ADAPTIVE THRESHOLD DETERMINATION FOR SPECTRAL PEAK CLASSIFICATION

Miroslav Zivanovic
Universidad Pública de Navarra
Pamplona, Spain
miro@unavarra.es

Axel Roebel and Xavier Rodet
IRCAM
Paris, France
roebel@ircam.fr,
rod@ircam.fr

ABSTRACT

A new approach to adaptive threshold selection for classification of peaks of audio spectra is presented. We here extend the previous work on classification of sinusoidal and noise peaks based on a set of spectral peak descriptors in a twofold way: on one hand we propose a compact sinusoidal model where all the modulation parameters are defined with respect to the analysis window. This fact is of great importance as we recall that the STFT spectra are closely related to the analysis window properties. On the other hand, we design a threshold selection algorithm that allows us to control the decision thresholds in an intuitive manner. The decision thresholds calculated from the relationships established between the noise power in the signal and the distributions of sinusoidal peaks assures that all peaks described as sinusoidal will be correctly classified. We also show that the threshold selection algorithm can be used for different types of analysis windows with only a slight parameter readjustment.

1. INTRODUCTION

The decomposition of audio spectra in sinusoids, transients and noise is a useful tool for improving the results of parameter estimation and/or signal manipulation applications. As has been shown for the case of transient detection [1] and sinusoidal and noise discrimination [2], the classification of spectral peaks is a beneficial approach to identify signal components. Such a classification scheme that makes optimal use of the information provided by spectral peaks can then be used to achieve a robust segmentation into higher level signal components, e.g. partials or unvoiced regions.

The basis for spectral peak classification is an adequate choice of criteria that would best describe sinusoidal and noise spectral peaks of audio signals. Ideally, those criteria (from now on descriptors) would be able to precisely detect the nature of each peak in the spectrum and thus provide for a complete separation between the corresponding peak classes in the descriptor domains. Consequently, the decision boundary for the classification process would be unambiguous and no misclassfication of spectral peaks would occur. This scenario, however, is purely hypothetical as the peaks corresponding to sinusoids (partials) in the spectra of real-world signals are usually subject to additive noise and some type of modulation. In these cases the descriptor distributions of the different peak classes overlap and the optimal determination of the decision boundaries will depend on the specific application.

The peak classification method proposed in [2] makes use of descriptors that were designed to adequately characterize non-stationary sinusoidal signals. These descriptors have proven to lead to superior classification performance than other approaches devoted to sinudoidal detection/estimation [3,4]. It was shown in [2] that the peak classes can be characterized by distributions in the descriptor domains, similar to probability density functions. Once the distributions have been generated, a simple decision tree can be derived that allows the classification of spectral peaks into sinusoids, noise and sidelobes.

The peak classification method has been used successfully in a number of applications. As examples we mention polyphonic F0 detection [5], adaptive noise floor determination [6] and voiced unvoiced frequency boundary determination. Another interesting application would be the pre-selection of the sinusoidal peaks to reduce the number of candidate peaks considered for partial tracking in additive analysis. A reliable classification of noise peaks could reduce the number of incorrect connections and for probabilistic approaches like [7] it would considerably reduce the computational cost. The major problem with the classification scheme in [2] is the control of the classification boundaries (classification thresholds) that generally need adaptation for the specific problem at hand. A further problem is that the descriptor boundaries of the different classes will depend on the analysis window that is used. Up to know there did not exist a high level control parameter that would allow to adjust the sensitivity of the algorithm in an intuitive manner. There are two signal parameters that directly affect the classification boundaries. The first is the maximum modulation depth and period of the sinusoids. The second is the minimum amplitude of the sinusoids above the noise floor. Both parameters influence the boundaries of the sinusoidal class and accordingly both can be used to control the decision boundaries. The problem using the modulation limits as control parameter is the fact that the modulation is not a single parameter but a parameter vector of at least 4 dimensions (period and depth for amplitude and frequency modulation). Therefore, it can not be used to provide and intuitive control of the classification boundaries. On the other hand the sinusoidal peak amplitude above the noise floor is a single parameter that for a given modulation limit would allow us to control the complex decision thresholds rather intuitively.

Accordingly, in this paper we investigate into the relation between the peak amplitude above the noise floor and the descriptor boundaries for the class of sinusoidal peaks. The descriptors are defined and their properties discussed thoroughly in [2] but for sake of clarity we will give a brief resume of the most prominent characteristics in the section 2. For the sinusoidal model described in section 3 we define the space of sinusoidal components by selecting particular limits of the amplitude and frequency modulation rate and depths, as well as the modulation laws. In section 4 we present the descriptor distributions for the different signal classes and in section 5 we establish the mathematical model for the descriptor limits of the sinusoidal class as a function of the peak amplitude level above the

noise floor. In the experimental part in section 6 we show that the threshold model successfully adapts to the limits of the distributions of sinusoidal peaks for different types of analysis windows.

2. SPECTRAL PEAK DESCRIPTORS - SUMMARY

Being an elementary classification object, we define a spectral peak as the normalized energy spectral density between two contiguous minima in the DFT modulus $|X(k)|$ of the signal $x(n)$ multiplied by the analysis window. The spectral peak descriptors proposed in [2] are the Normalized Bandwidth Descriptor, the Normalized Duration Descriptor and the Frequency Coherence Descriptor. The first two are well suited to distinguish between sinusoidal and noise peaks while the third can be used to detect the sidelobe structure that is an artifact of the windowing process.

2.1. Normalized Bandwidth Descriptor (NBD)

Energy distribution along the frequency grid provides useful information for identifying the nature of the signal related to a given spectral peak. Being $X(k)$ the DFT of the windowed signal and considering L to be the number of samples in the spectral peak, we have defined the NBD as a function of mean frequency $\bar{k}$ and root mean square bandwidth BW_{rms}:

$$\mathrm{NBD} = \frac{BW_{rms}}{L} = \frac{1}{L}\sqrt{\frac{\sum_k \left(k-\bar{k}\right)^2 \left|X(k)\right|^2}{\sum_k \left|X(k)\right|^2}}, \quad (1)$$

$$\bar{k} = \frac{\sum_k k\left|X(k)\right|^2}{\sum_k \left|X(k)\right|^2}. \quad (2)$$

The sums are performed over the L bins in the peak under consideration.

2.2. Normalized Duration Descriptor (NDD)

As with mean frequency and bandwidth, the mean time and root mean square duration give a rough idea of the distribution of the signal related to a spectral peak along the time grid. The time duration for continuous signals has been defined in [8] as the standard deviation of the time with respect to the mean time. For discrete signals, the following expressions characterize the duration T_{rms} and mean time $\bar{n}$ respectively:

$$T_{rms} = \sqrt{\sum_n \left(n-\bar{n}\right)^2 \left|x(n)\right|^2}, \quad (3)$$

$$\bar{n} = \sum_n n\left|x(n)\right|^2, \quad (4)$$

where $|x(n)|^2$ is the normalized signal's energy. It was shown in [8] that, from the duality of the Fourier transform, both mean time and duration can be expressed in terms of the spectrum. This important feature permits us to describe individual spectral peaks through the parameters generally employed in the time domain. Considering M to be the size of the analysis window, for discrete spectra the NDD can be obtained by means of:

$$\mathrm{NDD} = \frac{T_{rms}}{M} = \frac{1}{M}\sqrt{\frac{\sum_k \left(A'(k)^2 + \left(g_d(k)+\bar{n}\right)^2\right)\left|X(k)\right|^2}{\sum_k \left|X(k)\right|^2}}, \quad (5)$$

$$\bar{n} = -\frac{\sum_k g_d(k)\left|X(k)\right|^2}{\sum_k \left|X(k)\right|^2}, \quad (6)$$

where $g_d(k)$ is the group delay and $A'(k)$ is the frequency derivative of the continuous magnitude spectrum. 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 [8] and specifies the contribution of this frequency to the center of gravity of the signal related to the spectral peak. This property of the group delay has been used in [9] to derive the time reassignment operator, which together with the frequency reassignment aims to improve signal localization in the time-frequency plane. According to [9] the group delay can be calculated efficiently by:

$$g_d(k) = -\mathrm{real}\frac{X_t(k)X^*(k)}{\left|X(k)\right|^2}, \quad (7)$$

being $X_t(k)$ the DFT of the signal using a time weighted analysis window. It can be shown that $A'(k)$ is the imaginary counterpart of the group delay in (7):

$$g_d(k) = -\mathrm{imag}\frac{X_t(k)X^*(k)}{\left|X(k)\right|^2}. \quad (8)$$

As for the NBD all the summations are done over all the bins in the spectral peak.

2.3. Frequency Coherence Descriptor (FCD)

The frequency reassignment operator for constant amplitude chirp signals points exactly onto the frequency trajectory of the chirp at the position of the centre of gravity of the windowed signal. The frequency offset Δ_ω between the frequency at the center of a DFT bin and the reassigned frequency in radians is given by:

$$\Delta_w(k) = \mathrm{imag}\frac{X_{dt}(k)X^*(k)}{\left|X(k)\right|^2}, \quad (9)$$

where $X_{dt}(k)$ is the DFT of the signal windowed by the time derivative of the analysis window. The Frequency Coherence Descriptor is defined as a minimum absolute frequency offset $\Delta_\omega(k)$ for all the bins belonging to that peak:

$$\mathrm{FCD} = \frac{N}{2\mathrm{p}}\min_k\left|\Delta_w(k)\right|, \quad (10)$$

being N the number of bins in the DFT. The normalization factor in (10) ensures that the descriptor is expressed in bins of DFT.

3. SINUSOIDAL MODEL AND PEAK DISTRIBUTIONS

To be able to classify a sinusoidal component we need to define what we consider to belong to the sinusoidal class. As is common for sinusoidal modeling we are going to understand a sinusoidal component as a sinusoid with slowly varying amplitude and frequency parameters [10]. For an investigation into the properties of the spectral peak classes this requirement is not sufficient. To completely define the space of sinusoidal components we have to select concrete limits of the amplitude and frequency modulation rate and depths, and we have to specify a concrete form of the modulation laws.

For the present application there exists an obvious constraint for the modulation which is related to the fact that the spectrum of the sinusoidal component has to contain a dominant mainlobe. Otherwise the investigation of an individual spectral peak can not provide us with sufficient information about the underlying sinusoid. Accordingly, the modulation rate and depth have to be limited such that a dominant mainlobe is present in the Fourier spectrum of each sinusoidal component. Because frequency and time resolution are related to the window size and form, the modulation limits will depend on these two variables. A simple solution to ensure the modulation constraint described above for all window sizes is to determine the maximum modulation that respects the constraint for a given window size and to change the worst case modulation rate proportionally with the window size.

As the next step we need to define the worst case signal that is the signal that will be used to derive the descriptor limits of the sinusoidal class. From the wide range of possible modulation laws we have chosen the sinusoidal amplitude and frequency modulation in white Gaussian background noise as our worst case reference signal. The choice is motivated by the fact that a wide range of FM and AM conditions can be covered. If the window size is small compared to the vibrato rate for example, it is easy to see that the vibrato signal approximately creates linear FM and AM. Recent investigations have shown [11,12] that for real world vibrato signals the AM and FM will generally not be phase synchronous. Accordingly, the worst case signal model exhibits arbitrary phase relations between the amplitude and frequency modulation. A special feature of real world AM is the fact that the dominant AM rate may either be the same as the FM rate, or twice as high. As the letter case is more critical, we chose it for our worst case signal scenario.

In a view of the aforementioned discussion, the following mathematical expression for the sinusoidal model is proposed:

$$x(n) = \cos\left[2\pi F_0 n + A_{FM}\sin\left(2\pi F_{FM} n + \alpha\right)\right]\times \left[1 + A_{AM}\cos\left(2\pi F_{AM} n + \beta\right)\right] + r(n), \quad (11)$$

where $r(n)$ is additive Gaussian noise. The parameters are selected as follows. According to the previous discussion we set $F_{AM} = 2F_{FM}$. The frequency vibrato rate F_{FM} has to be selected such that the spectrum always contains a significant mainlobe, which is ensured by $F_{FM} = 1/(4.2M)$. Accordingly, the window covers less than the fourth part of the FM vibrato period. The values for the amplitude and frequency modulation depth have been chosen as $A_{AM} = 0.5$ and $A_{FM} = 10$.

These values ensure a dominant peak mainlobe for arbitrary phase angles (α and β) . The window length$M$, the sinusoidal frequency F_o, and the sample-rate R do not have any impact on the results. The size of the DFT N is chosen in such a way to assure that the Picket-Fence effect has minimal impact on a peak representation in the discrete spectrum. For completeness we note the values that we used for the following investigation into the descriptor distributions (M = 40ms, N = 4096, F_o = 880Hz, R = 44 kS/s).

It is clear that the present worst case signal does not cover all modulations that may be encountered in a real world setting, even if we respect the fact that a dominant mainlobe is required to detect a modulated sinusoid. The explicit inclusion of time varying sinusoids into the model will nevertheless lead to a classifier that has significant advantages in real world situations with time varying sinusoids.

Because the part of the sinusoidal peak that can be observed changes with the variance σ_r^2 of the background noise level $r(n)$ the peak descriptors will not only change with the modulation, but also with the SNR. For multicomponent signals the global SNR does not provide meaningful insight, and therefore, we will use the *Peak Signal-to-Noise Ratio* (SNR_p) as our noise level parameter. The SNR_p indicates the sinusoidal peak power level in dB over the noise floor (see Figure 1) and it presents a convenient parameter to control the limits of the sinusoidal class.

To experimentally create the descriptor distributions we proceed as follows. For the noise class distributions we calculate the descriptors for all spectral peaks in the DFT of white Gaussian noise processes using an analysis window of size M. For the sinusoidal class we create a grid of phase values covering all combinations α and β over the range $-\pi$ to π and we set $\sigma_r^2 = 0$. Then we calculate the descriptor values for the largest peak in each frame. This gives us the distributions for an infinite SNR_p. The sidelobe distributions are calculated from all but the strongest spectral peak in the spectrum of the worst case sinusoid. The resulting descriptor distributions are normalized by the maximum value and shown in Figure 2 for the Hanning window.

As we can see from Figure 2 the NBD distributions for modulated noise free sinusoidal peaks and for noise peaks do not overlap at all, making them a very good candidate for sinusoidal and noise separation. The sine and noise distributions for the NDD significantly overlap, but the sinusoidal distribution covers only a small range of descriptor values. This fact will be used to refine the sine/noise separation done by the NBD for signals of finite SNR_p as will be explained in the next section. Finally, the sidelobe structures can efficiently be distinguished by means of the FCD. Note that in Figure 2 the maximum of the sidelobe distribution is to be interpreted as a cumulus of all the sidelobe FCD values distributed out of the current axis range.

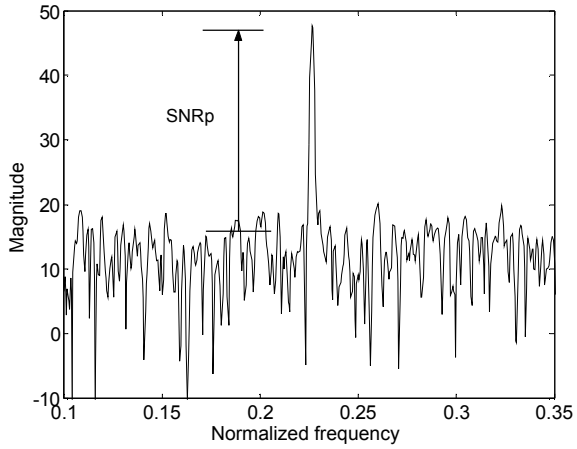

Figure 1: Illustration for the parameter SNR_p (peak signal-to-noise ratio)

4. CLASSIFICATION STRATEGY

The peak classification algorithm, based on the proposed peak descriptors, is established through a two-level decision tree as follows: in the first level the sidelobe and non-sidelobe classification is performed. Then in the second level the peaks previously declared non-sidelobes are classified as sinusoids and noise. The thresholds for both levels of classification are obtained by means of analyzing the distributions shown in Figure 2. For infinite SNR_p the classification could be obtained by simply using FCD and NBD thresholds to perfectly seperate all three peak classes. Note that only in this particular case the NBD attains almost perfect sine/non-sine classification, therefore the contribution of the NDD is negligible.

For a finite SNR_p the sinusoidal distributions experiment a spread proportional to the noise level in the worst case signal. In particular, the NBD sinusoidal distribution extends towards right while the NDD sinusoidal distribution spreads in both directions. The sinusoidal NBD distribution overlaps partially with the noise NBD distribution, which means that the NBD can no longer separate perfectly the peak classes. In order to reduce this ambiguity, we make use of the NDD. As mentioned before, the sinusoidal NDD distribution covers only a small range of descriptor values. Hence, by considering only the peaks within the limits of the sinusoidal NDD distribution as sinusoids, we can eliminate some of the noise peaks previously classified as sinusoids and thus refine the initial sine/noise classification. The classification scheme that is used for finite SNR_p is shown in Table 1. It is important to understand that a decreasing SNRp will modify the limits of the sinusoidal distribution in a similar manner as an increase in the modulation parameters would do. Therefore, the minimum SNRp can be used to control the decision thresholds in a rather intuitive manner.

In order to keep track of the limit values of the sinusoidal distributions we would need to regenerate all the sinusoidal distributions every time the minimum SNR_p that is selected by the user is changed. As shown below, however, the experimental evaluation of the distribution limits can be avoided, due to a simple approximate formula that expresses the relationship between the parameter SNR_p and the margins of the sinusoidal peak distributions in the descriptor domain. These can be used to adapt the classifier to the selected SNR_p. The thresholds to be adapted are the right margin of the NBD sinusoidal distribution and both margins of the NDD sinusoidal distribution. As for the FCD, the threshold may be kept fixed thanks to the good sidelobe separation from the rest of the peak classes.

sidelobe / non-sidelobe	$FCD \geq N/M$
sine / noise	$NBD \leq 0.13$ & $0.13 \leq NDD \leq 0.16$

Table 1: Peak classification thresholds for infinite SNR_p, the window is Hanning.

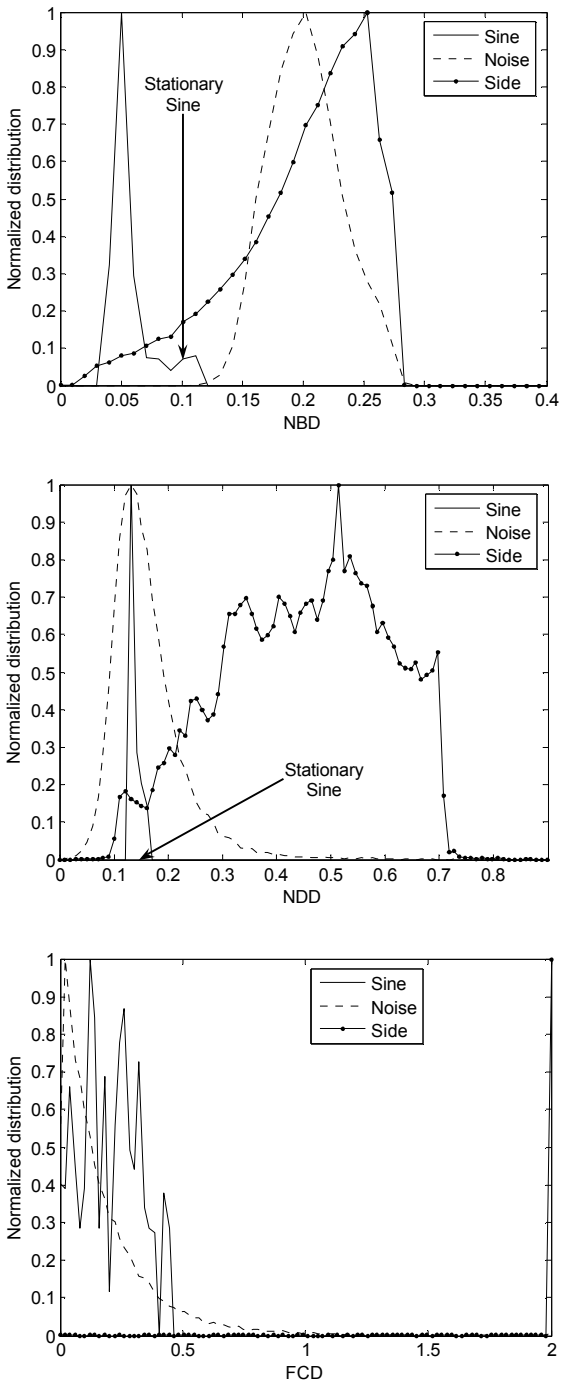

Figure 2: Normalized distributions for three peak classes in the descriptor domain; $\sigma_r^2 = 0$ and the window is Hanning.

5. MODELLING SNR_P DEPENDENCY

The relation between the classification threshold and the SNR_p is rather complex and to be able to achieve a model of these relations the problem requires a number of simplifications. The idea we pro-

pose is to first experimentally determine the signal pattern that is related to the descriptor limits for infinite SNR_p. Then we develop a simplified model of the effect of the additive noise to be able to achieve a mathematical formulation of the threshold dependency on the SNR_p. The relation does not take into account that the signal pattern at the descriptor limits may depend on the SNRp.

Window	α_{max}	β_{max}
Hanning	0.75π	0.50π
Blackman	0.75π	0.55π
Hamming	0.70π	0.45π

Table 2: The phase values of the sinusoidal model corresponding to NBD_{max} *for various analysis windows*

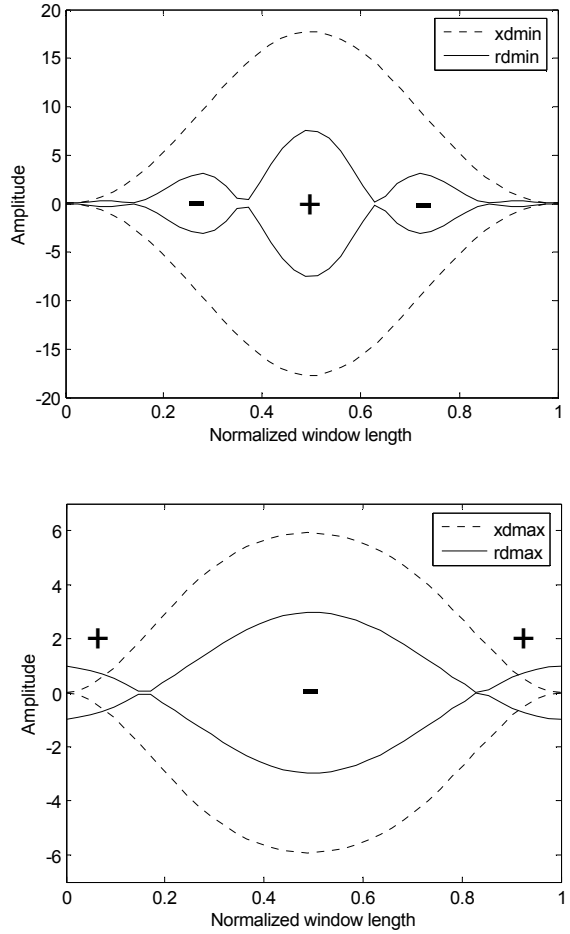

Figure 3: Envelopes of the signal patterns and noise patterns corresponding to the NDD thresholds for SNR_p = 10dB; the sign symbols mark the carrier phase relationship between the waveform; the analysis window is Hanning.

5.1. NBD threshold (NBD_{max})

Let us recall that the NBD is the ratio of the peak bandwidth and peak width. As described above we first need to determine the sinusoidal signal that will give rise to the maximum value of the descriptor $NBD_{max} = BW/L$. This can be done by means of a straightforward search over the two-dimensional grid of phase values α and β for a given analysis window (see Table 2 for some prominent analysis windows).

The presence of noise will affect both *BW* and *L*. It is clear that *L* will decrease because the peak local minima get closer to the peak maximum in terms of magnitude. In a simple approximation we may assume that *BW* will keep almost constant because the peak shape around the maximum is only slightly affected by additive noise. Accordingly, we may assume that the NBD_{max} is a function of *L* solely, which in turn depends on the SNR_p. Practically, for the given α_{max} and β_{max} we calculate the spectrum of the sinusoidal signal only once and store it in memory. Then, the NBD threshold can easily be calculated by taking into account only the DFT bins of the mainlobe that lie above the noise floor given by SNR_p. The validity of this simple approximation will be checked in the next section by comparing its values to those obtained by measuring NBD_{max} for different SNR_p and different analysis windows.

5.2. NDD threshold (NDD_{min} and NDD_{max})

The sinusoidal model in (11) is herein simplified in order to investigate into the NDD thresholds. More specifically, the FM can be disregarded because it does not modify the NDD of a sinusoid. Hence,

$$x(n) = \cos(2\pi F_0 n)\times \left[1 + A_{AM}\cos(2\pi F_{AM} n + \beta)\right] + r(n) \quad (12)$$

The phase β that gives rise to the minimum and maximum values of the NDD descriptor for the signal in (12) and after applying the analysis window can be calculated numerically. The solution shows that the maximum value is obtained when the minimum of the AM envelope is located in the signal center. The minimum of the NDD is obtained for a phase β that places the AM envelope maximum close to the window center. Due to the interactions between the analysis window and the envelope the AM envelope is not exactly aligned with the window center. To simplify the discussion and due to the fact that all values of beta in the range $-\pi \le \beta \le 0$ result in a variation of the NDD of less than 1% we will use the signal pattern with AM envelope maximum in the window center for the following discussion. Accordingly the (approximate) signal patterns for the shortest and longest signal in terms of the NDD are:

$$x_{d\min} = x(n;\beta = -0.5\pi)w(n),$$
$$x_{d\max} = x(n;\beta = 0.5\pi)w(n), \quad (13)$$

where $w(n)$ is the analysis window. The envelopes of the signal patterns $x_{d\min}$ and $x_{d\max}$ for the Hanning window are displayed in Figure 3. For finite SNR_p the patterns in (13) are superposed to a narrow-band Gaussian noise. Due to the small bandwidth of the signal peak the effective noise bandwidth is rather small. For each

SNR_p there exist two noise signal patterns, $r_{d\min}$ and $r_{d\max}$, that will maximally increase respectively decrease the NDD_{max} and NDD_{min} values. We will use a very simple signal model consisting of an amplitude modulated carrier as basis for our noise model. The noise model is band limited (reflecting the bandwidth of the spectral peak) but not necessarily time limited. Due to the small bandwidth the noise pattern may extend out of the signal window. Because for the simple model we are aiming at we don't want to take into account the length of the DFT we will limit the noise signal to the time segment of the analysis window.

In order to reduce NDD_{min} $r_{d\min}$ should narrow the width of the central maximum of $x_{d\min}$. To achieve this $r_{d\min}$ must be in-phase with $x_{d\min}$ around the window's centre and in counter-phase otherwise. Because a strong amplitude at the window boundaries would always enlarge the NDD we additionally assume that the noise pattern $r_{d\min}$ has the analysis window applied.

On the contrary, $r_{d\max}$ must be in counter-phase with $x_{d\max}$ around the window's centre and in-phase close to the window edges. The resulting waveform would have the energy more uniformly distributed along the analysis window and thus larger NDD_{max}. $r_{d\max}$ must not be tapered in order to contribute significantly to the energy concentration in $x_{d\max}$ around the window edges.

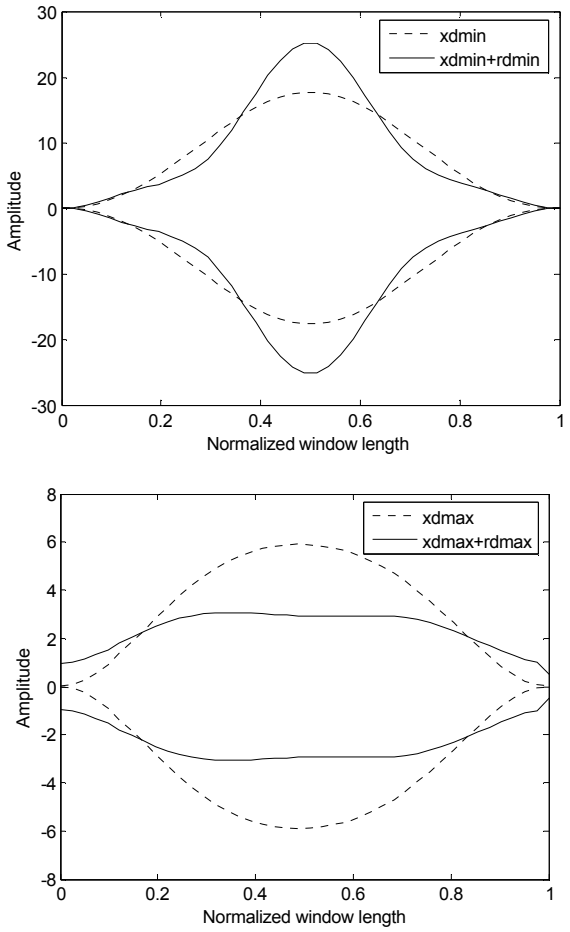

Figure 4: The resulting envelopes after the superposition of the signal patterns to the corresponding noise patterns for SNR_p = 10dB; the analysis window is Hanning.

According to the above discussion we have used the following model for the narrow-band Gaussian noise patterns:

$$r_{d\min}(n) = A\cos(2\pi F_o n)[1 + m_{\min}\cos(4\pi n/M)]w(n),$$
$$r_{d\max}(n) = -A\cos(2\pi F_o n)[1 - m_{\max}\cos(2\pi n/M)]. \quad (14)$$

The noise patterns are therefore sine-modulated waveforms. The modulation frequencies are different because the bandwidth of the peaks related to NDD_{min} and NDD_{max} are different. They have been selected such that they obey a simple relation to the window size. Note that the exact frequency values are not critical for the model and that the frequencies do not depend on the SNR_p.

The modulation indices $m_{\min}$ and $m_{\max}$ have to be greater than one in order to ensure the phase change of π in the crossover between contiguous modulation lobes. Both amplitude A and modulation indices are function of the SNR_p. A determines the total energy of each pattern while $m_{\min}$ and $m_{\max}$ control the distribution of that energy along the analysis window. The amplitude is simply a scaling factor that ensures the most of the spectral energy of the noise patterns lays SNR_p decibels under the mainlobe of the worst case signal. The values for the modulation indices are more difficult to estimate as they change in a non-linear fashion with the SNR_p. To obtain a mathematical model we have used the signal (13) and a wide range of SNR_p settings and have experimentally determined the maximum and minimum NDD as well as the values for $m_{\min}$ and $m_{\max}$ that would best match the experimental data. Finally, we derived a second order polynomial representation of the modulation indices by means of adapting a second order polynomial to the set of modulation indices. For various types of analysis windows the resulting functions are:

$$m_{\min} = \sum_i a_i SNR_p^i, \quad m_{\max} = \sum_i b_i SNR_p^i,$$

while the corresponding coefficients are given in Table 3. For the Hanning window, the envelopes of the corresponding noise patterns for SNR_p = 10dB are shown on Figure 3 while the envelopes of the resulting waveforms after the superposition are shown on Figure 4. We can observe that the energy distributions of the signal patterns have indeed been modified coherently to the aforementioned explanation. In practical applications, the signal patterns are calculated only once while the noise patterns are recalculated each time the SNR_p or type of analysis window is changed such that the new thresholds can be obtained. We will show in the following section the behavior of this model with respect to the measured NDD_{min} and NDD_{max} for different SNR_p and various analysis window types.

6. EXPERIMENTAL RESULTS

In this section we aim to check the validity of the proposed adaptive threshold selection algorithm. For different types of analysis windows and for a wide range of SNR_p values, the decision thresholds NBD_{max}, NDD_{min} and NDD_{max} were generated from the corresponding models (Section 5) and compared to their respective measured values. The measured values are obtained from the Gaussian noise added to the sinusoidal model in the proportion established by the SNR_p. The approximation errors are calculated as a difference between the measured and modeled values and are shown on Figure 5. Generally,

the approximation errors are larger for smaller SNR_p. In case of the NBD_{max} and the NDD_{min} thresholds the experimentally obtained errors show a systematic trend. This could be used to refine the model. For the NDD thresholds the error is generally overestimating the change of the boundaries that goes with the SNRp. For the NBD threshold the threshold change is underestimated. The overall approximation error is obtained by evaluating the correlation coefficient R between the measured and approximated curve for each threshold and various analysis windows. From Table 4 we can see that in almost all situations the correlation coefficient is above 0.95 which can be considered a very good approximation. Also, note that the largest approximation errors are committed in the NBD_{max} thresholding domain for the Hanning window. On the contrary, the Blackman window thresholding adapts well to the corresponding curve of measured threshold values.

Window	a_i (r_{dmin})	b_i (r_{dmax})
Hanning	0.0174 -0.5770 10.6280	-0.0006 0.1211 0.8279
Blackman	0.0081 -0.3903 9.0630	-0.0022 0.1472 0.7083
Hamming	0.0003 -0.2816 2.4716	-0.0037 0.1615 0.7230

Table 3: The coefficient values for modeling the m_{min} and m_{max} dependency on SNRp.

Window	Hanning	Blackman	Hamming
$R(NDD_{min})$	0.9567	0.9685	0.9604
$R(NDD_{max})$	0.9799	0.9792	0.9840
$R(NBD_{max})$	0.9139	0.9885	0.9585

Table 4: The correlation coefficient calculated between the measured and approximation threshold curves for various analysis windows

7. CONCLUSIONS

In this paper we have presented a new adaptive threshold selection algorithm that can be used for classification of spectral peaks. By means of the set of peak descriptors from previous work and a herein proposed compact sinusoidal model related to the analysis window, the limit values for the distributions of sinusoidal peaks in the descriptor domain can be explicitly obtained. Next, the variations of those limit values, due to the presence of noise in the sinusoidal model, are characterized in a deterministic fashion through only one parameter we refer to as the peak signal/noise ratio. By means of this user-defined parameter the descriptor limits of the classification algorithm can be controlled intuitively using as control parameter the peak signal to noise ratio.

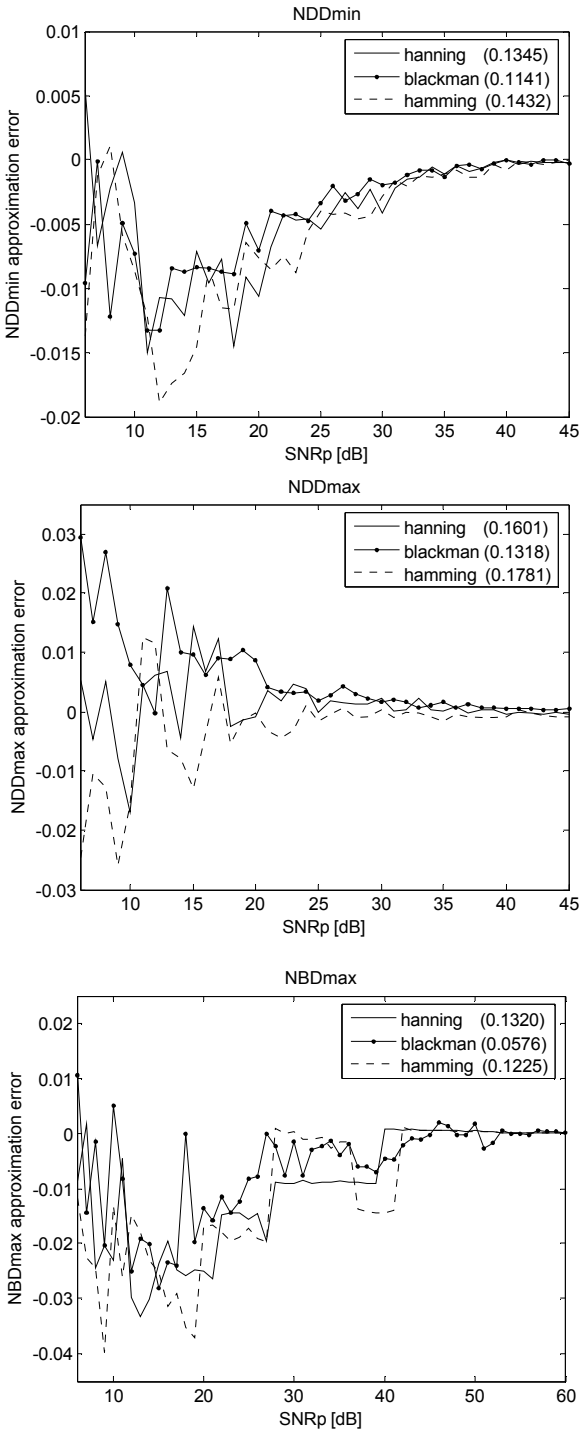

Figure 5: Approximation errors calculated as a difference between the measured and modeled values for each SNR_p and various analysis windows. The values in the legend correspond to infinite SNR_p.

The approximation accuracy given through the correlation coefficient is shown to be large for different types of analysis window. At the present state the new threshold selection method provides a control precision that can be considered sufficient for interactive control of a classification algorithm. Further investigation will be concerned with the improving the threshold models in order to reduce the approximation errors such that the precision of the control can be improved.

8. ACKNOWLEDGEMENTS

The first author of the paper would like to gratefully acknowledge the financial support of the Universidad Publica de Navarra, Spain.

9. REFERENCES

[1] A. Röbel, "A new approach to transient processing in the phase vocoder" in Proc. of the 6th Int. Conf. on Digital Audio Effects (DAFx03), 2003, pp. 344–349.

[2] M.Zivanovic, A. Röbel, X. Rodet, "A new approach to spectral peak classification", in Proc. of the 12th EUSIPCO, Vienna, Austria, September 2004, pp.1277-1280

[3] X. Rodet, "Musical sound signal analysis/synthesis: Sinusoidal + residual and elementary waveform models," in Proc IEEE Time-Frequency and Time-Scale Workshop 97, (TFTS'97), 1997

[4] D. J. Thompson, "Spectrum estimation and harmonic analysis", IEEE Proc. Vol.70, No.9, September 1982

[5] C. Yeh, A. Röbel, X. Rodet, 'Multiple Fundamental Frequency Estimation Of Polyphonic Music Signals", Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2005), pp. 225-228, Vol III, 2005.

[6] C. Yeh, A. Röbel, "Adaptive noise level estimation", Proc. of the 9th Int. Conf. on Digital Audio Effects (DAFx'06), Montreal, pp. 145-148, 2006.

[7] P. Depalle, G. Garcia, X. Rodet, "Tracking of partials for additive sound synthesis using hidden Markov models," in Proc. Int. Conf. on Acoustics, Speech and Signal Processing, vol. I, pp. 242–245, 1993

[8] L.Cohen, "Time-frequency analysis", Prentice Hall, 1995

[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, "Adaptive additive modeling with continuous parameter trajectories", IEEE Transactions on Speech and Audio Processing, Vol. 14, No. 4, pp.1440-1453, 2006.

[11] V. Verfaille, C. Guastavino, P. Depalle, "Perceptual evaluation of vibrato models", Proc. of the Conf. on Interdisciplinary Musicology (CIMOS), 2006

[12] I. Arroabarren, X. Rodet, A. Carlosena, "On the measurement of the instantaneous frequency and amplitude of partials in vocal vibrato", ", IEEE Transactions on Speech and Audio Processing, Vol. 14, NO. 4, pp.1413-1421, July 2006