Estimating Frequency, Amplitude and Phase of Two Sinusoids with Very Close Frequencies

Jayme G. A. Barbedo, Amauri Lopes

Abstract—This paper presents an algorithm to estimate the parameters of two closely spaced sinusoids, providing a frequency resolution that is more than 800 times greater than that obtained by using the Discrete Fourier Transform (DFT). The strategy uses a highly optimized grid search approach to accurately estimate frequency, amplitude and phase of both sinusoids, keeping at the same time the computational effort at reasonable levels. The proposed method has three main characteristics: 1) a high frequency resolution; 2) frequency, amplitude and phase are all estimated at once using one single package; 3) it does not rely on any statistical assumption or constraint. Potential applications to this strategy include the difficult task of resolving coincident partials of instruments in musical signals.

Keywords—Closely spaced sinusoids, high-resolution parameter estimation, optimized grid search.

I. INTRODUCTION

THE problem of estimating the parameters of sinusoidal signals has been intensively studied for many decades due to its importance in many practical situations. A few examples of applications are transient disturbance in power systems; channel prediction in communications; estimation of direction of arrival in radar and sonar; audio, speech and image processing; condition monitoring of engineering structures and systems; and nuclear magnetic resonance.

As a result of such a scenario, there is a huge amount of literature treating aspects ranging from the proposition of estimation methods to the development of performance bounds, analysis of accuracy and computational effort.

This paper deals with just one kind of sinusoidal estimation problem: the simultaneous estimation of amplitude, frequency and phase of two sinusoids with extremely close frequencies and using only one set of observations (just one snapshot). Although this particular problem also has significance in a variety of applications, the background motivation here is the complete identification of all sinusoids present in small pieces of musical signals, aiming applications such as the single channel source separation problem [1].

Signals produced by many musical instruments exhibit strong local periodicities modeled as a sum of harmonically related sinusoids. Since most songs are played by simultaneous harmonically related instruments, there may exist sinusoidal components almost coincident in frequency for a certain period of time. Such situation can occur in two main cases: 1) when two instruments play the same note, the fundamental frequency component and corresponding partials probably have very closely spaced frequencies; 2) if the instruments play harmonically related notes, then the nearly coincident frequencies can occur at some harmonic components. Exact spectral coincidence is highly unlikely due to different characteristics of instruments and musicians, and, if it occurs, it will last just for a few milliseconds. Therefore, any algorithm aiming to separate the signal of each instrument must deal with sinusoids with very close frequencies.

It is worth noting that estimating all three sinusoidal parameters is paramount for many musical signal processing tools, because the human perception of a song is usually closely linked to the way different sources and their respective partials interact, and such an interaction strongly depends on those parameters.

Even for this particular sinusoidal estimation problem there are several estimation methods which can be classified in a number of ways depending on their theoretical supports. One possible way to classify such methods into categories is presented next, together with some of related work: 1- correlation-based techniques [2-7]; 2- methods derived from the maximum likelihood function [8-15]; 3- methods based on rational models [4-6,16,17]; methods based on subspace properties [18-28]; algorithms that use spectral properties or filtering [5,13,14,17,29-36]; least squares-based methods [37,38] and so on. Many of those propositions also use iterative or adaptive procedures [7,8,30]. Finally, there are some papers that test and compare a variety of methods [39-42].

All those propositions have good performance for certain conditions in which their underlying assumptions hold, but they may fail when facing some specific conditions such as, for example, extremely closely spaced frequencies and/or limited number of observations.

In this context, this work presents the development of a new strategy to estimate the parameters of two sinusoids under the conditions aforementioned. The basic idea is to define a fine grid of points in the 6-dimensional parameter space (six parameters) and to compose a two-sinusoid mathematical model for each point. The choice of the best point is carried out by comparing the waveform corresponding to each model

Jayme Garcia Arnal Barbedo (corresponding author) and Amauri Lopes are with Department of Communication of the School of Electrical and Computer Engineering of the State University of Campinas (Unicamp), C.P. 6101, CEP: 13.083-970, Campinas - SP – Brazil (phone: +55-19-3521-3703; e-mails: {jgab,amauri}@decom.fee.unicamp.br).

with the observed data via least-square errors (LSE). Both the grid search and the LSE criterion were chosen because they are not restricted by any statistical assumption.

As the conventional grid search approach is usually very costly in computational terms, a number of sinusoidal properties and a carefully guided search were developed in order to make the algorithm almost one million times faster than the conventional unaltered grid search. The proposed grid search uses an intermediary step to estimate the best single-sinusoid approximation to the two-sinusoid target. The parameters of this optimum intermediary estimation determine an appropriate initial condition for the 6-dimensional grid search. Moreover, the single approximation is realized in a fast way using trained data, and a number of sinusoidal properties speed up the main grid search.

Although the proposed grid search is much faster than the conventional unaltered one, real-time implementations are still not possible given the average computational resources currently available, but such a relatively high computational effort is compensated by the great accuracy and broad applicability of the algorithm.

The proposal can be applied to a variety of practical problems. However, the scenario adopted here to present the technique is based in a future application in music processing. Accordingly, the signal sampling frequency is 48 kHz, the signal frequencies can vary from 50 Hz to 10 kHz and the target minimum frequency resolution is 0.05 Hz. A good time resolution for the intended application is achieved by using 25-ms analysis frames. This length is small enough to avoid excessive intra-frame sinusoidal parameter variations, which is important when dealing with signals whose parameters tend to fluctuate within a short time. It is also worth noting that applying a DFT to an excerpt of 25-ms length would result in a frequency resolution of about 40 Hz, which is 800 times worse than that desired here. The target amplitude resolution is 0.5% of the maximum amplitude (normalized to 1) and the target phase resolution is $\pi/200$ radians. The target resolutions together with the parameter ranges determine the grid of points in the 6-dimensional parameter space.

The performance of the proposed algorithm was assessed for a wide range of parameter values, achieving good accuracy, especially at low noise levels. The performance could not be compared with the Cramer-Rao lower bound (CRLB) and other estimation methods due to a number of reasons to be discussed in Section 3.

The paper is organized as follows. Section 2 describes the algorithm and the properties explored in order to get faster executions. Tests and results are presented in Section 3. Finally, Section 4 presents the conclusions and final remarks.

II. THE PROPOSED ALGORITHM

Figure 1 shows the general structure of the proposed algorithm, and each block is described in the following. As commented before, all procedures described from this point to the end of the paper will be applied over 25-ms analysis

frames. Although such a value is appropriate for most types of signals, it can be easily changed to meet any application characteristics.
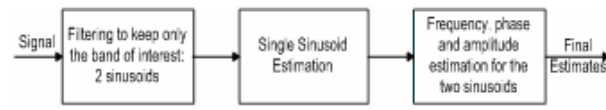

Fig. 1 Algorithm general structure.

A. Filtering

Because the proposed algorithm deals with only two sinusoids, an initial spectral selection must be done before starting the estimation procedures. There are a number of ways to determine the spectral band of interest. If the two target sinusoids are expected to be relatively isolated from other significant spectral components, very simple procedures can be adopted. However, if they are located in a populated area of the spectrum, more sophisticated procedures may be necessary. Since the main scope of this work is to provide a tool to discriminate very close sinusoids, and not to deal with tools to clean up the surrounding spectrum, it was assumed that the sinusoids of interest are at least f_s/N Hz apart from any other non-noise spectral component, where f_s is the sampling rate and N is the number of samples in a frame. Such an assumption allows using here a conventional band-pass filter and a simple procedure to determine the cut-off frequencies: a DFT is calculated, and the peaks provide a rough estimate of the central spectral location of the desired sinusoids (f_c); the filter cut-off frequencies are chosen as $f_c \pm f_s/N$. The filter used here is a second-order band-pass Butterworth. A narrower and sharper filter was avoided at this moment because the rough estimate provided by the DFT would not guarantee that the pass-band includes the sinusoids. More sophisticated procedures to improve the first frequency estimate and to determine the cut-off frequencies are one of the topics to be studied in a future research.

B. Grid Search Approach

Let the model for the observed data be

$$X(n) = A_{1x} \sin\left(\frac{2\pi f_{1x}n}{f_s} + \theta_{2x}\right) + A_{2x} \sin\left(\frac{2\pi f_{2x}n}{f_s} + \theta_{2x}\right) + r(n) \quad (1)$$

where A_{1x}, A_{2x}, f_{1x}, f_{2x}, θ_{1x} and θ_{2x} are the true sinusoid parameters to be estimated, f_s is the sampling frequency, $r(n)$ is additive Gaussian noise, $n = 1, 2, \ldots, N$, and N is the number of samples in a frame.

Let S be the sum of two closely spaced sinusoids as given by

$$S(n) = A_1 \cdot \sin\left(\frac{2\pi f_1 n}{f_s} + \theta_1\right) + A_2 \cdot \sin\left(\frac{2\pi f_2 n}{f_s} + \theta_2\right) , \quad (2)$$

whose parameters A_1, A_2, f_1, f_2, θ_1 and θ_2 must be adjusted to best fit $X(n)$, $n=0, 1, \ldots, N$, in the Least-Squares Error (LSE) sense.

The following parameter range constraints were adopted in

this work:

- A_1 and A_2 vary in the interval $(0 , 1]$ and the ratio $\bar{A} = A_2/A_1$ lies in the interval $[0.1 , 10]$;

- f_1 and f_2 vary in the interval $[50 , f_s/2]$;

- θ_1 and θ_2 vary in the interval $[0 , 2\pi)$.

The basic idea of the proposed algorithm is to estimate the parameters of S in (2) by means of a grid search. This kind of approach uses discrete values for all the parameters along with their respective ranges, and tests all the resulting points in the 6-dimensional parameter space to determine which one best satisfies the LSE optimization criterion. The criterion adopted here is: 1- the noise free version of $S(n)$ in (2) is calculated at each grid point; 2- the best grid point is the one that minimizes the quadratic error .

The 6-D grid space is composed of 480,000 frequency points, 400 phase points and about 200 amplitude points for each sinusoid. Such a number of points to be tested makes the conventional grid search approach prohibitive, and a strongly optimized search procedure must be developed. As will be seen along this section, the following measures were taken in order to reduce the computational complexity by a factor of one million: 1- an intermediary step based on the estimation of a single sinusoid was developed, which significantly narrows down the set of candidate points; 2- candidate testing is performed by using a smart grid search; 3- final refinements are introduced aiming some special situations. Next subsection describes the intermediary single sinusoid step.

C. Single Sinusoid Estimate

The optimized grid search is preceded by an intermediary step in which the algorithm determines the best single sinusoidal approximation to the available data in the LSE sense. As will be shown in the following, by adequately exploring the information provided by this optimum single sinusoid, the search for the best $S(n)$ can be greatly optimized. The underlying idea here is to deal with a simpler estimation problem, determining the parameters of just one sinusoid. The parameters of this best single sinusoidal approximation can be determined in a fast way, and provide a good starting point for the grid search associated to the two-sinusoid estimation problem.

In order to describe this intermediate step, consider $X(n)$ in Equation 1 for some given sinusoidal parameters values and $r(n) = 0$. Suppose that the corresponding $X(n)$ is scaled to a certain level (see Equation 2) and that a single sinusoid $S_v(n)$ is used to approximate $X(n)$ in the LSE sense. It is straightforward to verify that for each value of n the square error between $X(n)$ and the corresponding sample of any single sinusoidal approximation depends only on the frequency difference ($\Delta f_x = f_{2x} - f_{1x}$), phase difference ($\Delta\theta_x = \theta_{2x} - \theta_{1x}$), and amplitude ratio ($\bar{A}_x = A_{2x}/A_{1x}$) between the two sinusoids of $X(n)$. Therefore, the best LSE approximation S_v will also depend only on those three parameters. Moreover, those three parameters (Δf_x, $\Delta\theta_x$ and $\bar{A}_x$) also univocally determine the distance d in Hz between the frequency f_v associated to S_v and the mean frequency $f_m = (f_{1x} + f_{2x})/2$. Such

a distance will be positive if $f_v > f_m$, and negative otherwise. Therefore, the parameters of the optimum single sinusoid contain information about the parameters to be estimated and that information will speed up the grid search for the two-sinusoid problem.

In order to reduce the computational effort to estimate the single sinusoid, a kind of training procedure is carried out previously. The training consists in calculating and storing the LSE of such single sinusoidal approximation for a large number of possible combinations of frequency differences, phase differences and amplitude ratios in $X(n)$ in the absence of noise. The corresponding LSE are stored in a 3-D matrix **M**, which can be used later as a lookup table whose information can greatly narrow down the search space in the final part of the algorithm. The corresponding d values are also stored in the matrix **F**. It worth mentioning that the matrices **M** and **F** have to be determined just once for the parameter ranges employed here. If another application demands wider ranges, one just needs to determine the new values and add them to the matrices. Accordingly, if narrower ranges are enough, the size of the matrices can be reduced.

In this work, the normalized frequency difference ranges from 0 to $\pi/600$ (40 Hz for $f_s = 48$ kHz), with a $\pi/480000$ step (0.05 Hz for $f_s = 48$ kHz); the phase difference ranges from 0 to 2π with a $\pi/200$ step; and the amplitude ratio ranges from 0.1 to 10 with varying steps (smaller for small ratios). The resulting **M** matrix has a dimension of $800 \times 400 \times 500$. Such setting is only a suggestion and can be freely changed according to the desired application and available computational resources. The corresponding **F** matrix containing the distances between f_v and f_m is also stored; **F** has the same dimension of **M**. The way **M** and **F** are used is explained in Section 2.4.

After all these considerations, the intermediary estimation step can now be described from the beginning. After filtering the data to be analyzed, the algorithm performs a level scaling given by

$$\hat{X}(n) = \left(\frac{2}{N} \cdot \sum_{n=1}^{N} \left[X(n) \right]^2 \right)^{-0.5} \cdot X(n) , \qquad (3)$$

where is the scaled version of **X**. This scaling aims to eliminate the need for estimating the amplitude of the single sinusoid, which is made equal to 1. This is possible because the scaled amplitude matches the level of a single sinusoid with amplitude A = 1 (absolute power level = 0.5). Since the single sinusoid amplitude is fixed, only the frequency and phase of S_v have to be determined.

The easiest way to determine S_v (under an implementation point-of-view) would be the pure brute force approach, in which all possible frequencies and phases for the single sinusoid would be tested. However, a much faster and equally effective approach is possible by exploiting some characteristics of the search, as described in the following steps:

1) The candidate frequency f_t of the approximation sinusoid is forced to vary only from $f_c - f_s/1000$ to $f_c + f_s/1000$, with step

of $10^{-6} \times f_s$ Hz (f_c is the central spectral location determined by the initial filtering). Tests revealed that this range contains the actual fv in more than 99.9 % of the cases.

2) Initially, f_t is fixed to f_c - f_s/1000, and the phase is varied along its entire range. The phase θ_t associated to f_t that results in the lowest LSE is stored in the temporary vector $\mathbf{p}$, together with the corresponding LSE value itself; all of them are also stored in the best candidate vector $\mathbf{b}$.

3) The next f_t is considered, but instead of testing the entire phase range, only the phase values in the range $\theta_t - \pi/100$ to $\theta_t + \pi/100$ are tested. This is because the best phase does not vary much from one f_t to the next. However, if the step for f_t is greater than $10^{-4} \times f_s$ Hz, this rule tends to fail, and a phase range from $\theta_t - \pi/10$ to $\theta_t + \pi/10$ may be more appropriate. The vector p is then updated with the new values of f_t, θ_t and the corresponding LSE. If the current LSE is lower than the one in vector $\mathbf{b}$, such a vector is updated with the new values of f_t, θ_t and LSE.

4) The procedure in 3) is repeated for all possible f_t values. At the end, the vector $\mathbf{b}$ will have stored the frequency and phase of the best match sinusoid S_v, as well as the LSE value, which is essential for the main part of the algorithm, as described in the following.

D. Frequency, Phase and Amplitude Estimates for the Two Sinusoids

This subsection describes how the single sinusoid intermediary estimation is used to speed up the grid search for the estimation of the two sinusoids.

The first step to reduce the computational effort is narrowing down the search space by means of the LSE matrix $\mathbf{M}$, according to the following. Let m be the LSE value stored in vector $\mathbf{b}$. All elements for which Equation 4 holds are selected:

$$(1-\alpha) \cdot m < \mathbf{M}(i,j,k) < (1+\alpha) \cdot m , \qquad (4)$$

where

$$\begin{cases} \alpha = 8^{-m} & \text{if } m \leq 1, \\ \alpha = 0.125^{\left(\frac{(m-1)}{1000}+1\right)} & \text{if } m > 1. \end{cases} \qquad (5)$$

The values of α given in Equation 5 were determined experimentally in order to select as few elements of $\mathbf{M}$ as possible, but giving enough room to consider deviations due to noise. If only noise free signals were to be considered, the values of α could be considerably smaller, which would speed up the algorithm.

The indices i, j and k of each selected element of $\mathbf{M}$ have associated a Δf, a $\Delta \theta$ and a $\bar{A}$, respectively. Additionally, those indices provide the corresponding distance d in the F matrix. Those four values play a key role in the next steps of the algorithm, because they make possible to greatly reduce the dimensionality of the problem, as will be explained in the following.

From each selected $\mathbf{M}(i,j,k)$, one must obtain the best values of A_1, A_2, f_1, f_2, θ_1, θ_2 for S in Equation 2. Items a), b) and c) present three procedures to speed up this process. The

determination of the best among the selected $\mathbf{M}(i,j,k)$ will be described later.

a) Reducing the number of amplitude parameters to be estimated from 2 to 1

Instead of estimating one amplitude parameter for each sinusoid (A_1 and A_2 in Equation 2), it is possible to estimate only the ratio A_2/A_1. To do that, it suffices to set one of the amplitudes to one, and then scale the other one accordingly. Hence, Equation 2 becomes

$$S'(n) = \sin\left(\frac{2\pi f_1 n}{f_s} + \theta_1\right) + \bar{A} \cdot \sin\left(\frac{2\pi f_2 n}{f_s} + \theta_2\right) . \qquad (6)$$

After estimating the ratio $\bar{A} = A_2/A_1$ and the other parameters, the actual amplitudes A_1 and A_2 can be obtained according to

$$A_1 = \sqrt{P_o/P_r} , \quad A_2 = A_1 \cdot \bar{A} , \qquad (7)$$

where Po is the absolute power level of the original signal X as given by Equation 1, and P_r is the absolute power level of the estimated signal S' as given by Equation 6.

b) Reducing the number of phase parameters to be estimated from 2 to 1

Let S'' be a phase shifted version of S' given by

$$S''(n) = \sin\left(\frac{2\pi f_1 n}{f_s}\right) + A_r \cdot \sin\left(\frac{2\pi f_2 n}{f_s} + (\theta_2 - \theta_1)\right) . \qquad (8)$$

The waveform resulting from Equation 8 is identical to that resulting from Equation 6, except for a shift in the phase. Estimating only the difference $\Delta \theta = \theta_1 - \theta_2$, and then applying a compensation to match the phase of the target signal, is a much faster procedure than trying to estimate θ_1 and θ_2 separately.

The compensation can be done simply by determining the phase shift that minimizes the LSE between S'' and the target signal. A way to do that would be carrying out an exhaustive search in which several phase shifts between 0 and 2π would be tested. This approach, although effective, is computationally inefficient. A much faster procedure is presented in the following.

It can be shown that the LSE between a sinusoid and a phase shifted version of that same sinusoid depends only on the power level and on the phase displacement in radians. Since the sum of two closely spaced sinusoids behaves similarly to a single sinusoid, the LSE values obtained for a single sinusoid hold almost perfectly when considering a sum of two sinusoids. Additionally, the LSE values vary linearly with the power level of the signal, making it possible to store the values for a given power level, and then properly scale those stored values when applying them to other signals. In this context, a vector $\mathbf{v}$ containing LSE values for phase shifts ranging from 0 to π and step size of $\pi/50000$, was stored for a sinusoid with an absolute power level of 0.5. Figure 2 shows how the LSE values vary with the phase shift for that power level in a frame with 1200 samples.

As commented before, the vector $\mathbf{v}$ has to be scaled according to the level of the signal being considered. For example, if the signal has an absolute power level of 0.25,

vector $\mathbf{v}$ has to be multiplied by $0.25/0.5 = 0.5$. After that, the LSE between the original and estimated signals is calculated and the corresponding element v in vector $\mathbf{v}$ is identified. The shift to be applied to the estimated signal will be either $v \cdot \pi/50000$ if the first peak of such a signal is after the first peak of the original signal, or $- v \cdot \pi/50000$ if the first peak of such a signal is before the first peak of the original signal. Such a shift is a good approximation for the value of $\theta 1$ in Equation 1, and then $\theta 2 = \Delta\theta + \theta 1$.
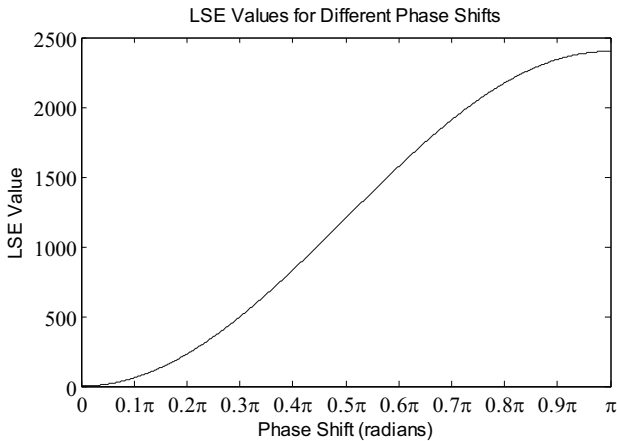

Fig. 2 LSE values for different phase shifts.

c) Reducing the number of degrees of freedom in the frequency estimation

The optimum single sinusoid estimation simplifies significantly the estimation of the frequencies of the two sinusoids. To understand this possibility, note that each element resulting from the search performed in matrix $\mathbf{M}$ (Equations 4 and 5) fixes a distance in Hz between both frequencies. Additionally, each element in matrix $\mathbf{M}$ will have associated a distance d in matrix $\mathbf{F}$ that reveals the probable distance in Hz between the estimated frequency f_v for the single sinusoid and the mean frequency f_m (see Section 3.3). Therefore, the value of each frequency to be tested will be

$$f_1 = f_v + d - 0.5 \cdot \Delta f_i \quad , \qquad (9)$$

$$f_2 = f_1 + \Delta f_i \quad , \qquad (10)$$

where Δf_i is the distance Hz between the sinusoids as given by the index i of the element in matrix $\mathbf{M}$ currently being considered and $d = \mathbf{F}(i,j,k)$. The result of this strategy is that the number of f_1 and f_2 combinations to be tested will be equal to the number of selected elements in matrix $\mathbf{M}$.

d) Grouping the procedures described in a), b) and c) into an effective estimate strategy

The final steps for the algorithm to estimate the frequencies, phases and amplitudes of the sinusoids can be summarized as follows:

1- After estimating the parameters of the single sinusoid, the elements in matrix $\mathbf{M}$ are selected. Each element will have four values associated: a frequency difference Δf_i, a phase difference $\Delta\theta_j$, an amplitude ratio $\bar{A}_k$, and a distance d given

by $\mathbf{F}(i,j,k)$, where i, j, k are the indices corresponding to each dimension of matrix $\mathbf{M}$. The selected elements roughly define a hexahedron whose dimensions are proportional to the tolerance α (see Equations 4 and 5).

2- Instead of testing all possible Δf_i at once, a few rules are applied. At first, only Δf_a is considered, where a is the lowest index i among the elements selected from $\mathbf{M}$ (usually a = 1), and all phase differences $\Delta\theta$ and all amplitude ratios $\bar{A}$ associated to this particular frequency difference will be considered according to the search performed over matrix $\mathbf{M}$. In other words, i is fixed to a given value (in this case a), and then all possible combinations of j and k associated to Δf_a that satisfy the search criteria given by Equation 4 are considered.

3- The candidate sum of sinusoids C_{ijk} for each i (in this case fixed to a), j, and k is then generated according to Equations 8 to 10.

4- The level of each C_{ijk} is then matched to the level of the target signal S_t.

5- The phases of C_{ijk} and S_t are aligned according to the procedure described in item b) of this section.

6- The LSEs between each C_{ijk} and X are calculated.

7- If the a LSE is lower than any other observed so far, the value of the LSE itself, together with the values of A_1, A_2, f_1, f_2, θ_1 and θ_2 are stored in the best estimate vector $\mathbf{e}$.

8- If the lowest LSE achieved for this first Δf_a is smaller than min(0.01 , $0.01 \cdot m$), it is considered that Δf_a is either the frequency difference Δf_r that leads to the smallest LSE value, or it is relatively close to Δf_r. If this is the case, the algorithm will skip to step 10 and continue the search for the minimum LSE value considering only the neighborhood of Δf_a; the so-called reference index that defines such a search neighborhood will be, in this case, $i_r = a$. If the condition is not satisfied, the search for i_r continues, as described in step 9.

9- If the condition in step 8 is not satisfied, a new Δf_i must be considered. Testing all possible values of i sequentially until the condition is satisfied would be computationally inefficient. Instead, only a number of key points following a are tested: [400, 800, 200, 600, 100, 300, 500, 700, 50, 150, 250, 350, 450, 550, 650, 750], in that order. If after all those values of i have been considered the condition is still not satisfied, i_r assumes the value of the index i that resulted in the lowest LSE.

The next steps of the algorithm depend on two variables (p_1 and p_2) whose values vary according to the value of m. The values that the variables can assume are:

- if $m < 0.2$, $p_1 = 50$ and $p_2 = 10$;
- if $0.2 \le m < 1$, $p_1 = 25$ and $p_2 = 15$;
- if $1 \le m < 5$, $p_1 = 15$ and $p_2 = 15$;
- if $m \ge 5$, $p_1 = 10$ and $p_2 = 20$.

The values of p_1 and p_2 constraint the next steps of the algorithm, and were chosen in such a way that the search is broader when m is small and narrower when m is large. This distinction is necessary because when m has small values, the sum of the two sinusoids behave almost as if only one sinusoid was present. This normally happens when the sinusoids amplitudes are very different and/or the phases and

frequencies are very close. Such a situation is very difficult to be dealt with, demanding a broader search in order to improve the chances of a good estimate.

In the next steps of the algorithm, the elements of vector e will be updated every time the new LSE value is smaller than that stored in the vector.

10- i_r (the final reference index obtained from either step 8 or step 9) determines the point where the final search procedure begins. Beginning with i_r, a forward search is performed by considering all Δf_i for which i is larger than i_r. The search stops either if $c - l = p_2$, where c is the index of the Δf currently being considered and l is the index of the last Δf for which the LSE has improved, or if $i = 800$ (last index). The same procedure is repeated backwards. If $i_r = 1$, only the forward search is performed, and if $i_r = 800$, only the backward search is carried out.

11- As commented before, each Δf_i has a number of phase differences $\Delta \theta_j$ and amplitude ratios $\bar{A}_k$ associated, according to the j and k indices resulting from the search over matrix **M**. Let j_t and k_t be the indices to be considered in a given moment. The rule is simply to consider all phase differences and amplitude ratios whose indices are between $t - p_1$ and $t + p_1$, where t is the index of the phase that resulted in the smaller LSE for the previous index i.

12- The last values stored in vector e reveal all parameters of the estimated sinusoids and the LSE with respect to the target signal.

Next section presents the results achieved by the algorithm.

III. TESTS AND RESULTS

One of the most common procedures to assess the performance of an estimation algorithm is the comparison with the Cramer-Rao Lower Bound (CRLB) [43]. However, a number of reasons prevent that such a comparison be a good option in this particular case. First, the discrete values of the parameters used by the proposed algorithm introduce a quantization error that is not modeled by the Cramer-Rao procedure. The average errors produced by simulation procedures depend on the parameter values and can vary from zero to a multiple of the half interval between adjacent discrete levels. This behavior impairs the comparison in situations with very low errors, for example, when the SNR and/or the number of samples are very high. Hence, it would be necessary to calculate an additional average of the errors, now varying the values of the parameters inside the range between two discrete levels. However, even this additional process would not eliminate the differences between the assumptions of the CRLB and those of the algorithm.

The second reason is related to some numerical problems to calculate the Cramer-Rao bound. The proposed algorithm is intended to operate at critical conditions of resolution, and it was not possible to achieve good Cramer-Rao numerical values for frequency differences close to 0.1 Hz using only 1,200 signal samples.

Due to those characteristics the performance tests were developed trying to simulate typical conditions expected to be found in the audio context.

The tests to validate the algorithm were performed using one thousand 25-ms excerpts sampled at 48 kHz, each one consisting of the sum of two sinusoids spaced by a minimum of 0.1 Hz and a maximum of 40 Hz. The frequencies, amplitudes and phases of each sinusoid were randomly generated, with the amplitude ratio between the strongest and weakest sinusoids varying from 1:1 to 10:1, the frequencies varying between 50 Hz and 10 kHz, and the phases varying between 0 and 2π. The noisy conditions were simulated by adding white Gaussian noise to each excerpt. Real world signals were avoided because their intrinsic variability makes it very difficult to determine which would be the target to be pursued by the algorithm, preventing a precise performance measurement.

Table 1 shows the root mean square errors (RMSE) between the estimates and the targets for the noise free case. The frequency error is given in Hz, the amplitude error is given in percentage of the target amplitude in order to assign equal importance to small and large amplitudes, and the phase error is given in radians.

TABLE I
MEAN RMSE BETWEEN ESTIMATES AND TARGET SIGNAL (NOISE FREE)

Frequency Band	Frequency (Hz)	Amplitude (%)	Phase (rads)
50 Hz – 10 kHz	0.25	2.16	0.053
50 – 200 Hz	0.38	3.16	0.058
200 – 500 Hz	0.28	2.37	0.066
500 – 1000 Hz	0.22	2.08	0.033
1 – 10 kHz	0.25	2.01	0.060

The first row in the table shows the overall errors, while the remaining ones show the results for certain sinusoid frequency ranges. As can be seen, the error levels for noise free signals are very low, especially taking into account the wide range of frequencies and amplitudes that are being considered. It is also useful to analyze the impact of the amplitude ratio in the mean errors. Table 2 shows the mean errors for different amplitude ratios (high ÷ low amplitudes) in the absence of noise.

As can be seen, the amplitude ratio seems to have nearly no impact in the overall accuracy for noise free signals, meaning that the errors have the same order no matter the amplitude ratio. However, when noise is added to the signals, the amplitude ratio between the sinusoids plays a central role, as can be seen in Table 3. The same behavior is observed when analyzing the impact of the distance in Hz between the sinusoids.

TABLE II
MEAN ERRORS FOR DIFFERENT AMPLITUDE RATIOS (NOISE FREE)

Amplitude Ratio	Frequency (Hz)	Amplitude (%)	Phase (rads)
1 - 2	0.25	2.25	0.030
2 - 5	0.21	2.15	0.062
5 - 10	0.31	2.07	0.066

TABLE III
MEAN ERRORS FOR DIFFERENT AMPLITUDE RATIOS (SNR OF 30 DB)

Amplitude Ratio	Frequency (Hz)	Amplitude (%)	Phase (ras)
1 - 2	1.51	15.78	0.304
2 - 5	2.02	22.01	0.388
5 - 10	3.01	79.86	0.536

It is also important to analyze the impact of noise in the overall performance of the algorithm. Table 4 shows the mean errors for a number of different SNRs.

TABLE IV
MEAN ERRORS FOR DIFFERENT SNRS

SNR (dB)	Frequency (Hz)	Amplitude (%)	Phase (rads)
>60	0.26	3.01	0.060
50	0.40	7.19	0.111
40	0.83	26.9	0.269
30	2.16	37.1	0.406
20	6.11	90.8	0.972
10	8.15	65.6	0.981
0	9.03	70.7	1.022

As can be seen in Table 4, the estimates start to degrade more rapidly for SNRs below 50 dB. It is worth noting that the frequency errors with respect to the absolute values of the sinusoid frequencies are small even for low SNRs. This is because the single sinusoid estimation presented in Section 2.3 is very robust and provides a good starting point even under severely noisy conditions. However, if the two sinusoids are very closely spaced, their frequency difference cannot be suitably estimated under such noisy conditions.

A direct comparison of this algorithm with its predecessors is difficult because their characteristics are actually distinct. The main purpose of this method is to simultaneously estimate all parameters for both sinusoids, while most of the other strategies are interested in estimating one of the parameters at a time. Also, most of them are based on some statistical assumptions, whereas the proposed algorithm is not restricted by any statistical assumption. As a consequence, the computational effort demanded by the algorithm is higher than that demanded by most of the other methods, and can vary significantly due to the rules and conditions presented in Section 3.4. The algorithm, implemented in Matlab® and running in a computer with Intel Centrino Duo processor of 1.83 GHz, has taken, in average, 10 seconds to process each 25-ms frame. Such time is certainly excessive to allow real-time implementations even using a more efficient programming language, but it is remarkably low for a grid-search approach, allowing a relatively fast offline processing of an audio signal. Additionally, future versions of the algorithm for this application are expected to include mechanisms to explore the information extracted from frames previously processed, speeding up the program.

IV. CONCLUSION

This paper presented an algorithm to estimate the frequencies, amplitudes and phases of two closely spaced sinusoids. It uses a highly optimized grid-search approach to provide high-resolution estimates while keeping the computational effort at relatively low levels. Its capability of resolving sinusoids with extremely close frequencies makes this algorithm ideal to be used in the musical signals context, whose simultaneous instruments often generate very close harmonics that are difficult to be resolved.

Future versions of the algorithm are expected to include several improvements. A strategy able to use the information of frames previously processed in order to speed up the estimates is expected to be included in the next versions of the algorithm, as well as a more efficient strategy to determine the spectral band of interest before the parameters estimation. There are also plans to use the algorithm as part of a monophonic audio source separation tool currently being designed.

REFERENCES

[1] L. Benaroya, F. Bimbot, R. Gribonval, "Audio source separation with a single sensor," *IEEE Tran. on Audio, Speech, and Language Proc.*, vol. 14, no. 1, pp. 191-199, Jan. 2006.
[2] S. Kay and S. L. Marple, "Spectrum analysis- A modern perspective," *Proc. IEEE*, vol. 9, pp. 1380-1419, Nov. 1981.
[3] *Proceedings of IEEE, Special Issue on Spectral Estimation*, vol. 70, Sept. 1982.
[4] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall Signal Processing Series, Englewood Cliffs, NJ, USA, 1988.
[5] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall Signal Processing Series, Englewood Cliffs, NJ, USA, 1987.
[6] P. Stoica, R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.
[7] K. Kitsios, A. Spanias, B. Welfert, P. Loizou, "An Adaptive Modified Covariance Algorithm For Spectral Analysis," in *Proc. of 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 56-59, Corfu, Greece, Jun 1996.
[8] [8] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-34, n. 5, pp. 1081-1089, Oct. 1986.
[9] P. Stoica, R. L. Moses, B. Friedlander and T. Söderström, "Maximum Likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 37, n. 3, pp. 378-391, March 1989.
[10] P. Stoica and C. Sharman "Novel eigenanalysis method for direction estimation," *IEE Proceedings*, vol. 137, No. 1, pp. 19-26, Feb. 1990.
[11] D. C. Rife, R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observation," *Bell System Technical Journal*, vol. 55, no. 9, pp. 1389 - 1410, Nov. 1976.
[12] P. Stoica, P. Händel, T. Söderström, "Approximate maximum likelihood frequency estimation," *Automatica*, vol. 30, no. 1, pp. 131 - 145, January 1994.
[13] M. D. Macleod, "Joint detection and high resolution ML estimation of multiple sinusoids in noise," in *Proc. of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 5, pp. 3065 - 3068, Salt Lake City, May 2001.
[14] J.-K. Hwang, Y.-C. Chen, "Superresolution Frequency Estimation by Alternating Notch Periodogram," *IEEE Transaction on Signal Processing*, vol. 41, no. 2, pp. 727 - 741, February 1993.

[15] P. Stoica, K.C. Sharman, "Maximum Likelihood Methods for Direction-of-Arrival Estimation", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 38, N. 7 pp. 1132-1143, July 1990.

[16] J. P. Burg, "Maximum Entropy Spectral Analysis," in *Proceedings of the 37th Annual International Meeting of the Society Exploration Geophysicists*, Oklahoma City, USA, 1967.

[17] P. Depalle, T. Helie, "Extraction of spectral peak parameters using a short-time Fourier transform modeling and no sidelobe windows," in *Proc.of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 19 - 22, New Paltz, USA, October 1997.

[18] D. Tufts and R. Kumaresan, "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood", *Proceedings of the IEEE*, vol. 70, No. 9, pp. 975-989, Sept. 1982.

[19] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, Stanford, CA, 1981.

[20] H. Wang and M. Kaveh, "On the Performance of Signal-Subspace Processing-Part I: Narrow-Band Systems", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 34, N. 5 pp. 1201- 1209, Oct. 1986.

[21] J. A. Cadzow, "Multiple Source Location-The Signal Subspace Approach", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 38, N. 7 pp. 1110-1125, July 1990.

[22] R. Roy, T. Kailath, "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, July 1989.

[23] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting", *IEEE Trans. on Signal Processing*, vol. 39, No. 5, pp. 1110-1121, May. 1991.

[24] M. Viberg, B. Ottersten and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting", *IEEE Trans. on Signal Processing*, vol. 39, No. 11, pp. 2436-2449, Nov. 1991.

[25] D. M. Wilkes, J. A. Cadzow, "The effects of phase on high-resolution frequency estimators," *IEEE Transaction on Signal Processing*, vol. 41, no. 3, pp. 1319 - 1330, March 1993.

[26] G. Feyh, "Inverse Eigenvalue Problem for Sinusoidal Frequency Estimation," in *Proc. of 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 3985 - 3988, Munich, Germany, April 1997.

[27] F.-X. Ge, H. Zhang, J. Yang, Y.-N. Peng, "Super-resolution frequency estimation of the sinusoidal signals with unknown lowpass envelopes," in *Proc. of 2003 IEEE Radar Conference*, pp. 273 - 273, Huntsville, USA, May 2003.

[28] M. G. Christensen, A. Jakobsson, S. H. Jensen, "Joint High-Resolution Fundamental Frequency and Order Estimation," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 5, pp. 1635 - 1644, July 2007.

[29] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, Vol. 57, pp. 1408 - 1418, August 1969.

[30] S. Nishimura, "Adaptive detection and enhancement of closely spaced sinusoids using multirate techniques," in *Proc. of 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 400 - 403, Minneapolis, USA, April 1993.

[31] C. I. Byrnes, T. T. Georgiou, A. Lindquist, "A New Approach to Spectral Estimation: A Tunable High-Resolution Spectral Estimator," *IEEE Transactions on Signal Processing*, Vol. 48, no. 11, pp. 3189 - 3205, November 2000.

[32] C. B. Lambrecht, M. Karrakchou, "Wavelet packets-based high-resolution spectral estimation," *Signal Processing*, vol. 47, no. 2, pp. 135 - 144, November 1995.

[33] H.-T. Li, P. M. Djuric, "A novel approach to detection of closely spaced sinusoids," *Signal Processing*, Vol. 51, No. 2, pp. 93 - 104, June 1996.

[34] P. O'Shea, "A High-Resolution Spectral Analysis Algorithm for Power-System Disturbance Monitoring," *IEEE Transactions on Power Systems*, Vol. 17, no. 3, August 2002.

[35] S. Hainsworth, M. Macleod, "On Sinusoidal Parameter Estimation," in *Proc. of the 6th International Conference on Digital Audio Effects*, London, UK, September 2003.

[36] I. Sarkar, A. T. Fam, "The interlaced chirp Z transform," *Signal Processing*, vol. 86, no. 9, pp. 2221 - 2232, September 2006.

[37] O. Besson, P. Stoica, "Nonlinear least-squares approach to frequency estimation and detection for sinusoidal signals with arbitrary envelope," *Digital Signal Processing*, vol. 9, no. 1, pp. 45-56, January 1999.

[38] T. Tolonen "Methods for Separation of Harmonic Sound. Sources using Sinusoidal Modeling," presented at the 106th AES Convention, paper 4958, Munich, Germany, 1999.

[39] R. H. Roy, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," Ph.D. dissertation, Stanford University, Stanford, CA, 1987.

[40] P. Stoica, T. Soderstrom, "Statistical analysis of MUSIC and ESPRIT estimates of sinusoidal frequencies," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3273 - 3276, Toronto, Canada, April 1991.

[41] O. Besson, P. Stoica, "Analysis of MUSIC and ESPRIT frequency estimates for sinusoidal signals with lowpass envelopes," *IEEE Transactions on Signal Processing*, vo1.44, no. 9, pp. 2359-2364, September 1996.

[42] P. Stoica, H. Li, J. Li, "Amplitude estimation of sinusoidal signals: survey, new results, and an application," *IEEE Transactions on Signal Processing*, Vol. 48, No. 2, pp. 338 - 352, February 2000.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing – Estimation Theory*, Prentice-Hall Signal Processing Series, Englewood Cliffs, NJ, USA, 1993.

Jayme Garcia Arnal Barbedo Jayme Garcia Arnal Barbedo received a B.Sc. degree in electrical engineering from the Federal University of Mato Grosso do Sul, Brazil, in 1998, and M.Sc. and Ph.D. degrees for research on the objective assessment of speech and audio quality from the State University of Campinas, Brazil, in 2001 and 2004, respectively.

From 2004 to 2005 he worked with the Source Signals Encoding Group of the Digital Television Division at CPqD Telecom & IT Solutions, Campinas, Brazil. From 2006 to 2007 he was with the School of Engineering and Applied Sciences at Harvard University, Cambridge, MA, as a postdoctoral fellow. Since 2007 he has been with the Department of Communications at the School of Electrical and Computer Engineering of the State University of Campinas as a researcher, conducting postdoctoral studies in the areas of content-based audio signal classification, automatic music transcription, and audio source separation. His interests also include audio and video encoding applied to digital television broadcasting and other digital signal processing areas.

Amauri Lopes received B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the State University of Campinas, Brazil, in 1972, 1974, and 1982, respectively.

He has been with the Electrical and Computer Engineering School (FEEC) at the State University of Campinas since 1973, where he has served as chairman in the Department of Communications, vice dean of the Electrical and Computer Engineering School, and currently is a professor. His teaching and research interests include analog and digital signal processing, circuit theory, digital communications, and stochastic processes.

Dr. Lopes has published over 100 refereed papers in some of these areas and over 30 technical reports about the development of telecommunications equipment.