DETECTION OF SINUSOIDAL SIGNALS IN NOISE BY PROBABILISTIC MODELLING OF THE SPECTRAL MAGNITUDE SHAPE AND PHASE CONTINUITY

Peter Jančovič and Münver Kökür

School of Electronic, Electrical & Computer Engineering, University of Birmingham, Birmingham, UK

{p.jancovic, m.kokuer}@bham.ac.uk

ABSTRACT

This paper presents a method for detection of sinusoidal signals corrupted by an additive noise in the short-time Fourier domain. The proposed method is based on probabilistic modelling of the spectral magnitude shape and phase continuity around spectral peaks and can deal with both stationary and non-stationary sinusoidal signals. Experimental results are presented for both sinusoidal signals of a constant frequency and frequency varying continuously over time. The performance is analysed in terms of the false acceptance and false rejection error rates of spectral peak detection and also compared to our previous method. Experimental results demonstrate very high detection accuracy in very even strong noisy conditions.

Index Terms— detection, sinusoidal signal, noise, short-time spectrum, DFT, spectral shape, phase continuity, speech, audio, music.

1. INTRODUCTION

Detection and parameter estimation of unknown sinusoidal signals from a finite set of signal observations is an important problem with applications in various areas, including speech and audio processing, image processing, and radar detection.

Sinusoidal modelling has been widely used to represent speech and audio signals for analysis, synthesis, and coding, e.g., [1] [2]. The most common technique for detecting sinusoidal components in audio processing is the spectral peak picking. However, due to spurious peaks, this produces a high number of falsely detected sinusoids. This may be improved by incorporating additional harmonicity constraints or temporal continuity criteria for peaks magnitude values or their frequencies. However, the use of such criteria may become problematic in many practical situations when there is a large number of strong peaks in the spectrum, such as, in polyphonic music or when a background noise is present in the signal.

There have been several approaches to classification whether a spectral peak corresponds to a sinusoidal signal or noise. In [3], the complex correlation between the DFT of the analysis window and each peak of the signal was used. The approach proposed in [4] calculates various statistics for each side of the peak and uses likelihood ratio test to make the decision. A sinusoidality criterion derived from the phase spectrum, which was proposed in [5] and employed for estimation of the fundamental frequency of clean speech signals, was employed in [6] to verify whether a spectral peak is considered as a sinusoid in musical signals. All the above works used the particular method for decision about the spectral peaks but did not provide any formal evaluation of the method and did not deal with noise-corrupted sinusoidal signals. The authors in [7] evaluated several algorithms for extraction of sinusoids in stationary sounds but used only several simple rules on frequency and magnitude continuity, such as magnitude varying by less than 3 dB, to decide on the validity of the peaks. In [8], the variation of the frequency reassignment operator and null-to-null bandwidth, i.e., the spectral mainlobe width, was employed for distinguishing spectral peaks of noise from sinusoidal signals. Evaluations were presented only in terms of the percentage of noise peaks classified as sinusoids. In our previous research presented in [9] we used the Euclidean distance to compare the spectral peak magnitude shape to the analysis window and evaluated the detection performance for noise corrupted simulated harmonic signals and speech. We employed the above method for speech and speaker recognition, speech alignment and enhancement, e.g., [10], and recently also for bird signal detection and recognition [11].

In this paper, we extend our previous research in [9] and present a method for detection of sinusoidal signals in noise which uses both the spectral magnitude shape and phase continuity and employs probabilistic modelling. The modelling is performed using Gaussian mixture models (GMM). The method is first presented for dealing with stationary sinusoidal signals but then we also discuss its employment to non-stationary sinusoids. Experimental evaluations are performed on both stationary and non-stationary sinusoidal signals corrupted by additive white Gaussian noise at various SNRs. The results are presented in terms of the false acceptance and false rejection error rates. It is demonstrated that the proposed method significantly outperforms our previous method and achieves very high detection accuracy even in very strong noisy conditions.

2. SINUSOIDAL DETECTION BY MODELLING THE SPECTRAL MAGNITUDE SHAPE AND PHASE CONTINUITY

We consider signal $x(n)$ consisting of an unknown number of sinusoidal components corrupted by additive zero mean white Gaussian noise. The short-time Fourier Transform of a frame of the signal $x(n)$ consisting of $N$ samples is defined as

$$X_l(k) = \sum_{n=0}^{N-1} x(n)L + n)w(n)e^{-j2\pi kn/N}$$

(1)

where $w(n)$ denotes the analysis window function, $L$ is the shift between adjacent signal frames in samples and $l$ is the frame-time index. $X_l(k)$ can be written as $|X_l(k)|e^{j\phi_l(k)}$, where $|X_l(k)|$ is the short-time magnitude spectrum and $\phi_l(k)$ is the short-time phase spectrum. The $X_l(k)$ can be expressed as the summation of scaled and shifted versions of the Fourier transform of the analysis window $W(k)$, which is the principle we have previously exploited in [9] and also further extend here.
2.1. Previous method based on spectral magnitude shape

In our previous research \[9\], the detection was based on a threshold decision performed on the distance \( D(k) \) between the signal short-time magnitude spectrum around the frequency index \( k \) of a spectral peak and the spectrum of the analysis window \( W(k) \) calculated as

\[
D(k) = \left[ \frac{1}{2M+1} \sum_{m=-M}^{M} \left| X(k+m) - |W(m)| \right|^2 \right]^{1/2}
\]

where \( M \) determines the number of components of the spectrum at each side around the peak \( k \) to be compared, and both, the \( X(k) \) and \( W(k) \) are normalised to have magnitude value equal to 1 at \( m=0 \) prior to their use in Eq. 2. This method does not take into account the fact that the DFT provides only samples of the spectrum and as such the shape of the short-time magnitude spectrum of a sinusoidal signal whose frequency does not correspond to an integer frequency index \( k \) will show difference to the shape of the \( W(k) \). Such sinusoid will then produce a non-zero distance even in noise-free conditions and this may contribute to an incorrect detection.

2.2. Proposed method

The proposed method tackles the detection problem as a pattern recognition problem. It employs the information about both the spectral magnitude shape and the phase continuity around the spectral peak. Each spectral peak is represented by a set of features extracted from the short-time spectrum. The detection system consists of training and recognition stage. In the training stage, the feature vectors are collected for peaks corresponding to the sinusoidal signals and noise. The distributions of these features are modelled by using Gaussian mixture models (GMMs). The outcome of the training provides an individual model for spectral peaks corresponding to sinusoidal signals and noise. A given spectral peak is then classified as a sinusoid or noise based on the maximum likelihood criterion. The following sections present details of the proposed method.

2.2.1. Spectral magnitude shape and phase continuity

In the proposed method, we capture the spectral magnitude shape around a spectral peak by considering it to be a multivariate random variable. The modelling of the magnitude shape is performed by using the feature vector \( y \) that contains the spectral magnitude values over the range of frequency bins from \( k-M \) to \( k+M \), i.e., \( y=(S(k-M), \ldots, S(k-1), S(k+1), \ldots, S(k+M)) \). Unlike the use of a fixed idealised shape of \( W(k) \) as in the case of the distance used in the previous method, the modelling of the entire magnitude vector \( y \) will allow us to model the variations of the \( W(k) \) shape due to DFT sampling.

In addition to modelling the spectral magnitude shape, we also incorporate the modelling of the phase continuity. We consider that a sinusoidal signal should demonstrate a phase continuity between adjacent signal frames, while noise signal would have a random phase. Let us define the phase difference between the current and previous signal frame at the spectral peak \( k \) as

\[
\Delta \phi_{\text{i}}(k) = \phi_{\text{i}}(k) - \phi_{\text{i}-1}(k) - 2\pi k L / N \tag{3}
\]

where \( \phi_{\text{i}}(k) \) and \( \phi_{\text{i}-1}(k) \) denote the phase at frame-time \( l \) and \( l-1 \), respectively. The term \( 2\pi k L / N \) is included in order to compensate for the shift of the sinusoidal signal between the adjacent signal frames, with \( L \) being the shift in samples. The modelling of the phase continuity is performed by using the feature vector \( y \) that contains the spectral phase difference values over the range of frequency bins from \( k-M \) to \( k+M \), i.e., \( y=(\Delta \phi_{\text{i}}(k-M), \ldots, \Delta \phi_{\text{i}}(k+M)) \).

Moreover, we also consider modelling of the joint spectral magnitude shape and phase continuity features, in which case the feature vector \( y \) will be formed by concatenation of the above magnitude and phase difference feature vectors.

An example of the distribution of the magnitude values \( |S(k-2)| \) and \( |S(k+2)| \) and of the phase differences \( \Delta \phi_{\text{i}}(k-2) \) and \( \Delta \phi_{\text{i}}(k+2) \) for noise and for sinusoidal signal corrupted at various local SNRs is depicted in Figure 1 (a) and (b), respectively. It can be seen that while the distributions are widely spread for noise signal, they are highly localised in a narrow area for sinusoidal signals at 40 dB and widens as the SNR decreases. Note that when the frequency of the sinusoidal signal does not fall exactly on a frequency bin \( k \), this will cause the compensation term in Eq. 3 being inaccurate and as such the phase difference in Figure 1 (b) for the 40 dB case is not localised around zero but is a narrow line.

![Fig. 1. Distribution of the magnitudes (a) and phase differences \( \Delta \phi_{\text{i}}(k) \) (b) for spectral elements \( X(k-2) \) and \( X(k+2) \) for noise (red) and sinusoidal signals corrupted by noise at the local SNR of 10 dB (blue), 20 dB (black), and 40 dB (green).](image)

The distributions in Figure 1 considered a stationary sinusoidal signal. When a sinusoidal signal is non-stationary, either the amplitude or frequency varies over time, the distributions of the spectral magnitude and phase difference features will be different to stationary sinusoids. For instance, let us consider a sinusoidal signal of a continuously varying frequency. The frequency variations of a sinusoid within a single frame will spread the energy over a larger number of frequency bins and the magnitude around the sinusoidal peak becomes of a wider shape than the \( W(k) \). However, it can be shown that the magnitude will still have a specific shape which is largely different to the shape exhibited by noise peaks. An example of this is illustrated in Figure 2. In the case of phase, the phase difference for frequency bins surrounding the peak will become non-zero but these will actually capture the type of frequency variations present in the signal. Thus, again these will be different to noise. In this paper, we compensate for the differences between the features of stationary and non-stationary sinusoidal signals by including the non-stationary sinusoidal signals in the modelling.

2.2.2. Probabilistic modelling

We model the distribution of the multivariate feature vector \( y \), which may represent the spectral magnitude shape or the phase continuity or both, by using Gaussian mixture model (GMM). An \( L \)-component GMM \( \lambda \) is a linear combination of \( L \) Gaussian probability density
functions and has the form
\[ p(y|\lambda) = \sum_{l=1}^{L} w_l b_l(y) \]  
(4)

where \( w_l \) is the weight and \( b_l(y) \) is the multivariate Gaussian density of the \( l \)-th mixture component. Each \( b_l(y) \) is defined by the mean vector \( \mu_l \) and covariance matrix \( \Sigma_l \). For computational reasons, we employed Gaussian densities with diagonal covariance matrix. Let us denote by \( \lambda \) the set of GMM parameters, i.e., \( \lambda = \{ w_l, \mu_l, \Sigma_l \}_{l=1}^{L} \).

In the training stage, the parameters of the GMM, modelling the distribution of the features \( y \), are obtained for noise signal, denoted by \( \lambda_n \), and for sinusoidal signals corrupted by noise at various SNRs, denoted by \( \lambda_{snr} \). A large collection of features \( y \) corresponding to spectral peaks of the noise and sinusoidal signal at the given local SNR are used as the training data to estimate the GMM parameters. In experimental evaluations, we used GMMs of 32 mixture components in all cases, i.e., for each types of features employed and sinusoidal or noise models.

In the testing stage, the decision whether a spectral peak \( k \) at a given signal frame corresponds to a sinusoidal signal or not is based on the maximum likelihood criterion, i.e., the peak \( k \) is detected as a sinusoid if \( p(y|\lambda_{s}) > p(y|\lambda_{n}) \). The \( \lambda_{s} \) is one of the GMMs of the sinusoidal signals from the set \( \{ \lambda_{s(sn)}; SNR=SNR_{min}, \ldots, SNR_{max} \} \) and its choice can be decided based on the required amount of the false acceptance error rate.

3. EXPERIMENTAL EVALUATIONS

This section presents experimental evaluation of the detection methods on both stationary and non-stationary sinusoidal signals corrupted by white Gaussian noise at various signal-to-noise ratios (SNRs). The signal is sampled at 8 kHz and is divided into frames of 256 samples with an overlap of 80 samples. Based on the results of our previous research in [9], the Hamming analysis window is used and the DFT size is set to 512 points, i.e., the signal is appended by 256 zeros in order to provide a finer DFT sampled spectrum. The detection results are obtained as the average over 1000 independent runs. In each run, the frequency of the sinusoidal signal was randomly generated from a uniform distribution in the range from 100 to 2500 Hz. In the case of non-stationary sinusoidal signals, the frequency of the signal continuously varied in a linear manner over time. The amount of frequency variation was fixed to 1000 Hz per second, which corresponds to 32 Hz variations within a single frame. The signal was corrupted by additive white Gaussian noise at various global SNRs. The local SNR for each spectral peak \( k \) was calculated as \( 20 \cdot \log_{10}(\frac{X_S}{X_N}) \), where \( X_S \) and \( X_N \) is the average magnitude of the clean signal and noise, respectively, each calculated using the magnitude values within the main lobe width of the analysis window around the frequency bin \( k \), i.e., four bins above and below \( k \) in our case. The performance is evaluated based on spectral peaks and presented in terms of false acceptance (FA) and false rejection (FR) error rates. For comparison, we also present results obtained using our previous method presented in [9].

3.1. Evaluations on stationary sinusoidal signals

This section presents parameter set-up and evaluation for stationary sinusoidal signals.

First, we evaluate the effect of the value of the parameter \( M \). An upper-bound for value of the parameter \( M \) can be considered to be the half of the main lobe width of the analysis window spectrum, which corresponds to four in our parameter set-up. The obtained results are presented for modelling the magnitude and phase continuity in Figure 3 (a) and (b), respectively. It can be seen that when modelling the spectral shape the value of \( M \) has a large impact on the detection performance, with the performance improving significantly with increasing the value for \( M \). In the case of phase continuity, there is a considerable improvement when \( M \) increases from 2 to 3 but the results are nearly the same for \( M \) set to 3 or 4. Thus, in all the following experiments we set \( M \) to 4.

3.2. Evaluations on non-stationary sinusoidal signals

This section presents evaluations for non-stationary sinusoidal signals whose frequency varies continuously over time. The obtained
results are depicted in Figure 5. In the figures, for comparison, the full line results depict the results for the stationary sinusoids obtained by models trained also on stationary sinusoids. In Figure 5 (a), we demonstrate the results obtained when using the models trained on the stationary sinusoidal signals but testing using non-stationary sinusoids, i.e., uncompensated models. It can be seen that the performance decreases largely in comparison to the stationary case. Figure 5 (b) presents the results obtained using compensated models, i.e., models trained on the non-stationary sinusoidal signals. It can be seen that the performance improves significantly in comparison to the use of stationary trained models. Indeed, the performance is not far from the stationary test case. This demonstrates that the proposed method can be employed also for detection of non-stationary sinusoidal signals by accounting for the non-stationarity in the probabilistic models.

4. CONCLUSION

In this paper, we presented methods for detection of sinusoidal signals in noisy conditions in the short-time Fourier transform domain. The proposed methods were based on the extraction of the spectral magnitude shape and phase continuity features around spectral peaks and probabilistic modelling was employed to model the distribution of these features for noise and sinusoidal signals. The modelling was performed using Gaussian mixture models. We discussed that the proposed method can be used for the detection of both stationary and non-stationary sinusoidal signals by including the non-stationarity in the modelling. Evaluations were presented for both stationary and non-stationary sinusoidal signals corrupted by additive white Gaussian noise. The performance was also compared to our previous method employing distance calculation of spectral magnitudes. The proposed methods showed very high detection performance even in low SNR conditions.

This work was partly supported by UK EPSRC grant EP/D033659/1.

5. REFERENCES


