ForumAcusticum
2 0 0 5

Two representation tools to analyse non-stationary sounds in a perceptive context

Valentin Emiya
ENST, Département de Traitement du Signal et des Images, 46, rue Barrault, 75634 Paris Cedex 13, France,
PHASE, Univ. Paul Sabatier, 31062 Toulouse Cedex 09, France, e-mail: **valentin.emiya@enst.fr**,

Bertrand David
ENST, Département de Traitement du Signal et des Images, 46, rue Barrault, 75634 Paris Cedex 13, France, e-mail: **bertrand.david@enst.fr**,

Vincent Gibiat
PHASE, Univ. Paul Sabatier, 31062 Toulouse Cedex 09, France, e-mail: **gibiat@cict.fr**

Two time-frequency representations are demonstrated. The first one is based on reassigned Fourier spectrogram while the second one belongs to the family of the ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) High Resolution (HR) algorithms. Both methods benefit from recent improvements and their capabilities are illustrated through the analysis of steeldrum sounds and synthetic sounds designed for a perceptive study.

1 Introduction

Choosing the appropriate tool is a recurrent issue for sound analysis and representation. The amount of available schemes as spectrogram or wavelet transform often designed and tuned for specific applications seems to be a strong indication that looking for an universal solution is a hopeless quest. Two time-frequency representations are used here with applications to relevant contexts. Those tools have been designed in order to offer both an efficient analysis stage and an accurate and easy-to-read representation. They are based on Fourier analysis for the first one and on High Resolution methods for the second and then are complementary in terms of their usage and their abilities.

Both methods are first described, pointing out their most significant features. They are then applied to sounds selected for their perceptive interest: synthetic sounds studied for roughness perception are used to show how modulations can graphically emerge and steeldrum recordings are finally analysed to highlight processing and transient analysis, and component separation.

2 SAFIR

The Spectrogram in Amplitude and Frequency, Instantaneous and Reassigned (SAFIR [1]) is based on Short Time Fourier Transform (STFT), like spectrogram, on reassignment principles and on a few more mechanisms. Its design has been tuned to reach a line spectrum representation in the continuity of numerous precedent works [2, 3, 4, 5] with the intention to improve the intuitive aspect while spectrogram suffers from presence of lobes, due to the finite time support, that are not processed.

STFT coefficients are first computed using a N_{fft}-point Fast Fourier Transform (FFT) over N-point signal windows. In the usual spectrogram, the energy spreads over the graphical representation due to the analysis window and to the use of a regular scale in time and in frequency. Reassignment is a post-processing method to compute time and frequency pertinent coordinates for each STFT coefficient. Mathematically, time (or frequency) reassignment consists in a direct relation [6] between the center of gravity of STFT energy along time (or frequency) dimension and group delay (or instantaneous frequency) of STFT. From a signal processing point of view, one can consider STFT coefficients along a frequency channel k, seen as a temporal signal since computing STFT coefficients can be written as a band-pass filtering of the original signal:

$$\begin{aligned} X(n,k) &= \sum_{m=-N/2}^{N/2} x(m+n)w(m)e^{\left(-2i\pi\frac{km}{N_{\text{fft}}}\right)} \\ &= \left[h_k^w * x\right](n) \end{aligned} \tag{1}$$

$$\begin{aligned} \text{with } h_k^w(n) &= w(-n)\exp\left(2i\pi\frac{km}{N_{\text{fft}}}\right) \\ w &\quad \text{the sliding window of length } N \\ N_{\text{fft}} &\quad \text{the Fourier transform length} \end{aligned}$$

The filter band is centered in kF_s/N_{fft} with a bandwidth proportional to F_s/N, depending on the windowing. $X(n,k)$ is the analytic signal resulting from this filtering and whom energy $|X(n,k)|^2$ is plotted in spectrographic representations. In usual spectrograms, $|X(n,k)|^2$ is attributed to frequency kF_s/N_{fft}, with a significant step of F_s/N_{fft}. Nevertheless, a more meaningful frequency value is given by the instantaneous frequency $f(n,k) = d_t \arg\left(X(n,t)\right)/(2\pi)$ extracted from $X(n,k)$. If effects of unperfect bandpass filtering are ignoring, instantaneous frequency gives the frequency be-

havior of the subband signal. This corresponds to the instantaneous frequency obtained after filtering the analytic signal extracted from the original sound.

Practically, this reassignment leads to different results according to the subband signal content, which depends on the analysis window length. For single frequency signals, the instantaneous frequency matches the signal frequency itself, which differs from the subband center frequency. For other more complex high-energy signals, the combination of all components results in a variable instantaneous frequency, which can be studied and interpreted as a frequency modulation. This aspect is illustrated through applications in this article. Another case appear considering nearby subbands. As filter bandwidth is greater than distance between center frequencies (F_s/N_{fft}), overlap between nearby subbands leads to estimate the same instantaneous frequency in several subbands. To avoid those multiple occurences, SAFIR does not plot instantaneous frequency extracted from $X(n,k)$ if it is out of the interval $\left[k\frac{F_s}{N_{\text{fft}}} - \frac{F_s}{2N_{\text{fft}}}; k\frac{F_s}{N_{\text{fft}}} + \frac{F_s}{2N_{\text{fft}}}\right]$. A similar phenomenon also appears in distant subbands, due to secondary lobes and causes the current band content to be disturbed by the distant one. When the main lobe energy content is much larger than the attenuated secondary lobe content, effects are not significant. Otherwise, the analysis window length, which is the main parameter chosen by the user, should be adjusted to try to minimize this consequence of the time-frequency resolution limit.

3 HR-ogram

HR-ogram [7] stands for High Resolution Spectrogram since the signal is represented in the time-frequency plane with the help of a High Resolution method. The Pisarenko [8] or Prony [9] methods and MUSIC (MUltiple SIgnal Characterization [10]) belong to this category of algorithms. For our purpose, a subspace based High-Resolution method is utilized, relying on the so-called rotationnal invariance property, and thus, as the whole class of ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques [11]) benefits from recent signal processing advances. This algorithm is designed for particular signals modeled by overlapped blocks composed of exponentially damped sinusoids with additive white noise. Block $u \in N$ is defined for $t \in [0; N-1]$ by:

$$\begin{aligned} x(t_u + t) &= \sum_{m=1}^{M} a_{u,m} e^{d_{u,m} t} \cos\left(2\pi f_{u,m} t + \phi_{u,m}\right) \\ &\quad + w(t_u + t) \end{aligned} \tag{2}$$

where $w(t)$ represents an additive white noise.

For each local damped sinusoid m of block u, the algorithm determines parameters $\{a_{u,m}, d_{u,m}, f_{u,m}, \phi_{u,m}\}$ which are respectively called amplitude, real damping factor, frequency and initial phase. High Resolution methods possess the suitable property of not beeing constrained by the Fourier resolution limit. As sounds are frequently composed by local exponentially damped sinusoids and additive noise, which can be made white noise after a pre-processing stage, they fit the model required for HR analysis. This enables to discriminate close frequencies with a very accurate time resolution and high digit precision.

4 Applications

4.1 Phase effects in roughness perception

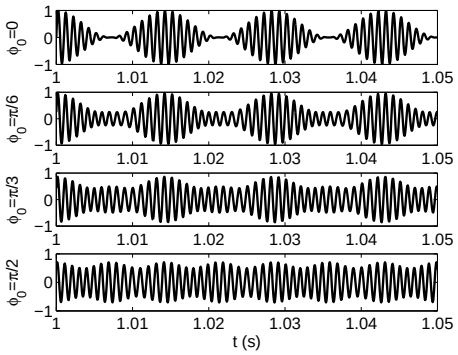

Figure 1: Roughness study: waveforms

SAFIR is here applied to the synthetic sounds proposed in the first experiment of [12], which establishes a relation between relative phases of components and roughness perception [13].

Sounds are defined by:

$$\begin{aligned} x_{\varphi_0}(t) &= \frac{1}{2}\sin(2\pi(f_c - f_m)t) + \sin(2\pi f_c t + \varphi_0) \\ &\quad + \frac{1}{2}\sin(2\pi(f_c + f_m)t) \\ \text{with } f_c &= 1000 \text{ Hz} \\ \text{and } f_m &= 70 \text{ Hz} \end{aligned} \tag{3}$$

They are composed of a center frequency with two side peaks with half amplitude. The side components are in phase whereas the central one has a relative phase φ_0 which is the object of the study and differs from one sound to the other. The distance between components at 1000 Hz is 70 Hz and causes a perceptive effect authors called roughness. When φ_0 varies, perceptive tests shows different degrees of roughness. The authors reports that the higher the roughness, the stronger the envelope modulation in the waveform. On figure 1, roughness is maximum when the 70-Hz modulation varies between 0 and 1

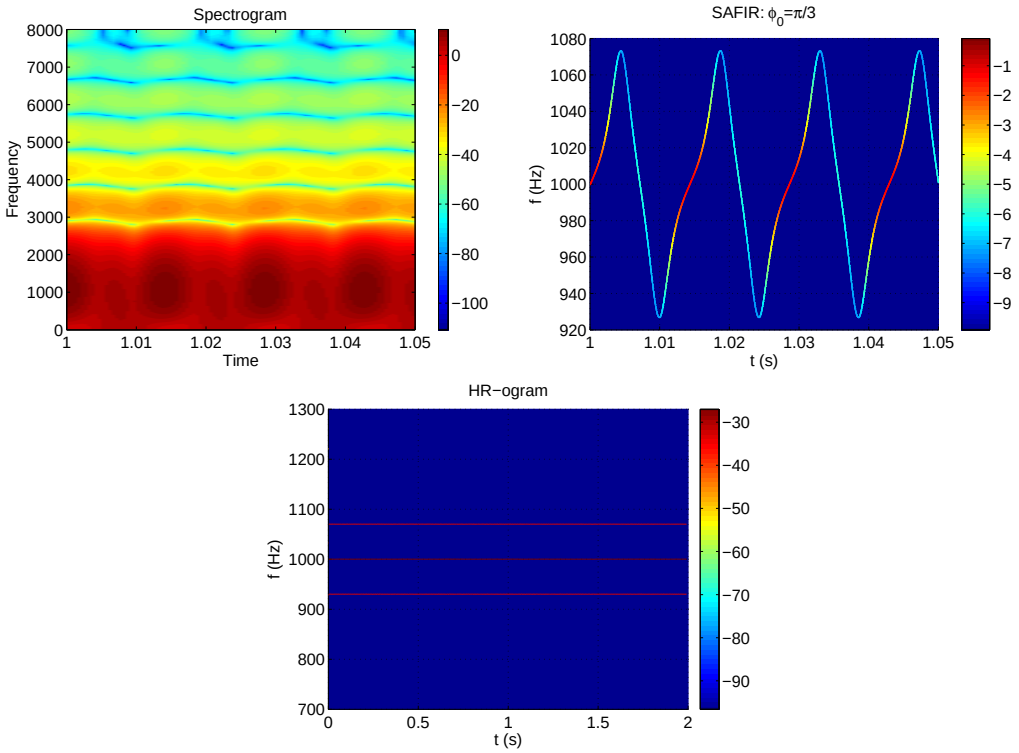

Figure 2: Roughness study: spectrogram (left), SAFIR (right) with 1ms hanning windows and HR-ogram (bottom) with 5.8 ms analysis windows, for $\varphi_0 = \frac{\pi}{3}$

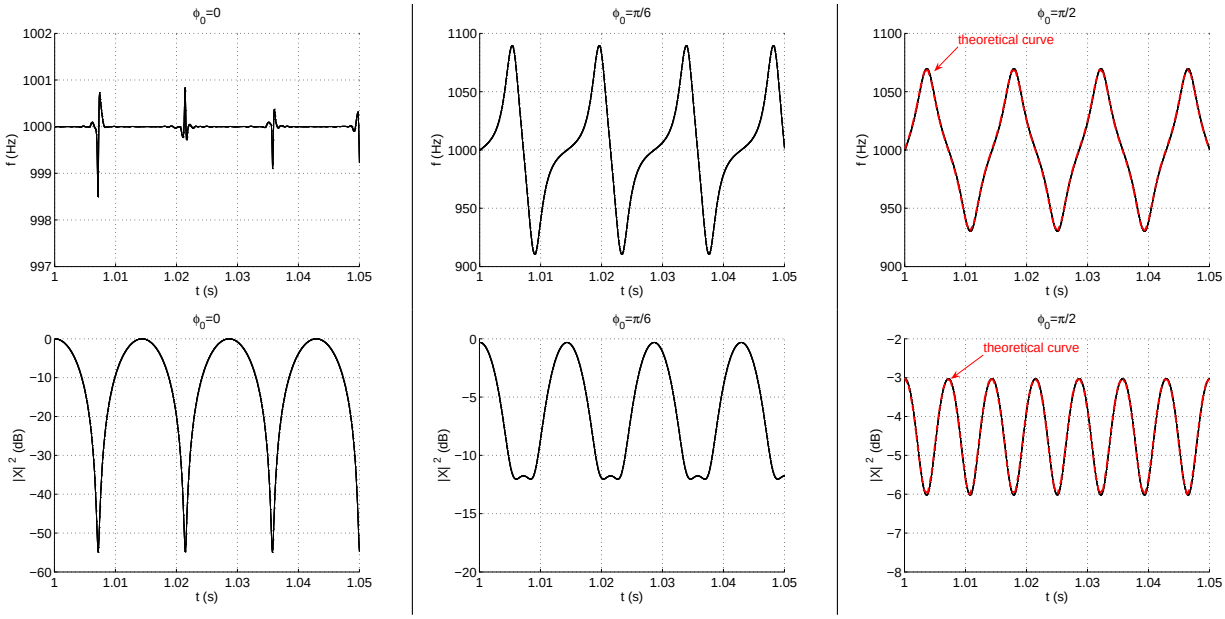

Figure 3: Extraction of frequency and amplitude modulations for $\varphi_0 = 0$, $\varphi_0 = \frac{\pi}{6}$ and $\varphi_0 = \frac{\pi}{2}$ (SAFIR with 1 ms hanning window)

($\varphi_0 = 0$) and is minimum when the enveloppe oscillates between 0.7 and 1 ($\varphi_0 = \pi/2$).

As shown in figure 2, in a short time analysis context, spectrogram does not give accurate results while HR-ogram and SAFIR leads to two complementary views: the former gives the decomposition of equation 3 and to find initial parameters, the latter allows to extend the perceptive results on roughness to the time-frequency domain by extracting amplitude and frequency modulations since all three components are grouped in the same subband. On figure 3, the theoretical modulation curves are plotted for $\varphi_0 = \pi/2$. In this case, rewriting equation 3 gives an amplitude modulation equal to $\sqrt{1+\cos(2\pi f_m t)}$ and a frequency modulation equal to $f_c + f_m \sin(2\pi f m t)/(1+\cos^2(2\pi f m t))$. When φ_0 varies, modulation frequency equals 70 Hz but modulation forms differ and can be detailed. For high roughness, modulations have high amplitudes, stiff slope and divergence points whereas they are much more regular for low roughness.

4.2 Beats in steeldrum sounds

In a musical context, beats are a particular case among all forms and auditory effects adopted under the generic term of modulation. Beats are found here in steeldrum sounds. Acoustical fundamentals of the steeldrum, which is considered as non-linear mode-localized oscillators, have been described in successive papers [14, 15, 16, 17]. Each note is localized to a specific area of the bottom of the barrel: the resulting sound is produced by a complex vibration of this surface, depending on material behavior, on imperfect isolation between each note area, on placement of areas and on stick hitting.

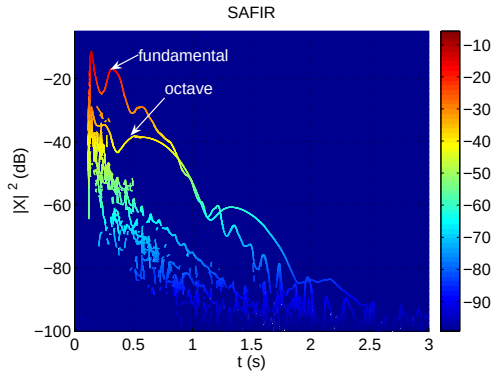

Figure 4: Amplitude behavior of first two partials

Figure 4 shows the analysis of a traditional tenorpan A3 note. The SAFIR is viewed in the amplitude-time plane and represents the frequency range between 200 Hz and 450 Hz, corresponding to the first two partials (fundamental and octave). In general, linear and non-linear couplings between more than two components make steeldrum sounds complex to analyse. They cause temporal effects on timbre that are part of specifities of steeldrum sounds [18]. One can note here that in the early part of the sound, amplitude modulations between fundamental and octave seem to be opposed in phase. Such behavior has been investigated by Achong in [14] where two or three coupled modes result in energy exchanges [19].

Beats are also present in steeldrum sounds in function of making: they are not similar from one instrument to another and vary from one steeldrum to another as well. A B4 note from a double pan has been selected to investigate the dissonance of this particular note. On figure 5, HR-ogram and SAFIR of this sound offer a complementary representation: HR-ogram shows pairs of components instead of single partials located along the harmonic series of B4 note. Distance between components of a pair is about 25 Hz, independently from the pair. As SAFIR is performed on subband wider than 25 Hz, a modulation is obtained at the location of partials, which better suits to the concept of dissonance one can have: one single object with more complexity than the line obtained with a single frequency.

By zooming on both figures, one can check that the modulation frequency on SAFIR corresponds to the distance between components in a pair on HR-ogram. On right part of figure 5, paired components are located at 976.5 Hz and 1000.7 Hz and measured modulation frequency is $1/0.0413 = 24.21$ Hz, which matches the difference $1000.7 - 976.5 = 24.2$ Hz. This example shows the interest to have several tools to analyse even the same sound. HR-ogram gives a reliable and objective decomposition into single frequencies whereas SAFIR is parametered to merge close components into subjective and perceptively meaningful variations in amplitude and frequency.

4.3 Attack transients

Time reassignment used by SAFIR improves analysis of non-stationary parts of sounds. Figure 6 focuses on the attack transient of a steeldrum note which is analysed by SAFIR. Without reassignment, at the attack time, the energy is spreading over the window size length, which is 32 ms long here, and causes multiple undesirable effects that can be avoided or at least minimized by time reassignment. First, localisation of peaks get more accurate and peaks themselves are narrower. Then time reassignment makes pre-echo effect disapear at all: the center of gravity of energy along time is estimated and cannot be located before the beginning of energy support. One can also observe that a similar effect makes amplitude behaviors of modal components get stiffer, which allows to estimate slopes, rise time and decay time to characterize transients.

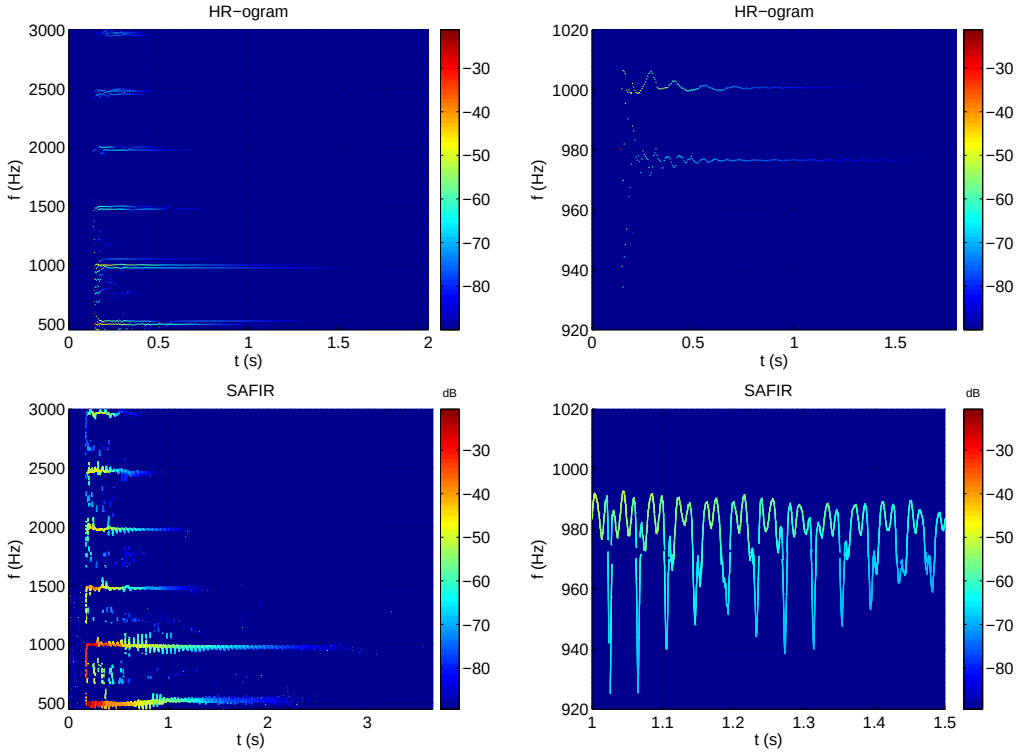

Figure 5: B4 played on second pan: HR-ogram (top) and SAFIR (bottom) with zoomed view (right).

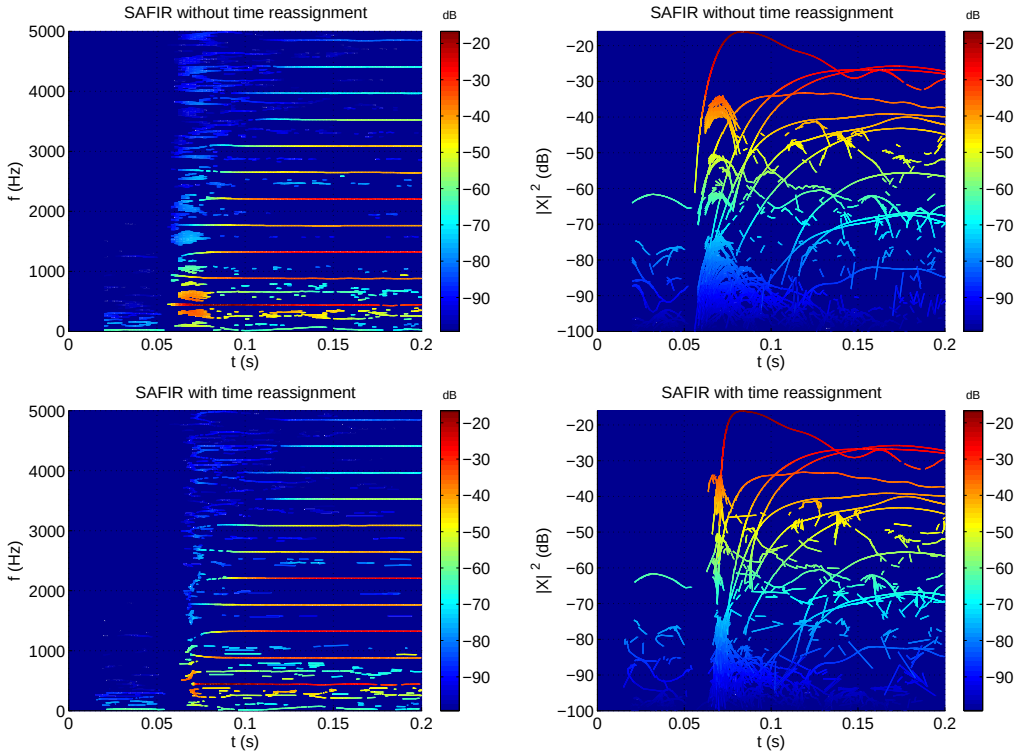

Figure 6: SAFIR of attack transient of A4 steeldrum sound: frequency (left) and amplitude (right) behaviors with (bottom) and without (top) time reassignment

Thus, time reassignment allows to avoid energy to spread over the sliding window by calculating the center of gravity of energy, which offers a good approximation for energy localisation. In addition to the precedent improvements, analysis can be performed with long windows to obtain both frequency accuracy and temporal event precision in non-stationary regions.

5 Conclusion

Two time-frequency representations have been shown here to offer an appropriate answer to several sound analysis issues: characterization of partials, decomposition sinusoidal components, extraction of amplitude and frequency modulations in relation with perception, time precision improvements, cancellation of pre-echo. This work has been done in the perspective to pay attention to quality of representation and to use the knowledge of several tools and techniques in order to benefit from complementary analysis and representations. This is all the more useful as perceptive studies of sounds need both usual, traditional tools like spectrogram and new ones that can take into account specificities and complexity of hearing.

References

[1] V. Gibiat, A. Padilla, V. Emiya, L. Cros, 'Phase characterization of soundscapes', *J. Acoust. Soc. Am.*, 117 (4), p. 2550 (2005)

[2] J. Ville, 'Théorie et applications de la notion de signal analytique', *Cables et transmission*, 2A (1), pp. 61-74 (1948)

[3] V. Gibiat, F. Wu, P. Perio, Sylviane Chaintreuil, André Banc-Lapierre, 'Signaux et Systèmes - Analyse Spectrale Différentielle (A.S.D.)', *C.R. Acad. Sc. Paris*, t.294, II, pp. 633-636 (1982)

[4] V. Gibiat, P. Jardin, F. Wu, 'Analyse spectrale différentielle - Application aux signaux de Myotis Mystacinus', *Acustica*, 63, pp. 90-99 (1987)

[5] F. Léonard, 'Spectrogramme de phase et spectrogramme de fréquence', *Traitement du Signal*, 17 (4), pp. 269-286 (2000)

[6] P. Flandrin, 'Temps-Fréquence', *Hermès* (1993)

[7] B. David, G. Richard, R. Badeau, 'An EDS modeling tool for tracking and modifying musical signals', *Proc. of SMAC'03*, 2, pp.715-718 (2003)

[8] V.F. Pisarenko, 'The retrieval of harmonics from a covariance function', *Geophysical J. Royal Astron. Soc.*, 33, pp. 347-366, 1973

[9] G.M. Riche de Prony, 'Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur d'eau et de la vapeur de l'alcool à différentes températures', *Journal de l'Ecole Polytechnique*, 1 (22), pp. 24-76 (1795)

[10] R.O. Schmidt, 'Multiple emitter location and signal parameter estimation', *IEEE Trans. Antennas Propagat.*, 34 (3), pp. 276-280 (1986)

[11] R. Roy, A. Paulraj, T. Kailath, 'ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise', *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, 34, pp. 1340-1342 (1986)

[12] D. Pressnitzer, S. McAdams, 'Two phase effects in roughness perception', *J. Acoust. Soc. Am.*, 105 (5), pp. 2773-2782 (1999)

[13] R.C. Mathes, R.L. Miller, 'Phase effects in monaural perception', *J. Acoust. Soc. Am.*, 19 (5), pp. 780-797 (1947)

[14] A. Achong, 'The steelpan as a system of nonlinear mode-localized oscillators, Part I: theory, simulations, experiments and bifurcations', *Journal of Sound and Vibration*, 197, pp. 471-487 (1996)

[15] A. Achong, K. A. Sinanan-Singh, 'The steelpan as a system of non-linear mode-localized oscillators, part II: coupled sub-systems, simulations and experiments', *Journal of Sound and Vibration*, 203, pp. 547-561 (1997)

[16] T.D. Rossing, U.J. Hansen, D.S. Hampton, 'Vibrational mode shapes in Caribbean steelpans. I. Tenor and double second', *J. Acoust. Soc. Am.*, 108 (2), pp. 803-812 (2000)

[17] B. Copeland, A. Morrison, T.D. Rossing, 'Sound radiation from Caribbean steelpans', *J. Acoust. Soc. Am.*, 117 (1), pp. 375-383 (2005)

[18] P. Gaillard, 'Etude de la perception des transitoires d'attaque des sons de steeldrums: particularités acoustiques, transformation par synthèse et catégorisation', Doctoral Thesis, Université de Toulouse Le Mirail (2000)

[19] , G. Weinreich, 'Coupled piano strings', *J. Acoust. Soc. Am.*, 62 (6), pp. 1474-1484 (1977)