



HAL
open science

Adaptive noise level estimation

Chunghsin Yeh, Axel Roebel

► **To cite this version:**

Chunghsin Yeh, Axel Roebel. Adaptive noise level estimation. 9th Int. Conf. on Digital Audio Effects (DAFx'06), Sep 2006, NA, France. pp.145-148. hal-01161313

HAL Id: hal-01161313

<https://hal.science/hal-01161313>

Submitted on 8 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE NOISE LEVEL ESTIMATION

Chunghsin Yeh

Analysis/Synthesis team
IRCAM/CNRS-STMS, Paris, France
cyeh@ircam.fr

Axel Röbel

Analysis/Synthesis team
IRCAM/CNRS-STMS, Paris, France
roebel@ircam.fr

ABSTRACT

We describe a novel algorithm for 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 slowly with frequency and that the magnitudes of the noise peaks obey a Rayleigh distribution. Our method is an extension of a recently proposed approach of spectral peak classification of sinusoids and noise, which takes into account a 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 estimating 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 such that non-stationary and colored noise can be dealt with. Noise level estimation, or noise power spectral density estimation, is usually done by explicit detection of time segments that contain only noise, or explicit 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] [2].

The other classical approach is to remove the sinusoids and estimate the underlying noise components afterwards [3]. This involves sinusoidal component identification, either in single frame [4] [5] or by tracking sinusoidal components across frames [6] [7]. We decide to follow this approach because the assumptions compared to the methods reviewed in [1] are relaxed. 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 [4] [5] 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 an iterative process of distribution fit and noise level estimation. Upon the termination of the iterative approximation, the estimated noise level is thus defined.

This paper is organized as follows. First the problem of noise level estimation is defined. In section 3, we explain how the distribution of the magnitudes of narrow band noise can be modeled. An iterative algorithm to approximate the noise level is then presented in section 4. Lastly, different types of noise are used to demonstrate the effectiveness of the proposed method.

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. A noise peak is defined as a peak that can not be explained as a stationary or weakly modulated sinusoid of the signal. The noise level 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 and also follows smoothly the variation of the observed spectral magnitudes.

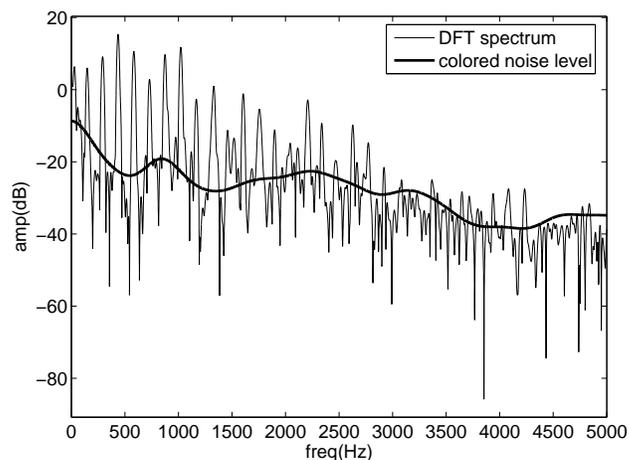


Figure 1: Colored noise level

3. MODELING NARROW BAND NOISE USING RAYLEIGH DISTRIBUTION

Under the assumption that noise is nearly white within a considered frequency band, we choose Rayleigh distribution to fit the distribution of the observed narrow band noise¹. 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 [8]:

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

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

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

and the p th percentile

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

In Figure 2, the probability density function is plotted for different values of σ ($\sigma = 0.5, 1, 1.5, 2, 2.5$ and 3). σ corresponds to the **mode** of the Rayleigh distribution, which is the most frequently observed value in X . Thus, $p(\sigma)$ corresponds to the maximum of the probability distribution. Notice that σ is not the usual notation for the variance of a distribution. The variance of Rayleigh distributed random variable is

$$\text{Var}(X) = \frac{4 - \pi}{2} \sigma^2 \quad (4)$$

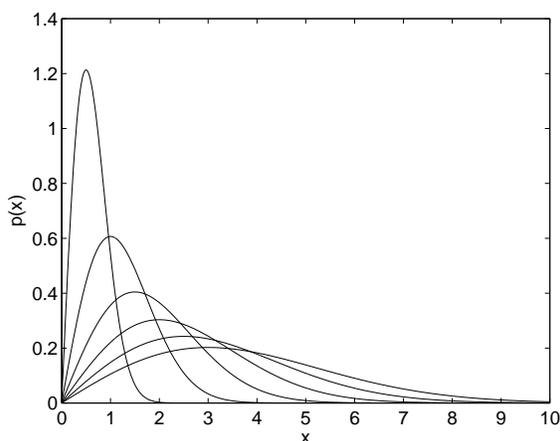


Figure 2: Rayleigh distribution with different σ

Consider the Rayleigh random variable X as the observed magnitudes of spectral peaks in a narrow band, then σ represents the most frequent magnitude values of noise peaks. The mode of the Rayleigh distribution can then be used to derive the probability of an observed peak to belong to the background noise process. Comparing the magnitude of the spectral peak to σ we may conclude that peaks having amplitude below σ are most likely noise

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

while for the spectral peaks having magnitudes larger than σ , the larger magnitudes they have, the less probable they are to be noise (and thus they are more likely related to the deterministic part of the signal).

4. NOISE LEVEL ESTIMATION

For a given narrow band, e.g. each frequency bin k , the noise distribution can be modeled by means of Rayleigh with mode $\sigma(k)$. Once $\sigma(k)$ has been estimated for all k , the curve passing through these σ -value magnitudes defines a reference noise level \mathcal{L}_σ . Using eq.(3) it is now possible to adjust the noise threshold to a desired percentage of misclassified noise peaks. The related *noise envelope* \mathcal{L}_n can be estimated by simply multiplying the *estimated Rayleigh mode* \mathcal{L}_σ with $\sqrt{-2 \log(1-p)}$. Therefore, the problem comes to estimating the frequency dependent $\sigma(k)$.

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

$$E[X] = \sigma \sqrt{\pi/2} \quad (5)$$

from which we have

$$\sigma = \frac{E[X]}{\sqrt{\pi/2}} \quad (6)$$

That is, the frequency dependent $\sigma(k)$ can be calculated if the mean noise magnitude $E[X]$, which is also frequency dependent, can be estimated.

However, estimation of the expected noise magnitude corresponding to each frequency bin requires sufficient observations for statistical evaluation. Most of the existing approaches [1] rely on observations from neighboring frames. Our approach relies on the assumption that the noise spectral envelope is changing only weakly with the bin index k such that we may use the observed spectral peaks in the predefined subbands² to estimate the (frequency dependent) *mean noise level* \mathcal{L}_m by means of a cepstrally-smoothed curve over the noise peaks. We describe the noise level estimation procedures in the following.

4.1. Spectral subtraction of sinusoids

In [4], 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 to classify sinusoidal and non-sinusoidal peaks in our algorithm. The sinusoidal peaks are then subtracted from the observed spectrum to obtain the residual spectrum that is assumed to contain mostly noise peaks.

To estimate the spectral parameters of each sinusoidal peak, the reassignment method proposed by F. Auger and P. Flandrin [9] is used to estimate the frequency slope [10]. Given a STFT (Short Time Fourier Transform), the frequency slope can be estimated by means of

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

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

²We divide equally the subbands with the bandwidth 312.5Hz.

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 S_R , the spectral peaks are again classified and then the iterative approximation of the noise level is carried out till the selected statistical measure of the noise distribution in all subbands 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.

We use skewness as the statistical measure for distribution fit. Skewness is a measure of the degree of asymmetry of a distribution [11]. 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

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

where μ_i is the i th central moment. And the skewness of Rayleigh distribution is independent of $\sigma(k)$:

$$Skw_{rayl} = \frac{2(\pi - 3)\sqrt{\pi}}{\sqrt{(4 - \pi)^3}} \approx 0.6311 \quad (9)$$

If the distribution of the noise magnitudes in a subband is assumed Rayleigh then we may test for misclassified sinusoids by means of the condition $Skw(X_n^b) > Skw_{rayl}$, where X_n^b are the noise magnitudes in the b th subband. Whenever this condition is true we assume that there are misclassified sinusoids that can be detected by observing their amplitude levels relative to the current estimate of $\sigma(k)$.

Note that the distribution of noise magnitudes in each subband will not be Rayleigh if $\sigma(k)$ in the subband is not constant. To improve the consistency of the skewness test we therefore rescale all noise magnitudes by means of normalizing with the current estimated Rayleigh mode \mathcal{L}_σ .

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

- I. 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 d th cepstral coefficient is formulated as

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

By truncating the cepstrum and using the first D cepstral coefficients, we reconstruct a smooth curve representing the

mean noise level \mathcal{L}_m as a sum of the slowly-varying components.

$$\mathcal{L}_m(\omega) = \exp(c_0 + 2 \sum_{d=1}^{D-1} c_d \cos(\omega d)) \quad (11)$$

The cepstral order D is determined in a way similar to that of [12]: $D = F_s / \max(\Delta f_{max}, BW) \cdot C$, where F_s is half the sampling frequency, Δf_{max} is the maximum frequency gap among all the noise peaks, BW is the subband bandwidth, and C is a parameter to set.

- II. Then we have the estimated Rayleigh mode $\mathcal{L}_\sigma = \mathcal{L}_m / (\sqrt{\pi/2})$ across the analysis frequency range.
- III. For each subband, check if the distribution fit is achieved. If the distribution fit is not achieved in the subband under investigation, that is, $Skw(X_n^b / \mathcal{L}_\sigma^b) > Skw_{rayl}$ where \mathcal{L}_σ^b denotes the estimated Rayleigh mode in the b th subband, then the largest outlier is excluded (re-classifying the largest outlier in the subband as sinusoid).

When all the subbands meet the requirement of the skewness measure, the estimated Rayleigh mode \mathcal{L}_σ can be used to derive a probabilistic classification of all spectral peaks into noise and sinusoidal peaks. For this we suggest the p th percentile of Rayleigh distribution

$$\mathcal{L}_n = \mathcal{L}_\sigma \sqrt{-2 \log(1 - p)} \quad (12)$$

with a user selected value for p . Notice that if the underlying noise level varies very fast in such a way that the proposed model cannot capture the noise level evolution then the procedure may not converge or may not converge to a reasonable estimate.

5. TESTING EXAMPLES

To demonstrate the effectiveness of the proposed algorithm, we have tested two types of signals: white noise and a polyphonic signal with background noise. In both cases, the sampling frequency is 16kHz and we set $C = 1$ for the cepstral order and $p = 0.8$ in eq.(12), that is, we allow 20% of the noise to be misclassified according to Rayleigh distribution.

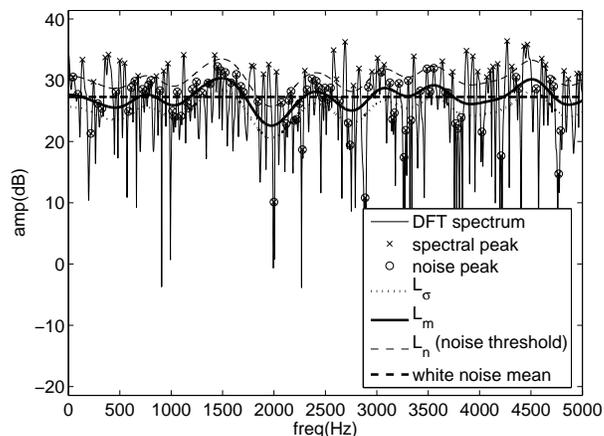


Figure 3: Estimated noise level for white noise (*test 1*)

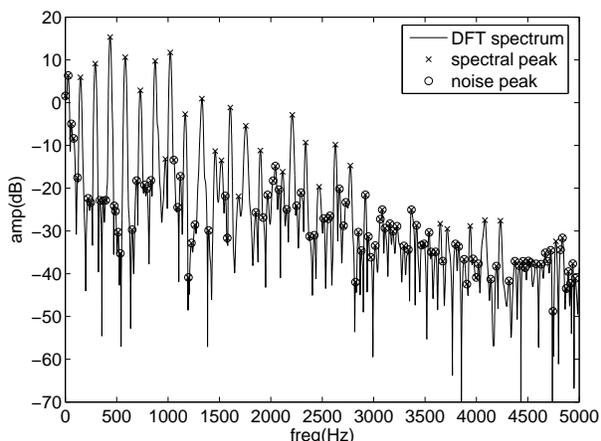


Figure 4: Initial spectral peak classification (*test 2*)

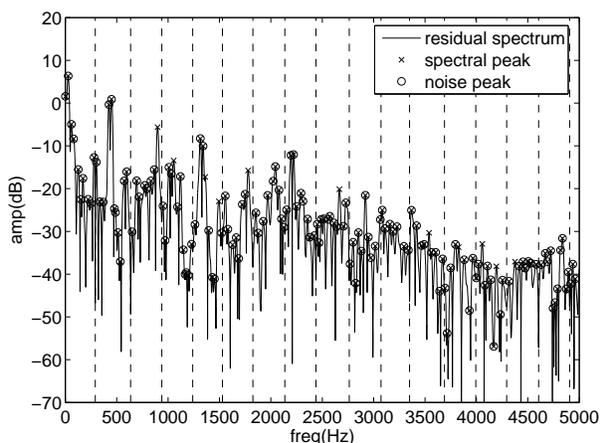


Figure 5: Residual Spectrum (*test 2*)

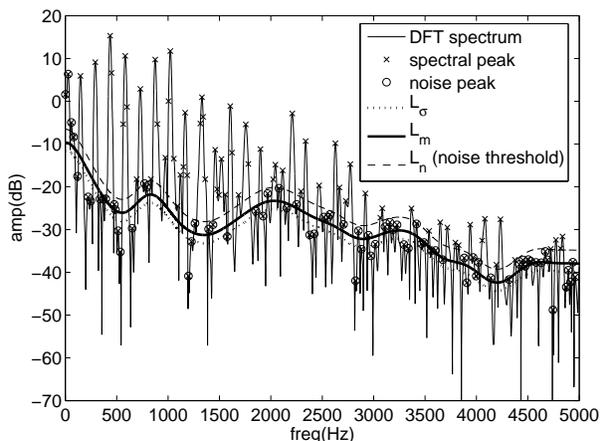


Figure 6: Estimated noise level for a polyphonic signal (*test 2*)

In Figure 3, a white noise spectrum is shown with the estimated noise level. The estimated mean noise level \mathcal{L}_m does approximate the constant white noise mean. The estimated noise envelope \mathcal{L}_n is noted as “noise threshold” to notify that this is a user-adjustable level. The noise peaks are finally re-classified as the spectral peaks having magnitudes below this threshold.

To further demonstrate how the proposed algorithm works for polyphonic signals, we estimate the colored noise level of a polyphonic signal. Figure 4 shows the initial spectral peak classification result and Figure 5 shows the residual spectrum after subtracting the sinusoidal peaks. The dotted vertical lines represent the boundaries of the equally divided subbands. The estimated noise level is shown in Figure 6³. The proposed noise envelope model does follow well the variation of the observed spectrum. Moreover, it provides us the control over misclassified spectral peaks at the first stage.

6. CONCLUSIONS

We have presented an iterative algorithm for approximating the noise level local in time and in frequency. This algorithm is adaptive to the dynamics of the spectral variation. It neither includes additional information from the neighboring frames or pure noise segments, nor makes use of harmonic analysis. The proposed noise envelope model represents the instantaneous noise spectrum, which can be used as a new feature for signal analysis.

Its ability to handle different types of 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 according to Rayleigh distribution.

The proposed algorithm is useful for many signal analysis and synthesis applications, such as partial tracking, signal enhancement, etc. It has been implemented by the authors for estimating the number of quasi-harmonic sources in connection with the problem of multiple fundamental frequency estimation.

7. REFERENCES

- [1] C. Ris and S. Dupont, “Assessing local noise level estimation methods: application to noise robust ASR,” *Speech Communication*, , no. 2, pp. 141–158, 2001.
- [2] V. Stahl, A. Fischer, and R. Bippus, “Quantile based noise estimation for spectral subtraction and Wiener filtering,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’00)*, Istanbul, Turkey, 2000, pp. 1875–1978.
- [3] 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)*, New Paltz, New York, 2003, pp. 95–98.
- [4] 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, pp. 446–449.

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

- [5] 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.
- [6] B. David, G. Richard, and R. Badeau, "An EDS modelling tool for tracking and modifying musical signals," in *Stockholm Music Acoustics Conference 2003*, Stockholm, Sweden, 2003, pp. 715–718.
- [7] M. Lagrange, S. Marchand, and J. Rault, "Tracking partials for the sinusoidal modeling of polyphonic sounds," in *Proc. of the IEEE International Conference on Speech and Signal Processing (ICASSP'05)*, Philadelphia, Pennsylvania, 2005, pp. 229–232.
- [8] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley & Sons, Inc, New York, 2nd. edition, 1994.
- [9] 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, pp. 1068–1089, 1995.
- [10] A. Röbel, "Estimating partial frequency and frequency slope using reassignment operators," in *Proc. of the International Computer Music Conference (ICMC'02)*, Göteborg, Sweden, 2002, pp. 122–125.
- [11] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory*, Oxford University Press, New York, 6th. edition, 1998.
- [12] A. Röbel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *Proc. of the 8th International Conference on Digital Audio Effects (DAFx'05)*, Madrid, Spain, 2005, pp. 30–35.