Adaptive noise level estimation

Chunghsin Yeh, Axel Roebel

ADAPTIVE NOISE LEVEL ESTIMATION

Chunghsin Yeh and Axel Röbel

IRCAM Analysis-Synthesis team
1, place Igor Stravinsky 75004 Paris

ABSTRACT

The topic of this article is the estimation of the colored noise level in audio signals with mixed noise and sinusoidal components. The noise envelope model is based on the assumptions that the envelope varies only slowly with frequency and that the noise amplitudes obey a Rayleigh distribution. The method is an extension of a recently proposed approach of classification of sinusoidal and noise spectral peaks, which takes into account the noise envelope model to improve the detection of sinusoidal peaks. By means of iterative evaluation and adaptation of the noise envelope model, the classification of noise and sinusoidal peaks is iteratively refined until the detected noise peaks are coherently explained by the noise envelope model. Testing examples of nearly white noise and colored noise are demonstrated.

1. INTRODUCTION

Many applications for audio signals such as speech and music require an estimation of the noise level that should be local in time and in frequency. Noise level estimation, or noise spectral magnitude estimation, is usually done by explicit detection of time segments that contain only noise or estimation of harmonically related spectral components (for nearly-harmonic signals). Since some of the noise is related to the signal, relying only on pure noise segments will not allow to properly detect the noise introduced with the source signal. Therefore, it has been proposed to include several consecutive analysis frames assuming that the time segment contains low energy portion and the noise present within the segment is more stationary than the signal [1].

The other classical approach is to remove the sinusoids and estimate the noise afterwards [2]. This involves sinusoidal peak identification, either in single frame [3] [4] or by tracking sinusoidal components across frames [5] [6]. We decide to follow this approach because the assumptions compared to the methods reviewed in [1] are released. We propose to classify the spectral peaks in each short-time spectrum independently because the costly tracking of sinusoidal components could then be avoided. Moreover, the spectral peak classification method proposed in [3] [4] allows to control the classification results such that a bias towards sinusoids or noise can be easily altered. After subtracting the sinusoidal peaks from the observed spectrum, we expect that there are few sinusoidal peaks left in the residual spectrum. Then, a bandwise noise distribution fit is performed using a statistical measure. The outliers of the observed noise peaks are excluded through the process of distribution fit. Finally, an average noise level is estimated from the remaining noise peaks.

This paper is organized as follows. First the problem of estimating noise level is defined. In section **3**, we explain how the narrow band noise can be modeled. An iterative algorithm to approximate the noise level is then presented in section **4**. Finally, nearly white noise and a polyphonic signal with colored noise are tested.

2. PROBLEM DEFINITION

A signal is called "white noise" if the knowledge of the past samples does not tell anything about the subsequent samples to come. The power density spectrum of white noise is constant. By means of filtering a white noise signal, correlations between the samples are introduced. Since in most cases the power density spectrum will no longer be constant, filtered white noise signals are generally called "colored noise". We define the "colored noise level" as the expected magnitude level of the observed noise peaks. It could be represented as a smooth frequency dependent curve approximating the noise spectrum as shown in Figure 1. The noise level should include most of the noise peaks below it and also follows smoothly the variation of the observed spectral magnitudes.

3. MODELING NARROW BAND NOISE USING RAYLEIGH DISTRIBUTION

Under the assumption that noise is nearly white within the considered frequency band, we choose Rayleigh distribution to fit the distribution of the observed noise

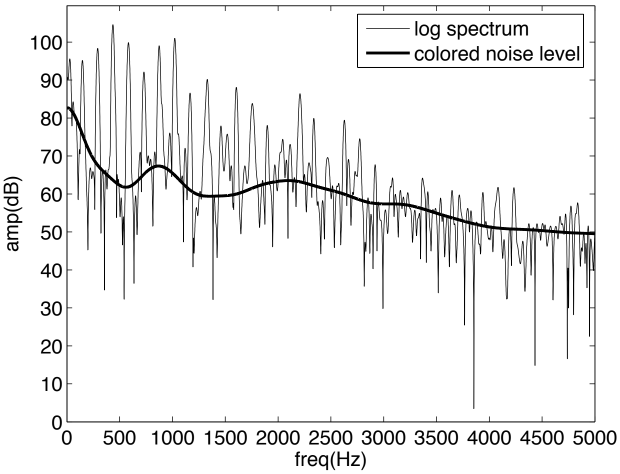

Fig. 1. Colored noise level

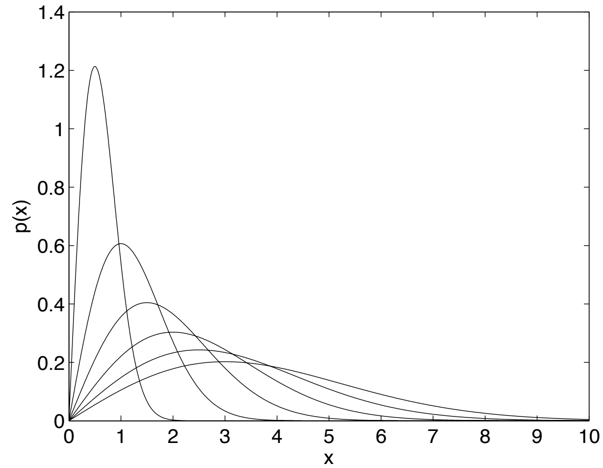

Fig. 2. Rayleigh distribution with different σ

peaks in each subband [1]. The Rayleigh distribution was originally derived by Lord Rayleigh in connection with a problem in the field of acoustics. A Rayleigh random variable X has probability density function [7]:

$$p(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} \tag{1}$$

with $0 \le x < \infty$, $\sigma > 0$, cumulative distribution function

$$F(x) = 1 - e^{-x^2/(2\sigma^2)} \tag{2}$$

and the pth percentile

$$x_p = F^{-1}(p) = \sigma\sqrt{-2\log(1-p)},\ 0 < p < 1 \tag{3}$$

In Figure 2, the probability density function is plotted for different values of σ. The larger the σ, the larger the maximum of the distribution. Notice that σ is not the usual notation for the variance of one distribution but the **mode** of the Rayleigh distribution. The variance of Rayleigh distributed random variable is

$$Var(x) = \frac{4-\pi}{2}\sigma \tag{4}$$

Consider a Rayleigh random variable X as the observed magnitudes of spectral peaks, σ represents the most frequent magnitude values for noise peaks in a narrow band. For the spectral components of magnitudes close to σ, they are of higher probability to be noise. On the other hand, for the spectral peaks of magnitudes larger than σ, those of larger magnitudes are less probable to be noise (and thus they are more probable to be signal).

[1] In fact, Rice has showned in the Bell Laboratories Journal in 1944 and 1945 that Rayleigh distribution is suitable for modeling the probability distribution of a narrow band noise.

4. NOISE LEVEL ESTIMATION

For a given narrow band, e.g. each frequency bin k, the noise distribution can be modeled by means of a Rayleigh distribution with a frequency dependent $\sigma(k)$. Once $\sigma(k)$ across the spectrum be estimated, the curve passing through these σ-valued magnitudes defines a reference noise level L_σ. By adjusting the percentage of noise to be included using eq.(3), the noise level L_n can be estimated by simply multiplying $\sqrt{-2\log(1-p)}$ with L_σ. Therefore, the problem comes to estimating the frequency dependent $\sigma(k)$.

It is known that the mean of a Rayleigh random variable X is

$$E[X] = \sigma\sqrt{\pi/2} \tag{5}$$

from which we have

$$\sigma = \frac{E[x]}{\sqrt{\pi/2}} \tag{6}$$

That is, the frequency dependent $\sigma(k)$ can be obtained if the mean magnitude of noise components, which is also frequency dependent, can be estimated. We propose an iterative approximation of the average noise level L_m (thus $L_\sigma \rightarrow L_n$) using the cepstrally-smoothed curve over the peaks classified as noise.

4.1. Spectral subtraction of sinusoids

In [3], four spectral peak descriptors have been proposed to classify spectral peaks. The descriptors are designed to properly deal with non-stationary sinusoids. This method serves as the first step of our algorithm to classify sinusoidal and non-sinusoidal peaks. The sinusoidal peaks are then subtracted from the observed spectrum to obtain the residual spectrum that contains mostly noise peaks.

To estimate the frequency of each sinusoidal peak, we rely on the reassignment method proposed by F. Auger and P. Flandrin [8]. Given a STFT (Short Time Fourier Transform) X_h using the analysis window h, the frequency slope can be estimated by means of [9]

$$\omega'(t,\omega) = \frac{\partial\hat{\omega}(t,\omega)/\partial t}{\partial\hat{t}(t,\omega)/\partial t} \tag{7}$$

, where $\hat{t}(t,\omega)$ and $\hat{\omega}(t,\omega)$ are the reassignment operators. Once the frequency and the frequency slope of each sinusoidal peak are estimated, the peak is subtracted from the observed spectrum. The optimal phase is estimated by means of the least square error criterion, i.e. the error between the original signal and the processed signal is minimized. However, if the estimated slope is larger than the maximal slope around the observed peak, it will not be considered as a consistent estimate and therefore be disregarded.

The main function of subtracting sinusoidal peaks is to provide sufficient residual peaks for a proper statistical measure of the magnitude distribution even if the frequency resolution is limited and sinusoidal peaks are very dense.

4.2. Iterative approximation of the noise level

After obtaining the residual spectrum, denoted as X_R, the spectral peak classification is re-performed and then the iterative approximation of the noise level is carried out till the selected statistical measure of the noise distribution in all subbands [2] fit that of Rayleigh distribution.

The reasons to use a statistical measure are: (i) the amount of the observed samples is usually not large enough to draw the underlying distribution, (ii) statistical measures are representative of a distribution and are more efficient for distribution fit.

[2] We divide equally 16 subbands for an analysis frequency range up to 5kHz.

We use skewness as the statistical measure for distribution fit. Skewness is a measure of the degree of asymmetry of a distribution [10]. If the right tail (tail at the large end of the distribution) extends more than the left tail does, the function is said to have positive skewness. If the reverse is true, it has negative skewness. If the two tails extend symmetrically, it has zero skewness, e.g. Gaussian distribution. The skewness of a distribution is defined as

$$Skw(X) = \frac{\mu_3}{\mu_2^{3/2}} \tag{8}$$

where μ_i is the ith central moment defined as the expected value of X^i. And the skewness of Rayleigh distribution is independent of $\sigma(k)$:

$$skw_{rayl} = \frac{2(\pi-3)\sqrt{\pi}}{\sqrt{(4-\pi)^3}} \cong 0.6311 \tag{9}$$

We define that a distribution fit is achieved if $Skw(X_n^b) \leq skw_{rayl}$, where X_n^b is the observed magnitudes of noise peaks in the bth subband.

Assuming that for each subband in X_R there are a greater proportion of noise peaks and only a few sinusoidal peaks remain with dominant magnitudes. Then the noise level approximation can be realized by iterating the following processes:

I. For each subband, check if the distribution fit is achieved. If the distribution fit is not achieved in the subband under investigation, that is, $Skw(X_n^b) > skw_{rayl}$, the largest outlier is removed (re-classifying the largest peak in the subband as sinusoids). Under the assumption that $\sigma(k)$ varies slowly with frequency, we expect that the skewness decreases while an outlier is removed. Therefore, convergence of this iterative procedure is expected.

II. Calculate the cepstrum of the noise spectrum constructed from interpolating the magnitudes of noise peaks. The cepstrum is the inverse Fourier transform of the log-magnitude spectrum and the dth cepstral coefficient is formulated as

$$c_d = \frac{1}{2}\int_{-\pi}^{\pi} \log\left|X_n(\omega)\right| e^{i\omega d}\, d\omega \tag{10}$$

By truncating the cepstrum and using the first D cepstral coefficients, we reconstruct a smooth curve representing the average noise level L_m as a sum of the slowly varying components.

$$L_m(\omega) = \exp\left(c_0 + 2\sum_{d=1}^{D-1} c_d \cos(p\omega)\right) \tag{11}$$

III. The noise peaks in each subband are updated w.r.t. the noise level:

$$\begin{aligned} L_n &= L_\sigma \sqrt{-2\log(1-p)} \\ &= \frac{L_m}{\sqrt{\pi/2}} \sqrt{-2\log(1-p)} \end{aligned} \tag{12}$$

The peaks below L_n are defined as the noise components for the next iteration.

While all the subbands meet the requirement of the skewness measure, the set of noise peaks at the end of iteration defines the final noise level. Notice that if the noise level varies very fast in such a way that the slope of the noise envelope is very large, the process might not converge.

5. TESTING EXAMPLES

To demonstrate the effectiveness of the proposed algorithm, we have tested two types of signals: nearly-white noise and signal with background noise. In both cases, we set the noise percentile of the Rayleigh distribution to 80%, that is, p=0.8 in eq.(12).

In Figure 3, a nearly-white noise spectrum is shown with the estimated noise level. The estimated noise level does approximate a flat curve that is expected for a nearly-white noise. To further demonstrate how the proposed algorithm works for polyphonic signals, we demonstrate a polyphonic signal with colored noise. Figure 4 shows the spectral peak classification result and Figure 5 shows the residual spectrum after subtracting sinusoidal components. The dotted line in Figure 5 represents the boundaries of the equally divided frequency bands. The estimated noise level is shown in Figure 6 [3]. The estimated noise level estimated by the proposed method does follow well the variation of the observed spectrum. Moreover, it provides us the control over misclassified spectral peaks at the first stage.

[3] Additional peaks are shown to indicate possibly hidden sinusoidal peaks.

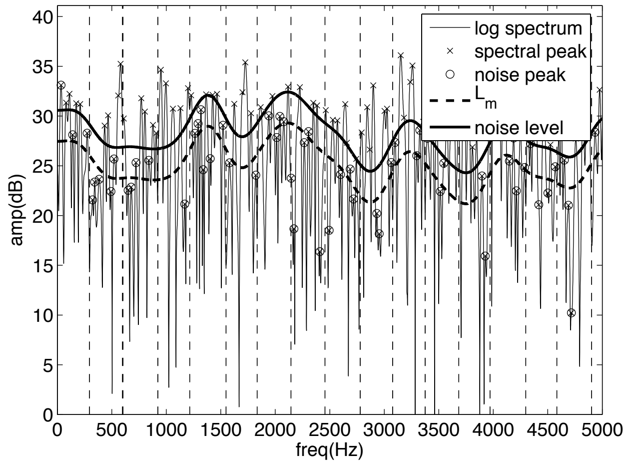

Fig. 3. Estimated noise level for nearly-white noise

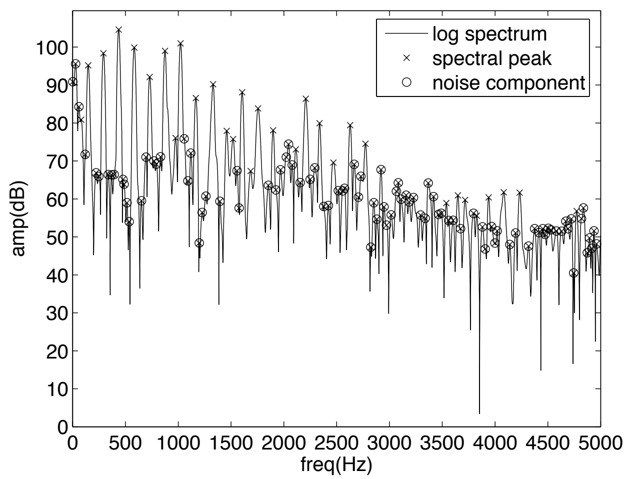

Fig. 4. Spectral peak classification

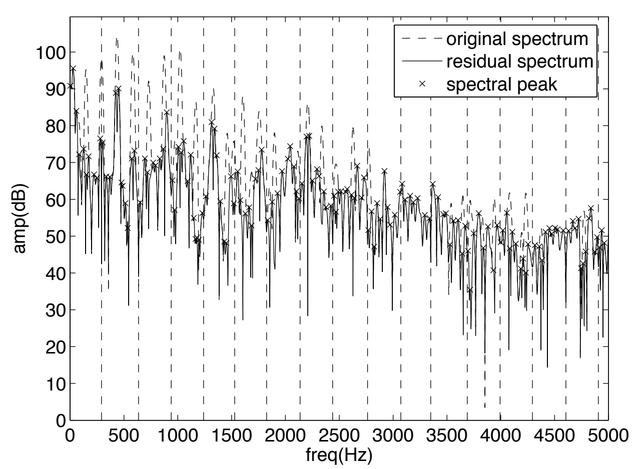

Fig. 5. Residual spectrum

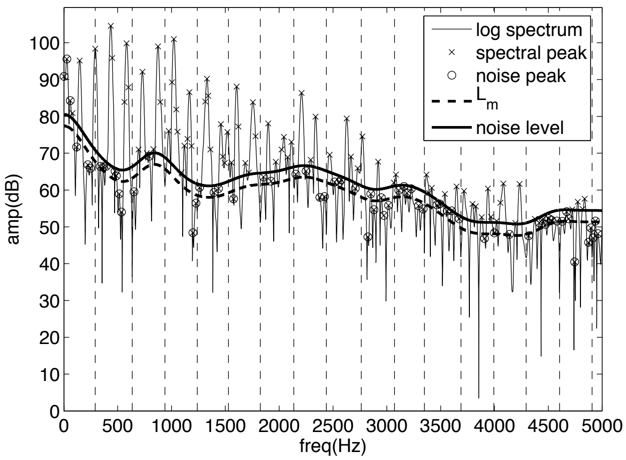

Fig. 6. Estimated noise level for a polyphonic signal

6. CONCLUSIONS

We have presented an iterative algorithm for approximating the local noise level. This algorithm is adaptive to the observed spectrum. It neither includes additional information from the neighboring frames or pure noise segments, nor makes use of harmonic analysis.

Its ability to handle polyphonic signals has been demonstrated. However, there are several parameters to be studied: the number of subbands, the order (the number of cepstral coefficients) of the noise level curve, and the percentage of the noise in eq.(12) to be included.

The proposed algorithm is useful for many signal analysis/synthesis applications. It has been implemented for multiple fundamental frequency estimation [11].

7. REFERENCES

[1] C. Ris and S. Dupont, "Assessing Local Noise Level Estimation Methods: Application to Noise Robust ASR," *Speech Communication*, 2000.

[2] M. Alonso, R. Badeau, B. David, and G. Richard, "Musical tempo estimation using noise subspace projection," in *IEEE Workshop on applications of signal processing to audio and acoustics (WASPAA '03)*, 2003, pp. 95–98.

[3] A. Röbel and M. Zivanovic, "Signal decomposition by means of classification of spectral peaks," in *Proc. of the International Computer Music Conference (ICMC'04)*, Miami, Florida, 2004.

[4] G. Peeters and X. Rodet, "Sinusoidal Characterization in terms of Sinusoidal and Non-Sinusoidal Components," in *Proc. of 1st international conference on Digital Audio Effects (DAFx'98)*, Barcelona, Spain, 1998.

[5] B. David, G. Richard, and R. Badeau, "An EDS modelling tool for tracking and modifying musical signals," in S*tockholm Music Acoustics Conference 2003*, Stockholm, Sweden, 2003, pp. 715–718.

[6] M. Lagrange, S. Marchand, and J. Rault, "Tracking Partials for the Sinusoidal Modeling of Polyphonic Sounds," in *Proceedings of the IEEE International Conference on Speech and Signal Processing (ICASSP'05)*, Philadelphia, USA, 2005.

[7] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley & Sons, Inc, New York, 2nd. edition, 1994.

[8] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. on Signal Processing*, vol. 43, no. 5, 1995.

[9] A. Röbel, "Estimating partial frequency and frequency slope using reassignment operators," in *Proc. of the International Computer Music Conference (ICMC'02)*, Göteborg, 2002, pp. 122–125.

[10] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory*, Oxford University Press, New York, 6th. edition, 1998.

[11] C. Yeh, A. Röbel, and X. Rodet, "Multiple fundamental frequency estimation of polyphonic music signals," in *Proc. IEEE, International Conference on Acoustics, Speech and Signal Processing (ICASSP'05)*, Philadelphia, 2005.