OPTIMAL SIGNAL RECONSTRUCTION FROM A CONSTANT-Q SPECTRUM

Anil Nagathil and Rainer Martin

Institute of Communication Acoustics, Ruhr-Universität Bochum, 44780 Bochum, Germany
email: {anil.nagathil,rainer.martin}@rub.de

ABSTRACT

In contrast to other well-known techniques for spectral analysis such as the discrete Fourier transform or the wavelet transform the constant-Q transform (CQT) matches the center frequencies of its sub-band filters to the frequency scale of western music and accounts for the requirement of frequency-dependent bandwidths. However, it does not possess a strict mathematical inverse. Therefore, we derive an optimal reconstruction method to recover a signal from its CQT spectrum. The reconstruction problem is posed as a segmented overdetermined minimization problem where the number of frequency bins is larger than the number of reconstructed signal samples in each segment. The framework also enables the reduction of the input-output latency of the CQT. The method is evaluated for different types of music signals as well as for speech and Gaussian noise. For classical music a reconstruction quality of 89 dB signal-to-noise ratio is achieved.

***Index Terms*—** Music, spectral analysis, signal reconstruction

1. INTRODUCTION

In music signal analysis many applications such as fundamental frequency estimation or source separation require the signal to be transformed into the spectral domain. Commonly used methods for spectral analysis such as the discrete Fourier Transform (DFT) [1] have fast implementations and have a mathematical inverse. However, since the frequencies of musical notes are distributed geometrically, a DFT with its uniformly arranged frequency scale either smears tones at low frequencies or computes too many components at high frequencies where the spacing between musical notes increases. Therefore, a spectral transform is required which offers a high frequency resolution at low frequencies and a low frequency resolution at high frequencies. While the wavelet transform [2] meets this condition, its tree-structured implementation only allows to represent the spectrum of music signals correctly up to an octave level. Recently, a wavelet transform with an adjustable resolution in the sub-octave level was proposed [3]. However, the center frequencies of the corresponding sub-band filters cannot be adapted to the frequency scale of musical notes.

The constant-Q transform (CQT) [4] is able to match the center frequencies of its sub-band filters to the frequency scale of musical notes and takes into account the requirement of frequency-dependent filter bandwidths. However, since it has a non-square transformation matrix, a strict mathematical inverse does not exist. While there exist different proposals towards a more efficient implementation of the CQT [5, 6, 7], approaches to reconstruct a signal from its CQT have been rare. If at all, only heuristic methods were proposed [8, 7]. Although in [7] a reasonable reconstruction quality of up to 60 dB signal-to-noise ratio for Gaussian noise was achieved, it is not clear if this approach optimizes any error criterion. In this paper, we therefore derive an optimal reconstruction method which estimates signal subsets from a constant-Q spectrum and recovers the original signal via an overlap-add procedure.

The paper is organized as follows. In Section 2 the mathematical formulation of the CQT and its relationship to the DFT is outlined. The proposed reconstruction method is derived in Section 3 which is evaluated in Section 4 with different types of audio signals considering symmetric and asymmetric analysis windows. Conclusions are drawn in Section 5.

2. CONSTANT-Q TRANSFORM

A conventional method to obtain the short-time spectrum of a signal $x(n)$ which is sampled at the sampling frequency f_s is to compute the short-time Fourier transform (STFT)

$$X_{\mathrm{dft}}(k,\lambda) = \sum_{n=0}^{N-1} w(n)\, x(n,\lambda)\, \exp\left(-j\frac{2\pi f_k n}{f_s}\right) \tag{1}$$

of a weighted signal segment $x(n,\lambda) = x(n+\lambda R)$ of size N with frame shift R and segment index λ. The parameter f_k denotes the center frequency of the k-th frequency bin with $k \in \{0, 1, ..., N-1\}$ and the weighting function $w(n)$ is usually defined as a tapered window. Typically the STFT is computed for $f_k/f_s = k/N$ which converts (1) into a sliding window DFT. While the DFT can be implemented very efficiently, e.g. by means of the fast Fourier transform (FFT) algorithm [1], and also can be inverted, the center frequencies $f_k = k\Delta f$ of the DFT bins are distributed uniformly and its frequency resolution $\Delta f = f_s/N$ is constant for all frequencies. However, musical notes exhibit an exponential frequency distribution $f_k = f_0\, 2^{\frac{k}{12b}}$ where f_0 denotes the lower bound of frequencies to be considered and for $b = 1$ we obtain twelve frequency bins per octave corresponding to the western musical scale of twelve semitones per octave. For $b \in \{2, 3, ...\}$ an even higher resolution than semitone resolution can be achieved which is beneficial if music instruments are not perfectly tuned. Therefore, for an appropriate analysis of musical notes a spectral transform is required which offers a high resolution at low frequencies and a low resolution at high frequencies. This is achieved if the filter bandwidths Δf_k are linearly related to the filter center frequencies $\Delta f_k = f_k/Q$ where Q is a quality factor defined as $Q = f_k/(f_{k+1} - f_k) = 1/(2^{\frac{1}{12b}} - 1)$. In order to achieve a frequency-dependent resolution, the window length therefore has to be chosen in accordance to the analyzed frequency $N_k = f_s/\Delta f_k$ yielding altogether $f_k = Q f_s/N_k$ which can be substituted into (1). Replacing $w(n)$ by frequency-dependent windows $w_k(n)$ and normalizing them to their length N_k we obtain the CQT

$$X_{\mathrm{cqt}}(k,\lambda) = \frac{1}{N_k} \sum_{n=0}^{N_k-1} w_k(n)\, x(n,\lambda)\, \exp\left(-j\frac{2\pi Q n}{N_k}\right) \tag{2}$$

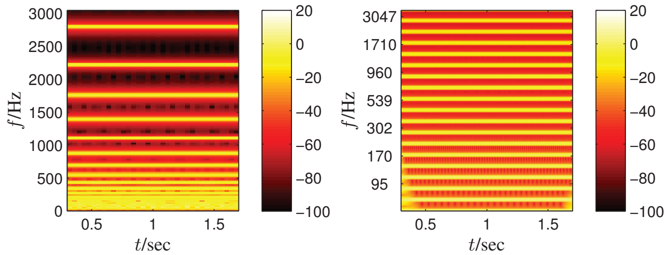

Fig. 1. STFT (left) vs. CQT (right) of multi-tone mixture with spacing of four semitones sampled at $f_s = 16$ kHz and $b = 2$.

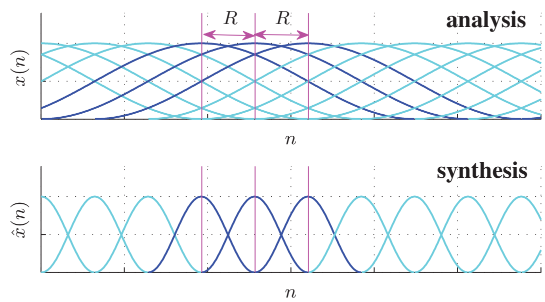

Fig. 2. Illustration of the analysis-synthesis procedure. A long input frame weighted by a symmetric analysis window (top) is analyzed by means of a sliding window CQT with frame shift R. In the synthesis process only a subset of the original input frame is reconstructed which is weighted by a shorter analysis window (bottom). An overlap-add procedure yields the reconstructed signal $\hat{x}(n)$.

as proposed in [4] with $k \in \mathcal{K} = \{0, 1, ..., K-1\}$ so that $f_{K-1} < f_s/2$ is ensured. By introducing the vector-matrix notation

$$\mathbf{X}_{\text{cqt}}^{(\lambda)} = [X_{\text{cqt}}(0,\lambda), X_{\text{cqt}}(1,\lambda), ..., X_{\text{cqt}}(K-1,\lambda)]^T, \qquad (3)$$

$$\mathbf{x}^{(\lambda)} = [x(0,\lambda), x(1,\lambda), ..., x(N_0-1,\lambda)]^T \qquad (4)$$

where $[\cdot]^T$ is the transpose operator and $\mathbf{C} \in \mathbb{C}^{K \times N_0}$ defined as

$$C_{kn} = \begin{cases} \frac{1}{N_k}\, w_k(n)\, \exp\left(-j\frac{2\pi Q n}{N_k}\right) & \forall (k,n) \in \mathcal{K} \times \mathcal{N}_k \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

with $\mathcal{N}_k = \left\{\frac{N_0 - N_k}{2}, \frac{N_0-N_k}{2}+1, ..., \frac{N_0+N_k}{2}-1\right\}$, (2) can be reformulated as

$$\mathbf{X}_{\text{cqt}}^{(\lambda)} = \mathbf{C}\,\mathbf{x}^{(\lambda)}. \qquad (6)$$

The benefit of the CQT is visualized in Figure 1. Here, a mixture of multiple equally contributing musical notes which have a spacing of four semitones are analyzed either by means of a STFT (left) or a CQT (right). While the STFT smears the spectral content at low frequencies and inefficiently represents the spectrum at high frequencies, the CQT clearly resolves all tones and facilitates their discrimination even at low frequencies.

3. OPTIMAL SIGNAL RECONSTRUCTION

While the DFT possesses a perfect inverse, the CQT is not invertible since the non-square transformation matrix $\mathbf{C}$ maps a signal segment of length N_0 onto a vector of spectral components of length K with $K < N_0$. Hence, the reconstruction of N_0 signal samples from their K CQT coefficients constitutes an under-determined problem which has an infinite number of solutions. However, restricting ourselves only to reconstruct a subset of L samples $\{y(l,\lambda) | l \in \mathcal{L}\} \subset \{x(n,\lambda)|n \in \widetilde{\mathcal{N}}\}$, with $\mathcal{L} = \{0, 1, ..., L-1\}$ and $\widetilde{\mathcal{N}} = \{\hat{n} - L/2 + 1, \hat{n} - L/2 + 2, ..., \hat{n} + L/2\}$, transforms the under-determined problem into an overdetermined problem if $L < K$. Here, $\hat{n}$ denotes the discrete time index for which all analysis window functions $w_k(n)$ attain their maximum, e.g. for symmetric windows $\hat{n} = N_0/2$ (Figure 3). This ensures that the signal subset is reconstructed which contributes most in the analysis stage of the CQT. As the analysis window $w_{K-1}(n)$ corresponding to the maximal analysis frequency f_{K-1} is the shortest of all analysis windows, it determines the minimal amount of data which is analyzed for all frequencies. In order to reconstruct the original data, we therefore have to choose the synthesis window length L so that $L \leq N_{K-1}$.

Setting the frame shift to $R \leq L/2$ and choosing a tapered synthesis window $s(l)$ with $l \in \mathcal{L}$ which satisfies a constant overlap-add constraint [9], the signal segment $x(n,\lambda)$ can be recovered approximately by overlap-adding the weighted segment subsets. The analysis-synthesis procedure is illustrated in Figure 2.

Note, that (6) computes a one-sided spectrum which in the DFT domain (1) implicitly corresponds to an analytic signal. Since the definition of the CQT (2) is related to the mathematical formulation of the DFT we can prevent $y(l,\lambda)$ to become complex-valued by extending the transformation matrix $\mathbf{C}$ to a matrix $\breve{\mathbf{C}} \in \mathbb{C}^{2K \times N_0}$ so that $\breve{\mathbf{C}} = [\mathbf{C} \quad \overline{\mathbf{C}}_{\updownarrow}]^T$ where $\overline{\mathbf{C}}_{\updownarrow}$ denotes the complex conjugate matrix of $\mathbf{C}$ with reversed row order. This yields a conjugate symmetric CQT spectrum $\breve{\mathbf{X}}_{\text{cqt}}^{(\lambda)} = \breve{\mathbf{C}}\,\mathbf{x}^{(\lambda)}$. Using the notation $\mathbf{y}^{(\lambda)} = [y(0,\lambda), y(1,\lambda), ..., y(L-1,\lambda)]^T$ and requiring $\mathbf{y}^{(\lambda)} \overset{!}{=} \mathbf{A}\breve{\mathbf{X}}_{\text{cqt}}^{(\lambda)} = \mathbf{A}\breve{\mathbf{C}}\mathbf{x}^{(\lambda)}$, where $\mathbf{A} \in \mathbb{C}^{L \times 2K}$ is a reconstruction matrix to be estimated, we introduce the error vector

$$\mathbf{e} = \mathbf{y}^{(\lambda)} - \tilde{\mathbf{I}}\mathbf{x}^{(\lambda)} = \left(\mathbf{A}\breve{\mathbf{C}} - \tilde{\mathbf{I}}\right)\mathbf{x}^{(\lambda)}. \qquad (7)$$

The matrix $\tilde{\mathbf{I}} = [\mathbf{0}_{L,P} \quad \mathbf{I} \quad \mathbf{0}_{L,P}]$ extracts L samples in the range $n \in \widetilde{\mathcal{N}}$ from each input data vector $\mathbf{x}^{(\lambda)}$ which is to be approximated by $\mathbf{y}^{(\lambda)}$ where $\mathbf{I}$ is an $L \times L$ identity matrix and $\mathbf{0}_{L,P}$ is an $L \times P$ zero matrix with $P = (N_0 - L)/2$.

We minimize the Frobenius norm of the error matrix $E_{\text{F}} = \|\mathbf{A}\breve{\mathbf{C}} - \tilde{\mathbf{I}}\|_{\text{F}}$ which involves the complex differentiation of a real-valued function with respect to $\mathbf{A}$, defined in e.g. [10, (B.4)] as

$$\frac{\partial}{\partial \mathbf{A}} = \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{A}_{\text{R}}} - j\frac{\partial}{\partial \mathbf{A}_{\text{I}}}\right) \qquad (8)$$

where $\mathbf{A}_{\text{R}} = \Re\{\mathbf{A}\}$ and $\mathbf{A}_{\text{I}} = \Im\{\mathbf{A}\}$. Exploiting the identities $\|\mathbf{Z}\|_{\text{F}} = \text{Tr}\left(\mathbf{Z}^H\mathbf{Z}\right)$ and $\text{Tr}\left(\mathbf{Z} + \mathbf{Z}^H\right) = 2\Re\left\{\text{Tr}\left(\mathbf{Z}\right)\right\}$ with $\text{Tr}(\cdot)$ being the trace operator, $[\cdot]^H$ being the Hermitian conjugate operator and $\mathbf{Z}$ being an arbitrary matrix we obtain

$$\frac{\partial E_{\text{F}}}{\partial \mathbf{A}} = \frac{1}{2}\frac{\partial}{\partial \mathbf{A}}\left[\text{Tr}\left(\breve{\mathbf{C}}^H\mathbf{A}^H\mathbf{A}\breve{\mathbf{C}}\right) + 2\Re\left\{\text{Tr}\left(\breve{\mathbf{C}}^H\mathbf{A}^H\tilde{\mathbf{I}}\right)\right\}\right] \overset{!}{=} 0 \qquad (9)$$

which by means of (8) yields

$$\mathbf{A} = \tilde{\mathbf{I}}\breve{\mathbf{C}}^H\left(\breve{\mathbf{C}}\breve{\mathbf{C}}^H\right)^{-1}. \qquad (10)$$

Introducing the vector notation $\mathbf{s} = [s(0), s(1), ..., s(L-1)]^T$ and denoting the synthesis matrix as $\mathbf{S} = \text{diag}(\mathbf{s})$ the reconstructed weighted signal segment subset becomes $\mathbf{y}_s^{(\lambda)} = \mathbf{S}\mathbf{y}^{(\lambda)} = \mathbf{S}\mathbf{A}\breve{\mathbf{X}}_{\text{cqt}}^{(\lambda)}$.

4. EVALUATION

In this section we will evaluate the performance of the reconstruction method presented in the previous section by computing the signal-to-noise ratio (SNR)

$$\frac{\text{SNR}}{\text{dB}} = 10 \log \left(\frac{\sum_n x^2(n)}{\sum_n \left[x(n) - \hat{x}(n)\right]^2} \right) \tag{11}$$

where $x(n)$ is the input signal and $\hat{x}(n)$ is the reconstructed signal obtained by overlap-adding $\mathbf{y}_s^{(\lambda)}$. As the CQT requires a minimal and maximal analysis frequency f_0 and f_{K-1} to be chosen, a direct comparison of the input signal and the reconstructed signal would not be fair because all spectral information outside of this frequency range is lost in the analysis process. Therefore, as in the evaluation stage of [7] the input signal is bandpass-filtered by means of a sixth order Butterworth filter before being fed into the CQT.

For the evaluation we choose the following experimental setup. The signals to be analyzed and reconstructed are sampled at $f_s =$ 16 kHz and have a duration of five seconds. The minimal and maximal analysis frequencies are chosen as $f_0 = 220$ Hz and $f_{K-1} =$ 7040 Hz, respectively, which spans a range of five octaves. We consider the cases $b \in \{1, 2, 4\}$ which correspond to 12, 24 and 48 frequency bins per octave, respectively, and the frame shifts $R \in \{16, 32\}$. We use Hann windows as analysis windows $w_k(n)$ to compute the transformation kernel (5). Specifications on the shape of the analysis windows will be made in the following subsections. The synthesis window $s(l)$ is chosen to be a Hann window as well, since it satisfies a constant overlap-add constraint if $R \leq L/2$ and the ratio L/R is a power of two. Further, as explained in Section 3 the constraint $L \leq N_{K-1}$ has to be satisfied. Therefore, since for $b \in \{1, 2, 4\}$ we obtain $N_{K-1} = \{38, 78, 156\}$, the synthesis window lengths will be $L = \{32, 64, 128\}$. We randomly selected ten classical music signals, ten popular music signals and four speech signals to evaluate the reconstruction performance.

4.1. Symmetric analysis windows

Using the experimental setup specified above we investigate the performance of the proposed synthesis system using conventional (symmetric) Hann windows $w_k(n)$ in the analysis stage. These are depicted in Figure 3 for $b = 2$ and a few selected center frequencies. In the synthesis stage the reconstructed signal segment subset $\mathbf{y}^{(\lambda)}$ approximately recovers L values of the original signal segment around $\hat{n} = N_0/2 \approx 1241$.

Table 1 summarizes the experimental results for a frame shift of $R = 16$. For all three audio types we can clearly observe an increase in SNR when the number of frequency bins per octave is increased. Ideally, twelve bins per octave should be enough to represent the spectral content of a music signal as each octave comprises twelve semitones in the western musical scale. However, perfectly tuning music instruments is not always feasible in practice. Therefore, allowing more than one frequency bin to resolve one semitone enhances the reconstruction quality at the cost of a higher computational complexity. We observe that classical music performs best showing an SNR between 52 and 89 dB on average, whereas for popular music we obtain SNR values between 47 and 75 dB. While according to (7) the reconstruction error is signal-dependent anyway, it further makes sense that popular music is not reconstructed as well as classical music as it more likely contains percussive and atonal sounds which do not match the frequency grid of the CQT. The same applies for speech sounds which show a similar reconstruction quality as popular music. The worst performance ranging from 38 to 65 dB SNR is achieved for Gaussian noise based on which the reconstruction system in [7] was evaluated. In [7] a maximal SNR of about 60 dB was achieved.

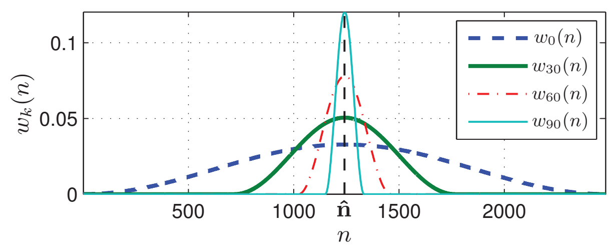

Fig. 3. Symmetric Hann windows used in the analysis stage of the CQT for $f_s = 16$ kHz, $f_0 = 220$ Hz, $f_{K-1} = 7040$ Hz, and $b = 2$.

Audio type	12 bins/oct.	24 bins/oct.	48 bins/oct.
Classical music	52.26 ± 2.08	69.56 ± 3.95	88.94 ± 4.38
Popular music	46.81 ± 3.15	60.06 ± 3.87	75.14 ± 4.68
Speech	47.52 ± 4.33	61.01 ± 5.58	76.73 ± 5.94
Gaussian noise	38.22	50.11	64.73

Table 1. Mean SNR values and standard deviations in dB of different reconstructed audio signals for $f_s = 16$ kHz, $f_0 = 220$ Hz, $f_{K-1} = 7040$ Hz, $R = 16$ and different numbers of frequency bins per octave.

Audio type	12 bins/oct.	24 bins/oct.	48 bins/oct.
Classical music	3.01 ± 0.01	67.06 ± 3.15	87.61 ± 4.35
Popular music	3.03 ± 0.02	57.08 ± 4.12	70.83 ± 4.65
Speech	3.05 ± 0.06	57.49 ± 5.14	70.73 ± 5.61
Gaussian noise	3.00	46.92	60.05

Table 2. Mean SNR values and standard deviations in dB of different reconstructed audio signals for $f_s = 16$ kHz, $f_0 = 220$ Hz, $f_{K-1} = 7040$ Hz, $R = 32$ and different numbers of frequency bins per octave.

Table 2 shows the corresponding results for a frame shift of $R = 32$ which halves the computational complexity. If 24 or 48 bins per octave are used in the analysis stage the performance slightly degrades compared to the case in Table 1. However, for 12 bins per octave the reconstruction quality strikingly degrades since the condition $R \leq L/2$ is violated.

4.2. Asymmetric analysis windows

While for symmetric analysis windows the reconstruction quality is very good, the analysis-synthesis system suffers from a high system delay which is unfavorable if a low input-output latency is required. As the length L of the synthesis window is relatively short and hence does not contribute much to the system delay, the main source of the high delay can be found in the design of the analysis windows which are centered around $\hat{n} = N_0/2$ if they are symmetric. That means, that the signal segment is mainly analyzed in its center as can be seen in Figure 3 which including the delay evoked by the synthesis

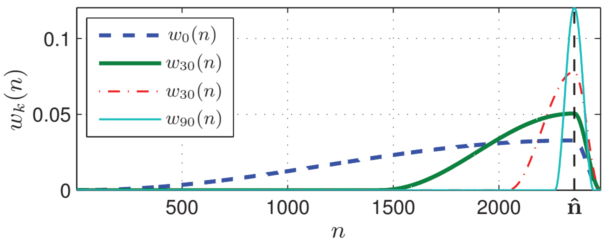

Fig. 4. Asymmetric analysis windows composed from a long Hann window and a short Hann window for $f_s = 16$ kHz, $f_0 = 220$ Hz, $f_{K-1} = 7040$ Hz, $b = 2$ and $\hat{n} = 0.95N_0$. The system delay is reduced to 10.8 ms.

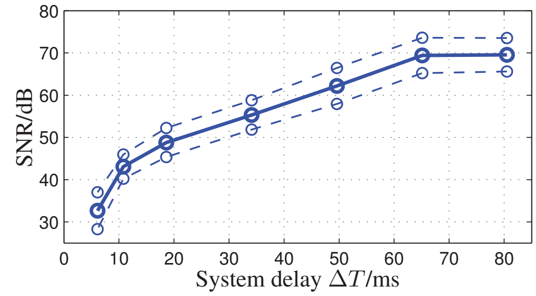

Fig. 5. Mean SNR (solid) and standard deviation (dashed) vs. system delay of ten reconstructed classical music pieces for f_s = 16 kHz, $f_0 = 220$ Hz, $f_{K-1} = 7040$ Hz, $R = 16$ and $b = 2$.

window and the frame shift amounts to the total system delay

$$\Delta T(\hat{n}) = \frac{N_0 - \hat{n} + L/2 + R}{f_s}. \qquad (12)$$

While N_0, L, R cannot be changed for a given experimental setup, the system delay can be reduced by increasing $\hat{n}$. For $\hat{n} = N_0 - L/2$ the system delay becomes $\Delta T(\hat{n}) = R/f_s$ and hence minimal. This requires to change the shape of the analysis windows from symmetric to asymmetric to ensure that the main contribution of the weighted signal segments in the analysis stage originates from the area around $\hat{n}$. While there are different approaches on the analytically optimal design of symmetric windows [11, 12, 13], asymmetric windows are - to the best knowledge of the authors - only designed in a heuristic [14] or numerically optimized [15] way so far.

An example of such asymmetric analysis windows is shown in Figure 4 for $b = 2$ and a few selected frequencies. These windows are composed from the left half of a long Hann window and the right half of a short Hann window.

The reconstruction quality of the proposed method with different sets of asymmetric analysis windows is shown in terms of the average SNR values and standard deviations over ten classical music signals as a function of their system delay, for $b = 2$ and $R = 16$. While the best signal quality is achieved for the highest system delay, which corresponds to the case of symmetric analysis windows, the performance degrades for decreasing system delays to about 33 dB SNR for $\Delta T(\hat{n}) = 6.1$ ms.

5. CONCLUSIONS

In this paper we propose a method for the optimal signal reconstruction from a constant-Q spectrum by restricting ourselves only to reconstruct a subset of an original signal segment instead of recovering the whole segment which would constitute an under-determined problem. The original signal can be approximately resynthesized by overlap-adding the reconstructed signal subsets. The robustness of the method is evaluated by analyzing and reconstructing classical music, popular music, speech signals and Gaussian noise for symmetric and asymmetric analysis windows. For symmetric windows we achieve an SNR of up to 89 dB, whereas spectrally unoptimized asymmetric windows, which decrease the system delay imposed by the analysis-synthesis system, achieve satisfactory performance of more than 50 dB SNR for a range of reduced delays $\Delta T > 30$ ms and lead to significant degradations of the reconstructed signal for very low delays.

6. ACKNOWLEDGMENTS

This work is funded by the German Research Foundation (DFG), Sonderforschungsbereich 823, Teilprojekt B3. The authors would like to thank Jan-Willem Schlattmann for implementing the reconstruction method and Sarmad Malik for helpful discussions.

7. REFERENCES

[1] P. Vary and R. Martin, *Digital Speech Transmission*, John Wiley & Sons Ltd., 2006.

[2] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1st edition, 1993.

[3] I.W. Selesnick, "Wavelet Transform With Tunable Q-Factor," *IEEE Trans. on Signal Processing*, vol. 59, no. 8, pp. 3560–3575, August 2011.

[4] J.C. Brown, "Calculation of a constant Q spectral transform," *Journal Acoust. Soc. of America*, vol. 89, no. 1, pp. 425–434, January 1991.

[5] J.C. Brown and M.S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *Journal Acoust. Soc. of America*, vol. 92, no. 5, pp. 2698–2701, November 1992.

[6] F.C.C.B. Diniz, I. Kothe, S.L. Netto, and L.W.P. Biscainho, "High-Selectivity Filter Banks for Spectral Analysis of Music Signals," *EURASIP Journal on Advances in Signal Processing*, 2007.

[7] C. Schörkhuber and A. Klapuri, "Constant-Q Transform Toolbox for Music Processing," in *Proc. Sound and Music Computing Conf.*, 2010.

[8] D. FitzGerald, M. Cranitch, and M.T. Cychowski, "Towards an Inverse Constant Q Transform," in *Proc. 120th Convention of the Audio Engineering Society*, 2006.

[9] M.M. Goodwin, "The STFT, Sinusoidal Models, and Speech Modification," in *Springer Handbook of Speech Processing*, J. Benesty, M.M. Sondhi, and Y. Huang, Eds. Springer, 2008.

[10] S.S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 4th edition, 2002.

[11] F.J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. of the IEEE*, vol. 66, no. 1, pp. 51–84, January 1978.

[12] J.W. Adams, "A New Optimal Window," *IEEE Trans. on Signal Processing*, vol. 39, no. 8, pp. 1753–1769, August 1991.

[13] B. Bell, D.B. Percival, and A.T. Walden, "Calculating Thomson's Spectral Multitapers by Inverse Iteration," *Journal of Comp. and Graph. Stat.*, vol. 2, no. 1, pp. 119–130, 1993.

[14] D.A.F. Florêncio, "On the Use of Asymmetric Windows for Reducing the Time Delay in Real-time Spectral Analysis," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 1991, vol. 5, pp. 3261–3264.

[15] D. Mauler and R. Martin, "Optimization of Switchable Windows for Low-delay Spectral Analysis-Synthesis," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 4718–4721.