

Signal decomposition by means of classification of spectral peaks

Axel Roebel⁺, Miroslav Zivanovic*, and Xavier Rodet⁺

⁺Analysis Synthesis Team, IRCAM, France

{roebel,rod}@ircam.fr

*Universidad Publica de Navarra, Spain

miro@unavarra.es

Abstract

In extending previous work on detecting transient spectral peaks we here investigate into the distinction between sinusoidal and noise components by means of classification of spectral peaks. The classification is based on descriptors derived from properties related to time-frequency distributions. In contrast to existing methods, the descriptors are designed to properly deal with non-stationary sinusoids, which considerably increases the range of applications. The experimental investigation shows superior classification results compared to the standard correlation-based approach.

1 Introduction

The decomposition of audio spectra in sinusoids, transients and noise is often used to improve the results of parameter estimation and/or signal manipulation algorithms. In the following we will investigate into the possibility to decompose signal spectra by means of classifying individual spectral peaks. As has been shown for the case of transients (Röbel 2003) signal decomposition can be achieved by means of integrating results obtained from the classification of spectral peaks. Complementing the transient detection algorithm mentioned above the present paper will deal with classification into noise and sinusoidal peaks, only.

There exist few approaches for the classification of spectral peaks. Among them we cite the widely used correlation based measure of sinusoidality (Rodet 1997) and a proposal based on the reassigned spectrogram (Hainsworth, Macleod, and Wolfe 1998). The former takes the maximum of the complex correlation between the DFT of the analysis window and each peak of the STFT of the signal and classifies the peak as steady state sinusoid if this value is above a threshold close to 1. The latter proposes the classification of the STFT peaks in sinusoids, unresolved sinusoids, transients and noise. Various statistics are calculated for each side of the peak separately and the traditional pattern classification method with a likelihood ratio test is applied to perform the classification.

The shortcoming of both approaches is the underlying assumption of quasi-stationary signals. Moreover, the method presented in (Hainsworth, Macleod, and Wolfe 1998) globally tries to minimize classification errors. Because the class

probabilities depend on the size of the analysis window (for a larger window more noise peaks will be observed in contrast to a constant number of sinusoidal peaks) it appears more appropriate to adjust classification thresholds such that class dependent error rates are achieved. Consequently, we derive our classification criteria by means of declaring a worst case signal and limit the error rate for each of the different classes.

There exist a number of audio signal processing applications that may benefit from a correct classification of spectral peaks. Musically interesting is the possibility to separate sinusoidal and noise components by means of grouping the classified spectral peaks. As a further example we mention the possibility to reduce the number of candidate peaks considered for partial tracking which could reduce the computational costs for probabilistic partial tracking algorithms (Depalle, Garcia, and Rodet 1993).

The paper is organized as follows. In section 2 we define the descriptors that will be used for classification of spectral peaks and discuss their properties if applied to different types of spectral peaks. In section 3 we describe the structure of the decision tree and derive the thresholds to be used for classification. An experimental demonstration of the superior performance of the new classifier compared to the correlation based peak classification is presented. We conclude the paper with a discussion of the achievements in section 4.

2 Spectral peak descriptors

Frequency coherence: The frequency reassignment operator has been derived in (Auger and Flandrin 1995) to improve signal localization in the time-frequency plane. For constant amplitude chirp signals it exactly points onto the frequency trajectory of the chirp at the position of the center of gravity of the windowed signal. The frequency offset Δ_ω between the frequency at the center of an DFT bin and the reassigned frequency in rad is given by

$$\Delta_\omega(k) = \text{imag} \frac{X_d(k)X^*(k)}{|X(k)|^2}. \quad (1)$$

Here k specifies the bin index of the DFT. $X(k)$ is the DFT of the signal windowed with the analysis window and $X_d(k)$ is the DFT of the signal windowed with the time derivative of

the same window. The operator X^* denotes complex conjugation. To characterize the frequency coherence of a spectral peak we select as descriptor the minimum value of $|\Delta_\omega|$ for all k belonging to this peak and normalize by $\frac{2\pi}{N}$ where N is the size of the DFT.

Energy location: The group delay $g_d(k)$ is defined to be the derivative of the phase spectrum with respect to frequency. For a single bin of the DFT spectrum it equals the mean time according to (Cohen 1995) and specifies the contribution of this frequency to the center of gravity of the signal related to the spectral peak. The mean time is the main feature to detect transient peaks (Röbel 2003). In the current investigation we found that due to the influence of neighboring peaks the mean time derived from the whole spectral peak is not sufficiently robust. The modified version used here

$$t_e = -\frac{g_d(k_1)|\Delta_\omega(k_2)| + g_d(k_2)|\Delta_\omega(k_1)|}{|\Delta_\omega(k_2)| + |\Delta_\omega(k_1)|}, \quad (2)$$

derives the energy location by means of investigating the peak center, only. The indices k_1 and k_2 correspond to the largest and second largest samples in the peak. The weighting by means of the frequency reassignment operator results in the fact that constant amplitude chirp signals will always have a mean time very close to zero even if their frequency trajectory does not exactly pass through a center frequency of a bin. We normalize $|t_e|$ by the length of the analysis window to obtain the energy location descriptor ELD .

Duration: The time duration of a signal as defined in (Cohen 1995) is the standard deviation of the time with respect to the mean time interpreting signal energy as distribution. For discrete spectra it can be obtained by means of

$$T = \sqrt{\frac{\sum_k (A'(k)^2 + (g_d(k) - \bar{t})^2) |X(k)|^2}{|X(k)|^2}}, \quad (3)$$

where the sum is performed over the spectral peak under consideration. \bar{t} is the mean time of the signal related to the peak and $A'(k)$ is the frequency derivative of the continuous magnitude spectrum. For normalization we divide the time duration T by means of the window size to obtain the duration descriptor DD . Note that it can be shown that $g_d(k)$ and $A'(k)$ are the real and imaginary part of the complex number

$$Y(k) = -\frac{X_t(k)X^*(k)}{|X(k)|^2}, \quad (4)$$

where $X_t(k)$ is the DFT spectrum obtained with a window that results from multiplying analysis window and a time ramp function (Auger and Flandrin 1995)

Normalized bandwidth: The mean frequency $\bar{\omega}$ and the bandwidth B for a single peak give a rough idea of the concentration of the spectral density along the frequency grid. Considering L to be the number of samples in the spectral peak then the normalized bandwidth descriptor NBD can be

defined as:

$$\bar{\omega} = \frac{\sum_k k |X(k)|^2}{\sum_k |X(k)|^2}, \quad (5)$$

$$NBD = \frac{B}{L} = \frac{\sum_k (k - \bar{\omega})^2 |X(k)|^2}{L \sum_k |X(k)|^2}. \quad (6)$$

The summation includes all the bins of the spectral peak.

2.1 Descriptor properties

The classification thresholds for the descriptors are determined by means of a worst case signal, a single AMFM-sinusoid in noise (SNR = 0dB) where both frequency and amplitude are modulated by means of sinusoids. To resemble natural vibrato, the period of the frequency modulation is twice the period of the amplitude modulation. The characteristics of the test signal are: AM index 0.5 and FM with 200Hz of frequency deviation. The analysis window is a 50ms Hanning window and the frequency modulation period is 100ms. For calculating the DFT we use a 4096-point FFT with the samplerate being 44100Hz. The signal roughly resembles the 10-th harmonic of a 333Hz tone with half tone vibrato extent.

In the initial investigation only the noise and sinusoidal peak classes have been taken into account and all but the sinusoid peaks have been considered to be noise. During the experiments we found that the noise distributions of the descriptors would change with the SNR. Further investigation revealed that this effect was due to the presence of sinusoidal sidelobes in the noise class. Therefore an additional class for sinusoidal sidelobes has been introduced.

The descriptor distributions for the peak classes that have been obtained for the test signal are shown in fig. 1. For the sinusoidal distributions the descriptors were applied only to the largest peak in the spectrum for a total of 1100 time frames. The noise distributions were obtained by analyzing all the peaks in the DFT of a white noise signal. For the sidelobe distributions we analyzed all the sidelobe peaks of a stationary noise-free sinusoid. Due to the very large variance in the distributions obtained already for stationary sidelobes we decided to assign every peak not in the sinusoid or noise class to the sidelobe class and a further analysis of nonstationary sidelobe peaks has not been considered. For ease of comparison all distributions are normalized such that their maximum value is equal to one which is well suited to determine thresholds that are related to fixed fractions of the distributions.

FCD: As long as the signal energy is located within the peak itself the FCD descriptor should be limited to be below 0.5. As shown in fig. 1 this is true for the sinusoidal and noise peak classes. The sidelobe peaks are related to signal energy located in the mainlobe peak which, as expected, results in a FCD distribution that extends nearly uniformly up to half the size of the DFT (distribution not completely displayed).

ELD: This descriptor will be close to zero for constant amplitude chirp signals. For amplitude modulation the ELD may increase, however, due to the normalization, its magnitude is always below 0.5. The signals corresponding to iso-

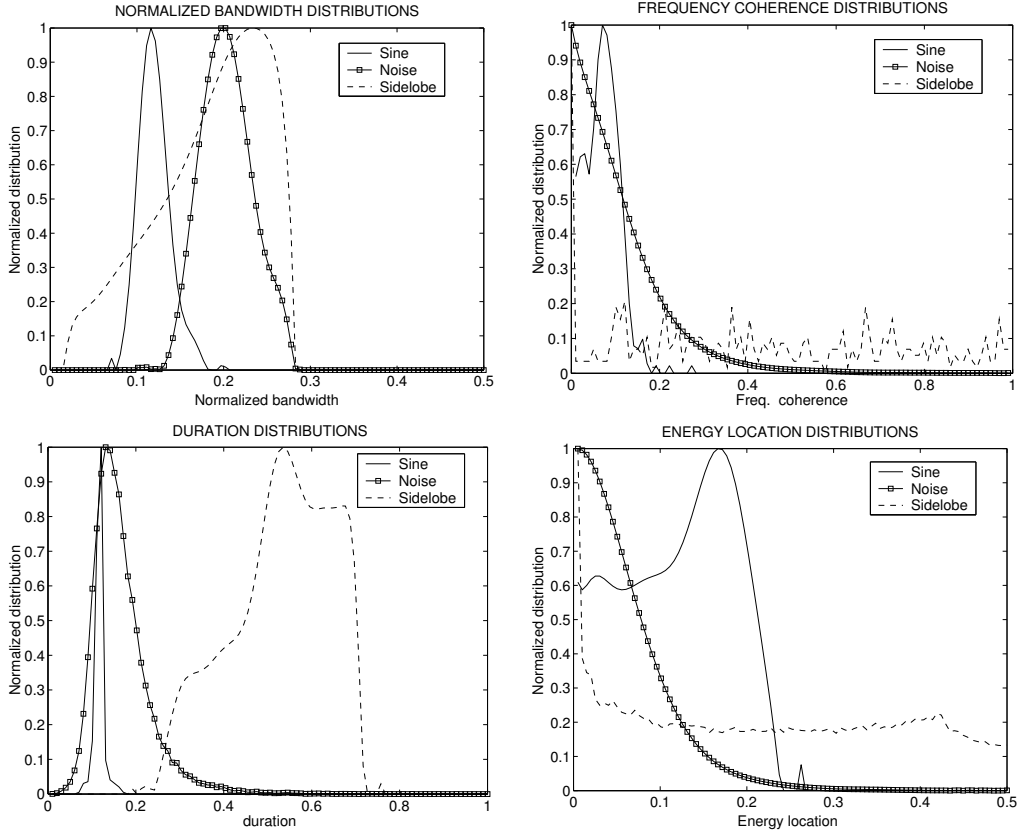


Figure 1: Distributions for the peak descriptors and the peak classes sinusoid/noise/sidlobe.

lated sidelobes are not limited to the duration of the analysis window but are confined to the region of the zero-padded analysis window. Therefore, the mean time covers a range limited by the zero padded signal (not completely displayed).

DD: For constant amplitude chirp signals the duration will always be close to the duration of the analysis window itself. For amplitude modulation the *DD* distribution of the sinusoidal peaks will spread and move its center thus covering a considerable part of the *DD* distribution of the noise peaks. As explained for the *ELD* sidelobe related signals extend outside the analysis window and, therefore, generally have a larger value of *DD* than noise and sinusoids.

NBD: This descriptor can be partly understood as a measure of the noise energy in the neighborhood of a sinusoidal spectral peak. The experimental investigation of the *NBD* distributions for modulated noise free sinusoidal peaks and for noise peaks shows that these distributions do not overlap at all making them a very good candidate for sinusoidal and noise classification. With increasing noise level the tail of the sinusoidal *NBD* distribution is moving right and, overlaps slightly with the distribution for the noise peak class.

3 Experimental results

To evaluate the performance of the proposed descriptors a preliminary binary decision tree for the peak classification

has been established as follows: in the first level a sinusoidal and non-sinusoidal classification is performed. Then in the second level the non-sinusoidal peaks are classified into sidelobes and noise. The thresholds for both levels of classification have been obtained by means of analyzing the distributions shown in fig. 1. The thresholds used for the evaluation are shown in table (1).

Because sidelobe and noise class distributions do hardly depend on the signal and analysis window the related thresholds need no adaptation. To adapt the *NBD* threshold used for sinusoid/noise classification we propose as user parameter the error rate of for noise-class peaks that can be easily transformed into a threshold using some frames of white noise signal and adjusting the *NBD* threshold accordingly. For the evaluation we requested 10% error rate for the noise class distribution which results in about 1% misclassification of sinusoidal peaks in our worst case signal.

The selected thresholds have been used to classify a number of artificial and real audio signals. Here we will present only one result of the algorithm applied to a flute signal with vibrato taken from the Iowa University Database. We use this example to compare the proposed classification method to the correlation method mentioned in the introduction. In order to make the comparison meaningful, we have adjusted the thresholds for the correlation method such that for the worst case scenario signal it achieves the same percentage of sinusoidal peaks correctly classified.

sinusoid/non-sinusoid:	$NBD \leq 0.17$ & $DD \leq 0.18$
sidelobe/noise:	$DD \geq 0.28$ $FCD \geq 0.35$ $ELD \geq 0.25$

Table 1: thresholds for sinusoid/nonsinusoid and for side-lobe/noise classification in the 2 levels of the decision tree.

In the top part of fig. 2 the spectrogram of the original signal is shown. Below it the classified spectrograms for both methods are drawn. The advantage of our approach (bottom) is evident. To reliably detect peaks related to non stationary sinusoids the threshold for the correlation based descriptor has to be lowered that much that nearly all noise peaks are considered sinusoids. Refined investigation showed that the results of the proposed method are always superior or equal to the correlation-based approach.

4 Conclusions

In this paper we have presented new descriptors for the classification of spectral peaks and have described preliminary results comparing the new classification method with a correlation-based approach. We have shown that the proposed descriptors achieve significantly better classification than the correlation-based descriptor if the signal contains non-stationary sinusoids. The thresholds are automatically adapted as a function of the desired noise classification error.

Acknowledgements

The second author of the paper likes to gratefully acknowledge the financial support of the Gobierno de Navarra, Spain.

References

- Auger, F. and P. Flandrin (1995). Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. on Signal Processing* 43(5), 1068–1089.
- Cohen, L. (1995). *Time-frequency analysis*. Signal Processing Series. Prentice Hall.
- Depalle, P., Garcia, and X. Rodet (1993). Tracking of partials for additive sound synthesis using hidden Markov models. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Volume I, pp. 242–245.
- Hainsworth, S., M. Macleod, and P. Wolfe (1998). Analysis of reassigned spectrograms for musical transcription. In *Proc. DAFX98 (Digital Audio Effects Workshop)*.
- Röbel, A. (2003). A new approach to transient processing in the phase vocoder. In *Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx03)*, pp. 344–349.
- Rodet, X. (1997). Musical sound signal analysis/synthesis: Sinusoidal+residual and elementary waveform models. In *Proc IEEE Time-Frequency and Time-Scale Workshop 97, (TFTS'97)*, pp. ??

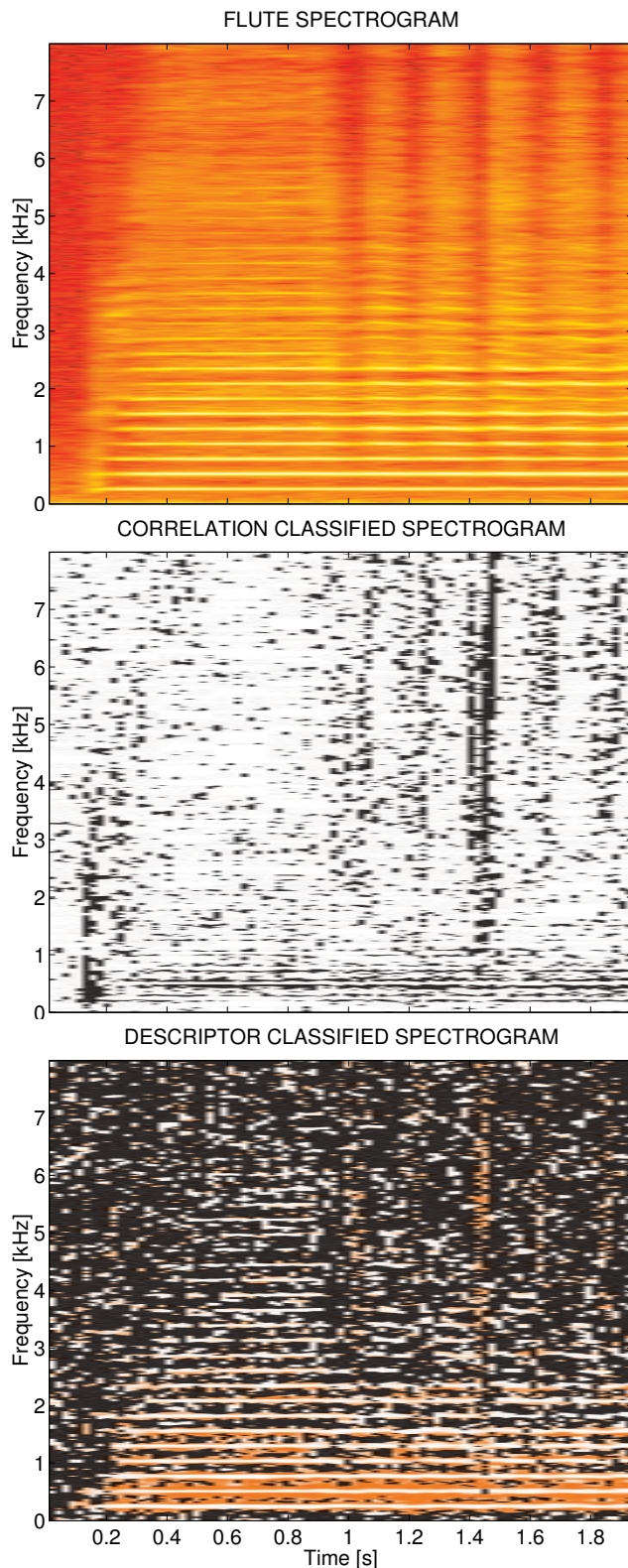


Figure 2: Flute vibrato signal: spectrogram (top), peaks classified by correlation (center), peaks classified by new descriptors (bottom). In the classified spectra the bins of all peaks are colored indicating the classification results as follows: white=sinusoid, black=noise, gray/orange=sidelobe.