SOME REMARKS ON INSTANTANEOUS AMPLITUDE AND FREQUENCY OF SIGNALS

Bernard Picinbono

Laboratoire des Signaux et Systèmes
Supélec, Plateau de Moulon, 91912, Gif sur Yvette, France. E-mail: picinbono@lss.supelec.fr.

ABSTRACT

The most common way to define instantaneous amplitude, phase and frequency of a real signal is to use its analytic signal. One of the main advantages of this procedure is that it establishes a one-to-one relationship between any real signal and the pair of functions defining its instantaneous amplitude and phase. However, there are some problems in its physical interpretation but it is shown that for narrow-band signals the analytical signal is well related to some physical procedures allowing the measurements of amplitude and frequency. As any analytical signal is a very specific function, any arbitrary pair of functions cannot be considered as the amplitude and phase of a real signal. This point is especially discussed in the case of phase signals, which means signals with constant instantaneous amplitude. The phase must satisfy very specific conditions related to the theory of Blaschke functions and analyzed in the regular as well as in the singular cases. These results are applied to hyperbolic and parabolic chirp signals. Hyperbolic signals are true phase signals corresponding to the singular case. On the other hand parabolic signals, often called signals with instantaneous frequency varying linearly in time, are not phase signals and their amplitude is not constant. This point is analyzed in detail and the structure of their amplitude and phase is explicitly calculated. Various calculations allow the evaluation of the errors appearing when assuming that these signals are true phase signals. Some extensions to other cases of chirp signals are discussed.

1. INTRODUCTION

Instantaneous amplitude, phase and frequency (IAPF) of signals seem common sense concepts widely used in radiocommunication systems and their designers do not take great care of mathematical concepts. In order to analyze the behavior of these systems they make use of these very simple ideas. Let us remember that a purely monochromatic signal such as $a\cos(\omega t + \phi)$ cannot transmit any information. For this purpose a modulation is required and one of the simplest possible to introduce is the *amplitude modulation*. Let $\mu(t)$ be a positive function corresponding to the information to be transmitted. By multiplying the carrier frequency signal $\cos(\omega_0 t)$ by $\mu(t)$ we obtain the signal

$$x(t) = \mu(t)\cos(\omega_0 t) \ , \qquad (1)$$

and it is commonly admitted that $\mu(t)$ is the *instantaneous amplitude* of the signal $x(t)$. By a similar reasoning it is commonly admitted that the signal

$$x(t) = a\cos[\phi(t)] \qquad (2)$$

has a constant amplitude a and an *instantaneous phase* $\phi(t)$. Furthermore the *instantaneous angular frequency* is given by the derivative of $\phi(t)$. This is a generalization of the procedure applied in the case where the phase is linear in time or in the form $\phi(t) = \omega_0 t + \phi$ giving the frequency ω_0.

Even if the previous definitions appear quite natural and are widely used in practical applications dealing with signal modulation, we immediately note that they cannot be satisfactory. Indeed it is possible to associate with any real signal $x(t)$ an infinite number of pairs $[a(t), \phi(t)]$ such that

$$x(t) = a(t)\cos[\phi(t)] \ . \qquad (3)$$

In order to arrive at a precise mathematical definition it is necessary to suppress this ambiguity. This introduces the concept of a *canonical pair* [1]. For such pairs there is a one to one correspondence between the real signal $x(t)$ and $[a(t), \phi(t)]$. This means that the pair is defined from $x(t)$ and conversely that this signal is given by (3). Canonical pairs are defined and analyzed in Section 2. It is especially shown that such pairs cannot be characterized only from spectral assumptions, as often believed. A good example of the constraints on such pairs appears when the amplitude is constant, which defines phase signals. The phase $\phi(t)$ such that the pair $[1, \phi(t)]$ is canonical must have a very specific structure related to Blaschke functions. This is especially analyzed in Sections 3 an 4. In order to verify the results of these sections we consider chirp signals on the form $\cos[(at)^n]$ and we calculate their IA and IF. It is shown that such signals are phase signals only for $n = \pm 1$. In the other cases the IF cannot be obtained by derivation of $(at)^n$ and numerical results are presented for $n = 2$ and $n = 3$.

2. DEFINITIONS AND CANONICAL PAIRS

2.1. Mathematical Definitions, Analytic Signal

The idea is to associate a complex signal $z(t)$ with any real signal $x(t)$ in such a way that the IA and IP of $x(t)$ are the

0-7803-5073-1/98/$10.00 ©1998 IEEE

modulus and the argument of $z(t)$. Therefore the canonical property is ensured if there is a one to one correspondence between $x(t)$ and $z(t)$. Furthermore it is necessary that this procedure gives the classical result for pure sinusoid signals. If we also impose that the relation between $x(t)$ and $z(t)$ is a linear filtering, the problem has a unique solution and $z(t)$ is the *analytic signal* (AS) of $x(t)$. Let us recall its precise definition.

The AS $z(t)$ is obtained by filtering $x(t)$ through a filter with the frequency response equal to 2 for $\nu > 0$ and to 0 for $\nu < 0$. Conversely it is obvious that $x(t) = \text{Re}[z(t)]$, where Re means the real part. So if $x(t)$ is real, there is indeed a one-to-one correspondence between $x(t)$ and $z(t)$. On the other hand, a complex function is an AS if its Fourier transform is zero for negative frequencies. It is clear that this function is the AS of its real part.

As $z(t)$ cannot be a real function because its Fourier transform $Z(\nu)$ is zero for $\nu < 0$, it can be written as

$$z(t) = a(t)\exp[j\phi(t)] \,, \tag{4}$$

where the phase $\phi(t)$ is defined modulo 2π and $a(t)$ is non-negative. As a conclusion, using the AS makes it possible to associate a unique pair $[a(t), \phi(t)]$ with any real signal. In the following it is called the *canonical pair* associated with $x(t)$. This pair defines in all what follows the IA and IP of the signal $x(t)$. The IF is the derivative of the phase with respect to time.

The introduction of the AS is not at all new. It is presented in [2] - [8] and it is shown in [7] and [9] that, starting from some *a priori* physical assumptions, the only possible definition of the instantaneous amplitude and phase is the one using the AS. The question which directly results from the definition is the following. As any AS is a very specific function, an arbitrary pair $[a(t), \phi(t)]$ has no reason to be canonical. It is therefore necessary to find conditions ensuring this property.

For this purpose we shall extensively use properties of *Hilbert transforms*. The function $z(t)$ is the AS of $x(t)$ if and only if it can be written as $z(t) = x(t) + jy(t)$, where $y(t)$ is the Hilbert transform (HT) of $x(t)$. This HT is obtained by filtering $x(t)$ through a linear filter with frequency response $-j\text{Sg}(\nu)$.

2.2. Spectral Characterization of a Canonical Pair

The canonical pair introduced in amplitude modulation is $[a(t), \omega_0 t + \phi]$. It is easy to verify that this pair of functions is canonical if and only if $a(t)$ is a positive band-limited signal, which means that its Fourier transform (FT) $A(\nu)$ is zero for $|\nu| > \nu_0$, where $\nu_0 = \omega/2\pi$. This is a very simple spectral characterization of a canonical pair. It is therefore tempting to try to use spectral methods for the characterization of more general pairs of functions $[a(t), \phi(t)]$. Unfortunately we shall see that this is impossible.

Saying that $a(t)\exp[j\phi(t)]$ is an AS is equivalent to saying that the Hilbert transform of $a(t)\cos[\phi(t)]$ is equal to $a(t)\sin[\phi(t)]$ (see p. 49 of [10]). It is therefore appropriate to make use of the so-called Bedrosian theorem [11] dealing with the Hilbert transform of a product of two real functions $x_1(t)$ and $x_2(t)$. A very simple derivation of this theorem and some extensions can be found in [12]. The main result is as follows: let $X_1(\nu)$ and $X_2(\nu)$ be the FTs of $x_1(t)$ and $x_2(t)$ respectively. If $X_1(\nu) = 0$ for $\nu > B$ and $X_2(\nu) = 0$ for $\nu < B$, then

$$H[x_1(t)x_2(t)] = x_1(t)H[x_2(t)] \,, \tag{5}$$

where $H[.]$ means the Hilbert transform. A direct application of this result shows that if $a(t)$ is a low-frequency signal (B) and $\cos[\phi(t)]$ a high-frequency signal (B), or if their spectra do not overlap, then

$$H\{a(t)\cos[\phi(t)]\} = a(t)H\{\cos[\phi(t)]\} \,. \tag{6}$$

However this does not at all imply that

$$H\{\cos[\phi(t)]\} = \sin[\phi(t)] \,, \tag{7}$$

as stated by many authors, and even recently in [8]. In fact this equation implies that $z(t) = \exp[j\phi(t)]$ is an AS or that the pair $[1, \phi(t)]$ is canonical. We shall see that this requires very specific properties of the structure of the phase $\phi(t)$ which cannot be characterized only by spectral considerations, as for amplitude modulation.

2.3. Physical Interpretations

There are various criticisms against the use of the AS especially to define the IF, and that is the reason why other attempts to define IF of signals have been presented [13], [14], [15], [16], [17]. The main criticisms usually presented are the following. (*a*) The IF defined by the AS is typically erratic and has wild variations. It can even take negative values. (*b*) The IF is not related to zero crossings of the signal, which seems a good approach to define the IF of a signal by extension of the situation valid for sinusoid signals. In this case the distance between successive zeros is the period, proportional to the inverse of the frequency. (*c*) For band-limited signals the IF does not always belong to the frequency band of the signal. (*d*) There is no clear relationship between the IF and the FT of the signal. (*f*) For periodic signals the IF has no relationship with the fundamental frequency.

We shall now discuss these points, and our objective is to show whether or not the quantities defined by the AS are related to some physical measurements.

However it is important to point out that even if the AS can be defined for any real signal, the convenient framework for its use is the case of *narrow-band signals*, and we shall admit this assumption unless otherwise specified. In this case the AS can be expressed as $a(t)\exp[j\omega_0 t + j\phi(t)]$, where ω_0 is the carrier angular frequency. Therefore $z(t)$ is the product of the monofrequency signal $\exp(j\omega_0 t)$ by the complex amplitude $a(t)\exp[j\phi(t)]$. The narrow-band assumption means that this amplitude varies very slowly in time intervals of the order of the period $T_0 = 2\pi/\omega_0$. In the complex plane this can be represented by a vector rotating with the velocity ω_0 slowly modulated by the complex amplitude, which introduces slow variations of the modulus and the phase.

Let us first consider the problem of the instantaneous amplitude. It is clear that the IA defined by the AS is the amplitude measured by a physical device taking the time average of $x^2(t)$ on an interval of the order of T_0.

Consider now the relationship between the IF given by the AS and the time intervals between two successive zeros of $x(t)$, which is a possible physical interpretation of the IF. This interval is defined by the fact that the IP $\Phi(t) = \omega_0 t + \phi(t)$ has a variation of 2π. As the local phase $\phi(t)$ has slow variations the time interval $T(t)$ corresponding to a variation of 2π of $\Phi(t)$ satisfies $[\omega_0 + \phi'(t)]T(t) = 2\pi$, which shows that the IF defined from the AS is related to the distance between successive zeros of the signal.

Let us now come to the problem of relationships between the FT of the signal and its IF. This question was discussed in [13], [14] and [15]. It is easy to show that the time average of the IF is the average frequency deduced from its FT [3] [13]. This means that if $Z(\omega)$ is the FT of $z(t)$ we have

$$\frac{\int \omega |Z(\omega)|^2 d\omega}{\int |Z(\omega)|^2 d\omega} = \frac{\int \omega(t) |z(t)|^2 dt}{\int |z(t)|^2 dt}, \qquad (8)$$

where $\omega(t)$ is the IF deduced from the AS. However this does not at all mean that if a real signal is band-limited in a frequency band B, its IF does not necessarily belong to this band, and this property is sometimes considered as indicating a lack of physical meaning of the IF deduced from the AS. In reality there is no reason for the existence of this kind of relationship between these two quantities.

In order to discuss this point more precisely, let us take the example of a two components signal, already analyzed in [13], [14] and [15]. Consider the signal

$$x(t) = a\cos[(1 - \frac{\Delta\omega}{\omega_0})\omega_0 t] + \cos[(1 + \frac{\Delta\omega}{\omega_0})\omega_0 t], \quad (9)$$

where ω_0 is the carrier frequency and $\Delta\omega$ satisfies $\Delta\omega << \omega_0$, which ensures the narrow-band property. This signal is composed of two spectral lines at the frequencies $\omega_0 - \Delta\omega$ and $\omega_0 + \Delta\omega$. It is easy to calculate the IF deduced from the AS, and the result is

$$\omega(t) = \omega_0 + \Delta\omega \, \frac{1 - a^2}{a^2(t)}, \qquad (10)$$

where $a^2(t)$ is the square of the IA of the signal given by

$$a^2(t) = 1 + 2a\cos(2\Delta\omega t) + a^2. \qquad (11)$$

If $a = 1$, this expression gives $\omega(t) = \omega_0$, which is satisfactory because in this case $x(t) = 2\cos(2\Delta\omega t)\cos(\omega_0 t)$, giving immediately the IF of the signal. On the other hand we have $\omega(t) > \omega_0$ for $a < 1$ and $\omega(t) < \omega_0$ for $a > 1$. This shows that there is a discontinuity for $a = 1$. It is clear that $\omega(t)$ is not necessarily in the frequency band of the signal. For example for $a < 1$, this property appears if $a^2(t) < 1 - a^2$. This yields $\cos(\Delta\omega t) < [(1-a)/2]^{1/2}$, and for any value of a ($0 < a < 1$) there are time instants where this inequality is satisfied.

This is not in contradiction with physical measurements and in order to verify this point, consider now the zeros of $x(t)$ defined by $x(t) = 0$ which can be put in the form

$$\tan(\omega t) = \frac{1 + a}{1 - a}\cot(r\omega t). \qquad (12)$$

By a calculation not reproduced here, it can be shown that the distance between two successive zeros introduces an instantaneous period $T(t)$ equal to $2\pi/\omega(t)$, where $\omega(t)$ is given by (10). Therefore a frequency discriminator using the distance between zeros will give an IF in accordance with the value obtained from the AS. As a consequence for some zeros this distance can be smaller than $2\pi/(\omega + \Delta\omega)$, which means that the IF is not inside the frequency band of the signal.

As there is a singularity for $a = 1$, consider a particular example of signal $x(t)$ represented in Fig. 1. This signal is calculated for $\omega_0 = 100\Delta\omega$ and the variable on the x axis is $x = \omega_0 t/\pi$. Therefore the zeros of $\cos(\omega_0 t)$ correspond to the numbers $x = n + 1/2$, where n is integer. The signal of the figure is calculated for $a = 0.995$. It appears that there is indeed a singularity at $x = 50$, and if we measure the IF by zero counting, we find that the IF can be twice as large as the frequency ω_0. In reality this figure shows that the concept itself of IF is not valid in the neighborhood of $x = 50$, or $t = \pi/(2\Delta\omega)$. Indeed for these values the IA $a(t)$ becomes very small and the figure shows that the signal cannot be locally approximated as a pure sinusoid signal. In reality the basic physical idea in the use of the AS is to admit that any narrow-band signal can be considered locally as a pure sinusoid signal, and the AS allows one to determine its amplitude and phase. These quantities must therefore vary slowly in time in order to be IAP of the signal.

Let us now consider the same signal as in Fig. 1 but with $a = 0.5$. The maximum and minimum values of the IA are 1.5 and 0.5, reached for $x = 0$ and $x = 50$ respectively. This appears in the Fig. 2 which also shows that the period is always smaller than 2, period corresponding to the frequency ω_0, and the minimum value of this period appears for $x = 50$. Theoretically the value of this minimum is 0.97×2, while the period corresponding to the upper bound of the frequency band is 0.99×2. This is of course difficult to evaluate in the figure and this is the result of the narrow-band assumption.

In conclusion of this discussion the fact that the IF does not belong to the frequency band of the signal deduced from its FT is not an argument against its physical meaning and this criticisms against the AS is not entirely valid in the narrow-band case. Mathematical quantities obtained from the AS can be in good agreement with possible physical measurements.

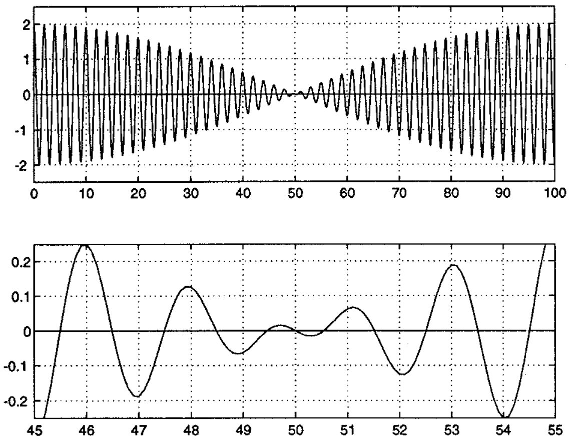

Figure 1: Signal with two spectral lines, $a = 0.995$

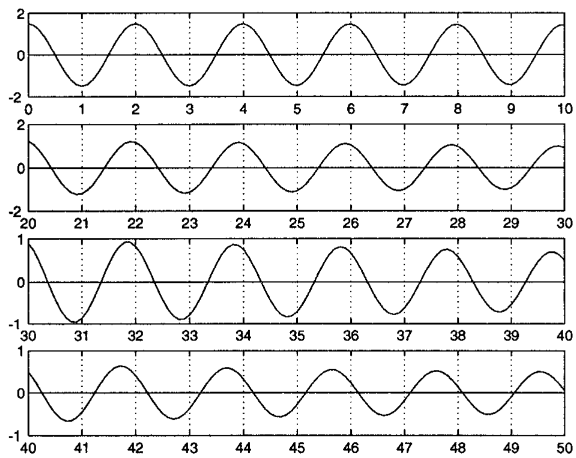

Figure 2: Signal with two spectral lines, $a = 0.5$

This is no longer the case if the assumption of narrow bandwidth does not hold. From the mathematical point of view it is always possible to introduce the AS and to deduce an IA and an IP from it. However the physical interpretation of these quantities looses any simple meaning. For example a broad-band signal can have no zeros and a device measuring the IF from the distance between two successive zeros give an IF equal to zero.

In front of this situation various approaches have been presented. The most recent one uses the so-called *homomorphic* AS related to the logarithm of the signal and well adapted to the structure of poles and zeros [16]. There is no space here to discuss this approach here and some comments will be presented in the oral presentation of this paper. We can only point out that this procedure cannot introduce the concept of IA and therefore only IF can be discussed. As an example let us take the case of the signal $x(t) = a + \cos(\omega t)$. The IF deduced from the AS is given by (10) where ω_0 and $\Delta\omega$ are replaced by $\omega/2$. It is clear that this IF has no real physical meaning because the signal is not narrow-band. By using the homomorphic AS the IF becomes ω if $|a| < 1$ and $\omega f(t)$ if $|a| > 1$, with $f(t) = 1 - \sqrt{a^2-1}[a + \cos(\omega t)]^{-1/2}$. It is difficult to give a physical interpretation of these 3 results and this comes from the fact that the assumption of narrow-band is not valid.

Another point sometimes discussed is the IF of a periodic signal [16] [17]. It is clear that looking at a music score for instruments playing only one note (flute or clarinet) gives the feeling of reading the IA and IF of the musical signal. This is however not the case. Indeed each note is not a pure sinusoid signal but only a periodic signal with various harmonics and the specificity or the quality of the sound depends precisely on the structure of these harmonics. The possible IF is only the frequency of the fundamental, and associating only one frequency at each instant with the complete signal is mathematically possible, for example by using the AS, but physically without real meaning. In reality there is a spectrum at each instant, and the concept of the instantaneous spectrum is better adapted to this situation than the one of IF. However this point is outside the scope of this talk.

3. PHASE SIGNALS, REGULAR CASE

3.1. General Structure

Phase signals are *real signals* with constant instantaneous amplitude. They can be expressed as (2), but with the condition that $\exp[j\phi(t)]$ is an AS. As a consequence (7) is satisfied. For such signals all the information is contained in the instantaneous phase (or frequency), and phase signals are then the basic elements of phase or frequency modulation.

The condition that $\exp[j\phi(t)]$ is an AS requires very specific properties on the phase $\phi(t)$. These properties have been analyzed in the framework of coherence problems in Optics [18] but more precisely in the framework of the study of analytic functions and especially in Chapter 17 of [19].

The most general structure of the AS of a regular (or nonsingular) phase signal is

$$z(t) = b(t)\exp[j(\omega_0 t + \theta)] , \qquad (13)$$

where θ is arbitrary, ω_0 is non-negative and $b(t)$ is a Blaschke function defined by

$$b(t) = \prod_{k=1}^{N} \frac{t - z_k}{t - z_k^*} , \quad z_k \in P_+ , \qquad (14)$$

P_+ being the half-plane of the complex plane defined by $\mathrm{Im}(z) > 0$. The quantity ω_0 is the carrier frequency and can be equal to zero. The expression *regular* (or nonsingular) means especially that the number N of factors in the product is finite. The interpretation of (14) is very simple.

In order to be an AS the function $b(t)$ for complex values of t must have all its poles in the half-plane $\mathrm{Im}(z) < 0$. In order to have a modulus equal to one each pole must be associated with a corresponding zero symmetric to this pole with respect to the real axis. This procedure is well known in filter theory: stable phase filters have the same number of poles and zeros and these zeros are symmetric of the poles with respect to the imaginary axis. The stability and causality conditions imply that all the poles are in the left half-plane of the complex plane.

It is obvious that $|b(t)| = 1$, which implies that $|z(t)| = 1$. Let us now explain why $z(t)$ is an AS. For this we must analyze the structure of the FT $B(\nu)$ of $b(t)$. As N is finite, $b(t)$ is a rational function in t. If all the z_ks are distinct we can write

$$b(t) = 1 + \sum_{k=1}^{N} \frac{c_k}{t - z_k^*} , \qquad (15)$$

where

$$c_k = \lim_{t \to z_k^*} (t - z_k^*) b(t) . \qquad (16)$$

As a consequence we have

$$B(\nu) = \delta(\nu) + \sum_{k=1}^{N} C_k(\nu) , \qquad (17)$$

where $C_k(\nu)$ is the FT of $c_k(t - z_k^*)^{-1}$. Because of the localization of z_k^* in the complex plane, we deduce that $C_k(\nu) = 0$ for $\nu < 0$, which implies that $B(\nu) = 0$ for $\nu < 0$, and ensures that $b(t)$ is an AS. Finally, as $\omega_0 > 0$, $z(t)$ also is an AS. The reasoning can be extended without difficulty when some poles z_k are no longer distinct.

The phase of $b(t)$ is of course

$$\phi_b(t) = \mathrm{Arg}[b(t)] \ , \mathrm{mod}(2\pi) , \qquad (18)$$

and, as a result, we can say that any phase signal can be written as (2) where $\phi(t)$ must have the form

$$\phi(t) = \theta + \omega_0 t + \phi_b(t) \ , \mathrm{mod}(2\pi) . \qquad (19)$$

In practice the continuity of the phase leads to suppress the term $\mathrm{mod}(2\pi)$ and this convention is adopted in all that follows.

This most general phase is defined by N complex parameters z_k and 2 real parameters ω_0 and θ. Furthermore it is obvious that the phase $\phi_b(t)$ is the sum of N phases of the factors appearing in the product (14). Let $b_k(t)$ be equal to $(t - z_k)(t - z_k^*)^{-1}$, and $\phi_k(t)$ be its phase. This gives

$$\phi(t) = \theta + \omega_0 t + \sum_{k=1}^{N} \phi_k(t). \qquad (20)$$

By introducing the real and imaginary parts of z_k, or $z_k = a_k + jb_k$, one obtains

$$\phi_k(t) = 2\mathrm{Arctg}\frac{b_k}{a_k - t} \ , -\pi/2 \le \phi_k(t) \le \pi/2 \ . \qquad (21)$$

At this step we return to the problem discussed in the introduction. The signal (2) is phase modulated only if its phase takes the form (19), and, as this is not in general the case, its amplitude is not constant and it must be expressed as in (3).

3.2. Properties of Regular Phase Signals

Having the most general structure of regular phase signals we shall now present some of their properties, which allows a better understanding of their structure. Note first that the product of phase signals is still a phase signal. This results directly from the definition of phase signals.

3.1. *A phase signal contains only two spectral lines corresponding to its carrier frequency.*

This is a direct consequence of the structure of the Blaschke function appearing in (14). Its FT $B(\nu)$ is given by (17) which can be written as

$$B(\nu) = \delta(\nu) + B_c(\nu) \ . \qquad (22)$$

The function $B_c(\nu)$ describes the continuous part of the FT of $B(\nu)$. It is a sum of N components $C_k(\nu)$ that are bounded and equal to zero for $\nu < 0$. This implies that $B_c(\nu)$ is also bounded and equal to zero for $\nu < 0$, and thus $B(\nu)$ exhibits only one Dirac component, or a spectral line, at the frequency zero. Because of the exponential term in (13), the FT of $z(t)$ is $Z(\nu) = e^{j\theta}B(\nu - \nu_0)$ and this means that there is only one spectral line at the carrier frequency $\nu_0 = \omega_0/2\pi$. By using the Hermitian symmetry, we deduce that $x(t)$ has only two spectral lines at the frequencies $\pm\nu_0$.

3.2. *A phase signal with a non-zero carrier frequency is a high-frequency (ν_0) signal.*

This is a direct consequence of the form of the FT $Z(\nu)$ analyzed just above. As $B(\nu) = 0$ for $\nu < 0$, $Z(\nu) = 0$ for $\nu < \nu_0$ and $X(\nu) = 0$ for $|\nu| < \nu_0$.

The converse is of course not true. There is no reason for a high-frequency signal to be a phase signal because this frequency condition does not imply the structure (13).

3.3. *The FT of the AS of a phase signal is zero for all the frequencies smaller than the carrier frequency ν_0 where the spectral line is located.*

This is a direct consequence of (22) and of the fact that the FT $Z(\nu)$ of $z(t)$ is proportional to $B(\nu - \nu_0)$.

3.4. *A phase signal cannot be a low-frequency (B) signal except when it is monochromatic.*

The monochromatic case appears when $b(t) = 1$, and $z(t)$ is therefore $\cos(\omega_0 t + \theta)$, that is, of course, a low-frequency signal. Except this case, the property means that it is impossible to find a frequency B such that $Z(\nu) = 0$ for $\nu > B$. This property is a consequence of the fact that the functions $C_k(\nu)$ of (17) are exponential functions for $\nu > 0$ when the poles are distinct. But it is well known that a sum of a finite number of exponential functions cannot be zero for all the frequencies satisfying $\nu > B$.

3.5. *Frequency shift*

If in (13) we replace ω_0 by ω_1, with $\omega_1 > \omega_0$, we obtain a complex signal that is still an AS. As a consequence if $x(t) = \cos[\phi(t)]$ is a phase signal, which means that its phase has the structure (19), $x'(t) = a\cos[\Delta\omega t + \phi(t)]$, where $\Delta\omega = \omega_1 - \omega_0$, is still a phase signal with the carrier frequency ω_1. This especially implies that $x'(t)$ is a high-frequency (ω_1) signal.

3.6. *Instantaneous frequency of a phase signal*

It is obtained by differentiating the instantaneous phase. The most general form of this phase is given by (20) and (21), and differentiating this equation yields

$$\omega(t) = \omega_0 + 2\sum_{k=1}^{N} \frac{b_k}{b_k^2 + (a_k - t)^2} \ . \tag{23}$$

As the coefficients b_k are positive, because of the localization of the zeros z_k, we deduce that the IF $\omega(t)$ is always greater than ω_0. This is another illustration of the fact that the FT of $x(t)$ is zero for $\nu < \omega_0$ and also this shows that the IF $\omega(t)$ belong to the frequency domain of this FT. If all the z_ks are zero, then $b(t)$ defined by (14) is equal to one, and the instantaneous frequency is simply ω_0.

Some other comments can be presented on the structure of the instantaneous frequency of a phase signal, and this is especially relevant in all those questions dealing with frequency or phase modulation of signals. The information carried by the instantaneous frequency of a phase signal is entirely in the term

$$\omega_m(t) = 2\sum_{k=1}^{N} \frac{b_k}{b_k^2 + (a_k - t)^2} \ , \tag{24}$$

where the index m stands for the modulation term. We note that this function tends to zero when $|t| \to \infty$. This especially means that $\omega_m(t)$ cannot be a periodic function, and this is related to the fact that a regular phase signal cannot have spectral lines, except those coming from the carrier frequency ω_0.

Furthermore we note that $\omega_m(t)$ is a rational function in t. The polynomials appearing in the numerator and the denominator have the degrees $2N-2$ and $2N$ respectively. As N is arbitrary, we deduce that by using the $2N$ parameters a_k and b_k it is possible to approximate a large class of functions. The most limiting constraint on these functions comes from the necessary behavior for $|t| \to \infty$. In fact $\omega_m(t)$ decreases at infinity in $|t|^{-2}$, which is a strong restriction on the instantaneous frequency.

4. PHASE SIGNALS, SINGULAR CASES

4.1. Infinite Product

The simplest example of a Blaschke function with infinite value of N appears if the poles of (14) are $z_k = kT + jb$, with $T = 2\pi/\omega$ and $b = (1/\omega)\ln(1/a)$ with $0 < a < 1$. In this case it can be shown that

$$b(z) = \prod_{k=-\infty}^{+\infty} \frac{z - z_k}{z - z_k^*} = \frac{a - e^{j\omega z}}{1 - ae^{j\omega z}} \ . \tag{25}$$

Apparently there is no other example of infinite Blaschke products giving an explicit and simple expression. This means that the function

$$b(t) = \frac{a - e^{j\omega t}}{1 - ae^{j\omega t}} \tag{26}$$

is the AS of a phase signal. This can be verified immediately because $|z(t)| = 1$ and the FT is zero for negative frequencies. In order to verify this point it suffices to note that $b(t)$ is periodic, and by using the geometric series one sees that the Fourier coefficients F_n are zero for $n < 0$ and furthermore we have $F_0 = a$ and $F_n = -(1-a^2)a^{n-1}$.

Consider now the signal defined by (13) and (14) where $b(t)$ is given by (26). It is the AS of the real signal

$$x(t) = \frac{-a^2\cos[(\omega_0 - \omega)t] + 2a\cos(\omega_0 t) - \cos[(\omega_0 + \omega)]t}{1 - 2a\cos(\omega t) + a^2} \tag{27}$$

This signal is therefore a phase signal and its IA is equal to 1. It is narrow-band if $(\omega/\omega_0) << 1$, and its IF is

$$\omega(t) = \omega_0 + \omega\,\frac{1 - a^2}{1 - 2a\cos(\omega t) + a^2} \ . \tag{28}$$

Note the analogy with (10). All the properties of phase signal indicated previously are verified, except those using the point that N is finite. In this singular case phase signals contain only spectral lines, but their number is infinite. There are various examples of phase signals that can be generated from (26). It suffices to recall that the product of phase signal remains a phase signal, and it is perfectly possible in these products to mix regular and singular phase signals.

4.2. Poles on the Real Axis

By using arguments that cannot be presented here (see theorem (17-15) in [19]) one can show that the function $z(t) = \exp(-j/t)$ is an AS. This means that the signal $x(t) = \cos(1/at), a > 0$, is a phase signal and its Hilbert transform is $-\sin(1/at)$. These signals are called hyperbolic chirps and are represented in Fig. 3. It is also possible to calculate the FT of $x(t)$, and the result is

$$X(\nu) = \delta(\nu) + u(\nu)\frac{\pi}{\sqrt{2\pi a\nu}}\, J_1\left(2\sqrt{\frac{2\pi a\nu}{a}}\right), \tag{29}$$

where $u(\nu)$ is the unit step function. The spectral line at the frequency 0 is due to the fact that $x(t)$ tends to 1 for infinite values of t. This FT is presented in Fig. 4 where $a = 1$ and $x = 2\pi\nu$.

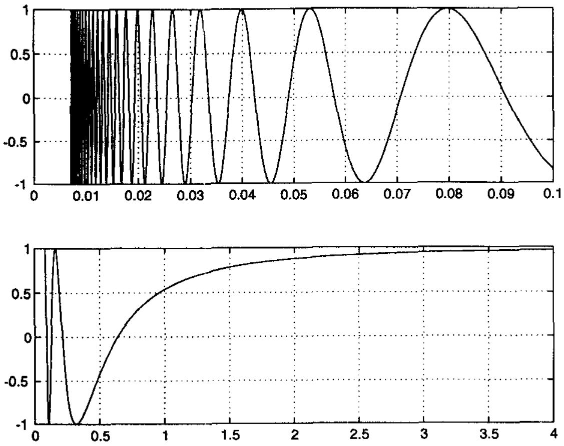

Figure 3: Hyperbolic chirp $y = \cos(1/t)$

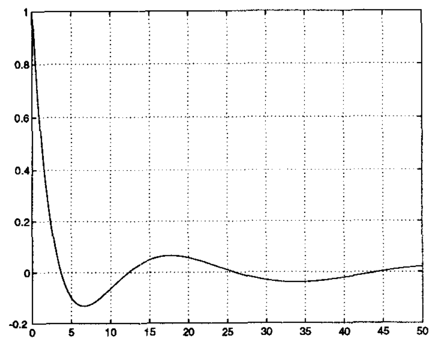

Figure 4: Fourier transform of the hyperbolic chirp

5. APPROXIMATE PHASE SIGNALS

We shall now study signals in the form of $\cos[\Phi(t)]$ where $\Phi(t)$ does not satisfy the conditions analyzed in the previous sections and ensuring that $\exp[j\Phi(t)]$ is an AS. As a consequence these signals do not have a constant IA and $\Phi(t)$ is not the IP. However under some circumstances these signals are approximately phase signals and we intend to study the meaning of this approximation. These signals are widely used in many areas of Signal Processing and especially when the phase is a polynomial in t. In order to simplify the discussion we shall restrict our analysis to the case of degree two and three.

Let us consider the signal $x(t) = \cos[\omega_0 t + a^2t^2]$. It is clear that it is not a phase signal, because its phase does not have the structure previously analyzed. This means that its Hilbert transform is not $\sin[\omega_0 t + a^2t^2]$, or that the complex signal

$$w(t) = \exp[j(\omega_0 t + a^2t^2)] \tag{30}$$

is not an AS. This point is easily verified by calculating its FT which is

$$W(\nu) = \frac{\sqrt{\pi}}{a}\exp(j\pi/4)\exp\left[-j\,\frac{\pi^2(\nu-\nu_0)^2}{a^2}\right], \tag{31}$$

where $\nu_0 = 2\pi/\omega_0$. This function satisfies $|W(\nu)|^2 = \pi/a^2$ and therefore cannot be zero for negative frequencies. Furthermore it does not decrease for large values of $|\nu|$. However this decreasing property is the argument used in [8] and [20] to indicate that any complex signal in the form $\exp[j\omega t + j\phi(t)]$ is an asymptotic AS, which means that its FT for negative values of ν tends to zero when $\omega_0 \to \infty$. It is therefore necessary to study more carefully the problem and for this purpose the simplest way is to calculate the Hilbert transform $y_a(t) = H[x(t)]$ of $x(t) = \cos[\omega_0 t + a^2t^2]$. After some algebra not presented here one finds

$$y_a(t) = \sqrt{2/\pi}\{[C(\theta) + S(\theta)]\sin[\omega_0 t + a^2t^2] + [C(\theta) - S(\theta)]\cos[\omega_0 t + a^2t^2]\}, \tag{32}$$

with $\theta = (\omega_0/2a) + at$. The functions C and S are defined by

$$C(x) = \int_0^x \cos u^2 du \ , \ S(x) = \int_0^x \sin u^2 du. \tag{33}$$

As $C(\infty) = S(\infty) = \sqrt{\pi/8}$, we see that when $\theta \to \infty$, $y_a(t)$ tends to $\sin[\omega_0 t + a^2t^2]$. This shows that $w(t)$ given by (30) is in fact an asymptotic AS, but not for the reasons indicated in [20].

In order to discuss this point we present some numerical results to evaluate the errors introduced when assuming that the AS of $\cos[\omega_0 t + a^2t^2]$ is $\sin[\omega_0 t + a^2t^2]$. To simplify the calculations we assume that $\omega_0 = 0$, which does not restrict the meaning of the results because the effect of ω_0 is the same as the one of t, according to the definition of θ.

The parabolic chirp signals $\cos(\pi t^2)$ and $\sin(\pi t^2)$ are represented in Fig. 5 and previous equations show that these two functions are not Hilbert transforms. This appears in Fig. 6 where the Hilbert transform $y_a(t)$ of $\cos[\pi(at)^2]$ and the difference $\epsilon(t) = \sin[\pi(at)^2] - y_a(t)$ for $a = \sqrt{\pi}$ are presented. This function is the error made when assuming that the Hilbert transform of $\cos[\pi(at)^2]$ is $\sin[\pi(at)^2]$. As a consequence the IA of $\cos[\pi(at)^2]$ is not 1, and this appears in Fig. 7 where the IA $a(t) = \sqrt{x^2(t) + y_a^2(t)}$ is presented. As the IF is an increasing function this figure presents the relative error when assuming a linear IF.

All these results show that the parabolic chirps are not phase signals, but tend to have this structure either when the time t or the frequency ω_0 are increasing. Therefore the general intuition that such signals are asymptotic phase signal is verified, but the reason is not the elementary argument presented in [20]. In reality the FT given by (31) has a constant modulus, but there are also very rapid oscillations as the signal itself. This appears on the figure for large values of t. Because of these oscillations, even if the FT is not vanishing, its integral in any domain of the negative frequencies tends to zero because of the rapid changes between positive and negative values.

Let us now consider the signal $w(t) = \exp(ja^3t^3)$. As $w(t) = w^*(-t)$, its FT $W(\nu)$ is real. Its expression can be obtained by using some properties of the Bessel functions $J(z)$ and $K(z)$ and after calculations not reproduced here one obtains the following results. For positive frequencies we have

$$W(\nu) = \frac{2\pi}{3a}\sqrt{x}[J_{1/3}(2x\sqrt{x}) + J_{-1/3}(2x\sqrt{x})] \quad (34)$$

with $x = 2\pi\nu/3a$. On the other hand for negative values of the frequency ν we have

$$W(-\nu) = \frac{2}{\sqrt{3}a}\sqrt{x}K_{1/3}(2x\sqrt{x}) \quad (35)$$

with $x = -2\pi\nu/3a$. This last equation shows that $w(t)$ is not an AS because its FT is not zero for negative frequencies. The form of this FT will be shown in the oral presentation of this talk. We can only note that there is no spectral line in this FT and that it decreases to zero for infinite frequencies, which is a strong difference with the FT of the parabolic chirp. On the other hand it is much more difficult to calculate the Hilbert transform of $x(t) = \cos(\omega_0 t + a^3 t^3)$.

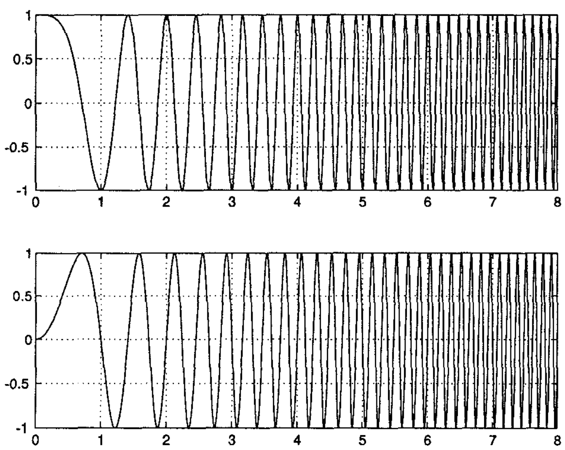

Figure 5: Parabolic chirps $\cos(\pi t^2)$ and $\sin(\pi t^2)$

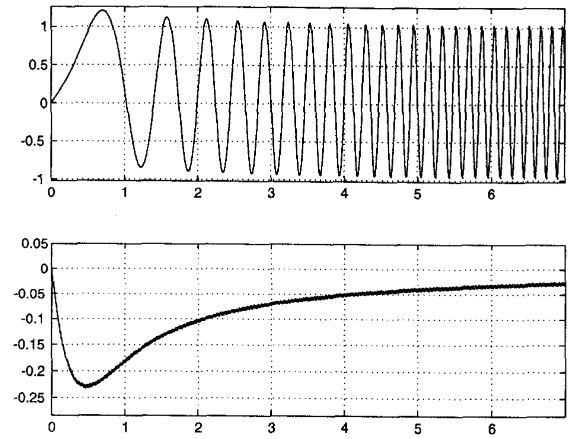

Figure 6: Hilbert transform of $\cos(\pi t^2)$ and error $\epsilon(t) = \sin(\pi t^2) - y_a(t)$

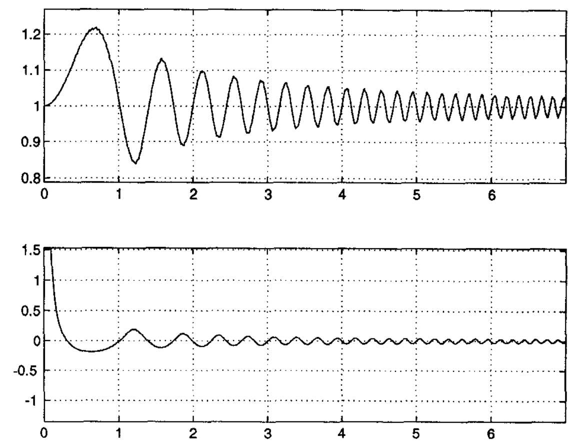

Figure 7: IA and relative error of the IF of $\cos(\pi t^2)$

6. REFERENCES

[1] Picinbono B., "On instantaneous amplitude and phase of signals", *IEEE Trans. on Signal Processing*, vol. 45, pp. 552-560, March 1997.

[2] Gabor D., "Theory of communications", *J. IEE*, 93, pp. 429–457, 1946.

[3] Ville J., "Théorie et applications de la notion de signal analytique", *Cables et Transmissions*, 2, pp. 61–74, 1948.

[4] Oswald J., "The theory of analytic band-limited signals applied to carrier systems", *IRE Trans. Comm. Theory*, 3, pp. 244–251, 1956.

[5] Dugundji J., "Envelopes and preenvelopes of real waveforms", *IRE Trans. Inf. Theory*, 4, pp. 53–57, 1958.

[6] Bedrosian E., "The analytic signal representation of modulated waveforms", *Proc. of the IRE*, 50, pp. 2071–2076, 1962.

[7] Vakman D., "On the definition of concepts of amplitude, phase and instantaneous frequency of a signal", *Radio Engineering and Electronic Physics*, pp. 754–759, 1972.

[8] Boashash B., "Estimating and interpreting the instantaneous frequency of a signal", *Proc. of the IEEE*, , vol. 80, pp. 520–538, 1992.

[9] Vakman D., "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency", *IEEE Trans. on Signal Processing*, vol. 44, pp. 791-797, April 1996.

[10] Picinbono B. *Principles of signals and systems*. London: Artech House, 1988.

[11] Bedrosian E., "A product theorem for Hilbert transforms", *Proc. of the IEEE*, 51, pp. 868–869, 1963.

[12] Picinbono B., "Représentation des signaux par amplitude et phase instantanées", *Ann. Télécommunications*, 38, pp. 179–190, 1983.

[13] Mandel L. "Interpretation of instantaneous frequency", *Amer. J. Phys.* , vol. 42, pp. 840-846, 1974.

[14] Loughlin P. and Tacer B. "On the amplitude and frequency-modulation decomposition of signals", *Journal Acoust. Soc. Am.*, vol. 100, pp. 1594-1601, September 1996.

[15] Loughlin P. "Comments on the interpretation of instantaneous frequency", *IEEE Signal Processing Letters*, vol. 4, pp. 123-125, May 1997.

[16] Poletti M. "The homomorphic analytic signal", *IEEE Trans. on Signal Processing*, vol. 45, pp. 1943-1953, August 1997.

[17] Kumaresan R. and Rao A. "Algorithm for decomposing an analytic signal into AM and positive FM components", *Proc. IEEE ICASSP*, 1998.

[18] Edwards S.F., Parrent G.B., "The form of the general unimodular analytic signal", *Optica Acta*, 6, pp. 367–371, 1959.

[19] Rudin W., *Real and complex analysis*. New-York: MacGraw-Hill, 1987.

[20] Nuttal A., "On the quadrature approximation to the Hilbert transform of modulated signals", *Proc. IEEE*, 54, pp. 1458-1459, 1966. 1959.