ANALYSIS OF SOUND PATTERNS
THROUGH WAVELET TRANSFORMS†

R. KRONLAND-MARTINET

*Faculté des Sciences, Luminy, and
CNRS-LMA, 31 Chemin Joseph Aiguier
B.P. 71, 13277 Marseille Cedex 9
France*

J. MORLET

*O.R.I.C. 371bis, Rue Napoléon Bonaparte
92500 Rueil-Malmaison
France*

and

A. GROSSMANN

*Centre de Physique Théorique
Section II, CNRS-Luminy, Case 907
Marseille 13288, Cedex 9
France*

This paper starts with a brief discussion of so-called wavelet transforms, i.e., decompositions of arbitrary signals into localized contributions labelled by a scale parameter. The main features of the method are first illustrated through simple mathematical examples. Then we present the first applications of the method to the recognition and visualisation of characteristic features of speech and of musical sounds.

Keywords: Analysis; Synthesis; Wavelet transforms; Speech; Musical sound; Transient signal.

1. INTRODUCTION

The main purpose of the procedures that have been called wavelet transforms is to decompose arbitrary signals into localized contributions that can be labelled by a "scale parameter".

In order to obtain an intuitive understanding of these procedures, it is convenient to start by a look at the standard methods of smoothing by convolution.

If $s(t)$ is a signal (a real-valued function of the real argument t), then one can eliminate its small-scale features by taking its convolution with a suitable function.

Consider such a function g; we allow g to be complex-valued but assume $\hat{g}$, the Fourier transform of g, to be real-valued. For consistency with later definitions, we shall convolve s with the function g^* defined as $g^*(t)=\bar{g}(-t)$, where $\bar{g}$ is the complex conjugate of g:

$$(g^* * s)(b) = \int \bar{g}(t-b)s(t)dt = \int e^{ib\omega}\hat{g}(\omega)\hat{s}(\omega)d\omega. \qquad (1.1)$$

Here the notation adopted is:

$$\hat{g}(\pi) = (2n)^{-1/2} \int e^{-i\omega t} g(t) \mathrm{d}t.$$

Let $\hat{g}(\omega)$ be negligible above a certain frequency ω_{max}, i.e., $\hat{g}(\omega) \approx 0$ if $\omega| >_{max}$. Then the function (1.1) is insensitive to the Fourier components $\hat{s}(\omega)$ with $|\omega| > \omega_{max}$. This is the elimination of small-scale features. Furthermore, let us assume that g is negligible outside an interval $[t_{min}, t_{max}]$ of the t-axis; then (1.1) is insensitive to the values $s(t)$ such that $t - b$ lies outside of $[t_{min}, t_{max}]$. This is localization (or better, concentration) in time.

Under the conditions above, the problem of recovering $s(t)$ from (1.1) is clearly ill-posed.

We can now describe the main ideas involved in wavelet transforms. One starts with (1.1), and performs the following three additional steps.

(i) *First step:* Choose g in such a way that it eliminates not only "small-scale features" of s (i.e. Fourier components with $|\omega| > \omega_{max}$) but also "sufficiently large-scale features of s" (i.e. Fourier components with $|\omega| < \omega_{min}$; of course $0 < \omega_{min} < \omega_{max} < \infty$). In other words, assume that $\hat{g}(\omega) \approx 0$ if $|\omega| < \omega_{min}$. The function (1.1) depends then only on features of $s(t)$ that are neither too fine nor too gross. Such a localization in scale can be obtained without excessive sacrifice of concentration in time.

It is important to realize, however, that this assumption on g runs against the intuitive idea that (1.1) should be viewed as a local average of the values of $s(t)$. Indeed, the absence of zero-frequency components in g entails $\int g(t) \mathrm{d}t = 0$. If a signal is essentially constant between t_{min} and t_{max}, then (1.1) is negligible even if the constant value of the signal is large, and so (1.1) is nowhere near the mean of s. Intuitively one should then think of (1.1) as a local average of some derivative of $s(t)$. This picture can be made precise for suitable choices of g.

(ii) *Second step:* Notice that there exists a perfectly natural way of changing the scale examined by g. Introduce a positive "scale parameter" a. Together with g consider the whole family of re-scaled functions

$$g_a(t) = (1/\sqrt{a})\, g(t/a),\ (a > 0)$$

(The factor $1/\sqrt{a}$ ensures that $\int |g(t)|^2 \mathrm{d}t = \int |g_a(t)|^2 \mathrm{d}t$).

To any signal $s(t)$, associate now the function $S(b,a)$ on the open (b,a) half-plane (b arbitrary, $a > 0$):

$$S(b,a) = (g_a^* * s)(b) = 1/\sqrt{a} \int \bar{g}((t-b)/a) s(t) \mathrm{d}t = \sqrt{a} \int \hat{g}(a\omega) \hat{s}(\omega) e^{ib\omega} \mathrm{d}\omega \qquad (1.2)$$

$$\approx \sqrt{a} \int_{\omega_{min}/a}^{\omega_{max}/a} \hat{g}(a\omega) \hat{s}(\omega) e^{ib\omega} \mathrm{d}\omega \approx 1/\sqrt{a} \int_{at_{min}+b}^{at_{max}+b} \bar{g}((t-b)/a) s(t) \mathrm{d}t\ .$$

The crucial point is the non-zero lower limit ω_{min}/a in one of the integrals (1.2); this lower limit is there because of (i).

For fixed a, the function (1.2) depends on features of $s(t)$ given by frequencies in the range between ω_{min}/a and ω_{max}/a. A small value $a=a_0$ of the scale parameter allows us to look at fine detail of s, but $S(b,a_0)$ is then quite insensitive to slow variations of s.

Furthermore, the value of $S(b,a)$ for a given b depends effectively only on the values of $s(t)$ in the interval $[b^*at_{min},b+at_{max}]$. If a is small, this interval is short.

(iii) *Third step:* Realize that the function $S(b,a)$ on the (b,a) half-plane, (which describes the signal at all scales) has properties that make it a practical tool for the analysis of s. We list here some of these properties; they will be described in more detail below.

(a) The total energy of the signal, namely $\int s(t)^2 dt$, can also be read off from $S(a,b)$. Except for a multiplicative constant, independent of s, it is:

$$\iint |S(b,a)|^2 (1/a^2) db da .$$

(b) The signal $s(t)$ can be reconstructed from $S(b,a)$ by straightforward formulas and the reconstruction procedure is quite stable, in contrast to reconstruction from $S(b,a_0)$ for fixed a_0. Moreover $s(t)$ can be reconstructed from just the values of $S(b,a)$ at points of a suitable grid of the (b,a) plane.

The function $S(b,a)$ defined by (1.2) is called the wavelet transform of the signal $s(t)$ with respect to the *analyzing wavelet* $g(t)$.

Sound signals are a natural domain for the application of wavelet transform techniques. We present here the first results on wavelet transform analysis of both computer-generated and natural sounds.

2. BASICS OF WAVELET TRANSFORMS

In this section we shall give a quick review of the main formulas, and illustrate them by simple examples. For proofs, details, other applications, and different points of view, see the references.

2.1. Analyzing Wavelets

A function g satisfying the conditions described in the introduction will be called an *analyzing wavelet*. More precisely, the minimal requirements for the label "analyzing wavelet" are":

$$\int |g(t)|^2 dt < \infty \tag{2.1}$$

(finite energy), and

$$\int |\hat{g}(\omega)|^2 d\omega/|\omega| < \infty , \tag{2.2}$$

(if $\hat{g}(\omega)$ is smooth, then (2.2) implies the condition $\hat{g}(0)=0$ of the introduction).

In practice, one often requires more than just (2.1) and (2.2): some smoothness of g and $\hat{g}$, and simultaneous concentration of g and $\hat{g}$, not too far from the limit imposed by the uncertainty principle.

A plane wave is not an analyzing wavelet. It extends to infinity and oscillates too much. A strictly positive function $g(t)$, e.g. a bump, cannot be an analyzing wavelet either since it cannot satisfy the condition $\int g(t)dt=0$.

In this paper, we shall explicitly make the following additional assumption:

$g(t)$ contains only positive-frequency components i.e.,
$\hat{g}(\omega)=0$ for $\omega<0$. $\qquad\qquad$ (2.3)

It follows from (2.3) that the real and imaginery parts of $g(t)$ are Hilbert transforms of each other.

All the numerical examples shown in this paper were computed with analyzing wavelets that have Fourier transform of the form:

$\hat{g}(\omega)=$const.$\exp(-(\omega-\omega_0)^2/2)$ + small corrections $\qquad$ (2.4)
(gaussian bump centered at ω_0).

The correction terms are theoretically necessary since $\exp(-(\omega-\omega_0)^2/2)$ does not vanish for $\omega\leq 0$, and so does not satisfy (2.2) and (2.3). The choice of ω_0 in this paper (between 5.0 and 6.0) makes however such correction terms negligible, and we shall not bother to write them down. Consequently, up to a constant normalization factor and negligible correction terms, our analyzing wavelet is the modulated gaussian

$$g(t)=e^{i\omega_0 t}.\exp(-t^2/2).$$

Its real part, imaginary part and modulus are shown in Figs. 1a-c for various values of the scale parameter.

2.2. Definition and Main Properties of the Wavelet Transform

If g satisfies the conditions of Sect. 2.1 and if $s(t)$ is an arbitrary (continuous-time, real-valued, deterministic) signal, then the wavelet transform of s with respect to g is defined as the following function $S(b,a)=S_g(b,a)$ on the open (b,a) half-plane (b arbitrary, $a>0$):

$$S(b,a)=(1/\sqrt{a})\int g-((t-b)/a)s(t)dt= \sqrt{a}\int_0^\infty \hat{g}(a\omega)\hat{s}(\omega)e^{ib\omega}d\omega. \qquad (2.5)$$

It is the scalar product of s with the function $g^{(b,a)}(t)$ obtained from $g(t)$ by re-scaling by a, and shifting by b.
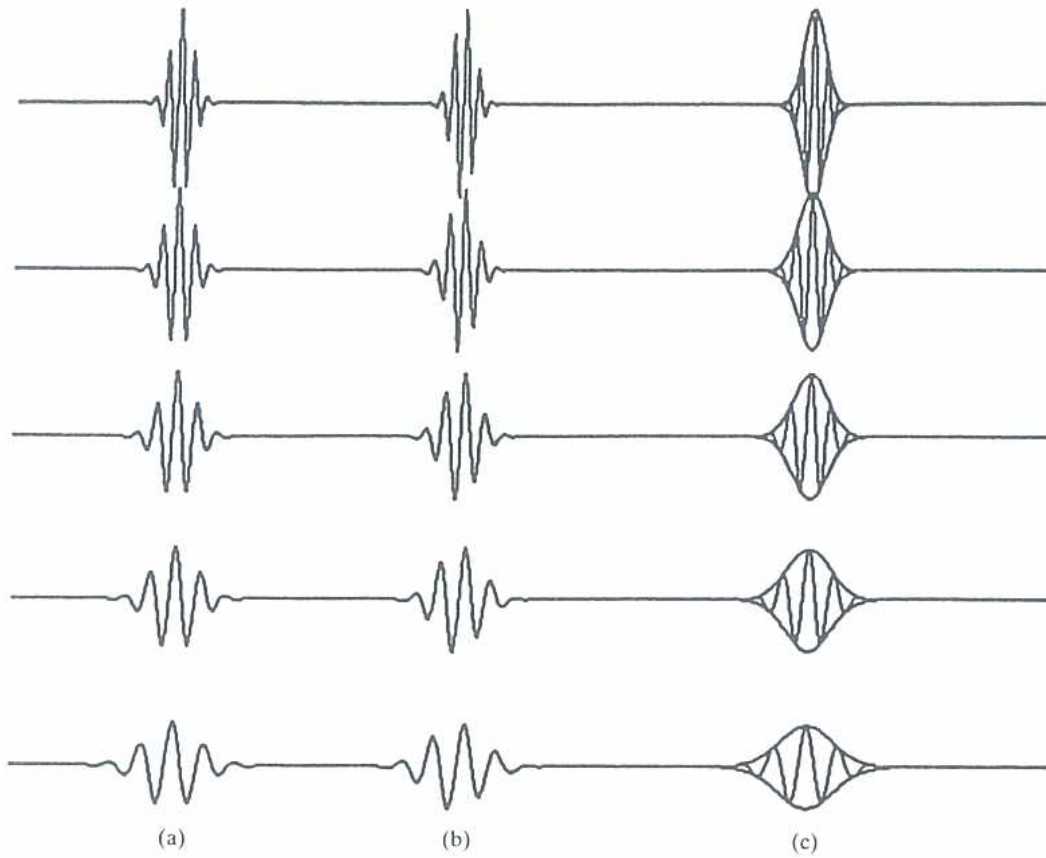
Fig. 1. An analyzing wavelet at various values of the scale parameter (a) real part, (b) imaginary part and (c) modulus and real part.

The main properties of the correspondence between s and S are:

a) Independence of choice of origin of time
If $s(t)$ is shifted in time by t_0 ($s(t) \rightarrow s(t-t_0)$), then $S(a,b)$ is transformed into $S(b-t_0,a)$.

b) Conservation of energy
It is convenient to introduce the positive number

$$c_g = 2\pi \int_0^\infty (1/\omega)\hat{g}(\omega)^2 d\omega .$$
(2.6)

which is finite by (2.2).

Let $s(t)$ be a signal of finite energy: $\int s(t)^2 dt < \infty$; then

$$\int s(t)^2 dt = (1/c_g) \int\int |S(b,a)|^2 1/a^2 da \, db .$$
(2.7)

101

At this point it is useful to point out an apparent paradox: Consider a sequence $s_n(t)$ of signals of finite energy which tend to a function that is constant: $s(t) = $ const. $\neq 0$. Then the energies on the l.h.s. of (2.7) tend to $+\infty$. On the other hand, the wavelet transform of the function $s(t) = $ const. is identically zero, since $\int g(t)dt = 0$: so we obtain $+\infty = 0$. We invite the reader to improve his understanding of the transform (2.5) by resolving this paradox.

c) Sampling of the transform by reproducing kernel

The sampling (evaluation) of an arbitrary continuous function of one or more variables is commonly written as the scalar product of this function with an appropriate δ-function of Dirac. If the function to be evaluated is not arbitrary, then it may happen that its evaluation at an arbitrary point can be obained by scalar products not only with a δ-function but also with another, smoother, function. A good example for this is provided by band-limited functions. A point evaluation of such a function can be obtained either with the help of a δ-function or with the help of an appropriate sinc-function.

Let g, the analyzing wavelet, be fixed. Then every (reasonable) signal $s(t)$ gives rise by (2.5) to a function on the open (b,a)-half-plane, but an arbitrary function on this half-plane is not, in general, the transform of some s. Being "somebody's transform" imposes constraints which are practically important. They can be described as follows:

Given g, consider on the (b,a)-half-plane the function

$$p(b,a)=(1/c_g)(1/\sqrt{a})\int \bar{g}(t-b)/a)g(t)dt \tag{2.8}$$

where c_g is defined by (2.6).

Then a function $S(b,a)$ is the wavelet transform (with respect to g) of some signal s, if and only if it satisfies, for every point b_0,a_0 in the (b,a) half-plane, the condition

$$S(b_0,a_0)=\int\int p((b_0-b)/a,a_0/a)S(b,a)(1/a^2)dadb. \tag{2.9}$$

This is an evaluation by reproducing kernel, of the type described above. The kernel p in (2.9) is significantly different from zero only if the point (b,a) is in a suitable neighbourhood of (b_0,a_0). For the wavelet (2.4), the modulus of the function $c_g p(b,a)$ is

$$(2\pi)^{1/2}(a/(a^2+1))^{1/2}\exp\{-(1/2)[b^2+\omega_0^2(a-1)^2]/(a^2+1)\}.$$

The phase is

$$b\omega_0(a+1)/(a^2+1).$$

So (2.9) describes correlations between values of S at a scale given by p. Because of these correlations, $S(a,b)$ is entirely determined by its values on suitable grid of the (b,a) half-plane. The qualitative features of such a grid can be seen in Fig. 2.
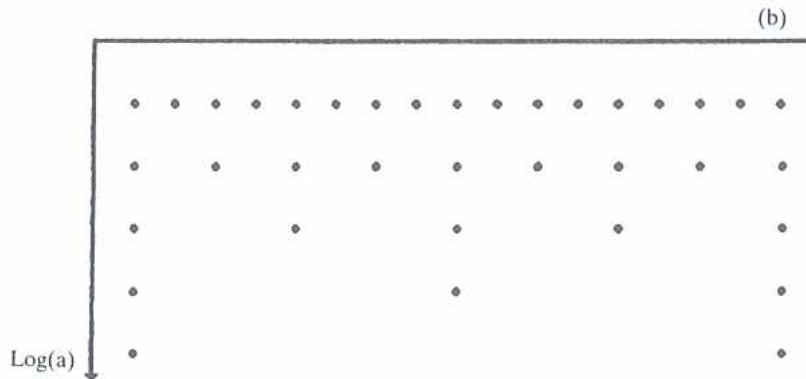
Fig. 2. Schematic representation of a grid on the (b,a) plane, allowing arbitrarily precise reconstruction of signals.

2.3. Reconstruction Formulas

A signal s can be recovered from its transform S with the help of the formula

$$s(t)=(1/c_g)\iint(1/\sqrt{a})g((t-b)/a)S(b,a)(1/a^2)\mathrm{d}a\ \mathrm{d}b. \qquad (2.10)$$

There are many other exact reconstruction formulas, some involving the values of S on a vertical line $b=$const. and others involving only the values of S on a suitable grid. The discussion of these formulas would take us too far from the main subject of this paper, which is the use of S as a "generalized sonogram". We refer, again, to the literature quoted in the references.

2.4. Detection of Discontinuities in the Signal or in Its Derivatives

Abrupt changes in s or in its derivatives can under suitable circumstances be clearly seen on its transform $S(a,b)$. This is a point with important implications, and will be illustrated in the pictures below. The main qualitative features are:

1) A discontinuity is signaled by a localized increase in the modulus $|S(a,b)|$ for small a, around a point b_0 (a discontinuity contains high frequencies).

2) A discontinuity is also signaled by the convergence of lines of constant phase towards a point at the edge of the (b,a) half-plane. In order to get an intuitive feeling for this, consider first the case where $s(t)=\delta(t)$ (Dirac's function concentrated at $t=0$). Then, by (1.1), $S(b,a)=(1/\sqrt{a})\ g(-b/a)$. Since a is real, the phase of $S(b,a)$ is constant along the straight lines $b/a=$const., coverging towards the point $b=0$, $a=0$, on the edge of the open (b,a) half-plane. One should realize next that this result depends only on the fact that $\delta(t)$ is homogeneous, $(\delta(\lambda t)=\lambda^{-1}\delta(t)$ for $\lambda>0)$ and that it remains true for any $s(t)$ that satisfies a homogeneity condition of the form $s(\lambda t)=\lambda^\gamma s(t)$, which implies, in (2.5), that $\hat{s}(a\omega)=a^{-\gamma-1}\hat{s}(\omega),a>0$, δ real. An example of such a function is an $s(t)$ which vanishes for $t\leq0$ and which is equal to t^γ fot $t>0$. Now, locally, (i.e. in a neighbourhood of $t=0$) this is precisely the behaviour of a function that has a

103

discontinuous derivative. In any true signal, this behaviour will be only approximate, and around certain points; the choice of $t=0$ in the above examples is of no special significance, by the independence of the choice of origin of time, discussed in Sect. 2.2.

These statements apply equally well to s and to its derivatives. The wavelet transform (with respect to g) of a derivative s is a times the wavelet transform of s with respect to g.

3. IMPLEMENTATION OF WAVELET ANALYSIS OF SOUNDS

The methods described above has been implemented in a real-time signal processor SYTER (Fig. 3).
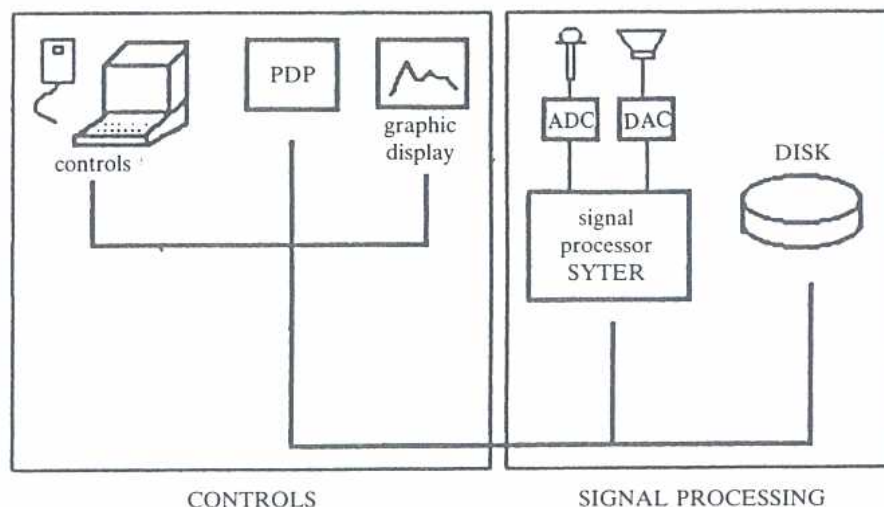
Fig. 3. Schematic description of the SYTER processor.

The system consists of a control part and of a part for real-time signal analysis. The controls consists of a set of I/O peripherals and of a PDP 11/73 host processor. The real-time part includes the SYTER processor, the A/D and D/A converters and the mass storage system.

The above components are connected by a bus.

The system—fully programmable and interactive—has been used to calculate wavelet transforms of signals, as follows:

First of all, the signal to be analyzed is sampled at a sampling rate of 32 kHz and stored on disk.

Next, for every chosen value of the scale parameter, (i.e. for each "voice"), the host processor calculates the sampled values $g_a(n)$ of the analyzing wavelet, and transmits them to the real-time signal processor.

This processor performs then the transform in real time, for each voice, and stores the result on a disk.

The processing of the results is then performed by the host processor. The graphic representations are then available at a graphic terminal.

The structure used in the processing of the signal is of transverse filter type. The convolution described above is

$$X(n) = \sum_{i=0}^{N-1} g_i s(n - i - 1)$$

where $X(n)$ is the n-th output sample, g_i the i-th value of the filter response (here the wavelet) and $s(n)$ the n-th sample of the input signal.

Such a structure requires N memory locations for data. For each output number, one has to perform N multiplications and $N-1$ additions (Fig. 4).
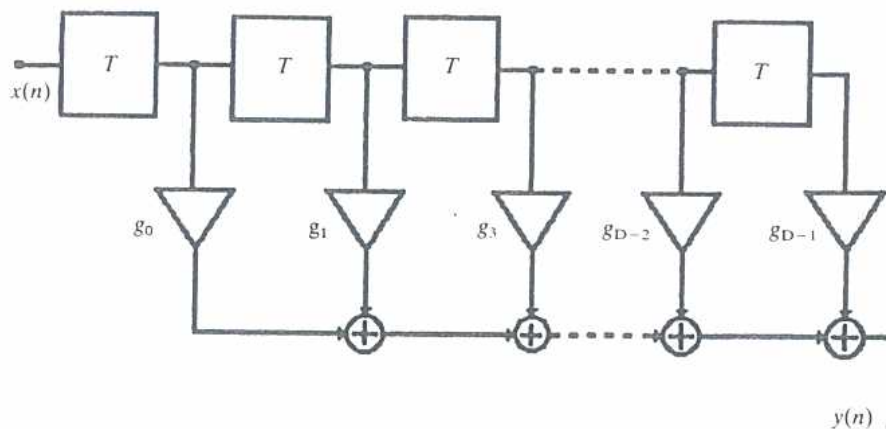
$$y_n = \sum_{i=0}^{N-1} g_i x(n - i - 1)$$

Fig. 4. Structure of the filter used for transformation.

Note that the SYTER systems allows the reconstruction of signals from their transform, and the listening to this reconstruction. This enables us to check the acoustic relevance of various approximations.

The availability of reconstruction formulas enables us to subject the signal—while transformed—to intimate modifications that are of interest in many fields, and in particular in the composing of electronic music.

4.   DISCUSSION OF SELECTED EXAMPLES

This section contains comments of pictures on wavelet transforms of a variety of signals, ranging from δ-functions to speech and to notes on a clarinet. The more academic examples (calculated on the system described above and on an IBM PC-compatible microcomputer) can be usefully compared to their real-life counterparts.

4.1.   Description of Graphic Conventions

The pictures below are representations of complex-valued functions $S(b,a)$ on the open (b,a) half-plane.

The positive b-axis points to the right and the a-axis points downward so that high frequencies (small a) are above low frequencies.

It is useful to represent separately the modulus and the phase of the complex-valued function $S(b,a)$ to be described. Both quantities are represented by shades of gray obtained through appropriate random printing of black dots. Furthermore, some phase pictures are cut off when the modulus is smaller than a prescribed number, and so provide also some information about the modulus.

4.2. δ-functions, Pure Harmonic Signals and Reproducing Kernels

A δ-function positioned at $t=t_0$ gives rise to the transform

$$S(b,a) = a^{-1/2} g((t_0-b)/a) \tag{4.1}$$

while the harmonic signal $\cos(\omega_0 t)$ has transform

$$S(b,a) = (1/2)a^{1/2}\hat{g}(\omega_0 a) \exp(i\omega_0 b). \tag{4.2}$$

Linear combinations of such functions show the expected interference effects. Figures 5a and 5b show the modulus and phase of the transform of a δ-function.

Notice the convergence of the straight lines of constant phase for the δ-function. This should be contrasted with the lines of constant phase of the reproducing kernel (Fig. 6c) which become parallel for small values of a.

The phase of the wavelet transform of a single-frequency signal shows the expected regular behaviour (Fig. 7).

It is quite instructive to look at the phase picture of the transform of a signal that is the sum of two harmonics; we choose the frequency ratio 1/1.6 (Fig. 8). The points where new lines arise are necessarily zeroes of the modulus; they correspond to destructive interference of terms in (4.2).

The constructive and destructive interference between transforms of two δ-functions is shown in Figs. 9a and 9b. The moth-eaten appearance of 9b is again due to the presence of zeroes of the modulus.

Finally Figs. 10 describe the transform (with cutoff on the modulus) of two weighted δ-functions together with two weighted cosines.

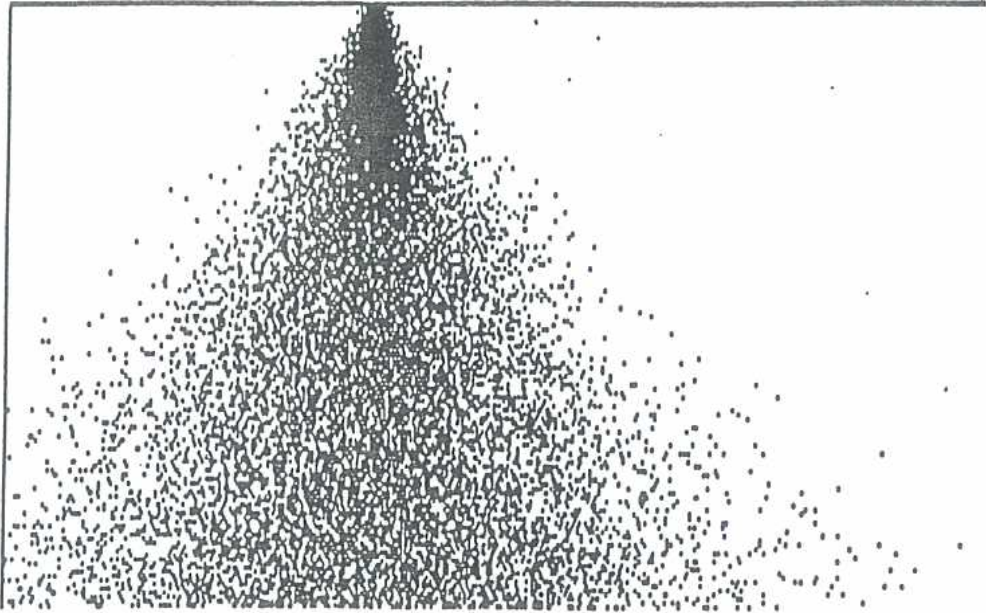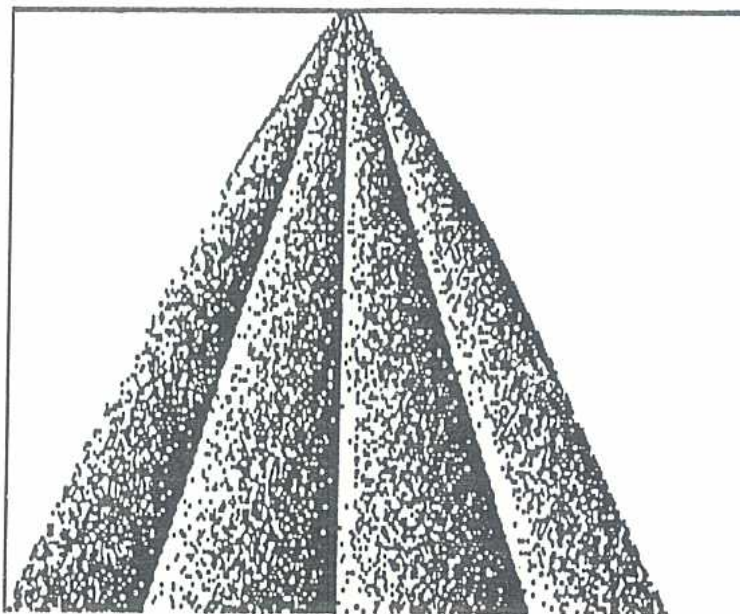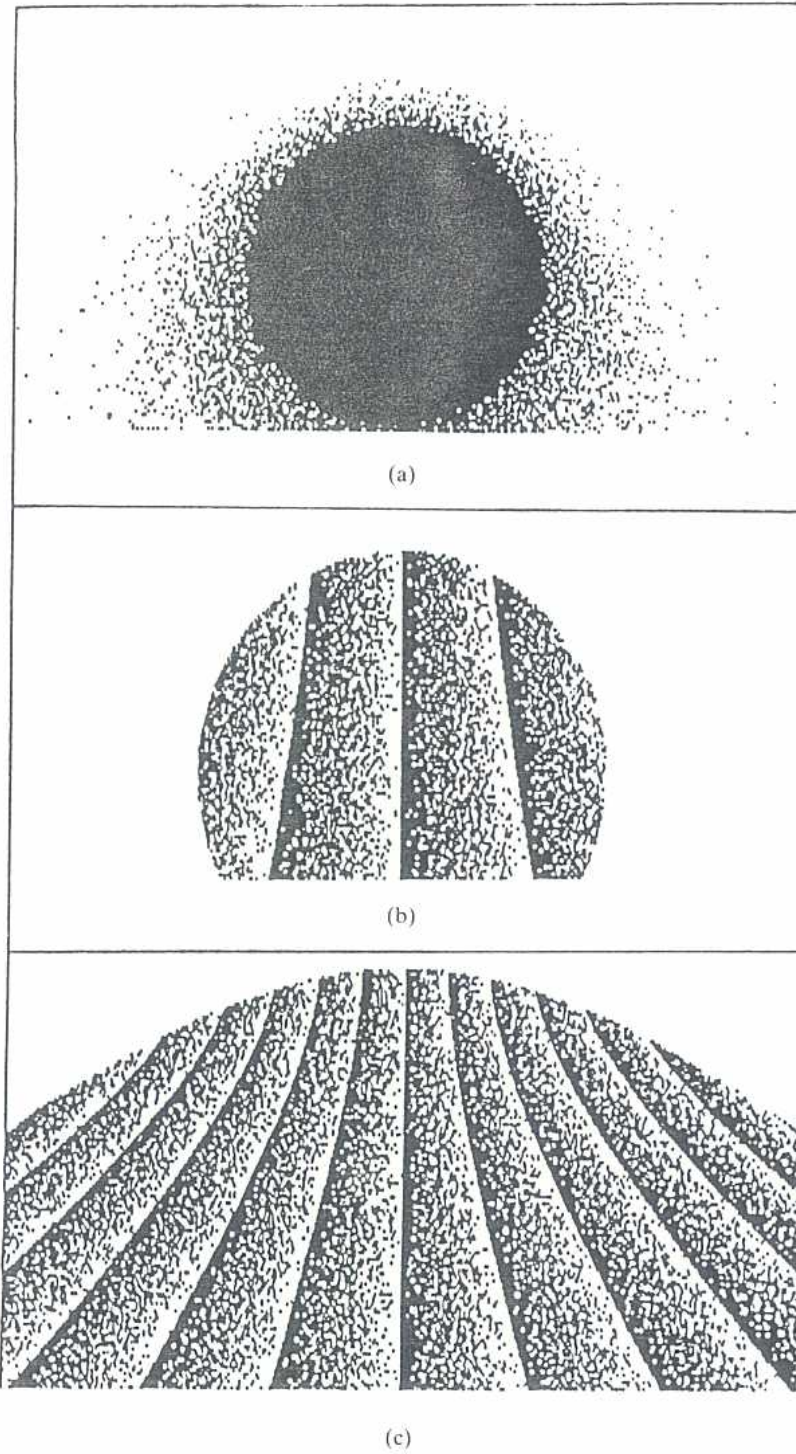Fig. 5a. Modulus of the transform of a δ-function.

Fig. 5b. Phase of the transform of a δ-function, with a cutoff on the modulus.

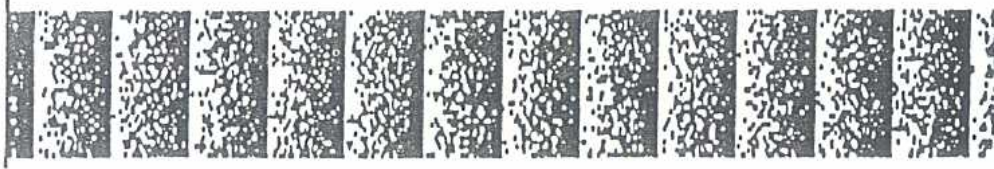Figs. 6. Reproducing kernel: modulus (a) and phase: (b), (c).

Fig. 7. Phase of the transform of a single-frequency signal. The corresponding modulus is constant on lines of constant a.
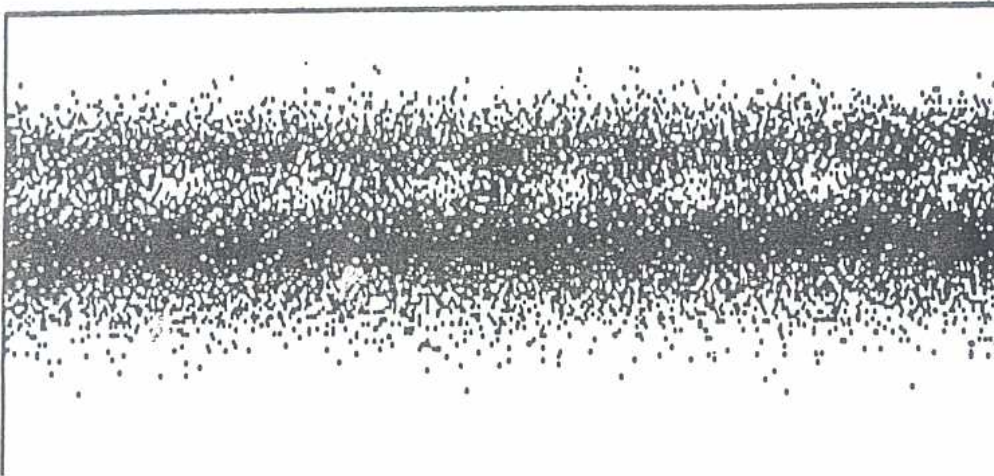

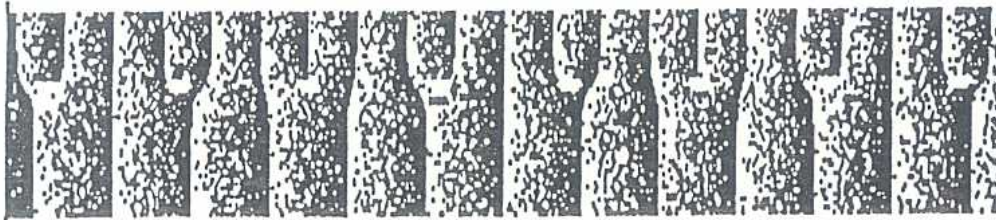Fig. 8a. Modulus of the transform of a two-frequency signal. The ratio of the frequency is 1.6. One sees the destructive interferences (with holes) that also appear on the phase picture 8b.

Fig. 8b. Phase of the transform of the two-frequency signal of Fig. 8a, with cutoff on the modulus.
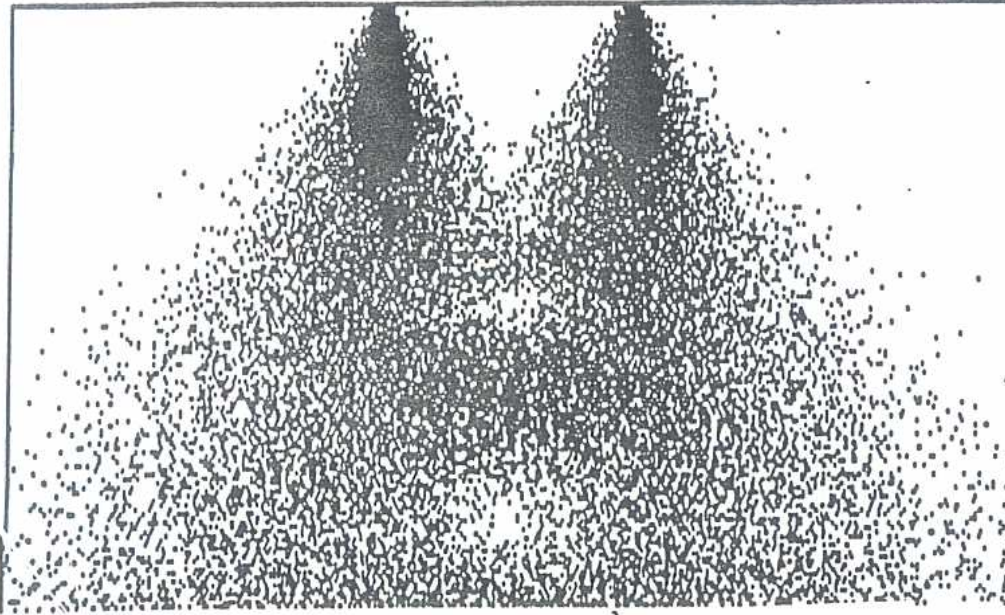
Fig. 9a. Modulus of the transform of the sum of two δ-functions of same weight.
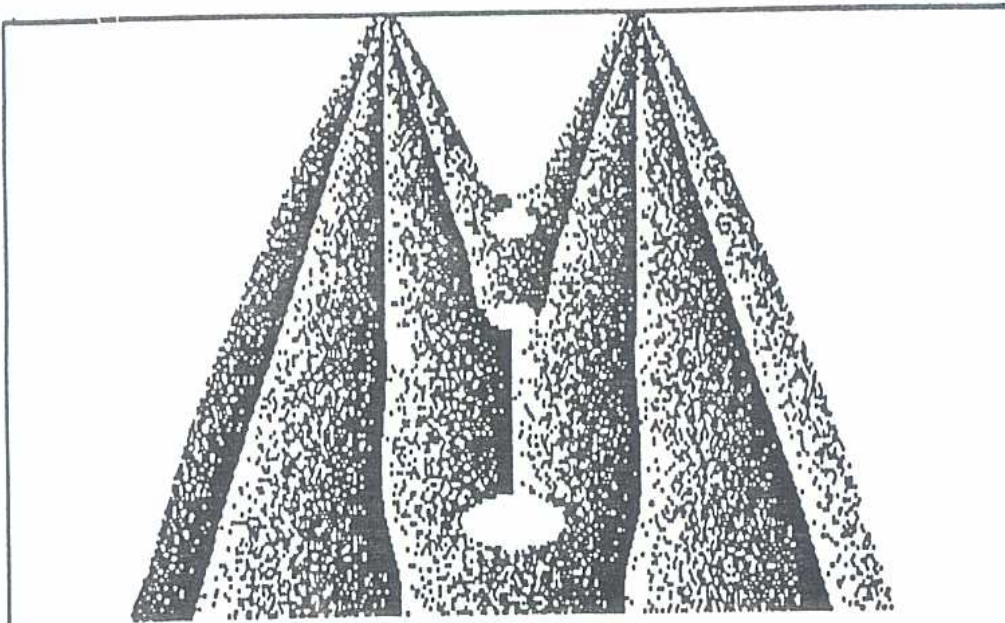
Fig. 9b. Phase of the transform of Fig. 9a, with cutoff on the modulus.
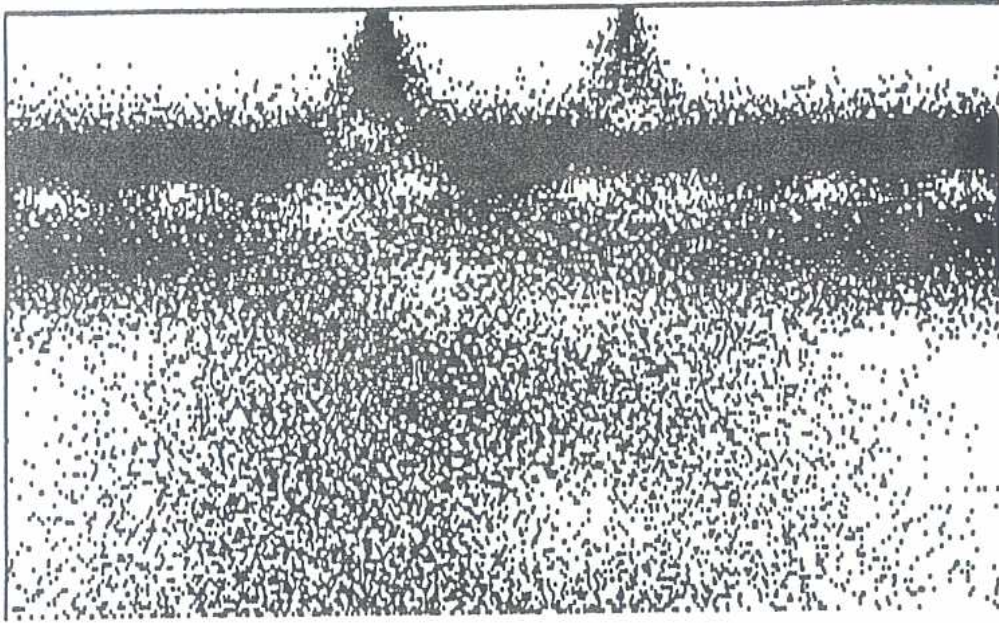
Fig. 10a. Modulus of the transform of the sum of two weighted δ-functions (ratio of weights 1:2) and of two weighted cosines (frequency ratio 1:1.6, ratio of weights 1:2).
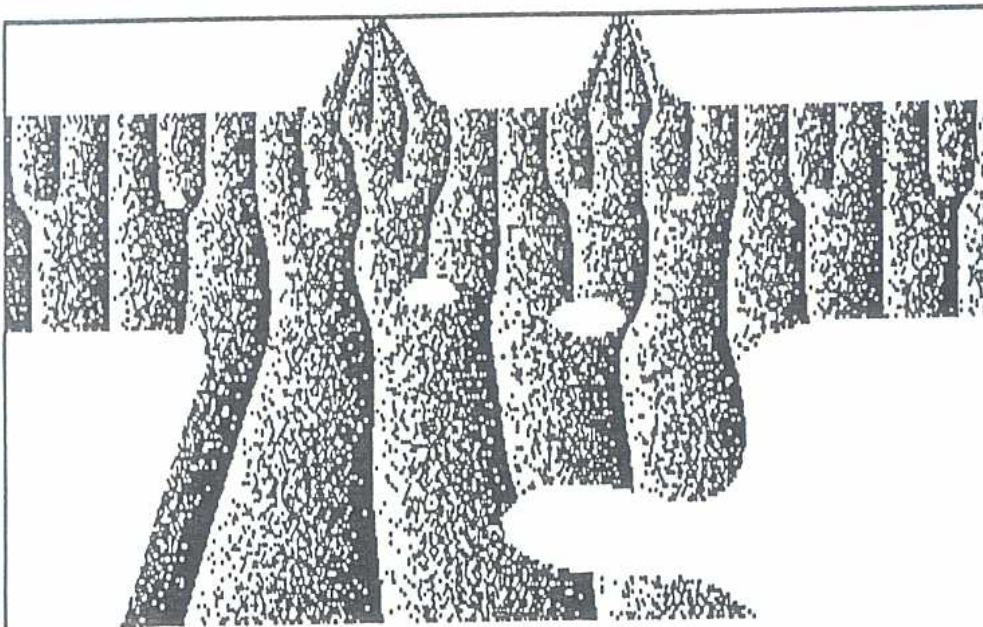
Fig. 10b. Phase of Fig. 10a, with cutoff on modulus.

111

The pictures to be shown from now on were calculated on the SYTER. From now on, we shall use the logarithmic scale on the *a*-axis, and write the phase pictures without cutoff on the modulus. The signal will appear above the transform, except in Fig. 11a.
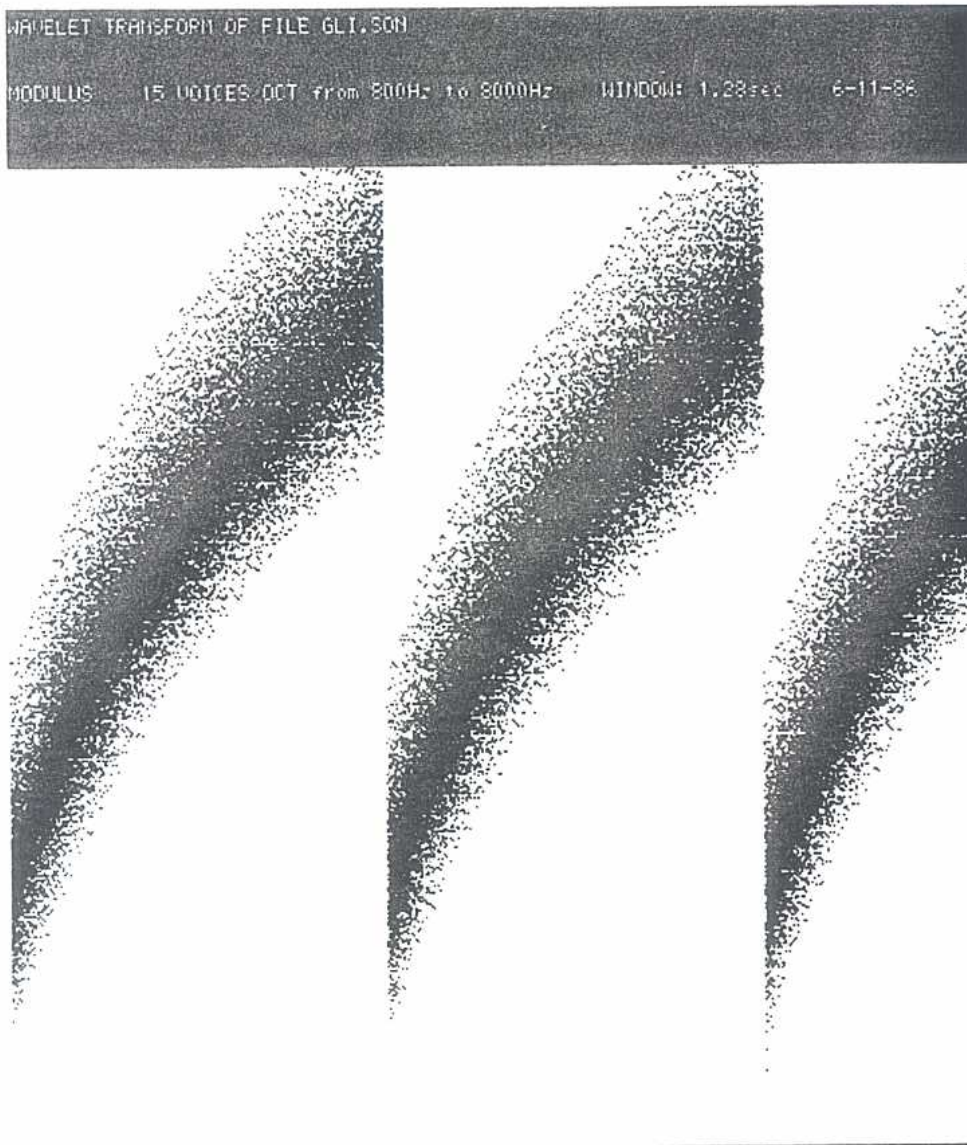
Fig. 11a. Modulus of the transform of three linear frequency sweeps. The total time is 1.28 seconds. Because of the proportionality of frequency to (1/*a*) and of the logarithmic scale in the *a*-direction, the lines of maximum modulus are curved.

4.2. A Chirp

Figures 11a to 11c represent the modulus of the transform of a sawtooth sweep over frequencies. The picture 11a corresponds to a total time of 1.28 seconds. The very sharp transition between frequencies can be seen in Figs. 11b and 11c which also shows the beginning of the rise in frequency.

The phase picture (Fig. 11c) is quite instructive. It shows the phase convergence, discussed previously, at the transition point.
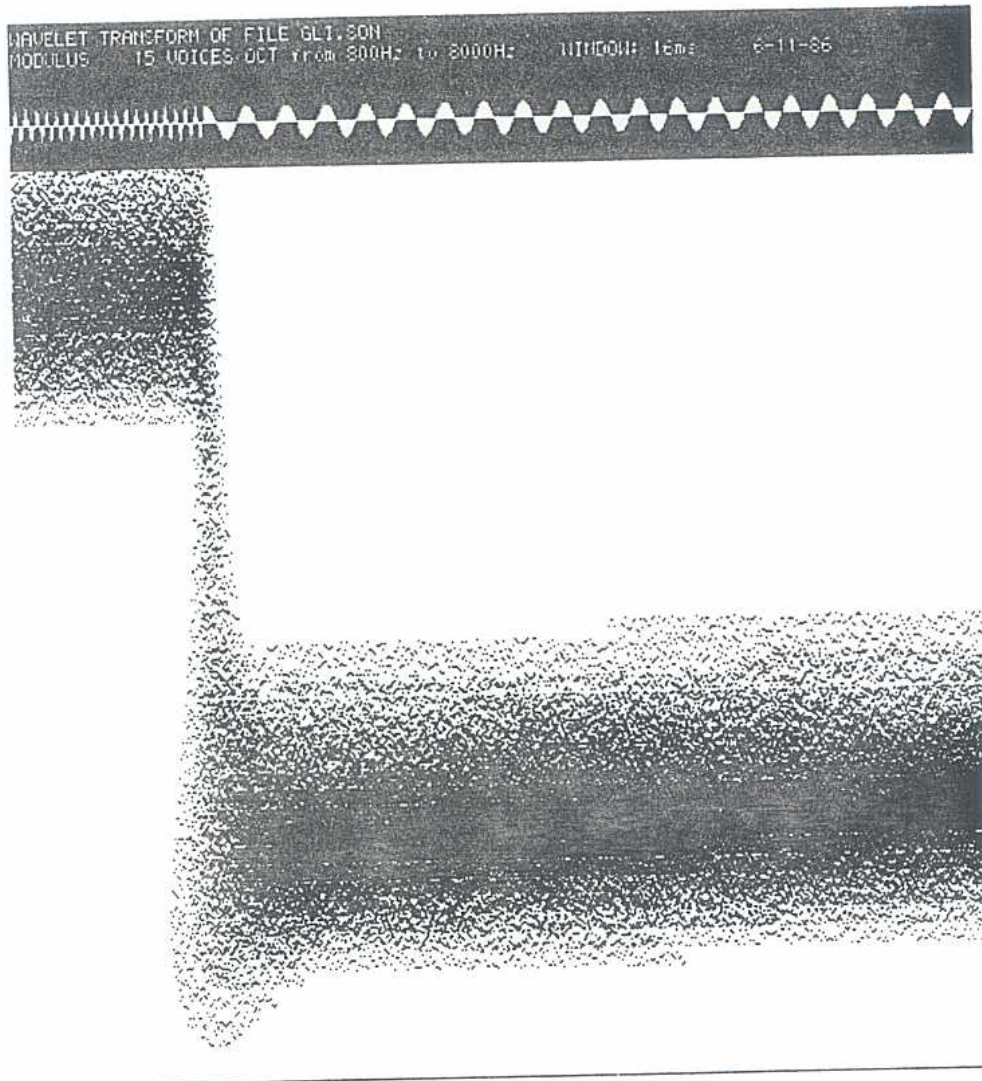
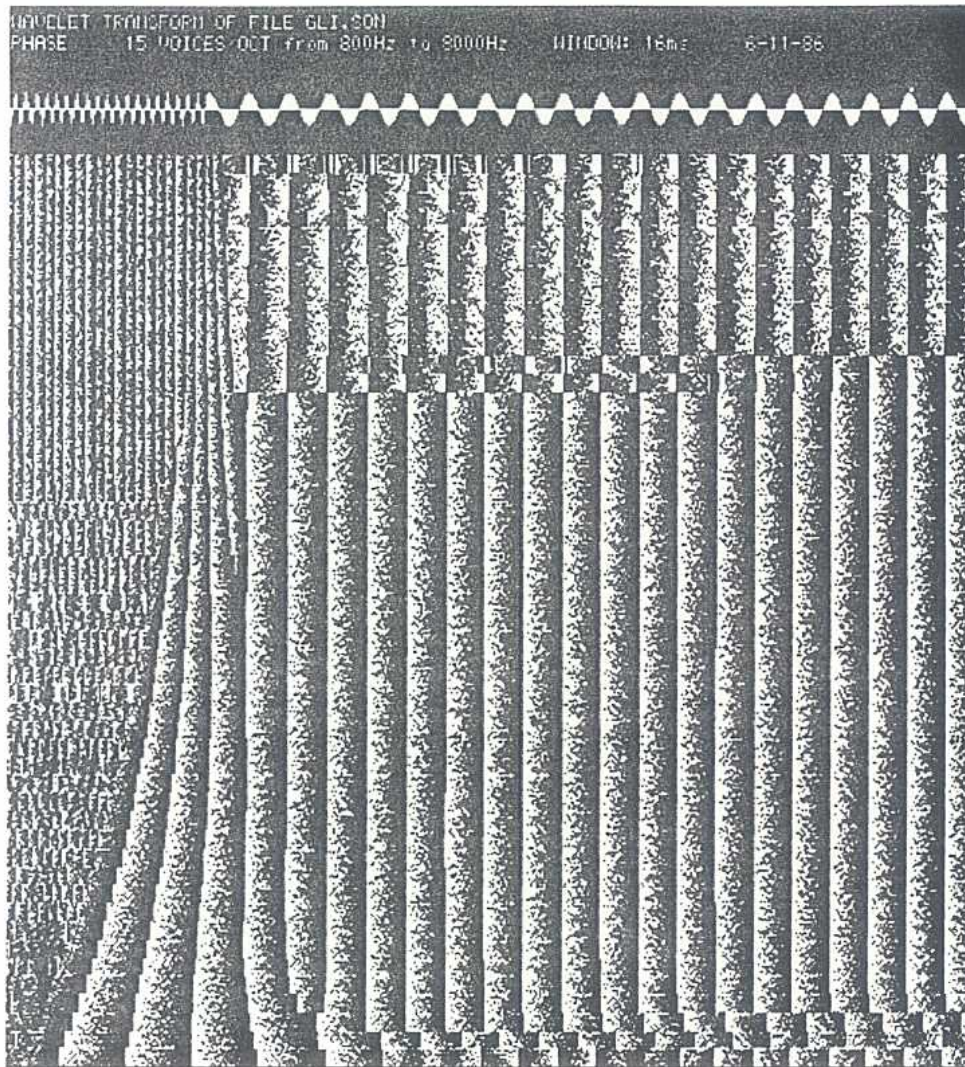Fig. 11b. Zoom of Fig. 11a to a total time of 16 ms, showing one of the transitions.

Fig. 11c. Phase picture corresponding to Fig. 11b and showing convergence of lines of constant phase toward the transition point.

4.3. Speech

We now leave computer-generated signals. Figure 12a is a modulus picture, over 0.16 seconds, of "pap" in "papy". The position of the maximum of the analyzing wavelet in frequency space ranges from 50 Hz (bottom of picture) to 8000 Hz (top of picture). The explosion of the "p", over 32ms, is shown in Fig. 12b (modulus), again between 50 and 8000 Hz. The same picture, in phase, (Fig. 12c) shows convergence of phase lines at the beginning of sound.

For the sake of comparison, Figs. 13a and 13b show the "ta" sound of "taty", during 0.128 seconds and the first 32 ms (modulus).

Fig. 12a. Modulus of the transform of "PAP". One can see, at low frequencies, the two explosions of "P" sound; the vowel "A", between the two "P"s, shows its resonant frequencies.
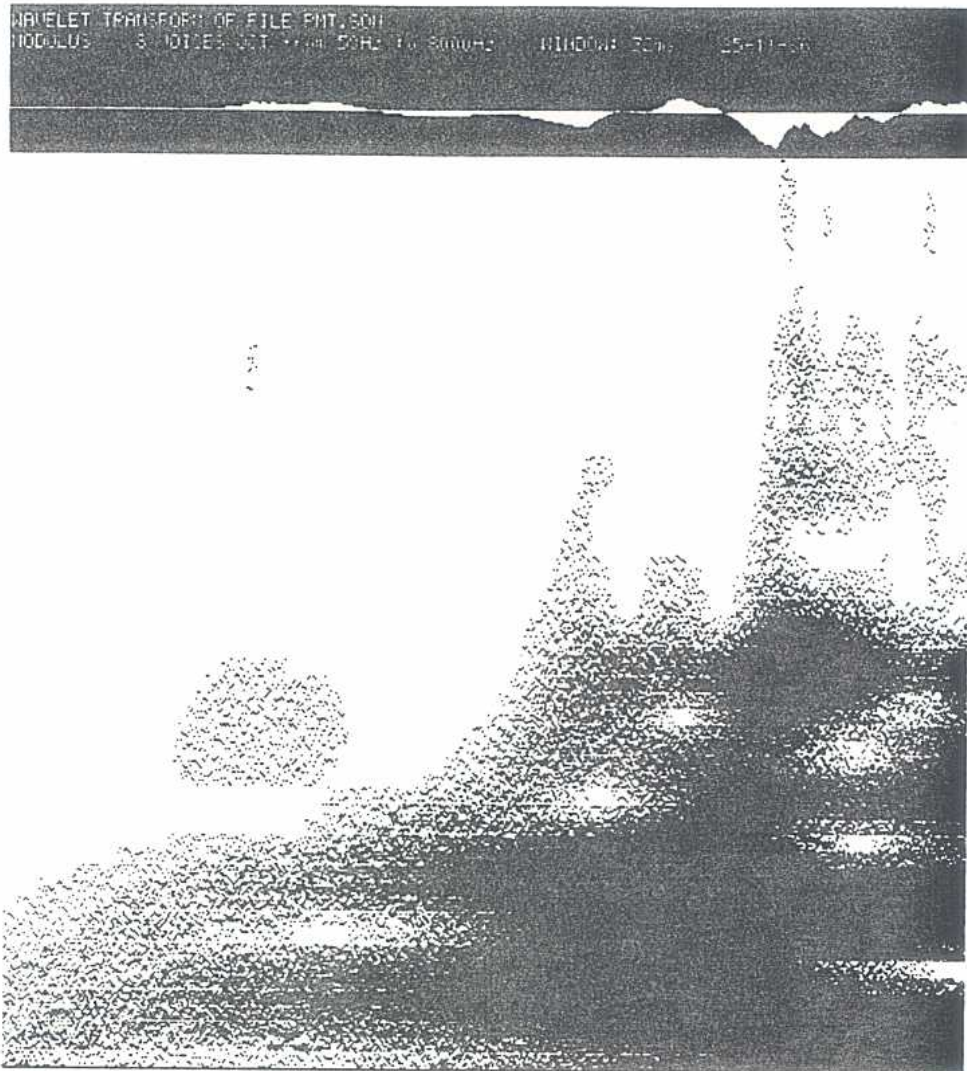
Fig. 12b. Modulus of the transform of the beginning of the first "P" of Fig. 12a, over 32 ms; with scale parameter corresponding to the range between 50 and 8000 Hz.
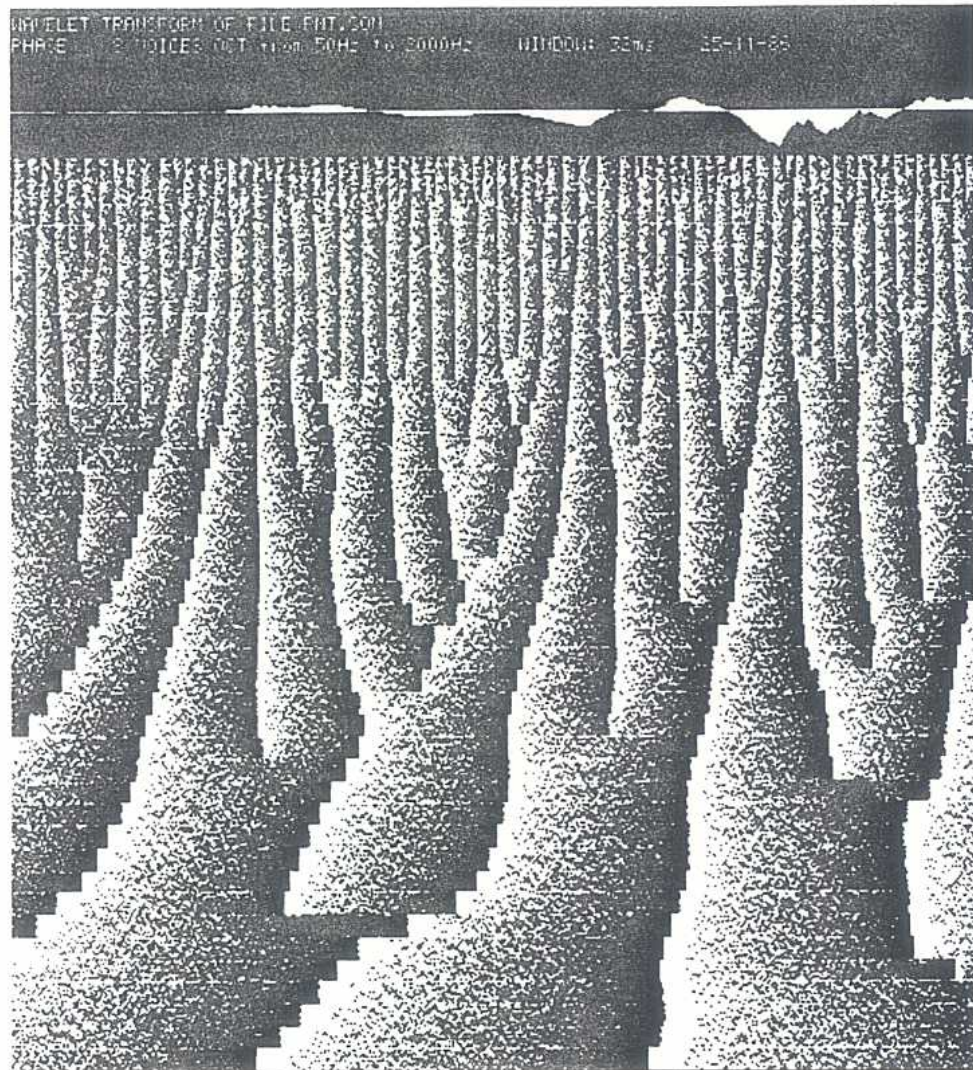
Fig. 12c. The same as Fig. 12b, in phase. The upper limit is now 2000 Hz. The points where new lines of constant phase originate correspond to zeroes of the modulus (compare, e.g., Figs. 8).
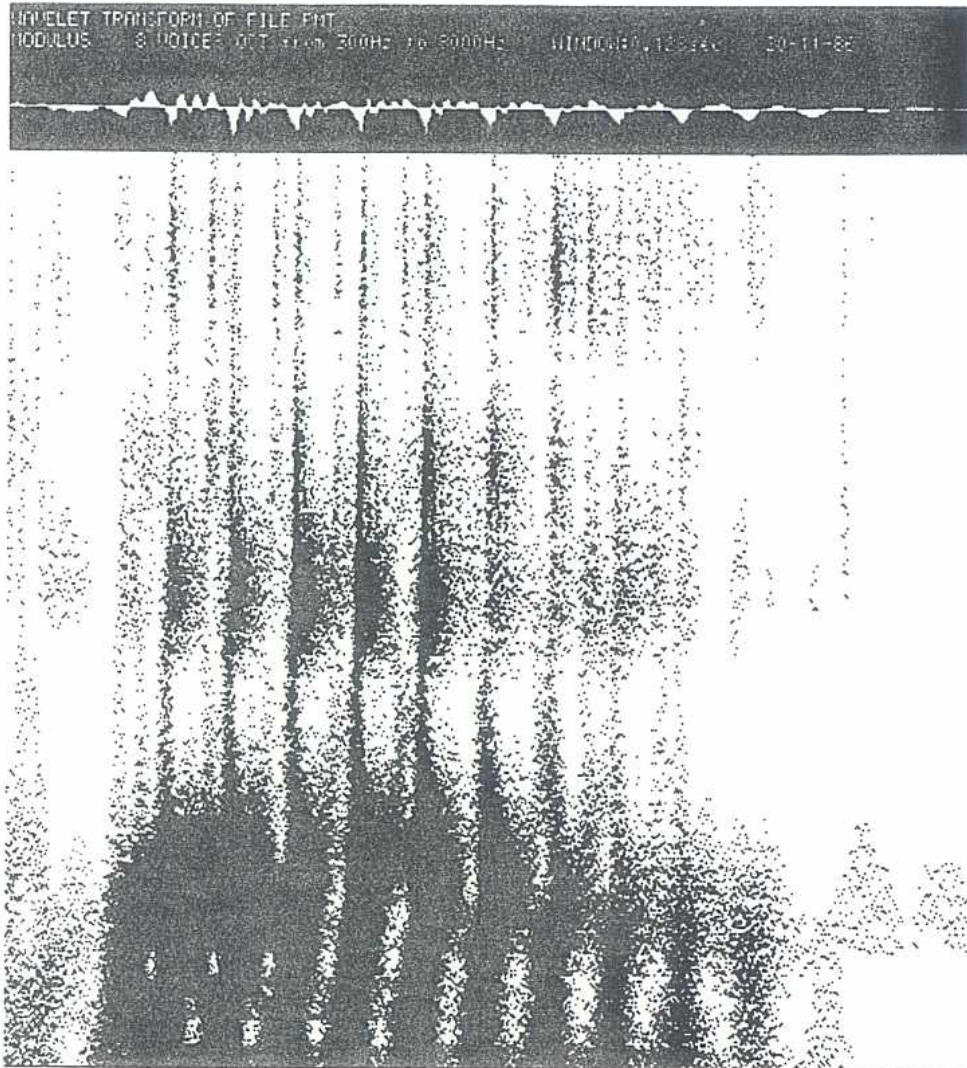
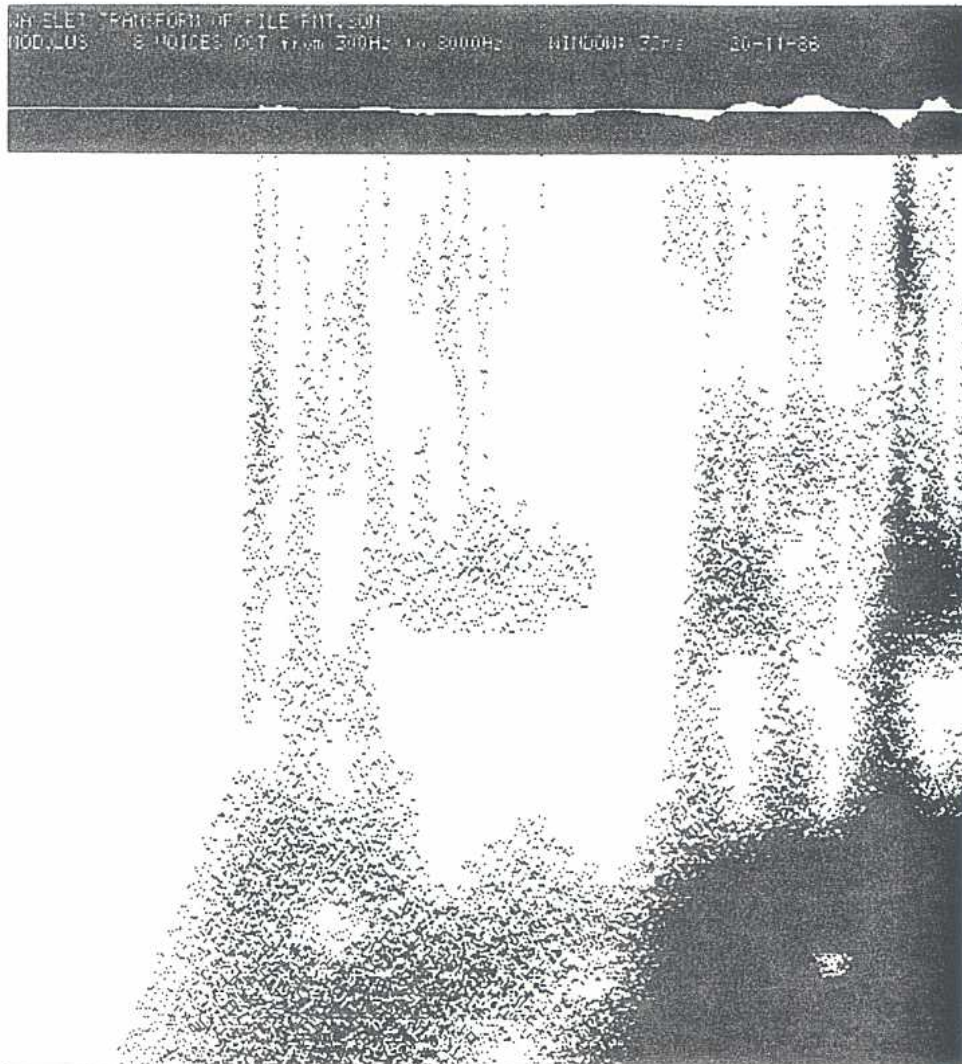Fig. 13a. Modulus of transform of "TA" of "TATY" during 0.128 seconds, between 300 and 8000 Hz.

Fig. 13b. The beginning of Fig. 13a, during 32 ms.

4.4. Notes on a Clarinet

Figures 14a to 14e represent sounds on a clarinet.

In Fig. 14a one sees (in modulus) the start of a high A. The duration of the picture is 0.128 seconds; the frequency range covered is 430 to 8000 Hz. The different harmonics are well-separated. The attack is shown, in the same frequency range, in Fig. 14b, which covers 16 ms. The phase picture (Fig. 14c) over 32 ms, shows gentle phase convergence.

The transition between a B and a C sound is shown in Fig. 14d (modulus) and 14e (phase), over 30 ms. One sees in Fig. 14d, both on the signal and on the transform, that C has more energy.
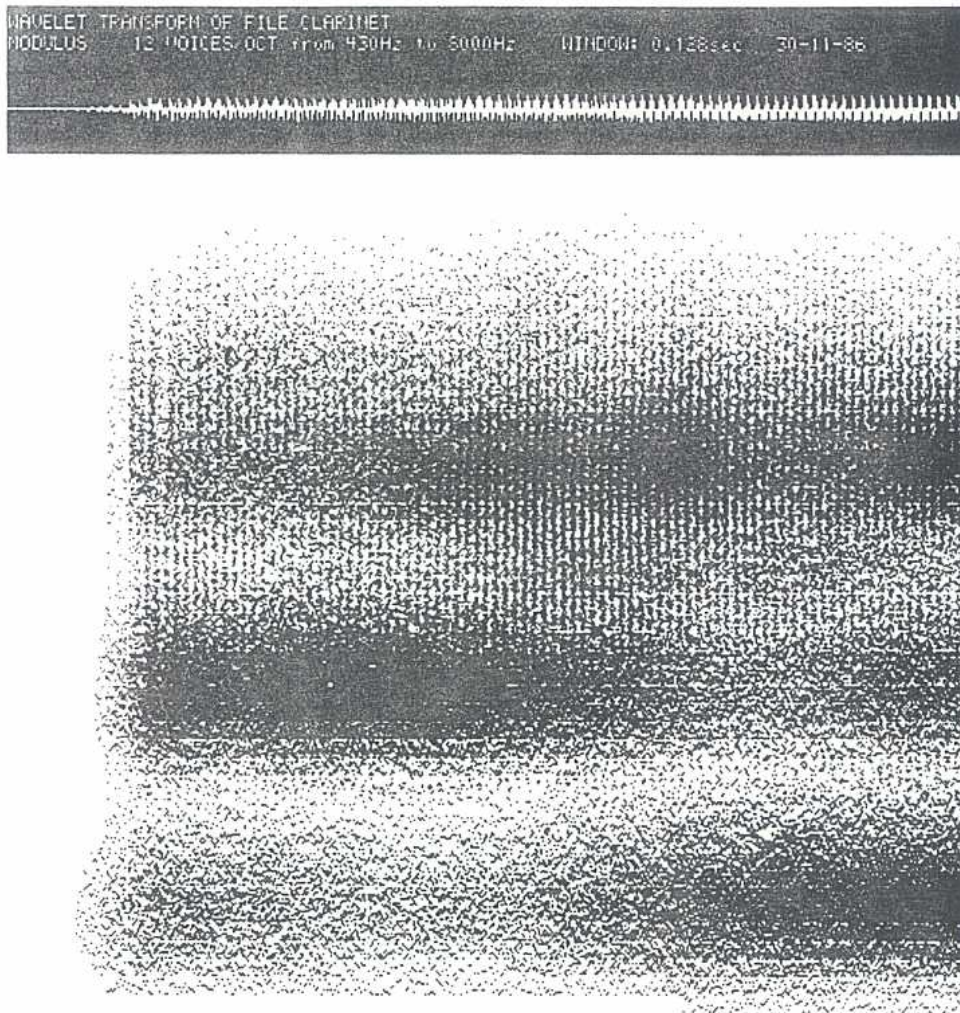

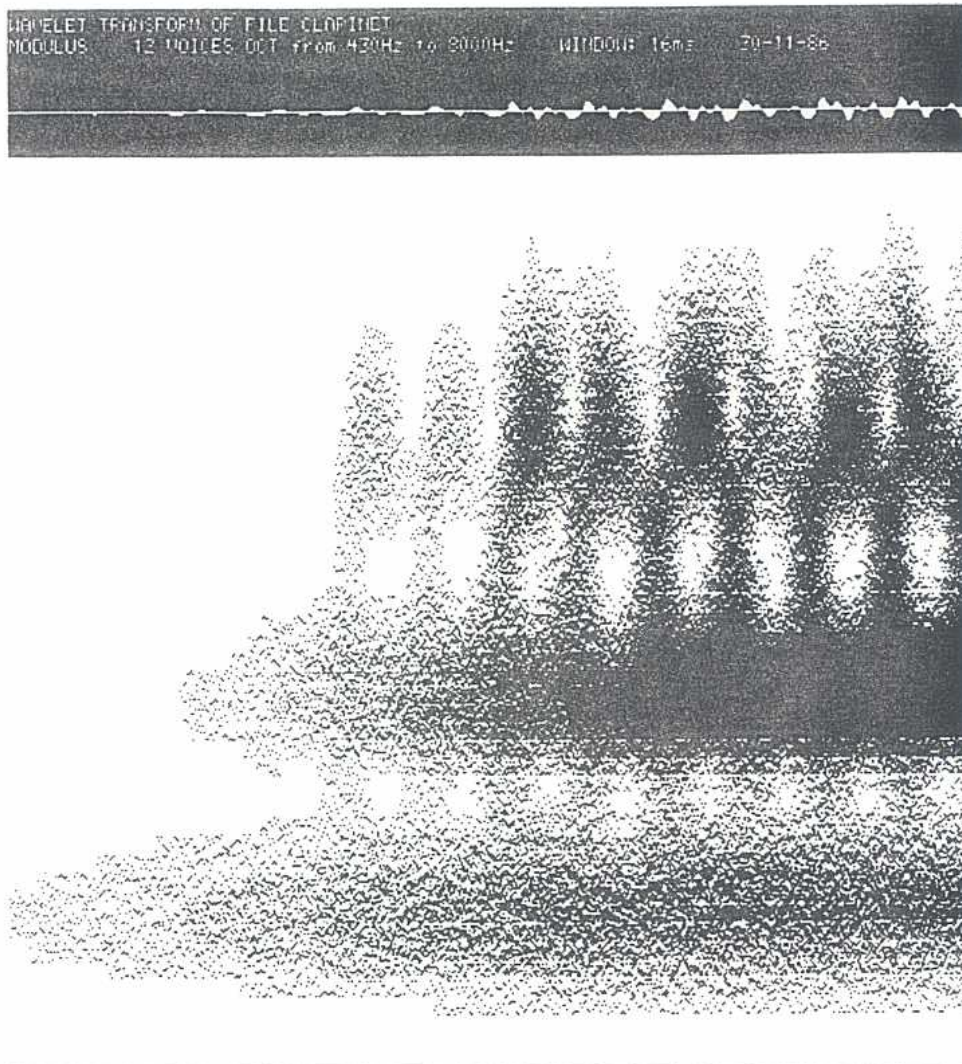Fig. 14a. The modulus of the transform of the high A on a clarinet, during 0.128 seconds, between 430 and 8000 Hz.
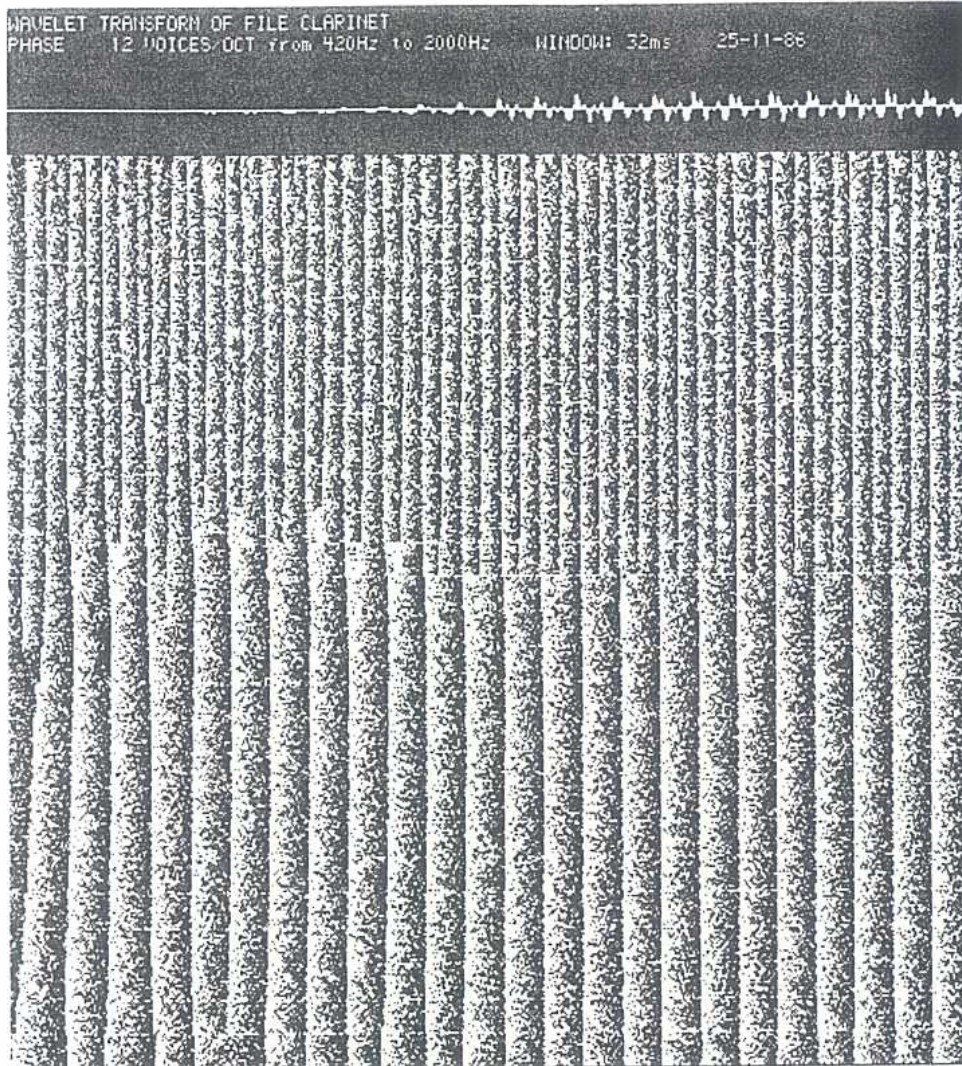
120

WAVELET TRANSFORM OF FILE CLARINET
MODULUS   12 VOICES OCT from 430Hz to 8000Hz   WINDOW 16ms   30-11-86

Fig. 14b. The first 16 ms of Fig. 14a.

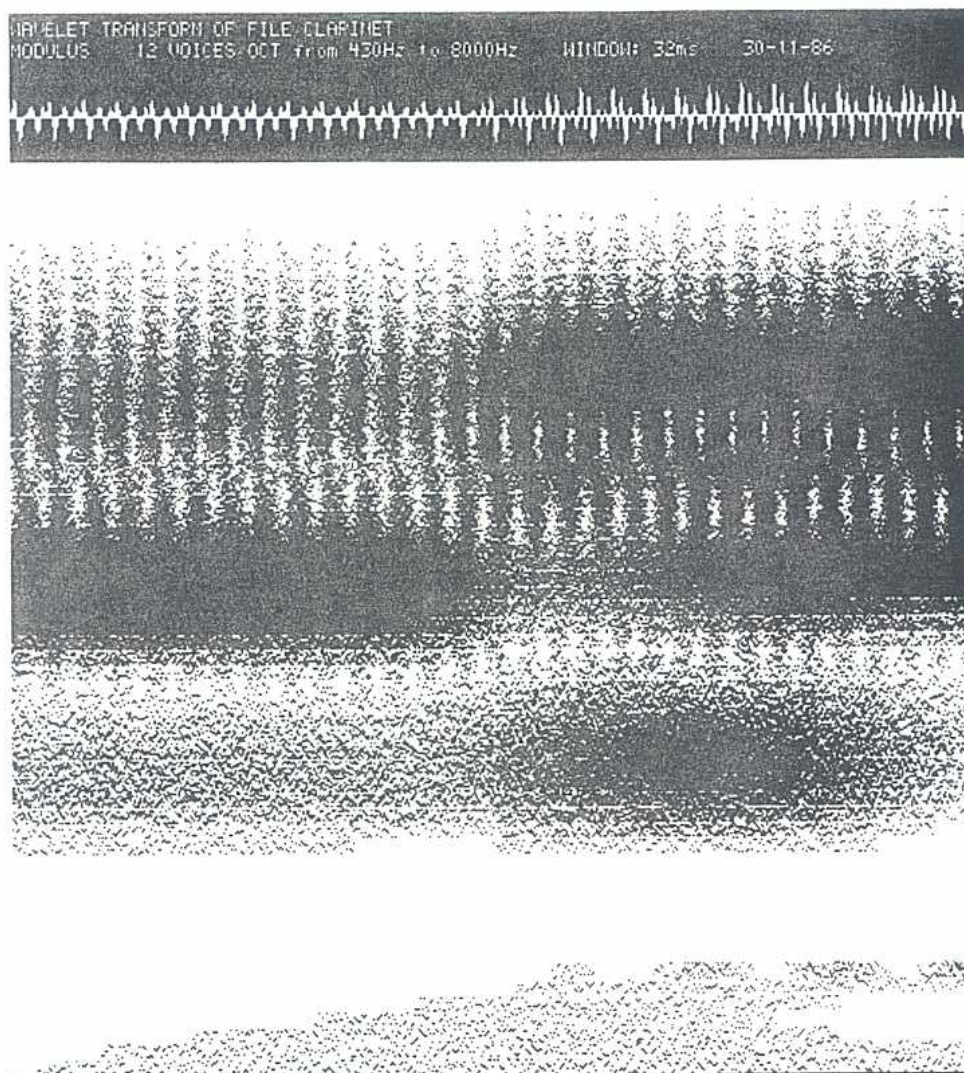Fig. 14c. Phase picture corresponding to the 32 ms of Fig. 14a.

Fig. 14d. Modulus of a legato transition between high B and high C, during 32 ms, between 430 and 8000 Hz.
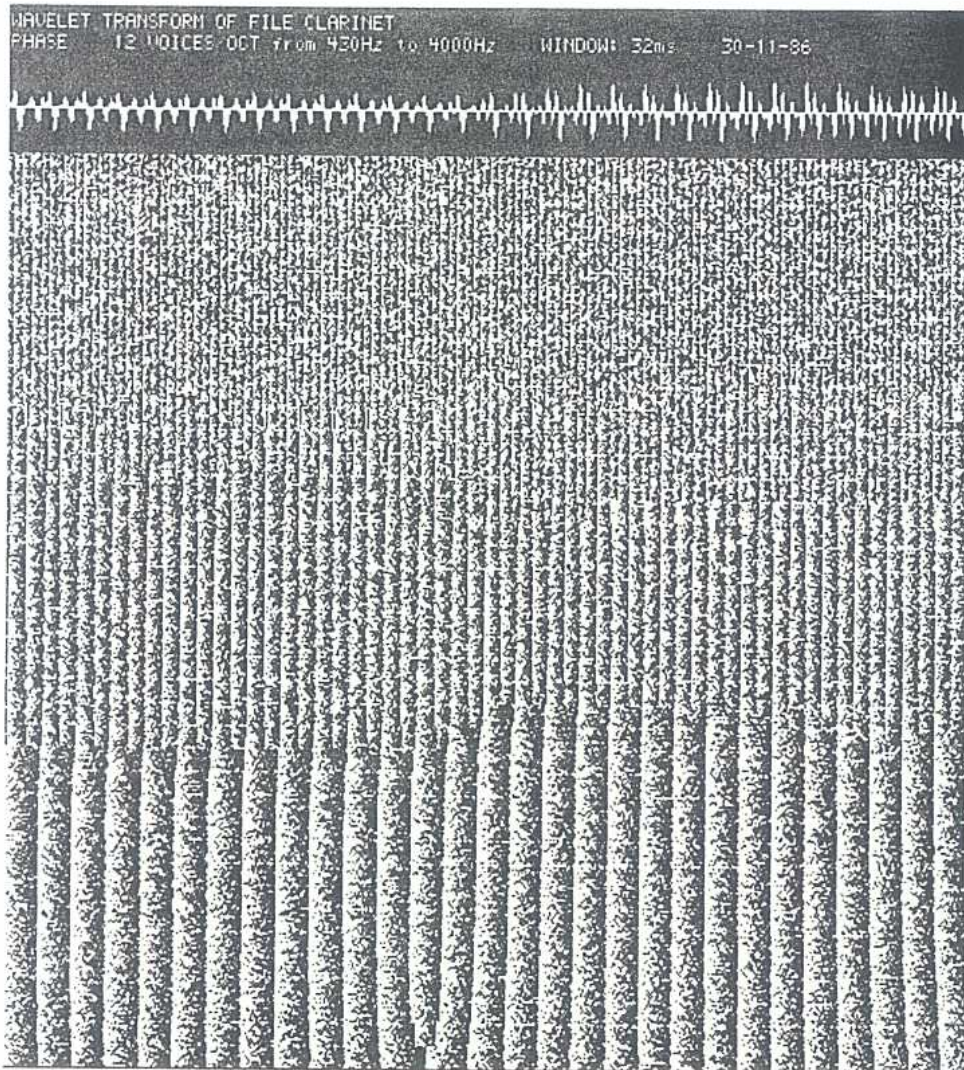
Fig. 14e. The transition of Fig. 14d, in phase picture. The transition, not entirely evident on the signal, can be pinpointed on the phase picture.

5. CONCLUSIONS

On the basis of our experience until now, we believe that various implementations of the method of wavelet transforms will find a place in a wide range of applications involving signal processing and pattern recognition. In the near future, we intend to describe the utilisation of these methods to problems of synthesis and modification of sounds. One of us (R.K.) has recently obtained very promising results in the old problem of sound transposition without change in duration. Concerning speech, the preliminary results shown in this paper encourage us to believe that the combined information on modulus and on phase of wavelet transforms is useful in the segmentation of speech sounds. The methods described here can be generalized to signals in more than one dimension. S. Mallat and Y. Meyer have informed us on applications on the processing of images.

REFERENCES

(This is a very sketchy list of references where the reader may find related information.)

1. P. Goupillaud, A. Grossmann and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration* **23** (1984/1985), 85–102.
2. A. L. Yuille and T. Poggio, "Scaling theorems for zero-crossings," Massachusetts Institute of Technology, A.I. Memo 722, June 1983.
3. A. Grossmann, "Wavelet transforms and edge detection," to appear in *Stochastic Processes in Physics and Engineering*, eds. Ph. Blanchard, L. Streit and M. Hazewinkel, Reidel Publ. Co.
4. I. Daubechies and T. Paul, "Wavelets and appliations," *Proceedings of the 8th International Congress on Mathematical Physics*, eds. M. Mebkhout and R. Séneor, (Marseille 1986), World Scientific, 1987, pp. 675–686.
5. P. G. Lemarié and Y. Meyer, "Ondelettes et bases Hilbertiennés," to appear in *Rev. Ibero-Americana Mat.*
6. A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. An.* **15** (1984) 723–736.
7. S. Mallat, "Scale space versus scale change representation," submitted for publication in ICCV 87 proceedings.

R. Kronland-Martinet works in the Laboratoire de Mecanique et Acoustique Centre National de la Recherche Scientifique, Marseilles, France. His activity has largely been in the domain of computer music, and in particular in the analysis and synthesis of audio signals.

Jean P. Morlet is a senior geophysicist (R&D) at O.R.I.C. (ELF Aquitaine) where he is employed since 1958. He is a member of the societies of SEG and EAEG. His areas of professional interest are in digital seismic processing, image processing, wave propagation and theoretical physics.

A. Grossmann is, by training, a theoretical physicist. Since 1967 he has been at the Centre de Physique Théorique, Centre National de la Recherche Scientifique, Marseilles, France. His previous work includes the study of phase-space quantum mechanics, which is related to time-frequency methods of signal analysis.