Bases that Discriminate Semitones for the Representation of Digitized Music

Juan M. Vuletich and Ana M. C. Ruedin

Departamento de Computación, FCEyN, Universidad de Buenos Aires
jmvuletich@sinectis.com.ar, anita@dc.uba.ar

Abstract. We present a new basis for representation of digitized music. The basis vectors are discretizations of a real waveform of fixed frequency inside a Gaussian envelope, and they 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, multimedia.

1 Introduction

Music can be incorporated into varied and numerous multimedia applications. Web pages may offer products with sound effects, music-jingles and video-clips. For online shopping, humming queries may be allowed for music retrieval, before ordering compact disks, music scores or songbooks. Music can be effectively added for successful distance learning, as well as for entertainment.

Having bases for music representation that can give the frequency and the time of the notes, as in a music score, will help to discriminate tones automatically, will ease voice separation and music transcription, and also enhance all kinds of digital audio applications. As a consequence, this property will in turn foster many novel and interesting applications over the web.

The construction of bases that distinguish musical notes, requires that their spectra be mainly supported on frequency intervals that have a constant relative bandwidth, and this constant is $a_0 = 2^{1/12}$, an irrational number. Also the support (in time) of these bases must vary with the frequency, to have good resolution in time: whereas a short time interval is sufficient to identify a high frequency, a longer interval is needed to identify a low frequency.

Different transforms have been employed for dealing with digitized music. The fast Fourier transform (FFT), the discrete cosine transform (DCT) and the modified discrete cosine transform (MDCT) – the latter used in MPEG-1 audio compression, layer III [1]– have poor, of any, time localization (that comes from processing the signal in blocks), and do not have a constant relative

bandwidth: on the contrary, their bandwidths are of fixed size. The discrete dyadic wavelet transform (DWT) [2] [3] does have good time localization and constant relative bandwidth, but its frequential resolution is poor: it is not good enough to discriminate semitones, it can only separate octaves. Wavelet packets [4] improve the frequency resolution, but cannot discriminate semitones.

Continuous transforms such as the Fourier transform, the windowed Fourier transform (as a special case the Gabor Transform), and the continuous wavelet transform (CWT), which are appropriate for signal analysis, generally provide redundant representations of the signal [5], [6].

Because of these drawbacks, several attempts have been made to make more general partitions of the time-frequency plane. In [7] Gabor systems were analyzed on an irregular grid, and in [8] there are constructions of orthogonal bases on arbitrary tilings of the time–frequency plane. Our paper is less general but presents a specific tiling for music processing.

We present a stable basis for the representation of digitized music signals. The basis vectors have excellent localization both in time and frequency. Written in terms of this basis, the signal is represented by means of coefficients that indicate the signals' contribution at a given time and at a given frequency. Our basis will give a new model for music signals, and help to understand them.

Essentially our system is a modification of a real Gabor wavelet, associated to a special tiling of the time–frequency plane, that reflects the properties of our musical scale: it is irrational, and has constant relative bandwidth. Our construction, which comes naturally from the need to break down an octave into semitones, gives tiles that are equal in area.

Our basis is composed of shifts in time and frequency of one fundamental wavelet. Both the tiling and the wavelet are later modified to reduce the inner products between time shifts in the same band.

In sections 2 to 4 we explain the construction of the basis, which was outlined in the first author's earlier work [9] and grade thesis [10]. In section 5 we present some preliminary tests, 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 central A (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 A as the first note of an octave: then the second note, a semitone above A, will have frequency 440 a_0 Hz, the third 440 a_0^2 Hz, and so on. The thirteenth note, an octave above A, 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

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

For an original signal sampled at a rate of F_s samples per second, our present work will address a limitted range of frequencies, the maximum frequency being below $F_s/2$ Hz.

We divide the frequency range (an octave for simplicity) 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 figure (1) we have an ideal tiling for an octave.

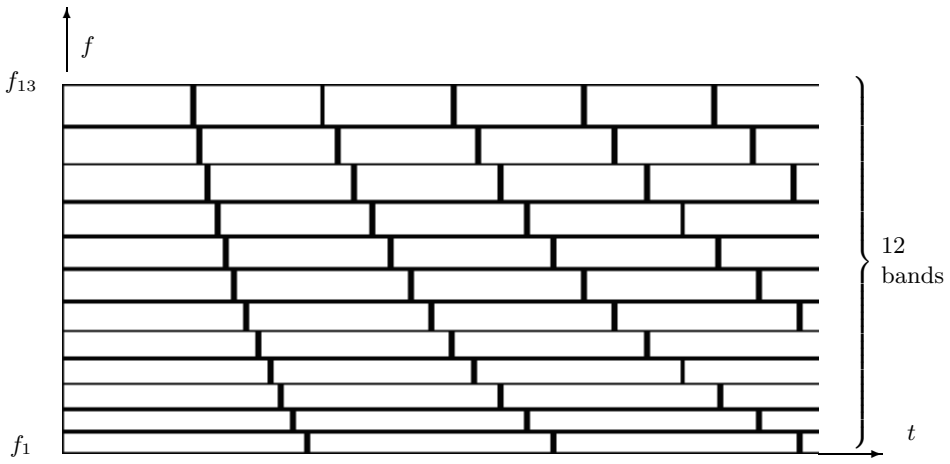

Fig. 1. Tiling of the time(t) -frequency(f) plane for an octave.

We briefly give the formulae to calculate the borders of a tile (see figure (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,
(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 $[\, t_{j,k}, t_{j,k+1}]$.

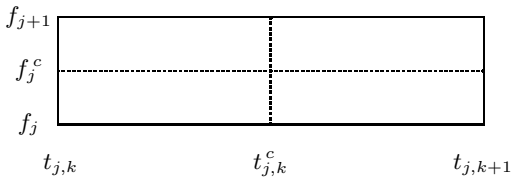

Fig. 2. A tile (j, k)

3 Construction of the Basis

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

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

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 [11] 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 $t = t_{j,\,k}^c$:

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

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

Figure (3)(a) shows two consecutive shifts of the wavelet. The wavelet $\Psi_{j,\,k}(t)$ has most of its energy concentrated on the time interval $[\,t_{j,k}, t_{j,k+1}]$. In Fig.

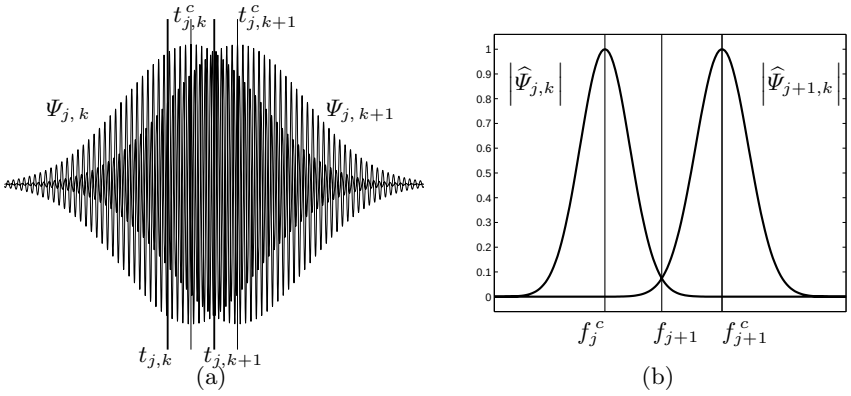

Fig. 3. (a) Two neighbouring wavelets in time, for the same note (b) Spectra of 2 wavelets corresponding to 2 consecutive notes

(3)(b) 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.

m	± 2	± 4	± 6	± 8	± 10
	-0.62	0.15	-0.014	0.0005	-0.0000071

Table 1. Correlations between basis vectors at m shifts.

We now strive to reduce correlations between basis vectors at even shifts in the same band, indicated in Table 1. 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 wavelet Ψ^*. By placing it on each tile of the rational tiling, we now obtain a stable basis.

The modified wavelet is plotted in Fig. (4)(a). In Fig. (4)(b) are the spectra of 2 modified wavelets on consecutive bands: the figure reveals good frequential localization.

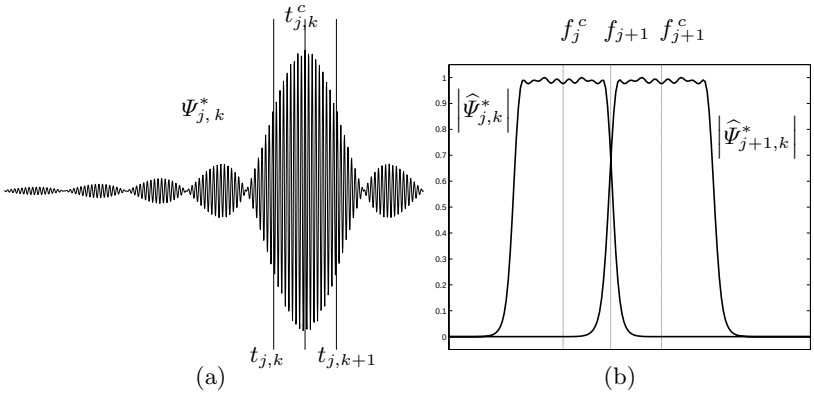

Fig. 4. (a) Modified wavelet (b) Spectra of 2 modified wavelets corresponding to 2 consecutive notes

5 Tests and Conclusion

We tested our algorithm on a short melody of 6 notes played by an electric bass guitar. We constructed a basis covering 3 octaves,and calculated the coefficients of this signal in terms of our basis.

In figure 5 we show the music score of the melody, and a grayscale map of the coefficients, where each coefficient is shown as a shade on its own tile. Darker shades stand for higher absolute values. Ellipses have been drawn by hand, around groups of larger coefficients.

Our basis discriminates the fundamental notes (and the overtones) perfectly well. The attack of notes are 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, the 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.

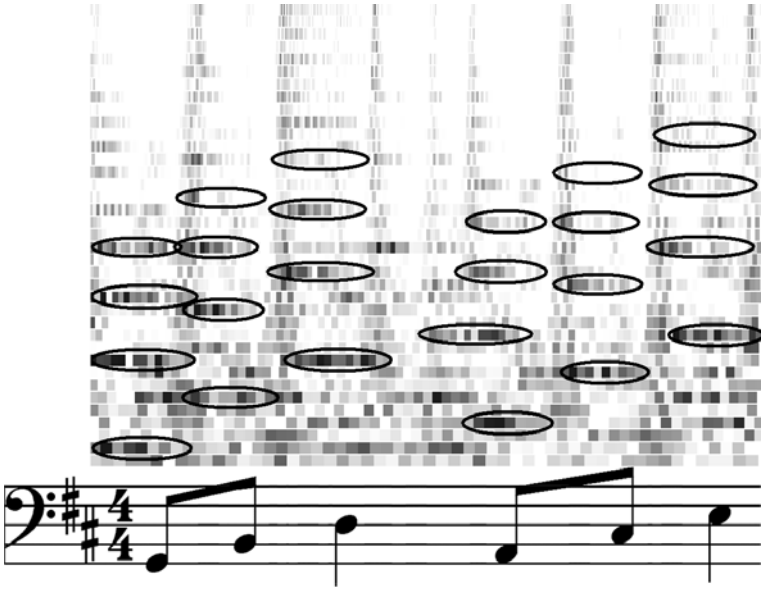

Fig. 5. Bass guitar recording: Map (above) and music score (below). See text

References

1. Pan, D.: A tutorial on mpeg/audio compression. Multimedia, IEEE **2** (1995) 60–74
2. Daubechies, I.: Ten lectures on wavelets. Society for Industrial and Applied Mathematics (1992)
3. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press (1999)
4. Wickerhauser, M.V.: Adapted Wavelet Analysis : FromTheory to Software. A K Peters (1994)
5. Torrésani, B.: An overview of wavelet analysis and time-frequency analysis. Proc. of the Int. Workshop in Self-Similar Systems (Dubna) (1998) 22
6. Benedetto, J., Heil, C., Walnut, D.: Differentiation and the Balian–Low theorem. Journal of Fourier Analysis and Applications **1**(4) (1995) 355–402
7. Feichtinger, H., Gröchenig, K.: Gabor Wavelets and the Heisenberg Group: Gabor Expansions and Short Time Fourier Transform from the Group Theoretical Point of View. C. K. Chui Ed., Academic Press (1992)
8. Bernardini, R., Kovacevic, J.: Arbitrary tilings of the time-frequency plane using local bases. IEEE Transactions on Signal Processing **47**(8) (1999) 2293–2304
9. Vuletich, J.: Orthonormal bases and tilings of the time-frequency plane for music processing. Proc. of SPIE, Wavelets X **5207** (2003) 784–793
10. Vuletich, J.: Nuevas bases para el procesamiento de música en el dominio tiempo-frecuencia. Tesis de Licenciatura, Univ. Buenos Aires, Dirección: A. Ruedin (2005)
11. Olmo, G., Dovis, F., Calosso, C., Passaro, P.: Instrument independent analysis of music by means of the continuous wavelet transform. Proceedings SPIE Wavelet Appl. Signal Image Proc. VII **3813** (1999) 716–723