EM-DRIVEN STEREO-LIKE GAUSSIAN CHIRPLET MIXTURE ESTIMATION

Luis Weruaga [1] *and Marián Képesi* [2]

[1] Commission for Scientific Visualization, Austrian Academy of Sciences, Vienna, Austria.
[2] Telecommunications Research Center Vienna, Austria.

ABSTRACT

This paper proposes a novel technique for the time-frequency (TF) estimation of a Gaussian chirplet mixture in a signal. The technique relies on the combination of one and a half-worth of Chirp transform of the signal and an Expectation Maximization-driven blind identification algorithm, in such a way that it carries out the chirplet estimation in-segment. The mechanism resembles stereo processing while keeping an interesting biological parallel with auditory cortex artificial models. This technique applied on a short-time processing basis is shown to be accurate at identifying the TF composition of a bat echolocation chirp. The proposed technique represents the initial basis for the development of a parallel hierarchic processing for more ambitious objectives.

1. INTRODUCTION

Several natural and artificial phenomena can be well described by a chirp function, for instance, signals produced by mammalian species, signals transmitted by radar systems [1], the human speech in natural intonation [2], etc. The so-called Gaussian chirplet [1] can be defined with a four-parameter model [3],

$$g_{\varrho,\tau,\nu,\gamma}(t) = \frac{1}{\sqrt[4]{\pi\varrho^2}}\, e^{-\frac{1}{2}\left(\frac{t-\tau}{\varrho}\right)^2} e^{j2\pi\left(\nu(t-\tau)+\frac{\gamma}{2}(t-\tau)^2\right)}\,, \quad (1)$$

where t is time, ϱ is the chirplet's effective duration, τ is the central time instant, ν is the instantaneous frequency at $t=\tau$, and γ is the chirp rate, which creates a linear variation of the frequency over t.

A main research interest in signal processing is to decompose signals into well-defined and localized components in the time-frequency (TF) plane. Since the set of Gaussian chirplets do not form an orthogonal basis, covering the TF plane in a redundant way, the chirplet-based decomposition cannot be easily addressed with a single-step transformation. This problem has attracted large interest in the last decade [3]-[6]. A well-accepted procedure for multicomponent signal analysis is the use of matching pursuits dictionaries [7], whose use with chirplet bases has been considered among others in [3][4]. A different approach is that based on the maximum likelihood estimation criterium [5], in which the computation of the chirplet parameters is estimated rather than 'searched'. An interesting adaptive technique was proposed by the inventors of the Chirplet transform [1], who address this estimation by means of an unsupervised Gaussian fitting mechanism on the TF energy [6]. In general this problem or related ones when addressed with state-of-the-art TF linear transforms require expensive decision procedures [9][8] that recall manual approaches.

Recent neurophysiological studies suggest modelling the cortical auditory processing with the Chirp transform and an unsupervised pattern recognition algorithm [10]. Somehow inspired by these ideas, in this paper we propose a parallel-computing mechanism for the estimation of a real-valued Gaussian chirplet mixture, this estimation relying on a single Chirp transform instance.

The problem is presented as follows: let $x(t)$ be a real signal,

$$x(t) = c(t) + n(t)\,, \quad (2)$$

where $n(t)$ is a random stationary noise and $c(t)$ is is a sum of complex conjugate Gaussian chirplets

$$c(t) = \sum_i g_{\varrho_i,\tau_i,\nu_i,\gamma_i}(t) + g_{\varrho_i,\tau_i,-\nu_i,-\gamma_i}(t)\,, \quad (3)$$

the goal of this work is to estimate the four parameters of each chirplet present in $x(t)$, while keeping up with the estimation of the background noise power spectrum. Section 2 contains the characterization of a chirplet in terms of the Chirp transform, Section 3 presents the proposed technique, Section 4 contain results on synthetic and real data, and the conclusions close the paper.

2. CHIRP ANALYSIS OF GAUSSIAN CHIRPLETS

A complex Gaussian chirplet is the only signal with a positive Wigner–Ville distribution. Chirplet (1) can be built by starting with the Gaussian signal of variance ϱ^2, whose Wigner–Ville distribution (WVD) is the bidimensional Gaussian

$$\mathcal{G}_{\varrho,0,0,0}(t,f) = 2\, e^{-\varrho^{-2}t^2}\, e^{-4\pi^2\varrho^2 f^2}\,. \quad (4)$$

The next step corresponds to the product of a γ-quadratic chirp [1]. It is straightforward to show that this step introduces the following shear in the Wigner plane: $t \to t$ and $f \to f-\gamma t$. This yields immediately to the WVD of $g_{\varrho,0,0,\gamma}(t)$

$$\mathcal{G}_{\varrho,0,0,\gamma}(t,f) = 2\, e^{-\varrho^{-2}t^2}\, e^{-4\pi^2\varrho^2(f-\gamma t)^2}\,. \quad (5)$$

The chirplet $g_{\varrho,\tau,\nu,\gamma}(t)$ is finally obtained after time and frequency shift to τ and ν respectively, yielding to the following WVD

$$\mathcal{G}_{\varrho,\tau,\nu,\gamma}(t,f) = 2\, e^{-\varrho^{-2}(t-\tau)^2} e^{-4\pi^2\varrho^2(f-\nu-\gamma(t-\tau))^2}\,. \quad (6)$$

The Chirp Transform (ChT) of $x(t)$ is computed as

$$X(f,\alpha) = \int_{-\infty}^{\infty} x(t)\, e^{-j\pi\alpha t^2} e^{-j\,2\pi f t}\, dt\,, \quad (7)$$

where f is frequency and α is the analysis chirp rate. Equation (7) is similar to the so-called "Chirplet" transform, defined by Mann and Haykin in [1], except that here the basis is not windowed by a Gaussian. The ChT can be understood as a time-frequency shear of the signal followed by a Fourier transform (FT), the time-frequency warping being described by

$$t \to t \quad \text{and} \quad f \to f + \alpha\, t\,. \tag{8}$$

Let $x(t)$ be the Gaussian chirplet $g_{\varrho,\tau,\nu,\gamma}(t)$; the WVD of the signal after the transformation (8) is

$$\mathcal{X}(t,f,\alpha) = 2\, e^{-\varrho^{-2}(t-\tau)^2} e^{-4\pi^2\varrho^2(f+\alpha t-\nu-\gamma(t-\tau))^2}. \tag{9}$$

The time marginal gives rise to the square magnitude of the Gaussian chirplet's Chirp transform, which results Gaussian shaped

$$|X(f,\alpha)|^2 = \int_{-\infty}^{\infty} \mathcal{X}(t,f,\alpha)\, dt = \tfrac{1}{\sqrt{2\pi}\sigma}\; e^{-\frac{1}{2}\frac{(f-\nu+\alpha\tau)^2}{\sigma^2}}\;, \tag{10}$$

where

$$\sigma^2 = \tfrac{1}{8\pi^2}\,\varrho^{-2} + \tfrac{1}{2}(\alpha-\gamma)^2\varrho^2\,. \tag{11}$$

In case of a real-valued chirplet (3), each complex chirplet results affected asymmetrically by the transformation (8): both Gaussians are centered respectively on f_+ and f_-

$$f_\pm = \pm\nu + \alpha\tau\,, \tag{12}$$

and their variance result σ_+^2 and σ_-^2 respectively

$${\sigma_\pm}^2 = \tfrac{1}{8\pi^2}\,\varrho^{-2} + \tfrac{1}{2}(\alpha\mp\gamma)^2\varrho^2\,. \tag{13}$$

Note that the WVD of a real-valued chirplet is not free of cross-terms, but its effect after the time projection (10) is depreciable. Thus the two-Gaussian model is accurate and only in cases of very high chirp rate α, both Gaussians may overlap in the projection direction and the cross terms be observed as zero-mean ripples on the Gaussian shapes. In general (13) yields to a double meaningful solution for both γ and ϱ. Given the quadratic nature of the equation system (13), a unique solution can be obtained with another reference, which can be provided by the FT (ChT for $\alpha = 0$). Then the following equation is added to the previous equation system

$$\sigma_0^2 = \tfrac{1}{8\pi^2}\,\varrho^{-2} + \tfrac{1}{2}\gamma^2\varrho^2\,. \tag{14}$$

Let us now assume that an accurate Gaussian fitting is available, that is, the centers and variances of the three Gaussians (two on each side of the ChT and another one from the FT) are at hand. Let these be μ_+, μ_-, μ_0, σ_+^2, σ_-^2 and ${\sigma_0}^2$ respectively. Then the chirplet parameters can be obtained as explained below.

Computation of ν *and* τ. Given that the positions of the Gaussians follow rule (12), ν and τ can be obtained by

$$\nu = \mu_0\,, \tag{15}$$

$$\tau = \frac{\mu_+ - \mu_0}{-\alpha} \quad \text{or} \quad \tau = \frac{\mu_- + \mu_0}{-\alpha}\,. \tag{16}$$

Computation of γ *and* ϱ. According to (13) the set of points $\mathcal{X} = \{-\alpha, 0, \alpha\}$, $\mathcal{Y} = \{\sigma_-^2, \sigma_0^2, \sigma_+^2\}$ lies on the parabola $y = ax^2 + bx + c$. The position of the minimum of that parabola corresponds to the estimation of the chirp rate γ,

$$\gamma = -\frac{b}{2a} \quad \text{and} \quad \varrho = \sqrt{2a}\,. \tag{17}$$

A constraint among the parabola parameters, derived from (13), is that the value of the parabola at its minimum is $\frac{1}{8\pi^2}\varrho^{-2}$, and since $a = \frac{1}{2}\varrho^2$, the parameter c is a function of the other two as

$$c = \tfrac{1}{4}\left(b^2 + \tfrac{1}{4\pi^2}\right)\, a^{-1}\,. \tag{18}$$

The values of a and b that keep the correspondence $\mathcal{X} \to \mathcal{Y}$ are obtained with the solution of the following least square fitting

$$\{a,b\} = \arg\min_{a,b} \sum_i \left(y_i - a{x_i}^2 - bx_i - c\right)^2\,. \tag{19}$$

3. ESTIMATION OF THE CHIRPLET MIXTURE

The asymmetric nature of the ChT, instead of being a limitation, provides a stereo view of the chirplet TF energy that can be exploited for estimating its parameters. Stationary wavelets are observed at both sides of the ChT with the same energy distribution, while a chirplet is observed asymmetrically, with an energy distribution of one side more concentrated than in the other. Based on the exposition so far we propose the mechanism depicted in Figure 1 for estimating a mixture of Gaussian chirplets: the real signal $x(t)$ goes through a bank of (at least two) chirplet transform blocks; the use of the α-Chirp and the 0-Chirp (Fourier) transforms provide three different projection views of the Wigner plane; since the chirplet projections result in Gaussian shapes, these from in each view are blindly identified by means of Radial Basis Function (RBF) neural network driven by an unsupervised learning algorithm (as suggested in [10]); the Gaussian fitting blocks do not carry out their respective tasks in an isolated way, but they cooperate among each other in order to achieve a meaningful description of the Gaussian chirplet.

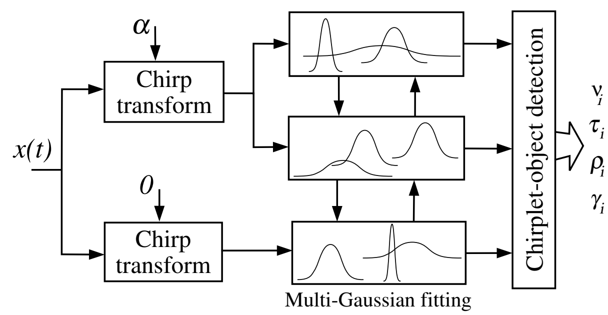

Fig. 1. Mechanism of the stereo-like Chirp estimation.

Let $x[n]$ be the discrete-time signal resulting from sampling $x(t)$ at a frequency rate f_s. In a real situation $x[n]$ has finite time support, then the Chirp transform is computed in discrete-time by the discrete Chirp transform[1] (DChT),

$$X[k,\hat{\alpha}] = \sum_{n=0}^{N-1} x[n]\, e^{-j\pi\hat{\alpha}(n-(N-1)/2)^2}\; e^{-j\frac{2\pi}{N}kn}\,, \tag{20}$$

where N is the signal length, $\hat{\alpha} = \alpha/{f_s}^2$ is the discrete-time chirp rate and $k = 0, \ldots, N-1$ is the frequency index. Let $S[k,\hat{\alpha}]$ be the normalized ChT power spectrum of signal $x[n]$, that is

$$S[k,\hat{\alpha}] = \left|X[k,\hat{\alpha}]\right|^2 / \sum_{i=0}^{N} \left|X[i,\hat{\alpha}]\right|^2\,. \tag{21}$$

[1] In order to avoid spectral aliasing from the product with the complex chirp in (20), $x(t)$ must be sampled at higher rate than the Nyquist rate.

Considering that $x[n]$ follows the chirplet mixture model described by (3), the power spectrum $S[k, \hat{\alpha}]$ is a uniform sampling of the following Gaussian mixture model

$$S_\alpha(x) = \sum_m \lambda_+^m \, G(x|\Theta_+^m) + \lambda_-^m \, G(x|\Theta_-^m) \; + \; N_\alpha(x) \,, \quad (22)$$

where

$$G\left(x|\{\mu, \sigma^2\}\right) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \,. \quad (23)$$

Here the superscript ^{m} denotes the m-th real-valued chirplet and $N_\alpha(x)$ contains the projection of the cross terms of the WVD and that of the noise $n(t)$. The cross terms result from the TF interference among the Gaussian chirplets and give rise to well-known undesired ripples in the power spectrum. The third view, computed with the FT, follows a similar model to (22) with the characteristic that the model is symmetric, $S_0(x) = S_0(-x)$ (the cross term projection is accordingly $N_0(x)$). Although the cross terms represent a disturbance in the Gaussian mixture model, this drawback is alleviated by the fact that the cross term projection has zero mean and is not observed in the same way in all projection views.

In order to identify the Gaussian chirplets present in $x[n]$ we use the Expectation Maximization (EM) algorithm [11] as the mechanism for drawing the Gaussian mixture model from the power spectra. The EM allows to estimate jointly the amplitude, variance and center of each Gaussian, and it pursues the minimization of the mean square error of the log-spectrum, fact that finds an interesting parallel with the log-sensitivity of the biological senses. Instead of a single EM algorithm fed with multidimensional data (as presented in [6]), the solution proposed here is totally different: several (at least two) EM algorithms running in parallel, each working on a different one-dimensional TF view; the resulting parameters from each algorithm at each iteration are coupled according to the chirplet constraints, making the EM mechanisms evolve jointly. This fitting mechanism is described with the following iterative procedure:

1. Perform an iteration on each EM algorithm.
2. Compute for each Gaussian object, from its centers and variances from all views, the four-chirplet parameters as described by equations (15)-(17).
3. Recompute the Gaussian parameters based on the estimated four-chirlet parameters, as described by (12)-(14).
4. If the approximation error is small stop; if not go to step 1.

The concept of background noise $n(t)$ is ill-defined versus the Gaussian chirplet model. Nevertheless the noise can be identified with Gaussian objects whose parameters are not constrained to the corresponding parameters from other views. Additionally, given that the expected noise energy is usually smooth in time and frequency, the variance of the noise-based Gaussians is clipped to a (high enough) minimum value in each EM iteration. These actions prevent the noise objects from competing for the chirplet energy. In absence of noise objects, the chirplet objects are attracted also by the noise energy (when this energy is relevant) and the mechanism hardly reaches a meaningful and stable solution. Finally, the number of chirplet objects has to be larger than the actual number of chirplets present in the signal. In case of a smaller number, not all objects may reach a stable state. A number of chirplet objects larger than the number of actual chirplets in the signal gives rise to a faster convergence.

4. RESULTS

Since the proposed method delivers a parametric chirplet representation of the signal, the results are shown graphically with the *adaptive spectrogram* (AD): a non-negative time-frequency energy representation free of cross term interference [4].

The first experiment addresses the problem of estimating multiple synthetic chirplets in noise. The gaussian objects were initially distributed in the frequency axis. The results shown in Figure 2 prove the good performance of the method at analyzing this difficult scenario. Nevertheless the estimation of the parameters suffers an error due to the random effect of the noise energy. The final solution is achieved after few EM iterations (around 15 in average for that experiment) with five chirplet objects. The chirplet time location τ is responsible for the shift in the Gaussian central position in each view (12), this fact disturbing the Gaussian fitting task. The use of the disjunctive (16) for estimating τ prevents the algorithm from getting trapped in stalled states.

An example on real data corresponds to the analysis of a bat

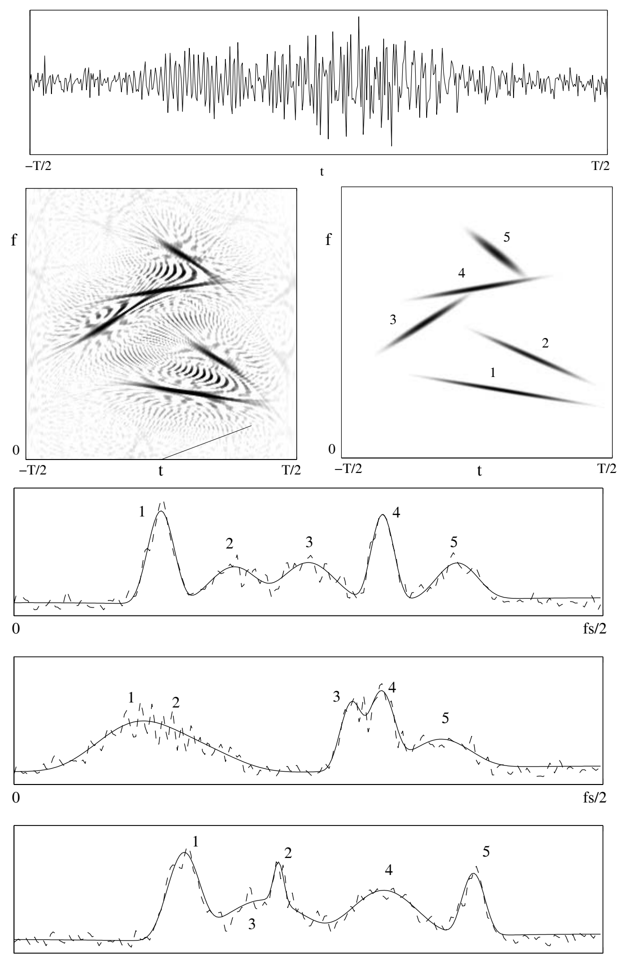

Fig. 2. Performance of the method on mixture of chirplets in Gaussian white noise (SNR=10dB). Figures from top to bottom: time samples, WVD (left) and estimated AD (right), magnitude of the three TF wiews (FT, ChT right and left side (mirrored)). The line in the WVD indicates the angle of projection of the ChT. The Gaussian objects are marked with numbers.

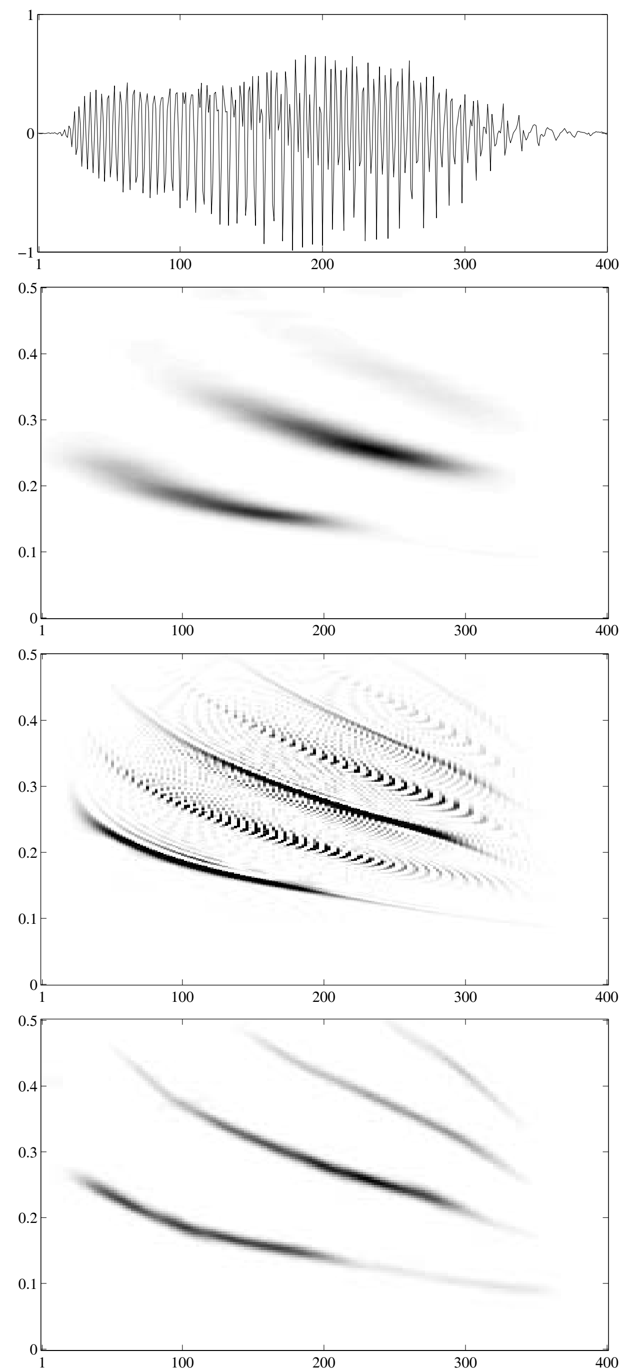

Fig. 3. Ultrasound bat signal. From top to bottom: time samples, spectrogram ($N = 128$), WVD, and estimated AD. The TF representation is in samples (horizontal axis) and normalized frequency (vertical axis). The intensity scale is linear.

echolocation ultrasound chirp, addressed in Figure 3. The signal (top) has 400 samples and the sampling period is 7 microseconds. The 128-point Hamming window spectrogram and the WVD are shown in the same Figure 3 (second and third from the top respectively). Since the harmonics are not Gaussian linear chirplets, a short-time based variant of the proposed method was used: the signal was segmented with a 128-point Gaussian window, using a shift of 32 samples between overlapped segments. This way of proceeding limits the frequency resolution to that of the window length but it prepares the signal according to the requirements of the proposed technique. The shift between segments, one fourth of the segment length, guarantees an AD-based TF energy representation with continuity. The result of the proposed technique is shown at the bottom of Figure 3: the TF energy localization is very precise, small details in the signal evolution being also visible. The number of chirplet objects in all segments is five (whilst the largest number of harmonics ever present in a segment is four). At any segment only chirp objects that represent significant energy result visible in the AD.

5. CONCLUSIONS

This paper has proposed a novel technique for the identification of a Gaussian linear chirplet mixture in a signal. The use of the Chirp transform and an EM-driven Gaussian fitting blind mechanism is a simple effective parallel-computing in-frame method with biological parallel. The results on synthetic as well as real signals prove the validity of the proposed technique, showing a clear advantage over the spectrogram and delivering a cross term free time–frequency representation that is competitive with the Wigner–Ville representation. Since in long signals the Gaussian shape assumption is likely not to hold, a short-time-based processing makes the technique applicable for general scenarios. This work does not aim to be a conclusive one, but to settle the basis for the development of a parallel hierarchic technique for more ambitious objectives.

6. ACKNOWLEDGMENTS

The authors wish to thank Curtis Condon, Ken White, and Al Feng of the Beckman Institute of the University of Illinois for the bat data and for permission to use it in this paper.

7. REFERENCES

[1] S. Mann, S. Haykin, "The chirplet transform: physical considerations," *IEEE Trans. SP*, pp. 2745-2761, Nov., 1995.

[2] L. Weruaga, M. Képesi, "Speech analysis with the short-time chirp transform," *Proc. Eurospeech*, pp. 53-56, Geneva, Sept. 2003.

[3] A. Bultan, "A four-parameter atomic decomposition of chirplets," *IEEE Trans. SP*, pp. 731-745, Mar. 1999.

[4] Q. Yin, S. Qian, A. Feng, "A fast refinement for adaptive gaussian chirplet decomposition," *IEEE Trans. SP*, pp. 1298-1306, June 2002.

[5] J.C. O'Neill, P. Flandrin, "Chirp hunting," *Proc. IEEE Int. Symp. Time-Freq & -Scale Anal.*, pp. 425-428, 1998.

[6] S. Mann, S. Haykin, "Adaptive 'chirplet' transform: an adaptive generalization of the wavelet transform," *Optical Engineering*, pp 1243-1256, June 1992.

[7] S. Mallat, Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. SP*, pp. 3397-3415, Dec. 1993.

[8] M. Alper, H.M. Ozaktas, "Optimal filtering in fractional Fourier domains," *IEEE Trans. SP*, pp. 1129-1143, May 1997.

[9] C. Capus, K. Brown, "Short-time fractional Fourier methods for the t-f representation of chirp signals", *J. ASoA*, pp. 3253-3263, 2003.

[10] E. Mercado III, et al., "Modeling auditory cortical processing as an adaptive chirplet transform," *Neurocomp.* 32-22, pp. 913-919, 2000.

[11] J.A. Bilmes, " A gentle tutorial of the EM algorithm ...", Tech. Rep. 97-021, *Intl. Comp. Sci. Instit.*, Berkeley, 1997.