INSTANTANEOUS FREQUENCY ATTRACTORS AND AUDITORY MASKING THRESHOLD CALCULUS

Amelia Ciobanu, Cristian Negrescu, Dumitru Stanomir

Telecommunication Department, University "Politehnica" of Bucharest

ABSTRACT

Based on the widely spread concept of instantaneous frequency (IF), we resumed the IF attractors (IFAs) method used for tonal components extraction from any given signal and we emphasized its essential advantages over traditional approaches by incorporating it into a popular application: the calculus of auditory masking threshold (AMT). Using MPEG1 Layer 1 psychoacoustic model, we propose to replace the classic algorithm for tonal/noise maskers identification with the IFAs method. The new approach proved a superior behavior regarding the accuracy estimation of tonal component confirmed by comparative tests. Also, we report notable AMT magnitude differences, between reference method and new approach.

Index Terms— Instantaneous frequency, instantaneous frequency attractors, auditory masking threshold

1. INTRODUCTION

The traditional approach for time-frequency analysis is the well known Fourier transform or its more convenient version the short time Fourier transform (STFT). Although this mathematical instrument provides very good results for stationary signals whose frequency content remains unchanged, it is not very well suited for nonstationary signals, due to the compelled compromise between time and frequency resolution which leads to poor frequency accuracy. Since nonstationary signals are quite present in mainstream applications (e.g. applications using chirp signals, audio signal, speech signal, etc.), a more appropriate time-frequency analysis was sought: the instantaneous frequency (IF) analysis.

The concept of IF relates to any process involving time-varying spectral features found in nonstationary signals. The usage of IF is found in diverse technical fields: from seismic field [3], to radar, sonar, and biomedical applications [4]. In general, the IF of a nonstationary signal indicates the spectrum position of the signal's spectral peak as it varies with time. Hence, the IF can be viewed as the frequency of a sine wave which best fits the signal being analyzed, at a given time instant. When dealing with signals with more than one spectral component, a preliminary subband decomposition is mandatory [4]. Regarding IF estimation, many approaches rely on Wigner-Ville distribution and its variations, while others concentrate on obtaining IF through the analytical signal [4-6]. The solution adopted in this paper is similar with the one presented in [8], and uses a complex band pass filter (BPF) bank in order to decompose the analyzed signal.

The concept of IF attractors (IFAs) was first introduced in [1], and improved in a particular way in [2]. IFAs combine the accuracy of the IF analysis with the subband spectrum exploration to successfully avoid false frequency components (given by the smearing artifacts of the windowing) and determine the actual tonal components, even for low signal to noise ratios (SNR). The importance of pinpoint accuracy for tonal components is crucial for any application which relies on tonal components analysis (e. g. speech and audio models [5], the calculus of the auditory masking threshold (AMT) [10]).

In this paper we stress the importance of IF analysis in identifying tonal components of a given signal and we provide as a meaningful example its application to the calculus of the AMT. We review a fairly new introduced method for tonal components estimation with great accuracy, based on IF attractors [1-2]. Also we perform a comparison between a classic method for computing the AMT based on STFT and the new method based on IFAs. The result of the comparison shows the superiority of the latter method.

The paper is organized as follows. Section 2 highlights the important aspects of IF analysis and IFAs, section 3 presents general considerations regarding AMT and describes the proposed solution of using IFAs for AMT calculus and section 4 provides conclusive results obtained for the new approach when compared with the reference. Finally section 5 is reserved for conclusions.

2. TIME-FREQUENCY ANALYSIS USING IF ATTRACTORS

The IF analysis in conjunction with IF attractors offers an elegant and accurate method for the estimation of tonal components of any given signal, especially when compared with the traditional STFT analysis whose accuracy is

bounded by the compelled compromise between time and frequency resolution and the uniformly spaced frequency bins of the Fourier spectrum. The first part of the current section presents briefly the main steps for estimating the IF spectrogram, while the second part resumes the IFAs theory and surveys the discrete version of the IFAs algorithm.

2.1. IF analysis

In the context of the well known sinusoidal model [7], Abe et al. [8] introduce the idea of estimating the tonal components of any given signal through the help of IFs.

The main idea of the solution adopted in [8] is to decompose the analyzed signal, $s(t)$ into single component signals or narrow frequency band signals, by using a complex BPF bank. For each signal obtained through complex filtering the IF is estimated. This frequency is equivalent to the frequency of the best cosine wave that approximates the real part of the band pass filtered signal. The filter bank is elegantly built by modulating a causal and real prototype low-pass filter (LPF), whose impulse response is denoted $w(t)$. The impulse response function of the p^{th} filter, characterized by the central frequency $\Omega_p > 0$ will be

$$h_p(t) = w(t)e^{j\Omega_p t} \in \mathbb{C}, \quad (1)$$

while the output of the p^{th} filter, when the input is fed with $s(t)$, will be denoted $s_{fp}(t)$. Assuming that in the p^{th} channel we have a tonal component, then $s_{fp}(t)$ is completely characterized by its instantaneous amplitude $A_p(t)$ and argument $\theta_p(t)$, which represents the instantaneous phase. Furthermore, the sought IF information, can be extracted using (2)

$$\omega_p(t) = \frac{d}{dt}\arg\{s_{fp}(t)\}. \quad (2)$$

However, the direct usage of (2) is not recommended, therefore the advantage of filter bank approach combined with STFT was preferred. Namely, the original signal can be analyzed as the Fourier transform, $S(\omega, t_a)$ of the particular signal $s(t)w(t_a - t)$. Considering time as variable and a fixed frequency Ω_n, $S(\omega, t_a)$ can be translated into (3)

$$S(\Omega_n, t) = (u * w)(t), \text{ with } u(t) = s(t)e^{-j\Omega_n t}. \quad (3)$$

Now, using the filter bank approach, $s_{fp}(t)$ can be viewed as the product between $S(\Omega_n, t)$ and $e^{j\Omega_n t}$. The major advantage of this solution resides from the fact that, when computing the IF based on (2), the derivative can be commuted to $w(t)$, rather than $u(t)$ (due to a convolution property). Thus, a priori knowing the analytical expression of the analysis window (which is identical to the LPF prototype), there is no need for numeric derivative. Finally, IF is computed using (4), as the derivative of the phase of the analytic signal [9]:

$$\omega_{inst,n}(t_a) = \frac{R_{\Omega_n}(t_a)\dot{I}_{\Omega_n}(t_a) - I_{\Omega_n}(t_a)\dot{R}_{\Omega_n}(t_a)}{R^2_{\Omega_n}(t_a) + I^2_{\Omega_n}(t_a)}, \quad (4)$$

where R_{Ω_n} and I_{Ω_n} represent the real and imaginary part of $s_{fp}(t)$. The notation $\dot{(\,)}$ stands for the derivative.

2.2. IF attractors

The problem addressed by this section is the identification of tonal components in any given signal. It is well known that simple peak picking (PP) method delivers poor results due to the imprecision of distinguishing between real and false tonal components. The IFAs method overcomes this problem and provides conclusive results, even for low SNR. Its results are boosted by the usage of IF spectrogram.

The main idea of the IFAs method is to explore the IF spectrogram with the help of a BPF bank, obtained through STFT. The task of each channel of the filter bank is to capture (or "to attract") a tonal component. The exploration of the spectrum should be done in very small steps (large number of channels), so that each channel will attract no more than one component. If no tonal component exists, then the output of the channel will be its central frequency, otherwise it will be the IF of the corresponding tonal component.

According to [1-2], the IFAs corresponding to the tonal components are the frequencies satisfying the following conditions:

$$\mu(\Omega_{IF}, t_a) = 0, \quad (6)$$

$$\left.\frac{\partial \mu(\Omega, t_a)}{\partial \Omega}\right|_{\Omega=\Omega_{IF}} \in [-1-\varepsilon, -1+\varepsilon] \quad (7)$$

where

$$\mu(\Omega, t_a) = \omega_{inst,\Omega}(t_a) - \Omega. \quad (8)$$

In the above equations Ω stands for the central frequency of a channel, t_a is the current analysis moment, and $\omega_{inst,\Omega}$ represents the sought IF. The positive constant ε strengthens the robustness to noise and adds flexibility to the method. Depending on the SNR or the type of application involved, the value of ε can be adjusted between 0 and 1.

The discrete version of the IFAs method follows the same idea as its analog version and includes four main stages: a) perform IF analysis; b) plot the IF against the central frequencies of the BPF bank; c) search the previous plot for plateau regions; d) for each plateau region compute the frequency, amplitude and phase of the IFA. In stage c), another parameter which increases the algorithm's robustness to noise was introduced. This parameter indicates the minimum number of channels which should attract the

same tonal component, denoted W. Details regarding each stage can be found in [2].

3. APPLICATION OF IF ATTRACTORS

In the context of the sinusoidal model [7], IFAs proved to be a superior alternative to the classical PP method (reduction of the arithmetic complexity, increase of the synthesized speech quality) [1-2]. In the current section we broaden the usage of IFAs, by introducing this concept into the AMT calculus. As it will be further shown, the impact of such an approach is at least notable.

3.1. General considerations regarding AMT estimation

Psychoacoustic considerations offer consistent support in various signal processing applications, such as loudness estimation, audio compression, noise suppression, speech recognition, speech and audio synthesis. One largely used perceptual feature is the AMT. For instance, in perceptual audio compression, the usage of such a feature enables the possibility to reduce the bitrates without affecting the high quality of the transmitted audio signal.

The general computation framework of the AMT is based on several perceptual concepts: critical band, absolute threshold of hearing (ATH), (non)simultaneous masking and spread of masking [11]. There are different approaches regarding the computation of the AMT (e.g. [12]), but in the context of IFAs, only the approaches which require a stage for tonal/noise maskers identification are of interest (e.g. [10-11], [13-14]). For the next part of this paper we will consider the psychoacoustic model of MPEG1 Layer 1 [14] as a reference for AMT calculus.

3.2. Incorporating IF attractors in AMT algorithm

The main stages of MPEG1 Layer 1 AMT algorithm are [11], [14]: 1) spectral analysis and sound pressure level (SPL) normalization; 2) identification of tonal and noise maskers; 3) decimation and reorganization of maskers; 4) computation of individual masking thresholds; 5) computation of global masking threshold.

Next we focus only on the second stage of this algorithm. Details regarding the rest of the stages can be found in [11]. The classic approach for stage 2 requires finding the tonal components, and then from the remaining components, a noise masker for each critical band is computed. The frequency of one noise masker is computed as the geometrical mean of the nontonal components in the corresponding critical band. In order to identify the tonal maskers, the classic algorithm first searches all the local maxima and then keeps only those peaks which fulfill a sharpness condition.

We propose to replace the traditional PP method used in stage 2, with the IFAs method described in section 2. As a consequence the set of tonal maskers, delivered by the IFAs, will be more accurate with respect to their total number, frequency, amplitude, and phase. Due to the modification of the number and position of the tonal maskers, the position and amplitude of the noise maskers will also be changed. Another important side effect is the calculus of the spreading function, which now can be placed on the exact frequency of the tonal maskers. Moreover, these changes do not remain without echo on the global masking threshold. It will be further shown that the usage of IFAs produces a modification, on average, of approximately 2 dB on the magnitude of the AMT, depending on the SNR scenario or the frequency position of the tonal components.

4. EXPERIMENTS AND RESULTS

In order to quantify the changes introduced by the IFAs in the calculus of AMT, we considered two types of scenarios. One type of scenario uses test signals generated with known parameters (e.g. number of tonal components, SNR), while the second type of scenario uses fragments of high quality audio signal or high quality excerpts taken from a speech database. The sampling frequency, F_s used for all the test signals was ≥ 24000 Hz.

Scenario 1. For this type of scenario we have generated signals obtained as the sum between a periodic signal (a finite sum of sinusoidal components with known frequencies, amplitudes and phases) and a nonperiodic signal (white noise, with known power). The test signals' SNRs were imposed. We divided the tests for this scenario into tests with high SNR (higher than 30 dB), which will further be referred to as S1H and tests with low SNR (less than 30 dB), referred to as S1L. For both types of tests we have investigated the frequency estimation accuracy of the tonal components for the IFAs method and its effects over the spreading function and the global masking threshold. During all tests the frequencies of the tonal components were placed between two frequency bins, as these are the cases known to produce the most estimation errors. The most unfavorable situation leading to the highest frequency estimation error – FEE, (defined as the difference between original and estimated frequency) is when the frequency is placed at half distance between two consecutive bins.

We considered an example with three tonal components placed at $f_1 = 555\,\text{Hz}$, $f_2 = 1234\,\text{Hz}$, and $f_3 = 5111\,\text{Hz}$, with corresponding frequency bins 5.92, 13.16 and 54.51 (for $F_s = 24\,\text{kHz}$ and a 12 ms analysis frame). Figure 1 displays the power spectrum for this example (gray solid line). We performed a large number of tests for this type of signal, using different SNR levels. First, we focused our attention on the frequency estimation accuracy. We estimated the frequency of the tonal components with both IFAs and PP method, and we compared the results with original values, resulting FEE. The tests results were

analyzed through the mean and standard deviation of FEE ($\overline{FEE_{f_3}}$ and $\sigma_{FEE_{f_3}}$).

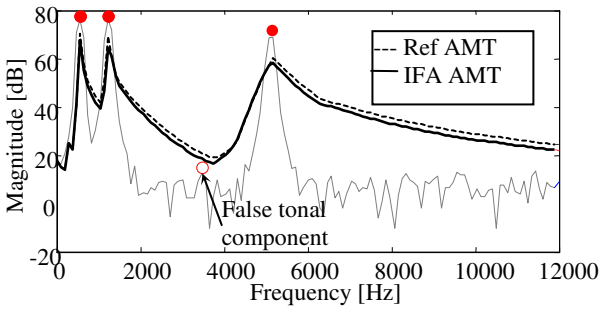

Figure 1 – Comparison between reference AMT (dotted line, ○) and IF AMT (dark solid line, ●) – S1H tests example

Table 1 presents a part of the results we obtained – the results for the most unfavorable case, frequency f_3 .

Table 1 Comparative test results – IFAs vs. classic approach

SNR	54 dB (S1H)		20 dB (S1L)		10 dB (S1L)	
	IFAs $\varepsilon = 0.2$ $W = 5$	PP	IFAs $\varepsilon = 0.35$ $W = 3$	PP	IFAs $\varepsilon = 0.45$ $W = 3$	PP
Frequency estimation accuracy						
$\overline{FEE_{f_3}}$ [Hz]	$7.6 \cdot 10^{-5}$	45.25	0.014	29.94	$-1.7 \cdot 10^{-3}$	10.8
$\sigma_{FEE_{f_3}}$ [Hz]	0.06	0	2.24	34.65	5.42	45.2
AMT magnitude differences						
$\overline{\Delta L}$ [dB]	−1.98		−1.026		−1.32	
$\sigma_{\Delta L}$ [dB]	0.967		2.5		3.41	

The results in Table 1 show that even for low SNR levels, the IFAs estimation is very accurate. We obtain an average error close to zero, and a standard deviation around 5 Hz, as opposed to the PP method where the error is higher with four magnitude orders. Regarding the number of false components, for high SNR levels, with IFAs method we always detect the exact number of tonal components. However the PP method, for most of the analyzed frames introduces false tonal components. This aspect is also visible in Figure 1. Our tests reveal that in 72.5% from the total number of analyzed frames the PP method leads to false tonal components. For low SNR levels, the IFAs method starts to introduce such components, but their number is less than the number introduced by the PP method and can be kept under control with the two parameters ε and W . Considering the findings above, we regard the IFAs method as an excellent choice for tonal components detection algorithms.

Next we investigate the impact of IFAs introduced in the calculus of AMT. One important effect, represented in Figure 2, is the influence over the spreading function (SF). The shape of SF for a tonal masker, when using the IFAs, follows closely the power spectrum of the analyzed frame, whereas for the PP method the peak of the SF appears displaced from its actual position. This effect appears regardless of the SNR level and is more prominent when the tone frequency is placed between bins (see Figure 2).

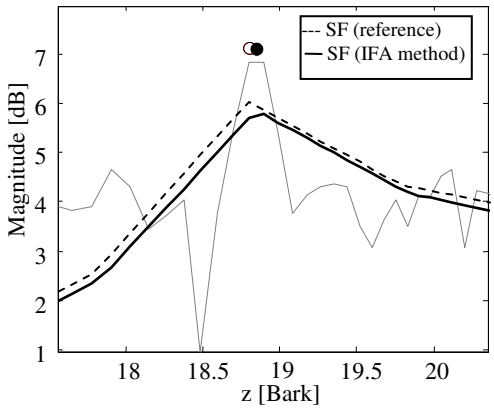

Figure 2 – Spreading function for a tonal component detected with IFAs method (dark solid line, ●) and with PP method (dotted line, ○)

The difference observed for the SFs, obtained when using the IFAs method and the classic method, is further visible in the AMT magnitude, especially in the frequency bands around the tonal maskers. For S1H tests, the magnitude difference, denoted ΔL , is kept constant from the position of the tonal masker until the end of the spectrum, if no other masker interferes. This is the case for f_3 frequency in Figure 1 (dotted line – reference AMT, solid dark line AMT obtained using IFAs – IFA AMT). If another tonal/noise masker follows (as it is the case for f_1 and f_2), then ΔL is preserved only for a certain frequency band. Outside that frequency band, ΔL has different degrees of variation, depending on the existence of false tonal components or the particular frequency content of the signal.

The effects mentioned above were observed during the tests we performed. The signals used in tests had a frequency content similar to the example in Figure 1. Second part of Table 1 provides the mean and standard deviation of ΔL ($\overline{\Delta L}$ and $\sigma_{\Delta L}$) for different SNR levels. For S1H test, $\overline{\Delta L}$ is around 2 dB. The small value for $\sigma_{\Delta L}$ signifies that the set of ΔL has tight grouped values, which confirms the fact that in the absence of another tonal masker, ΔL is kept constant throughout the spectrum. Also, these results indicate that, although the PP method introduces false tonal components, this aspect does not impact on the global masking threshold. These false components always lie below the AMT.

For S1L tests, $\overline{\Delta L}$ is around 1 dB, but $\sigma_{\Delta L} > \overline{\Delta L}$, which implies that the ΔL data contains values spread over a wide range. As expected, the existence of false tonal components now greatly influences the AMT, causing a

large variation of ΔL (see Figure 3). Again, the results are in agreement with the predicted behavior of ΔL variation.

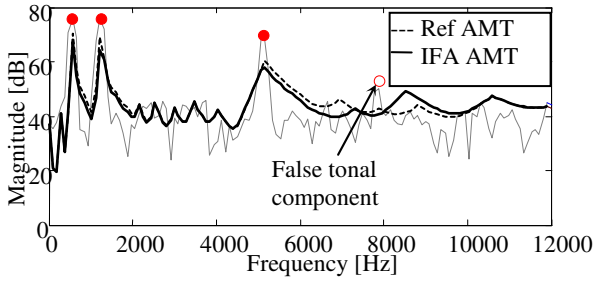

Figure 3. Comparison between reference AMT (dotted line, ○) and IFA AMT (solid dark line, ●) –S1L tests example

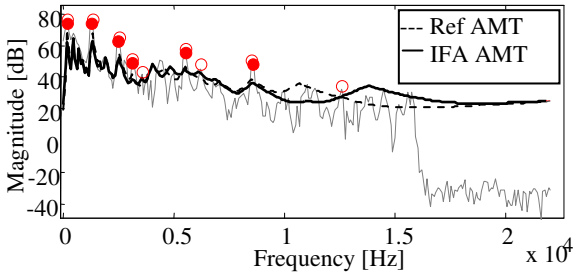

Figure 4. Comparison between reference AMT (dotted line, ○) and IFA AMT (solid dark line, ●) –Scenario 2 example

Scenario 2. In Figure 4 is represented the power spectrum of a high quality audio fragment ($F_s = 44.1\text{kHz}$). The comparison between reference AMT (dotted line) and IFA AMT (solid line) depicts a notable difference, of maximum 12 dB, which underlines the results obtained for the previous scenario. This difference can be crucial, for instance in perceptual compression applications, where the bitrate is obtained based on perceptual considerations.

5. CONCLUSIONS

This paper resumes the IFAs method used for tonal components extraction from any given signal and emphasizes its essential advantages over traditional approaches. We incorporated IFAs into a popular application: the calculus of AMT. Using MPEG1 Layer 1 psychoacoustic model, we proposed to replace the classic algorithm for tonal/noise maskers identification with the IFAs method.

We conducted extensive comparative tests which revealed multiple results: the extraordinary accuracy of IFAs method, even for very low SNRs, the reduced number of false components introduced by the IFA method, the influence of the new approach over the spreading function. Finally, when comparing IF AMT with the reference AMT we reported an average difference of 2dB in magnitude, with $\sigma_{\Delta L} = 0.967$, for signals with high SNRs. When SNR level decreases, the variation of ΔL increases, depending on the frequency content and the number of false tonal components.

6. REFERENCES

[1] T. Abe, and M. Honda, "Sinusoidal Model Based on Instantaneous Frequency Attractors", *IEEE Transactions on Audio, Speech and Language Processing*, pp. 1292-1300, 2006

[2] C. Negrescu, A. Ciobanu, and D. Stanomir, "Refined Spectral Estimation of Tonal Components", *SISOM*, , Bucharest, May 2008.

[3] A.E. Barnes, "The Calculation of Instantaneous Frequency and Instantaneous Bandwidth", *Geophysics Journal*, pp. 1520-1524, 1992.

[4] B. Boashash, *Time-Frequency Analysis and Processing*, Elsevier, Amsterdam, 2003.

[5] N.E. Huang, Z. Wu, S.R. Long, K.C. Arnold, X. Chen and K. Blank, "On Instantaneous Frequency", Advances in Adaptive Data Analysis, *World Scientific Publishing Company*, pp. 177-229, 2009.

[6] M. Lagrange, and S. Marchand, "Estimating the Instantaneous Frequency of Sinusoidal Components Using Phase-Based Methods", J.Audio Eng. Soc., pp. 385-397, May 2007.

[7] R.J. McAulay, T.F. Quatieri, "A Speech Analysis/Synthesis Based on a Sinusoidal Representation", *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp.744-754, 1986

[8] T.Abe, T. Kobayashi, and S. Imai, "Harmonic Tracking and Pitch Extraction Based on Instantaneous Frequency", *ICASSP 95*, pp. 756-759, 1995.

[9] H. W. Schuessler, *Digitale Signalverarbeitung I,* Springer, Berlin, 1988

[10] M. Ghulam, T. Fukuda, K. Katsurada, J. Horikawa, and T. Nitta, "PS-ZCPA Based Feature Extraction with Auditory Masking Modulation Enhancement and Noise Reduction for Robust ASR" IEICE Trans. Inf. and Syst., pp. 1015-,1023 2006.

[11] T. Painter, and A. Spanias, "A Review of Algorithms for Perceptual Coding of Digital Audio Signals" *International Conference of Digital Signal Processing*, pp. 179-205, July 1997.

[12] A. Natarajan, J.H. Hansen, K.H. Arehart, and J. Rossi-Kats, "An Auditory-Masking-Threshold-Based Noise Suppression Algorithm GMMSE-AMT[ERB] for Listeners with Sensorineural Hearing Loss", *EURASIP Journal on Applied Signal Processing*, pp. 2938-2953, 2005.

[13] J.D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria", *IEEE Journal on Selected Areas in Communications*, pp. 314-323, February 1988.

[14] ISO/IEC JTC1/SC29/WG11 MPEG, IS11172-3 "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s, Part 3: Audio" 1992.