On the Structure of the Phase around the Zeros of the Short-Time Fourier Transform *

F. Jaillet[1], P. Balazs[1], M. Dörfler[1,2], N. Engelputzeder[1,2]

[1] *Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria,*
Email: florent@kfs.oeaw.ac.at, peter.balazs@oeaw.ac.at

[2] *Numerical Harmonic Analysis Group, Faculty of Mathematics, University of Vienna, Austria*

Introduction

The short-time Fourier transform (STFT) is a time-frequency representation widely used in audio signal processing. Two conventions are commonly used for the definition of the short-time Fourier transform :

$$V_f^g(\tau,\omega) = \int f(t)\overline{g(t-\tau)}e^{-i\omega t}\mathrm{d}t \qquad (1)$$

and

$$W_f^g(\tau,\omega) = \int f(t)\overline{g(t-\tau)}e^{-i\omega(t-\tau)}\mathrm{d}t. \qquad (2)$$

With convention (1), the transform uses a fixed phase reference at $t = 0$, whereas in convention (2), the phase reference at $t = \tau$ is sliding with the window. The two conventions are thus related by a simple phase term following the relation $W_f^g(\tau,\omega) = e^{i\omega\tau}V_f^g(\tau,\omega)$.

The interpretation of the modulus of the STFT is relatively easy, considering the fact that the spectrogram (defined as the square absolute value of the STFT) can be interpreted as a time-frequency distribution of the signal energy. This interpretation led to the important success of STFT in signal processing. In particular, it has been widely used for applications in speech processing and acoustics as a graphical tool for signal analysis.

But the interpretation of the phase of the STFT is less obvious, and has thus hardly been considered in applications. Nevertheless the phase can be of particular interest for certain applications, as illustrated by important applications such as phase vocoder [1] or reasignment [2]. Furthermore, for applications requiring the perfect reconstruction of a signal from STFT coefficients, phase information is essential. For this type of applications, in particular for applications using Gabor multipliers [3], which motivated the present study, better understanding of the structure of the phase is necessary to improve the processing possibilities.

One first reason for the difficulty of interpretation of the phase is the fact that in applications a subsampled version of the STFT is used. This leads to a subsampled version of the phase, which introduces aliasing effects in the phase values. Therefore, in this study we use the full STFT to avoid this problem. That is to say that the numerical experiments are done using for the STFT a hop size of 1 signal sample between two consecutive windows, and using a FFT length equal to the length of the analyzed signal.

But even using a full STFT, phase is not easy to interpret directly. In fact, it is more interesting to consider the phase derivative over time or frequency. Indeed, these quantities appear naturally in the context of reassignment [2]. Their interpretation is easier, as the derivative of phase over time can be interpreted as local instantaneous frequency while the derivative of the phase over frequency can be interpreted as a local group delay.

The phase derivative over time is of particular interest for analysis of signals containing sinusoidal components, as often encountered in acoustics [1]. This can be seen when considering the case of a single (complex) frequency defined by:

$$f(t) = e^{i\omega_0 t}, \qquad (3)$$

for which the STFT is then given by:

$$W_f^g(\tau,\omega) = e^{i\omega_0\tau}\overline{\hat{g}(\omega_0-\omega)}. \qquad (4)$$

The derivative of the phase over time is then $\frac{\partial}{\partial\tau}\arg(W_f^g(\tau,\omega)) = \omega_0$, which give direct access to the value of the angular frequency of the signal.

More generally, this property remains approximately true for sinusoids that evolve sufficiently slowly over time and that are sufficiently well separated in the time-frequency plane, giving the possibility to precisely estimate the frequency of the different partials of a sound.

Figure 1 illustrates this phenomenon on a signal constituted of two cosines with angular frequencies of 0.5 and 1.5. The modulus of the transform shows local maxima at these frequencies and in a large regions around the local maxima, the derivative of the phase of the STFT computed using definition (2) takes the value of the angular frequency of the closest sinusoid. What is happening at the intersection of the two regions is less obvious and the results in the last section will give insight for this point.

It is interesting to compare the phase derivative obtained using the two different definitions for the STFT. First it should be noted that the relation between the two functions is simple, as

$$\frac{\partial}{\partial\tau}\arg(W_f^g(\tau,\omega)) = \frac{\partial}{\partial\tau}\arg(V_f^g(\tau,\omega)) + \omega, \qquad (5)$$

*This work was supported by the WWTF project MULAC ("Frame Multipliers: Theory and Application in Acoustics", MA07-025).

making it easy to switch from one convention to the other. Depending on the signal and the analysis or processing task of interest, one or the other of the two conventions can be more suitable. In the example shown in Figure 1, convention (2) appears to be be easier to interpret.

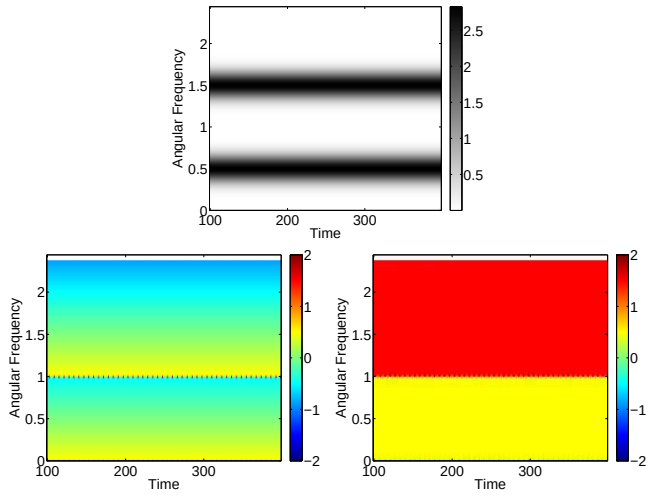

Figure 1: Representation for a sum of two sines. Top: modulus of the STFT. Bottom: derivative over time of the phase of the STFT using the definition (1) (left), and the definition (2) (right).

With this first result, we see that the interpretation of the phase derivative is achievable for simple signals such as the sum of slowly varying sinusoids. To extend ou understanding, we have to study the structure of the phase for complex signals, and in particular for noisy signals.

For this, we conducted numerical experiments on noise which are reported in the following section, and we checked that the observed patterns also appear for the continuous STFT on a simple analytical example in the last section.

Please note that in the following we focus on the study of the phase derivative over time of the STFT, but the same results can be obtained using the phase derivative over frequency of the STFT. In this case a very similar structure can be observed, with similar patterns in the neighborhood of the zeros of the STFT, but with a rotation of the patterns by a quarter turn in the time-frequency plane.

Observation for white noise

For noise, only statistical properties of the phase are accessible. Some interesting results for the phase derivative have been shown in the context of reassignment. In particular the study of the distribution of the phase derivative values appear in [4]. There, the following result is given. We consider a zero-mean Gaussian analytic white noise f such that

$$\mathrm{E}[\Re(f(t))\Re(f(s))] = \mathrm{E}[\Im(f(t))\Im(f(s))] = \frac{\sigma^2}{2}\delta(t-s) \tag{6}$$

and $\mathrm{E}[f(t)f(s)] = 0$ for any $(t,s) \in \mathbb{R}^2$, with its real and imaginary parts a Hilbert transform pair. Using a Gaussian window given by $g(t) = e^{-\frac{t^2}{2\sigma^2}}$, the phase derivative over time of V_f^g is a random variable with distribution of the form:

$$\rho(v) = \frac{1}{2(1+v^2)^{\frac{3}{2}}}. \tag{7}$$

This distribution is shown in Figure 2. As can be seen, it is a quite "peaky" distribution, indicating that the values of the phase derivative are mainly values close to zero, with some rare values with higher absolute values.

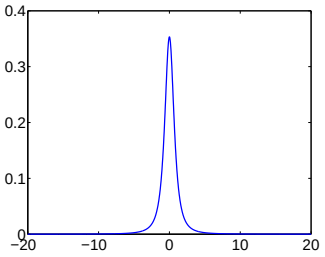

Figure 2: Distribution of the values of the phase derivative over time of the STFT for a white Gaussian noise.

But the information about the distribution of the values of the phase derivative does not give any clue about the spatial distribution of these values in the time-frequency plane. As accessing this information theoretically seems particularly difficult, we conducted systematic numerical experiments to study this spatial distribution.

For this, we need to compute the derivative of the phase in discrete settings. This can of course be done by approximating the derivative using a numerical scheme based on the discrete phase values (or even using the modulus values when using a Gaussian window, as shown in [5]), but an alternative efficient approach has been preferred.

We used the following expression given in [2] for the derivative:

$$\frac{\partial}{\partial\tau}\arg(V_f^g(\tau,\omega)) = \Im\left(\frac{V_f^{g'}(\tau,\omega)}{V_f^g(\tau,\omega)}\right), \tag{8}$$

with $g'(t) = \frac{\mathrm{d}g}{\mathrm{d}t}(t)$.

We see on this formula that we will face numerical difficulties when the denominator $V_f^g(\tau,\omega)$ is close to zero. But using double precision, these problems appear for really small values of the modulus (on the order of 10^{-13}), which allows us to reliably observe the values of the phase derivative even close to the zeros of the STFT. In the figures of this paper, the phase derivative values are ignored and represented as white at the points where the value of the modulus is too small. It is only noticeable in Figure 1 for high frequencies.

For computations, we used the linear time-frequency analysis toolbox [6], in which implementation for the STFT and the computation of the phase derivative may be found.

Our experiments gave surprising results, as illustrated by Figure 3. As can be seen on this figure, the time-frequency distribution of the values appears to be highly structured and in particular, the values of the phase derivative with high absolute values are concentrated around several time-frequency points, which can be identified as the zeros of the transform when looking at the modulus. Furthermore, the shape of the phase derivative seems to be very similar in the neighborhood of the zeros, with a typical pattern repeating at each zero. This typical pattern is represented on the third image of Figure 3. When going from low to high frequencies, it presents a negative peak followed by a positive one.

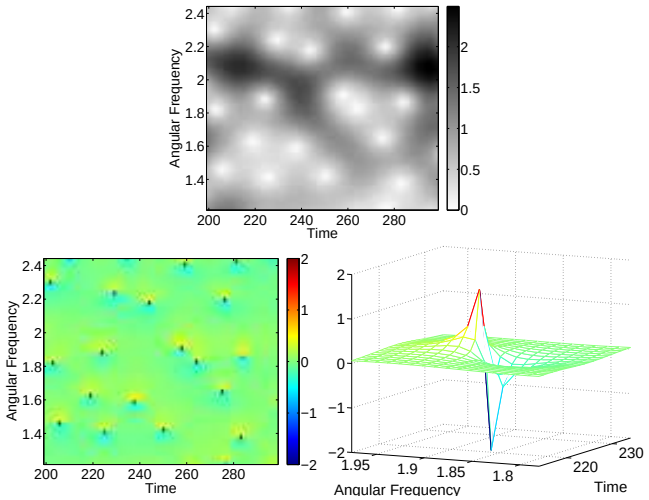

Figure 3: Observation for a Gaussian white noise, using a Gaussian window. Top: modulus of the STFT. Bottom-left: derivative over time of the phase of the STFT using the definition (1). Bottom-right: mesh plot of the derivative over time of the phase in the neighborhood of a zero of the STFT.

This phenomenon is related to the fact that the STFT of white noise is a correlated process, with a correlation determined by the window through the reproducing kernel of the transform (see part 6.2.1 of [7]). It is thus interesting to study the influence of the window choice on the observed structure of the phase derivative.

This influence is illustrated on Figure 4. Comparing Figure 3 and the first display of Figure 4, we can observe the effect of scaling of the window (or change of window length). We see that narrowing the window results in similar patterns around the zeros, but with a scaled shape: the resulting pattern is narrower over time, but wider over frequency.

Figure 4 also shows the influence of the window type. We see that for windows with a worse time-frequency concentration than the Gaussian window, the structure is more complicated. On the representation using a Hamming window, we still observe repeating patterns at the zeros of the transform, but the variability of the shape of this pattern seems higher, and the pattern orientation slightly varies, whereas it is fixed in the case of a Gaussian window.

For the case of the rectangular window, the zeros of the STFT form lines in the time-frequency plane instead of discrete points as for the two other types of window. This lead to much more variable patterns. Yet, we still interestingly observe that the values of the phase derivative with high absolute values concentrate around the zeros of the transform, whereas the phase derivative is close to zero in the regions of the STFT where the modulus is high.

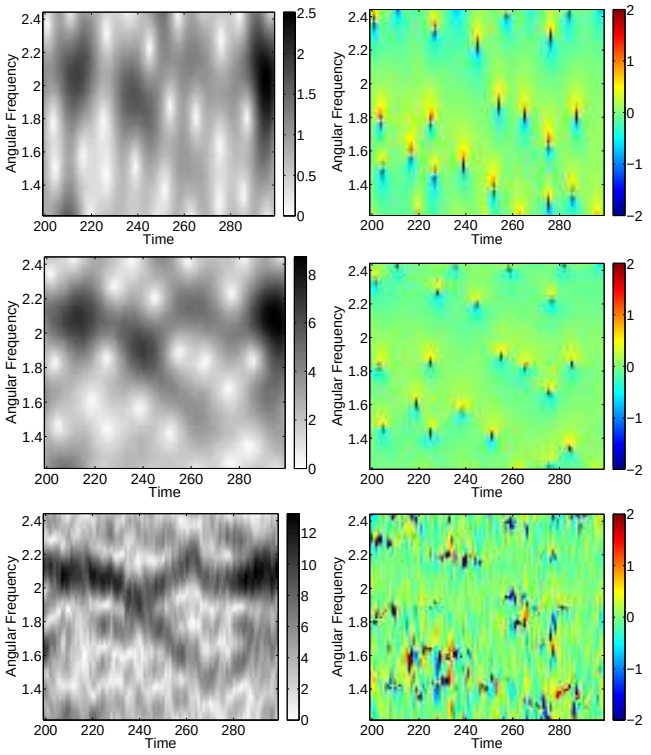

Figure 4: Influence of the window when analyzing a Gaussian white noise. For three different windows, on the left, modulus of the STFT, on the right, derivative over time of the phase of the STFT using the definition (1). From top to bottom, the windows are: a narrower Gaussian window, a Hamming window, a rectangular window.

The behavior that we observe for the noise is not specific to this kind of signal. Indeed further experiments on other synthesized and recorded complex sounds showed that the same characteristics can be observed for all signals: the values of the phase derivative of high absolute value are concentrated in the neighborhood of the zeros of the STFT, and for "nice" windows, a specific pattern appears in this neighborhood.

However, the numerical nature of these experiments limits the time-frequency resolution of the observation. Hence, we give a simple analytical example for which we can explicitly compute the expression of the phase derivative in the continuous case, and confirm the presence of the pattern observed numerically. This example is presented in the following section.

Analytical example

Considering the signal given by

$$f(t) = e^{i\omega_1 t} + e^{i\omega_2 t} \tag{9}$$

and using a Gaussian window $g(t) = e^{-\frac{t^2}{2\sigma^2}}$, we can explicitly compute the expression of the STFT, which results in the formula:

$$W_f^g(\tau,\omega) = \sqrt{2\pi}\sigma \left(e^{i\tau\omega_1} e^{-\frac{\sigma^2(\omega-\omega_1)^2}{2}} + e^{i\tau\omega_2} e^{-\frac{\sigma^2(\omega-\omega_2)^2}{2}} \right). \tag{10}$$

The zeros of this STFT are the points of coordinates (τ_k, ω_m) in the time-frequency plane, with $\omega_m = \frac{\omega_1+\omega_2}{2}$ and $\tau_k = \pi \frac{1+2k}{\omega_1-\omega_2}$ for $k \in \mathbb{Z}$.

The expression of the phase derivative for this signal, given in part VI-12 of [8], is:

$$\frac{\partial}{\partial\tau} \arg(W_f^g(\tau,\omega)) = \omega_m + \delta \tanh(s) \frac{1+\tan^2(\tau\delta)}{1+\tan^2(\tau\delta)\tanh^2(s)}. \tag{11}$$

with $\delta = \frac{\omega_2-\omega_1}{2}$ and $s = \sigma^2(\omega-\omega_m)\delta$.

The plot of this function around one of the zeros is visible in Figure 5. We see that the pattern observed is similar to the one obtained from numerical experiments in Figure 3.

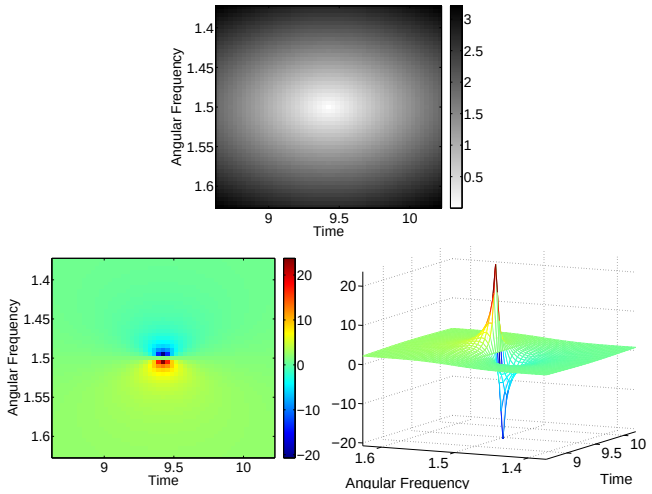

Figure 5: Observation for the signal defined in (9). Top: modulus of the STFT. Bottom: derivative over time of the phase of the STFT according to (11) represented as an image (left) and as a mesh (right).

From the expression of the phase derivative, we see that it is a continuous function everywhere except at the zeros of the STFT, and that at these points, we have the limits

$$\lim_{\omega\to\omega_m{}^-} \frac{\partial}{\partial\tau} \arg(W_f^g(\tau_k,\omega)) = -\infty \tag{12}$$

and

$$\lim_{\omega\to\omega_m{}^+} \frac{\partial}{\partial\tau} \arg(W_f^g(\tau_k,\omega)) = +\infty. \tag{13}$$

The results for this analytical example confirm that the structure observed in numerical experiments is present for the continuous transform. Furthermore, the expression of the phase derivative for this example may be used to look at other properties of the phase derivative near the zeros of the STFT.

Conclusion

While the phase derivative of the STFT allows interpretation for simple signals, its structure is not well understood for complex signals.

From numerical experiments, we have drawn the conclusion that the zeros of the STFT play a key role in its structure. In particular, high absolute values of the phase derivative seem to be concentrated, forming characteristic patterns in the neighborhood of the zeros. This fact should be taken into account when interpreting the phase derivative, as values are highly perturbed near the zeros.

A more theoretical description and characterization of the phenomenon would be suitable to proceed in the understanding of the phase structure of the STFT. We suggest to exploit the connection between the STFT computed using a Gaussian window and complex entire functions, as mentioned in [9].

References

[1] The Phase Vocoder: a Tutorial. M. Dolson. Computer Music Journal **10(4)** (1986), 14-27

[2] Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. F. Auger and P. Flandrin. IEEE Trans. Signal Process. **43(5)** (1995), 1068-1089

[3] A First Survey of Gabor Multipliers. H. G. Feichtinger and K. Nowak. In Advances in Gabor analysis. H. G. Feichtinger and T. Strohmer, editors. Chap. 5, 99-128. Birkhauser Boston, 2003

[4] On the Statistics of Spectrogram Reassignment Vectors. E. Chassande-Mottin, F. Auger and P. Flandrin. Multidim. Syst. and Signal Proc. **9(4)** (1998), 355-362

[5] Differential Reassignment. E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin. IEEE Signal Processing Letters **4(10)** (1997), 293-294

[6] Reference to the Linear Time/Frequency Toolbox (LTFAT) homepage. URL: `http://ltfat.sourceforge.net`

[7] Practical Time-Frequency Analysis: Gabor and Wavelet Transforms with an Implementation in S. R. Carmona, W.-L. Hwang, and B. Torrésani. Academic Press, Strasbourg, 1998

[8] Asymptotic Wavelet and Gabor Analysis: extraction of instantaneous frequencies. N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian and B. Torrésani. IEEE Transactions on Information Theory **38(2)** (1992), 644-664

[9] Sparse Time-Frequency Representations. T. J. Gardner and M. O. Magnasco. Proc. Natl. Acad. Sci. **103** (2006), 6094-6099