A novel basis for analysis of music

Ana Ruedin and Juan Vuletich

Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires.
ana.ruedin@dc.uba.ar

***Abstract*— We present a new basis for analysis of music. The basis vectors are sampled real waveforms of fixed frequency inside a Gaussian envelope. Their frequencies and time localizations are induced by a tiling of the time-frequency plane well adapted to music. Through a careful investigation of their properties they are subsequently slightly modified in order to give a stable system, without losing their time-frequency localization. Our new basis discriminates semitones, detects the overtones as well as the attack of notes, and gives a sparse representation of the signal. It will enhance the performance of all kinds of digital audio processors, and provide a useful tool for numerous multimedia applications.**

***Keywords*— time–frequency, music, tiling, Gabor, wavelets.**

1. INTRODUCTION

Having a representation system for music that gives information both on the time and the frequency of the notes that are played, will give an optimal representation of the signal. The search of bases for joint time-frequency representation of music is still an open problem. See for instance, the comparison of methods for onset detection in [1].

Transforms related to Fourier theory, such as FFT, DCT and MDCT, the latter being used in MPEG audio compressors[7], divide the spectrum of a signal into regular frequency intervals that do not match our auditory system, since our perception of pitch is logarithmic.

Gabor systems have long been used. Placed on a uniform time-frequency grid, they either give an unstable system or a redundant representation of the signal. Different approaches have been made to modify the grid [11], [5].

The construction of bases that discriminate musical notes, requires that their spectra be mainly supported on frequency intervals that have a constant relative bandwidth, and this constant is irrational. In order to have good resolution in time, the support of these bases must vary with the frequency.

Discrete dyadic wavelets [3] do have a constant relative bandwidth, and they provide good time-localization, but they cannot discriminate the notes in an octave.

These mentioned transforms do not offer an optimal represention of digitized music signals that can be posteriorly used for digital audio applications, such as pitch detection or melody identification.

We present a new basis for the representation of digitized music. It is induced by a special tiling of the time-frequency plane well adapted to music. Our tiling is less general than the tiling proposed in [2] for the construction of orthonormal bases, but is specifically built for music representation, and our basis has better frequency localization.

Essentially our basis is composed of shifts in time and frequency of one fundamental wavelet, which is a modification of a Morlet wavelet [8], having improved frequential resolution. Although the latter is known to be unstable on a uniform tiling of the time-frequency plane, we modify both the tiling and the wavelet in order to obtain stability.

In Sections 2-4 we develop the construction of our basis. This was outlined in the first author's earlier work [9] and grade thesis [10]. In Section 2 we define an ideal tiling for music analysis, which is irrational. In section 3 we construct our first basis. Subsequent modifications of the tiling (taking a rational approximation), and of the wavelet system (taking projections to reduce correlations between basis vectors) will give a stable basis (Section 4). In Section 5 we present a simple, illustrative test, and our concluding remarks.

2. An Ideal Tiling of the Time-Frequency Plane

The pitch or frequency of a note is measured in cycles per second, or Hertz (Hz); for example the note A_4 (La central) has 440 Hz. Our perception of pitch is logarithmic: to our ear, the three notes 220 Hz, 440 Hz and 880 Hz sound equally spaced apart, yet their frequencies are related by a multiplicative factor. When the frequencies of two notes differ by a factor of 2, as in this case, we have an octave. In the twelve-tone equal-tempered scale there are 12 notes or semitones per octave, and the frequency ratio of 2 adjacent notes is constant: we call this constant a_0. Take for example A_4 as the first note of an octave: then the second note, a semitone above A_4, will have frequency 440 a_0 Hz, the third 440 a_0^2 Hz, and so on. The thirteenth note, an octave above A_4, will have frequency 440 $a_0^{12} = 440 \times 2$ Hz. This means that $a_0 = \sqrt[12]{2}$, an irrational number.

Let f_1^c be the first note of an octave. Then the other notes can be obtained from

$$f_{j+1}^c = a_0 \; f_j^c \quad \text{for} \quad j = 1 \ldots 11. \tag{1}$$

In table 1 we list the notes and frequencies of an octave beginning with C_3 (Do).

Table 1: Notes and their frequency

Notes			Frequency [Hz]
f_1^c	*Do*	C_3	$440 \times 2^{-9/12}$
f_2^c	*Do♯*	$C_3♯$	$440 \times 2^{-8/12}$
f_3^c	*Re*	D_3	$440 \times 2^{-7/12}$
f_4^c	*Re♯*	$D_3♯$	$440 \times 2^{-6/12}$
f_5^c	*Mi*	E_3	$440 \times 2^{-5/12}$
f_6^c	*Fa*	F_3	$440 \times 2^{-4/12}$
f_7^c	*Fa♯*	$F_3♯$	$440 \times 2^{-3/12}$
f_8^c	*Sol*	G_3	$440 \times 2^{-2/12}$
f_9^c	*Sol♯*	$G_3♯$	$440 \times 2^{-1/12}$
f_{10}^c	*La*	A_4	440
f_{11}^c	*La♯*	$A_4♯$	$440 \times 2^{1/12}$
f_{12}^c	*Si*	B_4	$440 \times 2^{2/12}$
f_{13}^c	*Do*	C_4	$440 \times 2^{3/12}$
⋮	⋮	⋮	⋮

For an original signal sampled at a rate of F_s samples per second, our present work will address a limitted range of frequencies.

We divide the frequency range into frequency intervals or bands $[f_j, f_{j+1}]$, each having a semitone f_j^c as central frequency.

We next make a tiling of the time–frequency plane. The centers of the tiles are the points $(t_{j,\,k}^c, f_j^c)$, where

$$t_{j,\,k}^c = \left(\frac{q}{f_0^c}\right) a_0^{-j} \left(k + \frac{1}{2}\right), \tag{2}$$

$f_0^c = f_1^c / a_0$, and $q = (a_0 + 1)/[4(a_0 - 1)]$.

In Fig. 1 we have an ideal tiling for an octave.

We briefly give the formulae to calculate the borders of a tile (see Fig. 2). First set $f_j = 2\, f_j^c/(1 + a_0)$. Let $\Delta f_j = f_{j+1} - f_j$ be the bandwidth.

Call $[\, t_{j,k}, t_{j,k+1}]$ the interval in time for the same tile, and let

$\Delta t_j = t_{j,k+1} - t_{j,k}$. All tiles have the same area: we have chosen $\Delta f_j \;\; \Delta t_j = 0.5$; we use the latter to obtain Δt_j. Now set $t_{j,k} = k \;\Delta t_j$.

With a little calculation, it can be shown that (i) $f_{j+1} = a_0 \; f_j$,
(ii) the relative bandwidth $\Delta f_j / f_j = a_0 - 1$ is constant,

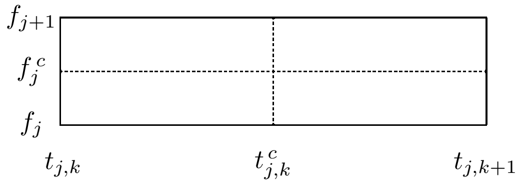

Figure 2: A tile (j, k)

(iii) $f_j^c / \Delta f_j = (a_0 + 1)/[2(a_0 - 1)]$,
(iv) f_j^c is the midpoint of $[f_j, f_{j+1}]$, and (v) $t_{j,\,k}^c$ is the midpoint of $[\, t_{j,k}, t_{j,k+1}]$.

3. Construction of the Basis

Once the tiling is constructed, we need a basis that is well localized over the tiles. We have chosen a real Gabor (or Morlet) wavelet,

$$\Psi(t) = 2\, b\, \sqrt{\pi}\; e^{-(b\;\pi\;t)^2} \cos\left(2\;\pi f\;\, t\right). \tag{3}$$

a waveform of fixed frequency inside a Gaussian envelope, whose Fourier transform has fast decay on neighbouring frequencies. It has often been used for music analysis [6] because it reaches the theoretical limit to time and frequency localization specified by Heisenberg's uncertainty principle.

Appropriate dilations and displacements of the wavelet will allow us to place it over any tile of the partition, to obtain all the elements of the basis. However, it is impossible to confine a wavelet strictly to a tile, because a function cannot be compactly supported both in time and frequency domain. We aim at having most of the energy of the wavelet concentrated on a tile, and have good decay on the neighbouring tiles.

To construct our basis, we place a wavelet on the center of each tile. For each tile (j, k) we have the corresponding wavelet $\Psi_{j,\,k}(t)$, having central frequency $f = f_j^c$, and centered at time $t = t_{j,\,k}^c$:

$$\Psi_{j,k}(t) = 2\, b_j\, \sqrt{\pi}\, e^{-(b_j\,\pi\,(t - t_{j,\,k}^c))^2} \cos\left(2\,\pi f_j^c\,(t - t_{j,k}^c)\right). \tag{4}$$

The spectrum (absolute value of the Fourier transform) of wavelet $\Psi_{j,k}$ is the sum of two Gaussian functions, one centered at $u = f_j^c$ (which is of interest) and the other centered at $u = -f_j^c$ (of no interest):

$$\left|\widehat{\Psi}_{j,\,k}(u)\right| = e^{-\left((u - f_j^c)/b_j\right)^2} + e^{-\left((u + f_j^c)/b_j\right)^2}, \tag{5}$$

where $\widehat{g}(u) = \int_{-\infty}^{\infty} e^{-\,i\,2\,\pi\,u\,t}\; g(t)\,dt$ is the Fourier transform of g.

Notice that parameter b_j is inversely proportional to the width of the Gaussian in Eq. (4), and proportional to the width of the Gaussian in Eq. (5). It controls the balance between time and frequency localization.

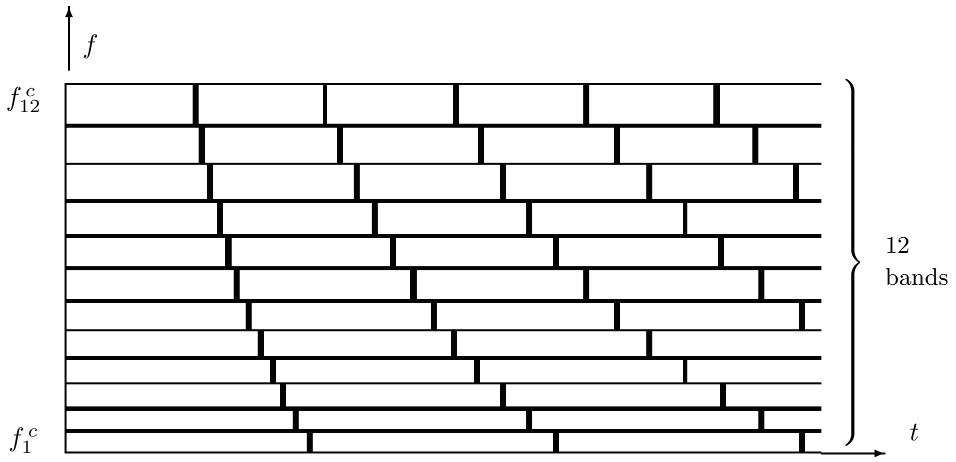

Figure 1: Tiling of the time(t) -frequency(f) plane for an octave.

Set $b_j = K\,\Delta f_j$. To priviledge frequency localization over temporal localization, we have chosen $K = 0.31$.

At the top of Fig. 3 are shown two consecutive shifts of the wavelet. The wavelet $\Psi_{j,\,k}(t)$ has most of its energy concentrated on the time interval $[\ t_{j,k}, t_{j,k+1}]$. At the bottom of Fig. 3 are shown the spectra of two wavelets belonging to neighbouring bands, i.e. having central frequencies f_j^c and f_{j+1}^c corresponding to two consecutive semitones: observe the very small overlap.

The wavelets are sampled over a larger time interval than the actual one, in order to avoid having clipped wavelets, and the signal is padded with zeros at both ends, so that their lengths are equal. The resulting basis vectors are then normalized, and completed to give a basis of the whole space. The vectors that are added to complete the basis correspond to frequencies lying outside the given frequency range.

4. Reduction of Correlation between Basis Vectors

The correlation between 2 basis vectors is equal to the correlation of their FFT's (Parseval). This indicates that basis vectors belonging to different bands will have small correlation, because of the small overlap of their spectra.

We want to reduce the correlation of wavelets in the same band. This will be done in 2 steps. First the correlation of wavelets at odd shifts is radically reduced through a modification of the tiling. Then the correlation of wavelets at even shifts is reduced through a modification of the wavelet itself.

A way to reduce the correlation of wavelets at odd shifts is to make the small oscillations of the wavelet be at a phase of 90° in neighboring displacements. The frequency of these small oscillations is the central frequency f_j^c of the tile. Therefore Δt_j, the width of the tile measured in seconds, must be an integer number of cycles of frequency f_j^c plus $\frac{1}{4}$ or $\frac{3}{4}$ of a cycle, i.e. the number of cycles should be $\frac{L}{2} + \frac{1}{4}$, with integer L. Since there are f_j^c cycles per second, there are $\frac{2L+1}{4}$ cycles in $\frac{2L+1}{4 f_j^c}$ seconds.

This means that $\Delta t_j = \frac{2\,L+1}{4\,f_j^c}$. Recall that $\Delta t_j = \frac{1}{2\,\Delta f_j}$, and substitute Δf_j from Eq. (iii) at the end of section 2. We get

$$\frac{2\,L+1}{4\,f_j^c} = \Delta t_j = \frac{1}{2\,\Delta f_j} = \frac{(a_0+1)}{4\,(a_0-1)\,f_j^c},$$

from which we obtain $a_0 = 1 + \frac{1}{L}$.

At this point, the orientation of this research needed to be changed. From a true musical scale tiling of the plane where the ratio of two consecutive frequencies was equal to an irrational number a_0, the focus was shifted to a tiling of the plane where the ratio of two consecutive frequencies is equal to the best rational approximation $1 + 1/L$ of a_0.

The fractions of the form $1 + 1/L$, closer to $a_0 = \sqrt[12]{2} \approx 1.05946$, are $1 + 1/17 \approx 1.05882$ and $1 + 1/16 \approx 1.0625$. We can therefore approximate the 12 bands of an octave by 10 slightly narrower bands and 2 slightly wider bands : instead of having ${a_0}^{12} = 2$, we have

$$\left(1 + \frac{1}{17}\right)^{10} \left(1 + \frac{1}{16}\right)^{2} \approx 1.9993725.$$

The error is negligible. We modify our tiling accordingly, with one rational approximation for a_0 in 10 bands of an octave, and another for 2 of the bands. With this rational tiling, we have orthogonality between all wavelets at odd shifts in the same band.

We now strive to reduce correlations between basis vectors at even shifts in the same band, indicated in Table 2. In an iterative process, we select the higher correlation between remaining basis vectors at even distances in the same band, say $2n$, and subtract the projection of one of the vectors (multiplied by an carefully choosen factor between 0 and 1) from the other, at distances $2n$ and $-2n$, in order to maintain symmetry of the basis.

In 3 steps of this process, the correlations become lower than 0.01, and we have calculated our modified

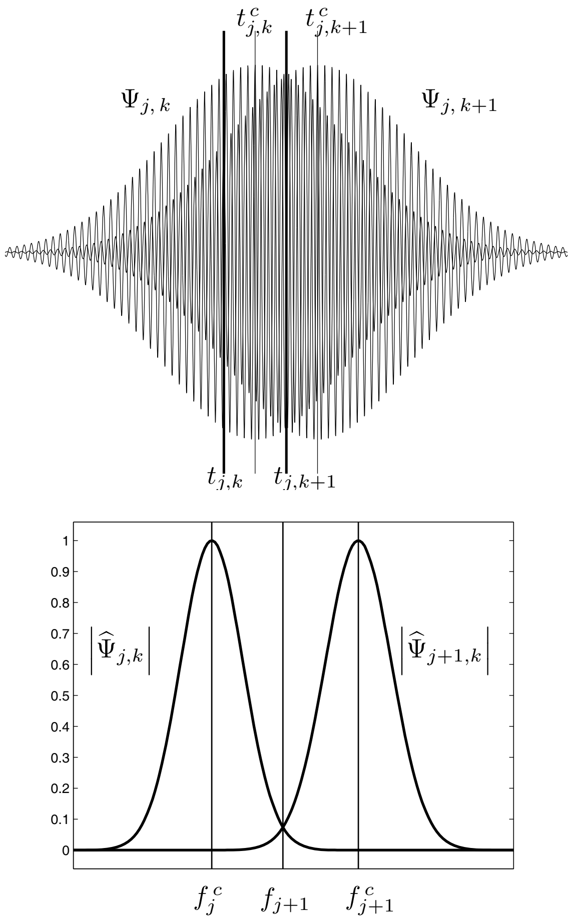

Figure 3: Top: two neighbouring wavelets in time, for the same note. Bottom: spectra of 2 wavelets corresponding to 2 consecutive notes

wavelet Ψ^*. By placing it on each tile of the rational tiling, we now obtain a modified basis. Let A be the matrix whose columns are the basis vectors. If we can find positive constants a, and b, such that for all vectors x,

$$a \|x\|_2 \leq \|A\,x\|_2 \leq b\,\|x\|_2\,, \tag{6}$$

then our basis is stable. It can be proved that the sums of all correlations between one basis vector and all the other basis vectors is less that 0.5, and this ensures that the basis is stable.

The modified wavelet is plotted at the top of Fig. 4. At the bottom of the same figure are the spectra of 2 modified wavelets on consecutive bands, revealing good frequential localization. Our basis looks much like an ideal filter, with the added property that correlations between basis vectors are very low.

Table 2: Correlations between basis vectors at m shifts.

m	Correlation
± 1	0
± 2	-0.62
± 3	0
± 4	0.15
± 5	0
± 6	-0.014
± 7	0
± 8	0.0005
± 9	0
± 10	-0.0000071

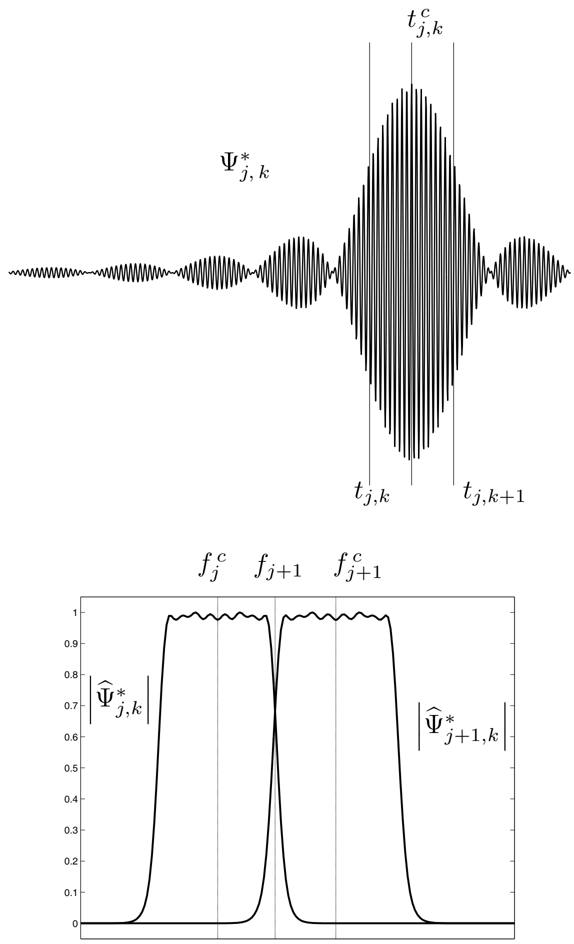

Figure 4: Top: Modified wavelet. Bottom: Spectra of 2 modified wavelets corresponding to 2 consecutive notes

5. Tests and Discussion

Generally Gabor frames are not stable under a perturbation of the tiling constants, that is, even for arbitrarily small changes of the parameters the frame property can be lost. However, Feichtinger et al. [4] proved that this is not the case when the envelope is Gaussian; the fast decay of the Gaussin prevents the system from losing its properties. That is, our rational approximation of an irrational tiling does not destroy the good properties of the original basis.

We tested our algorithm on a bass guitar recording of 7 notes, which belong to an arpeggio in the key of C major going up one octave, and down: (do, mi, sol, do, sol, mi, do) or (C3, E3, G3, C4, G3, E3, C3). The sampling frequency was 5512.5 samples per second, the audio recording is half a second. We constructed a basis covering 3 octaves, and calculated the coefficients of this signal in terms of our basis. In Fig. 5 we show a grayscale map of the coefficients, displayed on the time-frequency tiling, where each coefficient is shown as a shade on its own tile. Darker shades stand for higher absolute values. Frequencies are given on a logarithmic scale, and the total time of the tiling is 0.5 sec.

Our basis discriminates the fundamental notes (and the overtones) perfectly well. In the same figure, we often find, for the same frequency, an alternating succession of large and small coefficients; the reason is that two bases having the same frequency and situated at odd shifts are orthogonal – see Table 2. The attack of notes are also clearly distinguishable. Notice that most coefficients have a light shade: this indicates a sparse representation of the signal, and the suitability of our bases for signal compression, which we will address in future work.

The basis has excellent frequential localization and has good time localization. When applied to a signal, groups of larger coefficients corresponding to the fundamental notes look very much like a music score. This opens a wide scope in music processing: it includes a new model for music signals, helps to understand the music played, and paves the way for various interesting multimedia applications.

REFERENCES

[1] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on speech and audio processing*, 13(5):1035–1047, 2005.

[2] R. Bernardini and J. Kovacevic. Arbitrary tilings of the time-frequency plane using local bases. *IEEE Transactions on Signal Processing*, 47(8):2293–2304, 1999.

[3] I. Daubechies. *Ten lectures on wavelets.* Soc. Indus. Appl. Math., 1992.

[4] Hans Feichtinger and Norbert Kaiblinger. Varying the time-frequency lattice of gabor frames. *Transactions of the American Mathematical Society*, 356(5):2001–2023, 2004.

[5] M. Dörfler. *Gabor Analysis for a Class of Signals called Music.* PhD thesis, Universitat Wien, July 2002.

[6] G. Olmo, F. Dovis, C. Calosso, and P. Passaro. Instrument independent analysis of music by means of the continuous wavelet transform. *Proceedings SPIE Wavelet Appl. Signal Image Proc. VII*, 3813:716–723, 1999.

[7] D. Pan. A tutorial on mpeg/audio compression. *Multimedia, IEEE*, 2:60–74, 1995.

[8] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley Cambridge Press, 1996.

[9] J. Vuletich. Orthonormal bases and tilings of the time-frequency plane for music processing. *Proc. of SPIE, Wavelets X*, 5207:784–793, 2003.

[10] Juan Vuletich. Nuevas bases para el procesamiento de música en el dominio tiempo-frecuencia. *Tesis de Licenciatura, Univ. Buenos Aires, Dirección: A. Ruedin*, Abril 2005.

[11] Patrick Wolfe, Simon Godsill, and Monika Dörfler. Multi-gabor dictionaries for audio time-frequency analysis. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001.

Figure 5: Map of coefficients for arpeggio displayed on the time-frequency tiling. See text.