A NEW HIGH RESOLUTION TIME-BARK ANALYSIS METHOD FOR SPEECH

Unto K. Laine

Helsinki University of Technology, Acoustics Lab.,
Otakaari 5 A, 02150 Espoo, Finland

ABSTRACT

A new complex auditory filter bank based on a new class of orthogonal functions called FAM functions [2] is developed. The filter bank is used to produce time-Bark spectrograms where both the magnitude and the phase variations of the speech sample in the channels are seen. The temporal resoluion is of high quality: even phenomena inside one pitch period can be monitored. An analyzer is implemented with the TMS320C30 digital signal processor, which is programmed in a Macintosh IIci / CLOS Lisp environment [1].

1. INTRODUCTION

Auditory based representation of speech is proven to be an effective way to compress, code, enhance, analyse, describe, and visualize speech signals. Auditory analysis and compression seems to have many promising applications in digital audio, communication technology, and in audiology. The methodology provides also new tools for basic research in the fields of speech acoustics and production as well as speech perception and automatic speech recognition.

The auditory representation of audio signals is typically achieved by a filter bank with each filter having a bandwidth equal to one Bark (one critical band) which forms the limit for the auditory spectral resolution around the frequency in question [5].

Typically, one of the following methods is used to construct an auditory filter bank: a standard linear filter design method is used, modifying the short time Fourier transform (STFT), or applying quadrature filter (QF) techniques [3]. In the case of the modified STFT a proper frequency dependent window function is included in the Fourier transform in order to achieve the nonuniform frequency resolution in the uniform frequency (Hz) scale. Note that this is the same case as if we had a uniform resolution in a nonuniform frequency scale (in Barks). In the QF-approach a uniform resolution filter bank is first designed according to the highest frequency resolution, and the narrow channels are then combined at the higher frequencies to reduce the resolution according to the Bark scale.

This paper describes a new type of auditory filter bank, which is designed by applying the FAM-method [2]. The method is based on orthogonal functions of a FAM class, which leads to an auditory bank with **complex** and **orthogonal** one Bark bandpass channels. Each channel consists of two filters which form a Hilbert pair and give a complex signal as the channel output. This allows to define the channel energy at every sample as the magnitude of the complex output and to formulate a phase measure within each critical band (auditory phase modelling). The orthogonality between the channels means that the impulse responses of the channels do not correlate. In fact, such a bank performs an Orthogonal Auditory Transform (OAT) from the time domain to the Bark domain.

The magnitude calculation improves the time resolution of the bank since magnitude can be calculated immediately at every sample instead of rectification and low-pass filtering the output of one (real) channel. The latter introduce a delay and some unprecision and uncertainty along the time scale.

2. FAM-METHOD

The new class of orthogonal FAM functions which are formed from the circular functions by Frequency and Amplitude Modulation has been earlier published by Laine and Altosaar [2]. In FAM functions the frequency and amplitude modulations of sinusoids are combined in a way so that a set of orthogonal functions are produced. A generative function, g(ω), which defines the frequency modulation, defines the properties of the FAM function set.

Fig. 1 describes briefly the FAM-method used to form a complex, orthogonal auditory one Bark bank. The method and its use for computation of auditory spectrograms proceeds through the following steps:

1° The desired frequency resolution, or the new warped frequency scale, is fixed by the choice of the generative function g(ω). In the case of the auditory bank the choice equals to a Hz-to-Bark conversion [4,5].

2° The orthogonal sets of FAMsin(·) and FAMcos(·) functions e.g. FAMexp(j n g(ω)) are generated in the frequency domain in order to describe the complex spectra of the signal with the new resolution in the linear ω-scale, or with a uniform resolution in the new warped g(ω)-scale.

3° The FAM functions are inverse Fourier transformed into the time domain. The inverse Fourier transform retains the orthogonality, thus we get an orthogonal set of real time functions. Channel responses of this bank resemble phase distorted impulses. So, we could call the set as the **Bark pulse bank** (BPB) to emphasise their correspondence with the frequency domain (Bark-warped) FAM functions.

4° The speech signal can be convolved with the BPB to get the coefficients for the corresponding FAM-functions for spectrum composition.

5° The spectral picture in the Bark-domain is produced by Fourier transforming the output samples from the BPB. Note that this part of the spectrum computation can be avoided by forming the final one Bark filter bank as linear combinations of the FIRs in the BPB. Instead of the Fourier transform we can produce the final one Bark FIR filter bank by forming a cos(nx) weighted sum of the BPB FIRs to produce the real part of channel n, and sin(nx) weighted to produce the imaginary part of channel n.

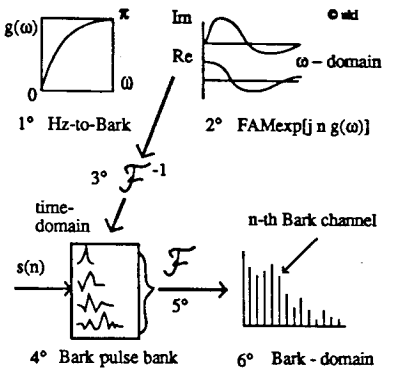

Fig. 1. FAM - method for nonuniform resolution spectral analysis.

6° The Fourier transform of the outputs of the BPB gives the desired complex spectra in the Bark-domain. The same information is available from the one Bark bank made in 5°.

3. IMPLEMENTATION

The one Bark auditory FIR filter bank designed by the FAM-method is implemented by using the TMS320C30 digital signal processor in a Macintosh IIci/ CLOS Lisp environment [1].

The C30 processor takes the speech samples in and performs the numerically intensive filtering, magnitude and phase computations, and finally, sends the data to the MacII for displaying the spectrogram and plotting with a laser printer.

The filter bank covers the frequency band from about 5 Hz to 11 kHz having one DC-type (real only) channel and 21 complex one Bark channels.

The magnitude and phase are processed for every sample in every channel. The prototype realization is relatively slow. However, we have estimated, that a real time implementation of the one Bark bank is possible by using all-pass structures for the filters.

The phase information from the one Bark channels can be processed in different

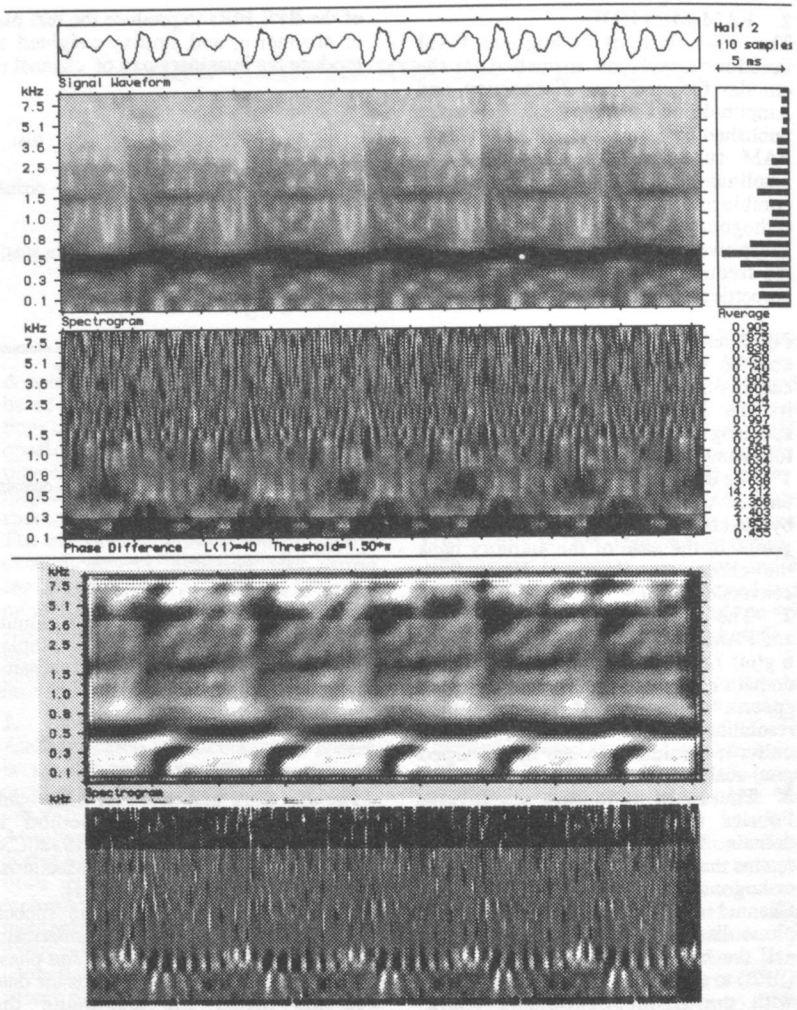

Fig. 2. Vowel /œ/ analysed by the auditory bank (see text).

ways in this auditory spectrograph: in each channel the phase information over time is first dewarped in order to cancel the 2π steps. Then the more or less linear ramp (corresponding to the time flow) is differentiated to remove the linear part and to emphasise the variations in phase. In another method a trivial prediction is made to estimate the coming phase sample from the previous samples. Then the difference of the estimated value and the true value is formed and plotted.

3. SPEECH ANALYSIS WITH THE AUDITORY BANK

Fig. 2 depicts results of the analysis made for the vowel /œ/. The waveform is shown in the topmost panel. Below it can be seen the time-Bark energy distribution and on the right hand side the average Bark-spectra. The auditory phase processed by the trivial predictor is shown in the next panel. In the following frame a graphically processed version of the previous spectrogram is

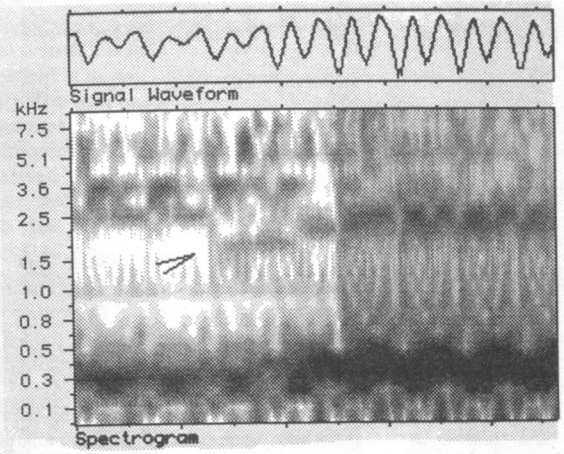

Fig. 3. Spectrogram of syllable /li/.

shown. The differentiated auditory phase is displayed in the lowest frame.

Some pitch-synchronous effects can clearly be seen: the higher formants will be primarily excitated at the closure of the glottis. In some cases a clear secondary excitation appears at the glottal opening. Note also the frequency modulation of the first formant.

Fig. 3 demonstrates a transient analysis of the syllable /li/. Typically a clear transient effect can be perceived in this syllable eventhough the steady state spectra of /l/ and /i/ do differ only very little. The arrow in the figure points to the place where the second formant suddenly appears at the moment the tongue opens the mouth cavity. The formant moves up very fast, about 2 Barks in 12 ms! Note that the contrast of the left hand side is improved by graphical postprocessing.

4. DISCUSSION

Preliminary results from the first implementation of the FAM-based auditory filter bank was presented. Many details can and must be improved. The way the phase information is processed is not yet robust enough. After a better processing it may reveal much more detailed and relevant information of the complicated phenomena called speech acoustics.

5. ACKNOWLEDGEMENTS

The author is grateful to his colleagues at the HUT Acoustics Lab. for the help and assistance during this study. Toomas Altosaar has made a great deal of programming work with Lisp in the QuickSig/Symbolics environment, work which made this study possible. Vesa Välimäki has implemented the algorithms in the C30/MacII CLOS Lisp environment creating a nice, user friendly package. The Lisp software for the environments has been developed by Professor Matti Karjalainen.

6. REFERENCES

[1] KARJALAINEN, M., "A Lisp-based high-level programming environment for the TMS320C30", *Proc. of IEEE ICASSP 89, Glasgow, Scotland, 1150-1153.*

[2] LAINE, U., and ALTOSAAR, T. (1990), "An orthogonal set of frequency and amplitude modulated (FAM) functions for variable resolution signal analysis", *Proc. of IEEE ICASSP 90*, Albuquerque, New Mexico, 1615-1618.

[3] SMITH, M., and BARNWELL, T. (1987), "A new filter bank theory for time-frequency representation", *IEEE Tr. ASSP*, **35**, 3, 314-327.

[4] TRAUNMÜLLER, H. (1981), "Perceptual dimension of openess in vowels", J. Acoust. Soc. Am., **69**, 1465-1475.

[5] ZWICKER, E., FELDKELLER, R. (1967), *"Das Ohr als Nachrichtenempfänger"*, Stuttgart: Hirzel Verlag.