# Time/Frequency Techniques for Signal Feature Detection

Adele B. Doser
Electrical Engineering Dept.
The University of Texas at Dallas
Richardson, Texas

Gerald D.T. Schuller
Bell Laboratories
Lucent Technologies
Murray Hill, New Jersey

## Abstract

*A method for detecting linear chirps in audio signals is presented. It uses a complex uniform spaced filter bank as a basis for the detection. The filter bank has an efficient implementation, good frequency localization, and delivers magnitude and phase of the subband signals. The complex filter bank uses two parallel real valued filter banks, each with critical sampling and perfect reconstruction capability. Detection is accomplished via a template based thresholding scheme. Various templates are used to account for differing amounts of linear chirp in the data sets. Experimental results using actual audio data are presented, which demonstrate the method is a useful tool for the detection of linear chirps in audio signals.*

## 1. Introduction

Linear chirp signals (signals whose frequency changes linearly over time) are found in many areas of nature and across many disciplines. One field where the detection of linear chirps is of major importance is audio coding. In this work, we address the detection of linear chirps encountered in audio signals, such as singing or speaking voices. In these circumstances, it is important to distinguish between narrow band noise and chirps, in order to control the right amount of quantization noise that can be introduced in subbands. In many cases, the assumption of linear chirps is a simplification. However, the supposition is justifiable here, since the changes in frequency of music signals, such as singing voices, come close to or have the characteristics of linear chirps.

Present high quality audio coders examine neighboring subbands of the audio signal to distinguish between tonal (sinusoidal) or noise-like parts. Unfortunately, this approach cannot distinguish between narrowband noise and a chirp that covers several subbands over a certain time period.

In very low bit rate audio coding, the frequency of sinusoids (and frequency changes of chirps) is estimated using an FFT and phase regression methods [1]. The FFT has the advantage that it delivers magnitude and phase of a subband signal. But since an FFT does not have good frequency localization properties, this method performs well only on signals with fairly sparse sinusoids and low noise.

In our approach, a filter bank which delivers magnitude and phase is used, as in the FFT approach, but with better frequency localization properties. Since we have linear chirps, we utilize a uniform filter bank, so that chirps become diagonal lines in the time/frequency plane (if the signals of interest were nonlinear chirps, an octave spaced wavelet filter bank would be used instead to create linearization). The filter bank has non-symmetric impulse responses, with a faster rise and a slower decay. In our goal of using the same filter bank for quantizing and encoding the subband signals (previously, a separate filter bank was used for encoding), the particular rise/decay feature is important, and it also better captures the relevant features of the signal. In addition, the property is desirable in audio coding applications, since the temporal masking properties of the inner ear have the same characteristic [2].

The output of the filter bank is a time/frequency representation of the audio signal. When the magnitude of the result is displayed, the locations of linear chirp signals can be clearly observed and distinguished from narrowband noise. Thus, an event classifier can be developed to detect the occurrence of the chirp events. The classifier employed consisted of a template matching scheme followed by a thresholding operation. Magnitude is a reliable detector for a sinusoidal or chirp signal in a subband, independent of its phase position, making the results independent of phase shifts and more trustworthy. In addition, the improved frequency localization in the utilized filter bank more effectively separated sinusoids or chirps closely spaced in frequency, making a wrong classification less likely.

Section 2 describes the complex filter bank used in the application. In section 3, details of the detection scheme are presented. The next segment discusses results of the combination filter bank/detection method. Finally, a summary is given.

## 2 The Filter Bank

The filter bank used for the chirp detection is a *complex filter bank*, as described in [3, 4]. Since we want use very non-symmetric filters, we use the design method of [5, 6] as a basis to construct a complex filter bank. The complex filter bank used is a uniform modulated filter bank, which is implemented as two real-valued filter banks, one for the real part and one for the imaginary part. Modulated filter banks obtain their filters by modulation, or multiplying a baseband filter with sinusoidal functions, shifting the baseband filter in frequency. In our application, the real part of the filter bank is modulated with a cosine function, and the imaginary part is obtained with a sine function. The phase shift in the modulating functions leads to a 90 degree phase shift between the frequency responses of the two parts, which is then reflected in a corresponding phase shift between the real and the imaginary parts of the subband signals. Therefore, these pairs of subband signals can be interpreted as complex exponentials, and hence it is straightforward to obtain phase and magnitude information. As an illustration, Fig. 1 shows the impulse responses of the cosine and sine modulated filter in the third band of the 1024 band filter bank used in our experiments. It can be seen that the filters have a length of 4096 taps, which is used to improve the frequency localization.

When the impulse responses of the filter bank are interpreted as a complex function, its magnitude displays that of the window function which was used to obtain them. This magnitude can be seen in Fig. 2, which also shows the non-symmetric character of the window function. The tail to the right is difficult to observe by inspection, but is primary to improve the frequency response, which can be seen in Fig. 3, with an enlargement for the passband. These figures demonstrate the good frequency localization (the narrow passband and the high stopband attenuation) of the used filter bank.

As a comparison, Fig. 4 shows the magnitude of the frequency response of a length 1024 DFT with no overlap (a requirement for critical sampling with a DFT) and rectangular window. It can be seen that the stopband attenuation is much lower, which is the reason for the low frequency localization properties of the DFT.

The number of bands (1024) was chosen because it is the most commonly used in audio coding (e.g., in MPEG and PAC coders [7]), since it fits reasonably well with the time/frequency characteristics of most audio signals. The high number of bands also shows that an efficient implementation is important for practical applications, because a direct implementation would require on the order of $N \times N$ multiplications per frequency frame. Fortunately, modulated filter banks can be implemented very efficiently using fast Discrete Cosine Transforms or Discrete Sine Transforms, for cosine and sine modulation respectively, which only require on the order of $N \times \log N$ multiplications per frequency frame. Further contributing to the efficient implementation is the fact that a so called *critical sampling* was used in our application. This means that for an $N$ band filter bank, the subbands are downsampled by a factor of $N$, so that the total number of samples in the subbands equal the number of samples in the signal. That means the $N \times \log N$ multiplications are only needed for every $N$ input samples. This representation still contains all information about the signal, if a corresponding synthesis filter bank can perfectly reconstruct the signal. This perfect reconstruction property is fulfilled not only for the real part in our application, but interestingly also for the imaginary part, although it uses a different modulation function.

The design method and the structure to obtain the window functions for perfect reconstruction and an efficient implementation for the real part, that is, the cosine modulated filter bank, is described in [5, 6]. The imaginary part is obtained by replacing the Discrete Cosine Transform in [5, 6] by a Discrete Sine Transform. The coefficients of the filter bank structure stay essentially the same, except for some suitable sign changes. The window function for perfect reconstruction was then obtained by numerical optimization of the coefficients of the filter bank structure.

Compared to a spectrogram using a STFT, the complex filter bank results in longer window functions with multiple overlap, without having to increase the number of samples in the subbands.

## 3 Detection Scheme

There were two basic signal types present in the audio files examined, linear chirps and narrowband sinusoidal components. The filter bank representation did an effective job of separating the two signal components, but the detection task was not trivial. Unlike sinusoids, which appear as horizontal signatures in the time/frequency plane, linear chirps display a degree of diagonality. In the same way that two lines can have different slopes, two chirps may contain diverse amounts of chirp yet still be linear. Thus, a linear chirp may look almost like a sinusoid in time/frequency, be nearly vertical like an impulse, or anything in between. In addition, we are not limited to one chirp per data file. An audio file may contain two chirps, a dozen, or none at all. Furthermore, the chirps will likely be located at widely varying portions of the time/frequency plane. A supplementary complication is that the chirps may be increasing in frequency, but they may also be decreasing.

If the magnitude of the filter bank output is displayed, the linear chirps present in the data become observable. The event classifier employed in this work consisted of a template matching scheme followed by a threshold detector. The classification method proceeds as follows. A template was constructed for the filter bank output in the pre-processing phase of the detection algorithm, which was based on a linear model. Different slopes were used to account for varying degrees of chirp in the signals. However, since we were using a discrete coordinate system, the templates appeared as jagged diagonal lines rather than strictly linear. We also had decreasing as well as increasing slopes, which effectively doubled the number of possible templates. The templates were binary, meaning they masked out all coefficients outside the desired region, and allowed values inside the zone to pass through unchanged, in much the same way as a stencil.

In the on-line processing, the magnitude of the filter bank output was first normalized. Due to the size of the output files (1024 frequency bins and scores of time bins), the data was examined in sections. In each section, several different lags were employed to each template, to allow the chirp to be located at various points within the time/frequency window of interest. For each lag/slope pair, the length of the prospective chirp was then adjusted. If the tails were too small relative to the maximum value that was passed through the template, the length was shortened. If the prospective chirp became too short, it was assumed not to be a chirp. If the potential chirp passed the length test, its coefficients were added together and compared against a threshold, selected to achieve a reasonable trade off between detection and false alarm. The locations of any detected chirps were then mapped and stored. Finally, the window in the time/frequency plane was moved, and the process repeated for a new section of data, until all sections had been examined. The product of the algorithm was a map in the time/frequency plane containing locations of any detected linear chirp signals.

## 4 Detection Results

Figure 5 displays the first 100 frequency bins of a typical filter bank output set. The frame has been normalized. In this particular example, the data coming into the filter bank consisted of approximately one second of an audio file with a female voice singing a cappella. Note the appearance of several chirps in the figure, some possessing the contours of an upside down U. Various strong sinusoidal components can be seen as well.

Figure 6 demonstrates the results of the detection scheme. Since the algorithm looks for chirps with several different lag/slope pairs, the templates do, necessarily, overlap. Therefore, it is possible that more than one lag/slope template may detect the same chirp signal. If this is the case, we can say with even stronger certainty that a chirp does exist in the region. The magnitude values in the chirp location map are explained the following way. If the magnitude of the location map is one, that means that exactly one lag/slope template found a chirp in the vicinity, a magnitude of two means that two lag/slope templates found the chirp, and so on. As can be seen in the figure, the magnitude values are as high as seven.

The detection algorithm was tested on seven different audio files, five contained a singing voice, two contained a speaking voice. Performance for the singing files exceeded that for the spoken files. This attribute was the result of the linear chirp edges not being as well defined in the expansions of the spoken data files. In addition, the labeling of chirps is somewhat subjective (do we call a slowly varying sinusoid a chirp, for instance?). If we call a *major chirp* a linear chirp which spans at least 5 frequency bins and has a coefficient at least a third the size of the largest coefficient in the filter bank data for that file, the scheme correctly detected 93% of the major chirps present in the singing data. One false alarm was observed.

## 5 Summary

In this work, we have employed a combination of a complex uniform filter bank with a template matching detection scheme in order to detect linear chirps present in audio signals. The filter bank utilized an efficient implementation, with two parallel real valued filter banks, each with critical sampling and perfect reconstruction capability. To account for various amounts of chirp that may be present in actual audio signals, a variety of templates were used in the detection method to accommodate different slopes as well as locations in the time/frequency window. Experi-

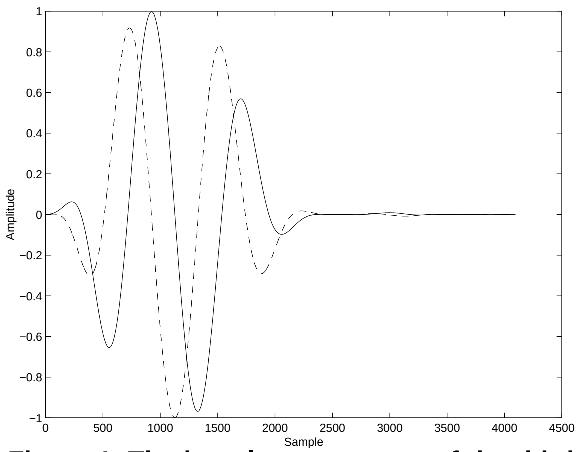


**Figure 1. The impulse responses of the third band of the sine and cosine modulated filter bank.**

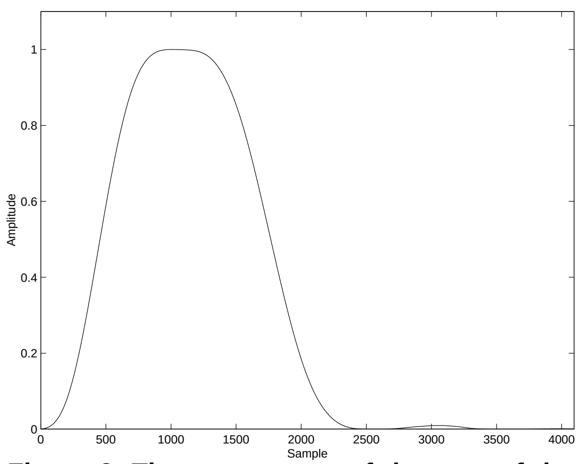


**Figure 2. The square root of the sum of the squares of the two impulse responses shown in Fig. 1.**

mental results demonstrated that complex filter banks, when used in conjunction with an appropriate detection scheme, can be useful tools in the detection of linear chirps in audio signals.

## References

[1] Bernd Edler, Heiko Purnhagen, Charalampos Ferekidis, "ASAC - Analysis/Synthesis Audio Codec for Very Low Bit Rates Preprint 4179," *Proceedings: 100th Audio Engineering Society (AES) Convention,* Copenhagen, May 1996.

[2] G. Schuller: "Time-Varying Filter Banks with Low Delay for Audio Coding", *Proceedings: 105th Audio Engineering Society (AES) Convention,* San Francisco, CA, Sep., 1998.

[3] H. S. Malvar, "A Modulated Complex Lapped Transform and its Applications to Audio Processing", *Proceedings: International Conference on Acoustics, Speech and Signal Processing (ICASSP99)*, Phoenix, AZ, March 1999.

[4] Nick G Kingsbury, "Shift Invariant Properties of the Dual-Tree Complex Wavelet Transform", *Proceedings: International Conference on Acoustics, Speech and Signal Processing (ICASSP99)*, Phoenix, AZ, March 1999.

[5] G. Schuller and M. J. T. Smith, "New Framework for Modulated Perfect Reconstruction Filter Banks", *IEEE Transactions on Signal Processing,* Vol.44, NO.8, August 1996.

[6] G. Schuller: "A New Factorization and Structure for Cosine Modulated Filter Banks with Variable System Delay", *Proceedings: Asilomar Conference on Signals, Systems, and Computers,* Pacific Grove, California, Nov. 3-6, 1996

[7] V.K. Madisetti, D.B. Williams, Editors, *The Digital Signal Processing Handbook*, Section IX, IEEE Press, 1998.

[8] A. Doser, "A New Time/Frequency Technique for Detecting Chirped Signals," *Proceedings: Thirty-Second Asilomar Conference on Signals, Systems and Computers,* Pacific Grove, CA, Nov. 1-4, 1998.

[9] Papandreou, A.; Boudreaux-Bartels, G.F.; Kay, S.M.,"Detection and estimation of generalized chirps using time-frequency representations" *Proceedings: Twenty-Eighth Asilomar Conference on Signals, Systems and Computers,* Pacific Grove, CA, Volume: 1 , 1994 , Page(s): 50 -54 vol.1

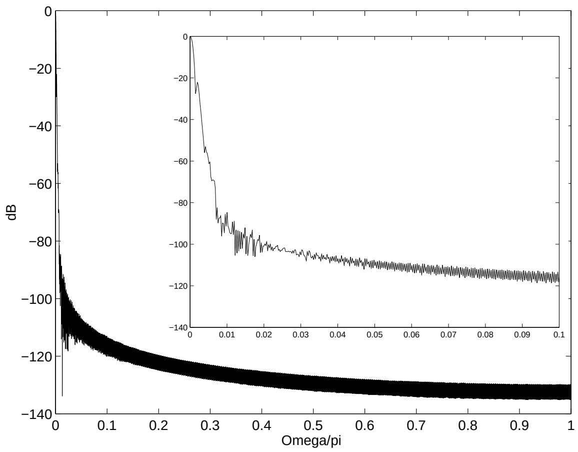


**Figure 3. The magnitude of the frequency response of the window function or baseband prototype. (The interior figure is an enlargement of the first tenth of the x axis.)**

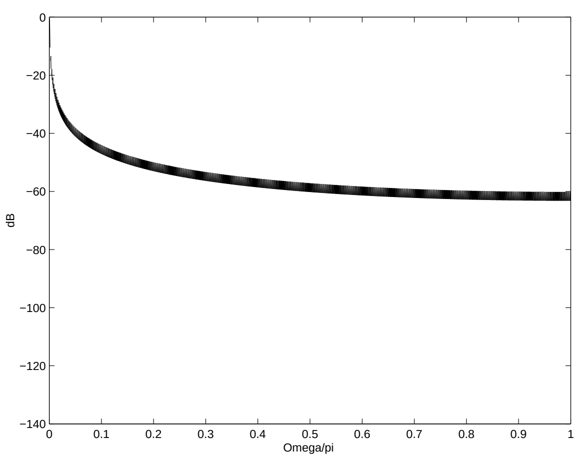


**Figure 4. The magnitude of the frequency response of the rectangular window function corresponding to a size 1024 DFT.**

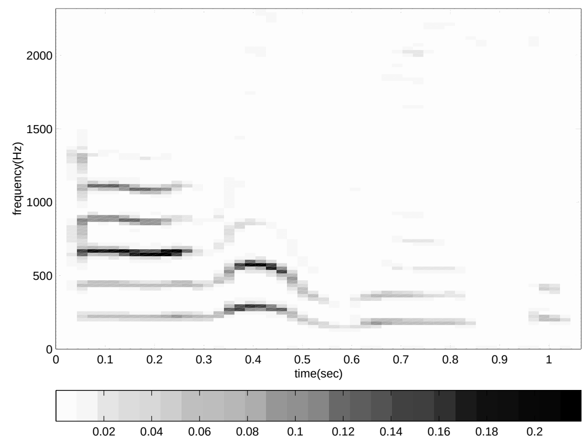


**Figure 5. Normalized filter bank output, female singing voice, a cappella (first 100 frequency bins).**

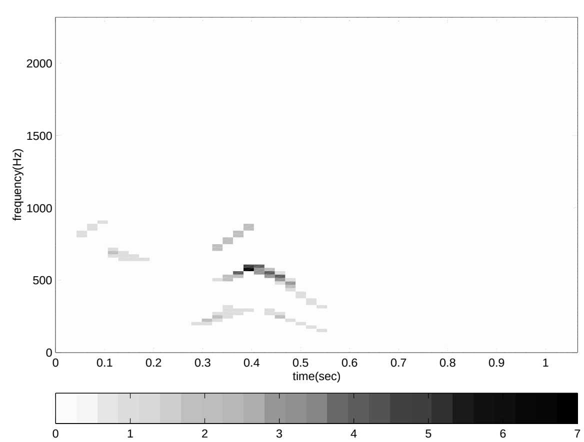


**Figure 6. Map containing locations of detected chirps for the data file of figure 5.**