Preprocessing for the Automated Transcription of Polyphonic Music: Linking Wavelet Theory and Auditory Filtering

Rolf Wöhrmann* and Ludger Solbach†
AB Vertete Systeme, TU Hamburg-Harburg
21071 Hamburg, Germany

Abstract. In this paper a fast method for the calculation of a linear time-frequency distribution based on the gammatone filter auditory model is introduced as a preprocessing step for the automated transcription of music and auditory source separation. Examples show that this method has a promising potential for the analysis of music pieces with limited spectral overlap of the different signal components.

1 Wavelet Analysis Using Auditory Filters

In the past few years wavelet transforms have become an important tool for signal processing. See [Rioul and Vetterli, 1991] for an overview. An important property of both the continuous wavelet transform (CWT) and the short-time Fourier transform (STFT) is their linearity, which makes them more suitable for the analysis of multicomponent signals than quadratic time-frequency distributions (TFD) suffering from cross-term artefacts. The CWT is given by

\[ W_s(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(r-a) \cdot \psi^* \left( \frac{r-b}{a} \right) \, dr, \quad a > 0. \]  

(1)

For admissibility of a time function \( g(t) \) as a mother wavelet, it must satisfy

\[ \int_{-\infty}^{\infty} |g(t)|^2 \, dt < \infty, \quad \text{and} \quad G(0) = 0, \]  

(2)

where \( G(f) \) is the Fourier transform of \( g(t) \). Functions satisfying these conditions look like short waves, which has been the reason for naming them wavelets.

The main difference between the STFT and the CWT lies in the fact, that for the STFT the analysis window remains unaltered, whereas the CWT changes its scale due to the scaling factor \( a \). The quasi-logarithmic organization of musical scales and of the frequency resolution in the human cochlea makes the CWT a more appropriate TFD of acoustic signals than the STFT. Since eq.1 cannot be evaluated everywhere, it has to be modified by picking certain fixed values for \( a \) and \( b \) yielding a discrete approximation of the CWT. Furthermore, the wavelets used throughout our work are close to being analytic (progressive), that is they satisfy \( \int_{-\infty}^{\infty} G(f) \, df = 0 \). Thus, given the complex-valued filter outputs the instantaneous frequencies can be estimated from the phases and the signal envelopes can be estimated from the modulus, if the signal components have a negligible overlap.

Due to the high computational burden of the quasi-continuous CWT, infinite impulse responses (IIR) realizations like our gammatone approach are desirable. As stated by Patterson (1992) the gammatone filter can be a good approximation of the filtering in the human cochlea if the parameters are properly adjusted. See fig.1 for an example of a gammatone impulse response. A quasi-analytic version of the gammatone filter is of the form

\[ g_{n,\lambda}(t) = k \cdot e(t) \cdot e^{-\lambda t^2 / 2} \cdot e^{i \lambda t}, \]  

(3)

with a Laplace transform of

\[ G_{n,\lambda}(s) = \frac{k \cdot (n-1)!}{(s + (\lambda - j2\pi f_0))^n}. \]  

(4)

*Email: rolf@rolfhz.hanse.de
†Email: solbach@tu-harburg.de

ICMC PROCEEDINGS 1995
where \( e(t) \) is the unit step function, \( n \) the filter order, \( \lambda > 0 \) the damping factor, \( C \) some normalization constant and \( f_c \) the center frequency of the filter. Closed expressions for the calculation of the coefficients of an IIR discrete-time analog gammatone filter approximation have been given in (Solbach et al., 1995). For energy normalization we set
\[
\lambda = (2\pi)^{2n-1} \Gamma(2n - 1) \Gamma^{-1}(n),
\]
where \( \Gamma(x) \) denotes the Gamma function. Fig. 2 visualizes an example of a bank of normalized gammatone filters.

A function cannot be arbitrarily concentrated in both time and frequency. The lower limit of the time-frequency window size is given by Heisenberg's uncertainty principle. The minimum RMS window area of \( 1/(4\pi) \) holds for the Gaussian function and its shifted variants in the time-frequency plane (Papoulis, 1987). In (Solbach et al., 1995) it has been shown that for the gammatone filter we have
\[
\frac{\Delta f}{\Delta t}(n) = \frac{1}{2\pi} \sqrt{\frac{n \cdot (2n - 1)}{4n - 6}}
\]
for the window area in the time-frequency plane, leading to the conclusion that \( n = 3 \) is the appropriate choice for minimum window size. Compared to the Gaussian window the resolution of the gammatone wavelet is considerably worse. In terms of wavelet theory, however, both the Gaussian and the gammatone wavelet are not admissible, because eq.2 is not satisfied. Practically, this can be neglected if the steepness of the filters is sufficiently high.

For further processing of the CWT output, a method for the detection of sound onsets has been introduced in (Solbach et al., 1995). It is based on a matched filtering procedure, making use of the fact that a characteristic phase pattern in the time-scale plane can be observed at the time of a sound onset.

2 Examples

The first example in fig.3 shows an analysis of the first four measures of Johann Sebastian Bach's Goldberg Variations played by Glenn Gould. The upper display was created by plotting bars proportional to the modulus of each band, centering around the instantaneous frequencies estimated from the phase. The lower display is a graph of the above mentioned sound onset detection algorithm. We expect to solve the problem of interference, strongly evident in the range between 12s and 17s, by removing identified steady components prior to the application of the algorithm. The analysis was performed over 3 octaves descending from 967.71 Hz using 15-equidistant filters per octave with \( \Delta f/f_c = 0.01 \). In order to enhance the graphical display, only filter output values with modulus above a certain threshold were plotted. In fig.4, harmonics of the base frequencies were suppressed to a great extent by a simple phase locking detection algorithm.

The second example shown in fig.5 is an excerpt of a Mbira improvisation played by Dumisani Mhazo from South-Africa. It shows the complex polytonic and polyrhythmic structure of African
Figure 3: Analytic gammatone filtering and onset detection, Goldberg Variations example

Figure 4: Analytic gammatone filtering with suppressed harmonics, Goldberg Variations example
Mbira music. The analysis was performed over 3 octaves descending from 1018.0 Hz using 9 non-equidistant filters per octave with $\Delta f / f_0 = 0.01$.

3 Conclusion

Examples showing the potential of our approach for the analysis of polyphonic music have been given. In order to alleviate the restriction of limited spectral overlap, the development of an adaptive lateral inhibition algorithm is envisaged.

References


