ANALYSIS/SYNTHESIS USING TIME-VARYING WINDOWS AND CHIRPED ATOMS

Corey Kereliuk,
SPCL* & CIRMMT†
McGill University, Montréal, Canada
corey.kereliuk@mail.mcgill.ca

Philippe Depalle,
SPCL* & CIRMMT†
McGill University, Montréal, Canada
depalle@music.mcgill.ca

ABSTRACT

A common assumption that is often made regarding audio signals is that they are short-term stationary. In other words, it is typically assumed that the statistical properties of audio signals change slowly enough that they can be considered nearly constant over a short interval. However, using a fixed analysis window (which is typical in practice) we have no way to change the analysis parameters over time in order to track the slowly evolving properties of the audio signal. For example, while a long window may be appropriate for analyzing tonal phenomena it will smear subsequent note onsets. Furthermore, the audio signal may not be completely stationary over the duration of the analysis window. This is often true of sounds containing glissando, vibrato, and other transient phenomena. In this paper we build upon previous work targeted at non-stationary analysis/synthesis. In particular, we discuss how to simultaneously adapt the window length and the chirp rate of the analysis frame in order to maximally concentrate the spectral energy. This is done by a) finding the analysis window that leads to the minimum entropy spectrum; and, b) estimating the chirp rate using the distribution derivative method. We also discuss a fast method of analysis/synthesis using the fan-chirp transform and overlap-add. Finally, we analyze several real and synthetic signals and show a qualitative improvement in the spectral energy concentration.

1. INTRODUCTION

A typical setup in time-frequency analysis consists of a) segmenting the time-domain audio signal into short duration slices; and, b) performing a spectral analysis on each slice. For example, let us use $h(t)$ to denote a compactly supported window function centered at the origin. We can extract a short slice of the audio signal, $y(t)$, via multiplication with the window function:

$$y_m(t) = y(t)h(t - am) \tag{1}$$

In this case a is the stride length and m is an integer index. The windowed slice $y_m(t)$ is a snapshot of the signal around time am. Provided that:

$$\sum_m h(t - am) = 1 \tag{2}$$

we can recover the original signal by overlap-adding the windowed slices. Windows which exhibit this property are often referred to as constant overlap-add (COLA) [1]. Taking the Fourier transform of the windowed slices yields the short-time Fourier transform (STFT). The STFT may also be viewed in terms of the inner products between the audio signal and a set of Gabor atoms:

$$g_{m,n}(t) = h(t - am)\exp(j2\pi bnt) \tag{3}$$

where the parameters a and b are the time and frequency sampling intervals, m and n are integer indices, $j = \sqrt{-1}$ is the imaginary unit, and t is time. Gabor atoms are a natural choice for modelling audio since they have a compact time-frequency footprint which can be used to approximate idealized rectangular 'tiles' in the time-frequency plane.

It is well-known that the time and frequency resolution of any waveform are inversely related (and this knowledge can be expressed using the uncertainty principle [2]). In other words, if we use a long duration window to analyze some phenomenon there will be some uncertainty as to when that phenomenon occurred. Likewise, a short window will yield accurate timing information at the cost of greater uncertainty in frequency. Although the uncertainty principle cannot be avoided, it is still advisable to choose a window length that is well-matched to the type of phenomenon we want to study. The problem with the STFT, however, is that once a window is chosen it cannot be changed in order to match the evolving structure of the audio signal.

In this paper we describe a simple adaptation of the COLA requirement for use with time-varying windows. We then show how this formulation can be used for time-frequency analysis/synthesis with time-varying chirped atoms. Some previous work has already been made in this direction, in particular, superposition frames [3] and non-stationary Gabor frames [4, 5, 6]. However, in [3] rather strong requirements were placed on the windows that can be used and in [4, 5, 6] the analysis was restricted to Gabor frames (e.g., atoms modulated by a constant frequency exponential). Our formulation, although similar, does not limit one to work with Gabor-type atoms. In fact, we suggest using chirped atoms in order to allow the frequency to evolve over the duration of the analysis window. This allows us to use sheared time-frequency tiles (as opposed to rectangular ones) to represent basic elements in time-frequency plane [7]. This can be quite beneficial in a musical context, for example, when analyzing glissando, vibrato, and other transient phenomena.

In the remainder of this paper we present a system for adaptive analysis of audio using time-varying windows and chirped atoms. This work brings together several separate contributions from the literature into a single analysis/synthesis system. The structure of this paper is laid out as follows. In the next section we introduce the concept of time-varying windows, and then discuss the necessary requirements for perfect resynthesis via overlap-add. We also examine the consequences these requirements have on the overall shape of the windows. Then in Sec. 3 we discuss a method for selecting a set of windows that match the underlying signal structures using the Renyi entropy. The resulting framework allows us

*Sound Processing and Control Laboratory
†Centre for Interdisciplinary Research in Music Media and Technology

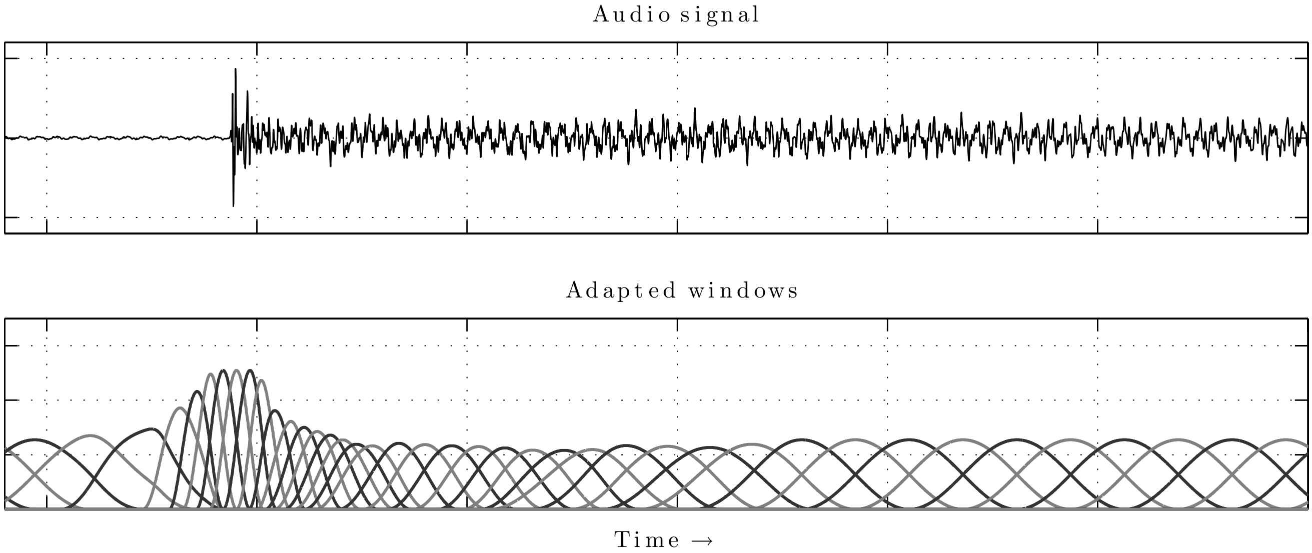

Figure 1: A set of time-varying windows adapted around the attack portion of a note onset.

to perform time-varying analysis/synthesis. In Sec. 4 and 5 we examine how to efficiently perform analysis/synthesis with chirped atoms by using the fan-chirp transform in conjunction with the distribution derivative method. Finally in Sec. 6 we present some experiments followed by a summary of our work.

2. TIME-VARYING WINDOWS

In order to allow the analysis window to vary over time we can replace $h(t-am)$ by $h_m(t-a_m)$. In other words, at time position a_m we window the signal with function h_m. This means that the set of windows is now time-varying, which affords us considerable flexibility. For example, we may adjust the window's scale and shape in order to better match a given phenomenon at a specific point in time. This adaptability can be quite beneficial in a musical context, for example, we could analyze a sharp attack followed by a sustained resonance using a short window followed by a set of longer windows. This type of scenario is depicted in Fig. 1.

2.1. Window restrictions

In order to allow for perfect re-synthesis we require the set of time-varying windows to satisfy the following set of inequalities

$$0 < A \leq z(t) \triangleq \sum_m |h_m(t-a_m)|^2 \leq B \quad \forall t \tag{4}$$

This requirement is tantamount to forcing the set of windows to overlap to some degree (as we will see in the following section more overlap is typically better). The bounds A and B ensure that $z(t)$ is bounded away from zero which allows $z(t)$ to be stably inverted [5]. When Eq. (4) is satisfied we can perfectly reconstruct the signal from its windowed segments $y_m(t) = y(t)h_m(t-a_m)$. That is to say, if we multiply each segment by the window a second time, overlap-add the segments, and finally divide by $z(t)$ (which is required to be non-zero) we will recover the original signal $y(t)$.

We can avoid the final division by $z(t)$ by normalizing the windows as follows:

$$\tilde{h}_m(t) = \frac{h_m(t-a_m)}{\sqrt{z(t)}} \tag{5}$$

This normalization is convenient because it allows us to use the same set of windows for both analysis and synthesis.

2.2. Window shapes

It is interesting to study how the degree of overlap changes the shape of the normalized windows. For example, Fig. 2 shows how the prototype window shape changes as a function of the overlap factor for a fixed (non time-varying) set of windows. As the amount of overlap decreases the normalized windows become increasingly rectangular (and as a result the height of the side-lobes will increase as well). Thus, not having enough overlap may lead to problems distinguishing between nearby components in the spectrum when using normalized windows. Furthermore, it is well known that insufficient overlap will lead to aliasing of the subband signals, which may not be perfectly cancelled if the spectral representation is modified (e.g., when implementing spectral audio effects) [8].

Insufficient overlap will lead to similar effects when the set of windows is time-varying. However, in this case the normalized windows may not be symmetric (even if the prototype windows were symmetric). This is because $z(t)$ is less regular when the set of windows and overlap factors are subject to change over time. For example, we can see that some of the normalized windows in Fig. 1 are not perfectly symmetric. Note that normalizing the windows as suggested in the previous section is only required if we want to use the same set of windows for both analysis and synthesis. If symmetric analysis windows are required (e.g., for a linear-phase filterbank analysis), then the windows $h_m(t-a_m)/z(t)$ may be used for synthesis.

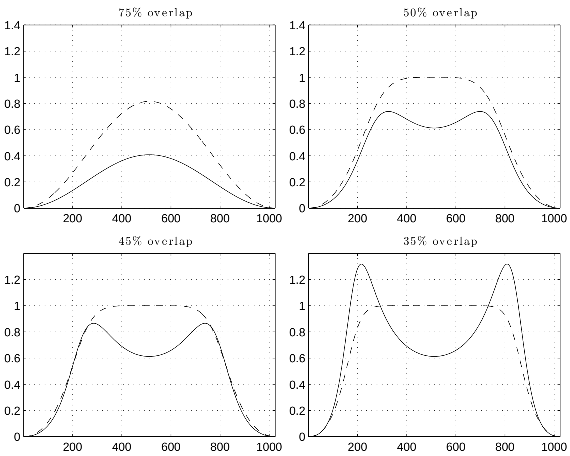

Figure 2: Illustration of the window shape vs. overlap factor for a fixed analysis window. The dashed line shows the shape of the normalized windows; the solid line shows the shape of the synthesis windows when no normalization is used.

3. WINDOW SELECTION

Although we have discussed a framework for time-varying analysis/synthesis, a central question that remains unanswered is: how do we select the set of time-varying windows? The method used to select the time-varying windows should be signal adaptive, since we want the windows to be well-matched to the signal content. In [5] a method for selecting windows was proposed based on onset detection. For example, we can use short duration windows near each onset and progressively longer windows as we move further away from each onset. Although this method works well for sounds with well-defined onsets it is more problematic for sounds with weaker forms of non-stationarity. Furthermore, an onset detector will make a hard (binary) decision with regard to the presence/absence of a note event, which can upset the continuity of the time-frequency representation (especially when false positives are made).

Another approach for window selection is to select the set of windows that leads to the most concentrated spectrum (e.g., the set of windows that spreads the energy the least among the time-frequency coefficients). For example, in [9] the spectral kurtosis was proposed as a tool for adapting time-frequency representations. The kurtosis can be used to measure the ‘peakiness’ of a probability distribution; therefore, selecting a set of windows in order to maximize the kurtosis should result in a high energy concentration.

In this work we propose using the Renyi entropy as a measure of spectral concentration. This same approach was recently proposed in [6] as well, although our use of the Renyi entropy was inspired by [9] and [10]. It is well-known that the more diffuse a density is the higher its entropy will be (and the maximum entropy distribution on an interval is achieved by the uniform distribution). Thus, like the spectral kurtosis, the Renyi entropy can be used as a measure of energy concentration. We employ the following definition of the discrete short-term Renyi entropy

$$H_n[m] = \frac{1}{1-\gamma}\log_2 \frac{\sum_{l=-d}^{d}\sum_k |S_m[l-n,k]|^{2\gamma}}{\left(\sum_{l=-d}^{d}\sum_k |S_m[l-n,k]|^2\right)^{\gamma}} \tag{6}$$

where γ is a non-negative real number, $S_m[n,k]$ is the STFT using the m^{th} window and n and k index points on the time-frequency grid. We require each STFT to be sampled on the same time-frequency grid so that the extent of the summation is the same in each case. This in turn requires some oversampling of the STFTs. The parameter d controls the number of time-frames over which the short-term entropy calculation ranges. We also note that a smoothing window could be included in this equation in order to de-emphasize the spectral frames further from the central time instant[1].

At a given time-frame we can select the window that leads to the minimum short-term entropy as

$$m_n = \arg\min_m H_n[m] \tag{7}$$

Thus at each time-frame we have a method by which to select a window (and this window should concentrate the spectral energy). The set of windows can be optionally pruned, for example, if the overlap between windows is too great due to the oversampling required for the entropy calculation.

When using the Renyi entropy to choose a set of windows we have found that setting $\gamma \leq 1$ tends to lead to more satisfying results. This is due to the fact that high frequency energy in audio tends to be quite weak relative to the low frequency energy, however, (owing to the large dynamic range of the human auditory system) this energy is still perceptually relevant. In the entropy calculation the spectral coefficients are raised to the power of γ, so when γ is greater than 1, small coefficients will be given less weight in comparison to large coefficients. This means that the entropy calculation will tend to preference the large coefficients, which are predominately in the lower frequency bands. Using a value of γ less than 1 gives small coefficients relatively more weight in the entropy calculation, which allows low energy coefficients in the upper frequency bands to have more of an influence on the window selection. The experiments in Sec. 6 help to verify this claim.

4. CHIRP BASIS

After we have chosen a set of analysis windows we may proceed to analyze the spectral content of each windowed slice. We could, for example, analyze these slices using the Fourier transform (as in [4, 5, 12]). In this work, on the other hand, we propose to perform the analysis using chirped atoms:

$$\phi_{f,\alpha}(t) \triangleq \sqrt{|1+\alpha t|}e^{j2\pi f(1+\frac{\alpha}{2}t)t} \tag{8}$$

As shown in [13] the atoms in Eq. (8) form a basis over the interval $[-T/2, T/2]$ if the chirp parameter α is restricted to $|\alpha| < 2/T$. This constraint can be easily satisfied by constraining the chirp parameter based on the duration of the analysis window. The atoms in Eq. (8) fan out in frequency and thus are ideal candidates for modelling and/or approximating non-stationary harmonic phenomena (e.g., glissando, vibrato, and other transient events).

[1] we have not shown this explicitly to avoid cluttering the notation [11].

Performing the analysis using the atoms in Eq. (8) is referred to as the fan-chirp transform (FChT) [13]. In continuous-time the FChT of a windowed slice is

$$\langle y_m, \phi_{f,\alpha} \rangle = \int y(t)\tilde{h}_m(t - a_m)\sqrt{|1+\alpha t|}e^{-j2\pi f(1+\frac{\alpha}{2}t)t}dt \tag{9}$$

$$= \int y_m(t)\sqrt{|1+\alpha t|}e^{-j2\pi f(1+\frac{\alpha}{2}t)t}dt \tag{10}$$

As recognized in [14] and outlined in [13] fast computation of the FChT can be carried out by appropriately warping the signal. This may be done by making the following change of variables: $\tau = \Upsilon(t) = (1 + \alpha t/2)t$. Thus the integral in Eq. (10) becomes

$$\int_{\tau=\Upsilon(-1/\alpha)}^{\infty} \left(\frac{y_m(\Upsilon^{-1}(\tau))}{\sqrt{1+\alpha\Upsilon^{-1}(\tau)}} \right) e^{-j2\pi f\tau} d\tau \tag{11}$$

where $\Upsilon(t)$ has the following inverse (on $t \geq -1/\alpha$)

$$\Upsilon^{-1}(\tau) = -\frac{1}{\alpha} + \frac{\sqrt{1+2\alpha\tau}}{\alpha} \tag{12}$$

In this form, the FChT can be seen as the Fourier transform of the warped (and normalized) signal. As such we may compute the discrete FChT by first warping the signal and then performing an FFT. However, the warped sample points $\Upsilon^{-1}(\tau)$ will not typically lie on the sampling grid. In this case we must use interpolation to find the intersample values. As was suggested in [13] we opt to use cubic spline interpolation, which is very fast, although less accurate than bandlimited interpolation for audio signals [15]. Imperfect interpolation in the signal warping is a source of error in the (fast) FChT. Furthermore, when working in discrete-time we must be careful to avoid aliasing of the chirped atoms. This can be done by oversampling the signal such that aliasing occurs in an unused portion of the spectrum. For example, if we limit $|\alpha| < 1/T$, we require $2\times$ oversampling in order to avoid aliasing.

Figure 3 illustrates the spectrum of the FChT on a windowed slice of a synthetic chirp signal with five harmonics. The fundamental frequency was swept between 400 Hz and 800 Hz over 256 ms. The fundamental frequency at the midpoint of the chirp was thus 600 Hz. We note that the highest harmonic was swept between 2000 Hz and 4000 Hz making it the most non-stationary component of the signal. The windowed slice for the FFT and FChT were each centered at the midpoint of the chirp and were 512 samples long (the sampling rate was 8000 Hz). The chirp parameter α was set by an oracle. In comparison with the FFT we see that the FChT better concentrates the energy of the chirp. Furthermore, we can observe a progressive widening of the main-lobe as the frequency increases with the FFT (this is due to the fact that the higher harmonics are increasingly non-stationary). The FChT main-lobes on the other hand are more clearly resolved and remain relatively constant in width.

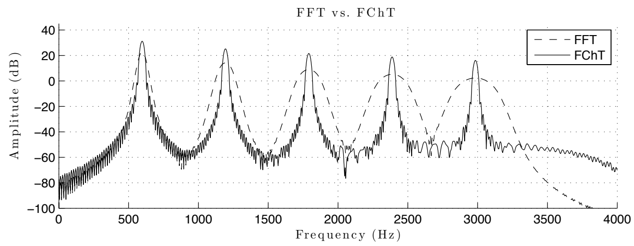

Figure 3: Comparison of FFT and FChT spectrums on synthetic chirped signal with five harmonics. The chirp parameters are described in the text.

5. ESTIMATING THE CHIRP PARAMETER

Since in practice we cannot set the chirp parameter using an oracle, we need a method by which to estimate it. There are several ways we could attempt to estimate the chirp rate, including intra-frame and inter-frame techniques. The later approach would involve trying to find spectral peaks in consecutive frames that are connected to the same underlying sinusoid [13, 16]. The chirp rate could then be estimated based on the trajectory of these peaks, much like in a partial tracking system. Another approach is to estimate the chirp rate based on a single frame of data. This approach relieves some of the complexities/difficulties of peak tracking and allows one to estimate a new chirp rate at every frame. This could become important in rapidly changing signals, where peak tracking algorithms would face difficulty. Intraframe approaches include the quadratically interpolated FFT (QIFFT) [17], reassignment (RA) method [18], derivative method (DM) [19], and distribution derivative method (DDM) [20]. An extensive review and comparison of these estimators can be found in [21].

In this work we have chosen to use the DDM method to estimate the chirp rate since it only requires us to perform a single derivative on the analysis window, and this can be accomplished analytically for many parametric windows. The RA and DM methods, to the contrary, require us to compute multiple derivatives either on the signal or on the window. In the former case, one must estimate the signal derivative from discrete-time samples, and in the latter case one must use windows that are continuously differentiable at high orders.

5.1. Distribution derivative method

The DDM is a method for estimating the parameters of a signal model based on the relation

$$\left\langle y', g_{m,n} \right\rangle = -\left\langle y, g'_{m,n} \right\rangle \tag{13}$$

which equates the signal derivatives with the derivative of a test function (in this case the test function is a non-stationary Gabor atom). For this relation to hold the test function must be supported on a finite interval and be once differentiable as described in [20]. In the DDM the signal, $y(t)$, is typically modelled as a polynomial phase sinusoid:

$$y(t) = \exp\left(\sum_{q=0}^{Q} \zeta_q t^q \right) \tag{14}$$

where ζ_q are complex coefficients whose real part corresponds to amplitude modulation and whose imaginary part corresponds to frequency modulation. In this case we may find $y'(t)$ analytically

$$y'(t) = \left(\sum_{q=1}^{Q} q\zeta_q t^{q-1} \right) y(t) \tag{15}$$

We can also find $g'_{m,n}(t)$ analytically, so long as the Gabor atoms we use are once differentiable. Subbing Eq. (15) into Eq. (13), we get

$$\sum_{q=1}^{Q} \zeta_q \left\langle q t^{q-1} y, g_{m,n} \right\rangle = -\left\langle y, g'_{m,n} \right\rangle \qquad (16)$$

where

$$g'_{m,n}(t) = h'_m(t - a_m)e^{j2\pi ft} + j2\pi f h_m(t - a_m)e^{j2\pi ft} \qquad (17)$$

In order to find the right-hand side of Eq. (16) at a given time and fixed set of frequencies we may compute two FFT's a) the FFT of the signal pre-windowed by the derivative window; and, b) the FFT of the windowed signal. The inner products on the left-hand side can be calculated using a third FFT, of the signal pre-multiplied by the ramp qt^{q-1}. For our application (chirp rate estimation), we consider a quadratic phase polynomial, i.e., $Q = 2$. In general we pick the largest magnitude peak[2] in the energy spectrum and a few adjacent bins to perform the estimation. We may then solve Eq. (16) by finding the best set of parameters $\zeta_i, i = 1, 2$ in the least-squares sense. We are ultimately interested in the quantity

$$\hat{\alpha} = 2\frac{\mathcal{I}\{\zeta_2\}}{\mathcal{I}\{\zeta_1\}} \qquad (18)$$

which is an estimate of the chirp parameter for our time-varying chirped atoms (the operator $\mathcal{I}\{\cdot\}$ returns the imaginary part of its argument).

6. EXPERIMENTS

In this section we analyze several real and synthetic signals. We have used the Renyi entropy in order to determine the set of time-varying windows and the DDM to estimate the chirp parameter.

We begin by examining the time-varying transform of a synthetic audio signal. This signal consists of a sequence of five different phenomena, each with a distinct time-frequency footprint:

1. a short pulse;
2. a medium duration harmonic oscillation;
3. a long duration harmonic oscillation;
4. a harmonic glissando; and,
5. a harmonic vibrato.

The sampling frequency of this synthetic signal was set to 8kHz. Figure 4 illustrates this signal and three different spectrograms computed using short, medium, and long window lengths. As can be seen, none of the spectrograms can simultaneously concentrate all of the different phenomena. For example, the long window smears the short pulse and the short window smears the longer oscillations. Figure 5 shows the time-varying analysis of the same signal. In this case the short pulse and longer oscillations are simultaneously concentrated. We note that, as expected, there is some spreading of energy in the glissando and vibrato since these phenomena do not remain constant over the duration of the analysis window. Figure 5 also shows how the window length changes in order to adapt to the degree of non-stationarity present in the signal at each time instant. The results for $\gamma \leq 1$ are slightly more satisfying than for $\gamma > 1$ as predicted in Sec. 3.

Figure 6 shows the same synthetic signal analyzed with time-varying windows and the FChT. The glissando and vibrato appear

[2]this is the maximum likelihood frequency estimate for a stationary sinusoid [22].

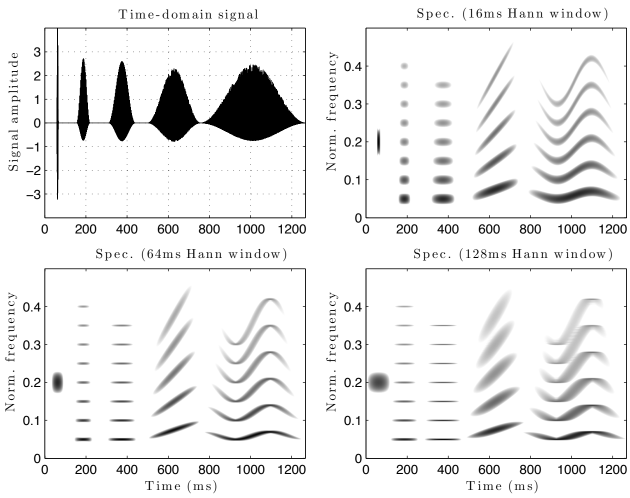

Figure 4: A time-varying synthetic signal and three spectrograms computed using short, medium, and long analysis windows, respectively.

to be more concentrated in this case (especially the higher harmonics). We can also see that the DDM is able to provide a relatively good estimate of the chirp parameter (despite the fact that the synthetic signal has several harmonics). We can observe a slight aberration of the chirp estimate near the boundaries of the different phenomena (e.g., near note onsets).

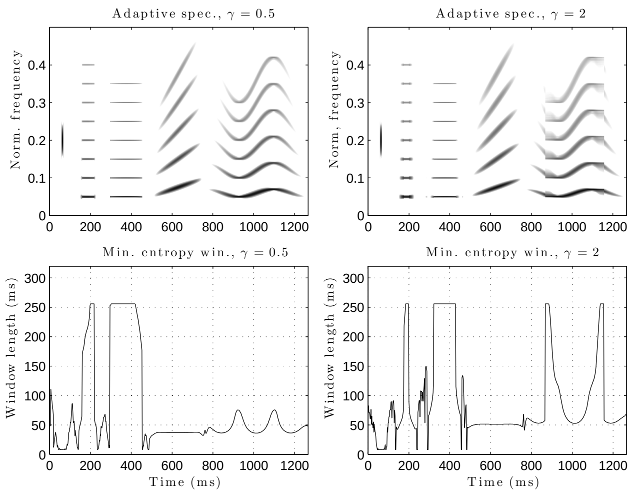

Figure 5: Analysis of the synthetic signal from Fig. 4 with time-varying windows selected using the Renyi entropy. Left column: $\gamma = 0.5$, Right column: $\gamma = 2$. Top row: time-varying spectrograms, Bottom row: selected window lengths at each time instant.

In Fig. 7 we consider a real glockenspiel signal which has well defined transient and tonal parts. We can observe that the time-varying spectrogram has as a frequency resolution comparable with that of the spectrogram calculated with a long duration window. Furthermore, the transients are well concentrated in the time-varying spectrogram (in contrast the note onsets are smeared by the long duration spectrogram).

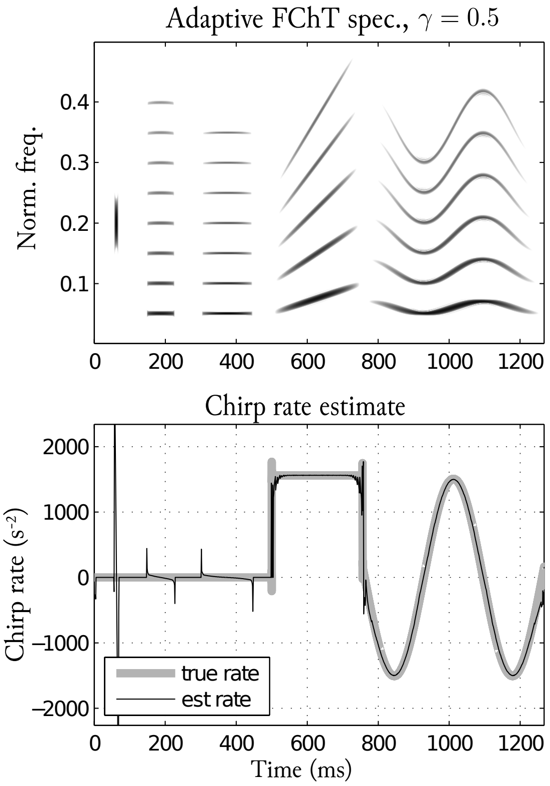

Figure 6: Analysis of a the synthetic signal from Fig. 4 using time-varying windows and fan-chirp atoms. Top row: time-varying spectrogram. Bottom row: estimated chirp rate (using the DDM).

Finally, we present an example of a real signal analyzed with time-varying windows and the FChT. Figure 8 illustrates the analysis of a male vocal excerpt. This figure shows a fixed window spectrogram, a time-varying spectrogram, and a time-varying FChT. In this example, the time-varying FChT seems to more clearly concentrate the signal information (especially the fast glissando and other vocal modulations). For example, consider the region between 2000 and 2500ms: the upper harmonics of the vibrato are most clearly resolved when using the FChT. A similar observation can be made regarding the upper harmonics of the glissando around 500ms.

7. CONCLUSIONS

In this paper we presented a new adaptive analysis system bringing together several techniques which have been discussed separately in the literature. First, we discussed how to perform analysis/synthesis with time-varying windows. A key point of our formulation was to present the time-varying analysis/synthesis setup such that any invertible transform could be used for the analysis (e.g., Fourier transform, fan-chirp transform, and so on). We then discussed how to adapt the set of analysis windows in order to maximally concentrate the spectral energy (i.e., using the Renyi entropy). Finally, we showed how to estimate the chirp parameter and then efficiently perform analysis/synthesis with time-varying chirped atoms using the FChT. We ended with several real and synthetic examples which showed qualitative differences between

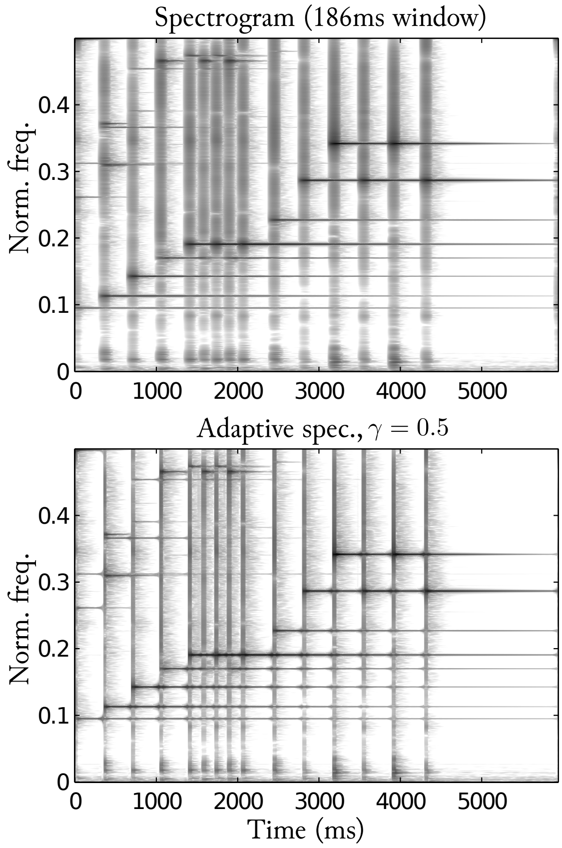

Figure 7: Analysis of glockenspiel with time-varying windows. (The sampling frequency was 11025 Hz).

the different analysis methods proposed (e.g., fixed window, time-varying window, and time-varying window with adapted chirp rate). We may conclude from this investigation that the time-varying FChT is indeed quite useful for the analysis of audio since we are able to a) adapt the window length to track the evolving structure of the audio signal; and b) adapt the chirp rate to model non-stationary and transient phenomena that evolve over the duration of the analysis window.

8. ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Centre for Interdisciplinary Research in Music Media and Technology (CIR-MMT).

9. REFERENCES

[1] J.O. Smith, *Spectral audio signal processing*, Stanford University, CCRMA, 2008.

[2] D. Gabor, “Theory of communication,” *Journal of the Institute of Electrical Engineers*, vol. 93, no. 4, pp. 429–457, 1946.

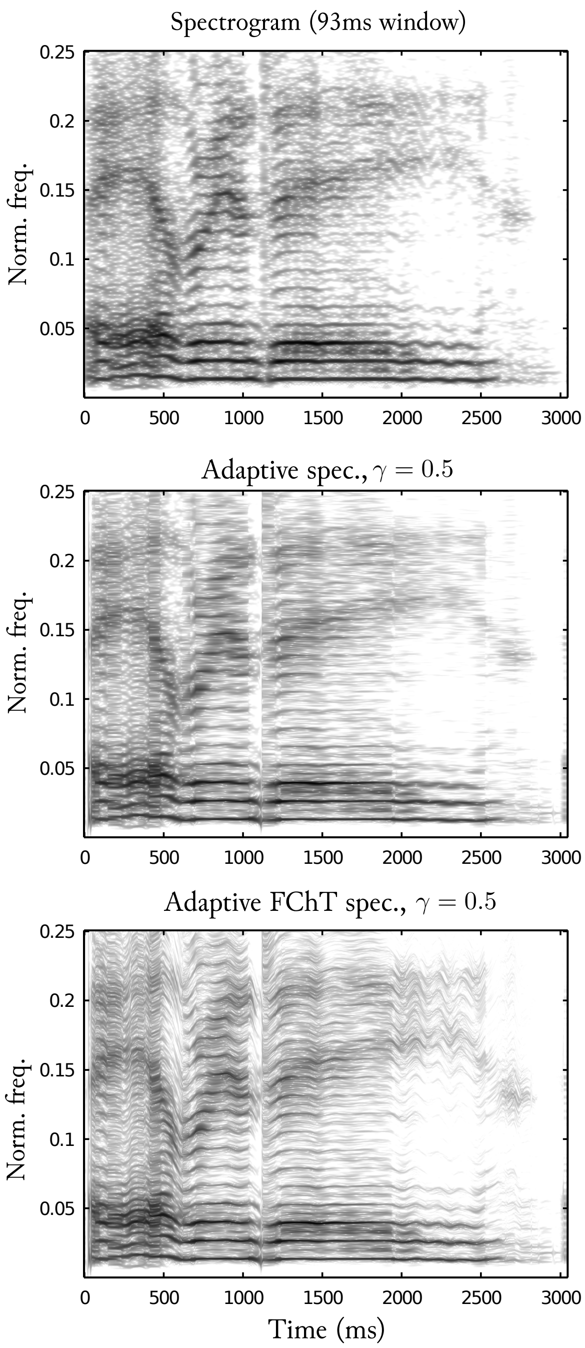

Figure 8: Analysis of male vocal with time-varying windows and the FChT. Top: spectrogram, Middle: time-varying spectrogram, Bottom: time-varying FChT. (The sampling frequency was 11025Hz).

[3] D. Rudoy, P. Basu, and P.J. Wolfe, "Superposition frames for adaptive time-frequency analysis and fast reconstruction," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2581–2596, 2010.

[4] F. Jaillet, P. Balazs, and M. Dörfler, "Nonstationary Gabor Frames," in *Proceedings of Sampling Theory and Applications (SAMPTA)*, 2009.

[5] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. Velasco, "Theory, implementation and applications of nonstationary Gabor frames," *Journal of computational and applied mathematics*, vol. 236, no. 6, pp. 1481—1496, 2011.

[6] M. Liuni, A. Roebel, E. Matusiak, M. Romito, and X. Rodet, "Automatic adaptation of the time-frequency resolution for sound analysis and re-synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 959–970, 2013.

[7] S. Mann and S. Haykin, "The chirplet transform: Physical considerations," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2745–2761, 1995.

[8] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 235–238, 1977.

[9] D.L. Jones and R.G. Baraniuk, "A simple scheme for adapting time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3530–3535, 1994.

[10] F. Jaillet and B. Torrésani, "Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered Gabor expansions," *International Journal of Wavelets Multiresolution and Information Processing*, vol. 5, no. 2, pp. 293–315, 2007.

[11] C. Kereliuk, *Sparse and structured atomic modelling of audio*, Ph.D. thesis, McGill University, defended on March 28th, 2013.

[12] M. Liuni, A. Roebel, M. Romito, and X. Rodet, "An entropy based method for local time-adaptation of the spectrogram," in *Lecture Notes in Computer Science: Computer Music Modeling and Retrieval. Exploring Music Contents*, S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds., pp. 60–75. Springer Verlag, Berlin, Germany, 2011.

[13] L. Weruaga and M. Képesi, "The fan-chirp transform for non-stationary harmonic signals," *Signal Processing*, vol. 87, no. 6, pp. 1504–1522, 2007.

[14] R.G. Baraniuk and D.L. Jones, "Unitary equivalence: A new twist on signal processing," *Signal Processing, IEEE Transactions on*, vol. 43, no. 10, pp. 2269–2282, 1995.

[15] J.O. Smith, *Digital Audio Resampling Home Page*, http://www-ccrma.stanford.edu/˜jos/resample/, Last accessed on October 1, 2012.

[16] P. Cancela, E. López, and M. Rocamora, "Fan chirp transform for music representation," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2010, pp. 54–61.

[17] M. Abe and J.O. Smith III, "AM/FM rate estimation for time-varying sinusoidal modeling," in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2005, vol. 3, pp. 201–204.

[18] A. Röbel, "Estimating partial frequency and frequency slope using reassignment operators," in *Proceedings of the International Computer Music Conference (ICMC)*, 2002, pp. 122–125.

[19] S. Marchand and P. Depalle, "Generalization of the derivative analysis method to non-stationary sinusoidal modeling," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2008.

[20] M. Betser, "Sinusoidal polynomial parameter estimation using the distribution derivative," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4633–4645, 2009.

[21] B. Hamilton and Ph. Depalle, "Comparisons of parameter estimation methods for an exponential polynomial sound signal model," in *Audio Engineering Society, Applications of Time-Frequency Processing in Audio*, 2012.

[22] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 591–598, 1974.