Efficient Spectral Power Estimation on an Arbitrary Frequency Scale

Filip ZAPLATA, Miroslav KASAL

Dept. of Radio Electronics, Brno University of Technology, Technická 12, 612 00 Brno, Czech Republic

Abstract. The Fast Fourier Transform is a very efficient algorithm for the Fourier spectrum estimation, but has the limitation of a linear frequency scale spectrum, which may not be suitable for every system. For example, audio and speech analysis needs a logarithmic frequency scale due to the characteristic of a human’s ear. The Fast Fourier Transform algorithms are not able to efficiently give the desired results and modified techniques have to be used in this case. In the following text a simple technique using the Goertzel algorithm allowing the evaluation of the power spectra on an arbitrary frequency scale will be introduced. Due to its simplicity the algorithm suffers from imperfections which will be discussed and partially solved in this paper. The implementation into real systems and the impact of quantization errors appeared to be critical and have to be dealt with in special cases. The simple method dealing with the quantization error will also be introduced. Finally, the proposed method will be compared to other methods based on its computational demands and its potential speed.

Keywords
Goertzel algorithm, Mel frequency cepstral coefficients, MFCCs, Q-constant transform, spectral power estimation

1. Introduction

Many techniques for spectral analysis are based on the short time Fourier spectra and since the Fast Fourier Transform (FFT) algorithms have been developed, the estimation of the spectra is even more suitable. Spectral resolution of the algorithms depends on the number of time domain samples that are processed, and the number of complex samples directly equals the resulted number of complex spectral lines (bins). The bins are equidistantly spread along the linear frequency scale in the range constrained by the sampling theorem.

There are situations in which the linear frequency scale is not convenient, e.g. speech or music analysis. Such inconvenience originates from the natural character of a human’s ear the frequency characteristic of which is far different from the linear spectral scale. The model of a human’s ear [1] says that the spectral resolution should be logarithmically distributed over the audible spectrum. Several techniques are used for the logarithmically distributed frequency scale spectrum estimation; the Mel-Frequency Cepstrum Coefficients (MFCC) [2] and the constant Q-transform [3] are the main ones.

The MFCC algorithm uses the spectral estimation of a sufficient frequency resolution computed by the FFT. The spectrum is then successively processed by triangular windows in accordance to spectral requirements. Windowing may be provided by a predefined filter bank. Filter bank outputs are converted to a logarithmic amplitude scale and processed by the Discrete Cosine Transform (DCT). Each output of the DCT then corresponds to the spectral line of the desired frequency scale. Over the years many computational forms have been developed which are more or less effective as e.g. in [4] or [5].

The constant Q-transform originates from the basic definition of the Discrete Fourier Transform (DFT), where the computation is not provided for all bins. J. Brown in [3] defined the algorithm for computing DFT parameters to fulfill predefined requirements. The original calculation can also be efficiently modified as in [6] or for audio signals in [7].

M. Tröbs and G. Heinzel also published a different approach [8], where they applied an averaging over modified periodogram. Their improved method computes the Fourier transform optimally for the logarithmic frequency scale, but is more convenient for a high resolution analysis and therefore is very complex.

The first step of mentioned algorithms is to estimate the spectrum of a huge frequency resolution to catch the differences between spectral components at low frequencies. The proposed algorithm uses a separate setting for each desired spectral component as similar to a bank of filters and is also useful for estimating the spectrum of an arbitrary frequency scale distribution while maintaining a high computational efficiency. On the other hand the price paid for its simplicity is a spectral leakage when using a rectangular weighting window. The main contribution of the text is a simple way to partially eliminate the leakage at frequencies in between the Fourier spectrum bins described in Sec. 3. For better possibility of comparison, the
logarithmic frequency scale distribution will be studied in the following text.

2. Efficient Approach to Spectral Power Estimation

In Fig. 1 the problem of using conventional Fourier transform algorithms for spectral estimation is depicted. Under the log-scale at low frequencies there is an obvious lack of information whereas at high frequencies there are many redundant bins that have to be averaged. Thus, even though the FFT is a very efficient algorithm for estimation of linearly distributed spectra, its computational demands grow rapidly if high resolution at low frequencies is required.

As obvious from the definition of the discrete Fourier transform

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}$$

(1)

individual bins can be calculated separately. The resolution of the spectrum depends on the number of processed samples $N$ and $k$ indexes the bins.

Obviously, this property helps to eliminate unnecessary bins and additionally set the resolution for each bin computation separately. The Fourier transform can be viewed as a bank of band-pass filters with an impulse response

$$h(n) = u_N(n) e^{j2\pi nk/N}$$

(2)

where $u_N(n)$ denotes the constraints of the sum (1), i.e. rectangular window. The power spectral density (PSD) characteristics of the filters can be derived by the method in [9] and an example is shown in Fig. 2.

From PSD it is clear that the length of the filter $N$ directly influences its bandwidth, further description of this relation is in [10]; notice also the zero transfer at frequencies of adjacent bins. The filters can then be set to analyze the signal under the desired scale; a nice example is a log-scale of base 2. The $k$ coefficient is constant over all bins and length $N$ is halved with each bin going to the highest frequencies and it causes the desired widening of its bandwidth as depicted in Fig. 3.

The direct evaluation of (1) can be replaced by the faster Goertzel algorithm. The algorithm is a convolutional form of (1) and the derivation can be found in [9]. It is used for fast calculation of selected bins and hence very suitable for our purpose. Its transfer function is shown in the following equation

$$H(z) = \frac{1 - z^{-1} e^{-j2\pi k/N}}{1 - 2z^{-1} \cos(2\pi k/N) + z^{-2}}$$

(3)

where the output sample is valid for every $N$th sample after which the filter has to be cleared. This condition provides the rectangular windowing of the impulse response.

The Goertzel algorithm is useful not only for discrete frequencies defined by $N$, i.e. $k \in \mathbb{Z}$, but as declared in [11] it can be used for frequencies in between the bins, i.e. $k$ may be real. This statement is correct, but suffers from spectral leakage more than for integer $k$. The reason why the leakage is higher is that since the filter is complex, the main passing lobe of the PSD is present only on one side of the complex spectrum. Thus, shifting the frequency response of non-integer $k$ causes the zero of the transfer function not to be at the position of the opposite frequency image, which then causes aliasing. The effect is clear from Fig. 4.

This effect is more apparent when the filter is tuned near the edges of a sampled region, where the images are very close, and when the spectral leakage is high, which is for a rectangular window. Another window with better

![Fig. 1.](image1) The upper image shows the linear distribution of the Fourier transform and the bottom image shows the same distribution under the log-scale.

![Fig. 2.](image2) PSD example for $N=10$ and $k=2$ (solid line) and $N=20$ and $k=8$ (dashed line).

![Fig. 3.](image3) Example of logarithmic frequency scale of base 2. From the left lobe $N=64 \ 32 \ 16 \ 8$ and $k=[2 \ 2 \ 2 \ 2]$. 

![Fig. 4.](image4) 


spectral leakage characteristics can be used by weighting the input signal because the impulse response is inaccessible within the effective realization of the Goertzel algorithm. The impact of weighting by Hanning window on the PSD has been presented in [10]. The additional windowing has to be applied before the filter and therefore it degrades useful power of the signal and significantly increases computational demands because each log-bin needs different $N$ and consequently the window size is also different.

3. A Novel Technique for Elimination of Additional Spectral Leakage for Arbitrary Frequencies

The spectral leakage is usually treated by windowing. Even in [11], dealing with arbitrary frequencies defined by real $k$, some weighting window is supposed to be used. Authors in [8] also met this issue and due to its minimization by windowing considered it as negligible. The following text presents a novel and more efficient approach to eliminating leakage of the mirror image of the desired frequency and therefore providing better characteristic for real $k$ and rectangular weighting window.

First, we have to find an analytic response of the bin calculation to a harmonic signal of the same frequency. The solution of the limited convolution is

$$Y_k(n) = \sum_{m=0}^{N-1} M \cdot \cos \left(2\pi \frac{k}{N} (m + \phi) \right) e^{j2\pi \frac{k}{N} (n - m)}$$

$$= \frac{M}{2} \left[ N e^{-j\phi} + e^{j\phi} \frac{1 - e^{-j4\pi k}}{1 - e^{-j4\pi \frac{k}{N}}} \right]. \quad (4)$$

The bin value is newly denoted as $Y$ and is proposed to be dependent on time samples $n$ instead of one time sample $X$ in (1). The applied harmonic signal has magnitude $M$ and phase $\phi$. Note that for integer $k$ the second term of the solution is zero and the result simplifies to a complex value expected for the Fourier bin power. Substituting the parasitic value

$$A = \frac{1 - e^{-j4\pi k}}{1 - e^{-j4\pi \frac{k}{N}}} \quad (5)$$

and the desired value

$$B = Me^{-j\phi} \quad (6)$$

the response (4) can be rewritten as

$$Y_k(n) = \frac{1}{2} \left( BN + B^* A \right). \quad (7)$$

Equation (7) can now be solved for unknown $B$ as

$$B = \frac{2}{AA^* - N^2} \left( AY_k - NY_k \right) \quad (8)$$

where $Y_k$ is the output of the uncompensated Goertzel filter.

The optimized way for signal power computation of the original Goertzel algorithm

$$\left| Y_k(n) \right|^2 = d(n)^2 + d(n-1)^2 - C d(n) d(n-1) \quad (9)$$

has been obtained by powering the complex Goertzel output $X_k$ [12]. Variables $d(n)$ and $d(n-1)$ denote delayed samples within the recursive part of the algorithm [12]. $C$ denotes the frequency coefficient (10), which can also be evaluated by corresponding absolute frequencies $f_c$ (tuned frequency) and $f_s$ (sampling frequency),

$$C = 2\cos \left(2\pi \frac{k}{N} \right) = 2\cos \left(2\pi \frac{f_c}{f_s} \right). \quad (10)$$

The solution of (8) can be put into (9) in the form of additional coefficients resulting in

$$\left| Y_k(n) \right|^2 = d(n)^2 Q_{M1} + d(n-1)^2 Q_{M2}$$

$$-C d(n) d(n-1) Q_{M3}. \quad (11)$$

The compensation coefficients are then

$$Q_{M1} = \frac{4 \left( A_k^2 + A_t^2 + N^2 - 2 A_R N \right)}{\left( A_k^2 + A_t^2 - N^2 \right)^2}, \quad (12)$$

$$Q_{M2} = \frac{4 \left( A_k^2 + A_t^2 + N^2 - 2 N \left( A_R \cos \left(4\pi \frac{k}{N} \right) + A_t \sin \left(4\pi \frac{k}{N} \right) \right) \right)}{\left( A_k^2 + A_t^2 - N^2 \right)^2}, \quad (13)$$

$$Q_{M3} = \frac{4 \left( A_k^2 + A_t^2 + N^2 - 2 N \left( A_R \sin \left(2\pi \frac{k}{N} \right) \cos \left(2\pi \frac{k}{N} \right) \right) \right)}{\left( A_k^2 + A_t^2 - N^2 \right)^2}. \quad (14)$$

For more convenient evaluation, the substituted complex constant $A$ is split into real and imaginary parts $A_R$ and $A_I$ respectively. Notice again the state where the frequency coefficient $k$ is integer and $A$ is then zero. All coefficients simplify to only weighting constant $(2/N)^2$ that is common for the Fourier transform power computation.

The evolution of the coefficients is shown in Fig. 6.
It is obvious that the most critical are the frequencies near the edges of the sampled region where the coefficients grow to infinity. Otherwise the coefficients slightly fluctuate around the weighting value \((2/N)^2\), which equals 1/9 in this example. The compensation effect is clearly seen on the waterfall spectrogram in Fig. 5; the upper image is for the original uncompensated algorithm, where strong fluctuations at the tuned frequency \((k = 2.3)\) are seen. Notice also the magnitude range, which goes even over the unity value. The bottom image shows the output signal after compensation, the fluctuations are effectively suppressed at the tuned frequency.

4. Quantization Error of the Coefficients

When implementing whatever signal processing algorithm, the issues about limited accuracy computing has to be taken into account. The general problem of overflowing and rounding error has been discussed e.g. in [12]. The solution of these issues mainly lies in properly set accuracy of the fixed point calculations and appropriate order of arithmetic operations. However, the quantization error of the coefficients is a more complicated issue which may cause total failure of the system and therefore it should be given attention.

In Fig. 7, there are two lines, the solid one shows the real quantization error of coefficient \(C\) when 4 bits of memory are allocated for its fractional part. The dashed curve shows the maximal error given by

\[
E_{qc} = \frac{\cos^{-1}\left(\cos\left(\frac{2\pi k}{N}\right) - \frac{0.5}{R}\right)}{2\pi}
\]

and is the envelope of the real error. The coefficient resolution is given by the quantum count \(R\), which is \(R = 2^4\) for 4-bit resolution. The equation was obtained by the reverse tuning coefficient computation, i.e. the absolute frequency error is the frequency calculated back from the ideal tuning coefficient distorted by a half quantum of fixed point num-

Fig. 5. Waterfall spectrogram of the original method and compensated method for \(N = 10\) and \(k = 2.3\).

Fig. 6. Evolution of the magnitude compensation coefficients: example for \(N = 6\).

Fig. 7. Absolute quantization error of coefficient \(C\) for quantum count \(R = 16\).
ber representation. It is now clear that the most critical are the edges of the frequency range, where the error grows rapidly. Note also, that equation (15) does not have a real solution for all frequencies, where the argument of the \( \cos^{-1}(x) \) function can run out of the range \((-1; 1)\). Such an occurrence has to be prevented during the design.

The quantization error of coefficient \( C \) is particularly critical for log-scale frequency spectral estimation, where an occurrence has to be prevented during the design. The need of accuracy in frequencies grows at low frequencies. A possible solution for this issue is to raise the quantization accuracy of the system, i.e. raise \( R \), or to move the ratio \( k/N \) to a less critical region. The second way can be maintained by pre-filtering and decimation.

A simple low-pass filter is the moving average, which in fact is a limiting case of the Goertzel algorithm and therefore the equation

\[
H_G(f) = \frac{1}{N} \sin \left( \frac{\pi N \left( f - \frac{k}{N} \right)}{N} \right) e^{j2\pi \left( f - \frac{k}{N} \right)^{1-N}}
\]  

(16)

[9] can be used for calculating the frequency response of the moving average filter

\[
|H_{MA}(f)| = |H_G(f)|_{k=0} = \frac{\sin(\pi N_{MA} f)}{N_{MA} \sin(\pi f)}.
\]  

(17)

The impulse response is then

\[
h_{MA}(n) = \frac{1}{N_{MA}} u_{NMA}(n)
\]  

(18)

which is a rectangular pulse in fact.

The moving average is realized as the summing of \( N_{MA} \) samples and weighting them by its number before letting them into the Goertzel filter. The design of the Goertzel coefficients then has to include \( N_{MA} \) times decimation as a substitution

\[
\frac{k}{N} \rightarrow \frac{N_{MA} k}{N} = \frac{N_{MA} f_c}{f_s},
\]  

(19)

i.e. to give the desired frequency response, the \( k/N \) ratio of the Goertzel filter is \( N_{MA} \) times raised.

The same approach can be used to design the high-pass filter resulting in the impulse response

\[
h_{HP}(n) = \frac{1}{N_{HP}} e^{j\pi n} = \frac{1}{N_{HP}} (-1)^n
\]  

(20)

and magnitude frequency response

\[
|H_{HP}(f)| = |H_G(f)|_{k=0} \frac{\sin \left( \pi N_{HP} \left( f - \frac{1}{2} \right) \right)}{N_{MA} \sin \left( \pi \left( f - \frac{1}{2} \right) \right)}.
\]  

(21)

From the impulse response it can be derived that realization of the high-pass filter lies also in summing of \( N_{HP} \) samples, but with alternating signs.

The design of the Goertzel filter coefficients then has to include substitution according to formula

\[
\frac{k}{N} \rightarrow \frac{N_{HP} k}{N} = \frac{N_{HP} f_c}{f_s}.
\]  

(22)

This proposed solution helps to eliminate the quantization error of the coefficient \( C \), however, very simple filters cause aliasing and distortion; therefore to reach the best performance, the system has to be optimized, or other better filters have to be used.

The compensation coefficients \( Q_{M1}, Q_{M2}, Q_{M3} \) do not affect the recursive part of the filter and therefore, their impact on the results is weak and can be easily predicted from their values, as seen in Fig. 6.

5. Efficiency Comparison

Finally, the proposed method is briefly compared to the FFT methods from an efficiency point of view. The efficiency or computation demands can be assessed from the theoretical number of multiplications and additions per frame or in our case per fixed number of samples.

The realization equations of the Goertzel algorithm are summarized as

\[
d(n) = x(n) + C d(n-1) - d(n-2)
\]  

(23)

and

\[
Y_k(n) = Q_{M1} d(n)^2 + Q_{M2} d(n-1)^2 - C Q_{M3} d(n) d(n-1).
\]  

(24)

From the equations the number of real multiplications per frame is \( N + 6 \) and the number of real additions is \( 2(N + 1) \). Assuming the log-scale distribution of base 2, as in Fig. 3, results in the number of multiplications

\[
M_G = 2^K \left( K 2^{n+1} + 6 \right) - 6
\]  

(25)

and additions

\[
A_G = 2^K \left( K 2^{n+4} + 4 \right) - 4
\]  

(26)

for \( K \) log-scale bins. The frame size of each bin is different, so the number of samples, to which the results are compared, is the length of the longest frame. The longest frame also defines the length of the FFT algorithm used for the same frequency resolution within other spectral analysis algorithms. The value

\[
p = \log_2 N_K
\]  

(27)

defines the bandwidth of the bin tuned to the highest frequency, i.e. with the widest bandwidth \( N_K \).

The fast Fourier transform radix-2 algorithm has the following number of multiplications
\[ M_{\text{FFT}} = N \log_2 N + 2N \] (28)

and additions
\[ A_{\text{FFT}} = 2M \log_2 N + N, \] (29)
when used for power spectrum computation.

In Tab. 1, there are a few values of the Goertzel algorithm computational demands in comparison with its equivalent FFT algorithm for \( p = 3 \). For example for the row for eight log-scale bins at relative frequencies \( \frac{2}{8}, \frac{2}{16}, \frac{2}{32}, \frac{2}{64}, \frac{2}{128}, \frac{2}{256}, \frac{2}{512}, \frac{2}{1024} \) the relative deviation is -20.9\% and -19.1\% for multiplications and additions respectively compared at 1024 samples of the input signal. The comparison was provided as a relative deviation of the computational demands against the reference FFT radix-2 power spectrum algorithm. In all cases, there appears the minus sign which means an improvement. Notice also that for very large number of bins \( K \), equation (25) and (26) converge to the FFT demands (28) and (29), because in such a case \( K = \log_2 N \).

<table>
<thead>
<tr>
<th>( K )</th>
<th>( M_C )</th>
<th>( M_{\text{FFT}} )</th>
<th>( \delta_{M} )</th>
<th>( M_A )</th>
<th>( A_{\text{FFT}} )</th>
<th>( \delta_A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>346</td>
<td>512</td>
<td>-32.4%</td>
<td>572</td>
<td>832</td>
<td>-31.3%</td>
</tr>
<tr>
<td>8</td>
<td>9722</td>
<td>12288</td>
<td>-20.9%</td>
<td>17404</td>
<td>21504</td>
<td>-19.1%</td>
</tr>
<tr>
<td>16</td>
<td>4587514</td>
<td>5242880</td>
<td>-12.5%</td>
<td>8650748</td>
<td>9699328</td>
<td>-10.8%</td>
</tr>
<tr>
<td>32</td>
<td>5.76E+11</td>
<td>6.18E+11</td>
<td>-6.9%</td>
<td>1.12E+12</td>
<td>1.19E+12</td>
<td>-5.8%</td>
</tr>
<tr>
<td>64</td>
<td>4.83E+21</td>
<td>5.02E+21</td>
<td>-3.7%</td>
<td>9.52E+21</td>
<td>9.81E+21</td>
<td>-3.0%</td>
</tr>
</tbody>
</table>

Tab. 1. An example of computational demands.

6. Conclusion

The Goertzel algorithm has been introduced in a role of an efficient bank of filters for an arbitrary frequency scale spectral analysis. The algorithm implicitly uses a rectangular weighting window and this causes strong spectral leakage, which is a price paid for its simplicity. The Goertzel algorithm is able to be tuned to whatever real frequency within the sampled range, but has stronger leakage caused by the mirrored central frequency. The novel method of effective mirror image elimination has been introduced. The elimination lies in compensation provided on the output power side with 3 additional, mostly non-critical, coefficients which are much more efficient compared to window weighting.

Implementation issues are a very important part of the system design and the Goertzel filter has a very critical issue in the quantization error of the tuning coefficient, which affects the recursive part. The proposed solution helps to significantly reduce the tuning error, but the price paid is the aliasing of leaking signals from the stop-band.

The efficiency of the algorithm was assessed by comparing the FFT radix-2 algorithm of equivalent length. The number of multiplication and addition operations per a fixed number of samples was compared. The result showed that the algorithm is slightly less demanding, but considering that the FFT spectral evaluation is only a part not giving the desired arbitrary frequency scale power spectrum, the proposed algorithm is even more efficient.

The algorithm has already been implemented within an audio amplifier to visualize an input signal spectrum in the logarithmic frequency scale. The display resolution is 16 bins of 4-bit depth. In such cases the spectral leakage of the rectangular window is not disturbing; moreover the speed of the algorithm is excellent.

Acknowledgments

The presented research was financed by the Czech Ministry of Education in frame of the National Sustainability Program, the grant LO1401 INWITE. For the research, infrastructure of the SIX Center was used.

References


About the Authors ...

Filip ZAPLATA was born in Chrudim, Czech Republic in 1987. He received the M.Sc. degree in Electronics and Communications from the Brno University of Technology (BUT) in 2011. From 2011, he was a combined Ph.D. student and simultaneously working as an R&D Engineer in Vesla s.r.o. company. Since 2012 he has been a full-time Ph.D. student at the Department of Radio Electronics, BUT. His research interests include digital signal processing algorithms for software defined radio and digital receiver technology.

Miroslav KASAL was born in Litomysl, Czech Republic. He graduated in Communication Engineering from the Faculty of Electrical Engineering, Brno University of Technology, in 1970. In 1984 he obtained his Ph.D. degree in Metering Engineering. He was the head of the NMR Department and Electronics Laboratory of the Institute of Scientific Instruments, Academy of Science of the Czech Republic (1991-2002). Since 2002 he has been with the Department of Radio Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology as a professor. He is a senior member of the IEEE. In 2007, prof. Kasal received the Prize for Research of the Minister of Education of the Czech Republic.