

# Orthonormal bases and tilings of the time-frequency plane for music processing

Juan M. Vuletich \*

Dept. of Computer Science, University of Buenos Aires, Argentina

## ABSTRACT

Conventional techniques for signal analysis and processing in the time-frequency domain are not well adapted to digital processing of music signals. This restricts the features and quality of applications. We present the current status of a research initiative on this problem. A novel family of wavelet-like bases allows a tiling of the time-frequency plane that is better adapted to digital music signals. This will allow performance enhancements in all kinds of digital audio applications.

**Keywords:** Time-frequency signal representations, Orthonormal bases, Non-dyadic discrete wavelets.

## 1. INTRODUCTION

When processing digital audio signals it is often necessary to control both the time and frequency intervals that are affected. Therefore, finding good bases for joint time-frequency representation of signals is a key problem. To attack this problem, a horizontal slice of the time-frequency plane is considered. This region is divided in small rectangular areas called tiles. Each tile corresponds to an element of the basis.

### 1.1. Tilings of the time-frequency plane and basis elements

The time frequency plane is a two-dimensional Cartesian coordinate system, with time on the abscissas and frequency on the ordinates. The region of the time-frequency plane considered is a horizontal slice, bounded by frequency zero and half the sampling frequency of the signal. This region is divided in small rectangular areas called tiles. These tiles have all the same area and they form a partition of the region. Their area is such that for any time interval the number of tiles needed to cover the corresponding region of the plane equals the number of signal samples in the interval (or half of it, if we choose to represent the transformed signals with complex coefficients). Therefore, the tiling forms a critical sampling of its region of the time-frequency plane.

The time-frequency plane can be divided, for example, in horizontal bands, each covering a different arbitrary frequency interval. Then, each band can be sliced in time independently of the others in rectangles of the same length. For music applications, the tiling of the time-frequency plane should reflect the construction of the musical scale. Therefore, the bands must have a constant relative bandwidth (also said to be "constant Q bands").

The purpose of the tiling is to specify the desired characteristics of each element of the basis. These are time-frequency atoms, because each one's energy is well concentrated both in time and in frequency over its tile. They should also form an orthonormal basis, to ease computation of the transform and reconstruction of signals.

### 1.2. Traditional linear time-frequency representations

Linear time-frequency representations developed before wavelets include the Discrete Fourier Transform, the Discrete Cosine Transform and the Short Time Fourier Transform. Their major drawback when used for music processing is that they can not provide analysis bands with any arbitrary constant relative bandwidth.

### 1.3. Wavelets and their drawbacks

Wavelets are the latest and most successful way to represent signals showing both time and frequency information. The Continuous Wavelet Transform (CWT) exhibits the properties that are needed for the analysis of music, but a discrete version that can provide bases is needed for processing (i.e. analysis / resynthesis). However, Discrete Dyadic Wavelet Transforms (DWT) can not be adjusted to the nature of the tones of the musical scale. Although they generate constant

---

\* [jmvuletich @ sinectis.com.ar](mailto:jmvuletich@sinectis.com.ar); <http://www.sinectis.com.ar/u/jmvuletich>

Q bands, the ratio between the width of adjacent bands (called basic dilation factor or  $a_0$ ) is fixed at 2. There exist ways to build wavelet bases for rational  $a_0$ 's. But the  $a_0$  needed to create the musical scale is  $2^{1/12}$ , an irrational number. This means that the standard tool for building wavelet bases (the multiresolution analysis) must be abandoned.

Regarding the restrictions of discrete dyadic wavelets Bruno Torresani says<sup>1</sup>: "The connection between continuous and discrete wavelet systems is not completely understood. ... The multiresolution approach seems to be also extremely constrained by algebraic arguments, which should be developed further." And Ingrid Daubechies says<sup>2</sup>: "Although the constructive method for orthonormal wavelet bases, called multiresolution analysis, can work only if  $a_0$  is rational, it is an open question whether there exist orthonormal wavelet bases (necessarily not associated with a multiresolution analysis), with good time-frequency localization, and with irrational  $a_0$ ." It is also worth noting that the cover of a key text book<sup>3</sup> shows a piece of sheet music with several notes, as a metaphor of a dyadic wavelet basis. But although each octave has 12 different notes, and 5 octaves are shown, only one note per octave is included (the first one, called C). So, all the notes displayed have a frequency  $2^i f_0$  for some  $f_0$  with integer  $i$ . This is enough to suggest the need of generalizing wavelet bases so we can also represent the other notes of the scale.

As a consequence of these problems, some audio and music applications currently use overcomplete representations, and others use techniques that do not match music signals well. And many can't perfectly reconstruct the original signal.

## 2. CRITERIA

The objective is to find bases that fulfill all the already mentioned requirements. The following are the main ideas kept present during the research.

### 2.1. Signals

We assume that we are processing a finite length signal sampled at a known sampling frequency  $F_s$ . A specific basis will be constructed for each signal.

### 2.2. Desired tiling of the time-frequency plane

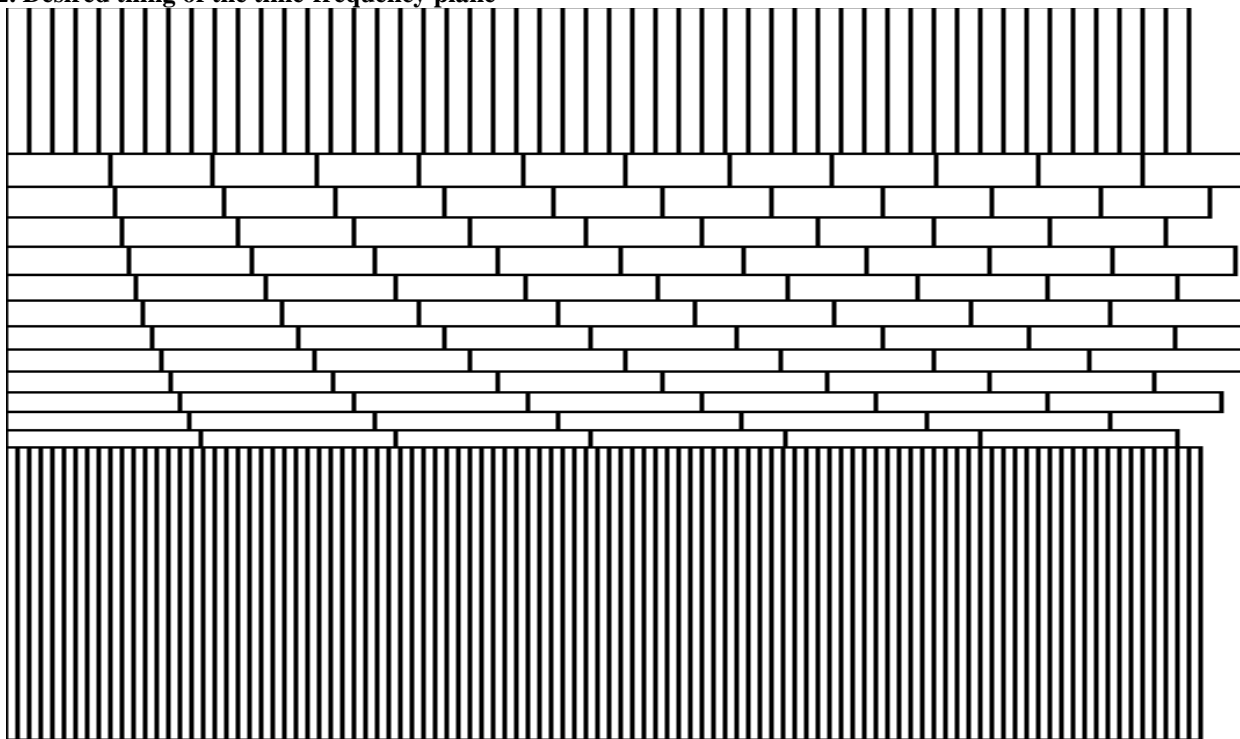


Figure 1: An example of the kind of tilings of the time-frequency plane used for music processing.

As stated in section 2.1, the first task is to choose a tiling of the time-frequency plane, with one band covering the frequencies of each tone of the scale, and where each band is critically sampled. The tones of the chromatic tempered scale used in most western music have a relative bandwidth of  $2^{1/12}$ . The basis elements and transform coefficients will be real. In addition, the frequency interval we are interested in might be narrower than determined by the sampling frequency of the input signal. For this reason we add two special bands, one covering frequencies from zero to the lowest tone we are interested in, and another one covering frequencies from the highest harmonic we'll consider up to  $F_s / 2$ . These special bands also allow us to tune the tiling to the specific frequencies used in music (i.e. central A at 440 Hz), regardless of the sampling frequency used.

Figure 1 shows example of such a tiling. It includes twelve analysis bands for twelve tones, a higher band covering 1/5 of the region of the plane, and a lower band covering 2/5 of the region of the plane. The signal length is 256 samples, and this tiling has 256 tiles.

### 2.3. Basis elements

The next problem is to find a wavelet that is well localized over such kind of tile. Appropriate dilations and displacements 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 no signal can have compact support both on time and on frequency as a consequence of Heisenberg's uncertainty principle. In fact, to have a good decay in one domain, infinite support is needed in the other. The best known wavelets for this kind of problems have infinite support in both domains. It is however possible to choose a better localization in one domain, accepting a worse localization in the other. For this work, frequency localization is privileged over time localization. The elements of the basis will have very little frequency overlap between different bands, to allow good discrimination of the tones in the signal being processed. This also means that there will be a significant temporal overlap between the elements of the basis that belong to the same band. This decision is made to match the characteristics of our hearing system.

### 2.4. Algebraic properties of the basis

The basis should be a Riesz basis. In addition, it is better if it is orthonormal, allowing easier and faster computation of the transform and reconstruction of signals. When building a basis with time-frequency atoms, an element of the basis can have significant correlation with those that are close in time and in frequency. Therefore, orthogonalization will somewhat dilute the localization of the elements to these neighbor bands and times.

#### 2.4.1. Orthogonalization between elements of the same band

If a wavelet is made orthogonal to all its displacements that are a multiple of the length of its tile, this property will be maintained for any Nyquist sampling of the wavelet. This means that the orthogonalization needs to be done just once on the wavelet, before building the basis.

#### 2.4.2. Orthogonalization between elements of different bands

The ratio between the bandwidth of different bands is irrational (except for whole octave distances). This means that the ratio between the sampling frequencies (the inverse of the temporal length of the tiles) of different bands will also be irrational (except for whole octave distances). This means that the exact configuration of the tiles that happen at a certain moment will never be repeated identically. This makes the analysis of correlation between elements of different bands a difficult problem: we will need to study and get rid of the correlation of every pair of basis elements a pair at a time.

## 3. AN EXPERIMENTAL APPROACH TO BUILDING BASES

Due to the lack of theoretical tools to meet these criteria, an experimental approach was taken. We need three fundamental functions: the wavelet, the scale function and the "mirror scale function". The wavelet is used to build the elements for the tiles that belong to a particular note of the musical scale. The scale function is used to build the elements of the lower special band. The mirror scale function is used to build the elements for the higher special band. Each of these fundamental functions should have little correlation to its displacements over different tiles. The three functions are then finely sampled. The elements of the basis are built by resampling the corresponding fundamental function to position it over each tile, and they are stored as columns of a square matrix. In order to do this; they are truncated at both ends as necessary. Then the correlation between all elements is attacked.

### 3.1. Fundamental functions

#### 3.1.1. Wavelet

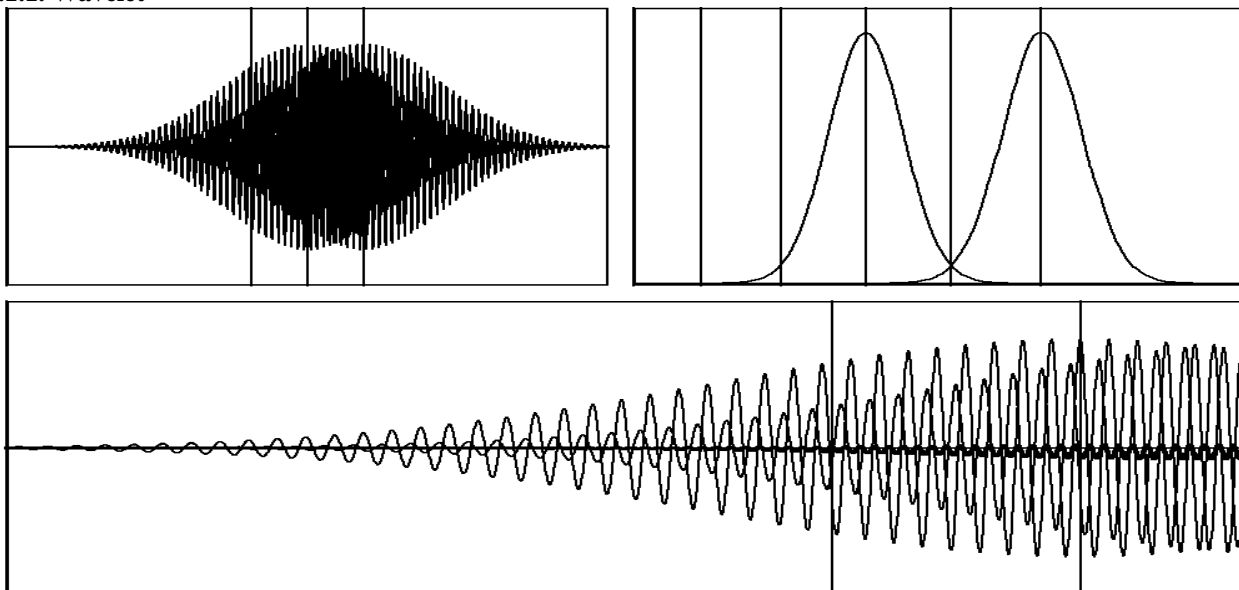


Figure 2: A Morlet wavelet.

The Morlet wavelet, a modulated Gaussian is the standard option for music analysis with discretized CWTs. This wavelet reaches the theoretical limit to time and frequency localization specified by Heisenberg's uncertainty principle. The Morlet wavelet is a complex signal, but a real version was chosen. The spectrum of the wavelet is the Gaussian:

$$e^{-((x-f_0)/b)^2}$$

As stated before, the frequency localization was privileged over temporal localization. The b factor controls this balance. A value of 0.3 was chosen for b. Figure 2 shows two consecutive displacements of the wavelet function at the left, a detail of them and their product at the bottom, and their spectra (the magnitude of the Fourier Transform) of two close bands at the right.

#### 3.1.2. Scale function

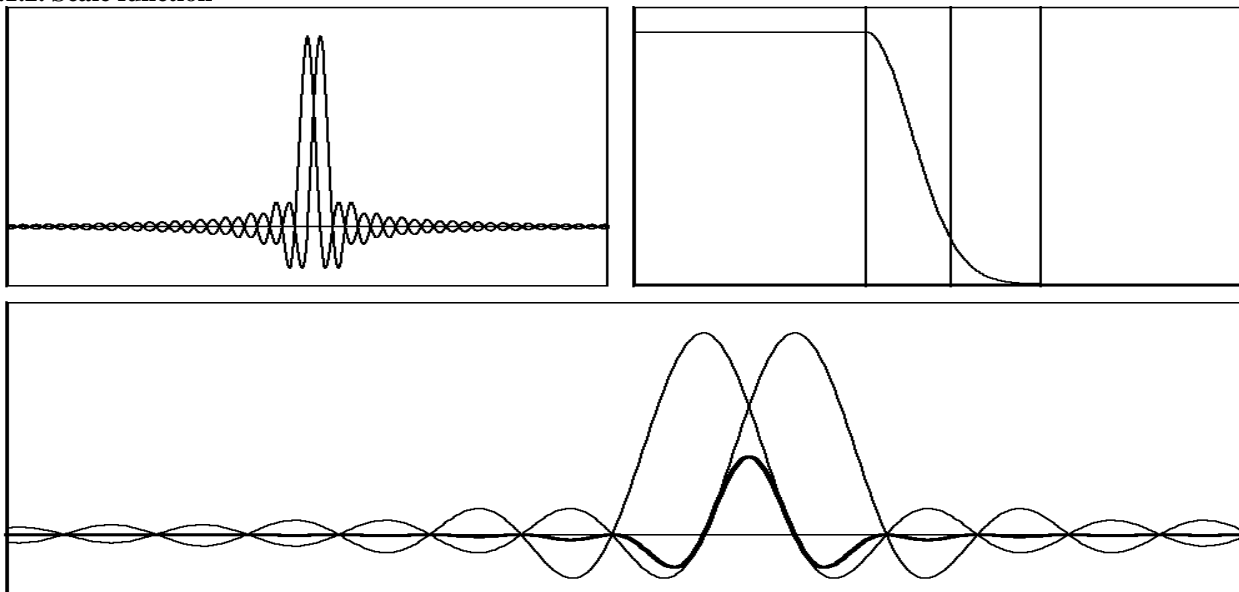


Figure 3: A scale function for the Morlet wavelet.

It is necessary to find a suitable scale function. It should cover the lower special band, that is all frequencies lower than a particular one. To do this, we took the higher half of the spectrum of the wavelet, and made the spectrum 1 from the central frequency to frequency zero. Figure 3 shows two consecutive displacements of the scale function at the left, a detail of them and their product at the bottom, and the spectrum at the right.

### 3.1.3. Mirror scale function

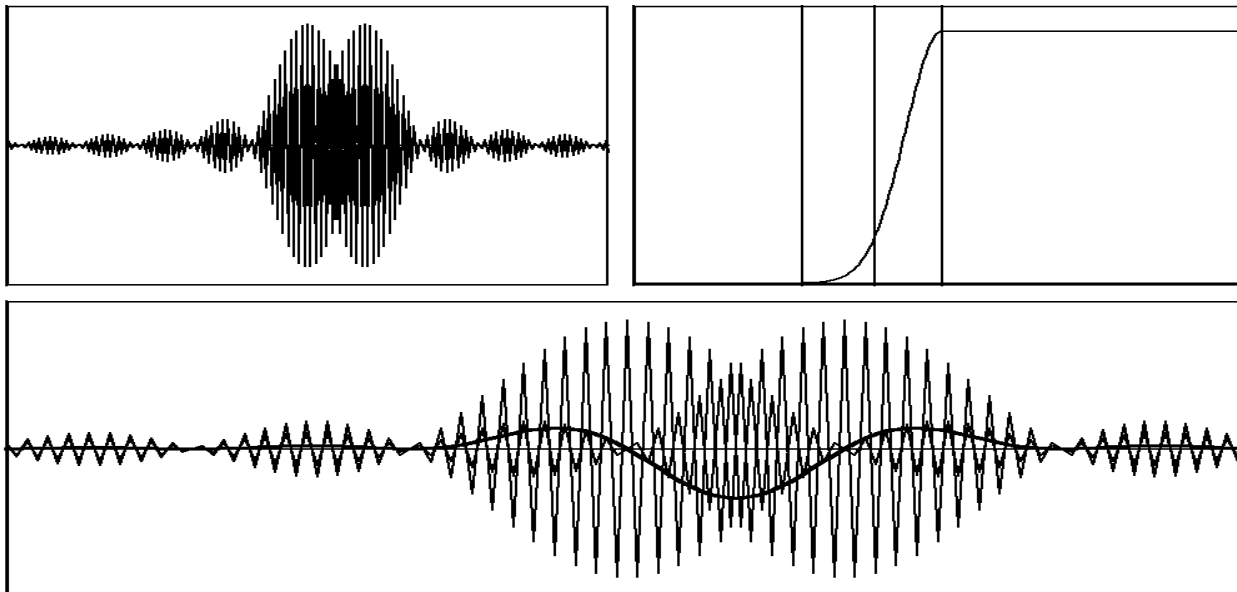


Figure 4: A mirror scale function for the Morlet wavelet.

It is also necessary to find a suitable function for the special higher band of the tiling. It should cover all frequencies higher than a particular one. It can be shown that if we take a signal and multiply every other sample by  $-1$ , we invert the signal's spectrum. This property is used to build the fundamental function for our higher band. Figure 4 shows two consecutive displacements of the mirror scale function at the left, a detail of them and their product at the bottom, and the spectrum at the right.

## 3.2. Correlation reduction

### 3.2.1. Reducing correlation of the fundamental functions

This wavelet shows a significant correlation between two successive displacements for the same dilation. This can be seen in figure 2, the product of two successive displacements lies mostly below zero. This means that the sum will be non-zero, implying a significant correlation. After experimentation, the only way to reduce this correlation that we found was to make the small oscillations of the wavelet to be at a phase of  $90^\circ$  in neighbor displacements. The frequency of these small oscillations is the central frequency  $f_c$  of the tile. Therefore, the tile's temporal length  $dt$  needs to be an integer number of cycles of frequency  $f_c$ , plus  $1/4$  or  $3/4$  of a cycle. This means:

$$dt = \frac{2k + 1}{4 f_c}$$

with integer  $k$ . This also implies that  $a_0$  can not be any arbitrary real number, it must be a rational number  $1 + 1/k$  with integer  $k$ .

At this point, the orientation of this research needs to be changed. From a true musical scale tiling of the plane, the focus is shifted to the best rational ( $a_0 = 1 + 1/k$ ) tiling of the plane.

The fractions of the form  $1 + 1/k$  closer to  $2^{1/12} \cong 1.05946$  are  $1 + 1/17 \cong 1.05882$  and  $1 + 1/16 \cong 1.0625$ . Then, we can approximate an octave (12 semitones at  $2^{1/12} \cong 1.05946$ ) by 10 bands lightly narrower and 2 bands slightly wider:  $(1 + 1/17)^{10} \cdot (1 + 1/16)^2 \cong 1.9993725$ . The error made is negligible.

We now have orthogonality between neighbor displacements. We also get orthogonality with all displacements at distances that are even multiples of  $dt$ .

The next step is to get orthogonality with displacements at odd distances. This is not trivial, because as we need to reduce correlation with some displacement of the very same function, we can not subtract one of them (multiplied by the correlation) to the other. Doing so, we would modify only one of them, but both need to be modified, as they are displacements of the same wavelet. Besides, to maintain wavelet symmetry, the same orthogonalization needs to be done with element at  $i$  distance to the left as with the element at  $i$  distance to the right. To further complicate things, adjusting correlation with elements at a particular distance will also modify the correlation with elements at other distances.

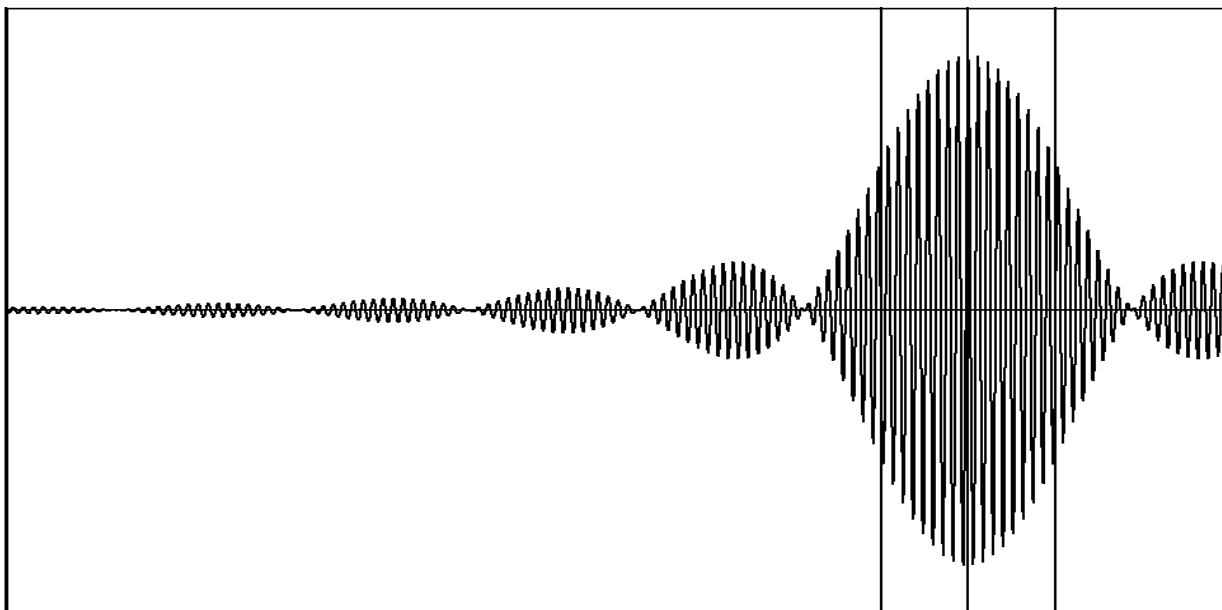


Figure 5: A Wavelet that is orthogonal to its displacements at multiples of  $dt$ .

To attack this problem, the following algorithm was developed:

- Compute correlation with relevant elements at odd distances. It will not be necessary to go beyond some distance (for example 24 tiles) because temporal localization of the wavelet means that the correlation will be low for long distances.
- If all computed correlations are negligible, end.
- Take the distance  $i$  that had greatest correlation.
- Try subtracting the projection at distances  $+i$  and  $-i$ , multiplying it by factors  $c$  between 0 and 1, obtaining a new wavelet for each value of  $c$ .
- Find the value of  $c$  that minimizes correlation between the corresponding wavelet and its displacements at distances  $+i$  and  $-i$ . Keep this wavelet and forget all the rest (including the one we started with).
- Go to the beginning.

This technique gave excellent results. We wanted to reduce correlation between elements of the same band below 1%. For this, it was only necessary to subtract displacements at  $i = 2, 4, 8$  (in this order) and the  $c$  coefficients used were 0.647453, 0.5556 and 0.5045. If a lower correlation is desired, it is possible to obtain it with a few more iterations of the algorithm. Figure 5 shows the wavelet built.

Correlation between displacements of the scale function and of the mirror scale function is reasonably low (below 1.5%). And there is no need to work on reducing it, because the whole basis will be orthogonalized anyway; and the correlation inside these special bands would only mean loss of temporal localization inside them. And as we do not intend to do any meaningful processing with them, this is not a problem.

### 3.2.2. Orthogonalization of the basis

When building the wavelet, frequency localization was privileged over temporal localization. Therefore, correlation between elements of different bands is reasonably low. It is below 1% except between bands that are result of different rational approximations. In these cases, correlation is below 7%. The only way we found to reduce this correlation is to orthogonalize the whole basis.

To eliminate this correlation (and any correlation that could be left inside each band), the whole basis is orthogonalized with a modified version of the Gram Schmidt algorithm. As previously stated, the elements of the basis are the columns of a matrix and they were truncated at the matrix bounds. This means that the elements that are close to the temporal bounds of the signal were severely truncated, therefore affecting their frequency localization. Doing the orthogonalization from left to right would destroy the frequency localization of the entire basis. The orthogonalization algorithm was modified to simulate moving the first columns to the right end of the matrix, doing the orthogonalization from left to right, and moving the columns back to the left end of the matrix. The orthogonalized basis has bad frequency behavior at both ends (i.e., it does not follow the tiling), but the central part results almost unaffected. Therefore it is important to pad the signals with zeroes at both ends, and to use corresponding bigger basis, to use only the “good” part of the basis for processing the relevant part of the signal.

## 4. RESULTS

To test these techniques in the real world, we built a big basis. The basis built one has three octaves of analysis, from central C (centered at 261.5 Hz) up to B two octaves above it (centered at 1974.585 Hz). This means we have 36 analysis bands. The sampling frequency is low: 5512.5 Hz. The basis was built for signals of 10,000 samples. At the chosen sampling frequency, this gives about 1.8 seconds of signal length. This is barely enough to get audible results, but the matrix that holds the basis needs almost 400Mb of memory space.

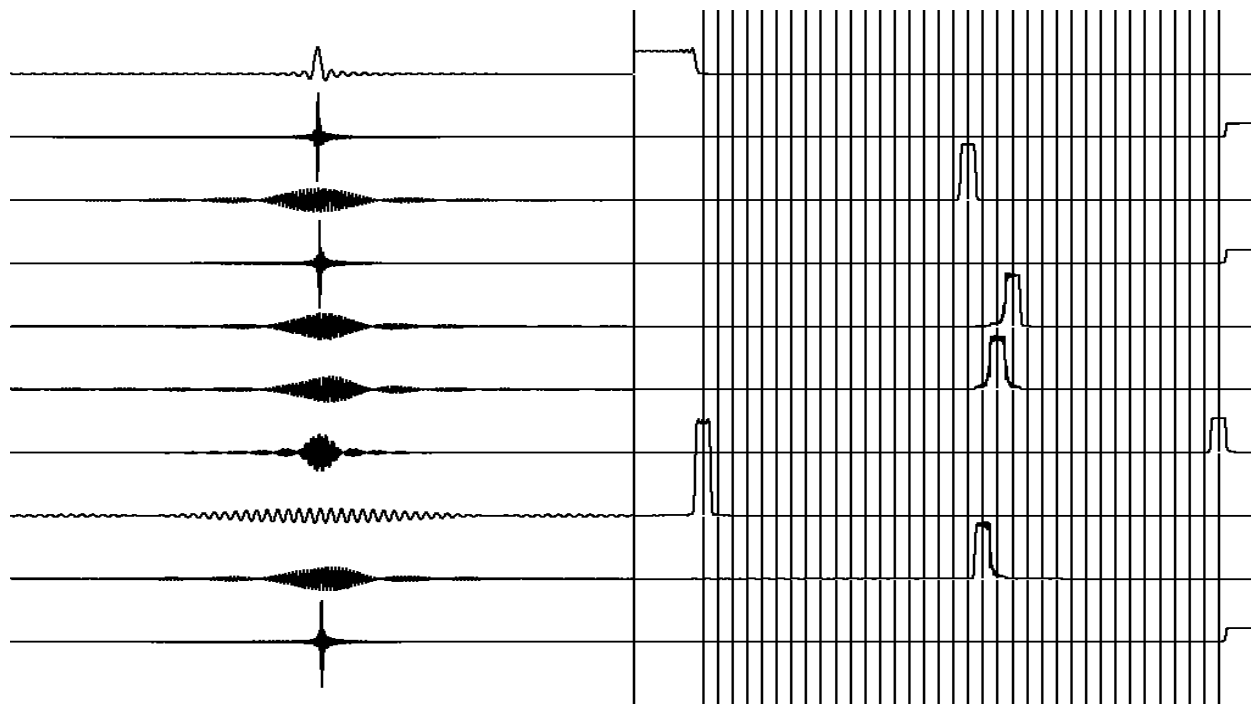


Figure 6: Part of the basis built.

Figure 6 shows part of the basis built. On the left, several consecutive elements are shown. On the right, the corresponding spectra (the magnitude of the Fourier transform) are shown. The vertical lines show the central frequency of each tone. The lower and higher special bands have been truncated, and as the frequency scale is logarithmic, the tones seem of equal width. The 6<sup>th</sup> element shown is for a tile made with a rational approximation of  $1 + 1/16$ . Before

orthogonalization, this element had significant correlation with its neighbor bands (the 5<sup>th</sup> and 9<sup>th</sup> elements shown). The frequency localization of these elements is affected by the orthogonalization of the basis, as it can be seen in the graph.

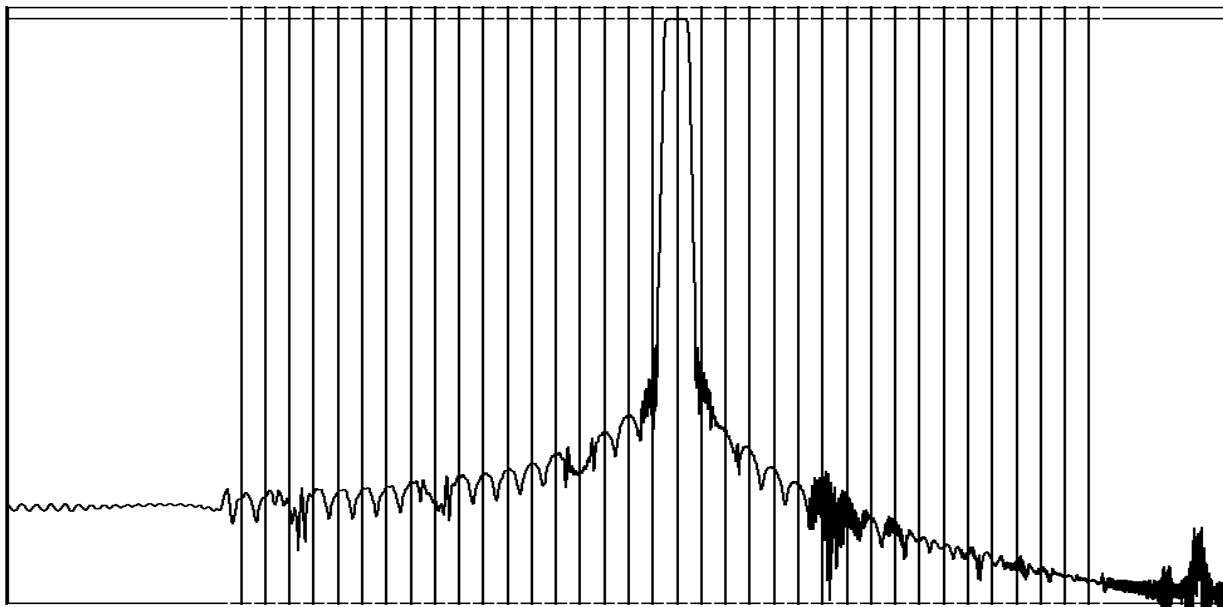


Figure 7: Frequency response of an element of the basis.

Figure 7 shows the outstanding frequency response of a typical element of the basis. The graph shows the magnitude of the Fourier Transform, expressed in db (a logarithmic scale used for audio applications), and the vertical lines show the central frequency of the analysis bands. The frequency scale is logarithmic. The element shown is the 3<sup>rd</sup> one in figure 6.

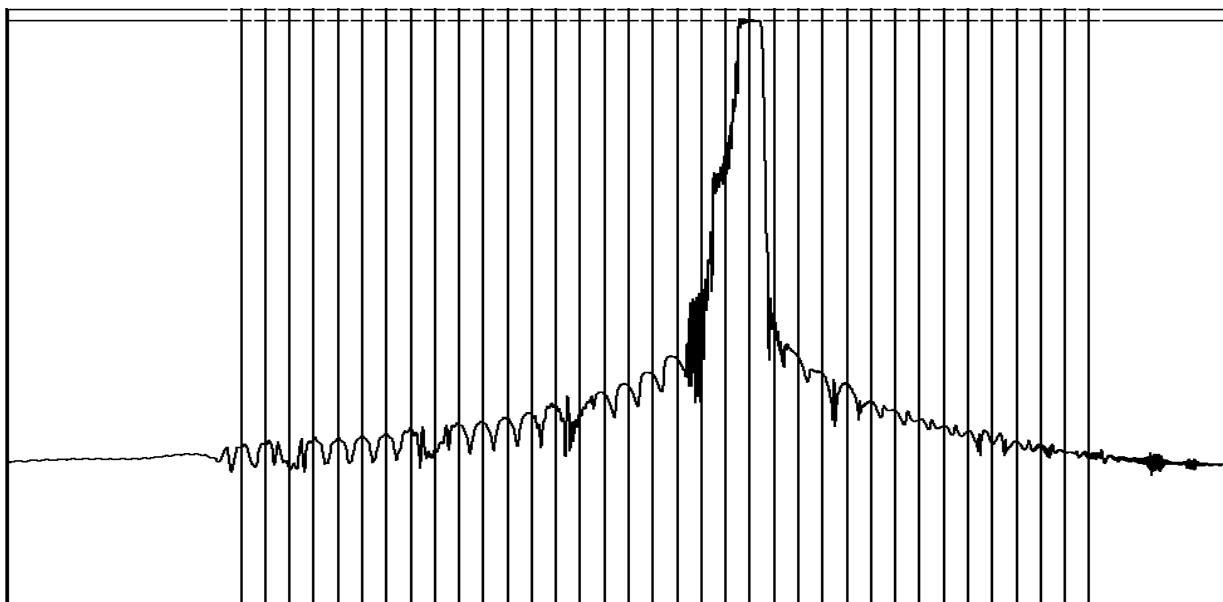


Figure 8: Frequency response of an element of the basis.

Figure 8 shows the magnitude frequency response of an element of the basis that has a neighbor band with different rational approximation (the 5<sup>th</sup> one in figure 6). The frequency response is not as good as in figure 7, but it is quite good anyway. The vertical scale is in db, and the vertical lines show the central frequency of the analysis bands. The frequency scale is logarithmic.



## 5. RELATED RECENT WORK

The results obtained can be compared with those from a recently published work<sup>4</sup>. In this work, the authors address the more general problem of taking any tiling of the time-frequency plane and building a basis that can approximate it. The work presented here is of more limited scope, focusing only on tilings that are suitable for music processing, but the frequency localization achieved is better. This means that the tone discrimination will be better and the results of processing will be closer to what the user expects.

Figure 9 is included for comparison with figure 7 in <sup>4</sup>. The frequency response graph is in linear frequency scale, and the vertical axis goes from -40 to 20 db. The element shown is the one in figure 8.

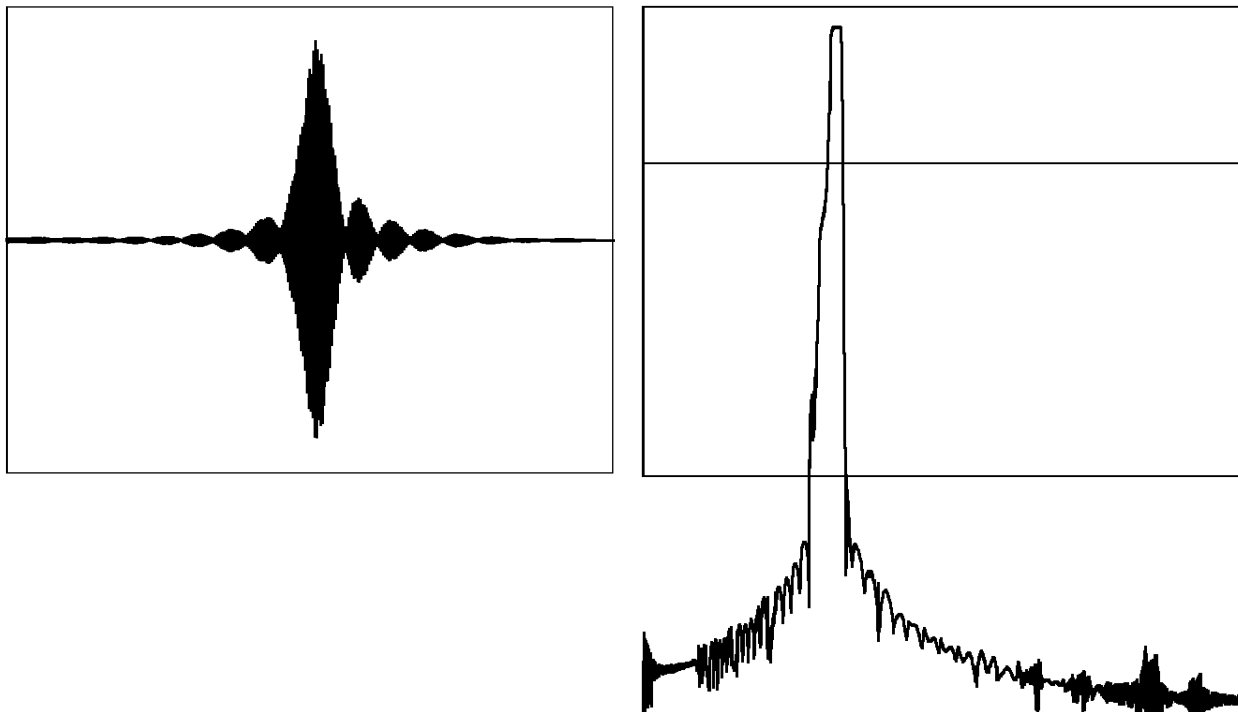


Figure 9: An element of the basis and its frequency response.

## 6. CONCLUSIONS

This work presents the current status of a research initiative on orthonormal bases for time-frequency representation of music signals. The bases built are the first ones developed specially for this problem, as far as we know.

Their virtues include:

- They are orthonormal
- They have excellent frequency localization
- They have reasonable temporal localization
- They are relatively easy to build

Their defects include:

- Each basis is built specifically for a signal length
- They have frequency localization problems at both ends, and they thus require zero padding of the signals
- There is no compact representation of the bases themselves
- For signal length, they require  $n^2$  space and computation takes  $O(n^2)$  time

These defects are a consequence of the rational approximations and the need for a global orthogonalization.

The next objective is to fix the defects shown by the bases described here, without losing their virtues. We are currently on the path towards bases that fulfill the criteria previously specified, but they are born orthogonal. They won't need orthogonalization. This will also mean that each element will be built independent of the rest, and it won't be necessary to store huge matrices. This will also open the door to computation in less than  $O(n^2)$  time.

## 7. REFERENCES

1. B. Torr sani, "An Overview of Wavelet Analysis and Time-Frequency Analysis", p. 22, proceedings of the International Workshop in Self-Similar Systems (Dubna, Russia), 1998.
2. I. Daubechies, *Ten Lectures on Wavelets*, p.16, Society for Industrial and Applied Mathematics, CBMS-NSF 61 Regional Conference Series in Applied Mathematics, 1992.
3. G. Strang, T. Nguyen, *Wavelets and Filter Banks*, cover, Wellesley-Cambridge Press, Wellesley, MA, 1997.
4. R. Bernardini, J. Kovacevic, "Arbitrary Tilings of the Time-Frequency Plane Using Local Bases", IEEE Transactions on Signal Processing, **vol. 47, number 8**, pp. 2293-2304, 1999.