THE ERBLET TRANSFORM: AN AUDITORY-BASED TIME-FREQUENCY REPRESENTATION WITH PERFECT RECONSTRUCTION

T. Necciari, Member, IEEE, P. Balazs, Senior Member, IEEE, N. Holighaus, and P. L. Søndergaard

Acoustics Research Institute
Austrian Academy of Sciences
Wohllebengasse 12–14, A-1040 Vienna, Austria

ABSTRACT

This paper describes a method for obtaining a perceptually motivated and perfectly invertible time-frequency representation of a sound signal. Based on frame theory and the recent non-stationary Gabor transform, a linear representation with resolution evolving across frequency is formulated and implemented as a non-uniform filterbank. To match the human auditory time-frequency resolution, the transform uses Gaussian windows equidistantly spaced on the psychoacoustic "ERB" frequency scale. Additionally, the transform features adaptable resolution and redundancy. Simulations showed that perfect reconstruction can be achieved using fast iterative methods and preconditioning even using one filter per ERB and a very low redundancy (1.08). Comparison with a linear gammatone filterbank showed that the ERBlet approximates well the auditory time-frequency resolution.

Index Terms— time-frequency representation, auditory filterbank, perfect reconstruction, non-stationary Gabor transform, preconditioning, ERB scale

1. INTRODUCTION

Sound signals such as speech or music are non-stationary by nature. Accordingly, audio processing techniques like sound design, audio coding, auditory scene analysis or speech recognition call for specific tools to analyze, process, and resynthesize sounds. In this context, linear time-frequency (TF) representations have become standard tools. They allow decomposing any signal into a set of elementary functions with good TF localization and achieving perfect reconstruction if the transform parameters are chosen appropriately (*e.g.*, [1]). Because speech and music signals are usually targeted at human listeners, using knowledge about the human auditory system in audio signal processing is natural. This is done for instance in perceptual audio coding where psychophysical data on auditory masking are used to reduce the size of digital audio files [2] or, similarly, in a recent perceptual sparsity approach [3]. Nevertheless, many audio applications (including audio coding) still rely on Fourier-based transforms such as the Gabor or discrete cosine transforms that have a fixed resolution over the whole TF plane. Consequently, their TF resolutions are not compatible with that of the auditory system (see Sec. 2). Several existing methods allow for obtaining a perceptually motivated TF representation of audio signals (see Sec. 3) but many properties desirable in analysis-synthesis systems, namely invertibility, computational efficiency, and adaptable redundancy are lost in the process. Our work introduces a new TF transform constructed to fulfill these requirements while providing perceptually motivated TF resolution. We rely on frame theory and the recent non-stationary Gabor transform (NSGT) to formulate a perfectly invertible representation with resolution evolving across frequency [4], resulting in a non-uniform filterbank. This approach has been used in [5] to implement an invertible transform with constant relative bandwidth ("constant-Q") frequency analysis. Here, we propose a transform matched to the psychoacoustic "ERB" frequency scale, thereby referring to the result as the *ERBlet transform*. To provide more control over the resolution and redundancy of the signal representation, we allow for non-compactly supported windows and large frequency-dependent down-sampling factors, potentially violating the "painless" conditions for perfect reconstruction given in [4]. We show that in this particular case, reconstruction is still possible using fast iterative methods and preconditioning provided the system constitutes a frame. Practical examples and simulations in Sec. 5 show that if some redundancy is retained, the proposed analysis-synthesis system is well-conditioned and converges fast to the correct solution.

2. THE AUDITORY TF RESOLUTION

The peripheral auditory system can be modeled in a first approximation as a bank of bandpass filters whose bandwidth corresponds to the spectral resolution of the ear. Many psychoacoustical studies have focused on the characterization of these "auditory filters"(see [6] for a review). The filters are commonly described by their equivalent rectangular bandwidth (ERB). The ERB (in Hz) of the auditory filter centered

This work was supported by the Austrian Science Fund (FWF) START-project FLAME ("Frames and Linear Operators for Acoustical Modeling and Parameter Estimation"; Y 551-N13).

at frequency F (in Hz) is [7]

$$ERB(F) = 24.7 + \frac{F}{9.265}. \quad (1)$$

Eq. (1) indicates that the auditory frequency resolution as described by the ERB is approximately constant-Q only at high frequencies (> 2 kHz). For the full range of audible frequencies (.02–20 kHz) the ERBs range from 27 Hz to 2.2 kHz. Using ERB units, the range of audible frequencies can be discretized as a bank of 39 adjacent filters whose ERB number is [7]

$$ERB_{\text{num}}(F) = 9.265 \ln\left(1 + \frac{F}{228.8455}\right) \quad (2)$$

and, reciprocally, $F = u(E_F) = 228.8455\ \left(e^{(E_F/9.265)} - 1\right)$. Eq. (2) corresponds to the ERB scale used to plot psychoacoustical data on a perceptual frequency axis. The partition of the frequency axis into filters leads to a partition of the time axis into time windows whose widths correspond to the temporal resolution at a certain frequency. In [8] the windows' shape was estimated using Gaussian stimuli with various spectro-temporal shapes. The results indicated that the spectral width of one window corresponds to one ERB and the temporal width approximately corresponds to four periods of the carrier frequency, *e.g.*, 4 ms at 1 kHz. Overall, the data in [8] suggests that the auditory systems performs a TF analysis using its own "internal" windows that are well approximated by Gaussians with frequency dependent spectro-temporal resolution.

3. AUDITORY-BASED TF REPRESENTATIONS

To date, two general approaches exist to achieve a perceptually motivated TF representation of an audio signal. The first approach includes models of auditory processing like in [9, 10]. Such models attempt to replicate the various stages of auditory processing and are useful to improve our knowledge about the auditory system. However, they feature many parameters, they are computationally demanding and *not invertible*. Approximately invertible models were proposed in, *e.g.*, [11, 12] as integrated audio coders. Consequently, the signal representation is not easily accessible. Overall, auditory models do not constitute TF analysis-synthesis tools. The second approach includes TF transforms tuned to mimic the auditory TF resolution (see Sec. 2). Wavelet and constant-Q transforms are used [5, 13, 14] in this context, but they mismatch the auditory spectral resolution at low frequencies. Further developments include a bilinear transform [15], linear [16, 17] and nonlinear gammatone filterbanks [18], and auditory-based non-uniform filterbanks [19, 20]. They approximate the auditory TF resolution nicely but fail at providing prefect reconstruction.

4. PROPOSED APPROACH

In the following, we consider real-valued signals of length L. The inner product of two signals f, g is $\langle f, g\rangle = \sum_{l=1}^{L} f[l] \cdot g[l]$ and the energy of a signal is defined from the inner product as $\|f\|^2 = \langle f, f\rangle$. We denote the Fourier transformation by $\mathcal{F} :\ f \mapsto \hat{f}$.

4.1. Basic concept: The non-stationary Gabor transform

An NSG system [4] with resolution evolving across frequency can be formulated as a non-uniform filterbank as

$$\mathcal{G}(\mathbf{g}, \mathbf{D}) := (g_{n,k}[l]) = (g_k\,[l - nD_k]) \quad (3)$$

where indexes $n, k \in \mathbb{Z}$ are related to time and frequency, respectively. This system features frequency-dependent filters g_k and down-sampling factors D_k. The NSGT relies on frame theory. The sequence $(g_{n,k})$ is called a *frame* if there exist positive constants A and B (called lower and upper frame bounds, respectively) that satisfy

$$A\|f\|^2 \leq \sum_{n,k} |\,\langle f, g_{n,k}\rangle\,|^2 \leq B\|f\|^2 \quad (4)$$

for any signal f. The coefficients $c_{n,k} = \langle f, g_{n,k}\rangle$ yield the analysis of the signal f and the synthesis is given by $\sum_{n,k} c_{n,k}\, g_{n,k}$. Frame theory gives a stable way to reconstruct the signal from the coefficients $c_{n,k}$ using the frame operator $\mathbf{S}$ given by $\mathbf{S}f = \sum_{n,k} \langle f, g_{n,k}\rangle\, g_{n,k}$. If $\mathbf{S}$ is invertible, then reconstruction is achieved using the *canonical dual frame* $\widetilde{\mathcal{G}}(\mathbf{g}, \mathbf{D}) = (\widetilde{g_{n,k}})$ defined by

$$\widetilde{g_{n,k}} = \mathbf{S}^{-1} g_{n,k} \quad (5)$$

and $f = \mathbf{S}^{-1}\mathbf{S}f = \sum_{n,k} \langle f, g_{n,k}\rangle\, \widetilde{g_{n,k}}$. Note that, in general, $\widetilde{\mathcal{G}}(\mathbf{g}, \mathbf{D})$ does not have the same structure as $\mathcal{G}(\mathbf{g}, \mathbf{D})$. As in standard Gabor theory, certain conditions must be fulfilled in NSGT to achieve perfect reconstruction (*i.e.*, invertibility of $\mathbf{S}$). Suppose the frequency response of $\hat{g}_k$ has a bandpass characteristic and $\operatorname{supp}(\hat{g}_k) = \mathcal{I}_k$ samples in the positive frequency domain. If the time sampling in each channel satisfies $\lceil L/D_k \rceil \geq 2\,\mathcal{I}_k$, then the operator

$$\hat{\mathbf{S}} := \mathcal{F}\,\mathbf{S}\,\mathcal{F}^{-1} \quad (6)$$

is diagonal and easily invertible. This is called the *painless case* [4]. Because this setting restrains the resolution and redundancy of the transform, in this paper we use NSG systems featuring non-compactly supported windows and large down-sampling factors. We propose a method using fast iterative methods and preconditioning [21, 22, 23] so that perfect reconstruction can be numerically efficient in this setting. This is verified by an experiment in Sec. 5.

4.2. Analysis and dual windows: ERBlets

The ERBlet transform consists of filters $g_k, k = 0, \ldots, K$, that are Gaussian windows constructed in the positive frequency domain according to

$$\hat{g}_k[m] = \Gamma_k^{-\frac{1}{2}} e^{-\pi\left[\frac{m-\nu_k}{\Gamma_k}\right]^2} \quad (7)$$

where $m \in \mathbb{Z}$ is the discrete frequency variable, ν is the center frequency (in Hz), and Γ is a shape factor that controls the effective bandwidth of $\hat{g}$ (in Hz). Let F_{min} and F_{max} denote the minimum and maximum analysis frequencies, respectively. Their corresponding ERB numbers are (Eq. (2)) E_0 and E_K. Linearly distributing $K+1$ filters from E_0 to E_K with a density of V filters per ERB leads to $E_k = E_0 + k/V$ with $K = V\,(E_K - E_0)$. Then $\nu_k = u(E_k)$ and $\Gamma_k = ERB(\nu_k)$. The factor $\Gamma_k^{-\frac{1}{2}}$ in Eq. (7) ensures that all filters have the same energy. Although Gaussians are not compactly-supported windows, they decay very fast. Thus, by truncating the filters so that $\text{supp}(\hat{g}_k) = \lceil 4\,\Gamma_k \rceil$ samples, the filters are close to zero at the borders. Finally, D_k and V can be chosen such that the frame operator associated with the ERBlets is invertible (see Sec. 4.1 and [4]).

4.3. Implementation

The ERBlet algorithms are available at `http://www.kfs.oeaw.ac.at/ICASSP2013_ERBlets`. This address is referred below to as the "web page".

Algorithms for computing an NSGT with resolution evolving over time are provided in the Matlab/Octave "LTFAT" toolbox [24], where NSG analysis and synthesis are handled by the algorithms `nsdgt` and `insdgt`, respectively. By applying these algorithms to $\hat{f}$ we obtain an NSGT with resolution evolving across frequency. To process the positive and negative frequencies the $K+1$ filters are mirrored to the negative frequency domain (note that if F_{min} and F_{max} are set at the 0 and Nyquist frequencies, respectively, then only $K-1$ filters are mirrored). The ERBlet transform is determined by the two parameters D_k and V that provide control over the resolution and redundancy of the transform, $red = \sum_{k=-K}^{K} D_k^{-1}$. The number of time samples in each channel is given by $N_k = \lceil L/D_k \rceil$. Choosing D_k such that $N_k \geq \text{supp}(\hat{g}_k)$ results in a painless system (see Sec. 4.1). Otherwise $\hat{\mathbf{S}}$ is not diagonal and iterative algorithms are required for efficient inversion. To do so, we use the equality

$$\sum_{n,k} c_{n,k} \widetilde{g_{n,k}} = \sum_{n,k} c_{n,k} \mathbf{S}^{-1} g_{n,k} = \mathcal{F}^{-1} \hat{\mathbf{S}}^{-1} \sum_{n,k} c_{n,k} \hat{g}_{n,k}$$

to solve the linear system

$$\hat{\mathbf{S}} x = \sum_{n,k} c_{n,k}\, \hat{g}_{n,k} \tag{8}$$

with an adapted *conjugate gradients* (CG) algorithm [21, 22] where the right-hand sum in Eq. (8) is computed by `insdgt`. If $\mathcal{G}(\mathbf{g}, \mathbf{D})$ is a frame, then $\hat{\mathbf{S}}$ is self-adjoint and CG converges to the desired solution. The convergence speed depends on the condition number $\kappa(\hat{\mathbf{S}})$ (*i.e.*, the frame bound ratio B/A [22]) and can be improved with a preconditioning step [23]. If the $\hat{g}_k$s decay fast enough then $\hat{\mathbf{S}}$ is diagonal dominant and the matrix

$$\mathbf{D}(\hat{\mathbf{S}})^{-1}_{m,l} = \begin{cases} \left(\sum N_k |\hat{g}_k|^2\right)^{-1}[m], & \text{if } m = l \\ 0, & \text{else} \end{cases} \tag{9}$$

Table 1. Parameters, redundancies and frame bound ratios of the NSG ERBlet systems used in Exp. 1.

$\mathcal{G}(\mathbf{g}, \mathbf{D})$	case	V	K	N_k	red	B/A
ERBlet1	painless	1	43	$\lceil 4\,\Gamma_k \rceil$	4.00	1.44
ERBlet2	painless	3	129	$\lceil 4\,\Gamma_k \rceil$	12.00	1.07
ERBlet3	CG	1	43	$\lceil \frac{32\,\Gamma_k}{9} \rceil$	3.53	1.44
ERBlet4	CG	1	43	$\lceil \frac{8\,\Gamma_k}{3} \rceil$	2.64	1.44
ERBlet5	CG	1	43	$\lceil 2\,\Gamma_k \rceil$	1.98	1.52
ERBlet6	CG	1	43	$\lceil \frac{4\,\Gamma_k}{3} \rceil$	1.32	2.56
ERBlet7	CG	1	43	$\lceil \frac{12\,\Gamma_k}{11} \rceil$	1.08	5.88

is an efficient preconditioner. Because CG requires self-adjoint matrices and applying $\mathbf{D}(\hat{\mathbf{S}})^{-1}$ to $\hat{\mathbf{S}}$ does not result in a self-adjoint matrix, we use $\mathbf{D}(\hat{\mathbf{S}})^{-1/2} \hat{\mathbf{S}} \mathbf{D}(\hat{\mathbf{S}})^{-1/2}$ instead. Since applying $\hat{\mathbf{S}}$ to a signal f is equivalent to performing analysis followed by synthesis with $\mathcal{G}(\mathbf{g}, \mathbf{D})$, we can use `nsdgt` and `insdgt` to solve Eq. (8). The preconditioner in Eq. (9) is realized by point-wise multiplication. Thus, one CG step involves one application of `nsdgt` and `insdgt` and L scalar multiplications for the preconditioning. Pseudo-code for the algorithms can be found on the web page. As experimental results in Sec. 5 show, only a few iterations are necessary for CG to converge to the correct solution up to numerical precision.

5. RESULTS AND DISCUSSION

Two experiments were conducted to evaluate the performance of the ERBlet transform. In Exp. 1 we tested the convergence of iterative reconstruction for several NSG ERBlet systems yielding different redundancies (see Tab. 1). In Exp. 2 we compared the ERBlet to two other auditory-based approaches in terms of signal representation, reconstruction error, and redundancy. The audio material consisted of a 5-sec musical excerpt from the band Manowar (song "Heart of Steel", studio version) in mono format, sampled at 44.1 kHz, 16 bits/sample. All analyses were performed for $F_{\text{min}} = 0$ and $F_{\text{max}} = 22.05$ kHz. Complementary results, colored figures, and simulation codes are available on the web page.

The results of Exp. 1 are depicted in Fig. 1 as a convergence plot. It can be seen that preconditioning had a considerable effect for the systems ERBlet3 to ERBlet5. Iterative synthesis for ERBlet6 and ERBlet7, not shown in the plot, converged in 21 and 45 iterations, respectively. Preconditioning had no effect in these cases. Noteworthy, the number of iterations does not depend on the signal length but only on the condition number.

In Exp. 2 we considered the system "ERBlet2" in Tab. 1, the NSG constant-Q transform in [5], and the linear gammatone filterbank in [16]. The constant-Q transform used 24 filters per octave distributed between 50 Hz and 22.05 kHz (212 filters in total) and Q = 9 ($\approx F/ERB(F)$ for $F > 2$ kHz).

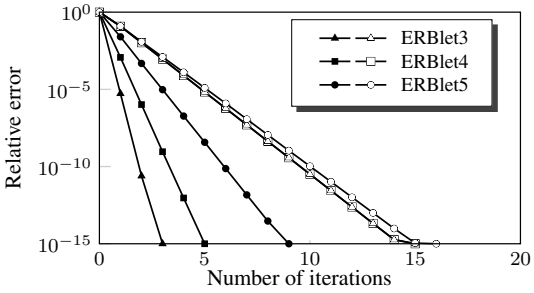

Fig. 1. Convergence of the iterative reconstruction with (filled markers) and without diagonal preconditioning (empty markers) for various ERBlet configurations indicated in Tab. 1.

These parameters for the constant-Q were chosen so that both the constant-Q and ERBlet transforms have approximately the same number of filters in the frequency range 2–20 kHz (84) and the same redundancy over the whole TF plane (12). The gammatone filterbank used 3 filters per ERB (128 filters in total). Signal representations are depicted in Fig. 2. Fig. 2a shows that the ERBlet captured both harmonic (voice vibrato) and transient (drums) parts in the broadband, rich background generated by drums and distorted guitars. Fig. 2b shows that ERBlet and constant-Q representations are very similar above 500 Hz but differ below. Below 500 Hz the ERBlet has a better time resolution while the constant-Q transform has a better spectral resolution. This is due to the fact that the constant-Q transform features a larger number of filters at low than at high frequencies. Consequently, the constant-Q transform required 212 filters to achieve the same (visual) high-frequency resolution as the ERBlet with 129 filters. Both the ERBlet and constant-Q transforms led to perfect reconstruction (relative errors $< 10^{-15}$). Fig. 2c shows that ERBlet and gammatone representations are very similar over the whole TF plane. Since gammatone filters are auditory filter models *per se*, this result indicates that the ERBlet approximates well the auditory TF resolution. While the ERBlet achieved perfect reconstruction, the gammatone filterbank led to a relative reconstruction error of about 10^{-3}. Note, however, that this error was perceptually irrelevant (as indicated by informal listening). Because the gammatone system in [16] features no down-sampling option, its redundancy was 128 compared to 12 for the ERBlet.

Overall, the proposed method provides a linear, auditory-based, and perfectly invertible TF transform that can be easily integrated in audio analysis-synthesis systems. An advantage of the current implementation is that resolution and redundancy are adaptable without affecting the reconstruction error. While the ERBlet achieves a perceptually motivated TF analysis comparable to that of linear gammatone filterbanks [16, 17], it allows perfect reconstruction even with a density of 1 filter per ERB (see Tab. 1). In comparison, a gammatone implementation designed to achieve near-perfect reconstruction (relative error = 10^{-7}) in [17] requires a minimum

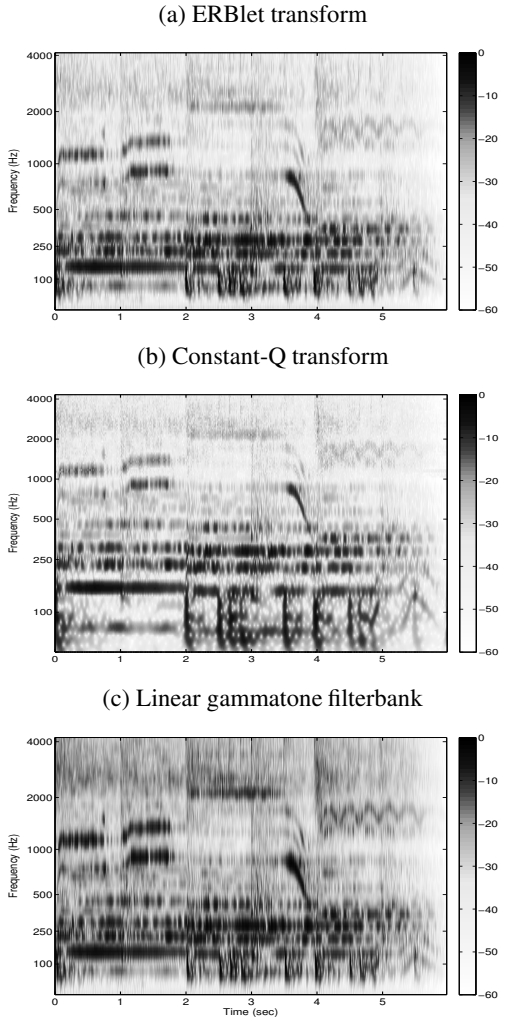

Fig. 2. TF representations (squared moduli, in dB) for (a) ERBlet, (b) constant-Q transform, and (c) gammatone filterbank (restricted to the relevant frequency band 0–4000 Hz).

density of 2.4 filters per ERB. Although our approach cannot substitute for physiologically plausible auditory models like [10], it could be useful to auditory modeling approaches in which a density of 1 filter per ERB is often desired [11].

To account for the level dependency of the auditory filters' bandwidth and the compressive response of the cochlea [6], an approximately invertible nonlinear gammatone filterbank was proposed in [18]. To further improve the match between the auditory and the transform resolutions while retaining the perfect reconstruction property of the ERBlet, future works include: inclusion of compression and use of windows with Gaussian shapes on the ERB scale (*i.e.*, using a warping function mapping the linear frequency axis to the ERB scale).

6. REFERENCES

[1] P. Flandrin, *Time-frequency/Time-scale analysis*, vol. 10 of *Wavelet analysis and its application*, Academic Press, San Diego, 1999.

[2] T. Painter and A. Spanias, "Perceptual coding of digital audio," in *Proceedings of the IEEE*, April 2000, vol. 88, pp. 451–515.

[3] P. Balazs, B. Laback, G. Eckel, and W. A. Deutsch, "Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking," *IEEE Audio, Speech, Language Process.*, vol. 18, no. 1, pp. 34–49, 2010.

[4] P. Balazs, M. Dörfler, N. Holighaus, F. Jaillet, and G. Velasco, "Theory, implementation and applications of nonstationary Gabor frames," *J. Comput. Appl. Math.*, vol. 236, no. 6, pp. 1481–1496, 2011.

[5] G. A. Velasco, N. Holighaus, M. Dörfler, and T. Grill, "Constructing an invertible constant-Q transform with nonstationary Gabor frames," in *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, Paris, France, September 19–23 2011, pp. 93–99.

[6] B. C. J. Moore, *An introduction to the psychology of hearing*, Emerald Group Publishing, sixth edition, 2012.

[7] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103–138, 1990.

[8] N. H. van Schijndel, T. Houtgast, and J. M. Festen, "Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency-time window," *J. Acoust. Soc. Am.*, vol. 105, no. 6, pp. 3425–3435, June 1999.

[9] E. A. Lopez-Poveda and R. Meddis, "A human nonlinear filterbank," *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3107–3118, December 2001.

[10] R. Meddis, W. Lecluyse, N. R. Clark, T. Jürgens, C. M. Tan, M. R. Panda, and G. J. Brown, "A computer model of the auditory periphery and its application to the study of hearing," in *Proceedings of the 16th International Symposium on Hearing (ISH 2012)*, Cambridge, UK, July, 23–27 2012.

[11] C. Feldbauer, G. Kubin, and W. B. Kleijn, "Anthropomorphic coding of speech and audio: A model inversion approach," *EURASIP J. Adv. Sig. Process.*, vol. 2005, no. 9, pp. 1334–1349, 2005.

[12] R. Pichevar, H. Najaf-Zadeh, L. Thibault, and H. Lahdili, "Auditory-inspired sparse representation of audio signals," *Speech Commun.*, vol. 53, no. 5, pp. 643–657, 2011.

[13] P. Philippe, F. M. de Saint-Martin, and M. Lever, "Wavelet packet filterbanks for low time delay audio coding," *IEEE Speech Audio Process.*, vol. 7, no. 3, pp. 310–322, May 1999.

[14] M. D. Abolhassani and Y. Salimpour, "A human auditory tuning curves matched wavelet function," in *Engineering in Medicine and Biology Society (EMBS 2008). 30th Annual International Conference of the IEEE*, Vancouver, Canada, August 20–24 2008, pp. 2956–2959.

[15] J. J. O'Donovan and D. J. Furlong, "Perceptually motivated time-frequency analysis," *J. Acoust. Soc. Am.*, vol. 117, no. 1, pp. 250–262, January 2005.

[16] V. Hohmann, "Frequency analysis and synthesis using a gammatone filterbank," *Acta Acust. united Ac.*, vol. 88, no. 3, pp. 433–442, 2002.

[17] S. Strahl and A. Mertins, "Analysis and design of gammatone signal models," *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2379–2389, November 2009.

[18] T. Irino and R. D. Patterson, "A dynamic compressive gammachirp auditory filterbank," *IEEE Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 2222–2232, November 2006.

[19] F. Baumgarte, "Improved audio coding using a psychoacoustic model based on a cochlear filter bank," *IEEE Speech Audio Process.*, vol. 10, no. 7, pp. 495–503, October 2002.

[20] Z. Cvetković and J. D. Johnston, "Nonuniform oversampled filter banks for audio signal processing," *IEEE Speech Audio Process.*, vol. 11, no. 5, pp. 393–399, September 2003.

[21] K. Gröchenig, "Acceleration of the frame algorithm," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3331–3340, December 1993.

[22] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*, SIAM, 1997.

[23] P. Balazs, H. G. Feichtinger, M. Hampejs, and G. Kracher, "Double preconditioning for Gabor frames," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4597–4610, December 2006.

[24] P. L. Søndergaard, B. Torrésani, and P. Balazs, "The linear time frequency analysis toolbox," *Int. J. Wavelets. Multi.*, vol. 10, no. 4, pp. 1250032, July 2012.