A REDUCED MULTIPLE GABOR FRAME FOR LOCAL TIME ADAPTATION OF THE SPECTROGRAM

Marco Liuni, [*]

Università di Firenze, Dip. di Matematica U. Dini,
Viale Morgagni, 67/a - 50134 Florence - ITALY
IRCAM - CNRS STMS, Analysis/Synthesis Team,
1, place Igor-Stravinsky - 75004 Paris - FRANCE
`marco.liuni@ircam.fr`

Axel Röebel,

IRCAM - CNRS STMS, Analysis/Synthesis Team,
1, place Igor-Stravinsky - 75004 Paris - FRANCE
`axel.roebel@ircam.fr`

Marco Romito,

Università di Firenze, Dip. di Matematica U. Dini,
Viale Morgagni, 67/a - 50134 Florence - ITALY
`marco.romito@math.unifi.it`

Xavier Rodet,

IRCAM - CNRS STMS, Analysis/Synthesis Team,
1, place Igor-Stravinsky - 75004 Paris - FRANCE
`xavier.rodet@ircam.fr`

ABSTRACT

In this paper we propose a method for automatic local time adaptation of the spectrogram of an audio signal, based on its decomposition within a Gabor multi-frame. The sparsity of the analyses within each individual frame is evaluated through the Rényi entropies measures. According to the sparsity of the decompositions, an optimal resolution and a reduced multi-frame are determined, defining an adapted spectrogram with variable resolution and hop size.

The composition of such a reduced multi-frame allows an immediate definition of a dual frame: re-synthesis techniques for this adapted analysis are easily derived by the traditional phase vocoder scheme.

1. INTRODUCTION

The quality of analysis and synthesis processes based on time-frequency transforms is highly affected by the frames used for the decomposition and the reconstruction of the signal. Traditional methods based on single frames of atomic functions have important limits: a Gabor frame imposes a fixed resolution over all the time-frequency plane, while a wavelet frame gives a strictly determined variation of the resolution: moreover, the user is frequently asked to define himself the analysis window features, which is not always a simple task even for normally experienced users.

The resolution of such analysis methods is linked to the time and frequency concentration of the basic functions involved in the decomposition. Frame Theory ([1],[2],[3]) extends the concept of orthonormal basis in a Hilbert space H: in our domain, it gives a unified model for the description of decomposing systems based on atomic functions. The set $\{\phi_\gamma\}_{\gamma \in \Gamma}$ is a *frame* for H if there exist two positive non zero constants A and B, called *frame bounds*, such that for all $f \in H$,

$$A\|f\|^2 \leq \sum_{\gamma \in \Gamma} |\langle f, \phi_\gamma \rangle|^2 \leq B\|f\|^2 . \qquad (1)$$

The time-frequency concentration of an atom ϕ_γ in a frame can be represented through its associated Heisenberg box: it is a rectangle drawn in the time-frequency plane whose dimensions are linked respectively to the time spread of a function and to the frequency spread of its Fourier Transform. In the Short Time Fourier Transform, the boxes associated to the transpositions of the window function g have fixed dimensions in every area of the time-frequency plane: the resolution is the same for all the components of the signal. In the Wavelet Transform, lower frequency components are represented with a higher time resolution, while a higher frequency resolution is given for the higher frequency ones. This limits are not motivated when analyzing a sound without an a priori knowledge of its features, as the best resolution tradeoff is neither unique nor depending only on a single variable. It is therefore useful to search for adaptive methods of sound analysis and synthesis, and for algorithms whose operations are designed to change locally according to the analyzed signal features.

Given $l \in \mathbb{R}^+$, the analysis resolution can be globally modified with a scaling operation

$$g^l(t) = \frac{1}{\sqrt{l}} \, g\left(\frac{t}{l}\right) , \qquad (2)$$

which has the effect of changing the ratio between the edges of the Heisenberg box associated to g while preserving its area: this means that the global time-frequency resolution is modified by privileging concentration in one dimension to the detriment of the other. The idea which has lead to the definition of *multiple Gabor frames* ([4]) is to consider a decomposing system where all these different resolution tradeoffs coexist, providing a more detailed description of the signal. The drawback is the introduction of a high redundancy which lowers the readability of the representation: therefore methods for appropriate reductions of these multiple frames are needed, typically using sparsity criteria.

A promising approach ([5]) takes into account *Rényi entropies*, a generalization of the Shannon entropy: given a unit-energy signal $f \in L^2(\mathbb{R})$ and a time-frequency representation $\Phi_f(u, \xi)$ of f the Rényi entropy of the representation is defined for an *order $\alpha > 0$*

as follows

$$H_\alpha(\Phi_f) = \frac{1}{1-\alpha} \, \log_2 \iint \Phi_f^\alpha(u,\xi) \mathrm{d}u \mathrm{d}\xi \, . \qquad (3)$$

In this paper, the time-frequency representation $\Phi_f(u,\xi)$ considered is the *spectrogram*, as detailed in the next section. The application to our problem is related to the concept that minimizing the complexity or information over a set of time-frequency representations of a same signal is equivalent to maximizing the concentration and peakiness of the analysis, thus selecting the best resolution tradeoff: a sparsity measure can consequently be defined through an information measure. Methods inspired by this approach have shown to give interesting results both analytically and numerically ([6]).

The proposed method of local time adaptation improves on the analysis multi-frame definition: the user can specify a finite arbitrary set of positive scaling factors $L \subset \mathbb{R}^+$ corresponding to the resolutions available; then the algorithm composes different frames $\{g_{n,k}^l\}_{(n,k)\in\mathbb{Z}^2}$ with $l \in L$ and g^l as in (2), and a multiple Gabor frame is obtained as the union of all the given frames. The main improvement in comparison with [7] is that we are not obliged to keep the same hop size within the individual frames analyses, thus avoiding unnecessary short hops for larger windows: our method employs frames which share the same redundancy, so that every analysis has the same overlap, with a significant gain in computational cost.

The limit of our approach in comparison with [6] is that we apply the entropy evaluation on the whole frequency dimension, thus providing analyses which are adapted only in the time dimension. On the other hand, the reduced multi-frame obtained with our method allows a perfect reconstruction of the signal which is not provided by [6]: in our scheme, for any analysis segment a single original frame is retained; therefore, a re-synthesis technique can be defined as a straightforward extension of the least square error estimation from the modified STFT presented in [8]. So our method can easily be used to provide common time-frequency processing frameworks with an adaptive analysis technique.

2. ENTROPY EVALUATION OF A SPECTROGRAM

We will now describe the application of the entropy sparsity measure on the spectrogram distribution. We will focus on discretized spectrograms, as dealing with digital signal processing requires to work with sampled signals and distributions, even if for the most part the results can be extended to the continuous case.

A *Gabor frame* is obtained by time shifting and frequency transposing a window function g according to a regular lattice. Given a time step a and a frequency step b we write $\{u_n\}_{n\in\mathbb{Z}} = an$ and $\{\xi_k\}_{k\in\mathbb{Z}} = bk$; these two sequences generate the nodes of the time-frequency lattice for the frame $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ defined as

$$g_{n,k}(t) = g(t - u_n)e^{2\pi i \xi_k t} \, ; \qquad (4)$$

the nodes are the centers of the Heisenberg boxes associated to the windows in the frame. The decomposition of a function $f \in L^2(\mathbb{R})$ in a Gabor frame is simply a sampling of its STFT according to such a lattice,

$$\mathrm{S}f[n,k] = \langle f, g_{n,k} \rangle = \int f(t)g(t-u_n)e^{-2\pi i \xi_k t}\mathrm{d}t \, , \qquad (5)$$

and the squared modulus of this decomposition is the discretized spectrogram,

$$\mathrm{PS}_f[n,k] = |\mathrm{S}f[n,k]|^2 \, . \qquad (6)$$

Given a discrete spectrogram with time step a and frequency step b as in (6), we look for an evaluation of its entropy over a certain rectangle of the time-frequency plane $[t_1,t_2] \times [\nu_1,\nu_2] \subseteq \mathbb{R}^2$. The rectangle identifies a sequence of points $G \subseteq \mathbb{Z}^2$ where $G = \{(n,k) \in \mathbb{Z}^2 : t_1 \le na \le t_2, \ \nu_1 \le kb \le \nu_2\}$. Through an appropriate normalization we obtain the sequence

$$\mathrm{PS}_f^G[n,k] = \frac{\mathrm{PS}_f[n,k]}{\sum_{[n',k']\in G} \mathrm{PS}_f[n',k']} \, , \qquad (7)$$

with $[n,k] \in G$, which can be seen as a discrete probability density. As a discretization of the original continuous spectrogram, every sample in PS_f^G is related to a time-frequency region of area ab; we thus obtain the Rényi entropy measure for (7) directly from (3),

$$H_\alpha(\mathrm{PS}_f^G) = \frac{1}{1-\alpha} \log_2 \sum_{[n,k]\in G} (\mathrm{PS}_f^G[n,k])^\alpha + \log_2(ab) \, . \quad (8)$$

General properties of Rényi entropies can be found in [9], [10] and [11]; we recall in particular those which have a closer relation with our problem. It is easy to show that for every finite discrete probability density P the entropy $H_\alpha(P)$ tends to coincide with the Shannon entropy of P as the order α tends to one. Moreover, $H_\alpha(P)$ is a non increasing function of α, so

$$\alpha_1 < \alpha_2 \Rightarrow H_{\alpha_1}(P) \ge H_{\alpha_2}(P) \, . \qquad (9)$$

As we are working with finite discrete densities we can also consider the case $\alpha = 0$ which is simply the logarithm of the number of elements in P; as a consequence $H_0(P) \ge H_\alpha(P)$ for every admissible order α.

A third basic fact is that for every order α the Rényi entropy H_α is maximum when P is uniformly distributed, while it is minimum and equal to zero when P has a single non-zero value. Given a generic P and its entropy $H_\alpha(P)$ for a certain order α, we have that for any $\beta \ge \alpha$

$$\frac{\alpha-1}{\alpha} H_\alpha(P) \le \frac{\beta-1}{\beta} H_\beta(P) \, . \qquad (10)$$

All of these results give useful informations on the values of different measures on a single density P as in (8), while the relations between the entropies of two different densities P and Q are in general hard to determine analytically; in our problem, P and Q are two spectrograms of a same signal in a same time-frequency area, based on two window functions with different scaling as in (2).

When the spectrogram of a signal does not depend on time it is easier to find such a relation, and it turns out to be the one expected: let PS_s^G be the sampled spectrogram of a sinusoid s over the region G with a window function h of compact support; then PS_s^G is simply a translation in the frequency domain of $\hat{h}$, the Fourier transform of the window, and it is therefore time-independent. We choose a bounded set L of admissible scaling factors, so that the discretized support of the scaled windows h^l still remains inside G for any $l \in L$. It is not hard to prove that the entropy of a spectrogram taken with such a scaled version of h is given by

$$H_\alpha(\mathrm{PS}_{sl}^G) = H_\alpha(\mathrm{PS}_s^G) - \log_2 l \, . \qquad (11)$$

The sparsity measure we are using looks for the window which minimizes the entropy measure: we deduce from (11) that it is the one obtained with the largest scaling factor available, so with the largest time-support. This is coherent with our expectation as stationary signals, such as sinusoids, are best analyzed with a high frequency resolution, because time-independency allows a small time resolution. Moreover, this is true for any order α used for the entropy calculus.

Symmetric considerations apply whenever the spectrogram of a signal does not depend on frequency, as for impulses.

A last remark regards the dependency of (8) on the time and frequency step a and b used for the discretization of the spectrogram. When considering signals as finite vectors, a signal and its Fourier Transform have the same length. Therefore in the STFT the window length determines the number frequency points, while the sampling rate sets frequency values: the definition of b is thus implicit in the window choice. Actually, the FFT algorithm allows to ask a number of frequency points larger than the signal length: further frequency values are obtained as an interpolation between the original ones by properly adding zero values to the signal. If the sampling rate is fix, such a procedure establishes smaller b as a consequence of a larger number of frequency points. We have numerically verified that such a variation of b has no impact on the entropy calculus, so that the FFT size can be set according to implementation needs.

Regarding the time step a, we are working on the analytical demonstration of a largely verified evidence: as long as the decomposing system is a frame the entropy measure is invariant to redundancy variation, so the choice of a can be ruled by considerations on the invertibility of the decomposing frame without losing coherence between the information measure of the different analyses. This is a key point, as it states that the sparsity measure obtained allows a total independence between the hop sizes of the different analyses: with the implementation of proper structures to handle multi-hop STFTs we have obtained a more efficient algorithm in comparison with the ones imposing a fixed hop size, as [7].

3. ALGORITHM AND EXAMPLES

We now summarize the main operations of the algorithm we have developed providing examples of its application. For spectrograms calculation we have used a *Hanning window*

$$h(t) = \cos^2(\pi t)\chi_{[-\frac{1}{2}, \frac{1}{2}]} , \qquad (12)$$

with χ the indicator function of the specified interval, but it is obviously possible to generalize the results thus obtained to the entire class of compactly supported window functions. We create a multiple Gabor frame as in (4) using as mother functions some scaled version of h, obtained as in (2) with a finite set of positive real scaling factors L.

Different spectrograms of segments of the signal are calculated with each one of the above frames: the length of the analysis segment and the overlap between two consecutive segments are given as parameters.

The different frames composing the multi-frame have the same frequency step b but different time steps $\{a_l : l \in L\}$: the smallest and largest window sizes are given as parameters together with $|L|$, the number of different windows composing the multi-frame,

and the global overlap needed for the analyses. The algorithm fixes the intermediate sizes so that for each signal segment the different frames have the same overlap between consecutive windows, and so the same redundancy. This generates an irregular time disposition of the multi-frame elements in each signal segment, as illustrated in figure 1. Such a disposition causes a different influence of the boundary parts of the signal on the different frames analyses: the beginning and the end of the signal segment have a higher energy when windowed in the smaller frames. This is avoided with a preliminary weighting: the beginning and the end of each signal segment are windowed respectively with the first and second half of the largest analysis window. Such a weighting does not concern the decomposition for re-synthesis purpose, but only the analyses used for entropy evaluations. For each signal segment we
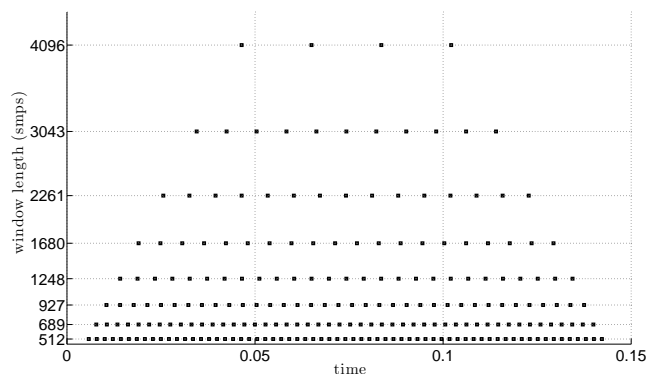

Figure 1: *An analysis segment: time locations of the Heisenberg boxes associated to the multi-frame used in our algorithm.*

calculate the entropy of every spectrogram as in (8), where G is the rectangle with the time segment analyzed as horizontal dimension and the whole frequency lattice as vertical. The sparsest local analysis is defined to be the one with minimum Rényi entropy: the best window is thus defined consequently. Adaptation is obtained over the time dimension as for every signal segment the selected analysis involve the whole frequency dimension. An interpolation is performed over the overlapping zones to avoid abrupt discontinuities in the tradeoff of the resolutions.

The time adapted analysis of the global signal is finally realized by opportunely assembling the slices of local sparsest analyses obtained with the selected windows.

In figure 3 we give a first example of an adaptive analysis performed by our algorithm with eight Hanning windows of different sizes on a real instrumental sound, a B4 note played by a marimba: this sound combines the need for a good time resolution at the moment of the percussion, with that of a good frequency resolution on the harmonic resonance of the instrument. This is fully provided by the algorithm, as shown in the adapted spectrogram at the bottom of figure 3. Moreover, we see that the pre-echo of the analysis at the bottom of figure 2 is completely removed in the adapted spectrogram.

In figure 5 we give a second example with a synthetic sound, a sinusoid with sinusoidal frequency modulation: as figure 4 shows, a small window is best adapted where the frequency variation is
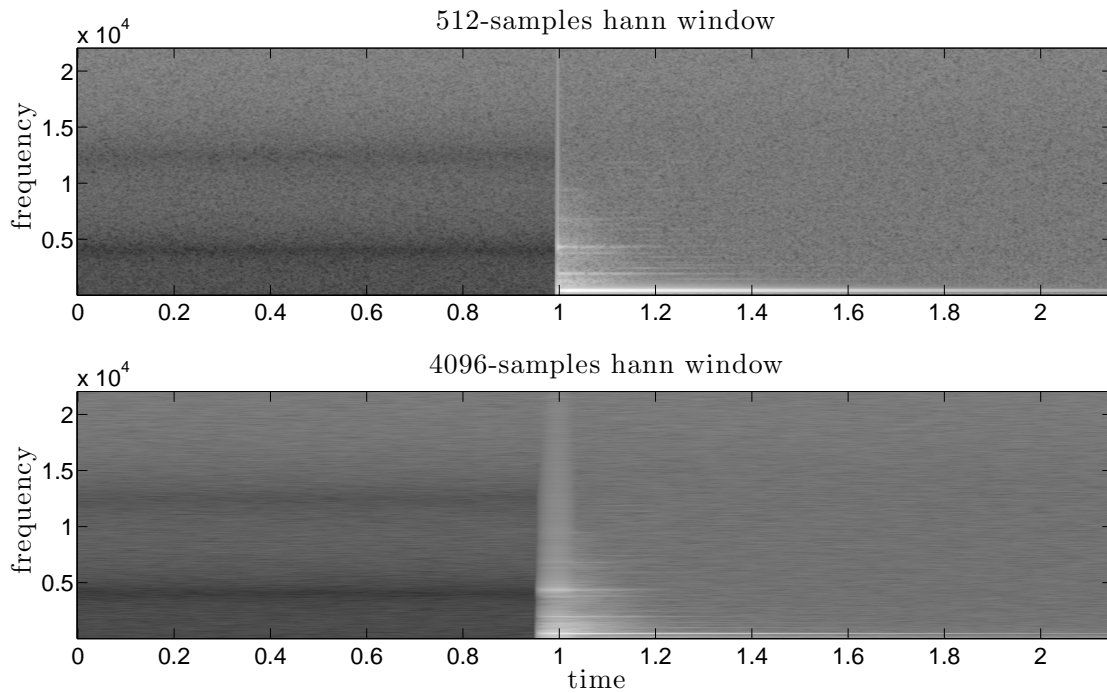
Figure 2: *Two different spectrograms of a B4 note played by a marimba, with Hanning windows of sizes 512 (top) and 4096 (bottom) samples.*

fast compared to the window length; on the other hand, the largest window is better where the signal is almost stationary.

The re-synthesis method introduced in [8] gives a perfect reconstruction of the signal as a weighted expansion of the coefficients of its STFT in the original analysis frame. Let $S_f[n,k]$ be the STFT of a signal f with window function h and time step a; fixing n, through an iFFT we have a windowed segment of f

$$f_h(n,l) = h(na - l)f(l) ,\qquad (13)$$

whose time location depends on n. An immediate perfect reconstruction of f is given by

$$f(l) = \frac{\sum_{n=-\infty}^{+\infty} h(na - l)f_h(n,l)}{\sum_{n=-\infty}^{+\infty} h^2(na - l)} . \qquad (14)$$

We extend the same technique using a variable window h and time step a according to the composition of the reduced multi-frame, obtaining a perfect reconstruction as well. The interest of (14) is that the given distribution needs not to be the STFT of a signal: for example, a transformation $S^*[n,k]$ of the STFT of a signal could be considered. In this case, (14) gives the signal whose STFT has minimal least squares error with $S^*[n,k]$.

The theoretical existence and the mathematical definition of the canonical dual frame for reduced multi-frames like the one we employ has recently been provided in [12]: the analysis and re-synthesis framework is thus entirely defined within the Gabor theory, but no automatic adaptation is employed. We are at present working on the interesting analogies between the two approaches, to establish a unified interpretation and develop further extensions.

4. CONCLUSIONS

We have presented an algorithm for time-adaptation of the spectrogram resolution, which can be easily integrated in existent framework for analysis, transformation and re-synthesis of an audio signal: the adaptation is locally obtained through an entropy minimization within a finite set of resolutions, which can be defined by the user or left as default. The user can also specify the time duration and overlap of the analysis segments where entropy minimization is performed, to privilege more or less discontinuous adapted analyses.

Future improvements of this method will concern the spectrogram adaptation in both time and frequency dimensions: this will provide a decomposition of the signal in several layers of analysis frames, thus requiring an extension of the proposed technique for re-synthesis.

5. REFERENCES

[1] K. Gröchenig, Ed., *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, Massachusetts, USA, 2001.

[2] O. Christensen, Ed., *An Introduction To Frames And Riesz Bases*, Birkhäuser, Boston, Massachusetts, USA, 2003.

[3] S. Mallat, Ed., *A wavelet tour on signal processing*, Academic Press, San Diego, California, USA, 1999.

[4] M. Dörfler, *Gabor Analysis for a Class of Signals called Music*, Ph.D. thesis, Institut für Mathematik der Universität Wien, 2002.
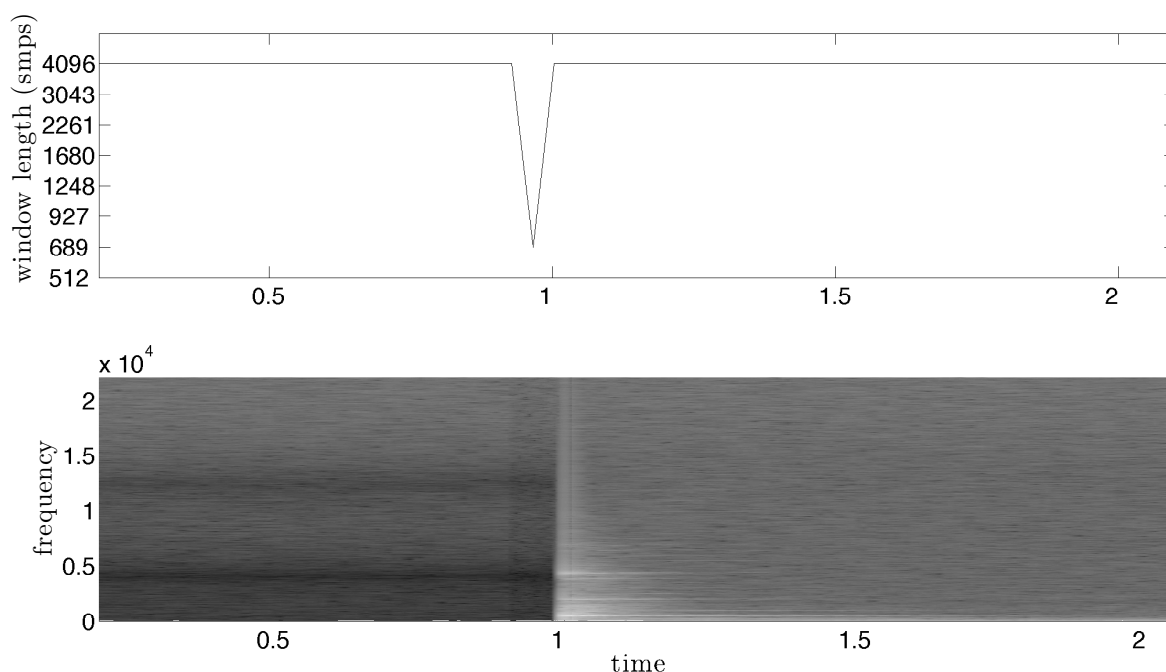
Figure 3: *Example of an adaptive analysis performed by our algorithm with eight Hanning windows of different sizes from 512 to 4096 samples, on a B4 note played by a marimba sampled at 44.1kHz: on top, the best window chosen as a function of time; at the bottom, the adaptive spectrogram. The entropy order is $\alpha = 0.7$ and each analysis segment contains four frames of the largest window analysis with a two-frames overlap between consequent segments.*

[5] R.G. Baraniuk P. Flandrin A.J.E.M. Janssen O.J.J. Michel, "Measuring Time-Frequency Information Content Using the Rényi Entropies," *IEEE Trans. Info. Theory*, vol. 47, no. 4, pp. 1391–1409, May 2001.

[6] F. Jaillet, *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*, Ph.D. thesis, Université de la Méditerranée - Aix-Marseille II, 2005.

[7] J.Todd A.Lukin, "Adaptive Time-Frequency Resolution for Analysis and Processing of Audio," 120th Audio Engineering Society Convention, Paris, France, May 2006. http://graphics.cs.msu.ru/en/publications/text/LukinTodd.pdf.

[8] D.W. Griffin J.S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 236–242, Apr. 1984.

[9] A. Rényi, "On Measures of Entropy and Information," in *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, Berkeley, California, June 20-30, 1961, pp. 547–561.

[10] F. Schlögl C. Beck, Ed., *Thermodynamics of chaotic systems*, Cambridge University Press, Cambridge, Massachusetts, USA, 1993.

[11] K. Zyczkowski, "Rényi Extrapolation of Shannon Entropy," *Open Systems & Information Dynamics*, vol. 10, no. 3, pp. 297–310, Sept. 2003.

[12] F. Jaillet P. Balazs M. Dörfler N. Engelputzeder, "Nonstationary Gabor Frames," in *Proc. of SAMPTA'09*, Marseille, France, May 18-22, 2009.
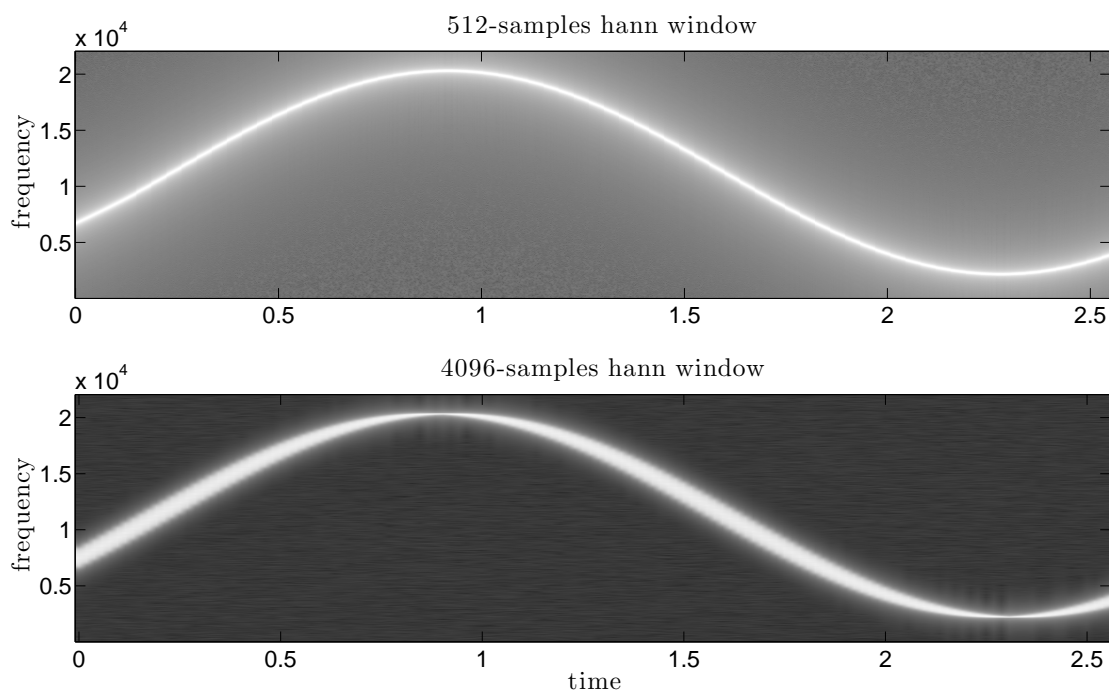
Figure 4: *Two different spectrograms of a sinusoid with sinusoidal frequency modulation, with Hanning windows of sizes 512 (top) and 4096 (bottom) samples.*
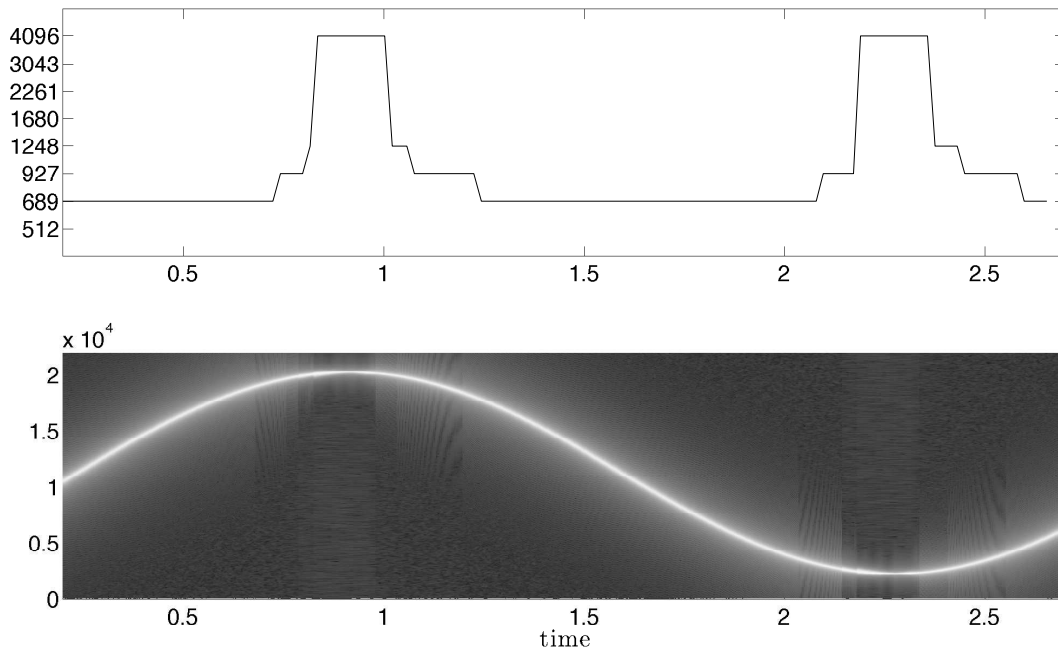

Figure 5: *Example of an adaptive analysis performed by our algorithm with eight Hanning windows of different sizes from 512 to 4096 samples, on a sinusoid with sinusoidal frequency modulation synthesized at 44.1 kHz: on top, the best window chosen as a function of time; at the bottom, the adaptive spectrogram. The entropy order is $\alpha = 0.7$ and each analysis segment contains four frames of the largest window analysis with a three-frames overlap between consequent segments.*