence of the harmonics of the voice from the cepstrum. A smooth lifter characteristic, being half of a Hanning function of length corresponding to the voice quefrency, was used to avoid introduction of discontinuities in the liftered cepstrum. Finally, the short pass liftered cepstrum was Fourier transformed back to the frequency domain, producing the log power spectrum shown in Fig.2.d, where the formant characteristics are clearly seen.

The cepstrum liftering procedure was performed using a small desktop calculator in conjunction with the FFT analyser, both communicating over their common IEC data bus.

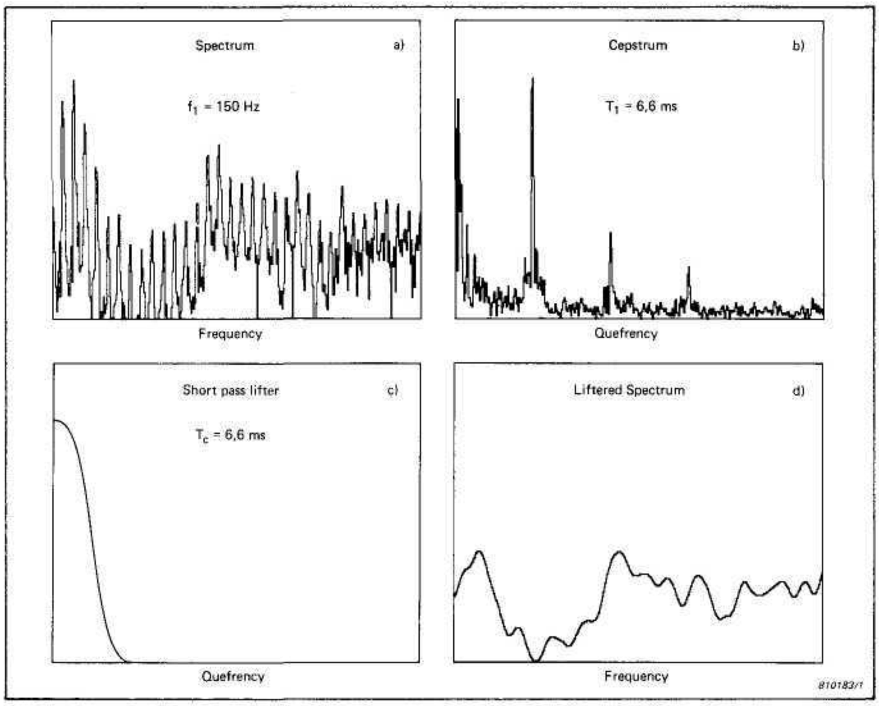


Fig.2. Cepstrum liftering. a) log power spectrum of vowel. b) magnitude of cepstrum. c) short pass lifter characteristic. d) short pass liftered log power spectrum

The final result is shown in Fig.3 as a 3-dimensional plot. In contrast to the analysis shown in Fig.1, the cepstrum liftering has emphasised the formant structures so that the development with time can now easily be followed and studied in detail.

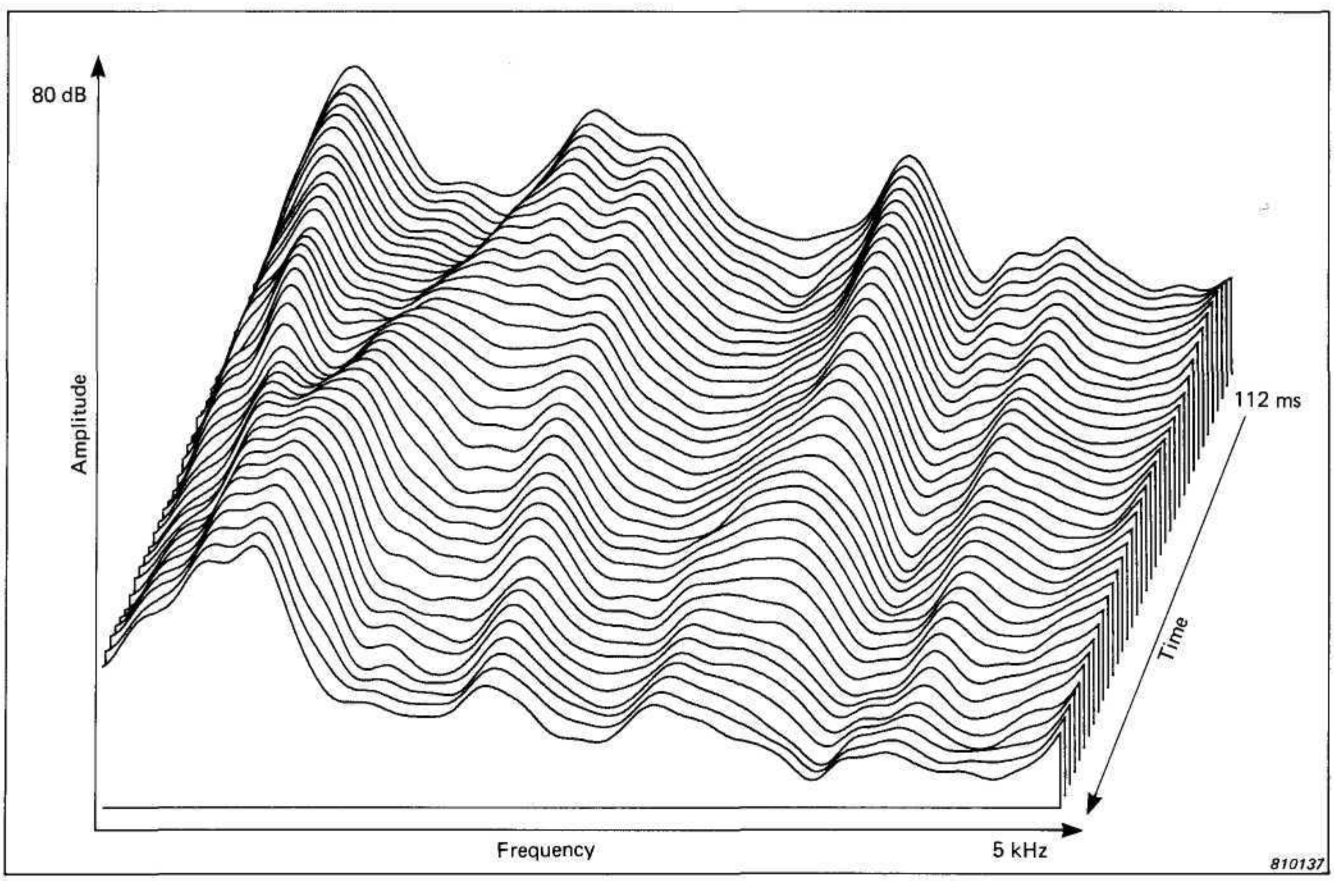


Fig.3. Short pass liftered scan spectrum of "ea" in "Montreal"

Conclusion

An FFT analyser having a long input memory can be used for a live "slow motion" scan analysis of speech. In conjunction with a small desktop calculator it provides a number of more refined analysis methods: scan analysis of short pass liftered spectra, scan of voice pitch either in the frequency or in the quenfrency domain, and scan analysis of signals changing very rapidly by use of time windows shorter than those normally available.

Reference

N. Thrane: Zoom, FFT, Technical Review No. 2, 1980, Brüel & Kjær.

The instrument described in this paper is the Brüel & Kjær Type 2033, High Resolution Signal Analyzer.

Further information on the cepstrum technique and its applications can be found in the following references:

A.V. Oppenheimer & R.W. Schafer, Digital Signal Processing, Prentice Hall, N.J., 1975.

D.G. Childers, D.P. Skinner, and R.C. Kemerait, "The Cepstrum: A guide to processing", Proc. IEEE, vol. 65, pp. 1428—1443.

R.B. Randall, and J. Hee, "Cepstrum Analysis", Technical Review, Brüel & Kjær, to be published.