FREQUENCY, PHASE AND AMPLITUDE ESTIMATION OF OVERLAPPING PARTIALS IN MONAURAL MUSICAL SIGNALS

Marco Fabiani, *

Dept. of Speech, Music and Hearing (TMH)
School of Computer Science and Communication (CSC)
Royal Institute of Technology (KTH)
Stockholm, Sweden
himork@kth.se

ABSTRACT

A method is described that simultaneously estimates the frequency, phase and amplitude of two overlapping partials in a monaural musical signal from the amplitudes and phases in three frequency bins of the signal's Odd Discrete Fourier Transform (ODFT). From the transform of the analysis window in its analytical form, and given the frequencies of the two partials, an analytical solution for the amplitude and phase of the two overlapping partials was obtained. Furthermore, the frequencies are estimated numerically solving a system of two equations and two unknowns, since no analytical solution could be found. Although the estimation is done independently frame by frame, particular situations (*e.g.* extremely close frequencies, same phase in the time window) lead to errors, which can be partly corrected with a moving average filter over several time frames. Results are presented for artificial sinusoids with time varying frequencies and amplitudes, and with different levels of noise added. The system still performs well with a Signal-to-Noise ratio of down to 30 dB, with moderately modulated frequencies, and time varying amplitudes.

1. INTRODUCTION

Separation of different instruments in a polyphonic music recording is a problem that has been extensively studied in recent years. This stems from the fact that many different tasks related to digital music, being extraction of cues for automatic analysis and classification, audio compression or sound manipulation, make use of source separation. Different applications have different specific assumptions, such as the number of available channels or the number of instruments recorded simultaneously, and requirements for the separation accuracy.

The introduction to a recent article by Li et al. [1] presents a concise but complete overview of different techniques for source separation in monaural signals. The various techniques are divided into three categories based on the underling general approach: psycho acoustical techniques (auditory scene analysis), statistical techniques (*e.g.* ICA, independent component analysis) and signal processing techniques (*e.g.* sinusoidal modeling). A particularly difficult task, especially in the case of single channel signals, is the separation of overlapping partials. Pitched acoustic instruments usually produce harmonic spectra (*i.e.* the frequencies of the different partials are multiples of that of the fundamental). Since Western music is based on a twelve-tone equal tempered scale, pitches in musical intervals are commonly in an integer ratio relationship. This leads to a very high number of partials from different instruments to have almost the same frequency. To obtain an accurate separation, these overlapping partials should also be resolved. A short summary of the overview from [1] is presented below, with particular emphasis on overlapping partials separation in monaural signals.

Computational auditory scene analysis (CASA) [2] aims at building computational models that mimic the complex auditory scene analysis performed in the human brain. Specific systems have been developed for monaural sound separation [3, 4]. These methods do not attempt to explicitly resolve overlapping partials, but simply assign the entire energy to one source.

A wide variety of statistical techniques for source separation have been proposed in recent years. Strong assumptions about the signals are made when using these techniques, such as the fact that sources are independent, or that a sparse representation of a source is possible (*i.e.* a signal can be represented by a weighted sum of bases from an over-complete set, with values of weights mostly zero for most of the time). Independent Subspace Analysis (ISA) [5] and Nonnegative Matrix Factorization (NMF) [6, 7] have been used for monaural source separation. Statistical methods handle overlapping partials implicitly, but since they work solely on the magnitude of the spectrum, they disregard the phase information. It has to be pointed out that if two sinusoids have very close frequencies but opposite phases, they tend to cancel each other out, while if they have the same phase, their magnitudes add up. This means that phase information must be explicitly taken into account if accurate partials separation is needed.

Sinusoidal modeling is based on the assumption that a musical signal can be divided into the sum of time-varying sinusoidal components, and eventually, a stochastic residual component (a comparison of different techniques can be found in [8]). Sinusoidal modeling methods allow for both source separation and resynthesis of audio signals, and thus are also known as Analysis-Synthesis techniques. The problem of source separation reduces to the estimation of the parameters of the single sinusoidal components (*i.e.* frequency, amplitude and phase).

One of the first attempts to resolve overlapping partials based on a sinusoidal model was described in [9]. The method is based on the fact that, if the sinusoidal components are assumed to be stationary, their transforms can be expressed using the transform of the analysis window by which the time signal is multiplied before being transformed to the frequency domain. The author makes

* This work was supported by the SAME project (FP7-ICT-STREP-215749)

two strong assumptions: there must be a dominant signal (*i.e.* the amplitude of one component must be much larger than the other), and the two components must be separated by an appreciable difference in frequency. Also in this method, only the magnitude of the transform is used, disregarding the phase information.

More recent methods [10, 11, 12] use information about adjacent, non-overlapping partials, to estimate the amplitude of the overlapping ones, assuming that the spectral envelope of an instrument is smooth. This assumption, for real instrument sounds, is often violated. Instrument models have also been used that contain information about relative amplitudes between partials [13], but the limit here is that different dynamic levels, playing styles, instrument specimen, etc. have a big impact on the models. Finally, common amplitude modulation (CAM), *i.e.* the fact that amplitude envelopes of different harmonics of the same source tend to be similar, is used in [1] in a least-squares framework to resolve overlapping partials. The method makes also use of phase information in the signal's transform.

The primary use of the method described in the present paper is in a system, under development, that aims at real-time, rule-based expressive modification of audio musical recordings, and which is based on sinusoidal modeling. A preliminary description of the system can be found in [14]. In this system, the technique described by Ferreira in [15] has been used for the estimation of the partials' parameters because of the accuracy of the frequency estimation. This technique works well for single sinusoidal components, but does not address the problem of overlapping components, which is one of the main obstacles for an effective modification of audio such as the one described in [14]. The present paper addresses this problem by combining the technique described in [15] for sinusoidal model analysis with the approach proposed by Parsons in [9] for overlapping partials separation.

2. METHOD

2.1. Single sinusoid approximate estimation [15]

An important aspect of sinusoidal modeling is to obtain accurate model parameters (*i.e.* frequency, amplitude and phase of the sinusoidal components). In [15], Ferreira describes a method to accurately estimate these parameters from the Odd Discrete Fourier Transform (ODFT) of the signal. In this section, his method is briefly described, as a starting point for the new methods presented in Sec. 2.2 - 2.4.

The audio signal is first divided into overlapping time frames (75% overlap in the examples shown in this paper, $N = 4096$, $f_s = 44100$) and multiplied by the sine window

$$h(n) = \sin \frac{\pi}{N}(n + \frac{1}{2}) \qquad n = 0, \ldots, N-1 \tag{1}$$

where N is the length of the frame in samples. The windowed signal is then transformed using the Odd Discrete Fourier Transform (ODFT), which is defined as:

$$X(k) = \sum_{n=0}^{N-1} h(n)x(n) \exp^{-j\frac{2\pi}{N}(k+\frac{1}{2})n} \tag{2}$$

where k is the frequency bin.

The sine window and the ODFT were chosen by Ferreira since the method was used in a MDCT-based perceptual audio codec. The sine window is used in many perceptual audio codecs (*e.g.* MP3, AAC, MPEG-4) since it is the square root of a Hann window: when applied twice (before the analysis and after the synthesis), the result is perfect reconstruction for 50% overlap. Furthermore, the conversion from ODFT to MDCT, a filter bank commonly used in audio coding, is very simple, but the frequency estimation is more accurate in the ODFT domain [16].

The sinusoidal component $x(n)$ is represented as a discrete sinusoid of the form:

$$x(n) = A \sin\left(\frac{2\pi}{N}(l + \Delta l)n + \Phi\right) \tag{3}$$

where A and Φ are the amplitude and initial phase of the sinusoid, and the frequency is written as the sum of an integer part l/N, which corresponds to the lth frequency bin in the ODFT, and a fractional part $\Delta l/N$. It is possible to accurately estimate the three parameters A, Φ and Δl by observing that the amplitude and phase of the ODFT in the lth bin (*i.e.* the bin with the maximum amplitude) and the two adjacent bins $l-1$ and $l+1$ can be directly expressed as a function of the normalized magnitude of the window's transform $|\widehat{H(w)}|$, which for the window defined in Eq. (1), is analytically expressed as [15]:

$$|\widehat{H(w)}| = 2\sin\left(\frac{\pi}{2N}\right)\left|\cos\left(N\frac{\omega}{2}\right)\right| \ldots \left|\frac{1}{\sin\left(\frac{1}{2}(\frac{\pi}{N}-\omega)\right)} + \frac{1}{\sin\left(\frac{1}{2}(\frac{\pi}{N}+\omega)\right)}\right| \tag{4}$$

Since $|\widehat{H(w)}|$ is not continuous for $\omega = \pm\pi/N$, in [15] the author uses a continuous approximation of the main lobe

$$|\widehat{H(w)}| \approx \left[\cos\left(\frac{n}{6}\omega\right)\right]^G \qquad |\omega| < \frac{3\pi}{N} \tag{5}$$

where G is a numerical constant, obtained by minimizing the difference between Eq. (4) and (5). Using this approximation, an analytical solution for the instantaneous frequency (expressed as Δl), phase and amplitude of the sinusoid (assumed to be stationary) is obtained as

$$\Delta l \approx \frac{3}{\pi}\arctan\left(\frac{\sqrt{3}}{1 + 2\left(\frac{|X(l-1)|}{|X(l+1)|}\right)^{\frac{1}{G}}}\right) \tag{6}$$

$$A \approx \frac{4|X(l)|}{N}\left|\frac{\sqrt{3}}{2\cos\left(\frac{\pi}{6}(2\Delta l - 1)\right)}\right|^F \tag{7}$$

$$\Phi = \angle X(l) + \pi\left(1 - \frac{1}{2N}\right) - \pi\Delta l\left(1 - \frac{1}{N}\right) \tag{8}$$

where F is another numerical constant, not necessarily equal to G, but obtained in a similar way. The maximum fractional frequency estimation error, as shown in [15], is approximately 1%.

2.2. Single sinusoid numerical estimation

In this section, a numerical approach to the estimation of a single sinusoidal component in described. This approach improves the accuracy of the estimation at the cost of an increased estimation time. Furthermore, it generalizes the method in [15] by allowing any analysis window with a computable $H(\omega)$ to be used. It is also a building block of the technique described in Sec. 2.2 - 2.4

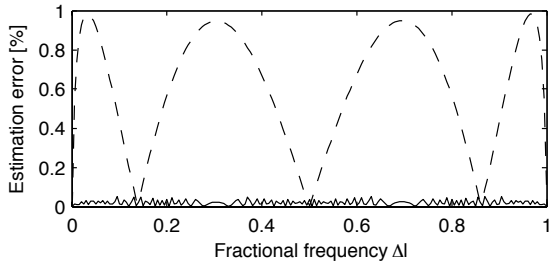

Figure 1: *Error in the estimation of the fractional frequency Δl for a single sinusoid, expressed as percentage of the bin width. Comparison between the approximate analytical method [15] (dashed line) and the numerical method (Eq. (9), dotted line). The peak error for the numerical method is 0.05%.*

for overlapping components separation.

Although the approximation made in [15] for the main lobe of $|\widehat{H(w)}|$ (Eq. (5)) allows for an analytical solution to the estimation problem, and gives accurate results, Eq. (4) is numerically computable, except for $\omega = \pm\pi/N$. An important expression derived in [15] which relates the amplitudes of two adjacent frequency bins through Δl is

$$\frac{|H\left(\frac{2\pi}{N}(\Delta l + \frac{1}{2})\right)|}{|H\left(\frac{2\pi}{N}(\Delta l - \frac{3}{2})\right)|} - \frac{|X(l-1)|}{|X(l+1)|} = 0 \qquad (9)$$

This relation was used to derive Eq. (6). By substituting the exact transform $H(\omega)$ (Eq. (4)) into Eq. (9), we can in theory improve the accuracy of the estimation, compared to the original method described in [15]. Unfortunately, it was not possible to find an analytical solution for Δl to the implicit equation obtained combining Eq. (4) and (9). A numerical solution was then computed using, as a starting point, the approximate $\overline{\Delta l}$ obtained with Eq. (6). Throughout this paper, to numerically solve implicit functions (in this case Eq. (9)), a simple grid search was used. In the point of discontinuity of Eq. (4), the function was linearized and an interpolated real value was returned.

A comparison of the estimation accuracy of the original method from [15] and of the one based on the numerical solution to Eq. (9) is shown in Fig. 1. It can be observed from Fig. 1 that the maximum estimation error is 20 times smaller for the numerical method, down to 0.05%. The accuracy is limited by the step size in the grid search (8×10^4 in this paper). The drawback of the numerical solution is the estimation time, which is proportional to the number of steps in the grid search, compared to a single operation for the original method. The phase is estimated from Eq. (8) (which is already an exact solution). The amplitude is estimated using a formula similar to Eq. (7), were Eq. (4) is used instead of Eq. (5) to compute $|H(\omega)|$.

2.3. Estimation of amplitude and phase of two overlapping sinusoids

For the remainder of this paper, the word *overlapping* will indicate two sinusoids, $x_1(n)$ and $x_2(n)$, whose frequencies are so close that, if taken separately, the peak in their ODFT would fall in the same frequency bin l. The case in which the peak falls into two adjacent bins (*e.g.* l and $l-1$) is not considered in this paper, but the method described in the following sections can be easily extended to cover also that case. Thus, by using the same notation as in Eq. (3), the separation of two overlapping partials reduces to the estimation of the two fractional frequencies Δl_1 and Δl_2, the two amplitudes A_1 and A_2, and the two phases Φ_1 and Φ_2, from the ODFT of the signal $x(n) = x_1(n) + x_2(n)$ (assuming for now that there is no noise in the signal).

The method described in this section combines the results from Sec. 2.1 - 2.2 with the approach to overlapping sinusoidal components separation used by Parsons in [9]. From the few details given by the author it was not possible to reproduce the original algorithm, and thus compare the results. Nevertheless it seems safe to affirm that the method presented here is an improvement over Parsons' approach because it removes the two main constraints described in [9]: the two components must have appreciably different amplitudes, and appreciably different frequencies. The only limitation of the present method is that the two components cannot have the *exact* same frequency. The improvement is a consequence of two factors: first, the parameters of both components are estimated simultaneously instead of estimating the largest component first and subtracting it from the total spectrogram, as in [9]; second, the phase information in the transform of the signal, which was discarded by Parsons, is also taken into account.

The problem of estimating the two partials' parameters Δl_1, Δl_2, A_1, A_2, Φ_1 and Φ_2 is equivalent to that of estimating their respective ODFT transforms $X_1(k)$ and $X_2(k)$, and then estimating the parameters using the method described in Sec. 2.1. First, only an expression for the amplitudes and phases will be derived, assuming the fractional frequencies Δl_1 and Δl_2 are known. This is in fact a realistic situation if the fundamental frequencies of two harmonic signals are known.

Let us start by defining a few symbols that will simplify the notation. The four unknowns we want to estimate are

$$r_j = \Re\{X_j(l-1)\}\,, \qquad j = 1,2 \qquad (10)$$
$$i_j = \Im\{X_j(l-1)\}\,, \qquad j = 1,2 \qquad (11)$$

We know from Eq. (9) that there is a relation between the real parts (and similarly for the imaginary parts) of the signal's transform in two different frequency bins through the ratio between the transform of the time window $H(\omega)$ evaluated at their corresponding values of $\omega_j(k)$

$$\omega_j(k) = \frac{2\pi}{N}(l + \Delta l_j - k - \frac{1}{2})\,, \qquad j = 1,2 \qquad (12)$$

Let us define the relation between frequency bins l and $l-1$ as

$$P_j = \frac{|H(\omega_j(l))|}{|H(\omega_j(l-1))|}\,, \qquad j = 1,2 \qquad (13)$$

This is constant, since Δl_1 and Δl_2 are assumed to be known for the time being.

We can now express the real and imaginary parts of $X_1(l)$ and $X_2(l)$ as functions of the four unknowns defined in Eq. (10) - (11) and, knowing that [15]

$$\angle X_j(l) = \angle X_j(l-1) - \pi\left(1 - \frac{1}{N}\right)\,, \qquad j = 1,2 \qquad (14)$$

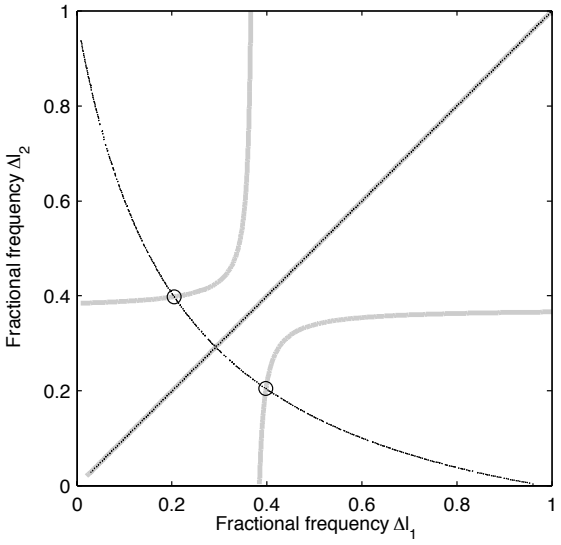

Figure 2: *Eq. (29) (black line) and (30) (grey line) evaluate numerically for* $\Delta l_1 = 0.4$ *and* $\Delta l_2 = 0.2$*, when* $\angle X_1(l) \neq \angle X_2(l)$ *(intersection at* $[0.2048, 0.3980]$*). The two curves coincide on the diagonal.*

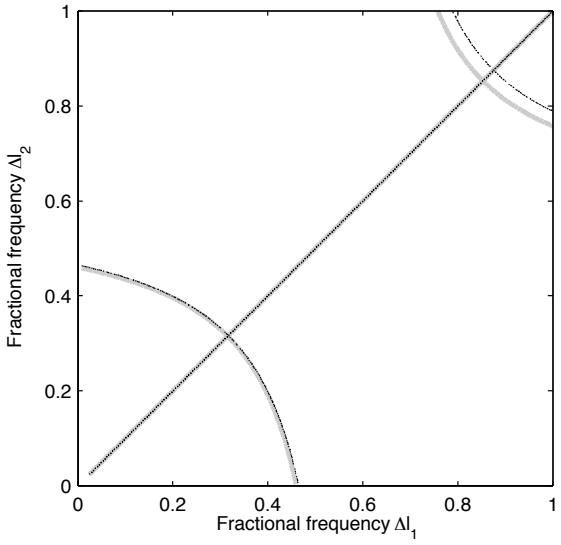

Figure 3: *Eq. (29) (black line) and (30) (grey line) evaluate numerically for* $\Delta l_1 = 0.4$ *and* $\Delta l_2 = 0.2$*, when* $\angle X_1(l) \approx \angle X_2(l)$*: note how the two curves intersect only on the diagonal.*

we obtain

$$\Re\{X_j(l)\} = P_j \left(-\cos\left(\frac{\pi}{N}\right) r_j + \sin\left(\frac{\pi}{N}\right) i_j\right) \quad (15)$$

$$\Im\{X_j(l)\} = P_j \left(-\sin\left(\frac{\pi}{N}\right) r_j - \cos\left(\frac{\pi}{N}\right) i_j\right) \quad (16)$$

Observing that $X(k) = X_1(k) + X_2(k)$, the following system of equations can be written:

$$\begin{cases} r_1 + r_2 - X_R(l-1) = 0 & (17) \\ i_1 + i_2 - X_I(l-1) = 0 & (18) \\ P_1 (C_1 r_1 - S_1 i_1) + P_2 (C_1 r_2 - S_1 i_2) + X_R(l) = 0 & (19) \\ P_1 (S_1 r_1 + C_1 i_1) + P_2 (S_1 r_2 + C_1 i_2) + X_I(l) = 0 & (20) \end{cases}$$

where $X_R(\cdot)$ and $X_I(\cdot)$ are the real and imaginary parts of $X(\cdot)$, $C_1 = \cos(\pi/N)$ and $S_1 = \sin(\pi/N)$.

The solutions are easily found:

$$r_1 = -\frac{C_1 X_R(l) + S_1 X_I(l) + P_2 X_R(l-1)}{P_1 - P_2} \quad (21)$$

$$i_1 = -\frac{C_1 X_I(l) - S_1 X_R(l) + P_2 X_I(l-1)}{P_1 - P_2} \quad (22)$$

$$r_2 = \frac{C_1 X_R(l) + S_1 X_I(l) + P_1 X_R(l-1)}{P_1 - P_2} \quad (23)$$

$$i_2 = \frac{C_1 X_I(l) - S_1 X_R(l) + P_1 X_I(l-1)}{P_1 - P_2} \quad (24)$$

The amplitudes and phases of the two sinusoids are then separately evaluated using the method described in Sec. 2.1 - 2.2.

2.4. Overlapping sinusoids frequency estimation

In real situations, it is possible that the signals we are trying to separate are not perfectly harmonic (*e.g.* piano tones). For this reason, it is important to be able to also estimate the frequency of the two overlapping partials. This can be done by extending (17) - (20) with two additional equations.

Let us begin by observing that $\omega_j(k)$ (Eq. (12)) depends on the fractional frequency Δl_j. Thus, if Δl_1 and Δl_2 are unknown, P_1 and P_2 (Eq. (13)) are no longer constants. Using the same reasoning behind Eq. (13), we can define the relation between frequency bins $l-1$ and $l+1$ as

$$Q_j = \frac{|H(\omega_j(l+1)|}{|H(\omega_j(l-1)|}, \qquad j = 1, 2 \quad (25)$$

Furthermore, it can be shown, by computing $X(k)$ as in Eq. (2) and evaluating it at $k = l-1$ and $k = l+1$, that

$$\angle X_j(l+1) = \angle X_j(l-1) + \frac{2\pi}{N}, \qquad j = 1, 2 \quad (26)$$

We can thus express $X_j(l+1)$ as a function of r_j and i_j

$$\Re\{X_j(l+1)\} = Q_j \left(\cos\left(\frac{2\pi}{N}\right) r_j - \sin\left(\frac{2\pi}{N}\right) i_j\right) \quad (27)$$

$$\Im\{X_j(l+1)\} = Q_j \left(\sin\left(\frac{2\pi}{N}\right) r_j + \cos\left(\frac{2\pi}{N}\right) i_j\right) \quad (28)$$

The two additional equations to the system describe by Eq. (17) -

(20) are then

$$Q_1\left(C_2 r_1 - S_2 i_1\right) + Q_2\left(C_2 r_2 - S_2 i_2\right) - X_R(l+1) = 0 \tag{29}$$

$$Q_1\left(S_2 r_1 + C_2 i_1\right) + Q_2\left(S_2 r_2 + C_2 i_2\right) - X_I(l+1) = 0 \tag{30}$$

where $C_2 = \cos(2\pi/N)$ and $S_2 = \sin(2\pi/N)$.

By substituting Eq. (21 - 24) into Eq. (29) and (30) we obtain two equations that are only dependent on Δl_1 and Δl_2. Again, as happened for Eq. (9), attempts to solve them analytically (both using the Matlab Symbolic Math Toolbox and by hand) failed. To estimate Δl_1 and Δl_2, thus, a numerical solution has been used. The two implicit functions are evaluated in the intervals $0 \leq \Delta l_1 < 1$ and $0 \leq \Delta l_2 < 1$, and the points of intersection, corresponding to the two fractional frequencies, determined. Fig. 2 shows a typical example of the two curves described by Eq. (29) and (30). Two interesting observations can be made: there is a trivial solution ($\Delta l_1 = \Delta l_2$), which means that the two curves have an infinite number of intersections (the diagonal). The other solution (indicated by circles) is symmetrical with respect to the diagonal, since the two signals are interchangeable, the numbering being just a convention.

A simple grid search was again used to find the intersections of the two curves. Assuming that the functions were evaluated at M equally spaced points, the complexity of the numerical solution is roughly proportional to $M^2/2$ (because of the symmetry of the problem, only one half of the points is needed).

3. EVALUATION

In most situations, the method described in Sec. 2.3 and 2.4 gives an accurate solution to the problem of separating two overlapping partials, given that their frequencies fall both into the same frequency bin l. Unfortunately, there are few particular cases in which the system does not work.

When the two partials have the exact same frequency, their sum results in a single sinusoid with the same frequency as the two components. This makes the problem underdetermined, since $P_1 = P_2$ and $Q_1 = Q_2$, and thus Eq. (21) and (22) are equivalent to (23) and (24), except for a constant factor.

It can also happen that in a particular time frame, the transforms of the two partials (*i.e.* $X_1(l)$ and $X_2(l)$) have the same phase or opposite phases, even though the frequencies are different. This also makes the system under-defined. In this case the two curves do not intersect in the region where $\Delta l_1 \neq \Delta l_2$, as can be seen in Fig. 3, but only on the diagonal. To test the effect of this situation on the estimation error, the phase difference between the two transforms $\angle X_1(l) - \angle X_2(l)$ was systematically varied, while keeping the amplitudes and frequencies of the two sinusoids constant. The result is shown in Fig. 4. As can be observed, around 0, π and 2π, the error increases considerably. Tests with different combinations of Δl_1 and Δl_2 showed that the amount of this error was very unpredictable. To solve in part this problem, the estimated values from previous time frames were taken into consideration by using a smoothing filter (moving average). In Fig. 5, a signal composed of two sinusoids with constant Δl_1 and Δl_2 is displayed. The estimated values have been filtered using a moving average filter (10 frames). It can be seen that the error is nicely reduced. Notice that the occurrence of this error can be somehow

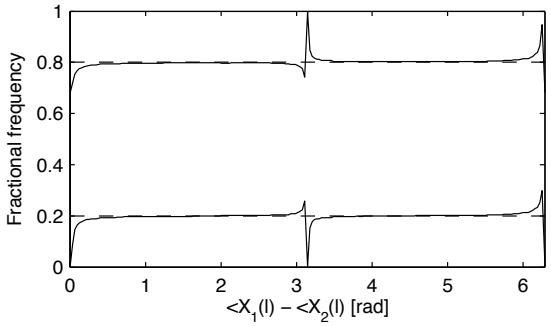

Figure 4: *Effect of the difference between $\angle X_1(l)$ and $\angle X_2(l)$ on the estimation accuracy.*

Table 1: *Estimation error (% of the bin width) as a function of the Signal-to-Noise ratio (best case scenario).*

SNR	Error (%)
50	0.24%
40	0.25%
30	0.37%
20	1.03%
10	4.22%
0	7.27%
-10	12.34%

predicted by looking at the phase difference between $X(l)$ and $X(l-1)$: the closer this difference is to 0 or π, the closer $\angle X_1(l)$ and $\angle X_2(l)$. This observation can be used to design an adaptive filter that corrects the error more consistently.

Another source of error is the numerical instability of $H(\omega)$ in the vicinity of $\Delta l = 0.5$. To study this problem, Δl_1 and Δl_2 were varied systematically between 0 and 1, and the mean estimation error computed and plotted in Fig. 6 (the phase difference between $X_1(l)$ and $X_2(l)$ was kept constant at $\pi/2$ in order to reduce the effect of the previously mentioned phase problem). The mean error was defined as

$$\overline{e}(\Delta l_1, \Delta l_2) = \frac{|\Delta l_1 - \overline{\Delta l_1}| + |\Delta l_2 - \overline{\Delta l_2}|}{2} \tag{31}$$

where Δl_j is the correct fractional frequency and $\overline{\Delta l_j}$ is the estimated one ($j = 1, 2$). It can be seen from Fig. 6 that the error increases appreciably only when $\Delta l_1 \approx \Delta l_2$

In reality, signals are hardly stationary and without noise. It this thus very important to test how well the method copes with these two factors. In Tab. 1, mean errors as defined in Eq. (31) are compared for different values of the Signal-to-Noise Ratio (SNR), when white noise was added to the signal. Down to SNR = 30 dB, there is no appreciable variation in the estimation error (the error remains below 1% of the bin width). After this point, the error starts increasing, and reaches the value of ∼10% for SNR= 0 dB. Again, the phase difference between $X_1(l)$ and $X_2(l)$ was held constant at $\pi/2$ (best estimation without noise). The sinusoids had parameters $\Delta l_1 = 0.4$, $a_1 = 0.8$, and $\Delta l_2 = 0.1$, $a_2 = 0.4$.

Finally, Fig. 7 shows, the result of the analysis of a signal

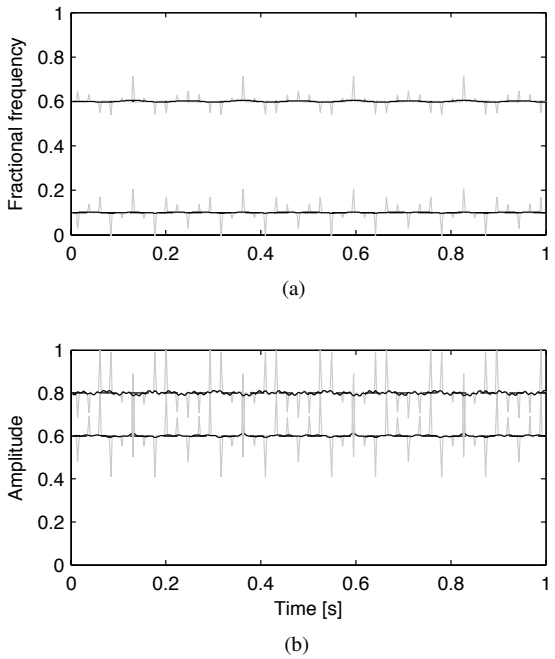

Figure 5: *Separation performed on a 1 second long signal composed of two sinusoids: one with constant fractional frequency* $\Delta l_1 = 0.6$ *and constant amplitude* $a_1 = 0.8$*, one with* $\Delta l_2 = 0.1$ *and amplitude* $a_2 = 0.6$*. In Fig. 5(a) the estimated (grey line) and moving average filtered (black line) fractional frequencies are displayed, together with the correct frequencies (dashed line). In Fig. 5(b), the amplitudes computed using Eq. (7) from the estimated frequencies (grey line) and from the moving average filtered frequencies (black line) are plotted.*

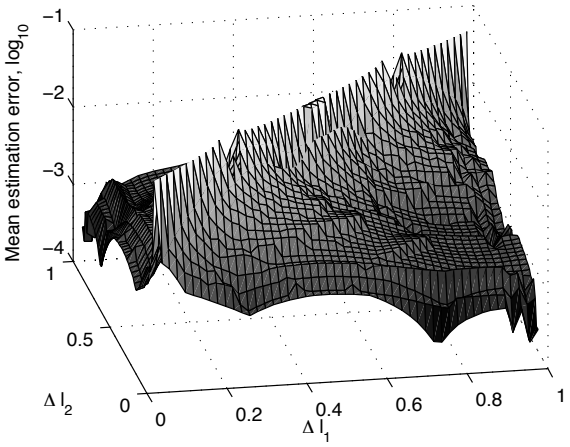

Figure 6: *Mean estimation error (Eq. (31)) for different combinations of* Δl_1 *and* Δl_2*, on a logarithmic scale (-1 corresponds to 10% error, -2 to 1%, and so on).*

with two non-stationary components (the carrier frequency of signal $x_1(n)$ was modulated with a frequency of $0.5\,Hz$, the amplitude of $x_2(n)$ was linearly varied over time). White noise (SNR = 20 dB) was also added to the signal. Using the moving average correction, the system was able to estimate the two sinusoids with relatively small errors. It was also observed that increasing the modulation frequency quickly reduces the performance of the method.

4. CONCLUSIONS AND FUTURE WORK

In this paper, a method to estimate the parameters (frequency, amplitude and phase) of a single, and of two overlapping sinusoidal components (*i.e.* closely spaced in frequency) was described. The method is based on the assumption that, if the component is stationary, its Fourier transform corresponds to the transform of the analysis window, centered around its frequency.

For a single component, the estimation was obtained by numerically solving a system of equations with three unknowns, based on the magnitude and phase of two frequency bins in the signal's transform. The maximum frequency estimation error was $0.05\,\%$ of the bin width. This is 20 times smaller than the error obtained using the technique described in [15], from which it derives. Furthermore, the present method can be used with any analysis window, whereas the previous method worked only with a sine window. The drawback with the method described here is the computational cost, which depends on the efficiency of the numerical evaluation of an implicit function.

For two components that have closely spaced frequencies (*i.e.* falling in the same frequency bin), a system of equations with six equations and six unknowns is solved numerically to estimate the parameters of the two components, this time using the magnitude and phase of the three frequency bins centered around the peak in the spectrogram. The proposed method represents an improvement over a similar method described in [9] because it removes its two main constraints, *e.g.* that the two components have appreciably different amplitudes, and appreciably spaced frequencies. Because

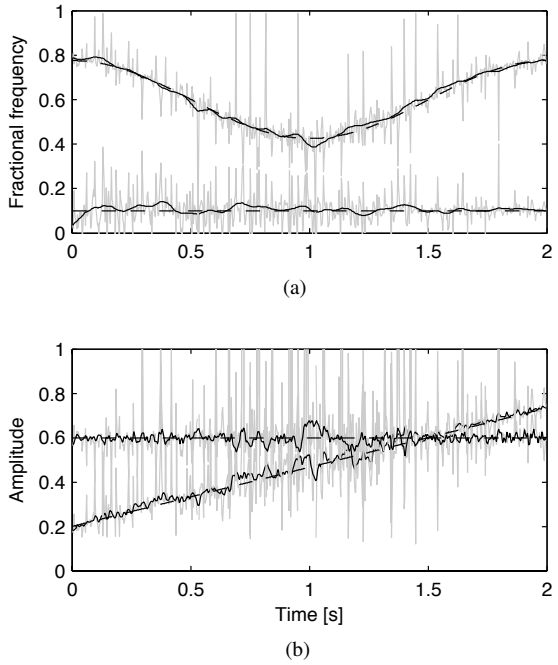

Figure 7: *Estimation performed on a 2 seconds long signal composed of two sinusoids, plus white noise with SNR = 20 dB. One component had a carrier fractional frequency of 0.6, modulated at 0.5 Hz, and a constant amplitude $a_1 = 0.6$, the other a constant fractional frequency $\Delta l_2 = 0.1$ and an amplitude a_2 increasing from 0.2 to 0.7. In Fig. 7(a) the estimated (grey line) and moving average filtered (black line) fractional frequencies are displayed, together with the correct frequencies (dashed line). In Fig. 7(b), the amplitudes computed using Eq. (7) from the estimated frequencies (grey line) and from the moving average filtered frequencies (black line) are plotted.*

of the few details given in [9], it was not possible to reproduce the previous algorithm for an objective comparison.

In Sec. 3, several sources of error were identified, and their effect evaluated with specific tests. Except for a few special cases, the present method gives very good separation of the two components, even when white noise is added, or the amplitude of the signals changes over time. Even if the frequency is moderately modulated (up to $1\ Hz$), the algorithm still performs rather well, as can be seen in Fig. 7. A moving average filter was used to smoothen the errors caused by particular combinations of phases and frequencies.

Regarding the performance of the algorithm, the grid approach to the numerical estimation of implicit functions is far from being optimal. Computational speed can be greatly improved by using optimization algorithms. Initial estimates can be gathered from previous time frames, helping the optimization to converge more quickly. An even better solution to the problem would be to find an analysis window that leads to an analytical solution to the implicit function. This will be further investigated in the future.

Future work will also be directed to the integration of the method into a larger system, described in [14], which aims at real-time expressive modifications of recorded music. The system requires source separation, which is done using a score-driven partials tracking method. The method described in the present paper will be used to separate the overlapping partials. Furthermore, the method will be extended to solve the case when the two frequencies fall into adjacent bins. Another area of improvement is the moving average filtering, which will be made more selective by using information about phase differences between different frequency bins as a measure of the reliability of the estimate. Finally, the system must be extensively tested on real instrument signals.

5. ACKNOWLEDGMENTS

This study was funded by the SAME project (FP7-ICT-STREP-215749), `http://sameproject.eu/`.

6. REFERENCES

[1] Yipeng Li, John Woodruff, and DeLiang Wang, "Monaural musical sound separation based on pitch and common amplitude modulation," *Trans. Audio, Speech and Lang. Proc.*, vol. 17, no. 7, pp. 1361–1371, 2009.

[2] D. L. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, Wiley/IEEE Press, Hoboken, NJ, 2006.

[3] G. J. Brown and M. P. Cook, "Perceptual grouping of musical sounds: a computational model," *J. New Music Res.*, vol. 23, pp. 107–132, 1994.

[4] Yipeng Li and DeLiang Wang, "Musical sound separation using pitch-based labeling and binary time-frequency masking," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 2008, pp. 173–176.

[5] M.A. Casey, "Separation of mixed audio sources by independent subspace analysis," in *Proc. of the International Computer Music Conference (ICMC00)*, Berlin (Germany), 2000.

[6] Paris Smaragdis and Judith C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY (USA), 2003.

[7] Tuomas Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparness criteria," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[8] M. Wright, J. Beauchamp, K. Fitz, X. Rodet, A. Robel, X. Sierra, and G. Wakefield, "Analysis/synthesis comparison," *Organized Sound*, vol. 5(3), pp. 173–189, 2000.

[9] Thomas W. Parsons, "Separation of speech from interfering speech by means of harmonic selection," *J. Acoust. Soc. Am.*, vol. 60, no. 4, pp. 911–918, 1976.

[10] T. Virtanen and A. Klapuri, "Separation of harmonic sound sources using sinusoidal modeling," in *Proc. of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP00)*, IEEE, Ed., Istanbul (Turkey), 2000.

[11] A.P. Klapuri, "Automatic transcription of music," in *Proc.of the Stockholm Music Acoustics Conference (SMAC03)*, Stockholm (Sweden), 2003.

[12] Mark R. Every, *Separation of Musical Sources and Structure from Single-Channel Polyphonic Recordings*, Ph.D. thesis, University of York, Department of Electronics, York (UK), February 2006.

[13] Mert Bay and James W. Beauchamp, "Harmonic source separation using prestored spectra," in *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA2006)*, Charleston, SC (USA), March 2006, Springer-Verlag Berlin Heidelberg.

[14] Marco Fabiani and Anders Friberg, "A prototype system for rule-based expressive modifications of audio recordings," in *Proceedings of ISPS 2007 (Int. Symp. of Performance Science 2007)*, Aaron Williamon and Daniela Coimbra, Eds., Porto, Portugal, November 2007, pp. 355–360, AEC (European Conservatories Association).

[15] Anibal J.S. Ferreira, "Accurate estimation in the ODFT domain of the frequency, phase and magnitude of stationary sinusoids," in *Proceedings of the IEEE Workshop in Applications of Signal Processing in Audio and Acoustics*, New Paltz, NY (USA), October 2001.

[16] Anibal J.S. Ferreira, "Combined spectral envelope normalization and subtraction of sinusoidal components in the ODFT and MDCT frequency domains," in *Proceedings of the IEEE Workshop in Application od Signal Processing to Audio and Acoustics*, New Paltz, NY (USA), October 2001.