

# Time-frequency Analysis for Music Signals

## A Mathematical Approach\*

Monika Dörfler

Institut für Mathematik, Universität Wien, Austria

### Abstract

This paper establishes a link between certain problems that arise when digital audio signals are processed, and a branch of mathematical theory, called Gabor analysis. Going back to the work of D. Gabor in 1946, Gabor analysis is based on the idea of representing arbitrary signals of finite energy in terms of building blocks which have a well-defined “center of gravity” in a time-frequency sense. Indeed, various widely used spectral domain methods correspond to the use of particular Gabor frames. The interpretation of this situation from the point of view of Gabor analysis makes the calculation of the so called dual window computationally efficient and post-processing by a weighting function unnecessary. Furthermore the idea of locally adaptive sets of building blocks will yield a generalisation of spectral domain methods especially helpful in the investigation of music signals.

### Introduction

Time-frequency analysis<sup>1</sup> is omnipresent in the processing of music. It might often not be called by this name, but in fact any method cutting a signal into pieces and doing Fourier analysis or filtering on the single pieces is time-frequency analysis. This paper aims to introduce the main features of Gabor analysis, see (Feichtinger & Strohmer, 1998), which align in many ways with “practical” time-frequency analysis. Gabor analysis yields a mathematically well-defined and precise framework helping to understand many issues in the processing of audio signals. Since in the last 20 years the basic problems of Gabor analysis (arising from the fact that

the collection of building blocks is non-orthogonal and even redundant) have been analysed in great detail by mathematicians and engineers, we believe that it makes sense to exhibit the potential of Gabor analysis to applied scientists, in particular to electronic musicians who certainly are familiar with very similar ideas. Ambiguities arising from the rather arbitrary choice of window, sampling constants in time and frequency and the consequences of these choices are better understood within this framework.

The merits of a mathematical approach discussed in this work in more detail are the following.

- It facilitates certain steps in implementation by imbuilding knowledge gained from knowing about the theoretical background.
- It offers a framework for desirable generalisation and adaption as will be discussed for a certain class of signals.

### Time can be frequency

One fundamental idea in a mathematical approach to time-frequency analysis is looking at time and frequency in a symmetric way, which makes the theory more unified and flexible. Instead of thinking of cutting a signal into pieces, Fourier analysing it and putting it back together after some processing, the Gabor approach thinks of a signal being projected onto basic functions, which are concentrated in certain regions of the time-frequency plane.

### Excursus: Three kinds of looking at signals

Mathematicians spend a great amount of time thinking about function spaces, their properties, members and mutual relations. One of the reasons is the fact that different spaces often

\* Supported by the science grant of the National Bank of Austria and the FWF, Project P14485.

<sup>1</sup> Material on time-frequency analysis can be found e.g., in Qian & Chen (1996) or Teolis (1998), among many others.

allow taking different points of view. When thinking about natural signals such as music, the first idea is to think of them as continuous signals. Of course they can't have infinite energy, so it is appropriate to assume they are members of  $L^2(\mathbb{R})$ , the space of square-integrable functions  $\mathbb{R} \rightarrow \mathbb{C}$ . This is a space with many nice properties, e.g., an inner product can be defined and it is in fact a Hilbert space. A lot of theory, e.g., on uncertainty, continuous dependencies or approximation is done in  $L^2$  or more general spaces of continuous functions.

On the other hand, signal processing is mostly done on digital computers, using discrete signals, so that

$$l^2(\mathbb{Z}) := \left\{ f: \mathbb{Z} \rightarrow \mathbb{C}: \sum_{n=-\infty}^{\infty} |f(n)|^2 < \infty \right\}$$

would be the more appropriate choice of Hilbert space. However, in real life, we can only process signals of finite length, say  $L$ , i.e., they are most easily understood as members in  $\mathbb{C}^L$ , a complex vector space of finite dimension  $L$ . This choice simplifies some questions and complicates others. To give an example for the kind of problems that can arise, think of how to define convolution, an operation with an important role in signal processing. On  $l^2$  it is defined as

$$(f * g)(n) = \sum_m f(m)g(n-m)$$

so the computation of  $f * g$  may ask for values such as  $f(-1)$ , which are not defined in  $\mathbb{C}^L$ . A trick to solve this kind of problem is to think of periodic sequences as members of the group  $\mathbb{Z}_L = \mathbb{Z} \bmod L$ , corresponding to a *circular extension* of the signal.

Subsequently, we will sometimes refer to one of the introduced spaces depending on the context. Generally, though, we'll be talking about functions  $f \in \mathbb{C}^L$ .

### Bases and frames

The Gauss function  $g(t) = e^{-\pi t^2} \in L^2$  is invariant under Fourier transform and has minimal "extension in the time-frequency plane" in the sense that it achieves equality in Heisenberg's uncertainty principle inequality.<sup>2</sup> If the energy in a signal is now thought of as being distributed over the time-frequency plane, indicating the amount of energy a certain frequency

<sup>2</sup>Heisenberg's inequality states that for all functions  $f \in L^2(\mathbb{R})$  and for all points  $(t_0, w_0)$  in the time-frequency plane

$$\|f\|_2^2 \leq 4\pi \|(t-t_0)f(t)\|_2 \|(w-w_0)\hat{f}(w)\|_2$$

where equality is achieved only by functions of the form

$$g(t) = C e^{2\pi i w_0 t} e^{-s(t-t_0)^2}, \quad C \in \mathbb{C}, s > 0$$

which are modulated and translated (i.e., time-frequency shifted) Gaussians.

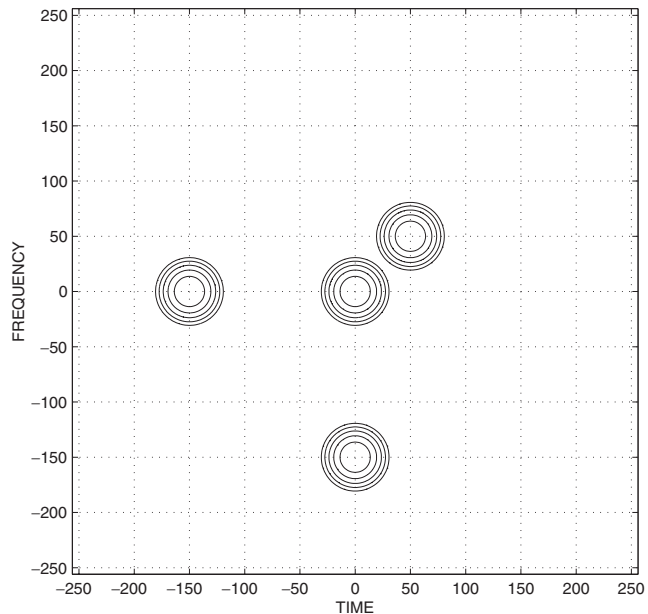


Fig. 1. Time-frequency shifted versions of a Gauss function in the time-frequency plane.

contributes at a specific time, then the Gabor approach can be described as "cutting" out pieces of at least the size of the Gauss function's time-frequency spread from this distribution. The energy thus captured can be thought of as the *inner product* of the given signal  $f$  with the Gaussian or any other window function.

### Definition 1

Let  $f$  and  $g \in \mathbb{C}^L$ . Their inner product is defined as

$$\langle f, g \rangle = \sum_{n=0}^{L-1} f(n) \overline{g(n)}$$

$g$  will from now on be called window function.<sup>3</sup> Let us now consider how a function  $g$ , centered around zero, can be moved to other points in the in the time-frequency plane, or, as we may prefer to call it in the discrete case, time-frequency lattice  $\mathbb{C}^L \times \mathbb{C}^L$ .

### Definition 2 (Time-frequency shifts)

$T_k f(t) := f(t+k)$  is called translation operator or time shift.

$M_l f(t) := e^{\frac{-2\pi i l t}{L}}$ ,  $l \in \mathbb{Z}$  is called modulation operator or frequency shift.

The composition of these operators,  $M_l T_k$  is a time-frequency shift operator.

<sup>3</sup>The Gaussian function is optimal in a certain, especially theoretical sense. In engineering Hanning or Hamming windows are most frequently used.

Generally, we are not interested in calculating the inner product in every point of the time-frequency lattice.<sup>4</sup> This would yield a redundancy of  $L$ , the length of the given signal. We downsample in time by  $a$  and in frequency by  $b$ , so that the redundancy reduces to  $\frac{L}{ab}$ .  $a$  and  $b$  are referred to as time-frequency lattice constants or time and frequency shift parameters. The family

$$g_{m,n} := M_{mb} T_{na} g$$

for  $m = 0, \dots, M-1$  and  $n = 0, \dots, N-1$ , where  $Na = Mb = L$ , is called the set of *Gabor analysis functions*.

Let us assume the  $g_{m,n}$  were an orthogonal basis for a moment. In this case, the inner products  $\langle f, g_{m,n} \rangle$  uniquely determine  $f$ , each representing a single and unique coefficient in the expansion

$$f = \sum_m \sum_n \langle f, g_{m,n} \rangle g_{m,n}$$

Together with Plancherel's formula  $\|f\|_2^2 = \sum_m \sum_n |\langle f, g_{m,n} \rangle|^2$  this gives a beautiful split of  $f$  in pieces, preserving the signal's energy in the coefficients.

Of course there is a problem. Theory<sup>5</sup> tells us, that the members of a basis of the above form can never be well-localised in both time and frequency.<sup>6</sup> Therefore we have to play a trade-off between nice properties of the representation on the one hand and satisfactory mathematical properties, similar to those of a basis, on the other hand.

The theory of *frames* gives the appropriate framework.

### Definition 3

A set of functions  $f_k$  in  $L^2(\mathbb{R})$  is called a *frame*, if there exist constants  $A, B > 0$ , so that

$$A\|f\|^2 \leq \sum_{k \in \mathbb{Z}} |\langle f, f_k \rangle|^2 \leq B\|f\|^2 \text{ for all } f \in L^2(\mathbb{R})$$

Note that above inequality can be understood as an ‘‘approximate Plancherel formula’’. For any frame a dual frame  $\tilde{f}_k$  exists, allowing an expansion of  $f$  as:

<sup>4</sup>Calculating the inner product in every point of the time-frequency lattice would yield the *full short-time Fourier transform*, a representation of redundancy  $L$ . The term ‘‘short-time Fourier transform’’ is often used for sampled short-time Fourier transforms as well. The spectrogram, its modulus squared, is one of the most popular time-frequency representations.

<sup>5</sup>The *Balian-Low theorem* is a key result in time-frequency analysis. It states that if a Gabor system in  $L^2(\mathbb{R})$  forms an orthogonal basis for  $L^2(\mathbb{R})$ , then

$$\left( \int_{-\infty}^{\infty} |tg(t)|^2 dt \right) \left( \int_{-\infty}^{\infty} |\omega g(\omega)|^2 d\omega \right) = +\infty.$$

There exist several extensions, first of all to the case of *exact frames*, which are frames that cease to be complete when any element is deleted. See (Benedetto et al., 1998) for more details.

<sup>6</sup>Pure sinusoid form an orthogonal basis and are a nice example for functions having unbounded support in time.

$$f = \sum_k \langle f, f_k \rangle \tilde{f}_k = \sum_k \langle f, \tilde{f}_k \rangle f_k \quad (1)$$

If the frame is not a basis, as will be the case in most applications, then the coefficients  $c_k = \langle f, f_k \rangle$  are not unique, still optimal in the sense of minimising  $\sum_k |c_k|^2$ .

The special case  $f_k = g_{m,n}$  is called Gabor or Weyl-Heisenberg frame.

### Framebounds

The frame bounds  $A$  and  $B$  are the infimum and supremum, respectively, of the eigenvalues of the *frame operator*  $S$ , defined as

$$Sf = \sum_k \langle f, f_k \rangle f_k$$

In the finite discrete case of  $f \in \mathbb{C}^L$  a collection  $\{g_{m,n}\} \in \mathbb{C}^L$  with  $k = NM$  can only be a frame, if  $L \leq k$  and if the matrix  $G$ , defined as the  $k \times L$  matrix having  $\overline{g_{m,n}}$  as its  $(m + nM) - th$  row, has full rank. In this case the frame bounds are the maximum and minimum eigenvalues, respectively. They yield information about numerical stability. The closer the frame-bounds are, the closer the frame operator will be to a diagonal matrix. If the  $A$  and  $B$  differ too much, the inversion of the frame operator is numerically unstable. Why are we interested in the inversion of the frame operator?

The canonical *dual frame*  $\tilde{g}_{m,n}$ , which yields reconstruction as in (1), is given by

$$\tilde{g}_{m,n} = S^{-1} g_{m,n}$$

as

$$f = S^{-1} Sf = \sum \langle f, g_{m,n} \rangle S^{-1} g_{m,n} = \sum \langle f, g_{m,n} \rangle \tilde{g}_{m,n}$$

The case  $A = B$  is called *tight frame*. In this case  $S = AI$ . ( $I$  denotes the identity operator) and therefore  $S^{-1} = \frac{1}{A}I$ . Tight frames will be further discussed in section 3.1.2.

The next section introduces the special case arising from applications in audio signal processing. We will see that in this case the frame operator takes a simple form.

### The bridge to applications

Let us from now on assume that we are given a signal  $f \in \mathbb{C}^L$ . This signal represents a piece of music or a spoken sentence etc., which we are interested to investigate and/or modify. Modifications might aim to achieve noise reduction in old or degraded recordings, see e.g. (Godsill & Rayner, 1998). Another issue might be the extraction of certain features of the signal, for example single instrument components. Let us further assume that an engineer approaches the problem by using a Fourier transform of length  $l$  in a first step. This implies that the window used for cutting out the part of interest must have this length. Looking at the definition of the Gabor coefficients:

$$c_{m,n} = \langle f, g_{m,n} \rangle = \sum_{j=0}^{L-1} f(j) \overline{g_{m,n}(j)}$$

as an inner product, which can be interpreted as correlation between the window and the respective part of the signal, we can see that the signals  $f$  and  $g_{m,n}$  must have the same length, at least theoretically. Practically, of course, as  $l \ll L$ , most of the “theoretical”  $g$  would be zero. As we don’t tend to waste computation time on multiplying with 0, only the *effective* length of  $g$ , here  $l$ , is multiplied with the part of interest of  $f$ . This procedure implicitly introduces a frequency lattice constant  $b = \frac{l}{L}$ . The time constant  $a$  is related to what is often called *overlap*. If  $a = \frac{1}{2}$  or  $a = \frac{1}{4}$ , the overlap is  $\frac{1}{2}$  and  $\frac{3}{4}$ , respectively. The redundancy of the representation is thus given by  $red = \frac{1}{a}$ , e.g., if the overlap is half the windowlength, we get twice as many datapoints as in the original signal. This is in accordance with the general case where

$$red = \frac{\frac{L}{a} \frac{L}{b}}{L} = \frac{L}{ab} = \frac{L}{a \frac{L}{l}} = \frac{l}{a}$$

**Remark:**

The reduction of redundancy from  $L$  in the case of the full short-time Fourier transform to a reasonable amount of redundancy in the Gabor setting ensures a balance between computability on the one hand and sufficient localisation on the other hand. The choice of a reasonable window-length and overlap common in applications corresponds roughly to such a rather balanced situation in the Gabor setting. Gabor theory, though, allows for more general choices of lattices, concerning the redundancy as well as the distribution of the lattice-points. It also provides detailed knowledge about the dependance of results on the choice of analysis parameters. This is especially decisive in the case of modification of the synthesis coefficients, which are non-unique due to the redundant nature of the expansion. Modification on the coefficient level corresponds to applying *Gabor multipliers*, a field which promises more insight in the properties of such modifications, see e.g., Zheng and Feichtinger (2000).

Let us now look at the calculation of the inner products  $c_{m,n} = \langle f, g_{m,n} \rangle$  more closely. They can also be written as

$$(c_{m,n})_{m=1, \dots, M; n=1, \dots, N} = G \cdot f$$

where  $G$  is the operator (matrix) introduced in Section 2.2.  $G$  consists of blocks

$$G_n, n = 0, \dots, N-1$$

each corresponding to one time-position of the window  $g$ . If we define  $g^l$  as the restriction of  $g \in \mathbb{C}^L$  to its non-zero part of length  $l$ , we get the following. The block  $G_n$  acts on the samples  $f(na + 1), \dots, f(na + l) =: f_{na}(t)$  by taking inner products of this slice  $f_{na}$  of the signal with each of the  $l$  modulated windows

$$\begin{aligned} M_{mb} g^l(t) &= e^{\frac{-2\pi i m b t}{L}} g^l(t) \\ &= e^{\frac{-2\pi i m \frac{L}{l} t}{L}} g^l(t) = e^{\frac{-2\pi i m t}{l}} g^l(t) \\ m &= 0, \dots, M-1 \quad \text{and} \quad t = 0, \dots, l-1 \end{aligned}$$

The coefficients  $e^{\frac{-2\pi i m t}{l}}$  are exactly the entries of the Fourier matrix  $\mathcal{F}_l$  of the FFT of length  $l$  with  $\hat{f} = \mathcal{F}_l f$ . Therefore

$$\begin{aligned} G_n f_{na}(t) &= \mathcal{F}_l(f_{na} \cdot g^l)(t) \\ t &= 0, \dots, l-1 \quad \text{and} \quad n = 0, \dots, N-1 \end{aligned}$$

and the action of  $G_n$  on  $f_{na}$  corresponds to multiplying  $f$  with  $g$ , skipping zero entries and taking the Fourier transform of the non-zero part.

**Remarks:**

1. Although for implementation in real-life situations, the FFT-approach is always preferred, it is useful to look at the expansion from an operator point of view. Many important theoretical issues, yielding better understanding also for the applications, can be investigated more easily. For example the dependance of the coefficients on the choice of window  $g$  (see Feichtinger & Zimmermann, 1998), or the sensitivity of an expansion to perturbations of the signal or sampling grid, see (Christensen, 1998), are questions also of practical relevance. Additionally, as mentioned before, this approach is more flexible in allowing adaption to cases requiring other than the regular product lattice (see Feichtinger et al., 1995; Kozek et al., 1996).
2. As mentioned before, all operators in Gabor theory generally act on the whole signal length  $L$ . In the definition of the building blocks  $g_{m,n}$ , the modulation operator is therefore defined as

$$\begin{aligned} M_{mb} g(t) &= e^{\frac{-2\pi i m b t}{L}} g(t) \\ &\text{for } m = 0, \dots, N-1 \text{ and } t = 0, \dots, L-1 \end{aligned}$$

The blocks  $G_n$ , as opposed to the situation arising from implementation as discussed above, will not have identical entries, as the zero entries are in different positions.

*Example:*

Let  $g \in \mathbb{C}^{32}$  with

$$g(t) \begin{cases} \neq 0 & \text{for } t = 0, \dots, 7 \\ = 0 & \text{else} \end{cases}$$

Then  $\left( \text{by assumption } b = \frac{L}{l}, \text{ so that } e^{\frac{-2\pi i m b t}{L}} = e^{\frac{-2\pi i m t}{l}} \right)$

$$M_{mb} g(t) = \left( g(0), e^{\frac{-2\pi i m}{l}} g(1), e^{\frac{-2\pi i 2m}{l}} g(2), \dots, e^{\frac{-2\pi i 7m}{l}} g(7), 0, \dots, 0 \right)$$

whereas

$$M_{mb}T_a g(t) = \left( 0, \dots, 0, e^{\frac{-2\pi i m a}{l}} g(a), e^{\frac{-2\pi i m (a+1)}{l}} g(a+1), \dots, e^{\frac{-2\pi i m (a+7)}{l}} g(a+7), 0, \dots, 0 \right)$$

$e^{\frac{-2\pi i m a}{l}}$  is not necessarily 1, so that the blocks will differ by a phase factor. The following statement makes these considerations precise.

### Theorem 1

Let  $g \in \mathbb{C}^L$  be a window function with

$$g(t) \begin{cases} \neq 0 & \text{for } t = 0, \dots, l-1 \\ = 0 & \text{else} \end{cases}$$

where  $\frac{l}{l} \in \mathbb{N}$ .  $g^l$  is the restriction of  $g$  to its non-zero entries. Let furthermore  $b = \frac{l}{l}$  and  $a$  be a divisor of  $l$ .

Then the blocks in the matrix  $G = (g_{m,n})$  differ from the identical blocks  $G_n$  in the matrix  $G^l$  arising from FFT-implementation only by a phase factor

$$w_0^{(d-1)ak}$$

where  $d$  is the block index,  $k$  is the row index in the respective block  $G_m$  of  $G$  and  $w_0 = e^{\frac{-2\pi i}{l}}$ .

- The restriction that  $a$  be a divisor of  $l$  is also due to the usual choice of parameters in applications. Two common cases would be  $a = \frac{l}{2}$  and  $a = \frac{l}{4}$ , in which cases the number of different kinds of blocks reduce to 2 and 4, respectively.

The difference only concerns the phase spectrum, which is usually not considered in further processing, except for reconstruction. The dual window does not depend on the phase factor in the case discussed in the theorem as will be seen below.

### Mastering the frame operator – the Walnut representation

Let us now come back to the central question of how to find a set of windows  $\tilde{g}_{m,n}$  for reconstruction as in (1). If it is possible to find a window  $\tilde{g}$  which is smooth and similar to the original window  $g$  especially in decaying to zero and if the rest of the dual family can be deduced in analogy to the Gabor analysis function set by time-frequency shifts, this will make reconstruction in a kind of overlap-add process easier. Infact, all the above conditions can be fulfilled. Generally, the elements of the dual frame  $(\tilde{g}_{m,n})$  are generated from a single function (the dual window  $\tilde{g}$ ), just as the original family. This follows from the fact that  $S$  and  $S^{-1}$  (the frame operator and its inverse) commute with the modulation and translation operators  $M_{nb}$  and  $T_{ma}$ , for  $m = 1, \dots, M$  and  $n = 1, \dots, N$ , see e.g., Daubechies (1990).

The higher redundancy, the closer the shape of the dual window gets to the original window's shape. As in applications redundancy of 2, 4 or even higher are common, well

localised dual windows can be found. Even more is true. The special situation in which the effective length of the window  $g$  equals or is shorter<sup>7</sup> than the FFT-length, the frame operator takes a surprisingly simple form.

From the definition of the frame operator

$$Sf = \sum_{m,n} \langle f, g_{m,n} \rangle g_{m,n}$$

we deduce that the single entries of  $S$  are given by

$$S_{j,k} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} M_{mb}T_{na}g(j) \overline{M_{mb}T_{na}g(k)}$$

Looking at the inner sum, note that  $\sum_{m=0}^{M-1} e^{\frac{2\pi i m b (j-k)}{L}} = 0$  if  $(j-k)$  is not equal to 0 or a multiple of  $M$ . In these cases

$$\sum_{m=0}^{M-1} e^{\frac{2\pi i m b (j-k)}{L}} = \sum_{m=0}^{M-1} e^{\frac{2\pi i m b M}{L}} = M$$

This leads to the *Walnut representation* (Hel and Walnut, 1989) of the frame operator for the discrete case:

$$S_{jk} = \begin{cases} M \sum_{n=0}^{N-1} T_{na}g(j) \overline{T_{na}g(k)} & \text{if } |j-k| \bmod M = 0 \\ 0 & \text{else} \end{cases} \quad (2)$$

There will obviously be non-zero entries in the diagonal,  $j = k$ , but as  $M = l$ , i.e., the window-length,  $j = k$  is infact the only case for which  $|j-k| \bmod M = 0$  holds and  $g(j)$  and  $g(k)$  both have non-zeros values. Therefore, the frame operator is diagonal and the dual window  $\tilde{g}$  is calculated as

$$\tilde{g}(t) = g / \left( M \sum_{n=0}^{N-1} T_{na} |g(t)|^2 \right)$$

### Some dual windows

Figures 2 and 3 show a Gaussian window of length 2048 and a Hanning window of length 512, respectively, their duals and their Fourier transform. Note that the frequency concentration is very good for the Gaussian window and its dual (to 40 db down). The weaker frequency concentration of the Hanning window's dual is due less overlap which leads to some more smearing in frequency. Note that this detail is one of the reasons why it has proved advisable to use a redundancy of 4 in overlap-add procedures.

### Tight frames: synthesising with the analysis window

The definition of a *tight frame* was given in section 2.2. As a matter of fact for any given Gabor frame a corresponding tight frame can be found for which  $\tilde{g} = g$ .

<sup>7</sup>E.g., in the case of zero padding.

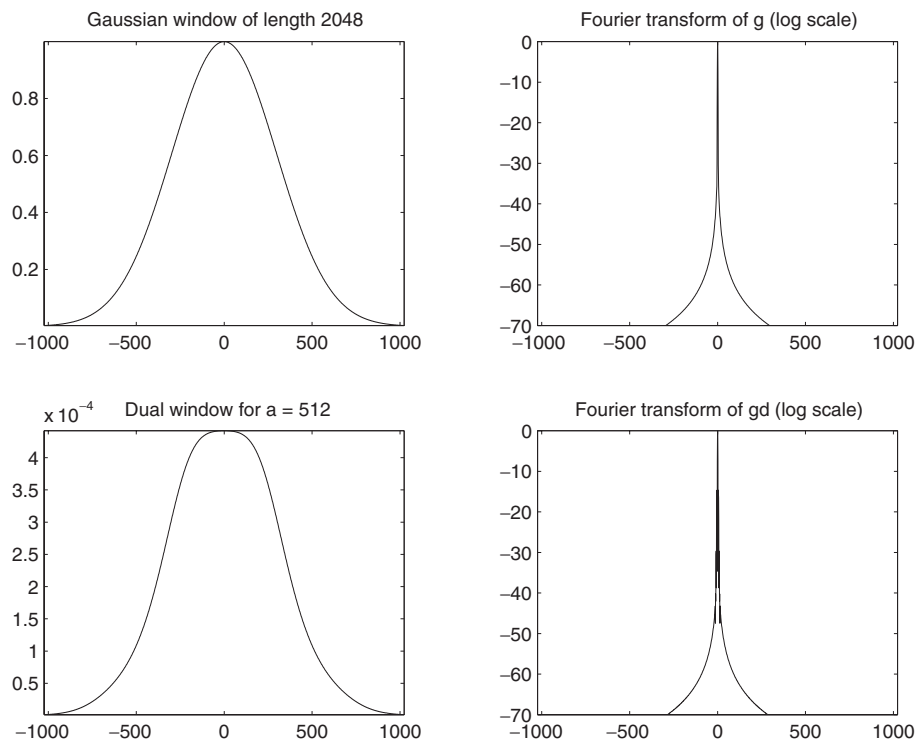


Fig. 2. Gaussian window and dual window (Redundancy 4).

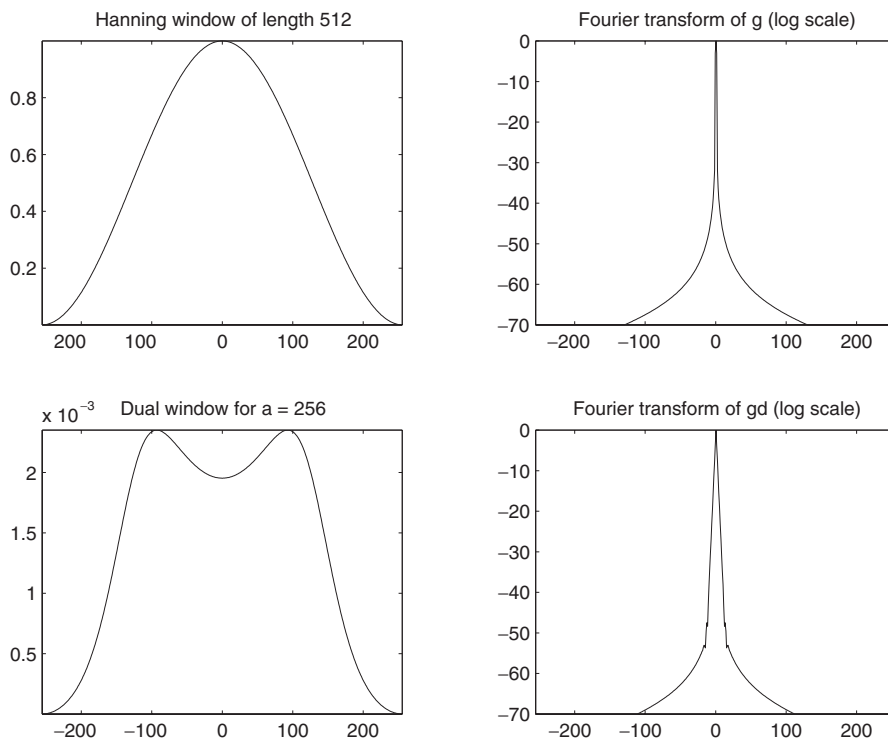


Fig. 3. Hanning window and dual window (Redundancy 2).

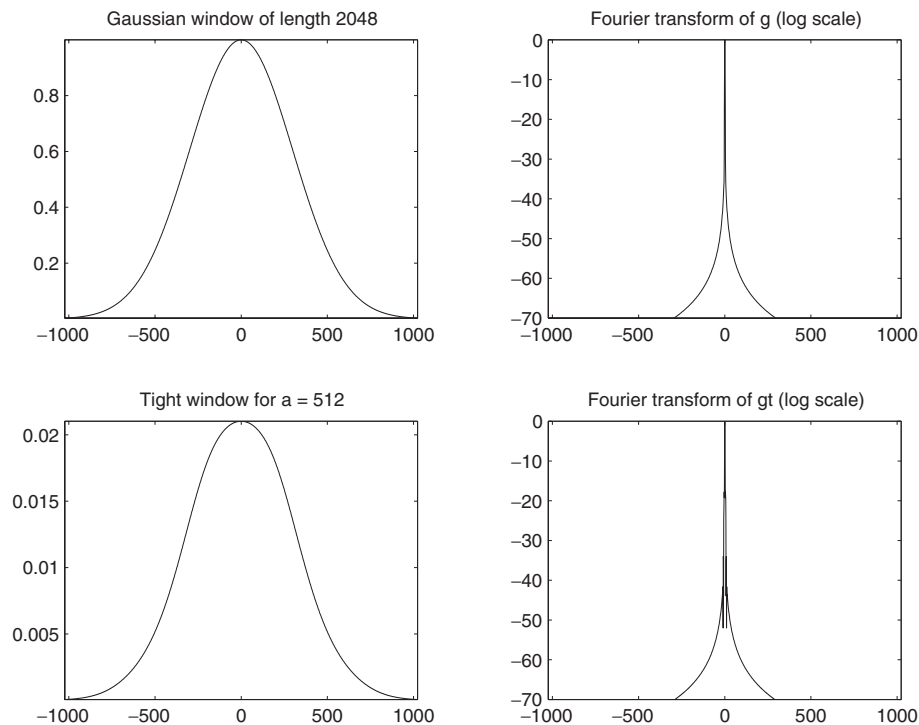


Fig. 4. Gaussian window and tight window (Redundancy 4).

Note that the frame operator  $S$  is a positive and symmetric and therefore selfadjoint operator, from which it follows that  $S^{-1}$  and  $S^{-\frac{1}{2}}$  are selfadjoint as well.

These properties allow the following manipulations of the expansion (1):

$$\begin{aligned} \sum_{m,n} \langle f, g_{m,n} \rangle \tilde{g}_{m,n} &= S^{-1} S f \\ &= S^{-\frac{1}{2}} S S^{-\frac{1}{2}} f = \sum_{m,n} \left\langle f, S^{-\frac{1}{2}} g_{m,n} \right\rangle S^{-\frac{1}{2}} g_{m,n} \end{aligned}$$

#### Remark:

Note that the tight window given by  $g_t = S^{-\frac{1}{2}} \cdot g$  is closest to the original window in the following sense.

Let  $g$  be a window generating a frame for lattice constants  $a$  and  $b$  and let  $g_t$  be the tight window given as  $g_t = S^{-\frac{1}{2}} \cdot g$ . Then for any function  $h$  generating a tight frame for lattice constants  $a$  and  $b$ , the following holds (Janssen & Strohmer, 2000):

$$\|g - g_t\|_2 \leq \|g - h\|_2$$

This result shows that the tight window calculated as  $g_t = S^{-\frac{1}{2}} \cdot g$  combines the advantage of using the same window for analysis and synthesis with optimal similarity to a given analysis window. At the same time no “correction” by multiplication with a gain function is necessary after processing, which

makes processing more efficient and the results less ambiguous in the case of modification of the synthesis coefficients. The tight window  $g_t$  corresponding to a given window  $g$  and the time constant  $a$  can therefore be calculated as:

$$g_t = S^{-\frac{1}{2}} g = g / \sqrt{\left( M \sum_{n=0}^{N-1} T_{na} |g(t)|^2 \right)}$$

#### Some tight windows

Figures 4 and 5 show the same windows as before, the corresponding tight windows and their Fourier transform.

### A lattice that reads the music

When analysing any class of signals, one is naturally interested in features specific to that particular class. In the case of music signals, for example, transients are important for several reasons. They give important cues for onset timing, and they carry information about instrument timbre (often instrument perception hinges on the perception of transients.) As another example, in low-frequency regions, very fine frequency resolution is required, because notes in this region lay the harmonic basis, musically speaking. This is especially true in music such as Jazz, where the function of the bass determines the harmonic structure and function of the whole piece.

Although at first glance the above description suggests the usage of wavelets for the given problem, they have not yet

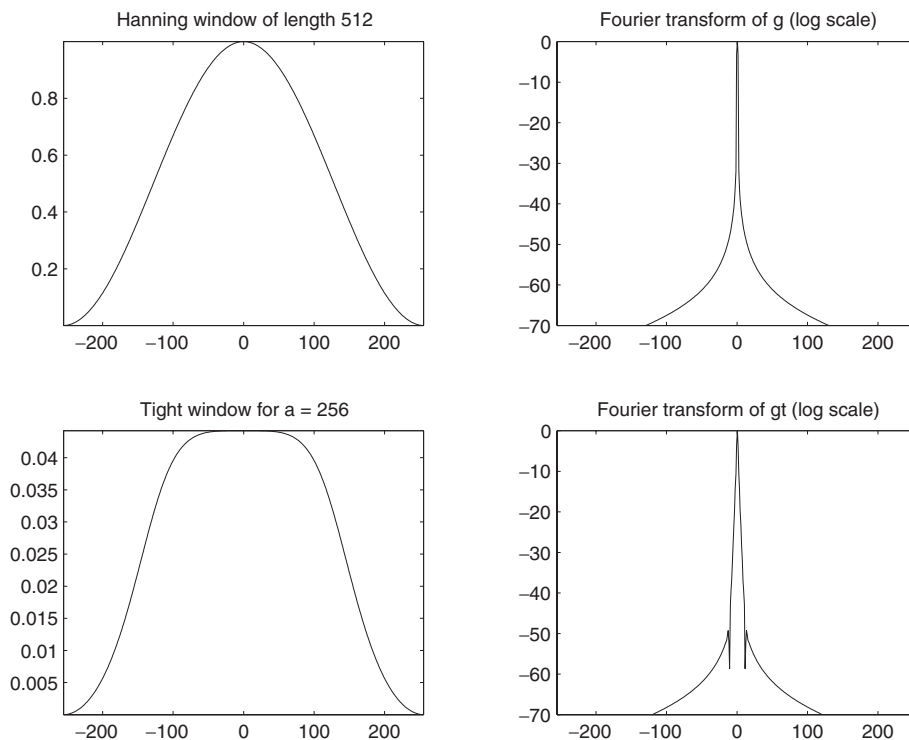


Fig. 5. Hanning window and dual window (Redundancy 2).

been shown to be well-adapted to the processing of music signals (Dörfler & Feichtinger, 1999). Several other approaches to deal with this problem are well-known; e.g. the use of wavelet packets, or certain classes of smoothed Wigner distributions (Pielemeier & Wakefield, 1996).<sup>8</sup>

Our approach aims to use the background and tools of Gabor analysis in order to incorporate knowledge about the class of signals in question, by constructing flexible, signal-adaptive tilings of the time-frequency plane. We can thus achieve appropriate results while maintaining the appealing advantages, both practical and theoretical, of Gabor analysis. Preliminary results, which we outline in the following sections, show this to be a promising approach.

### Adapting window and grid

In order to achieve a setting adapted to music as discussed above, it will be necessary to use wider windows with good frequency concentration in low-frequency regions, whereas in the high-pass regions, where mainly transients and broadband signals components occur, rather short windows, which don't have to be very localised in frequency, will be of use.

We can construct frames corresponding to this idea, by using sets of building blocks corresponding to the frequency

<sup>8</sup>In fact, the modulus squared of the short-time Fourier transform itself (the spectrogram) may be shown to be a smoothed Wigner distribution, in which the smoothing kernel is the Wigner transform of the window function used in the analysis.

region they cover. In order to keep reconstruction feasible, the frequency regions corresponding to the different sets of building blocks will be overlapping. The reconstruction can then be accomplished by means of *local biorthogonal families*. This approach is currently under study and the next section shows some properties of local biorthogonal families.

### Local biorthogonal families

The reconstruction of a signal in the context of Gabor analysis is accomplished by first finding a dual family of building blocks and then effecting an expansion similar to a basis expansion. It is possible to find *local biorthogonal families*, i.e. families of building blocks biorthogonal<sup>9</sup> to a family of atoms corresponding to a subset of the time-frequency plane. Such a family can be found by calculating the *pseudo-inverse* of the matrix comprised of the original family's members. For any  $k \times L$ -matrix  $M$  exists a *singular value decomposition*, which allows to write  $M$  as

$$M = V \cdot D \cdot U^T$$

where  $D$  is a  $k \times L$ -matrix with non-zero values only for  $M_{jj}$ ,  $j \leq \text{rank}(M)$  and  $V$  and  $U$  are unitary matrices comprised of the eigenvectors of  $M \cdot M^T$  and  $M^T \cdot M$ , respectively. Now in the case of a local set of building blocks,  $k$ , the number of blocks, will be a lot smaller than  $L$ , the length of the signal.

<sup>9</sup>On the adjoint grid, this statement corresponds to the *Wexler-Raz principle* (see Feichtinger & Strohmer, 1998).

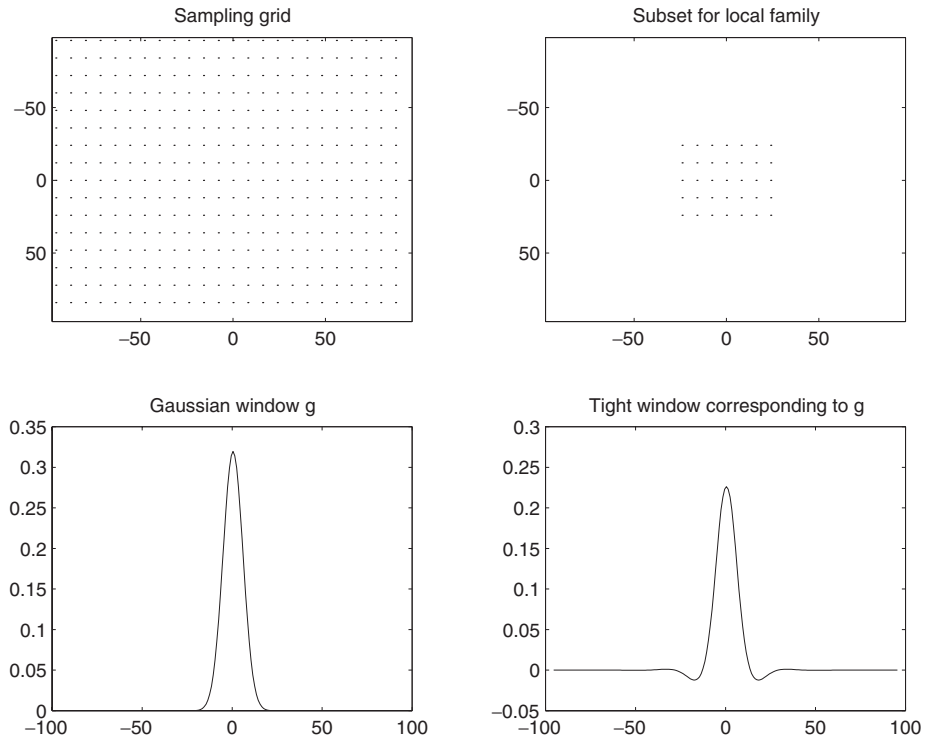


Fig. 6. Sampling grids for full Gabor family and subset.

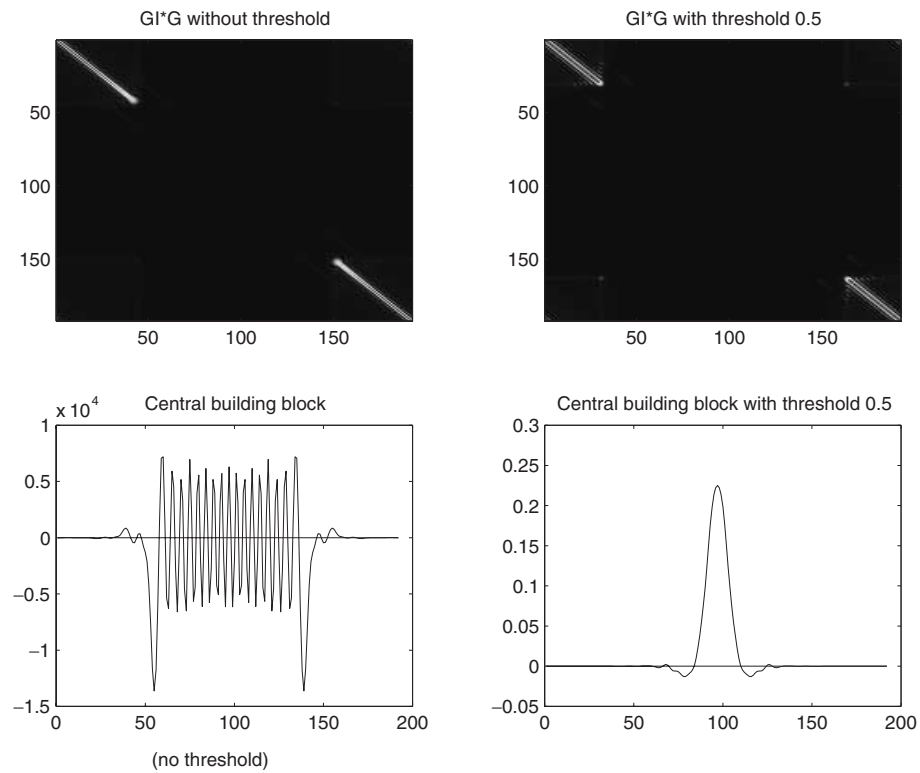


Fig. 7. Exact and approximate local dual window.

Furthermore, due to the redundancy of the (local) representation resulting from overlap of the building blocks, which is always desired, only  $r_1$  singular values for  $r_1 < \text{rank}(M)$  will be greater than a given threshold  $th$ :  $0 < th < 1$ .

Considering that only these singular values correspond to salient eigenvalues in the construction of the pseudoinverse, a threshold can be introduced in order to set all singular values below a certain level to zero in the inversion. The eigenvectors of  $M' \cdot M$  corresponding to small singular values are concentrated at the boundary of the area of interest, whereas eigenvectors corresponding to 0 will be concentrated outside this region. The pseudoinverse of  $M$  is given by

$$M^+ = U' \cdot (D')^{-1} \cdot V$$

where  $(D')^{-1}$  is the matrix  $D$  with inverted non-zero entries. From this we can see that the eigenvectors corresponding to low singular values will blow up in the calculation of  $M^+$ , so that all members of an exactly local biorthogonal family would be concentrated at the boundary of the region of interest. This can be avoided by introducing a threshold, below which all singular values are set to 0. The resulting family of atoms is thus only approximately biorthogonal to the original one, but shows time-frequency behavior similar to the original building blocks' behavior. In this manner a collection of overlapping patches can be constructed according to requested local resolution. Figures 6 and 7 illustrate this situation for a subfamily of building blocks in a Gabor family. Figure 6 shows the full sampling grid and the subset corresponding to the given local family, the analysis window and corresponding tight window. Figure 7 shows the operator corresponding to reconstruction with exactly and approximately biorthogonal families and a typical member of the respective families. Note that the approximate dual window is very similar to the original tight window. This means that inside the regions covered by one adaptive sub-family, the overall dual frame will have members similar to those of the dual frame of a regular Gabor analysis family, whereas at the boundaries modifications will be necessary.

## Conclusions

We have shown how mathematical tools can be used to facilitate and adapt the processing of audio signals. Future work will show the merits of the flexibility of an analysis of audio signals based on Gabor theory by adapting the properties of analysis window and grid to knowledge about the nature of audio and especially music signals.

## References

- Benedetto, J.J., Heil, C., & Walnut, D.F. (1998). Gabor systems and the Balian-Low theorem. In *Gabor analysis and algorithms* (pp. 85–122). Birkhäuser Boston, Boston, MA.
- Christensen, O. (1998). *Perturbations of frames and applications to Gabor frames*, volume Research monograph “Gabor Analysis and Algorithms: Theory and Applications” (pp. 193–209). Birkhäuser, Boston, 1st edition.
- Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Info. Theory*, 36, 961–1005.
- Dörfler, M. & Feichtinger, H.G. (1999). Quantitative description of expression in performance of music, using gabor representations. In H.G. Feichtinger & M. Dörfler (Eds.), *Proceedings of the Diderot Forum on Mathematics and Music: Computational and Mathematical Methods in Music* (pp. 139–144), Vienna.
- Feichtinger, H., Christensen, O., & Strohmer, T. (1995). Group theoretical approach to gabor analysis. *Optical Engineering*, 34(6), 1697–1704.
- Feichtinger, H. & Zimmermann, G. (1998). A Banach space of test functions for Gabor analysis. In H. Feichtinger & T. Strohmer (Eds.), *Gabor Analysis and Algorithms: Theory and Applications* (pp. 123–170). Birkhäuser, Boston.
- Feichtinger, H.G. & Strohmer, T. (Eds.) (1998). *Gabor Analysis and Algorithms: Theory and Applications*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston.
- Godsill, S. & Rayner, J. (1998). *Digital Audio Restoration, A statistical Model based Approach*. Springer-Verlag, London.
- Heil, C. & Walnut, D. (1989). Continuous and discrete wavelet transforms. *SIAM Review*, 31(4), 628–666.
- Janssen, A. & Strohmer, T. (2000). Characterization and computation of canonical tight windows for gabor frames. *J. Four. Anal. Appl.*, submitted.
- Kozek, W., Feichtinger, H., Prinz, P., & Strohmer, T. (1996). On multidimensional nonseparable Gabor expansions. In M. Unser, A. Aldroubi, & A. Laine (Eds.), *Proc. SPIE: Wavelet Applications in Signal and Image Processing IV*. to appear.
- Pielemeier, W.J. & Wakefield, G.H. (1996). A high-resolution time-frequency representation for musical instrument signals. *Journal of the Acoustical Society of America*, 99(4), 2382–2396.
- Qian, S. & Chen, D. (1996). *Joint Time-Frequency Analysis: Method and Application*. Prentice Hall, Englewood Cliffs, NJ.
- Teolis, A. (1998). *Computational Signal Processing with Wavelets*. Birkhäuser, Boston–Basel–Berlin.
- Zheng, Z. & Feichtinger, H. (2000). Gabor eigenspace time-variant filter. In *Proc. 2000 IEEE Electro/Information Technology Conference*, Chicago, USA.