EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Blind separation of overlapping partials in harmonic musical notes using amplitude and phase reconstruction

Jesús Ponce de León* and José Ramón Beltrán

Abstract

In this study, a new method of blind audio source separation (BASS) of monaural musical harmonic notes is presented. The input (mixed notes) signal is processed using a flexible analysis and synthesis algorithm (complex wavelet additive synthesis, CWAS), which is based on the complex continuous wavelet transform. When the harmonics from two or more sources overlap in a certain frequency band (or group of bands), a new technique based on amplitude similarity criteria is used to obtain an approximation to the original partial information. The aim is to show that the CWAS algorithm can be a powerful tool in BASS. Compared with other existing techniques, the main advantages of the proposed algorithm are its accuracy in the instantaneous phase estimation, its synthesis capability and that the only input information needed is the mixed signal itself. A set of synthetically mixed monaural isolated notes have been analyzed using this method, in eight different experiments: the same instrument playing two notes within the same octave and two harmonically related notes (5th and 12th intervals), two different musical instruments playing 5th and 12th intervals, two different instruments playing non-harmonic notes, major and minor chords played by the same musical instrument, three different instruments playing non-harmonically related notes and finally the mixture of a inharmonic instrument (piano) and one harmonic instrument. The results obtained show the strength of the technique.

Introduction

Blind audio source separation (BASS) has been receiving increasing attention in recent years. The BASS techniques try to recover source signals from a mixture, when the mixing process is unknown. *Blind* means that very little information is needed to carry out the separation, although it is in fact absolutely necessary to make assumptions about the statistical nature of the sources or the mixing process itself.

In the most general case, see Figure 1, separation will deal with N sources and M mixtures (microphones). The number of mixtures defines each particular case, and for each situation the literature provides several methods of separation. Probably due to the absence of interesting problems in the over-determined case, which has properly been solved, the most extensively studied case is undetermined separation, where $N > M$ ($N > M$ does not always imply poorer results). For example, in stereo separation (through the DUET algorithm [1] and other time-frequency masking evolutions [2-4]), the delay and attenuation between the left and right channel information can be used to discriminate the sources present and some kind of *scene* situation [5].

In other applications, when a monaural solution is needed (i.e., when $M = 1$), the mathematical indetermination of the mixture significantly increases the difficulties of the task. Hence, monaural separation is probably the most difficult challenge for BASS, but even in this case, the human auditory system itself can somehow segregate the acoustic signal into separate streams [6]. Several techniques for solving the BASS problem in general (and the monaural separation in particular) have been developed.

Psychoacoustic studies, such as computational auditory scene analysis [7,8], inspired by auditory scene analysis [6], attempts to explain the mentioned capability of the human auditory system in selective attention. Psychoacoustic also suggests that temporal and spectral coherence between sources can be used to discriminate between them [9].

*Correspondence: jponce@unizar.es
Department of Electronic Engineering and Communications, Universidad de Zaragoza, María de Luna 1, 50018 Zaragoza, Spain

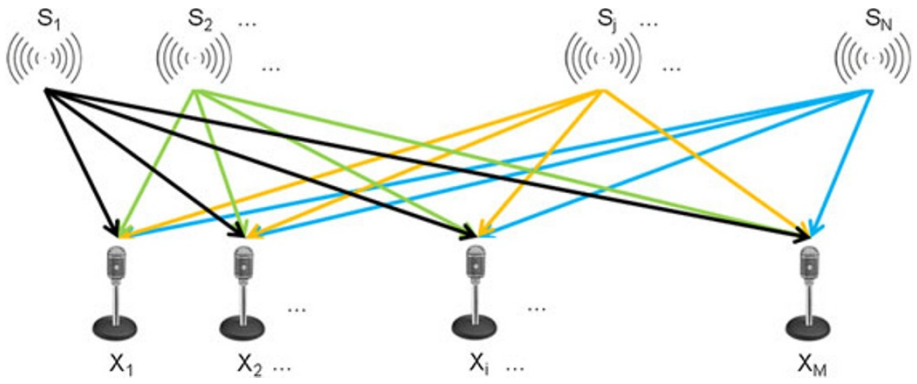

Figure 1 The general BASS task, with *N* mixed sources and *M* sensors.

Within the statistical techniques, independent component analysis (ICA) [10,11] assumes statistical independence among sources, while independent subspace analysis [12] extends ICA to single-channel source separation. Sparse decomposition [13] assumes that a source is a weighted sum of bases from an overcomplete set, considering that most of these bases are inactive most of the time [14], that is, their relative weights are presumed to be mostly zero. Non-negative matrix factorization [15,16] attempts to find a mixing matrix (with sparse weights [17,18]) and a source matrix with non-negative elements so that the reconstruction error is minimized.

Finally, sinusoidal modeling techniques assume that every sound is a linear combination of sinusoids (partials) with time-varying frequencies, amplitudes, and phases. Therefore, sound separation requires a reliable estimation of these parameters for each source present in the mixture [19-21], or some *a priori* knowledge, i.e., rough pitch estimates of each source [22,23]. One of the most important applications is monaural speech enhancement and separation [24]. These are generally based on some analysis of speech or interference and subsequent speech amplification or noise reduction. Most authors have used STFT to analyze the mixed signal in order to obtain its main sinusoidal components or partials. Auditory-based representations [25] can also be used.

One of the most important and difficult problems to solve in the separation of pitched musical sounds is overlapping harmonics, that is, when frequencies of two harmonics are approximately the same. The problem of overlapping harmonics has been studied during the past decades [26], but it is only in recent years that there has been a significant increase in research on this topic. Given that the information in overlapped regions is unreliable, several recent systems have attempted to utilize the information from neighboring non-overlapped harmonics. Some systems assume that the spectral envelope of the instrument sounds is smooth [27-29]; hence, the amplitude of an overlapped harmonic can be estimated from the amplitudes of non-overlapped harmonics from the same source, via weighted sum [20], or interpolation [21,27]. The spectral smoothness approximation is often violated in real instrument recordings. A different approximation is known as the common amplitude modulation (CAM) [22], which assumes that the amplitude envelopes of different harmonics from the same source tend to be similar. The authors of [30] propose an alternate technique for harmonic envelope estimation, called harmonic temporal envelope similarity (HTES). They use the information from the non-overlapped harmonics of notes of a given instrument, wherever they occur in a recording, to create a model of the instrument which can be used to reconstruct the harmonic envelopes for overlapped harmonics, allowing separation of completely overlapped notes. Another option is the average harmonic structure (AHS) model [31] which, given the number of sources, creates a harmonic structure model for each present source, using these models to separate notes showing overlapping harmonics.

In this study, we use an experimentally less restrictive version of the CAM assumption within a sinusoidal model generated using a complex band pass filtering of the signal. Non-overlapping harmonics are obtained using a binary masking approach obtained from the complex wavelet additive synthesis (CWAS) algorithm [32], which is based on the complex continuous wavelet transform (CCWT). The main advantage of the proposed technique is the capability of synthesis of the CWAS algorithm. Using the CWAS wavelet coefficients, it is possible to synthesize an output signal which differs negligibly (numerically and acoustically) from the original input signal. Hence, the non-overlapped partials can be obtained with accuracy. The separated amplitudes of overlapping harmonics are reconstructed proportionally from the non-overlapping harmonics, following energy criteria in a least-squares framework. This way, it is possible to relax the phase

restrictions, and the instantaneous phase for each overlapping source can also be constructed from the phase of non-overlapping partials. At its current stage, the proposed technique can be used to separate two or more musical instruments, each one playing a single note.

The rest of the article is divided as follows. "Complex bandpass filtering" section provides a brief introduction to the CCWT and the CWAS algorithms, including the interpretation of their results and the additive synthesis process. The proposed separation algorithm and its main blocks (as the fundamental frequency estimation) will be presented in "Separation algorithm " section, with a detailed example. The numerical results of the different experiments and tests are shown in "Experimental results" section. Finally, the main conclusions and current and future lines of work are presented in "Conclusions" section.

Complex bandpass filtering

The CCWT

The CCWT can be defined in several ways [33]. For a certain input signal $x(t)$, it can be written as

$$W_x(a,b) = \int_{-\infty}^{+\infty} x(t)\Psi^*_{a,b}(t)dt \tag{1}$$

where * is the complex conjugate and $\Psi_{a,b}(t)$ is the *mother wavelet*, frequency scaled by a factor a and temporally shifted by a factor b:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \tag{2}$$

In our case, we will choose a *complex* analyzing wavelet, (specifically the Morlet wavelet). The Morlet wavelet is a complex exponential modulated by a Gaussian of width $2\sqrt{2}/\sigma$, centered in the frequency ω_0/a. Its Fourier transform is

$$\hat{\psi}_a(\omega) = Ce^{-\sigma^2\frac{(a\omega-\omega_0)^2}{2}} \tag{3}$$

where C is a normalization constant which can be calculated independently of the input signal in order to conserve the energy of the transform [34].

A general audio signal (and in particular a monocomponent signal) can be modeled as

$$x(t) = A(t)\cos[\,\phi(t)] \tag{4}$$

From the module and the argument of the complex wavelet coefficients of Equation (1) [32] it is possible to obtain a complex function, which can be written as

$$\rho(t) \approx A(t)e^{j[\phi(t)]} \tag{5}$$

This result can locally be applied to every detected partial of the analyzed signal, providing a model of the audio signal close to its canonical pair. The output (synthetic) signal is the real part of $\rho(t)$ (the real part of the additive synthesis of the detected partials in the general case). This synthetic signal remains very close to the original input signal $x(t)$ in numerical and acoustical terms [32].

The CWAS algorithm

In the CWAS algorithm [32], a complex mother wavelet allows us to analyze the complex coefficients of Equation (1), stored in a matrix (the CWT matrix), in module and phase, obtaining directly the instantaneous amplitude and the instantaneous phase of each detected component [34,35]. A single parameter, the *number of divisions per octave D* (a vector with as many dimensions as octaves present in the signal's spectrum), controls the frequency resolution of the analysis.

Figure 2 (bottom left) depicts the module of the complex wavelet coefficients (also called the wavelet spectrogram) of the mixture of a tenor trombone playing a *C*5 note and a trumpet playing a *D*5 vibrato. In the figure, the dark zones are associated with the main trajectories of information (each one related with a partial).

The addition in the time axis of the module of the wavelet coefficients represents the scalogram of the signal. The scalogram presents a certain number of peaks, each one related to a detected component of the signal. We found that the quality of the resynthesis significantly improves by extending the definition of partial not exclusively to the scalogram peaks, but to their *regions of influence*. So, in our model, a partial contains all the information situated between an upper and a lower frequency limits (the region of influence of a certain peak). These regions of influence can be seen in Figure 3, which shows the scalogram of a guitar playing an *E*4 note (330 Hz). Each maximum of the scalogram is marked with a black point. The associated upper and lower frequency limits for each partial are marked with red stars. They are located at the minimum point between adjacent maxima.

For a certain peak i of the scalogram, its complex partial function P_i can be defined as the summation of the complex wavelet coefficients obtained through Equation (1) between its related frequency limits [32]. Hence, we can write

$$P_i(t) = \sum_{m_i=m_{i_{\text{low}}}}^{m_{i_{\text{up}}}} W_x(a_{m_i},t) \quad \forall i = 1,\ldots,n \tag{6}$$

where $W_x(a_{m_i},t)$ are the wavelet coefficients $W_x(a_m,t)$, related with the ith peak (partial).

Studying the complex-valued function $P_i(t)$ in module and phase, we can obtain the instantaneous amplitude $A_i(t)$ and the instantaneous phase $\Phi_i(t)$ of each detected

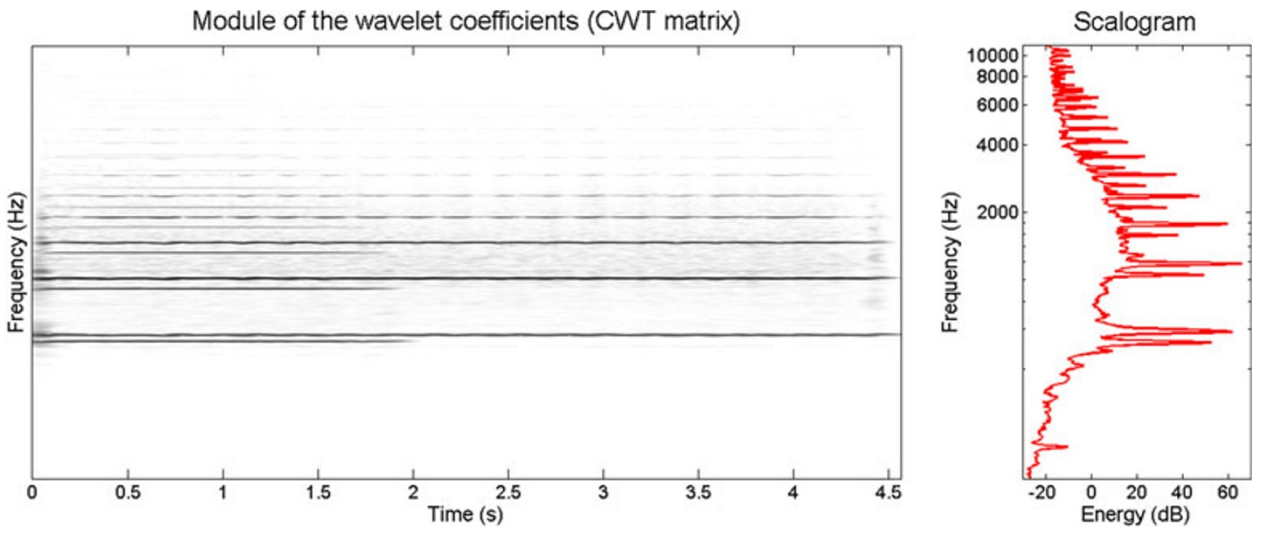

Figure 2 Left: The wavelet spectrogram, that is, the module of the CCWT coefficients (module of the CWT matrix). The dark zones are the different detected partials. **Right:** Scalogram. The analyzed signal is the mixture of a tenor trombone playing a C5 note and a trumpet playing a *D5* vibrato.

partial. The instantaneous frequency of the partial can be written [36] as

$$f_{i,\text{ins}}(t) = \frac{1}{2\pi}\frac{d[\,\Phi_i(t)]}{dt} \tag{7}$$

The global contribution of $P_i(t)$ to the scalogram of $x(t)$ can be approximated by

$$E_i = \sum_{m=1}^{l_i} \|P_i(t_m)\| \tag{8}$$

where t_m is the mth sample of the temporal duration of the partial i (whose length is l_i, in samples). Obviously, E_i is a measure of the energy of the partial.

As Equation (5) is true for every detected partial of the signal, the original signal $x(t)$ can be obtained through a simple additive synthesis method, performing the summation of the n detected partials, as was advanced in the previous section

$$x(t) = \Re\left(\sum_{i=1}^{n} P_i(t)\right) = \sum_{i=1}^{n} A_i(t)\cos[\,\Phi_i(t)]\,. \tag{9}$$

The objective of this study is to be able to use this information to somehow separate a signal composed of two or more mixed notes into the original isolated sources. The only input of the system is the mixed signal (no additional data is needed).

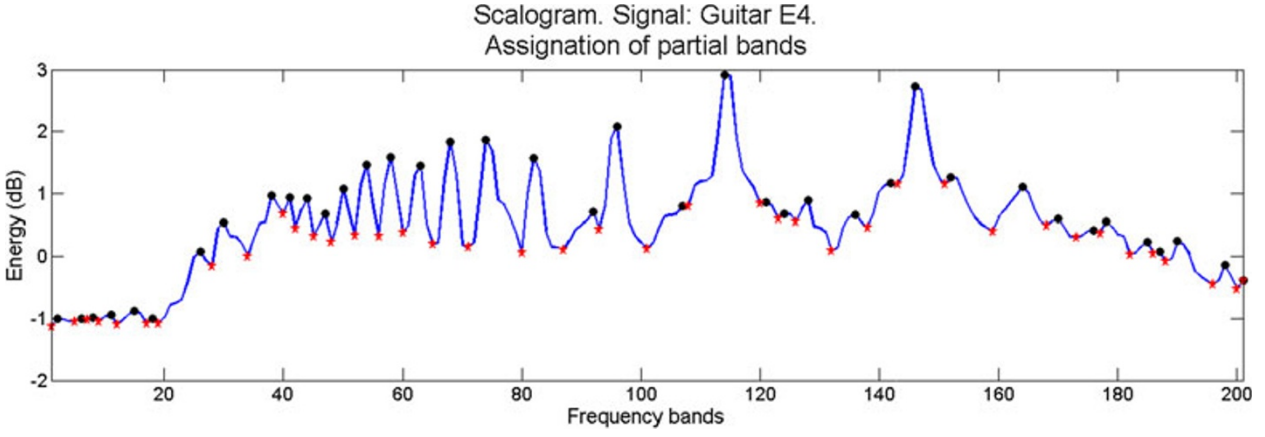

Figure 3 Scalogram of a guitar playing an *E4*. Each peak (black points) is related to a detected partial. Each partial has an upper and lower frequency (band) limit (red stars), located at the minimum point between two adjacent maxima.

BASS

The monaural mixed signal $x(t)$ can be written as

$$x(t) = \sum_{k=1}^{N} s_k(t) \tag{10}$$

As stated above, BASS attempts to obtain the original isolated sources $s_k(t)$ present in a certain signal $x(t)$, when the mixture process is unknown.

As we do not know *a priori* the number N of sources present in $x(t)$, the first problem is to divide the detected partials into as many different families or categories as sources, having a minimum error between members of a class [19]. A first approximation to the BASS task using the CWAS technique was performed and presented in [37]. There, we used an onset detection algorithm [38] to find a rough division of the partials, grouping them into the different sources. The main advantage of using the CWAS algorithm instead of the STFT is its proven ability of high-quality resynthesis. As explained, the time and frequency errors in the synthesis of signals using the CWAS algorithm is remarkably small, and the acoustical differences between the original and synthetic signals are negligible for most of people [32]. This high fidelity synthesis converts the CWAS algorithm in a very useful tool for source separation.

In the general case, when there are two or more audio sources present in the analyzed signal, a certain partial can be part of one of the sources, it can be shared by two or more sources, or it can be part of none of them (i.e., inharmonic or noisy partials). The algorithm will search for any fundamental frequency present in the mixed signal, and each f_0 will be considered as an indicator of the presence of a source (see "Multiple $\boldsymbol{f_0}$ estimation" section). A harmonic analysis will find the set of partials which belongs to each source, and the set of overlapping partials (and which sources are overlapping for each case). Then, the information of the isolated partials will be used to reconstruct an estimation of the contribution from each source to every overlapping partial, and the separated sources will be generated by additive synthesis (see "The separation process" section). This idea was used in [22], but in this study the only input information is the mixed signal (we do not need the estimated pitch, because the f_0 estimator gives us this information). The quality of the separation (see "Quality separation measurement" section) will be measured using the standards proposed in [39].

Separation algorithm

In this section, we will detail the proposed separation technique, and in particular its two main blocks: the estimation of the fundamental frequencies present in the mixed signal, and the separation process. A detailed example will be developed in parallel, in order to clarify the specified separation process. In this example, we will use a signal chosen arbitrarily from the set of analyzed signals (see "Experimental results " Section). In this case, the separation of a mixture composed of a trumpet playing a $D5$ (587 Hz) vibrato and a tenor trombone playing a $C5$ (523 Hz) note. For this signal, Equation (10) becomes

$$x(t) = s_1(t) + s_2(t) \tag{11}$$

The waveform, module of the CWT matrix, and scalogram of this signal can be seen in Figure 2. The numerical quality separation measurement of this signal can be seen in the following section. In the example, we will concentrate on a single overlapping partial. The isolated original partials will also be used to test the robustness of the method.

The main steps of the separation algorithm are summarized below.

- From $x(t) \rightarrow P_i(t) \rightarrow A_i(t)$, $\Phi_i(t)$ (CWAS).
- From $\Phi_i(t)$, through Equation (7) $\rightarrow f_i(t)$.
- Estimation of f_{0k} and their harmonic partials $\forall k$.
- Separation of overlapping partials.
- Additive synthesis $\rightarrow s_k(t)$.

It is important to remark that, at its actual stage, the separation process is performed using the information of the whole signal.

Multiple f_0 estimation

In this study, we have considered that a musical instrument cannot play more than one note simultaneously (i.e., we work mainly with monophonic instruments). If an instrument plays two or more notes simultaneously (polyphony), the developed algorithm will consider that each note comes from a different source. With such an approximation, the present fundamental frequencies f_{0j}, $j = 1, \ldots, N$ become the natural parameter which will be used to calculate the number of sources present in the mixture, and the reliability in the f_0 estimator acquires capital importance.

An algorithm of multipitch analysis based on the work of Klapuri [28,29], specially adapted for the CWAS algorithm, which uses the spectral smoothing technique of Pérez-Sancho et al. [40] has been developed. Figure 4 shows a block diagram of this algorithm.

The input (mixed) signal is analyzed using the CWAS algorithm, which provides as results the n complex functions that define the temporal evolution of each detected partial. Using Equations (7) and (8), the instantaneous frequencies for each partial (and their respective average values, $\overline{f_j}$ $\forall j = 1, \ldots, n$) and the energy distribution of the signal are obtained. This information is equivalent to the scalogram of the signal clustered around the set of detected partials. Only the partials with energy

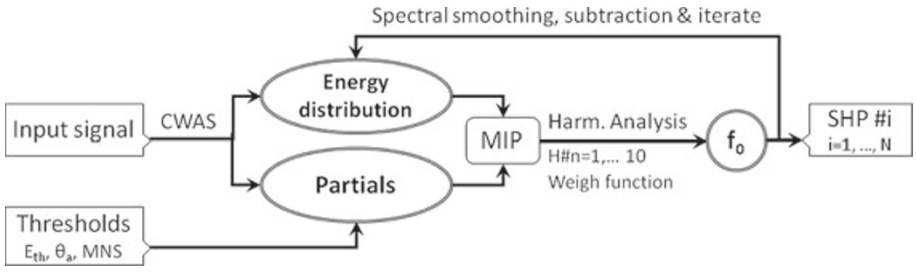

Figure 4 Block diagram of the fundamental frequencies estimation algorithm. *MIP* is the most important (energetic) partial. The output *SHP#i* is the set of the harmonic partials corresponding to each detected source ($i = 1, \ldots, N$). See text for details.

greater than the threshold $E_{\text{th}} = 1\%$ will be considered in the search of the harmonic sets associated with each source. From the remaining energy distribution, the most energetic partial (MIP in Figure 4) is selected, and the harmonic analysis is computed next.

Starting from the average frequency of the most important partial $\overline{f}_j$, it is assumed that this partial is harmonic of a certain fundamental frequency f_{0k}, that is

$$f_{0k} = \frac{\overline{f}_j}{k}, \qquad \forall k = 1, 2, \ldots, N_A \tag{12}$$

In this study, we have taken $N_A = 10$. In other words, the MIP will be at most the 10th harmonic of its related fundamental frequency. From the fundamental frequencies so obtained, the set of harmonic frequencies regarding each one is calculated.

$$f_{k,m} = m f_{0k}, \quad \forall m = 1, 2, \ldots, N_k \tag{13}$$

where N_k is the higher natural such that satisfies $N_k f_{0k} \leq f_s/2$, being f_s the sampling rate.

In the next step, for each $f_{k,m}$, its related partial is searched. A partial of mean frequency $\overline{f}_i$ is the mth harmonic of a certain fundamental frequency f_{0k} if

$$\left| \frac{\overline{f}_i}{f_{0k}} - \frac{f_{k,m}}{f_{0k}} \right| \leq \theta_a \tag{14}$$

where θ_a is the *inharmonicity* threshold. Taking $\theta_a = 0.03$, the partials of an inharmonic instrument like the piano are correctly analyzed.

The decision on which is the fundamental frequency associated with the current MIP is taken through a weight function w_k calculated for each of the candidates. This weigh function is proportional to the energy contribution of its set of partials:

$$w_k = \frac{n_{ip,k}^2}{n_{a,k}} \sum_{i=1}^{n_{a,k}} E_{i,k} \tag{15}$$

where $n_{a,k}$ is the total number of harmonics associated with f_{0k} and $n_{ip,k}$ is the number of partials with energy above the threshold E_{th}. $E_{i,k}$ is the energy of the ith partial associated with f_{0k}.

The fundamental frequency related to the current MIP is the one whose weight w_k is maximum. The algorithm stores the set of harmonic partials or spectral pattern, $\mathbf{P}_k = \{P_{1,k}, P_{2,k}, \ldots P_{n_a,k}\}$, which includes the obtained fundamental frequency, and proceeds to apply the spectral smoothing [40] to its energy distribution $\mathbf{E}_k = \{E_{1,k}, E_{2,k}, \ldots E_{n_a,k}\}$

$$\widetilde{\mathbf{E}}_k = G_w \star \mathbf{E}_k \tag{16}$$

where $G_w = \{0.212, 0.576, 0.212\}$ is a truncated normalized Gaussian window with three components and $\star$ is the convolution product operator. The smoothed energy for each harmonic partial is calculated as

$$E'_{i,k} = \begin{cases} E_{i,k} - \widetilde{E_{i,k}} & \text{if } E_{i,k} - \widetilde{E_{i,k}} > 0 \\ 0 & \text{if } E_{i,k} - \widetilde{E_{i,k}} \leq 0 \end{cases} \tag{17}$$

Substituting these new energy values into its corresponding partials of the original energy distribution, a new MIP can be obtained. The process is iterated until the energy of the distribution descends under a threshold or the maximum number of sources (MNS in Figure 4) has been reached. In this study, we have limited the number of sources to MNS = 5. Using this technique, it is possible to obtain the fundamental frequencies even in the most difficult cases, for example when a fundamental frequency is overlapped with a harmonic corresponding to other source or in the case of suppressed fundamentals. Overlapping fundamentals will not be detected using this technique.

This algorithm has been tested using a set of more than 200 signals, most of them extracted from the musical instrument samples of the University of Iowa [41].[a] Experimental results are shown in Table 1. In this table, the accuracy of the multipitch analysis is shown in four categories: Isolated instruments, synthetically generated mixtures of two and three harmonic instruments and mixtures of one harmonic and one inharmonic instrument.

Table 1 Accuracy results of the fundamental frequency estimation algorithm

	Analyzed signals (#)	Succes. det. (#)	Estim. error (%)
1 instr.	106	106	0
2 instr.	75	74	1.34
1H + 1I instr.	4	4	0
3 instr.	50	49	2
Total	235	233	0.85

Errors can be due to missed detections, wrong estimations, or false fundamentals.

In the signal of the example (Figure 2), the exact results given by the fundamental frequency estimator are $f_{01} = 589.25$ Hz for the trumpet and $f_{02} = 525.96$ Hz for the trombone. The instantaneous amplitude from these partials is shown in Figure 5. The continuous line comes from the fundamental partial of the trumpet and the dashed line from the tenor trombone fundamental. The set of harmonic partials from each instrument will be shown later.

The separation process

Once the fundamental frequencies present in the mixed signal (and the number of sources N) have been obtained, the separation process begins. A detailed block diagram of the process is shown in Figure 6.

Analyzing the sets of harmonic partials for each source, it is easy to distinguish between isolated harmonics (that is, partials which only belong to a single source) and overlapping harmonics (partials shared by two or more sources). The isolated harmonics and the fundamental partial of each source will be used later to separate the overlapping partials, through their onset and offset times, instantaneous envelopes, and phases. The separated source is eventually synthesized by the additive synthesis of its related set of partials (isolated and separated).

The inharmonic limit

Inharmonicity is a phenomenon occurring mainly in string instruments due to the stiffness of the string and non-rigid terminations. As a result, every partial has a frequency that is higher than the corresponding harmonic value. For example, the inharmonicity equation for a piano can be written [42] as

$$f_n = nf_0\sqrt{1+\beta n^2} \tag{18}$$

where n is the harmonic number and β is the inharmonicity parameter. In Equation (18), β is assumed constant, although it can be modeled more accurately by a polynomial up to order 7 [43]. It means that the parameter β has different values depending on the partials used to calculate it. Partials situated in the 6–7 octave provide the optimal result. Using two partials of order m (lower) and n (higher), it is

$$\beta = \frac{\delta - \varepsilon}{\varepsilon n^2 - \delta m^2} \tag{19}$$

where $\delta = (mf_n/nf_m)^2$ and ε is an induced error due to the physical structure of the piano which cannot be evaluated [42]. If partials m and n are correctly selected, $\varepsilon \approx 1$.

With the inharmonic model of Equation (19), it is possible to calculate the inharmonicity parameter β for each detected source, using (when possible) two isolated partials situated in the appropriate octaves. *A priori*, this technique includes inharmonic instruments (like piano) in the proposed model. Unfortunately, the obtention of the parameter β do not improve significantly the quality separation measurements evaluated in the tests.

Assumptions

In order to obtain the envelopes and phases of an overlapping partial related to each source, we will assume two approximations. The first one is a slightly less restrictive version of the CAM principle, which asserts that the amplitude envelopes of spectral components from the same source are correlated [22].

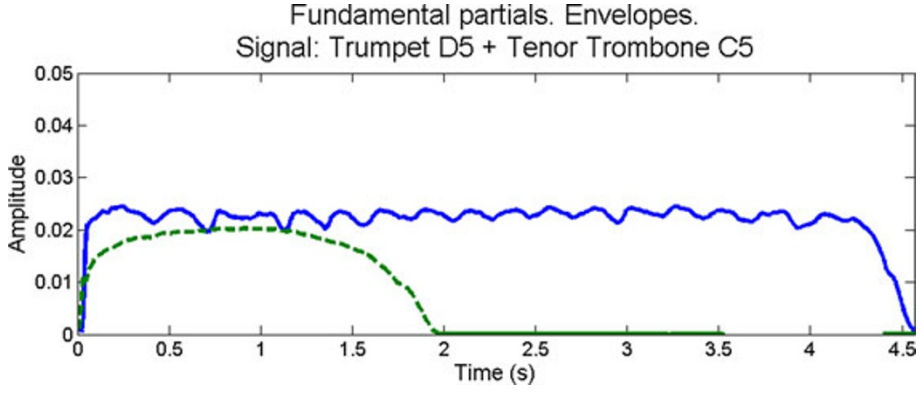

Figure 5 Envelopes of the fundamental partials. Continuous line, trumpet fundamental. Dashed line, tenor trombone fundamental.

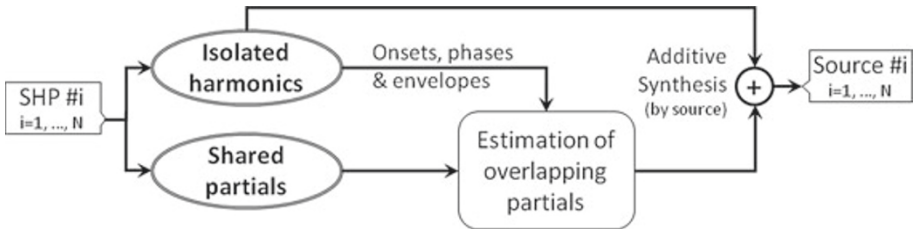

Figure 6 Block diagram of the separation algorithm. The input *SHP#i* is the set of the harmonic partials corresponding to each detected source ($i = 1, \ldots, N$). See text for details.

- The amplitudes (envelopes) of two harmonics P_1 and P_2, *with similar energy* $E_1 \approx E_2$, both belonging to the same source, have a high correlation coefficient.

As long as this approximation is true, we will have better separation results. As we are using the global signal information, the correlation coefficient between the strongest harmonic (and/or the fundamental partial) and the other harmonics decreases as the amplitude/energy differences between the involved partials increase [22]. Hence, the choice for the reconstruction of non-overlapping harmonics whose presence is energetically similar to the energy of the overlapping harmonic suggests that the correlation factor between the involved partials will be higher. In fact, as the correlation between high-energy partials tends also to be high, while the errors related with this assumption in lower energy partials tend to be energetically negligible, in most cases the quality measurement parameters have a high value, and the acoustic differences between the original and the separated source are acceptable.

The second approximation is

- The instantaneous phases of the pth and the qth harmonic partials belonging to the same source are approximately proportional with ratio p/q, except an initial phase gap, ϕ_0. That is

$$\phi_2(t) \approx \frac{p}{q}\phi_1(t) + \Delta\phi_0 \tag{20}$$

  where $\Delta\phi_0 = 0$ means that the initial phases of the involved partials are equal, that is, $\phi_{0p} = \phi_{0q}$.

We have found that in our model of the audio signal and even knowing the envelopes of the original overlapping harmonics, a difference in the initial phase $\Delta\phi_0 = 10^{-3}$ is enough to make impossible an adequate reconstruction of the mixed partial. Each partial has an aleatory initial phase (i.e., there is not a relation between ϕ_{0p} and ϕ_{0q}). However, as the instantaneous frequency of the mixed harmonics can be retrieved with accuracy independently of the value of the initial phase, the original and the synthetically mixed partials (using the separated contribution from each source) present similar sounds (provided that the first assumption is true).

Reconstruction process and additive synthesis

As mentioned above, in the proposed technique we use the information of the isolated partials to reconstruct the information of the overlapping partials. The output of the multipitch estimation algorithm is the harmonic set corresponding to each source present in the mixture. With this information, it is easy to distinguish between the isolated partials (partials belonging to a single given source) and the shared partials. For each overlapping partial, it is immediate to know the interfering sources. We can write

$$\mathbf{P}_k = \mathbf{P}_k^{(\mathrm{iso})} \bigcup \mathbf{P}_k^{(\mathrm{sh})} \tag{21}$$

In the example of the tenor trombone and the trumpet mixture, Figure 2, the instantaneous amplitudes of the isolated partials are shown in Figure 7. The instantaneous amplitudes of the fundamental partials are depicted in bold lines.

Using the information of the isolated partials and through an onset detection algorithm [38], it is easy to detect the beginning and the end of each present note. This information is necessary to avoid the artifacts and/or noise caused by the mixture process which tends to appear before and after active notes. This noise is acoustically annoying and makes worse the numerical quality separation measurement results.

Consider a certain mixed partial P_m of mean frequency $\overline{f_m}$. The mixed partial can be written as follows

$$\begin{aligned} P_m(t) &= A_m(t)e^{j[\phi_m(t)]} = \sum_{s_k} P_{s_k}(t) \\ &= \sum_{s_k} A_{s_k}(t)e^{j[\phi_{s_k}(t)]} \end{aligned} \tag{22}$$

where $P_{s_k}(t)$ are the original harmonics which overlap in the mixed partial. In Equation (22), the only accessible information is the instantaneous amplitude and phase of the mixed partial, that is, $A_m(t)$ and $\phi_m(t)$. The aim is to recover each $A_{s_k}(t)$ and $\phi_{s_k}(t)$ as accurately as possible.

To do this, it is necessary to select a partial belonging to each overlapping source s_k in order to separate the different contributions to P_m. From each isolated set of partials $\mathbf{P}_k^{\mathrm{iso}}$ corresponding to the interfering sources, we will search for a partial j with an energy E_j as similar to

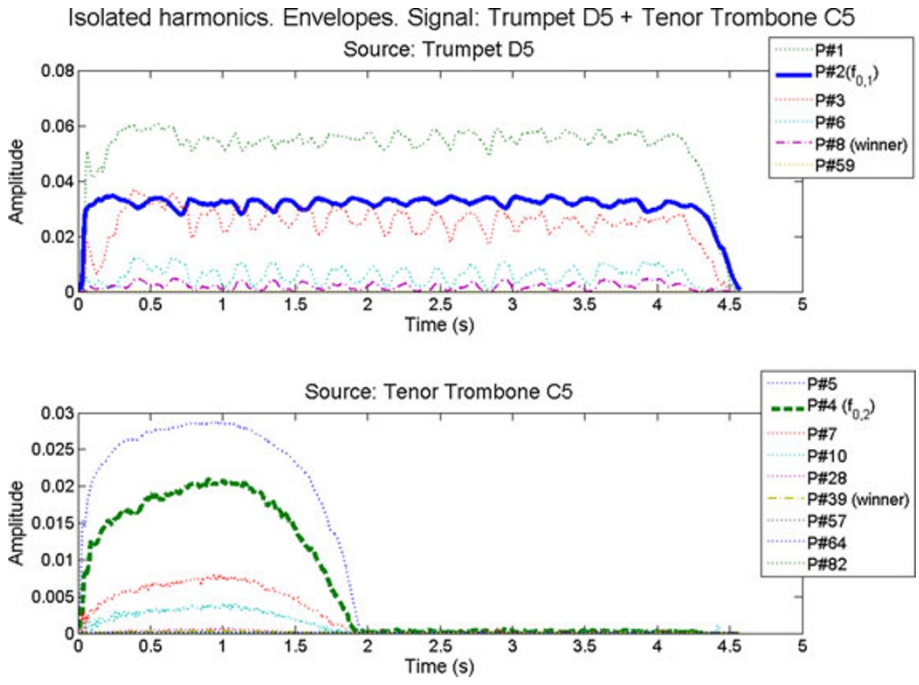

Figure 7 Envelopes of the isolated set of harmonics from each source (dotted lines). The fundamental envelopes are marked with bold lines. Continuous trace, trumpet; dashed line, tenor trombone.

the energy of P_m as possible, and with a mean frequency $\overline{f_j}$ as close to $\overline{f_m}$ as possible. If $\Delta(E_{j,m}) = |E_j - E_m|$ and $\Delta(f_{j,m}) = |\overline{f_j} - \overline{f_m}|$, these conditions can be written as

$$P_{k,\text{win}} = \text{argmin}(\Delta E_{j,m})|_{P_j \in \mathbf{P}_k^{(\text{iso})}} \tag{23}$$

and

$$P_{k,\text{win}} = \text{argmin}(\Delta f_{j,m})|_{P_j \in \mathbf{P}_k^{(\text{iso})}} \tag{24}$$

The energy condition, Equation (23), is calculated in the first place. Only in doubtful cases, the frequency condition of Equation (24) is evaluated. However, both conditions often lead to the same winner. For the purposes of simplicity, let P_{wk} denote the selected (winner) isolated partials of each source k. This can be written

$$P_{wk}(t) = A_{wk}(t)e^{j[\phi_{wk}(t)]} \quad \forall k \tag{25}$$

If $\overline{f_{wk}}$ is the mean frequency of the winner partial of the k source, it is easy to see that

$$\frac{\overline{f_{wk}}}{\overline{f_m}} = \frac{p_k}{q_k} \tag{26}$$

for some $p_k, q_k \; in \; \mathbb{N}$.

In fact, the same ratio p_k/q_k can be used to reconstruct the corresponding instantaneous frequency for each interfering source with high accuracy. In Figure 8, the instantaneous frequencies of the original (interfering) partials and the estimated instantaneous frequency of each separated contribution are shown, for a certain case of overlapping partial. In this figure, the original instantaneous frequencies are depicted in blue, and the reconstructed instantaneous frequencies in red. Note the accuracy in the estimation of each instantaneous frequency. The blue line corresponding to the tenor trombone is shorter due to the signal duration.

Hence, it is possible to use Equation (20) to reconstruct the phases ϕ_{s_k} of the separated partials for each overlapping source.[b]

Unlike other works [22,23], to reconstruct the envelope of the partials separated it is assumed that the instantaneous amplitude of the mixed partial $A_m(t)$ is directly a linear combination of the amplitudes of the interacting components $A_{wk}(t)$ (hence, unlike other existing techniques [22,23], the phases of the winner partials are not taken into account in this process). Therefore,

$$A_m(t_i) = \sum_{s_k} \alpha_k A_{wk}(t_i) \quad \forall t_i \tag{27}$$

The solution of Equation (27) that minimizes the error in the sum is equivalent to the least-squares solution in the presence of known covariance of the system.

$$\mathbf{A} * \alpha = b \tag{28}$$

where $\mathbf{A}$ is a matrix which contains the envelopes of each selected (winner) partial described by Equations (23) and (24), α is the mixture vector and $b = A_m(t)$.

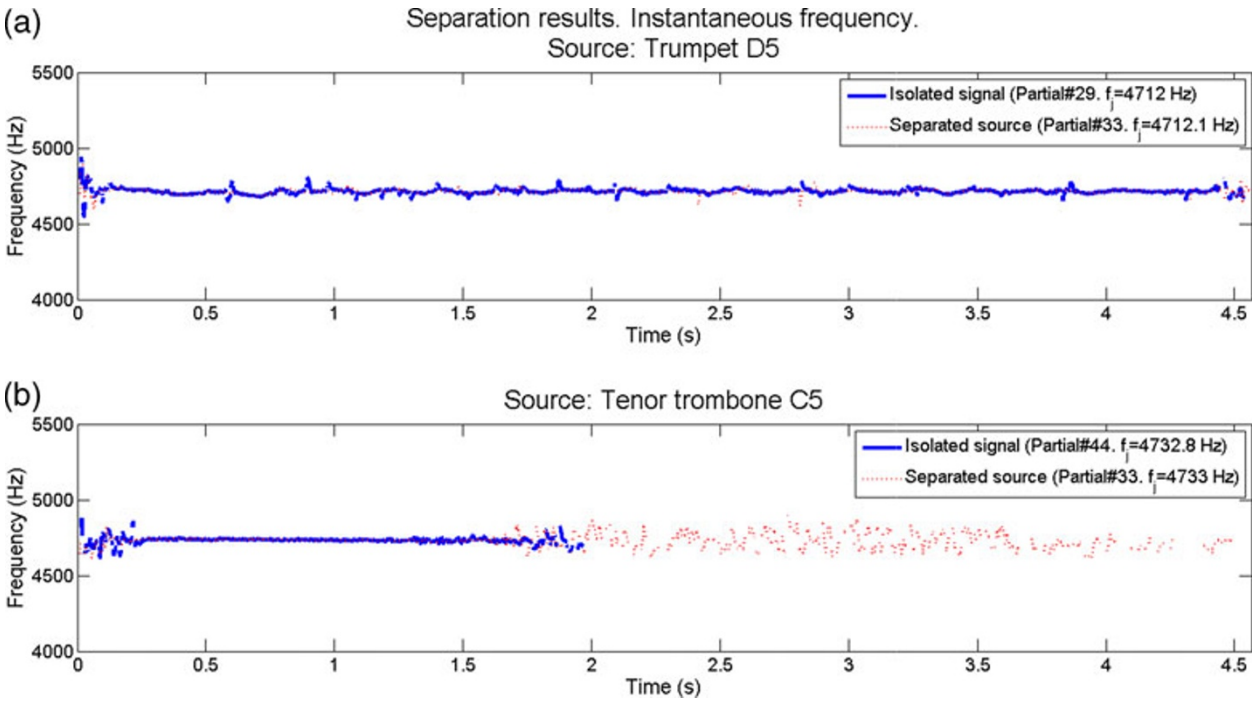

Figure 8 Comparison between the original (isolated) instantaneous frequencies and the estimated (separated) instantaneous frequencies. (a) Results for the trumpet source (continuous blue line, the original f_{ins}; red dotted line, the estimated one). **(b)** Results for the tenor trombone source.

Once found α_k, p_k, and q_k for each source k, the overlapping partial can be written as

$$P_m(t) = \sum_{s_k} \alpha_k A_{wk}(t) e^{j\left[\frac{p_k}{q_k}\phi_{wk}(t)\right]} \tag{29}$$

and the separated contributions of each present source are of course

$$P_{s_k}(t) = \alpha_k A_{wk}(t) e^{j\left[\frac{p_k}{q_k}\phi_{wk}(t)\right]} \tag{30}$$

Once each separated partial is obtained using the technique described, it is added to its corresponding source. This iterative process eventually results in the separated sources.

Figures 9 and 10 show the wavelet spectrograms and scalograms (obtained from the CWAS algorithm) corresponding to the isolated signals (tenor trombone and trumpet, respectively) and their related separated sources. From the spectrograms (module of the CWT matrix), it can be observed that most of the harmonic information has properly been recovered. This conclusion is reinforced using the scalogram information. Note that the harmonic reconstruction produces an artificial scalogram (red line) harmonically coincident with the original scalogram (blue line).

In the figures, the separated wavelet spectrogram shows that only the harmonic partials have been recovered.

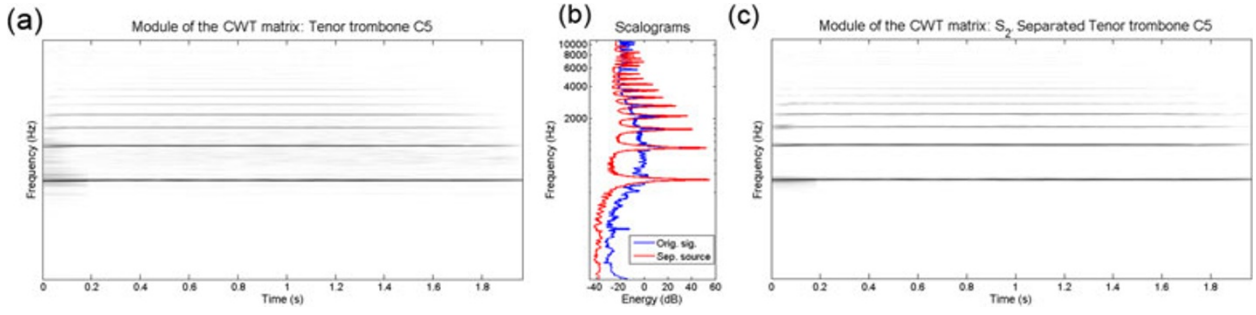

Figure 9 Spectrograms of the tenor trombone signals. (a) Wavelet spectrogram of the original (isolated) tenor trombone. **(b)** Blue line: Original scalogram. Red line: Scalogram of the separated source. **(c)** Wavelet spectrogram of the separated source.

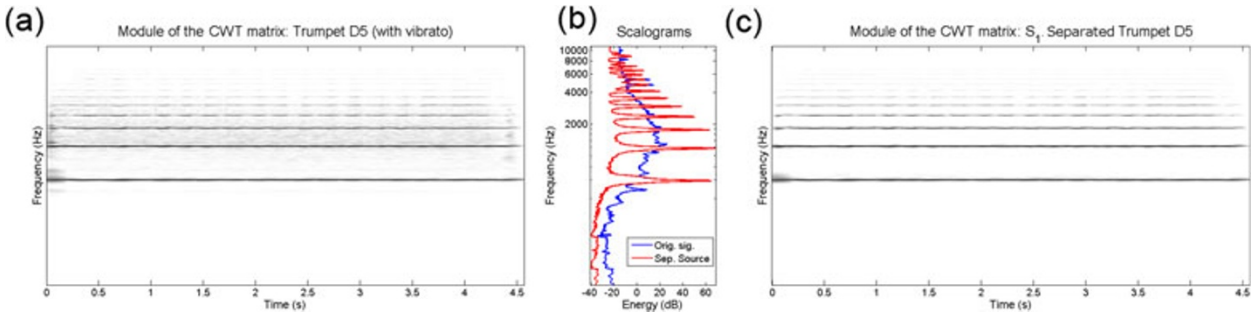

Figure 10 Spectrograms of the trumpet signals. (a) Wavelet spectrogram of the original (isolated) trumpet. **(b)** Blue line: Original scalogram. Red line: Scalogram of the separated source. **(c)** Wavelet spectrogram of the separated source.

When the inharmonic partials carry important (non noisy) information, the synthetic signal can sound somewhat different (as happened with the possible envelope errors in the high-frequency partials).

The values of the standard quality measurement parameters for this example and the rest of the analyzed signals will be detailed in "Summarizing: graphical results" section.

Main characteristics, advantages, and limitations

The reconstruction of overlapping partials causes that there is no information wrongly assigned to the separated sources using this technique, except the existing interference in the set of isolated partials. This means that the interference terms in the separation process will be in general negligible. This result will be numerically confirmed in "Experimental results " section.

The advantages of this separation process are mainly two. First, the process of separation of overlapping harmonics (multi-pitch estimation, calculus of the best linear combination for reconstruction, additive synthesis) is not computationally expensive. In fact, the obtention of the wavelet coefficients and their separation into partials uses much more computation time. The second advantage of this process is that the separation is completely blind. That is, we do not need any *a priori* characteristic of the input signal, neither the pitch contour of the original sources nor the relative energy, number of present sources, etc.

One of the most important limitations of this method is that is not valid for separating completely overlapping notes. Although the detailed algorithm of estimation of fundamental frequencies is capable of detecting overlapping fundamentals, in such a case the set of isolated partials of the overlapped source would be essentially empty, and therefore no isolated information would be available to carry out the reconstruction of phases and amplitudes of the corresponding source. To solve this problem (assuming the separation of musical themes of longer duration), it is possible to use models of the instruments present in the mixture, or previously separated notes from the same source. These ideas are the basis of HTES and AHS techniques (see "Introduction" section).

On the other hand, as was advanced in "Introduction" section, at its current stage, the proposed technique can be used to separate two or more musical instruments, each one playing a single note. The final quality of the separation depends of the number of mixed sources. This is due to the accuracy of the estimation of fundamental frequencies, and to the use of isolated partials to reconstruct the overlapping harmonics. The higher the number of sources, the lower the number of isolated harmonics and the poorer the final musical timbre of the separated sources.

Experimental results

The analyzed set of signals includes approximately 100 signals with two sources and 60 signals with three sources. All the analyzed signals are real recordings of musical instruments, most of them extracted from [41]. The final set of musical instruments includes flute, clarinet, sax, trombone, trumpet, oboe, bassoon, horn, tuba, violin, viola, guitar, and piano.

All the analyzed signals have been sub-sampled to $f_s = 22050$ Hz, then synthetically mixed. The number of divisions per octave D and all the thresholds used in the CWAS and the separation algorithms are the same for all the analyzed signals. Specifically, $D = \{16; 32; 64; 128; 128; 100; 100; 100; 100\}$, θ_{th}=0.03, E_{th}=1%. Observe that the number of divisions per octave depends on the octave, so we have a variable resolution.

We have developed eight experiments with two and three synthetically mixed sources. In each experiment, we have analyzed 20 signals. These experiments are listed in Table 2. In the next paragraphs, we will explain these experiments. Graphical and numerical results are given in "Summarizing: graphical results" section.

Experiment 1: harmonic and inharmonic instruments

In the first experiment, we have mixed a inharmonic instrument (piano) with one harmonic instrument.

Table 2 List of BASS experiments developed

Experiment (#)	Sources (#)	Instruments involved	Experiment characteristics
1	2	Different	1 Harm. +1 Inharm.
2	2	Same	Same octave
3	2	Same	5th & 12th intervals
4	2	Different	5th & 12th intervals
5	2	Different	Inharmonic notes
6	3	Same	Major chord
7	3	Same	Minor chord
8	3	Different	Inharmonic notes

Numerical data are presented in the first column of Figures 11, 12, and 13. The numerical separation results are not as good as results of Experiment 5, which is otherwise similar to this one (acoustically the situation is better). It is probably due to the uncertainty in the obtention of the inharmonicity parameter, β [43] (see "The inharmonic limit" section).

Experiment 2: single instrument, same octave

In the second test, two musical instruments (Alto Sax and Flute, respectively) were taken randomly from the original database. We have generated a total of 11 signals with each instrument, with two notes of the fourth octave (considering $A4 = 440\,\text{Hz}$) played by the same instrument. One of the notes is always a $C\#4$ (277 Hz), the other note corresponds to the same octave ($C4$, $D4$, $D\#4$, etc.). The experimental values of SDR, SIR, and SAR are presented in the second column of Figures 11, 12, and 13.

Experiment 3: single instrument, harmonic-related notes

In the third experiment, we mixed two harmonic note intervals from the same instrument. The used harmonic relations are: $C-G$, $D-A$, $E-B$, $F-C$, $G-D$, $A-E$, and $A\#-F$ from the same or different octave. That is, 5th and 12th intervals. We have generated three sets of signals, each one corresponding to one musical instrument (concretely, Alto Sax, Flute and *Bb* Clarinet), and seven mixtures from each one. Numerical results of this experiment are shown in the third column of Figures 11, 12, and 13.

Experiment 4: two instruments, harmonic-related notes

In the next experiment, we have mixed in 20 signals the same harmonic intervals of the previous experiment, this time executed by different musical instruments: Alto sax, guitar, bassoon, *Bb* and *Ee* clarinets, horn, oboe, and flute. The experimental values of the quality separation measurement are presented in the fourth column of Figures 11, 12, and 13.

Experiment 5: two instruments, inharmonic notes

In this experiment, each analyzed signal contains the mixture of two aleatory chosen musical instruments playing aleatory (non-harmonically related) notes. The experimental values of the quality separation parameters are presented in the fifth column of Figures 11, 12, and 13.

Experiment 6: one instrument, major chord

A major chord is the mixture of three notes, concretely $C-E-G$. We have generated 20 of these chords, played by the same musical instrument, concretely Bassoon, Alto Sax, *Bb* Clarinet, Flute and Trumpet. Numerical data are presented in the sixth column of Figures 11, 12, and 13.

Experiment 7: one instrument, minor chord

A minor chord is the mixture of $A-C-E$ notes. We have analyzed 20 signals, each one played by a single musical instrument: Bassoon, *Bb* Clarinet, Horn, Oboe, and Trumpet. The SDR, SIR, and SAR values for this experiment are depicted in the seventh column of Figures 11, 12, and 13.

Experiment 8: three instruments, inharmonic notes

Finally, 20 signals with three aleatory instruments playing aleatory (non-harmonically related) notes have been analyzed. These signals are randomly distributed from octaves 2 to 6, and 10 of the signals present widely separated notes. The experimental values of the quality

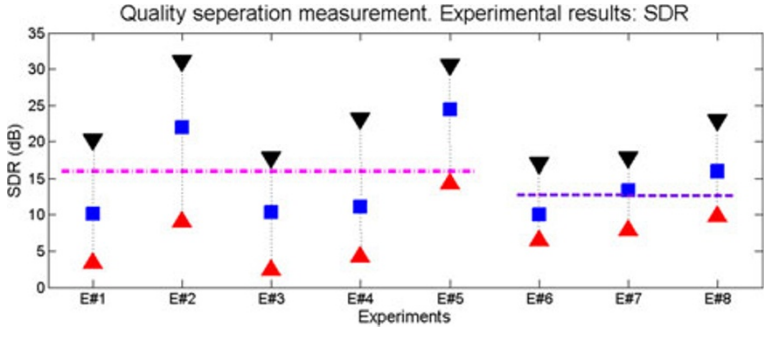

Figure 11 Experimental results of the SDR parameter for the eight separation experiments.

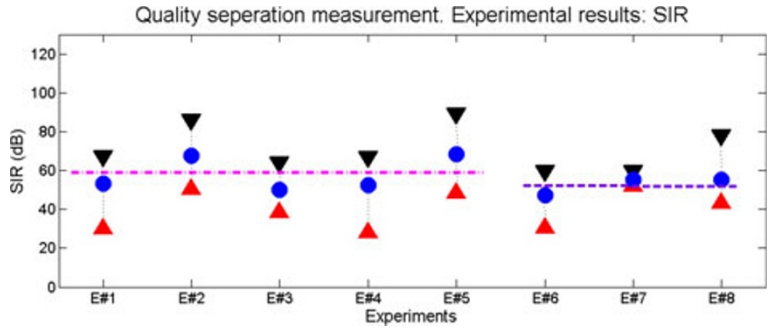

Figure 12 Experimental results of the SIR parameter for the eight separation experiments.

separation measurement parameters are presented in the last column of Figures 11, 12, and 13.

Quality separation measurement

We will assume that the errors committed in the separation process can have three different origins: they can be due to interference between sources, to distortions inserted in the separated signal, and to artifacts introduced by the separation algorithm itself.

We have used three standard parameters to test the final quality of the separation results using the proposed method related to these distortions. These parameters are the *signal-to-interference-ratio, (SIR)*, the *signal-to-distortion-ratio, SDR*, and the *signal-to-artifacts-ratio, SAR* [39,44,45]:

$$\mathrm{SIR} = 10\log_{10}\left(D_{\mathrm{interf}}^{-1}\right) \tag{31}$$

$$\mathrm{SDR} = 10\log_{10}\left(D_{\mathrm{total}}^{-1}\right) \tag{32}$$

and

$$\mathrm{SAR} = 10\log_{10}\left(D_{\mathrm{artif}}^{-1}\right) \tag{33}$$

where D_{interf}, D_{total}, and D_{artif} are energy ratios involving the separated signals and the target (isolated, supposed known) signals. The quality separation measurements of the next sections have been obtained within the *MATLAB*® toolbox *BSS_EVAL*, developed by Févotte, Gribonval, and Vincent and distributed online under the GNU Public License [44].

Summarizing: graphical results

As advanced before, in Figures 11, 12, and 13, we show the numerical results of the detailed tests. In Figure 11, the experimental values of the SDR parameter for each experiment are presented. In Figure 12, we have depicted the obtained SIR values. Finally, in Figure 13, the experimental values of the SAR parameter are shown.

In Figure 11, marked with squares, the SDR mean result for each test; with triangles, the maximum and minimum value of the parameter. These results show significant differences in the quality separation measurements for the experiments of separation involving two sources. In the case of experiments with three sources, the differences are smaller.

In Figure 12, the SIR mean result for each test is marked with circles; with triangles, the maximum and minimum value of the parameter. As can be seen in the figure, the experimental values of SIR present less variations than in the previous case. It means that the proposed technique does not present significative tendency to high interference terms.

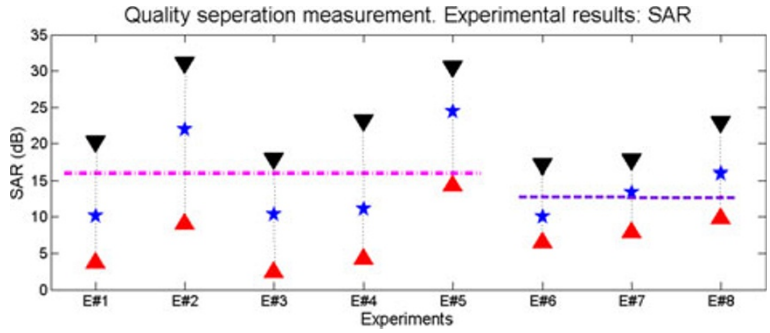

Figure 13 Experimental results of the SAR parameter for the eight separation experiments.

Finally, in Figure 13, the SAR results for each test are marked with stars. The maxima and minima of the experiments are depicted with triangles. The conclusions are the same that in Figure 11.

If we consider globally the whole set of signals with two mixed sources, the mean values of the quality separation measurement parameters can be used in some way to measure the final quality of the separation. These values (represented in Figures 11, 12, and 13 with horizontal dashed-dotted lines) are

- $\overline{SDR_{2s}} \approx 16.07$ dB.
- $\overline{SIR_{2s}} \approx 58.85$ dB.
- $\overline{SAR_{2s}} \approx 16.08$ dB.

The average of the standard parameters in the case of three mixed sources (horizontal dashed lines in Figures 11, 12, and 13) are

- $\overline{SDR_{3s}} \approx 12.81$ dB.
- $\overline{SIR_{3s}} \approx 52.03$ dB.
- $\overline{SAR_{3s}} \approx 12.82$ dB.

These results are consistent with the increasing number of sources in the mixture. Under the same degree of precision in the frequency axis, the higher the number of sources, the lower the separation between partials and the higher the probability of interference (lower SIR). Hence, the final distortions and artifacts tend to increase.

Conclusions

In this study, a BASS technique for monaural musical notes has been presented. There are two main differences between the proposed algorithm and the existing ones: first of all, the time–frequency analysis tool is not based on the STFT but in the CCWT, which offers a highly coherent model of the audio signal in both time and frequency domains. This tool allows us to obtain with great accuracy the instantaneous evolution (in time and frequency) of the isolated harmonics, easily assignable to the sources present in the mixture. Second, the separation algorithm only needs the mixed signal as input, no additional information is needed. The overlapping partials can entirely be reconstructed from the isolated partials searching for the best linear combination which minimizes the amplitude error in the mixture process, assuming the CAM principle. Using non-overlapping partials with similar energy to the overlapping partials, if the overlapping partial has high energy, the correlation factor tends to be high, and if the energy is low, errors associated with the low correlation are usually acceptable. The phase reconstruction is not as important as in other techniques, obtaining separated sources which have both high-quality separation measurement values and high-acoustic resemblance with respect to the original signals.

At its actual stage, the proposed technique can be used to separate two or more (monophonic) sources playing a single (and no proportional) note each. As the polyphony of the mixture signal increases, the acoustic performance of the separated signals tend to show a less resemblance timbre with respect to the original signals, because the set of isolated partials is decreasing in number of elements and therefore, in the reconstruction, the information used is smaller and less varied. Regarding the results of numerical quality, the SDR and SAR parameters descend with respect to the shown results from polyphony 5, while the SIR parameter, although it has a clear downward trend, remains high.

To develop a complete source separation algorithm, several improvements are needed.

First, it is necessary to implement this technique into an algorithm frame-to-frame to address the separation of long duration signals. The fundamental frequency, onset, and offset estimation algorithms presented in "Separation algorithm" section and [38] are able to work dynamically, obtaining the parameters of pitch, starting, and ending time of each note present in the mixture.

There are several useful techniques to properly assign each separated note to its corresponding source. For example, to use a rough estimation of the pitches of the mixture [22] or the score of the analyzed signal. Other possibility is to develop an algorithm of timbre classification. This method has the advantage of maintaining the blindness of the system, but the drawback of a potential loss of generality. Both methods could also be used to solve the limitation of the presented technique for the separation of polyphonic instruments.

Finally, as discussed briefly in "Main characteristics, advantages, and limitations" section, the appearance of completely overlapping notes is statistically inevitable in real recordings. This problem (one of the core problems in BASS) must be addressed to develop a complete separation algorithm. Therefore, future challenges remain to be tackled.

Endnotes

[a]Each original archive consists of a certain number of notes. Each note is approximately 2-s long and is immediately preceded and followed by ambient silence. The instruments are recorded in an anechoic chamber. Some instruments are recorded with and without vibrato. All samples are in mono, 16 bit, 44.1 kHz, AIFF format. Resampled at 16 bits, 22.05 kHz, wav format, excerpts consist of isolated notes. Some of these notes have synthetically been mixed. [b]We will suppose $\Delta\phi_0 = 0$ in Equation (20), but in fact an aleatory initial phase can be inserted without any significant difference in either the numeric or in the acoustical results.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This study was supported by the Spanish government project TEC2009-14414-C03-01 (*Analysis, Classification and Separation of Sound Sources, AnclaS^3 v2.0*). Many thanks to the reviewers for their insightful comments.

Received: 20 May 2011 Accepted: 2 July 2012
Published: 16 October 2012

References

1. S Rickard, *Blind Speech Separation, chapter 8. The DUET Blind Source Separation Algorithm* (Springer, Netherlands, 2007), pp. 217–241
2. T Melia, Underdetermined blind source separation in echoic environments using linear arrays and sparse representations. Ph.D. thesis, School of Electrical, Electronic and Mechanical Engineering University College Dublin, National University of Ireland, 2007
3. M Cobos, JJ Lrópez, Stereo audio source separation based on time-frequency masking and multilevel thresholding. Dig. Signal Process. **18**, pp. 960–976 (2008)
4. M Cobos, Application of sound source separation methods to advanced spatial audio systems. Ph.D. thesis, Universidad Politrécnica de Valencia, 2009
5. Ö Yilmaz, S Rickard, Blind separation of speech mixtures via time-frequency masking. IEEE Trans. Signal Process. **52**(7), pp. 1830–1847 (2004)
6. AS Bregman, *Auditory Scene Analysis: The perceptual organization of sound* (MIT Press, Boston, 1990)
7. GJ Brown, M Cooke, Computational auditory scene analysis. Comput. Speech Lang. Elsevier. **8**(4), pp. 297–336 (1994)
8. D Wang, GJ Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley-IEEE Press, Hoboken, 2006)
9. G Cauwenberghs, in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems-ISCAS '99,* Monaural separation of independent acoustical components. vol. 5, (Orlando, Florida, USA),1999, pp. 62–65
10. S Amari, JF Cardoso, Blind source separation—semiparametric statistical approach. IEEE Trans. Signal Process. **45**(11), pp. 2692–2700 (1997)
11. JF Cardoso, Blind signal separation: statistical principles. Proc. IEEE. **86**, pp. 2009–2025 (1998)
12. MA Casey, W Westner, in *Proceedings on International Computer Music Conference,* Separation of mixed audio sources by independent subspace analysis. vol. 2000, (Berlin, Germany, 2000), pp. 1–8
13. MG Jafari, SA Abdallah, MD Plumbey, ME Davies, Sparse coding for convolutive blind audio source separation. Lecture Notes in Comput. Sci.-Independent Component Anal. Blind Signal Sep. **3889**, pp. 132–139 (2006)
14. SA Abdallah, Towards music perception by redundancy reduction and unsupervised learning in probablistic models. PhD thesis, King's College London, 2002
15. T Virtanen, Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. IEEE Trans. Audio Speech Lang. Process. **15**(3), pp. 1066–1074 (2007)
16. NM Schmidt, M Mørup, in *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation, ICA'06,* Nonnegative matrix factor 2-D deconvolution for blind single channel source separation. Lecture Notes in Computer Science, vol. 3889, (Charleston, SC, USA), 2006, pp. 700–707
17. DD Lee, HS Seung, Learning the parts of objects by nonnegative matrix factorization. Nature. **401**, pp. 788–791 (1999)
18. MN Schmidt, RK Olsson, in *Internationnal Conference on Spoken Languaje Processing, ICSLP'06,* Single-channel speech separation using sparse non-negative matrix factorization. (Pittsburgh, Pennsylvania, USA) 2006, pp. 2614–2617
19. T Virtanen, A Klapuri, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP '00,* Separation of harmonic sound sources using sinusoidal modeling. vol. 2, (Istanbul, Turkey) 2000, pp. 765–768
20. T Virtanen, Sound source separation in monaural music signals. PhD thesis, Tampere University of Technology, 2006
21. MR Every, JE Szymanski, Separation of synchronous pitched notes by spectral filtering of harmonics. IEEE Trans. Audio, Speech Lang. Process. **14**(5), pp. 1845–1856 (2006)
22. Y Li, J Woodruff, D Wang, Monaural musical sound separation based on pitch and common amplitude modulation. Trans. Audio, Speech Lang. Process. **17**(7), pp. 1361–1371 (2009)
23. J Woodruff, Y Li, D Wang, in *Proceedings of the International Conference on Music Information Retrieval,* Resolving overlapping harmonics for monaural musical sound separation using pitch and common amplitude modulation. (Philadelphia, Pennsylvania, USA), 2008, pp. 538–543
24. G Hu, D Wang, Monaural speech segregation based on pitch tracking and amplitude modulation. IEEE Trans. Neural Netw. **15**(5), pp. 1135–1150 (2004)
25. JJ Burred, T Sikora, in *Fifth International Conference on Information, Communications and Signal Processing, ICICS05,* On the use of auditory representations for sparsity-based sound source separation. (Bangkok, Thailand) 2005, pp. 1466–1470
26. TW Parsons, Separation of Speech from interfering speech by means of harmonic selection. J. Acoust. Soc. Am. **60**(4), pp. 911–918 (1976)
27. T Virtanen, A Klapuri, in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics,* Separation of harmonic sounds using multipitch analysis and iterative parameter estimation. (New Paltz, NY, USA) 2001, pp. 83–86
28. A Klapuri, in *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'01),* Multipitch estimation and sound separation by the spectral smoothness principle. vol. 5, (Salt Lake City, Utah, USA) 2001, pp. 3381–3384
29. A Klapuri, Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. IEEE Trans. Speech Audio Process. **11**(6), pp. 804–816 (2003)
30. J Han, B Pardo, in *Proceedings of the IEEE International Conference on Acoustics, Speech ans Signal Processing (ICASSP'11),* Reconstructing completely overlapped notes from musical mixtures. (Evanston, IL, USA) 2011, pp. 249–252
31. Z Duan, Y Zhang, C Zhang, Z Shi, Unsupervised single-channel music source separation by average harmonic structure modeling. IEEE Trans. Audio Speech Lang. Process. **16**(4), pp. 766–778 (2008)
32. JR Beltrán, J Ponce de León, Estimation of the instantaneous amplitude and the instantaneous frequency of audio signals using complex wavelets. Signal Process. **90**(12), pp. 3093–3109 (2010)
33. I Daubechies, Ten lectures on wavelets, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, CBMS-NSF Series Appl. Math., SIAM, (Pasadena, California, USA), 1992
34. JR Beltrán, J Ponce de León, in *Proc. of the 118th Convention of the Audio Engineering Society (AES'05),* Analysis and synthesis of sounds through complex bandpass filterbanks. (Preprint 6361), (Barcelona, Spain), May 2005
35. JR Beltrán, J Ponce de León, in *Proceedings of the XX Simposium Nacional de la Unión Científica Internacional de Radio (URSI'05),* Extracción de Leyes de Variación Frecuenciales Mediante la Transformada Wavelet Continua Compleja. (Valencia, Spain), 2005
36. B Boashash, Estimating and interpreting the instantaneous frequency of a signal. Part 1: fundamentals. Proc. IEEE. **80**(4), pp. 520–538 (1992)
37. JR Beltrán, J Ponce de León, in *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09),* Blind source separation of monaural musical signals using complex wavelets. (Como, Italy) 2009, pp. 353–358
38. JR Beltrán, J Ponce de León, N Degara, A Pena, in *Proceedings of the XXIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI'08),* Localización de Onsets en Señales Musicales a través de Filtros Pasobanda Complejos. (Madrid, Spain), 2008
39. R Gribonval, E Vincent, C Févotte, L Benaroya, in *Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA),* Proposals for performance measurement in source separation. (Nara, Japan), 2003, pp. 763–768
40. C Pérez-Sancho, D Rizo, JM Illescas, Genre classification using chords and stochastic language models. Connection Sci. **21**(2-3), pp. 145–159 (2009)
41. L Fritts, *Electronic Music Studios*. University of Iowa, Musical Instrument Samples Database, http://theremin.music.uiowa.edu/MIS.html, [Online]
42. LI Ortiz-Berenguer, FJ Casajús-Quirós, M Torres-Guijarro, JA Beracoechea, in *Proceedings of the 7th Conference on Digital Audio Effects (DAFx'04),* Piano

transcription using pattern recognition: aspects on parameter extraction. (Naples, Italy), 2004, pp. 212–216
43. Ll Ortiz-Berenguer, Identificacirón Automrática de Acordes Musicales. PhD thesis, Escuela Trécnica Superior de Ingenieros de Telecomunicacirón, Universidad Politrécnica de Madrid, 2002
44. C Févotte, R Gribonval, Vincent E, *BSS EVAL Toolbox User Guide—Revision 2.0*. Technical Report, IRISA Technical Report 1706, Rennes, France, 2005
45. E Vincent, R Gribonval, C Févotte, Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**(4), pp. 1462–1469 (2006)

doi:10.1186/1687-6180-2012-223
Cite this article as: Ponce de León and Beltrán: **Blind separation of overlapping partials in harmonic musical notes using amplitude and phase reconstruction.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:223.