A flexible multi-resolution time-frequency analysis framework for audio signals

T. Fillon, J. Prado

A FLEXIBLE MULTI-RESOLUTION TIME-FREQUENCY ANALYSIS FRAMEWORK FOR AUDIO SIGNALS

*Thomas Fillon**

Jacques Prado

Institut Telecom, Telecom-ParisTech, CNRS LTCI,
37/39 rue Dareau, 75014 Paris, France

ABSTRACT

In this article, a new Constant-Q transform implementation is proposed together with a generalization of both the Short-Term Fourier Transform and the Constant-Q transform. The purpose of this generalization is to provide a discrete time-frequency analysis tool with arbitrary center frequency and frequency resolution for each bin of the transform. This new analysis framework is very flexible and can be related to many common time-frequency transforms. Furthermore, new interesting transforms with specific time-frequency resolutions can be defined. To illustrate and validate the corresponding approach, experimental results and examples are provided for different configurations.

1. INTRODUCTION

The Short-Term Fourier Transform (STFT) is widely used in audio signal processing since it provides an efficient analysis of non-stationary signals with varying components in time and frequency. The STFT analysis is performed over an equally spaced frequency distribution with a constant frequency resolution. The Constant-Q Transform (CQT) is an alternative time-frequency analysis method that enables to analyze musical signals with a frequency resolution higher enough to separate the different notes within an octave. As introduced by Brown in [1], the Constant Q transform is defined as being equivalent to a 1/24-octave filterbank but it can be easily extended to any $1/B$-octave frequency resolutions. Other typical choices for B are 12, 36 or 48 bins per octave. What makes CQT appealing for music signal processing is the fact that the center frequencies of the analysis are not uniformly distributed like in the Discrete Fourier Transform (DFT) but are aligned with the notes of the equal tempered scale. Moreover the frequency resolution of the CQT is not constant over the whole spectrum but has a constant Q-factor.

Another alternative time-frequency transform has been proposed in [2], the Multi-Resolution Fast Fourier Transform (FFT). It provides a way to enhance the frequency resolution in the low frequency bands by adapting the analysis windows length in a band-wise fashion and keeping the hop-size constant for every band.

*This work was supported under the research programme Quaero, funded by OSEO, the French State agency for innovation.

In this paper, in section 2, some properties of the CQT implementations are first discussed and some improvements are proposed. In section 3, a generalized formalism is derived from the CQT. This new formalism allows us to define a new time-frequency analysis framework with arbitrary center frequencies and arbitrary frequency resolution. All parameters of this framework are explicitly defined in order to provide the proper configuration to achieve the right frequency resolution at any given frequency. In section 4, some particular examples of implementation are given to illustrate the advantage of such a flexible framework.

2. CONSTANT-Q TRANSFORM

2.1. CQT implementations

The Discrete Fourier Transform of a sequence $x \in \mathbb{R}^N$ at an arbitrary frequency f_k (in Hz) is given by :

$$x^{dft}[k] = \sum_{n=0}^{N-1} w[n]x[n]e^{-j2\pi n\frac{f_k}{f_s}} \qquad (1)$$

where f_s is the sampling frequency and $\{w[n]\}_{n\in[0,N-1]} \in \mathbb{R}^N$ is an analysis window.

The spectral resolution (or bandwidth) Δf of the DFT is inversely proportional to the number of samples and vary according to the chosen temporal window function.

In order to provide a constant Q-factor analysis for each frequency bin, the CQT was proposed in [1] and further developped in [3]. The CQT adjusts the frequency resolution by allowing the use of a different analysis window $\mathbf{w}_k$ for each frequency bin k. According to the original definition, the k^{th} bin of the CQT of a discrete signal sequence $x \in \mathbb{R}^N$ is given by :

$$x^{cqt}[k] = \frac{1}{N_k}\sum_{n=0}^{N-1} w_k[n]x[n]e^{-j2\pi n\frac{f_k}{f_s}} \qquad (2)$$

where

- $\{w_k[n]\}_{n\in[0,N_k-1]}$ is an analysis window of length $N_k \in \mathbb{N}$,
- N is the length of the signal block under analysis such as $\forall k \in [0; K-1], N \geq N_k$,

978-1-4673-0382-8/12/$31.00 ©2012 IEEE

- $f_k = (2^{1/B})^k \cdot f_{min}$ are the geometrically spaced center frequencies aligned with the frequencies of the equal tempered scale,
- f_{min} is the minimal frequency to be analyzed, corresponding to the first bin, $k = 0$,
- B is the number of bins per octave (typically $B = 12$ or 24).

Each window $\{w_k\}_{n\in[0,N_k-1]}$ is derived from a common canonical window function (*e.g.* Rectangular, Hamming, Blackman). As the Q-factor is required to be constant for every bin of the CQT, then the analysis window length N_k must vary according to the center frequency f_k. In [1], N_k is given by :

$$N_k = \widehat{Q} \cdot \frac{f_s}{f_k} \tag{3}$$

where $\widehat{Q}$ depends on B, the number of bins per octave, and is given by :

$$\widehat{Q} = 1/(2^{1/B} - 1) \tag{4}$$

The formulation of the CQT has been further developed in [3, 4] mainly to provide fast computation methods.

2.2. Discussion

The Q-factor is defined by $Q = \frac{f}{\Delta f}$ where f is the center frequency of the filter and Δf is the $-3dB$ bandwidth. From classic theory on digital filter design by windowing [5, 6], the $-3dB$ bandwidth of the filter corresponding to the k^{th} bin of the CQT is given by :

$$\Delta f_k = \Delta\omega \cdot fs/N_k \tag{5}$$

where $\Delta\omega$ is a real factor in bin value depending on the window type. Common values of $\Delta\omega$ are indicated in Table 1.

Window type	$\Delta\omega$	$\delta\omega$	$\Pi\omega$
Rectangular	0.89	2	1.00
Hann	1.44	4	1.50
Hamming	1.30	4	1.36
Blackman	1.68	6	1.73

Table 1. Characteristics of some common window functions from [5, 6]. $\Delta\omega$, $\delta\omega$ and $\Pi\omega$ are respectively the $-3dB$ bandwith, the main lobe width and the Equivalent Noise Bandwith of the filter expressed in bin.

In (3), the expression of N_k does not depend on the type of window and does not take into account the true $-3dB$ bandwidth of the filter. Thus, the parameter $\widehat{Q}$ as given by [1, 3] in (3) and (4) is not the proper Q-factor of the filter. The proper one is given by Schörkhuber and Klapuri in [4] :

$$Q = \frac{N_k \cdot f_k}{\Delta\omega \cdot f_s} \tag{6}$$

In the same article, the CQT is designed such as the value of the Q-factor is :

$$Q = \frac{q}{\Delta\omega} \cdot \widehat{Q} \tag{7}$$

where $0 < q \leq 1$ is a scaling factor. Hence, the corresponding values for the windows length N_k is given by :

$$N_k = q \cdot \widehat{Q} \cdot \frac{f_s}{f_k} \tag{8}$$

For $q < 1$, the frequency resolution would be lowered but this would improve the time resolution and could improve the accuracy of the signal reconstruction after inversion of the CQT. One can notice that equations (8) and (3) only differ by the factor q and that both expressions are independant of the windows type. Thus for different analysis windows, the Q-factor would be different as shown in (7).

According to the previous remarks, a new implementation of the CQT is proposed in the following in order to provide some improvements over the various implementations in [1, 3, 4].

Frequency resolution In the context of audio signal analysis, if the requirement is to separate two neighboring tones, it could be more relevant to take into consideration the main lobe width rather than the $-3dB$ bandwidth when designing the analysis window. For the most common analysis windows, the main lobe width is given by :

$$\delta f_k = \delta\omega \cdot \frac{fs}{N_k} \tag{9}$$

where $\delta\omega$ is a real factor depending on the window type. Common values of $\delta\omega$ are indicated in Table 1.

In the proposed CQT implementation, the frequency resolution corresponds to the main lobe width of each filter. Contrary to previous implementations, we propose here to define N_k in order to provide the chosen frequency resolution independently of the type of the analysis window. N_k is thus given by :

$$N_k = \delta\omega \cdot \widehat{Q} \cdot \frac{f_s}{f_k} \tag{10}$$

It should be noticed that δf_k and Δf_k are both inversely proportional to N_k, thus if $f_k/\delta f_k$ is constant then the Q-factor is also constant.

Normalization In (2), the normalization factor for the k^{th} bin is N_k. In order to make the CQT coefficients amplitude almost independent from the choice of the analysis window, it could be more relevant to normalize by γ_k defined by :

$$\gamma_k = \sum_{n=0}^{N_k-1} w_k[n] \tag{11}$$

Temporal Alignment of the kernels The last proposed modification of the CQT implementation, consists in enabling different alignments of the temporal kernels inside the large N-length window.

Therefore, the final expression of the proposed CQT is given by :

$$x^{cqt}[k] = \frac{1}{\gamma_k} \sum_{n=0}^{N-1} a_k[n] x[n] e^{-j2\pi n \frac{f_k}{f_s}} \tag{12}$$

where the temporal window a_k is defined by :

$$a_k[n] = \begin{cases} 0 & \text{if } n < \tau_k \\ w_k\left[n - \tau_k\right] e^{j2\pi\tau_k \frac{f_k}{f_s}} & \text{if } \tau_k \leq n \leq \tau_k + N_k - 1 \\ 0 & \text{if } \tau_k + N_k \leq n \end{cases} \tag{13}$$

and where $\tau_k \in \mathbb{N}$ is a time delay that enables to control the temporal alignment of the K atoms inside the large N-length frame such that for :

$$\tau_k = \begin{cases} 0, & \text{All kernels are left aligned} \\ \frac{N-N_k}{2}, & \text{All kernels are centered} \\ N - N_k, & \text{All kernels are right aligned} \end{cases} \tag{14}$$

3. GENERALIZED FRAMEWORK FOR A MULTI-RESOLUTION SIGNAL ANALYSIS

3.1. Definition

The definition of the CQT given in (12) can be easily extended to arbitrary center frequencies $\{f_k\}$ and frequency resolutions $\{\delta f_k\}$. Indeed, it can be useful to adapt the discrete time-frequency analysis to the requirements of various application contexts. To that purpose, a flexible generalization of the CQT formalism is developed in the following.

Let us consider a discrete time-frequency analysis at the sampling frequency f_s. This multi-resolution time-frequency transform is determined by :

- an arbitrary set of K frequencies $\{f_k\}_{k\in[0;K-1]}$, with $K \in \mathbb{N}$.
- an arbitrary set $\{\delta f_k\}_{k\in[0;K-1]}$ of K frequency resolutions.
- a type of analysis windows $\mathcal{W}$ (Rectangular, Hamming, Blackmann, ...).

Following the formalism of the CQT in (12) this Multi-Resolution Transform (MRT) can be defined by :

$$x^{mrt}[k] = \frac{1}{\gamma_k} \sum_{n=0}^{N-1} a_k[n] x[n] e^{-j2\pi n \frac{f_k}{f_s}} \tag{15}$$

where γ_k is a normalization factor as defined in (11) and where the temporal window a_k is defined in (13).

Given the desired frequency resolution δf_k and the analysis windows $\mathcal{W}$, one can determine the appropriate length N_k of the analysis window $\{w_k\}$ to fulfil the frequency resolution at the k^{th} bin of the transform. To enable a further flexible design, the frequency resolution can correspond either to the main lobe width or to the $-3dB$ bandwidth or to the equivalent noise bandwidth of the filter. According to (5), (9) and Table 1, N_k is then defined by :

$$N_k = \begin{cases} \delta\omega \cdot f_s/\delta f_k & \text{, for the main lobe width} \\ \Delta\omega \cdot f_s/\delta f_k & \text{, for the } -3dB \text{ bandwidth} \\ \Pi\omega \cdot f_s/\delta f_k & \text{, for the equivalent noise bandwidth} \end{cases} \tag{16}$$

where $\Pi\omega$ is a real factor depending on the window type. Common values of $\Pi\omega$ are indicated in Table 1.

For the CQT, to compute the transform, the signal is segmented in N-length overlapping frames where $N \geq N_k, \forall k \in [0; K-1]$. Thus, N is determined by the highest resolution value in δk.

In (15), it should be noted that the frequency resolutions of the MRT is defined by the type of analysis windows and by their length N_k. In order to ensure that the frequency resolution of each bin k of the MRT matches the required frequency resolution δf_k a proper value of N_k has to chosen.

3.2. Vector formulation of the proposed Multi Resolution Transform

Let us consider a vector formulation of the proposed Multi-Resolution Transform. Given a column vector of a real sequence $\mathbf{x} \in \mathbb{R}^N$, the MRT of $\mathbf{x}$ is given by :

$$\mathbf{x}^{mrt} = \mathbf{A}^H \mathbf{x} \tag{17}$$

where, from (15), the transform matrix $\mathbf{A}^H$ of the MRT is defined by :

$$A[n,k] = \frac{1}{\gamma_k} a_k[n] . e^{j2\pi n \frac{f_k}{f_s}} \tag{18}$$

A fast and efficient implementation of the CQT was proposed in [3]. It consists of implementing the matrix product in the Fourier domain.

$$\mathbf{x}^{mrt} = A^H \mathbf{x} = A^H W_F^{-1} W_F \mathbf{x} = \mathcal{A}_F{}^H \mathbf{x}^{dft} \tag{19}$$

where W_F is the N-point DFT matrix and $\mathcal{A}_F$ and $\mathbf{x}^{dft}$ are the respective Discrete Fourier Transform of A and $\mathbf{x}$.

In [7], the authors introduced a tensor product formulation of the Short-Term Fourier Transform (STFT). Following this approach, the Multi-Resolution Transform can be expressed for a set of successive and overlapping frames. Given a sequence of $L \in \mathbb{N}$ samples in $\mathbb{R}^L$ with $L > N$:

$$\underline{\mathbf{x}} = [x[0], \ldots, x[L-1]]^T \tag{20}$$

To compute the STFT with a hop-size of R samples, this sequence is segmented into P segments that overlap by $N - R$ samples. The segmenting and overlapping processes inside the STFT computation can be expressed by the product :

$$\mathbf{X} = \mathbf{O}\underline{\mathbf{x}} \tag{21}$$

where $\mathbf{O}$ is the $[PK \times PN]$ overlap matrix operator defined by :

$$\mathbf{O} = \begin{bmatrix} \boxed{\mathbf{I}_N} & & & \\ \underset{R}{\rightarrow}\boxed{\mathbf{I}_N} & & \mathbf{0} & \\ \underset{R}{\rightarrow}\underset{R}{\rightarrow}\boxed{\mathbf{I}_N} & & & \\ & & \ddots & \\ \mathbf{0} & & & \boxed{\mathbf{I}_N} \end{bmatrix} \quad (22)$$

and $\mathbf{X}$ is the $[PN \times 1]$ column vector that concatenates all the successive signal frames. According to (22), the STFT of $\underline{\mathbf{x}}$ is given by :

$$\begin{aligned} \mathcal{X}^{dft} &= (I_K \otimes W_F)\, \mathbf{O}\underline{\mathbf{x}} \\ &= (I_K \otimes W_F)\, \mathbf{X} \end{aligned} \quad (23)$$

where $\otimes$ is the kronecker product and $\mathcal{X}^{dft}$ is the $PN \times 1$ vector concatenating the successive STFT frames. The equivalent formulation for the MRT is :

$$\mathcal{X}^{mrt} = \left(I_K \otimes \mathbf{A}^H\right) \mathbf{O}\underline{\mathbf{x}} \quad (24)$$

It can easily be derived from (19) and (23) that :

$$\mathcal{X}^{mrt} = \left(I_K \otimes {\mathcal{A}_F}^H\right) \mathcal{X}^{dft} \quad (25)$$

Despite the fact that $I_K \otimes {\mathcal{A}_F}^H$ is a high-dimension matrix of size $[PK \times PN]$ it should be noticed that it is sparse and most of its coefficients are zeroes.

Equations (24) and (25) point out the interesting property that the MRT can be derived from $\underline{\mathbf{x}}$ and its STFT by linear algebraic matrix product.

Several methods for the inversion of the CQT have been proposed in [4, 8]. The tensor product formulation introduced here aims at providing a convenient algebraic framework which can be used to elaborate inversion methods.

From (25) one can easily derive a least-mean square estimation of the inverse transform :

$$\widehat{\mathcal{X}^{dft}} = \left(I_K \otimes {\mathcal{A}_F}^H\right)^{\dagger} \mathcal{X}^{mrt} \quad (26)$$

where $(\,)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse operator [9]. By property of the pseudo-inverse, one can simplify the previous expression in :

$$\widehat{\mathcal{X}^{dft}} = \left(I_K \otimes ({\mathcal{A}_F}^H)^{\dagger}\right) \mathcal{X}^{mrt} \quad (27)$$

The time-frequency analysis performed by the CQT only considers the positive frequencies $f_k \leq f_s/2$. In order to provide a proper signal reconstruction for the MRT, the negative frequencies should be taken into account. As it is the case in the DFT, this can be done by augmenting f_k with the corresponding set of negative frequencies.

The inversion method provided by (27) has been tested on different signals and for different configuration of the MRT and it provides satifactory results. A complete analysis of these results and a comparison with existing methods will be provided in a future extension of this work.

4. EXAMPLES OF IMPLEMENTATIONS

4.1. Links to existing transforms

The Multi Resolution Transform[1] introduced in the previous sections is very flexible and can be used in many different configurations.

For a uniformly spaced center frequency set $\{f_k\}_k \in [0; K-1]$ for $K \in \mathbb{N}$ and a constant frequency resolution for all the MRT bins, $\delta f_k = \delta\omega f_s/K$, it is straightforward to see that the MRT is equivalent to the K-length uniform DFT.

Now, suppose that the frequencies are grouped in several sub-bands, the frequency resolution can be kept constant for every bin inside a given sub-band but vary across the sub-bands. The corresponding MRT would then be similar to the Multi-Resolution FFT in [2].

In an other fashion, the frequency distribution $\{f_k\}$ can be chosen to match the 1/24-octave decomposition of the original CQT. This particular frequency distribution has many advantages for musical signal analysis and it is widely used for multi-pitch estimation where harmonic template kernels can be defined and where a translated version of a given kernel can model several notes (see [10]). Given this configuration, one can choose to keep the frequency resolution constant for each bin to evaluate the DFT coefficients over a musically relevant frequency distribution. If the frequency resolution is required to have a constant Q-factor then the MRT is the CQT. The relation between different configurations of the MRT and some classic time-frequency analysis methods is illustrated in Table 2. Similarly, in Figure 1, different configurations of the frequency distribution and frequency resolution of the MRT are displayed to illustrate the flexibility of such a framework.

δf_k	f_k	
	Uniform	1/B-octave
Constant	DFT	log-DFT
Bandwise Constant	MR-FFT	-
Constant Q-factor	-	CQT

Table 2. Example of implementation of the proposed Multi-Resolution Transform for different choices of frequency distribution $\{f_k\}_{k\in[0;K-1]}$ and frequency resolutions $\{\delta f_k\}_{k\in[0;K-1]}$ and the corresponding classic transforms

4.2. Mixed Resolution CQT

One advantage of the CQT is its ability to perform analysis over long time windows for low frequency bins. Nevertheless, for very low frequencies, the length of the windows can be irrelevant and can generate temporal smearing. Indeed, the temporal resolution is defined by $\Delta t_k = N_k/f_s$. For a typical CQT implementation over 7 octaves with

[1] The code corresponding to this Time-Frequency analysis framework is avalaible at `http://www.tsi.telecom-paristech.fr/aao/?p=676`

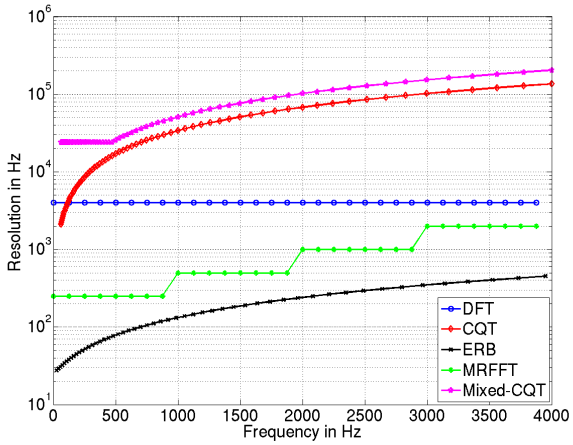

Fig. 1. Resolution vs frequency values for different implementations of the Multi-Resolution Transform

$B = 24$ and $f_{min} = 55Hz$, the time resolution for the lowest frequency bin would be of $0.65s$ and about $4ms$ for the highest frequency bin. Both values may not be appropriate for musical applications in regards of the potential duration of notes. The temporal smearing artefact is illustrated in Figure 2 which displays the CQT representation for a 5s musical excerpt of a bass pattern. In this figure, the low frequencies of the CQT exhibits a lot of temporal smearing and make it very difficult to analyze because the duration of the windows exceeds the note durations and superimposes successive notes.

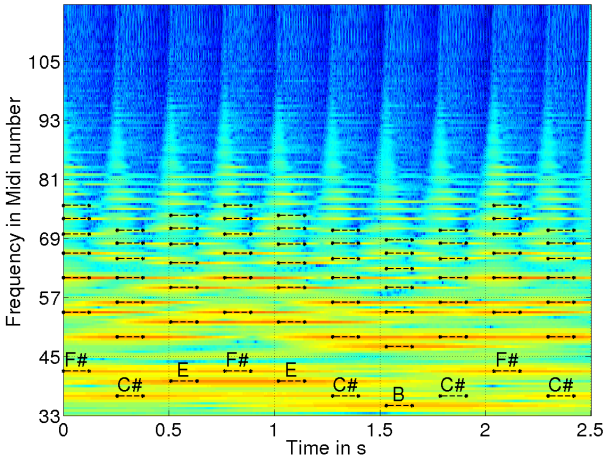

Fig. 2. MRT analysis of a musical excerpt (Bass pattern) with a Constant-Q resolution. The audio has been generated from a midi file. The note names, onset times and durations are indicated with the corresponding harmonics.

One of the justifications for the use of the CQT is the approximative constant Q-factor of the auditory filter above 500Hz. But below that frequency, the Q-factor is no more constant and is in fact decreasing [11]. Thus the psychoacoustic justification of the CQT is not relevant at those frequencies. According to [11], the bandwith of the auditory filter expressed in Equivalent Rectangular Bandwidth (ERB) is related to the center frequency f in Hz by :

$$ERB_f = 0.108 * f + 24.7 \quad (28)$$

Thus the equivalent Q-factor for the auditory filter is :

$$Q = \frac{f}{ERB_f} = Q_{HF} - \frac{\lambda}{f + \lambda} \quad (29)$$

where $Q_{HF} = 1/0.108$ and $\lambda = 24.7/0.108$. From (29), one can see that $Q \approx Q_{HF}$ for the high frequencies and that Q is not constant but decreasing in the low frequencies.

According to this consideration and the temporal smearing of the CQT in low frequencies, it could be appropriate in some applications to alleviate the constant-Q constraint in the low frequencies. To illustrate this purpose, a MRT with constant-Q resolution filter above 500Hz and with constant resolution frequency below 500Hz has been computed on the previous musical excerpts. The results are displayed in Figure 3. In this figure, one can notice that the temporal resolution has been improved for the lower part of the spectrum and the temporal smearing artefact has been reduced.

Of course, this improvement of the temporal resolution in the low frequencies is counterbalanced by a decrease in the frequency resolution and a problem of frequency smearing which prevents a quarter-tone resolution at those frequencies. Nevertheless, in the case of music applications, it can be noticed that the potential ambiguity in the frequency estimation for the fundamental frequency could be resolved thanks to its higher harmonics that are properly separated from neighboring tones. The counterpart does not hold in the time domain because harmonics can be shared between successive notes. This phenomenon is illustrated in the examples of Figures 2 and 3 where the first two notes F# and C# share respectively their third and second harmonics.

4.3. Auditory filter analysis

Because of its flexibility, the MRT can also be very useful to perform an analysis on a psychoacoustic frequency scale like the ERB-scale [11] described above. Let us suppose that the set of frequencies $\{f_k\}_{k\in[0;K-1]}$ is chosen to map a uniform frequency distribution over the ERB scale and that the frequency resolutions $\{\delta f_k\}_{k\in[0;K-1]}$ are chosen such that $\delta f_k = ERB_f(f_k)$, thus providing the same resolution that the auditory filters. In this configuration, the MRT can mimic the time-frequency analysis performed by the auditory system.

An example of such an implementation is illustrated in Figure 4 on a speech signal.

5. CONCLUSION

In this paper, classic CQT impementations have been discussed and commented. According to this comments, some modifications of the CQT implementation have been proposed. In particular, the relation between the Q-factor

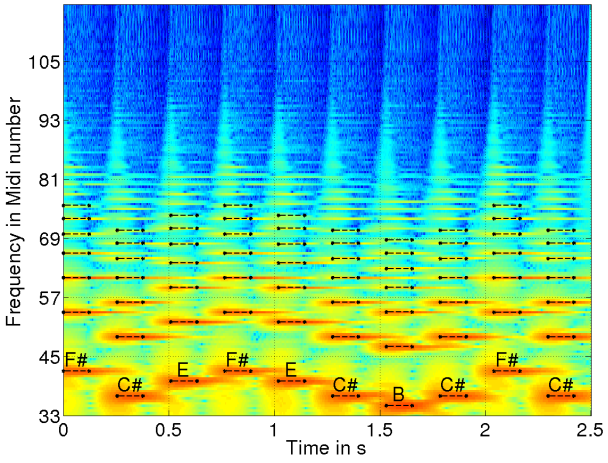

Fig. 3. MRT analysis of a musical excerpt (Bass pattern) with a mixed frequency resolution. The audio has been generated from a midi file. The note names, onset times and durations are indicated with the corresponding harmonics.

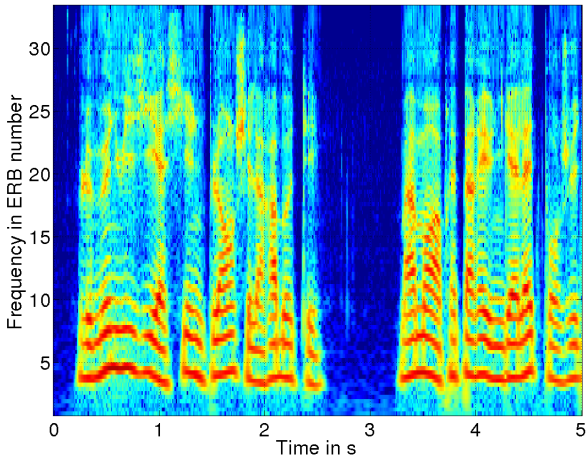

Fig. 4. MRT analysis of a speech signal with an ERB resolution. $f_{min} = 25Hz$, $f_{max} = 8kHz$, K=100 channels

and the analysis window have been clarified. Based upon the formulation of this CQT, a theoretical framework for the generalization of both the STFT and the Constant-Q Transform has been proposed. Its purpose is to provide a flexible time-frequency transform that performs an analysis with arbitrary center frequencies and frequency resolutions. To help the design of the analysis windows, explicit formulations of the relation between frequency resolution and analysis windows have been provided.

From the tensor formalism, practical algorithms can also be easily derived to provide efficient and fast ways to compute the corresponding signal transform and will be published in a future extension of this work.

Further work on this Multi-Resolution Transform framework can also study more in depth different inversion methods by taking advantage of the algebraic formulation introduced here.

6. REFERENCES

[1] Judith C Brown, "Calculation of a constant Q spectral transform," *Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991.

[2] Karin Dressler, "Sinusoidal extraction using an efficient implementation of a multi-resolution FFT," in *Proceeding of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada, September 18-20*, 2006, pp. 247–252.

[3] Judith C Brown, "An efficient algorithm for the calculation of a constant Q transform," *Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2698, 1992.

[4] C Schörkhuber and A Klapuri, "Constant-Q Transform Toolbox For Music Processing," in *7th Sound and Music Conf*, 2010.

[5] A.V. Oppenheim and R.W. Schafer, *Discrete-time signal processing*, Prentice-Hall signal processing series. Prentice Hall, 2010.

[6] Fredric J Harris, "On the use of windows for harmonic analysis," *Proceedings of the IEEE*, vol. 6, no. 1, pp. 51–83, 1978.

[7] B Yang, "A study of inverse short-time fourier transform," in *Proceedings of Acoustics Speech and Signal Processing 2008 ICASSP 2008 IEEE International Conference on*. 2008, p. 35413544, IEEE.

[8] D Fitzgerald, M Cranitch, and M T Cychowski, "Towards an inverse Constant Q Transform," *Proceedings of the 120th AES Convention*, vol. 1, no. 2, pp. 1–5, 2006.

[9] Adi Ben-Israel and Thomas N E Greville, *Generalized inverses: Theory and applications*, vol. 18-2, Springer-Verlag, 2003.

[10] Benoit Fuentes, Roland Badeau, and Gaël Richard, "Adaptive Harmonic Time-Frequency Decomposition of Audio Using Shift-Invariance PLCA," in *Proceedings of the 36th International Conference on Acoustics, Speech, and Signal Processing ICASSP'11*, Prague, Czech Republic, May 2011, pp. 401–404, IEEE.

[11] Brian R Glasberg and Brian C.J Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 12, pp. 103 – 138, 1990.