MULTI-CHANNEL MONO-PATH PERIODIC SIGNAL EXTRACTION WITH GLOBAL AMPLITUDE AND PHASE MODULATION FOR MUSIC AND SPEECH SIGNAL ANALYSIS

*Mahdi Triki, Dirk T.M. Slock**

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Email: {triki,slock}@eurecom.fr

ABSTRACT

A key building block in music transcription and indexing operations is the decomposition of the music signal into notes. We model a note signal as a periodic signal with (slow) global variation of amplitude (reflecting attack, sustain, decay) and frequency (limited time warping). Also voiced speech admits such a representation. The bandlimited variation of global amplitude and frequency gets expressed through a subsampled representation and parameterization of the corresponding signals. The periodic signal is assumed to arrive at a set of sensors with different amplitude and delay. Assuming additive white Gaussian noise, a Maximum Likelihood approach is proposed for the estimation of the model parameters and the optimization is performed in an iterative (cyclic) fashion that leads to a sequence of simple least-squares problems. Particular attention is paid to the estimation of the basic periodic signal, which can have a non-integer period. Simulation results reveal that the proposed approach allows to extract such signals accurately from an underdetermined mixture of several, using iterated successive interference cancellation.

1. INTRODUCTION

The majority of blind separation algorithms are based on the theory of Independent Component Analysis. The idea is to estimate the inverse mixing matrix using statistical independence of source signals. However, one area of research in Blind Source Separation, the Underdetermined BSS, is relatively untouched. It refers to the case when there are less mixtures than sources. The underdetermined BSS poses a challenge because the mixing matrix is not invertible and the traditional ICA methods does not work. And, contrary to most blind separation algorithms, the source extraction itself requires additional assumptions on the source statistics or structure. Several approaches in the literature are proposed to solve the problem exploiting essentially the time-frequency sparcity of the source signals [1, 2]. However, all the proposed solutions perform separation, independently, on each time-frequency frame; and do not take advantage of signal correlation in different frames.

On the other hand, Sinusoidal model based speech/music analysis/synthesis has received considerable interest in the signal processing community [3, 4, 5]. The sinusoidal transform represents a signal as a sum of discrete time-varying sinusoids or partials. The estimation of the model parameters is typically carried out using a short-time Fourier transform (STFT) with a fixed analysis frame size and a fixed stride between frames. The sinusoids are extracted by peak-picking in the STFT magnitude spectrum. Intermediate values are obtained by interpolation. A fundamental problem faced by the traditional sinusoidal-model based techniques is that, since the sppech/music signal is strongly non-stationary , it is not always possible to find a good tradeoff between time and frequency resolution. Another drawback of these techniques is that they ignore the harmonic structure of the music signal. For treating periodic signals, the state of the art is limited to the estimation of pure periodic signals with period equal to an integer number of samples [6, 7]. In these references, the authors propose a Maximum Likelihood approach to analyze pure periodic signals. They show that the resulting procedure can be interpreted as a signal projection onto suitable subspaces. In [8], we extend the results of those references, and we try to merge the modulated sinusoidal modeling and the periodic signal analysis techniques, by considering periodic signals with non-integer period and global amplitude variation and time warping. And, we show that the previous model gives a good tradeoff between modeling and estimation noise.

In all cases, all the previous references treat instantaneous mixtures and ignore propagation environment. In this paper, we assume a mono-path propagation environment, and we will focus on the underdetermined convolutive audio signal separation problem.

*The Eurecom Institute's research is partially supported by its industrial members: Bouygues Télécom, Fondation d'entreprise Groupe Cegetel, Fondation Hasler, France Télécom , Hitachi, Sharp, ST Microelectronics, Swisscom, Texas Instruments, Thales.

2. SIGNAL MODEL

In the sinusoidal modeling, the signal is modeled as a sum of evolving sinusoids:

$$s(t) = \sum_{k=0}^{P} A_k(t) \cos\left(\theta_k(t)\right) \quad . \tag{1}$$

where $\theta_k(t)$ represents the instantaneous phase of the k^{th} partial. As the music signal is quasi-periodic, $\theta_k(t)$ can be decomposed into

$$\theta_k(t) = 2\pi k t f_0 \; + \; 2\pi\varphi_k(t) \tag{2}$$

where $\varphi_k(t)$ characterizes the evolution of the instantaneous phases around the k^{th} harmonic; and can be assumed to be low-frequency.
The Global Modulation assumption implies that all harmonic amplitudes evolve proportionally in time; and that the instantaneous frequency of each harmonic is proportional to the harmonic index:

$$\begin{cases} A_k(n) = A_k \; A(n) \\ 2\pi\varphi_k(n) = 2\pi k \; \varphi(n) \; + \; \Phi_k \end{cases} \quad . \tag{3}$$

In summary, we model an audio signal as the superposition of harmonic components with a global amplitude modulation and time warping (that can be interpreted in terms of phase variations):

$$\begin{aligned} y(n) \; &= \; s(n) \; + \; v(n) \\ &= \; \textstyle\sum_k A_k(n) \; \cos\left(2\pi k n f_0 + 2\pi\varphi_k(n)\right) + v(n) \\ &= \; A(n) \textstyle\sum_k A_k \cos\left(2\pi k f_0 \left(n + \frac{\varphi(n)}{f_0}\right) + \Phi_k\right) + v(n) \end{aligned}$$

where

- v_n is an additive white Gaussian noise.
- $A(n)$ represents the amplitude modulating signal. It allows an evolution of the note power, reflecting attack, sustain, and decay.
- $\varphi(n)$ denotes the phase modulating signal (that can be interpreted in terms of time warping). The time warping focuses on the time evolution of the instantaneous frequency, and allows the modeling of several musical phenomena (vibrato, glissando ...)

In [8], we have expressed the time warping in term of interpolation operation over a basic periodic signal. In sum, the audio signal can be written as:

$$Y = \underbrace{A \, F \theta}_{= S} + V \tag{4}$$

where :
- $Y = [y(1) \cdots y(N)]^T$, represents the observation vector.
- $S = [s(1) \cdots s(N)]^T$, represents the signal of interest.
- $V = [v(1) \cdots v(N)]^T$, denotes the noise vector.
- $\theta = [\theta(1) \cdots \theta(\lceil T \rceil)]$, characterizes the harmonic signature over essentially one period
- $A = diag[A(1) \cdots A(N)]$, represents the global amplitude modulation signal.
- F is an $N \times \lceil T \rceil$ interpolation matrix characterizing the time warping. See [8] for a detailed description.

3. AUDIO SEPARATION IN MONO-PATH ENVIRONMENT

As a first approximation to the propagation environment, we use the delay-mixing model. In this model, only direct path signal components are considered. Signal components from one source arrive with a given attenuation and a fractional delay between the time of arrivals at two receivers. By fractional delays, we mean that delays between receivers are not generally integer multiples of the sampling period. The signal attenuation and delay depend on the position of the source with respect to the receiver axis and the distance between receivers. Under the previous propagation assumptions, observations can be written as:

$$\begin{aligned} y_1(t) &= \sum_{i=1}^{L} s_i(t) + v_1(t) \\ y_k(t) &= \sum_{i=1}^{L} \beta_{ki} s_i(t - \tau_{ki}) + v_k(t) \qquad k = 2 : M \end{aligned}$$

where $\{s_i(t)\}_{i=1:L}$ represent L distinct audio source signals following (4)(with distinct periodicities); $\{v_k(t)\}_{k=1:M}$ a spatially and temporally white gaussian noise signals; β_{ki} the relative attenuation of the i^{th} source at the k^{th} sensor; and τ_{ki} the propagation delay (function of the direction of arrival ϕ_i, and the sensor geometry (that we suppose fix but unknown)).

As in [8], the time delay operation can be expressed using an interpolation matrix H_τ (as it can be interpreted as a particular time warping):

$$S_\tau = H_\tau S$$

where H_τ is an $N \times N$ toeplitz, band matrix characterizing the time delay operation.

Thus, the total observation vector can be written as

$$Y = HS + V \tag{5}$$

where
- $Y = [Y_1^T \cdots Y_M^T]^T$, is a $MN \times 1$ vector representing the observation vector
- $S = [S_1^T \cdots S_L^T]^T$, $NL \times 1$ vector representing the signals of interest.
- $V = [V_1^T \cdots V_M^T]^T$, is a $MN \times 1$ vector denoting the noise vector

- $H = \begin{bmatrix} I_{N\times N} & \cdots & I_{N\times N} \\ \beta_{2,1}H_{\tau_{2,1}} & \cdots & \beta_{2,L}H_{\tau_{2,L}} \\ \vdots & & \vdots \\ \beta_{M,1}H_{\tau_{M,1}} & \cdots & \beta_{M,L}H_{\tau_{M,L}} \end{bmatrix}$ is an $NM \times NL$ interpolation matrix characterizing the propagation environment.

The previous model is linear in H, and S (separately); H, and S, being parameterized nonlinearly. Trying to estimate all factors jointly is a difficult nonlinear problem. Indeed, as the noise is assumed to be a white Gaussian signal, the ML approach leads to the following least-squares problem:

$$\min_{H,S} \|Y - HS\|^2 \tag{6}$$

The estimation can easily be performed iteratively though.

3.1. Channel Estimation

Under the current estimate of the source signal $\widehat{S}$, the Channel coefficients are optimized using

$$\min_{\tau_{ki},\beta_{ki}} \left\| Y - H\widehat{S} \right\| \tag{7}$$

On the other hand,

$$\left\| Y - H\widehat{S} \right\|^2 = \sum_{k=1}^{M} \left\| Y_k - \sum_{i=1}^{L} \beta_{ki} H_{\tau_{ki}} \widehat{S}_i \right\|^2$$

and,

$$\left\| Y_k - \sum_{i=1}^{L} \beta_{ki} H_{\tau_{ki}} \widehat{S}_i \right\|^2 = \|Y_k\|^2 + \sum_{i=1}^{L} \beta_{ki}^2 \left\| \widehat{S}_{\tau_{ki}} \right\|^2 - \sum_{i=1}^{L} \beta_{ki} \widehat{R}(y_k, s_{\tau_{ki}}) - \sum_{i\neq j} \beta_{ki}\beta_{kj} \widehat{R}(s_{\tau_{ki}}, s_{\tau_{kj}})$$

where $\widehat{S}_{\tau_{ki}} = H_{\tau_{ki}}\widehat{S}_i$ denotes the estimate of the i^{th} source delayed by τ_{ki}; and $\widehat{R}(x,y) = \frac{1}{N}\sum_{j=1}^{N} x(j)y(j)$ represents the estimate of the correlation between signals x and y.

Note that the quantities $\widehat{R}\left(s_{\tau_{ki}}, s_{\tau_{kj}}\right) \;\; i \neq j$ can be neglected, as source signals are assumed independent, having different periodicities. Then, the optimization problem in (7) is separable; and can be solved, independently, for each channel parameter.

The optimization over a given time-lag τ_{ki} can be interpreted in terms of maximizing the correlation $\widehat{R}\left(y_k, s_{\tau_{kj}}\right)$ between the observed signal on the sensor and the i^{th} source signal delayed by τ_{ki} .

$$\tau_{ki} = \arg\max_{\tau_{ki}} \widehat{R}\left(y_k, s_{\tau_{kj}}\right) \tag{8}$$

Ones the different time lags are estimated, the optimal attenuation coefficients are computed using:

$$\widehat{\beta}_{ki} = \frac{\widehat{R}(y_k, s_{\tau_{ki}})}{\left\| \widehat{S}_{\tau_{ki}} \right\|^2} \tag{9}$$

3.2. Source Signal Estimation

If we assume that the channel parameters known, the ML source estimation is given by:

$$\widehat{\widehat{S}} = H^{\#}Y$$

where $H^{\#} = \left(H^T H\right) H^T$ denotes the pseudoinverse of H.

On the other hand, the source signal is supposed to be a pseudo-periodic signal (as in (4)). Thus, it can be written as

$$\widehat{\widehat{S}}_i \;=\; \widehat{A}_i \widehat{F}_i \widehat{\theta}_i + V_i \;=\; \widehat{S}_i + V_i \qquad i = 1 : L$$

where $\widehat{A}_i$, $\widehat{F}_i$, and $\widehat{\theta}_i$ are estimated in an iterative (cyclic) fashion (as in [8]) from $\widehat{\widehat{S}}_i$:

3.2.1. Periodic Signature Estimation

If we assume that the matrices $\widehat{A}_i, \widehat{F}_i$ are given, the periodic signature θ_i can be isolated as

$$\widehat{\widehat{S}}_i = \widehat{A}_i \, \widehat{F}_i \, \theta_i + V_i = G_i \, \theta_i + V_i \tag{10}$$

Then minimizing (10) w.r.t. θ leads to

$$\widehat{\theta}_i = \left(G_i^T G_i\right)^{-1} G_i^T \widehat{\widehat{S}}_i \;\; . \tag{11}$$

Hence the periodic signature gets estimated by using the data over the whole note duration.

3.2.2. Instantaneous Amplitude Estimation

The instantaneous amplitude gets estimated based on the noisy data and noise energies estimation. By assuming the instantaneous amplitude be piecewise constant, $A_i(n)$ gets estimated using:

$$\widehat{A}_i(n) = \sqrt{\frac{1}{\frac{}{\theta_i^2}} \left\langle \widehat{\widehat{s}}_i^2(n) - \left(\widehat{\widehat{s}}_i(n) - \widehat{s}_i(n)\right)^2 \right\rangle_n} \tag{12}$$

where $\langle \, . \, \rangle_n$ denotes temporal averaging over the piecewise interval containing n; $\widehat{S}_i = \widehat{A}_i \widehat{F}_i \widehat{\theta}_i$ denotes the latest estimate of the signal of interest.

3.2.3. Instantaneous Frequency Estimation

As for the instantaneous amplitude, the instantaneous frequency gets estimated on a frame-by-frame basis. In each frame, the instantaneous frequency is optimized using (10):

$$\begin{cases} \min\limits_{f} \left\| \widehat{\widehat{S}}_i - \widehat{A}_i \widehat{F}_i(f) \widehat{\theta}_i \right\| \\ \frac{\Delta f}{f_0} \leq \alpha_{max} \end{cases} \tag{13}$$

where Δf denotes the maximum relative frequency variation in the current frame compared to the previous frame, reflecting an assumed limited frequency variation rate. The optimal instantaneous frequency value for the current frame gets determined from a finite set of discrete values within the thus limited range.

4. A ISIC IMPLEMENTATION FOR THE AUDIO SOURCE SEPARATION TECHNIQUE

In previous, we have proposed an audio separation scheme tacking into account simultaneously the source signal structure and the propagation environment model. The inherent complexity, however, is cubic on MN (as the technique requires the inversion of the non Toeplitz matrix H). For practical implementation, Iterated Successive Interference Cancellation (ISIC) approach can been used to approximate the previous technique; with only a linear complexity.

Iterated Successive Interference Cancellation is a non-linear type of parameters estimation scheme in which parameters are estimated successively. The approach successively cancels concurrent parameters using their current estimate. The ISIC audio separation algorithm appears in the table below.

Iterated SIC Multichannel Audio Source Separation
Computation
Initialization

for $i = 1 : L$ do
$s_i \leftarrow$ Periodic Source Extraction(y_1, T_i)
for $k = 1 : M$ do
$\tau_{ki} = \arg\max_{\tau_{ki}} \widehat{R}(y_k, s_{\tau_{ki}})$
$\beta_{ki} = \frac{\widehat{R}(y_k, s_{\tau_{ki}})}{\|s_{\tau_{ki}}\|^2}$
end for
end for

Iteration

for $i = 1 : L$ do
Interference Cancellation
for $k = 1 : M$ do
$y_k \leftarrow y_k - \sum_{p \neq i} \beta_{kp}\, s(t - \tau_{kp})$
end for
Channel Estimation
for $k = 1 : M$ do
$\tau_{ki} = \arg\max_{\tau_{ki}} \widehat{R}(y_k, s_{\tau_{ki}})$
$\beta_{ki} = \frac{\widehat{R}(y_k, s_{\tau_{ki}})}{\|s_{\tau_{ki}}\|^2}$
end for
Non parametric source estimation
$s_i(t) \leftarrow \frac{1}{\sum_k \beta_{ki}^2} \sum_k \beta_{ki} y(t - \tau_{ki})$
parametric source estimation
$s_i \leftarrow$ Periodic Source Extraction(s_i, T_i)
end for

cost/update ($2p \div$): $O(MN)$ $(\times)$

Table 1. Iterated SIC Multichannel Audio Source Separation

Note that the non parametric source estimation (computed using a simple matched filter) can be interpreted as a delay and sum beamformer. It leads then to a second level of interference cancellation.

Using the proposed approach, we perform separation using a single musical record. The proposed signal represents a synthesized mixture of three notes played by an acoustic guitar. The record has a duration of 1s and is sampled at 22.050 kHz (see figure 1). Their pitch frequencies are respectively 82 Hz, 92 Hz, 116 Hz. The SNR of the input signal is 26 dB.

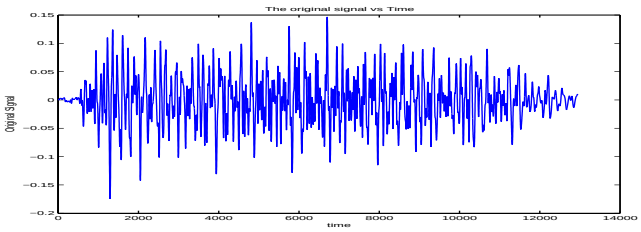

Fig. 1. Original guitar signal

In figure 2, we plot the different outputs of the algorithm concerning the note 1: original signal ($s_1(t)$), synthesized signal ($\widehat{s}_1(t)$) according to the global amplitude and frequency modulation model, signal error ($v_1(t)$) (difference between the previous two), and instantaneous amplitude signal $\widehat{A}_1(n)$.

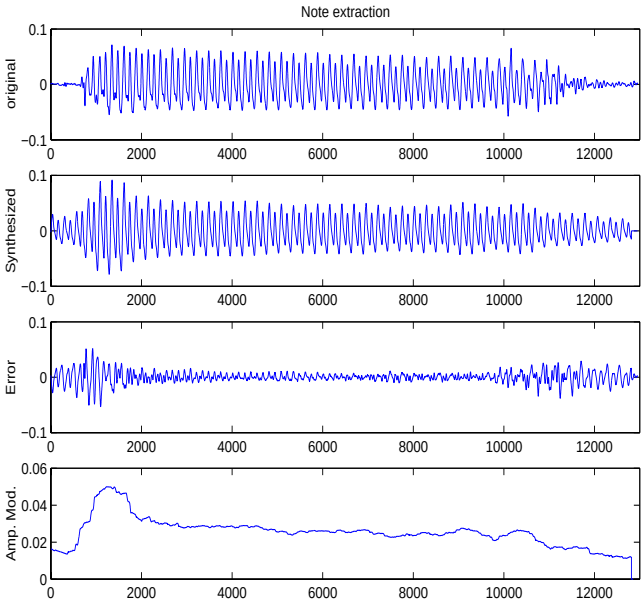

Fig. 2. "Note 1" Extracted parameters.

In order to analyze the extraction quality of our algorithm, we plot the Fast Fourier transform of the original signals (S_i) and the residual error signals ($\widehat{V}_i$) (figure 3).

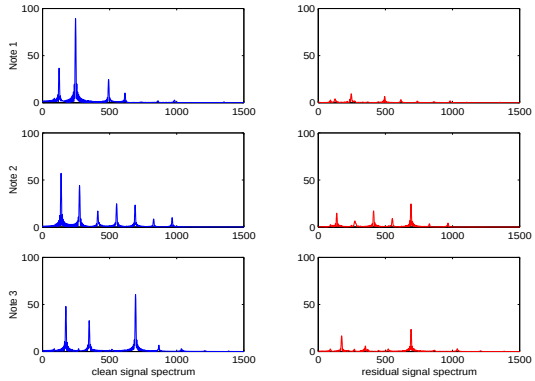

Fig. 3. The FFT of the original and residual error signals for the notes 1, 2, and 3.

We also calculate the signal to (measurement plus approximation) noise ratio for the estimated model (for the total note duration, and for the steady-state region). Results are summarized in the table 2.

	Total SNR (in dB)	SNR for steady-state portion only (in dB)
Note 1	8.2	16.8
Note 2	3.5	7.3
Note 3	4.4	8.9

Table 2. Total and steady-state SNR.

We see that, even using a single mixture, the extraction performances on the steady state portions are quite good: taking only periodicity into account is sufficient to perform separation. However, on the attack and decay portions, the extraction technique achieves poor performances. In fact, as the different note signals are asynchronous (don't start and finish at the same time), amplitude estimation based on a simple energy detection fails in tracking the note variations.

5. A MODIFIED MULTICHANNEL AUDIO SOURCE EXTRACTION TECHNIQUE

The reason for which the previous algorithm fails to track instantaneous amplitude variations is that amplitude estimation relies only on energy considerations. If only one audio source is present in the mixture, in presence of an additive stationary noise (signal enhancement problem), such approach is sufficient to extract the audio source. However, if more than one source is present (and/or the noise is non-stationary), instantaneous amplitude estimation should take into account some additional information (such as the signal periodicity, the harmonic signature...).
In this paper, we suggest using a Least Squares estimator for the instantaneous amplitude estimation.

In fact, if we assume that $(\widehat{F}_i, \widehat{\theta}_i)$ are given, the audio source estimate $\widehat{\widehat{S}}_i$ can be written as

$$\begin{aligned} \widehat{\widehat{S}}_i &= A_i \widehat{F}_i \widehat{\theta}_i + V_i \\ &= \breve{G}_i A_i + V_i \end{aligned}$$

On the other hand, the instantaneous amplitude (A_i) is supposed to be lowpass band. Then, it can be down-sampled. The remaining samples can be estimated using linear interpolation. Linear interpolation can be formalized as a linear transform:

$$\begin{bmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ \vdots \\ A_{iN} \end{bmatrix} = \begin{bmatrix} 1 & 0 \cdots & 0 \\ P_{21} & P_{22} \cdots & 0 \\ P_{31} & P_{32} \cdots & 0 \\ 0 & 1 \cdots & 0 \\ & \vdots & \\ 0 & \cdots \cdots & 1 \end{bmatrix} \begin{bmatrix} a_{i1} \\ \vdots \\ a_{ip} \end{bmatrix} = P\, a_i$$

Where $\{a_{ij}\}_{j=1:p}$ denote the freedom degrees of our model; and P represents the interpolation matrix.

In our simulation, we propose using Hamming window for linear interpolation. In fact, the linear interpolation can be interpreted as a linear filtering operation of the upsampled signal with a low-pass filter (triangular for linear interpolation, rectangular for nearest-neighbor interpolation...). By using a smooth window (with energy concentrated essentially in the principal lobe), the estimation error gets amplified less. Thus, we do better estimation.

We can also vary the interpolation window length with time. In fact, in the transient state the instantaneous amplitude is much more large-band than in harmonic steady-state; the fact that allows using larger windows (then increase estimation performance).

In sum, the estimation problem can be formalized as follow

$$\widehat{\widehat{S}}_i = \left(\breve{G}_i P\right) a_i + V_i$$

and, $\hat{A}_i$ gets estimated using the least-squares technique (via the $\hat{a}_i$).
The main drawback of this technique is its non robustness to the initialization. In fact, the estimation of the instantaneous amplitude relies strongly on the quality of estimation of the periodic signal θ_i. So, we suggest using the first version (with the non-coherent amplitude estimation) for the algorithm initialization.

We simulate the previous audio mixture using this modified version and we plot the different algorithm outputs (concerning the note 1) on figure 4. We observe that the algorithm achieves better performances; and that was able to detect the begin and the end of the musical note. Figure 5

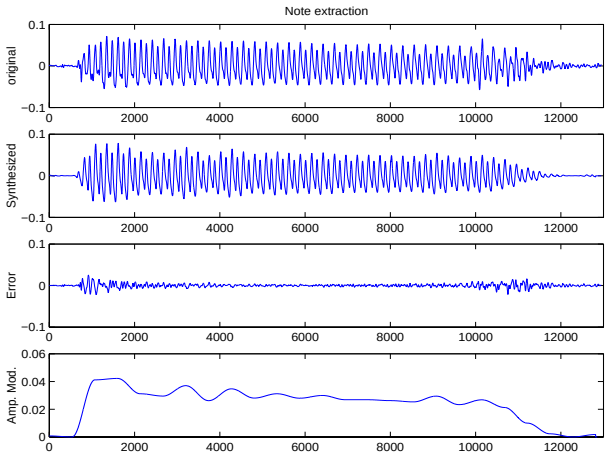

Fig. 4. "Note 1" Extracted parameters.

shows curves of the estimation SNR (for the total note duration, and for the steady state portion). Once again, we observe that the second version achieves better performances (not only on transition regions, but also on steady state region).

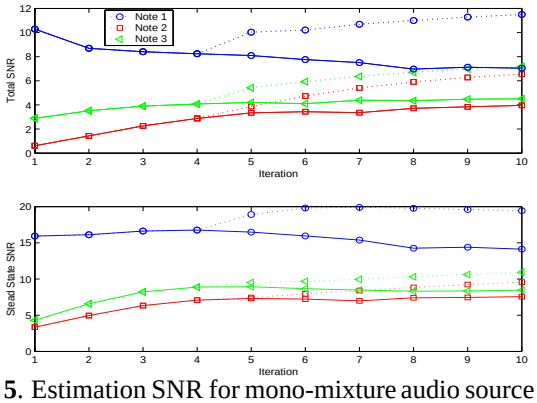

Fig. 5. Estimation SNR for mono-mixture audio source separation (version 1 on solid line, modified version on dotted line).

We consider now the Multi-Input Multi-Output problem (figure 6). We assume that the audio source signals are captured by two microphones (spaced by $d = 0.2m$). The angles of arrival of the three source signals are respectively $\theta_1 = -\frac{\pi}{3}$, $\theta_2 = 0$, and $\theta_3 = +\frac{\pi}{3}$. The relative attenuations at the second microphones are respectively $\beta_{21} = 0.9$, $\beta_{22} = 1$, and $\beta_{23} = 1.1$.

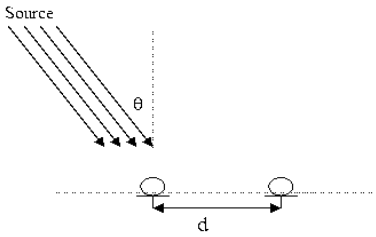

Fig. 6. Multi-Input Multi-Output propagation scenario.

Figure 7 shows curves of the estimation SNR (for the total note duration) for MISO (slide line) and MIMO (dotted line) scenarios. As it was expected, we observe that, ones relative delays and attenuations are well estimated, using multiple output enable algorithm to achieve better performances.

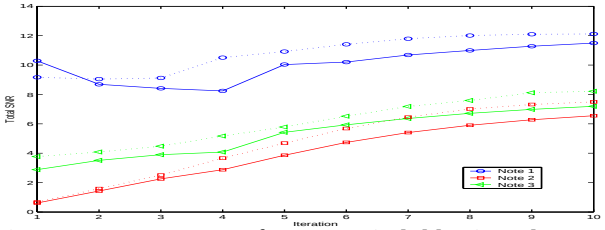

Fig. 7. Estimation SNR for MISO (solid line), and MIMO (dotted line) audio source separation.

6. CONCLUSION

In this paper we have investigated the underdetermined convolutive source separation of audio mixtures. We have considered the periodic signal model with a slow global amplitude and phase variation. We have proposed a separation technique that takes into account simultaneously the source signal structure and the propagation environment model. Simulations show that the extraction technique is suitable for the analysis of musical notes, and produces good auditive synthetic results.

7. REFERENCES

[1] F. Abrard, Y. Deville, and P. R. White."From Blind Source Separation to Blind Source Cancellation in the Underdetermined Case: a new Approach Based on Time-Frequency Analysis," *In Proceedings of ICA*, December 2001.

[2] L-T Nguyen, A. Belouchrani, K. Abed-Meraim, and B. Boashash."Separating more sources than sensors using Time-Frequency Distributions," *In Proceedings of ISSPA*, August 2001.

[3] M. Goodwin, M. Vetterli. "Time-Trequency Signal Models for Music Analysis, Transformation, and Synthesis," *In Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Pages: 133-136, June 1996.

[4] M. Goodwin, M. Vetterli. "Atomic Decompositions of Audio Signals," *In Proceedings of WASPAA*, October 1997.

[5] Yinong Ding, Xiaoshu Qian. "Estimating Sinusoidal Parameters of Musical Tones Based on Global Waveform Fitting," *In Proceedings of MMSP*, Pages:95 - 100, June 1997.

[6] D.D. Muresan, and T.W. Parks. "Orthogonal, Exactly Periodic Supspace Decomposition," *IEEE Transactions on Signal Processing*,Vol. 51, No. 9, September 2003.

[7] J.D. Wise, J.R. Caprio and T.W. Parks. "Maximum Likelihood pitch estimation," *IEEE Transaction on Acoustics, Speech, and Signal Processing* ,Vol. 51, May 1976.

[8] Mahdi Triki, Dirk T.M. Slock. "Periodic Signal Extraction with Global Amplitude and Phase Modulation for Music Signal Decomposition," *In Proceedings of ICASSP*, March 2005.