ADAPTIVE TRACKING IN THE TIME-FREQUENCY PLANE AND ITS APPLICATION IN CAUSAL REAL-TIME SPEECH ANALYSIS

Minh Ta, Hieu Thai and Victor DeBrunner

Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering

Florida State University

{taminh, thaihieu, victor.debrunner}@eng.fsu.edu

***Abstract**: This paper proposes a causal approach to adaptive estimation of time-frequency localized signals using Adaptive Notch Filters (ANF). By adaptively estimating the envelope of each sinusoidal component, it is possible to specify the tracking quality and restart the ANF unit whenever the tracked sinusoid disappears, as well as preventing the ANF from "tracking" non-existent sinusoids (the frequency mis-lock situation). Employing multiple ANFs allows an efficient approach to tracking time-frequency localized signals such as speech.*

Index Terms – Adaptive filters, Notch filters, Time-varying filters, Time-Frequency Tracking, Signal Resolution.

I. INTRODUCTION

The tracking aspect of adaptive filtering researchers has traditionally been focused on time-varying signals or systems with fixed or slowly-varying characteristics [1][2]. The problem of tracking non-stationary signals remains mostly open and is still a difficult obstacle for the research community.

An important class of non-stationary signals is the so called natural signals that are obtained from audio or visual sources. It is a well-used premise that these natural signals possess the time-frequency localization property [3]. The prospect of sparse representations in the time-frequency plane has led to an expansion in non-causal time-frequency representation processing methods like the Gabor filter [3] approaches as well as the multiresolution-wavelet analysis and compression [4] [5] approaches.

The approach of adaptive (causal) tracking of these signals is largely based on the similar approaches of transform-domain and sub-band filtering [2]. These approaches are limited in the sense that the sub-bands are fixed, making the task of tracking frequency changing signals such as chirps difficult [6]. In any case, a lesson is drawn, which is in order to track the local time-frequency signals simple "slow time-varying" models such as ones in the traditional tracking literature must be discarded and replaced by ones that describe fully the time-frequency localization property.

A better candidate to achieve that modeling goal is the Adaptive Notch Filter (ANF), such as the well-known adaptive filter proposed by Nehorai [7]. ANF has been shown to track *stationary* sinusoidal signals exceptionally well, as well as those with slow frequency variation. Improvements to the theory have been made from various aspects, such as the asymptotic convergence property [7] [8], filter structure [7][9][10], impact on the convergence by properly choosing the forgetting factor and the pole contraction factor [11][12], and recently the internal prediction of frequency variation to improve the steady state tracking performance [6].

An ANF, when being used to track a localized signal in the time-frequency plane, has a significant advantage over the transform domain characterizations such as the STFT (short-time Fourier Transform) or wavelet-like transforms in the sense that the frequency resolution is much higher with the ANF due to the flexible time-bandwidth of the filter. More importantly, this high resolution frequency is allowed to vary within the limited time-duration. Thus, the approach is significantly different from the transform-domain approaches where the frequency is considered to be a single pure tone for each time-frequency box with a frequency estimation variation bounded by the associated Heisenberg uncertainty of the sub-band filter. Another significant advantage of the ANF approach over the transform-domain methods is that each ANF unit is free to occupy any time-frequency region without any rigid partitioning of the phase-plane (including the time-domain) such as would occur with a wavelet-like transform method.

In order for an ANF to track a signal locally constrained in the time-frequency plane, it is important to emphasize that the local frequency content is usually time-limited. Thus it is absolutely important for any ANF structure to achieve 2 major goals: fast and *correct* convergence when a sinusoidal signal is detected, and switching-off from the tracking state when the sinusoidal signal is no longer available. The first goal of convergence, however, might be difficult to achieve with any ANF structure due to the IIR structure and the associated non-quadratic error surface. In fact, the error surface contains local minima which cause the filter to mis-lock on the wrong frequency every now and then. The internal adaptive gradient descent mechanism usually cannot distinguish between the ANF being locked onto the right frequency and the ANF mis-locking onto the wrong frequency. Therefore, non-linear approaches have been sought to improve the performance in convergence, such as the methods described in a recent paper by the authors [13]. In this paper a new thresholding scheme is proposed that can solve both the problem of mis-lock (thereby improving convergence) and the detection of a sinusoidal signal that suddenly disappears. The approach utilizes a quality index based on the magnitude of the extracted sinusoidal signal that is adaptively estimated online. Extended from that single unit a full set of cascaded ANFs can then be employed to analyze signals in the time-frequency plane.

Recent research in [14] [15] [16] have similar cascaded ANF structures to track multiple, but pure sinusoids. Our approach is generalized enough to also track time-varying, frequency band-limited signals. We start the paper by briefly reviewing the concept of the ANF, its application in adaptive spectral line enhancement, and its role as a time-frequency primitive element.

II. THE ANF AS A TIME-FREQUENCY ATOM

Consider a single-sinusoid signal embedded in noise:

$$y(t) = a_t \cos\big(\varphi(t)\big) + e(t) \triangleq c(t) + e(t) \tag{2.1}$$

where $\varphi(t)$ is the phase of the sinusoid and $e(t)$ is ZMWG noise.

An ANF, such as one proposed by Nehorai [7], is a 2^{nd}-order IIR filter whose zeros are on the unit circle and whose poles are inside the unit circle, but on (or near) the same radial line as the zeros. The structure of this filter is described by the IIR filtering equation:

$$\varepsilon(t,\theta_t) = \frac{1-2\alpha_t q^{-1}+q^{-2}}{1-2\alpha_t\rho_t q^{-1}+\rho_t^2 q^{-2}} y(t) \tag{2.2}$$

where parameter $\alpha_t = \cos(\omega_t)$ is the cosine of the instantaneous frequency ω_t that the filter is notching and $0 < \rho_t < 1$ is the pole contraction factor that indicates the location of the poles on the radial lines of the zeros. Since the zeros are on the unit circle and the poles are located close to, but in from, the zeros, the filter notches a single frequency while allowing all other frequencies to pass. Allowing the parameters $(\alpha_t, \rho_t) \triangleq \theta_t$ of the filter to be adaptive by the prediction error method [7] [17], the ANF can even notch time-varying sinusoidal signals such as chirps.

In spectral line enhancement, the ANF can be used to recover $c(t)$ from $y(t)$ by subtracting $\varepsilon(t,\theta_t)$:

$$c(t) \approx y(t) - \varepsilon(t,\theta_t) \triangleq \hat{c}(t) \tag{2.3}$$

Equation (2.2) shows that each 2^{nd}-order notch filter is in fact a cascade of two complex-parameter 1^{st}-order notch filters:

$$\frac{1-2\cos(\omega_t)q^{-1}+q^{-2}}{1-2\cos(\omega_t)\rho_t q^{-1}+\rho_t^2 q^{-2}} = \left(\frac{1-e^{j\omega_t}q^{-1}}{1-\rho_t e^{j\omega_t}q^{-1}}\right)\cdot\left(\frac{1-e^{-j\omega_t}q^{-1}}{1-\rho_t e^{-j\omega_t}q^{-1}}\right) \tag{2.4}$$

Notice that the two 1^{st}-order notch filters are of the same form, with only differing notching frequencies at ω_t and $-\omega_t$. The mechanism of recovering sinusoidal signals by an ANF can be further examined by focusing on the partial fraction expansion of each segment of the 1^{st}-order notch filter. Define $\varepsilon_+(t)$ to be the filtered signal of $y(t)$ from the first notch:

$$\varepsilon_+(t) = \frac{1-e^{j\omega_t}q^{-1}}{1-\rho_t e^{j\omega_t}q^{-1}} y(t) \tag{2.5}$$

Then by recursive calculation we get:

$$\varepsilon_+(t) = y(t) - \sum_{\tau=0}^{\infty}\left[\left(\prod_{i=0}^{\tau-1}\rho_{t-i}\right)(1-\rho_{t-\tau})\right]e^{\left(j\sum_{i=0}^{\tau}\omega_{t-i}\right)}y(t-\tau-1) \tag{2.6}$$

Define $\gamma(t,\tau) = \left(\prod_{i=0}^{\tau-1}\rho_{t-i}\right)(1-\rho_{t-\tau})$ where $\gamma(t,0) = 1-\rho_t$ and $\xi(t,\tau) = \frac{1}{\tau+1}\sum_{i=0}^{\tau}\omega_{t-i}$. Then $\varepsilon_+(t)$ can be represented as:

$$\varepsilon_+(t) = y(t) - \sum_{\tau=0}^{\infty} y(t-\tau-1)\gamma(t,\tau)e^{j\xi(t,\tau)(\tau+1)} \tag{2.7}$$

Note that because $\rho_t < 1, \forall t$, $\gamma(t,\tau)$ decreases polynomially when $\tau \to \infty$ and can be considered as a (varying) time window function. Similarly $\xi(t,\tau)$ acts as the time averaged frequency.

Perform a change of variable $u = t-\tau-1$, with $u \in (-\infty, t-1]$ and let $g(t,u) = \gamma(t, t-u-1)$, $\bar{\omega}(t,u) = \xi(t, t-u-1)$. From $y(t)$ and $\varepsilon_+(t)$ a complex component of $c(t)$ can be recovered by:

$$\begin{aligned} c_+(t) &= y(t) - \varepsilon_+(t) \\ &= \textstyle\sum_{u=-\infty}^{t-1} g(t,u)e^{j\bar{\omega}(t,u)(t-u)}y(u) \\ &= e^{j\bar{\omega}(t,u)t}\textstyle\sum_{u=-\infty}^{t-1} y(u)g(t,u)e^{-j\bar{\omega}(t,u)u} \end{aligned} \tag{2.8}$$

As we can see, $c_+(t)$ is just the projection of vector $\vec{y}(t-1) = [\dots, y(t-2), y(t-1)]$ (the past values of $y(t)$ on function $e^{-j\bar{\omega}(t,u)t}$) being windowed by the function $g(t,u)$. Thus, it can be seen that recovering a sinusoid using an ANF is in fact taking a causal inner product of $\vec{y}(t-1)$ with $g(t,u)e^{-j\bar{\omega}(t,u)u}$, a causal function similar to a Gabor localized time-frequency atom. One advantage obtained from this analysis when compared to the fixed transform is that the time-bandwidth of the windowing function can vary (increase or decrease) by simply controlling the pole contraction factor ρ_t, and hence the frequency uncertainty of this atom can be changed from imprecise (low time-bandwidth) to very precise (high time-bandwidth, $\rho_t \approx 1$).

III. Non-absolute Convergence of the ANF and the Quality index of the convergence

Due to the IIR structure, any ANF structure is known to have the non-absolute convergence problem[13]. Fig. 1 illustrates the error surface of the ANF structure (against the cosine frequency α_t and the bandwidth control parameter ρ_t) when tracking a single-sinusoid signal. It can be seen from the figure that local minima exist near the region of $\rho \approx 1$ for almost all frequencies different from the true frequency. This phenomenon causes the gradient to decrease to 0 and the algorithm ceases to adapt to the frequency. Thus it is imperative to detect such kind of frequency mis-lock and re-start the ANF to other parameter regions where the error surface is still convex allows correct frequency adaptation.

Another scheme that requires attention is when the currently tracked frequency component suddenly disappears, which is quite common for signals local in the time-frequency plane. If the ANF is still tracking this non-existing sinusoid, the recovered signal is just noise (with extremely small frequency bandwidth). Thus, it is also necessary for the ANF to get out of the current parameter region and start tracking other frequencies.

The two aforementioned problems require some kind of index to indicate when the ANF is no longer tracking any useful signal. An approach taken in [13] in the case of a single-sinusoid signal is to correlate the predicted signal $\hat{c}(t)$ and the input signal $y(t)$. If the ANF locks onto the right frequency then $\hat{c}(t)$ will strongly correlate to $y(t)$. On the other hand if the ANF does not lock onto the right frequency, the estimated signal $\hat{c}(t)$ will consist of a filtered version of the past white noise and therefore does not correlate with the present measurement signal $y(t)$ (consisting of the sinusoidal signal and the white measurement noise, neither of which correlates to the past white noise). The correlation can be performed online using a zero-order LMS structure. However this approach does not scale well to the multiple-sinusoid scheme.

In this paper, we propose a better index, which is the online estimation of the narrow band signal's magnitude a_t. Inspired by the adaptive phase-lock loop [1], a_t can be estimated by the much simpler unit demonstrated in Fig. 2. Note that when $\rho_t \approx 1$, the ANF locks onto a right frequency, or it mis-locks onto a wrong frequency, or the ANF tracks a sinusoid that is no-longer available. In these cases the reconstructed signal $\hat{c}(t)$ will have the form

$$\hat{c}(t) = \hat{a}(t)\cos(\hat{\omega}_t t + \hat{\varphi}_0) \tag{3.1}$$

which is a narrow band signal. By modulating $\hat{c}(t)$ with itself, the spectrum of the modulation will consist of a component of $\hat{a}_t$ at DC and a component at $2\hat{\omega}_t$. In fact, ideally

$$\hat{c}^2(t) = \hat{a}^2(t)\cos^2(\hat{\omega}_t t + \hat{\varphi}_0) = \tfrac{1}{2}\hat{a}^2(t)[1+\cos(2\hat{\omega}_t + \phi_t)] \tag{3.2}$$

Thus, recovering $\hat{a}_t$ is a task of filtering out the high frequency component of $\hat{c}^2(t)$ (also note that ρ_t is known, hence the bandwidth of $\hat{a}_t$ is also known). A complimentary filter can be employed in this case to obtain

$$\frac{1}{2}\hat{a}^2(t) = \left(1 - \frac{1-q^{-1}}{1-\rho_{DC}q^{-1}}\right)\hat{c}^2(t) \tag{3.3}$$

The estimation of $a(t)$ can be used to indicate the quality of the currently estimated sinusoidal signal $\hat{c}(t)$ by comparing $\hat{a}(t)$ to a threshold in a similar fashion to non-causal thresholding methods in spectral subtraction or wavelet thresholding. If $\hat{a}(t)$ is above a certain threshold then we have a usable sinusoidal signal. Otherwise, the filter is tracking noise and it needs to be re-started.

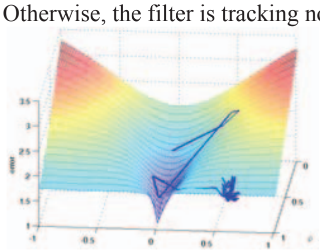


Fig. 1: Nehorai ANF error surface with evolution of adaptive parameters (α, ρ) in the case of frequency miss-lock (horizontal axis extended from -1 to 1 is α)

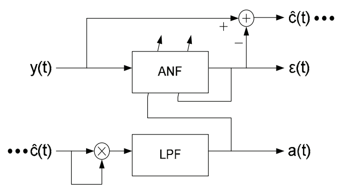


Fig. 2: ANF with conditional reset

In order to avoid the case when the estimation of $\hat{a}(t)$ is very small when the ANF is still in the convergence state, the restarting condition will be issued only when $\hat{a}(t)$ is less than the threshold *and* the pole contraction factor ρ_t is greater than a certain value, indicating that the ANF has "locked-in" a frequency. The restarting condition is summarized as:

$$\begin{cases} \text{if } \hat{a}(t) > T_a : \text{Continue Adapting} \\ \text{if } \hat{a}(t) < T_a \; AND \; \rho_t < T_\rho : \text{Converging State} - \text{Cont. Adapting} \\ \text{if } \hat{a}(t) < T_a \; AND \; T_\rho < \rho_t < 1 : \text{Mislock} - \text{Reset the ANF} \end{cases}$$

Restarting an ANF unit to a parameter region allowing the frequency tracking is a simple task of setting ρ_t to a number sufficiently less than 1 to open the notch bandwidth, and hence moving the ANF to a non degenerative region (convex region) of the error surface. Experiments show that a greatly improved convergence ability when the thresholding of $\hat{a}(t)$ scheme is incorporated.

IV. Multiple Frequency Tracking

A signal local in the time-frequency plane such as speech can have multiple local frequencies. Thus multiple notches are required. In our implementation, we employ the same cascading architecture as discussed in [10], [18] with some modification. These modifications allow us to use a limited number of notch elements and reuse them whenever one finishes its tracking. Assuming that the tracked signal is sparse in the Time-Frequency plane, the approach of using a limited number of notch elements makes sense.

The first modification is the estimation and thresholding of the sinusoidal envelop $\hat{a}_t$ as part of the non-linear adaptation of the ANF coefficients as discussed in Section III. Another modification is the organization of the set of ANFs that allow units that have reached the end of a time-frequency component to restart without "stealing" the signal from the other ANFs that are currently tracking viable sinusoids. Consequently, those ANF units that have lost tracking should be moved to the end of the cascading chain.

The estimation of the quality index $\hat{a}(t)$ also allows us to know when the sinusoidal component extracted from an ANF unit has enough power to be considered a valid signal (not just noise). The value of ρ_t indicates whether the signal is a pure sinusoidal signal or if it has a certain bandwidth. Thus, at any time step, it is possible to identify each sinusoidal component in a mixture of local time-frequency signals.

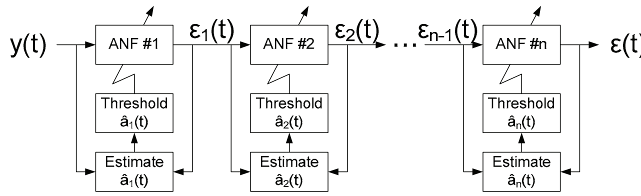


Fig. 3: Cascading ANF chain

<u>Parameters</u>: μ=1, λ_ρ=0.99, T_α=0.01, T_ρ=0.9

<u>Initialization</u>:

$\alpha_1, \alpha_2 \ldots \alpha_n$ is uniformly distributed over the spectrum
for each ANF $i = 1, 2 \ldots n$: initialize adaptive parameters [6]
$\rho_i(0) = 0.7$, $\lambda_{\alpha,i}(0) = 0.8$, $R_{\alpha,i}(0) = R_{\rho,i}(0) = 1$
$\psi_\alpha(0) = \psi_\rho(0) = 0$
$a_i(0) = 0$, $flag_i(0) = 0$

<u>Computation</u>:

for each time $t = 1, 2 \ldots N$
  $y_1(t) = y(t)$
  for each ANF in the cascading chain $i = 1, 2 \ldots n$
  - filter $y_i(t)$ by the ANF i (Section II)
  - $y_{i+1}(t) = \varepsilon_i(t)$
  - calculate $a_i(t)$ (Section III)
  - if $a_i(t) < T_a$ and $\rho_i(t) > T_\rho$: $flag_i(t) = 1$

  end for

  for each ANF in the cascading chain $i = 1, 2 \ldots n$
    if $flag_i(t) = 1$
    - Reinitialize: $\rho_i(t)$, $\lambda_{\alpha,i}(t)$, $R_{\alpha,i}(t)$, $R_{\rho,i}(t)$, $\psi_\alpha(t)$, $\psi_{\rho,i}(t)$ $a_i(t)$, $flag_i(t)$
    - Reinitialize $\alpha_i(t)$ (Section IV)
    - Move ANF i to the end of the cascading chain

    end if
  end for
end for

Table 1: The Time-Frequency tracking algorithm using multiple ANF units

Finally, when ANF units restart, their initial frequencies can be set in one of the two ways:

- Distributing the resetting frequencies between those frequencies that are tracking real signals.
- Uniformly distributing the resetting frequencies in the interval between [0, 1].

The time-frequency tracking algorithm using multiple ANFs is illustrated in Fig. 3 and summarized in Table 1. Each unit is equipped with a reset flag to indicate the current state of the ANF.

V. Simulation

A. *Tracking a single switching sinusoidal signal*

We test our algorithm with a single-sinusoid signal embedded in noise with SNR = 17dB. At one point, we switch off the signal (setting the amplitude to be 0) and let the ANF follow. The result is depicted in Fig. 4 to show the amplitude tracking capability. In our experiment, it initially takes about 20 samples for the amplitude to converge to a high value of 1. After the signal is switched off, it takes about 30 samples for the amplitude to reach the threshold, which is set to be the noise variance in our case. The MSE of the tracked signal compared to the true underlying sinusoidal is 0.0049 and the tracking MSE of the amplitude is 0.0036.

B. *Tracking of a speech signal*

Speech and Acoustic signals are well-known examples of signals local in the time-frequency plane. In our experiments we track the spoken word "hello" [16] with 9 ANF units. The frequency tracking capability of a dominant component is illustrated in Fig. 5. A comparison between the original signal's spectrogram and those of the reconstructed signal and the residual error are shown in Fig. 6.

VI. Conclusion

In this paper, we propose a time-frequency tracking scheme utilizing the ANF. It is shown that an ANF can be considered to be similar to a causal version of the Gabor Time-Frequency atom. An ANF with the special properties of being flexible in allowing time-varying time-uncertainty (hence allowing time-varying frequency-uncertainty) and the ability to change the frequency at each iteration is developed. The crucial point in adapting to a sinusoidal signal that disappears in time is the ability to assign a quality index to the tracked signal. This is accomplished by adaptively estimating the magnitude (envelope) of the sinusoidal signal and thresholding it. The same approach can be scaled to multiple units, each adapting to a single time-frequency element. We apply this scheme in tracking a speech signal. It is possible to extract frequency components of much higher resolution than the traditional spectrogram approach.

VII. References

[1] O. Macchi, *Adaptive Processing - The LMS Approach*. John Wiley & Sons, 1995.

[2] S. Haykins, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 2002.

[3] D. Gabor, "Theory of communication," *Journal of the Institute of Electrical Engineers*, vol. 93, no. 26, pp. 429-457, 1946.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1999.

[5] D. Dohono and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, 1994.

[6] M. Ta and V. DeBrunner, "Adaptive Notch Filter with Time-Frequency Tracking of Chirp Signals," in *ICAASP*, Las Vegas, 2008.

[7] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 4, pp. 983-996, Aug. 1985.

[8] D. V. B. Rao and S. Y. Kung, "Adaptive notch filtering for retrieval of sinusoids in noise,," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 283-295, Mar. 1984.

[9] A. G. Chambers and A. G. Constantinides, "Frequency tracking using constrained adaptive notch filters synthesised from allpass sections," *IEE Proceedings*, vol. 137 pt. F, no. 6, pp. 475-481, Dec. 1990.

[10] V. DeBrunner and S. Torres, "Multiple Fully Adaptive Notch Filter Design Based on Allpass Sections," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, Feb 2000.

[11] M. Dragosevic, "An Adaptive Notch Filter with Improved Tracking Properties," *IEEE Transactions on Signal Processing*, vol. 43, no. 9, 1995.

[12] P. Handel and A. Nehorai, "Tracking Analysis of an Adaptive Notch Filter with Constrained Poles and Zeros," *IEEE Transactions on Signal Processing*, vol. 42, no. 2, Feb. 1994.

[13] M. Ta, H. Thai, and V. DeBrunner, "Stochastic Search Methods to Improve the Convergence of Adaptive Notch Filter," *13th DSP Workshop* , 2009 (submitted).

[14] M. Mojiri and A. R. Bakhshai, "Estimation of n Frequencies Using Adaptive Notch Filter," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 54, no. 4, pp. 338-342, Apr. 2007.

[15] M. Niedźwiecki and A. Sobociński, "On Tracking Properties of Real-Valued Generalized Adaptive Notch Filters," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1688-1695, May 2007.

[16] J. E. Cousseau, S. Werner, and P. D. Dońate, "Factorized All-Pass Based IIR Adaptive Notch Filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5225-5236, Nov. 2007.

[17] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[18] T. Kwan and K. Martin, "Adaptive detection and enhancement of multiple sinusoids using a cascade IIR filters," *IEEE Trans. Circuits Syst.* , vol. 36, no. 7, pp. 937-947, 1989.

[19] P. A. Regalia, "An improved lattice-based adaptive IIR notch filter," *IEEE Transactions on Signal Processing*, vol. SP-39, Sep. 1998.

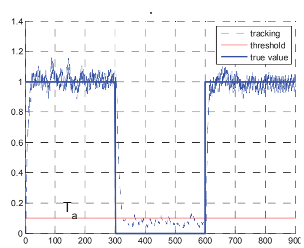


Fig. 4: Amplitude of a pure sinusoidal signal (—blue) and of the tracked signal (-- blue). The red line is the threshold.

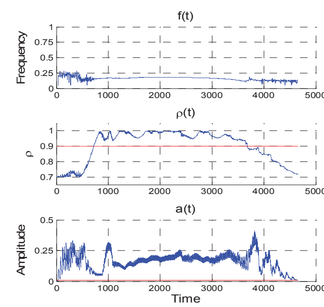
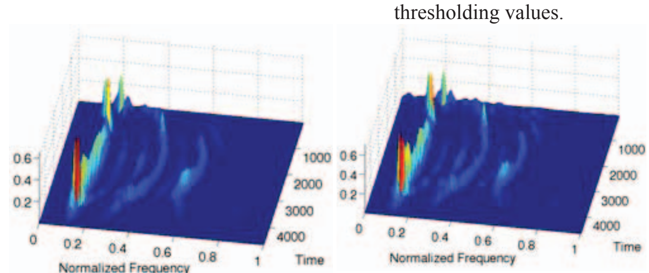
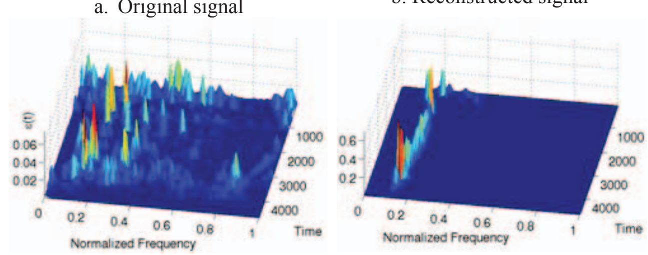


Fig. 5: First major component of 'hello' signal – Evolution of frequency (top), of ρ(t) (middle) and of $\hat{a}(t)$ (bottom). Red lines indicate thresholding values.



a. Original signal

b. Reconstructed signal



c. Residual Error

d. A dominant component extracted

Fig. 6: Spectrogram of the original signal, the reconstructed signal, the residual error, and the first major component extracted.