A Wiener Filter Approach to Microphone Leakage Reduction in Close-Microphone Applications

Elias K. Kokkinis, Joshua D. Reiss, and John Mourjopoulos, *Member, IEEE*

Abstract—Microphone leakage is one of the most prevalent problems in audio applications involving multiple instruments and multiple microphones. Currently, sound engineers have limited solutions available to them. In this paper, the applicability of two widely used signal enhancement methods to this problem is discussed, namely blind source separation and noise suppression. By extending previous work, it is shown that the noise suppression framework is a valid choice and can effectively address the problem of microphone leakage. Here, an extended form of the single channel Wiener filter is used which takes into account the individual audio sources to derive a multichannel noise term. A novel power spectral density (PSD) estimation method is also proposed based on the identification of dominant frequency bins by examining the microphone and output signal PSDs. The performance of the method is examined for simulated environments with various source–microphone setups and it is shown that the proposed approach efficiently suppresses leakage.

Index Terms—Microphone leakage, multichannel audio enhancement, noise suppression, power spectral density (PSD) estimation, source separation, Wiener filter.

I. INTRODUCTION

THE production of modern music often involves a number of musicians performing together inside the same room with a number of microphones set to capture the sound emitted by their instruments. Ideally, each microphone should pick up only the sound of the intended instrument, but due to the interaction between the various instruments and room acoustics, each microphone picks up not only the sound of interest but also a mixture of all other instruments. This is known as *microphone leakage* and is an undesirable effect, common in most multiple microphone multiple instrument applications (see Fig. 1). The *close-microphone* technique, in which the microphone is placed in close proximity to the source of interest, is typically used in order to enable the microphone to capture as much of the sound of interest as possible (i.e., increase the signal to noise ratio) [1] and reduce the effect of microphone leakage. It is also used to minimize the effect of room acoustics on the received signal
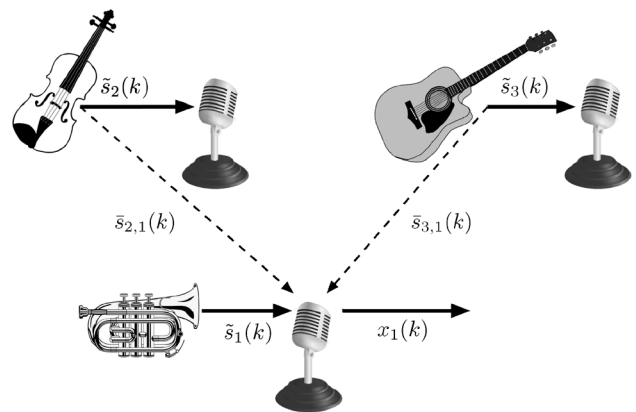

Fig. 1. Illustration of the microphone leakage effect for close microphone applications. The leakage for only one microphone is shown, for the case of three sources and three microphones.

(i.e., increase the direct to reverberant ratio) [2] in cases where the room acoustic properties are poor or where the sound engineer would like to later add artificial reverberation.

Even with the close-microphone technique, microphone leakage is difficult to eliminate, especially in live sound applications where the acoustic environment and the placement of instruments and microphones are far less controlled than in a recording studio. It is therefore reasonable to consider the introduction of advanced signal processing frameworks to address this problem.

One possible approach would be the use of the blind source separation (BSS) framework. BSS methods are attractive for audio applications since they treat the mixing process as a "black box" and do not require access to the original source signals [3]. However, a number of problems arise when considering the application of BSS methods to audio. First, some of the most common assumptions in BSS methods, such as statistical independence [4] do not always hold [5]. A more significant problem is reverberation [6] since in many audio applications and especially live sound, reverberation times around or even over 1 second are not uncommon. Combined with the high sampling rates (44.1 kHz or higher) required for preserving audio quality, the room impulse responses (RIRs) describing the mixing system are given by FIR filters with tens of thousands of coefficients. Therefore, BSS methods are required to estimate a set of comparably long filters [7], [8] that will invert the mixing process and produce separated signals. However, such long filters will slow down convergence and increase computational cost [9], [10]. Finally, the output signals of such methods are typically scaled and reordered versions of the original source signals. While the permutation problem can

E. K. Kokkinis and J. Mourjopoulos are with the Audio and Acoustic Technology Group, Department of Electrical and Computer Engineering, University of Patras, 26504, Patras, Greece (e-mail: ekokkinis@upatras.gr; mourjop@upatras.gr).

J. D. Reiss is with the Centre for Digital Music, Department of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: josh.reiss@elec.qmul.ac.uk).

be easily addressed in the case of close-microphone applications, the problem of scaling, especially in live sound, may lead to significant problems in the audio gain structure resulting in feedback [11] and/or distortion.

For all the above reasons, the alternative noise suppression framework seems a more plausible approach. This is because in practice the microphone signal consists of a signal of interest corrupted by additive noise, which in this case is the sound from all other audio sources. Furthermore noise suppression does not usually require any information or assumptions about room acoustics, although such knowledge could prove beneficial in some applications. The Wiener filter is one of the most common methods employed within this framework [12], [13]. The main issue here is the estimation of the noise signal properties and several approaches have been proposed to accomplish this [14]–[17]. More recently, multichannel Wiener filters have been proposed [18], [19] that assume the use of microphone arrays and exploit the spatial properties of noise. However, apart from the use of arrays with specific geometries, such methods assume a single source inside a noise field, rather than several sources corrupted by noise that consists of the interference between them.

The concept of using Wiener filters for microphone leakage reduction was considered in previous work [1], where it was shown that this approach can effectively address the problem in close-microphone recordings in real environments with two sources/microphones. Here, this concept is extended for an arbitrary number of sources and microphones. An extension of the single channel Wiener filter is derived, leading to a multichannel Wiener filter in the sense that the noise term depends on several interfering signals. A time–frequency domain implementation is used, where the power spectral densities (PSDs) involved are estimated from the microphone and output signals based on the identification of dominant frequency bins and an iterative procedure controlled by an energy-adaptive forgetting factor. The results presented show that the proposed method achieves satisfactory performance, effectively reducing microphone leakage even at long source–microphone distances, while being robust with respect to the number of sources and reverberation time.

The rest of the paper is organized as follows. In Section II, a straightforward extension of the single-channel Wiener filter is given while in Section III the PSD estimation method is described. In Section IV, simulation results are presented for two different source–microphone setups and various parameters and finally in Section V some conclusions are drawn from the analysis of the results.

II. PROBLEM FORMULATION AND FILTER DERIVATION

Consider N sources $s_n(k)$ located inside a reverberant enclosed space and M microphones producing the signals $x_m(k)$. Let $h_{mn}(k)$ be the FIR filter that models the response of the acoustic path (namely the RIR) between the nth source and the mth microphone, including microphone properties. Each microphone signal is given by

$$x_m(k) = \sum_{i=1}^{N} [\rho_s(\theta_{mi})s_i(k)] * [\rho_c(\theta_{mi})h_{mi}(k)] \quad (1)$$

for $m = 1, 2, \ldots, M$. θ_{mi} is the angle between the ith source and mth microphone, $\rho_s(\theta)$ is the directivity function of the source and $\rho_c(\theta)$ is the directivity function of the microphone. These functions add a further degree of freedom in source-source and source–microphone interaction and can even prove beneficial in some settings when the angles are suitably chosen by the sound engineer. In this work however, both sources and microphones will be considered omnidirectional, i.e., $\rho_s(\theta) = \rho_c(\theta) = 1$ for all θ. Also the number of microphones is assumed equal to the number of sources ($N = M$).

Since the use of the close-microphone technique is assumed here, then each microphone picks up primarily the sound of the source of interest and to a lesser extent the sound of all other sources. Therefore,

$$x_m(k) = s_m(k) * h_{mm}(k) + \sum_{\substack{i=1 \\ i \neq m}}^{M} s_i(k) * h_{mi}(k) \quad (2)$$

for $m = 1, 2, \ldots, M$. Now define the *direct* source as

$$\tilde{s}_m(k) = s_m(k) * h_{mm}(k) \quad (3)$$

and the *leakage* source as

$$\bar{s}_{i,m}(k) = s_i(k) * h_{mi}(k) \quad (4)$$

with $i \neq m$. The term *source* may refer to the anechoic source signal $s_m(k)$, the direct source $\tilde{s}_m(k)$ or the leakage source $\bar{s}_{i,m}(k)$. Also note that all of the following equations hold for $m = 1, 2, \ldots, M$ and $i \neq m$. Microphone leakage can be expressed as

$$u_m(k) = \sum_{\substack{i=1 \\ i \neq m}}^{M} \bar{s}_{i,m}(k) \quad (5)$$

and the microphone signal can be written as

$$x_m(k) = \tilde{s}_m(k) + u_m(k) \quad (6)$$

Equation (6) now describes microphone leakage as a signal in additive noise problem, where a filter must be calculated that will provide an estimate of the signal of interest. For a fixed m a single-channel Wiener filter can be applied, provided an adequate estimation of the noise term can be obtained.

For the following derivation of the extended Wiener filter, the original sources $s_m(k)$ are assumed to be uncorrelated with each other and to be wide-sense stationary (WSS) processes. While the assumption of uncorrelated sources holds for the case of audio signals [20], the WSS assumption does not hold in practice and will be addressed later in the text [see (15), (16)]. Also note, that due to the linearity of the convolution operation, the uncorrelated sources assumption holds for the direct and leakage sources as well, i.e.,

$$E\left\{\tilde{s}_m(k)\tilde{s}_n^*(k + \tau)\right\} = 0 \quad (7)$$

for $-\infty < \tau < \infty$, $m, n = 1, 2, \ldots, M$, and $m \neq n$. The $*$ superscript denotes complex conjugation. However since all the

signals considered in this work are real, the conjugation can be dropped.

Let $w_m(k)$ be a filter that suppresses the interference and provides an estimate of the source in the mean squared error (MSE) sense. Then, the estimated source is given by

$$\hat{s}_m(k) = \sum_{l=-\infty}^{\infty} w_m(l)x_m(k-l). \tag{8}$$

The infinite sum in (8) implies that $w_m(k)$ is an infinite impulse response (IIR) filter. The error signal between the actual source and the estimated one is

$$e_m(k) = \tilde{s}_m(k) - \hat{s}_m(k). \tag{9}$$

The optimum filter in the MSE sense can be obtained by minimizing

$$\frac{\partial E\left\{|e_m(k)|^2\right\}}{\partial w_m(\tau)} = 0 \quad \forall \tau. \tag{10}$$

Equation (10) after necessary computations becomes

$$E\left\{\tilde{s}_m(k)x_m(k-\tau)\right\} -$$
$$\sum_{l=-\infty}^{\infty} w_m(l)E\left\{x_m(k-l)x_m(k-\tau)\right\} = 0. \tag{11}$$

It is easy to see that (11) involves auto- and cross-correlation functions and can be written as

$$\sum_{l=-\infty}^{\infty} w_m(l)r_{x_m x_m}(\tau - l) = r_{\tilde{s}_m x_m}(\tau) \quad \forall \tau. \tag{12}$$

By invoking the assumption of uncorrelated sources the correlation functions above can be expressed as

$$r_{\tilde{s}_m x_m}(\tau) = r_{\tilde{s}_m \tilde{s}_m}(\tau) \tag{13}$$

$$r_{x_m x_m}(\tau) = r_{\tilde{s}_m \tilde{s}_m}(\tau) + \sum_{\substack{i=1 \\ i \neq m}}^{M} r_{\bar{s}_{i,m} \bar{s}_{i,m}}(\tau). \tag{14}$$

Substituting (13) and (14) to (12) and applying the Fourier transform, we obtain

$$W_m(\omega) = \frac{P_{\tilde{s}_m \tilde{s}_m}(\omega)}{P_{\tilde{s}_m \tilde{s}_m}(\omega) + \sum_{\substack{i=1 \\ i \neq m}}^{M} P_{\bar{s}_{i,m} \bar{s}_{i,m}}(\omega)}. \tag{15}$$

The derived filter $W_m(\omega)$ is non-causal since the Fourier transform was applied to infinitely long correlation sequences [21]. Furthermore, recall that the sources have been assumed wide-sense stationary, which is not the case for audio sources. Hence, for practical applications involving non-stationary signals, $W_m(\omega)$ is approximated [12] as

$$\widehat{W}_m(\kappa, \omega_b) = \frac{P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b)}{P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) + \sum_{\substack{i=1 \\ i \neq m}}^{M} P_{\bar{s}_{i,m} \bar{s}_{i,m}}(\kappa, \omega_b)} \tag{16}$$

where $P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b)$ and $P_{\bar{s}_{i,m} \bar{s}_{i,m}}(\kappa, \omega_b)$ are the short-time PSDs of $\tilde{s}_m(k)$ and $\bar{s}_{i,m}(k)$, respectively, which are obtained through the short-time Fourier transform (STFT) of the source

signals. The index $\kappa = 1, 2, \ldots, K$ describes the STFT frame index with K the total number of frames and $\omega_b = 1, 2, \ldots, B$ is the discrete frequency bin index with B the total number of frequency bins.

Assuming the signals are stationary for the duration of the STFT frame, a Wiener filter can be calculated for each microphone signal which will provide an estimate of the respective direct source, i.e.,

$$Y_m(\kappa, \omega_b) = \widehat{W}_m(\kappa, \omega_b)X_m(\kappa, \omega_b) \tag{17}$$

where $Y_m(\kappa, \omega_b)$ and $X_m(\kappa, \omega_b)$ are the STFT of the mth output signal and microphone signal, respectively.

The main problem now is how to estimate the PSDs of the direct and leakage sources. For this purpose, a novel method is introduced and described in detail in Section III, while the overall proposed method can be described by the block diagram of Fig. 2.

III. POWER SPECTRAL DENSITY ESTIMATIONS

A. Estimation of the Direct PSD

A fairly straightforward estimation method for the PSD of the direct source will be employed here, based on the assumption of the close-microphone technique and hence the assumption that the source of interest $s_m(k)$ is dominant in the microphone signal $x_m(k)$ which will form the basis of the estimation method. At first approximation, the microphone signal PSD can be used, that is

$$P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) \approx P_{x_m x_m}(\kappa, \omega_b) \tag{18}$$

However, due to the presence of interference, the actual microphone PSD is a sum of the direct source and the interference PSD. Hence, an error term ϵ is introduced

$$P_{x_m x_m}(\kappa, \omega_b) = P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) + \epsilon. \tag{19}$$

Clearly, the amount of interference present in the microphone signal controls the accuracy of the PSD estimation and consequently the performance of the resulting Wiener filter. It was shown in previous work [1] that for close-microphone recordings of a setup with two sources and two microphones inside real reverberant environments, this crude approximation results in an effective Wiener filter that successfully suppresses microphone leakage.

Here, this concept will be extended and a more robust method will be introduced based on the following observations. By examining Fig. 3 it can be seen that the PSD of the microphone signal, apart from the source, contains a large amount of energy in frequencies higher than 2.5 kHz that are due to the interference reaching the microphone. However, by looking at the PSD of the output signal, this energy has been attenuated by the Wiener filter and thus a better estimation of the desired PSD can be obtained. What is more important to note is that there exist PSD regions that are almost the same between the original source, the microphone and the output signal and that they are almost unaffected by interference. Hence, if these regions can be effectively identified, then they can be utilized in order to
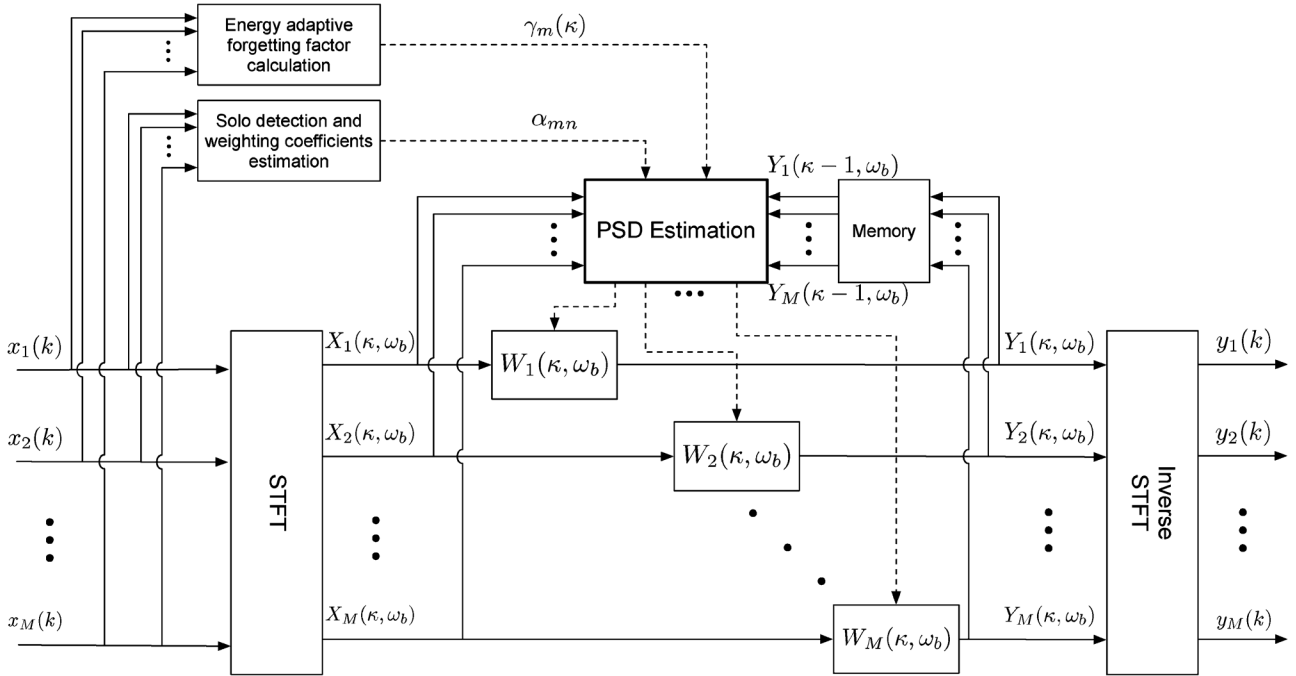
Fig. 2. Block diagram describing the proposed method. The PSD estimation process is detailed in Section III. The energy-adaptive forgetting factor is discussed in Section III-A and the solo detection and weighting coefficient estimation method is detailed in Section III-C. The dashed lines denote multichannel information.
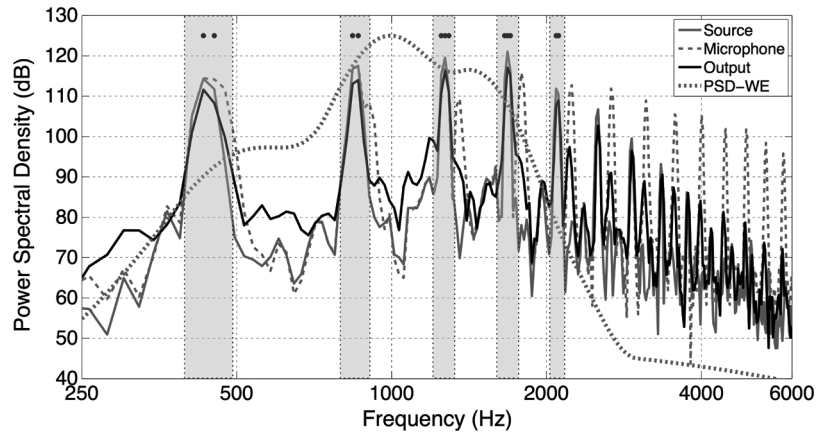

Fig. 3. PSD of the direct source, the microphone signal and the output, along with the PSD-WE (see Section III-D) for a setup with six sources, a source–microphone distance of 20 cm, and a reverberation time of 1.2 s. The gray shaded areas denote regions which are the same between the original source, the microphone and the output signal. The dots denote dominant frequency bins. All PSDs have been scaled for illustration purposes.

form a more accurate PSD estimation, since they most probably belong to the original source PSD.

In order to identify these regions, the set of "active" frequency bins are first identified in the microphone and output PSDs. These are defined as those frequency bins having an amplitude larger than the root mean squared (rms) amplitude of the PSD. In other words, let Ω_B be the set of all frequency bins and define $\Omega_{A_x}(\kappa)$ as the set of active frequency bins at the microphone PSD for the κth frame:

$$\Omega_{A_x}(\kappa) = \{\omega_b \in \Omega_B : P_{x_m x_m}(\kappa, \omega_b) \geq \mathrm{E}_{x_m}\} \qquad (20)$$

where E_{x_m} is the rms amplitude of $P_{x_m x_m}(\kappa, \omega_b)$. Similarly, the active frequency bins of the output signal $y_m(k)$ are defined as

$$\Omega_{A_y}(\kappa) = \{\omega_b \in \Omega_B : P_{y_m y_m}(\kappa - 1, \omega_b) \geq \mathrm{E}_{y_m}\} \qquad (21)$$

where E_{y_m} is the rms amplitude of the previous frame of the output PSD $P_{y_m y_m}(\kappa - 1, \omega_b)$.

Since as is observed in Fig. 3, the regions that should be identified are common to both microphone and output PSDs, the "dominant" frequency bins are chosen as those frequency bins being active in both signals:

$$\Omega_D(\kappa) = \Omega_{A_x}(\kappa) \cap \Omega_{A_y}(\kappa). \qquad (22)$$

The characteristic function of the set $\Omega_D(\kappa)$ is

$$\chi_{\Omega_D}(\kappa, \omega_b) = \begin{cases} 1 & \omega_b \in \Omega_D(\kappa) \\ 0 & \omega_b \notin \Omega_D(\kappa) \end{cases}. \qquad (23)$$

The characteristic function of the complement set $\Omega_D^c(\kappa)$, which contains the "residual" bins, is defined as

$$\chi_{\Omega_D^c}(\kappa, \omega_b) = 1 - \chi_{\Omega_D}(\kappa, \omega_b). \qquad (24)$$

The "dominant" PSD component is based on the dominant frequency bins and essentially provides a weighted version of the microphone PSD

$$P_m^D(\kappa, \omega_b) = \delta_D \chi_{\Omega_D}(\kappa, \omega_b) P_{x_m x_m}(\kappa, \omega_b). \qquad (25)$$

The reasoning behind using the dominant PSD components of the microphone and not the output signal is the fact that the processed output signal may be distorted with respect to the microphone signal and using it for the PSD estimation process may introduce further distortions.

The rest of the PSD estimation can be formed from the "residual" part of microphone and input PSDs, as

$$P_m^R(\kappa, \omega_b) = \chi_{\Omega_D^c}(\kappa, \omega_b) \left(\delta_{R_x} P_{x_m x_m}(\kappa, \omega_b) \right.$$
$$\left. + \delta_{R_y} P_{y_m y_m}(\kappa - 1, \omega_b) \right). \qquad (26)$$

The parameters $\delta_D, \delta_{R_x}, \delta_{R_y}$ are introduced in (25) and (26) to enable fine-tuning of the PSD estimation, by controlling the relative importance of the microphone and output PSD components with respect to the dominant component of the microphone PSD. They take values in [0, 1] and their sum should equal to unity (i.e., $\delta_D + \delta_{R_x} + \delta_{R_y} = 1$).

The final PSD estimation is achieved via an iterative procedure, controlled by an energy-adaptive forgetting factor

$$P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) = \gamma_m(\kappa) P_{\tilde{s}_m \tilde{s}_m}(\kappa - 1, \omega_b)$$
$$+ (1 - \gamma_m(\kappa)) \left[P_m^D(\kappa, \omega_b) + P_m^R(\kappa, \omega_b) \right]. \qquad (27)$$

The forgetting factor $\gamma_m(\kappa)$ controls the "memory" of the estimation or equivalently its sensitivity to sudden changes. The one-pole smoothing procedure of (27) is commonly used in cases where a PSD estimation is affected by noise or interference and smooths out abrupt fluctuations that may result from a high energy interfering signal.

For each microphone signal the respective value of the forgetting factor should follow the signal's energy changes, while taking into account the energy of all other signals. When the energy of the microphone signal is low compared to all other microphone signals, implying that the microphone receives a significant amount of interference and hence the current PSD estimation may not be reliable, the forgetting factor should have a high value in order to steer the iterative procedure towards previous values. On the other hand, when the energy of the microphone signal is quite larger compared to all other microphones, indicating that the interfering energy at the microphone will be low and hence the current PSD estimation adequate, the forgetting factor should take a low value in order for the iterative process to take into account mostly the current and to a lesser degree the previous estimations. This can be accomplished employing a time-varying forgetting factor for each microphone signal

$$\gamma_m(\kappa) = \exp\left(-\frac{\mathcal{E}[\mathrm{x}_{m,\kappa}]}{\sum_{\substack{i=1 \\ i \neq m}}^{M} \mathcal{E}[\mathrm{x}_{i,\kappa}]} \right) \qquad (28)$$
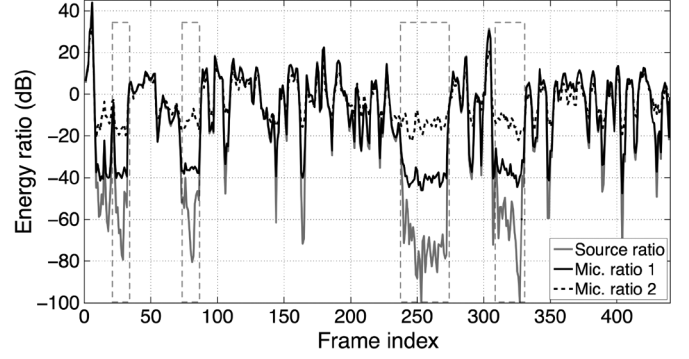

Fig. 4. Energy ratio of a source and the respective microphone energy ratio in low (ratio 1) and high (ratio 2) interference settings. The frames during which the source has low energy or is silent are marked by dashed rectangles.

which is based on the energy ratio between the microphone signals. $\mathcal{E}[\cdot]$ is the energy operator defined as

$$\mathcal{E}[\mathrm{x}_{m,\kappa}] = \frac{1}{L_b} \sum_{l=0}^{L_b - 1} |x_m(\kappa L_b + l)|^2 \qquad (29)$$

with $\mathrm{x}_{m,\kappa}$ being the κth block of the mth microphone signal of length L_b. The use of the exponential function bounds the values of $\gamma_m(\kappa)$ to (0,1], thus making the forgetting factor adequately robust with respect to amplitude differences due to varying source–microphone distance and/or source–microphone settings.

The energy of the microphone signal calculated by (29) does not correspond to the true energy of the respective source due the presence of leakage at the microphone. However, the energy ratio of the microphone signals will closely follow the energy ratio of the sources, since interference is a relatively constant factor for a given setup. This is illustrated in Fig. 4 where the energy ratio of a source is shown, along with the respective microphone energy ratio for low and high interference settings. The energy ratios are almost identical for all frames, except those for which the source has really low energy, where the microphone ratios have increased values due to the presence of interference.

B. Estimation of the Leakage PSD

The previous section provided an estimation method for the direct PSD $P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b)$. However, the estimation of leakage PSDs $P_{\tilde{s}_{i,m} \tilde{s}_{i,m}}(\kappa, \omega_b)$ is even more critical as they constitute the noise term to be suppressed.

Ignoring interference for the moment, the problem is to estimate

$$P_{\tilde{s}_{i,m} \tilde{s}_{i,m}}(\kappa, \omega_b) \propto |\mathcal{F}\{s_i(k) * h_{mi}(k)\}|^2 \qquad (30)$$

when the direct PSD

$$P_{\tilde{s}_i \tilde{s}_i}(\kappa, \omega_b) \propto |\mathcal{F}\{s_i(k) * h_{ii}(k)\}|^2 \qquad (31)$$

is known, where $\mathcal{F}$ is the Fourier transform operator. It can be seen that the relation between $P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b)$ and $P_{\tilde{s}_{i,m} \tilde{s}_{i,m}}(\kappa, \omega_b)$ boils down to the relation between $h_{ii}(k)$ and $h_{mi}(k)$. It has been shown in previous work [1], [2] that the

close-microphone response $h_{ii}(k)$ is almost ideal and can be reduced to a simple gain and delay

$$h_{ii}(k) \approx g_{ii}\delta(k - d_{ii}) \qquad (32)$$

where g_{ii} is the amplitude of the contribution of the direct sound in the RIR and d_{ii} is the delay in samples that represents the distance between the ith source and its respective microphone.

It was also shown [1] that leakage responses may involve only a few or even a single significant reflection, especially in large rooms where reflective surfaces are far away from the sources and microphones. Hence, the leakage response is decomposed as

$$h_{mi}(k) = g_{mi}\delta(k - d_{mi}) + r_{mi}(k) \qquad (33)$$

where the term $g_{mi}\delta(k - d_{mi})$ describes the direct part of the impulse response which consists of a gain and a delay, both depending on the distance between the ith source and the mth microphone and $r_{mi}(k)$ describes the rest of the impulse response. If only the direct part is taken into account, then a set of "weighting coefficients" can be calculated as

$$\alpha_{mi} = \left(\frac{g_{mi}}{g_{ii}}\right)^2 \qquad (34)$$

and using these coefficients, the leakage PSD can be written as

$$P_{\tilde{s}_{i,m}\tilde{s}_{i,m}}(\kappa, \omega_b) = \alpha_{mi}P_{\tilde{s}_i\tilde{s}_i}(\kappa, \omega_b). \qquad (35)$$

The estimation of leakage PSDs is now directly linked to the accuracy of the direct PSD estimation. In effect, the weighting coefficients are a scalar gain that accounts for the energy reduction of sound propagation. When setting $\alpha_{mi} = 1$, the multichannel noise term of (16) is overestimated since the interference contributed by each source is equally considered regardless of its proximity to the microphone. This will in turn result to the introduction of distortion and processing artifacts. Of course in practice the energy reduction due to distance is not the same for all frequencies; however, since leakage RIRs include only a few prominent reflections, the approximation of a scalar gain generally holds. The coefficients of (34) can be also seen as the multichannel equivalent of the noise weighting term introduced in the generalized Wiener filter [12], [22] as a means to balance the amount of suppression applied versus the amount of distortion introduced. Furthermore, setting the weighting coefficients to zero for sources that have very small contribution to leakage and may be perceptually unobtrusive reduces the amount of frequency domain processing by the Wiener filter and preserves signal quality.

The problem now relates to the estimation of those coefficients in a *blind* way, since in general there is no knowledge of the room impulse responses. In Section III-C, a method for estimation of the weighting coefficients will be presented.

C. Estimation of the Weighting Coefficients

In music performances there are often time intervals of varying duration during which only one instrument is active.

For such a "solo interval," when only source $s_i(k)$ is active, all microphone signals can be expressed as

$$x_m(k) = s_i(k) * h_{mi}(k) \qquad (36)$$

for $m = 1, \ldots, M$. Note that here it can be $m = i$. Modeling $h_{mi}(k)$ as in (33) and assuming that $g_{mi} \gg \sum_k |r_{mi}(k)|^2$ and since the delay does not affect the energy then the energy of each microphone signal is

$$\mathcal{E}[\mathrm{x}_{m,\kappa}] = \frac{g_{mi}^2}{L_b}\sum_{l=0}^{L_b-1}|s_i(\kappa L_b + l)|^2. \qquad (37)$$

Taking the ratio of each microphone signal $\mathrm{x}_{m,\kappa}$ with respect to $\mathrm{x}_{i,\kappa}$ during a solo interval and provided that the assumptions about the RIR decomposition hold, the weighting coefficients of (34) are estimated:

$$\frac{\mathcal{E}[\mathrm{x}_{m,\kappa}]}{\mathcal{E}[\mathrm{x}_{i,\kappa}]} = \left(\frac{g_{mi}}{g_{ii}}\right)^2 = \alpha_{mi} \qquad (38)$$

for $i \neq m$. The authors have previously proposed a method to detect "solo intervals" [23] based on the energy ratio which was discussed in Section III-A. Using a sigmoid bounding function

$$f_s(x) = \frac{2}{1 + \exp(-\beta x)} - 1 \qquad (39)$$

the energy ratio for solo detection is

$$ER(m, \kappa) = f_s\left(\frac{\mathcal{E}[\mathrm{x}_{m,\kappa}]}{\sum_{\substack{i=1 \\ i \neq m}}^{M}\mathcal{E}[\mathrm{x}_{i,\kappa}]}\right). \qquad (40)$$

Following the same reasoning as for the energy-adaptive forgetting factor and under the close-microphone assumption the bounded energy ratio of (40) will take values equal to unity for the microphone that corresponds to a solo source and quite low values for all other microphones, while it will generally have low values for all microphones when all sources are active simultaneously. Hence, by examining the values of $ER(m, \kappa)$ for all microphones at each frame, solo intervals can be detected. The process is described by the flowchart of Fig. 5. Parameter β of (39) controls the sensitivity of the detection process.

It is clear that the performance of this method depends heavily on the close microphone assumption. Using the solo detection ratio (that is the number of correctly identified solo frames to the total number of solo frames) to assess the detection performance it can be seen from Fig. 6, that it depends significantly on source–microphone distance and reverberation time. For short distances, below 10 cm, the method performs well with a detection ratio over 60% for all reverberation times examined. For longer distances and reverberation times the performances decreases quite fast. However, as it is shown in Section IV-B the overestimation effect is more evident in shorter distances and there the estimation of the weighting coefficients is more critical. Hence, the performance of the method is adequate for the purpose considered in this work.
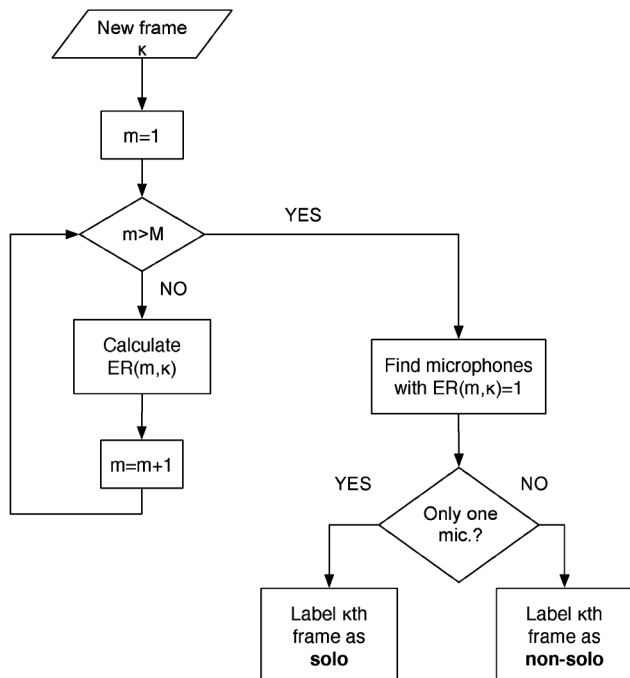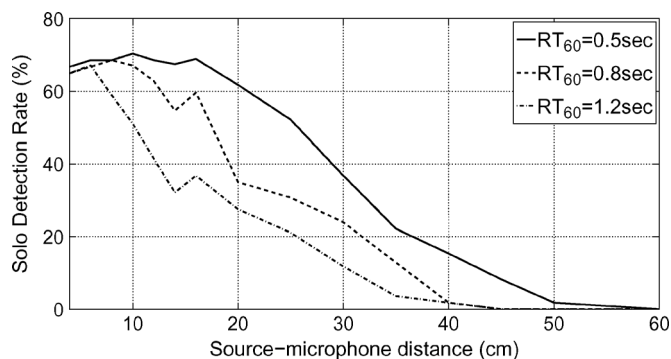
Fig. 5. Flowchart describing the process of detecting solo frames.

Fig. 6. Performance of the solo detection method for variable source–microphone distance and various reverberation times ($\beta = 8$).

D. Power Spectral Density Weighting Envelope (PSD-WE)

While the method presented in Section III-A can produce a fairly accurate estimation of the source PSD, in highly interfering environments and longer source–microphone distances the residual component of the PSD may contain significant energy which will bias the calculated Wiener filter and result in distorted output signals or even reduced leakage suppression.

By observing again Fig. 3 it can be argued that regions of significant interference tend to be at the extremes of the spectrum or in between the dominant bins. A power spectral density weighting envelope (PSD-WE) is introduced here, which is essentially a weighting function that attempts to attenuate components belonging to interfering sources and forces the PSD estimation to more closely follow the overall trend and shape of the actual PSD. An example of such an envelope is shown in Fig. 3.

Let $\{\omega_{D,j}(\kappa)\}$ be the tuple that corresponds to $\Omega_D(\kappa)$ when the elements of the set are given in ascending order, with $\beta_D(\kappa)$

being the number of elements in $\Omega_D(\kappa)$. Define the "global shaping" function as

$$f_0(\kappa, \omega_b) = \begin{cases} c_w e^{\lambda_L \omega_b}, & \text{if } \omega_b < \omega_{D,1}(\kappa) \\ c_w, & \text{if } \omega_{D,1}(\kappa) \leq \omega_b \leq \omega_{D,\beta_D(\kappa)}(\kappa) \\ c_w e^{-\lambda_H \omega_b}, & \text{if } \omega_b > \omega_{D,\beta_D(\kappa)}(\kappa) \end{cases} \tag{41}$$

where $0 < c_w \leq 1$ controls the overall amplitude while λ_L, λ_H control the steepness and shape of the rising and decaying slopes of the function. Furthermore, let us define the "dominant shaping" functions as

$$f_j(\omega_b) = \begin{cases} 0.5 \left[1 - \cos\left(2\pi \frac{\omega_b}{B_j} \right) \right] & \text{for } |\omega_{D,j}(\kappa) - B_j| \leq \omega_b \\ 0 & \text{else} \end{cases} \tag{42}$$

with

$$B_j = \frac{P_{x_m x_m}\left(\kappa, \omega_{D,j}(\kappa)\right)}{\max_j \left[P_{x_m x_m}\left(\kappa, \omega_{D,j}(\kappa)\right) \right]} B_{\max} \tag{43}$$

Equation (42) describes a Hanning window centered around the jth dominant bin with a size (bandwidth) relative to the ratio of the jth dominant bin PSD magnitude to the maximum PSD magnitude of all dominant bins and a maximum bandwidth of $B_{\max}$.

Finally, the PSD-WE is defined as

$$\Phi_m(\kappa, \omega_b) = f_0(\kappa, \omega_b) + \sum_{j=1}^{\beta_D(\kappa)} f_j(\kappa, \omega_b). \tag{44}$$

By applying (44) to (27) the PSD estimation process becomes

$$P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) = \gamma_m(\kappa) P_{\tilde{s}_m \tilde{s}_m}(\kappa, \omega_b) + (1 - \gamma_m(\kappa)) \Phi_m(\kappa, \omega_b) \left[P_m^D(\kappa, \omega_b) + P_m^R(\kappa, \omega_b) \right]. \tag{45}$$

The global function $f_0(\kappa, \omega_b)$ is an overall window applied to the PSD estimation that smooths out extremely low and high frequencies where the estimation is less definite. The steepness parameters λ_L, λ_H control the amount of smoothing and attenuation applied to spectrum extrema, as well as the bandwidth of this process. The dominant shaping functions $f_j(\kappa, \omega_b)$ are narrow windows that are overlayed on $f_0(\kappa, \omega_b)$ and attempt to preserve the information around dominant frequency bins and suppress possible interference in between.

IV. Tests and Results

In order to investigate the performance of the proposed method, two different source–microphone settings were studied via simulations inside a room with dimensions $12 \times 8 \times 4.5$ m and variable reverberation time, employing an image source method [24] (Fig. 7). For the first setup (A), the sources are separated by a fair distance, which suggests that the interference between them is less pronounced, a setting that is often used with acoustic instrument sources. The sources for the second setup (B) are placed closer together, in a positioning similar to the one used for rock/pop bands sound reinforcement. The microphones are assumed to be placed directly in front of the respective sources at a source–microphone distance d equal
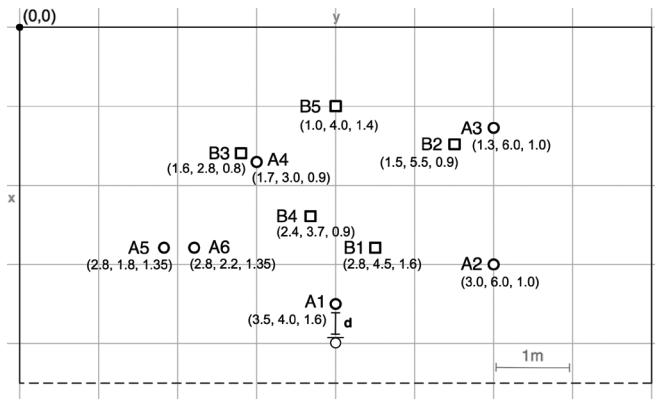
Fig. 7. Diagram of the source–microphone positions used in the simulations. Setup A (circles) consists of source–microphone positions with a distance typical for acoustic sources, while setup B (squares) describes an arrangement of sources closer together, similar to those used in sound reinforcement for rock/pop bands. Note that only the area around the "stage" is shown.

TABLE I
DETAILS OF THE TYPE OF SOURCES USED IN EACH SETTING

Source Number	Setting A	Setting B
1	Male Speaker	Female Vocals
2	Cello	El. Bass
3	Classic Guitar	El. Guitar
4	Female Speaker	Acoustic Guitar
5	Trumpet	Snare Drum
6	Cornet	-

for all source/microphone pairs (as shown in Fig. 7). The specific sources used for each setting are described in Table I. Given these source/microphone setups, the performance of the proposed method was investigated for a variety of acoustic environments, interference levels and source spectral profiles.

The performance assessment of audio signal enhancement algorithms such as source separation and noise suppression, is not a straightforward task, especially when the outputs of these methods are to be assessed and presented to human listeners [25]–[27]. In this work, the set of objective performance measures (signal-to-interference ratio—SIR; signal-to-distortion ratio—SDR) proposed in [28] and [29] for the evaluation of source separation algorithms. The reason for using these metrics to assess the performance of the proposed method, which is derived from a noise suppression framework, is that "noise" here is a mixture of audio signals which have different properties to typical noise interference. Furthermore, the segmental signal-to-noise ratio (segmental SNR) is also used together with the perceptual evaluation of audio quality (PEAQ) measure [30], [31], which provides a perceptually relevant assessment of the method's performance as it indicates improvement with respect to the MOS (mean opinion score) scale, to complement the above metrics.

For the results presented an STFT frame of $L_b = 2048$ samples was used with 50% overlap and a Hanning window. The values of all the parameters are summarized in Table II. For the estimation of the weighting coefficients, it is assumed that in the beginning of each source set, each source is assigned a solo interval of 300 ms during which solo frames are detected using the method of Section III-C to provide the estimated weighting coefficients.

TABLE II
PARAMETER VALUES

Parameter	Value
δ_D	0.6
δ_{R_x}	0.25
δ_{R_y}	0.15
c_w	1
λ_H	0.1
λ_L	0.001
B_{max}	50

A. Effect of Number of Sources

The number of simultaneously active sources in a given setting is a parameter that significantly affects the amount interference while it further increases the excitation of room acoustics contributing to increased leakage in the microphone signals. In Fig. 8, the performance of the proposed method is shown for variable number of sources for a reverberation time of $RT_{60} = 0.8$ s and source–microphone distances of 10 cm and 20 cm.

For the case when $\alpha_{mi} = 1$ ("no weights") there is a significant SIR improvement for both 10 cm and 20 cm which results partly from the suppression of leakage but also from the suppression of components from the source of interest due to the overestimation of the multichannel noise term. In turn, all other measures indicate a relative degradation. Especially for a small number of sources, this degradation is more prominent since there is a lesser amount of interference present and the overestimation is more severe. On the other hand, when the estimated weighting coefficients are used, less SIR improvement is achieved, while there is a significant improvement for all other measures. The method seems to perform well for any number of sources examined here. The improvement provided by the method increases for an increasing number of sources in terms of SIR, SDR, and segmental SNR while PEAQ follows the opposite trend. This indicates that for low interference cases, the method provides sufficient suppression while preserving the output signal quality, but when interference increases, while there is more suppression, the perceptual quality decreases.

B. Effect of Source–Microphone Distance

The most important factor that determines the performance of the proposed method is clearly the distance between the sources and the microphones, which also determines how valid is the close-microphone assumption and the related approximations. Here, the performance of the method will be examined for increasing source–microphone distances starting from 4 cm (where the close-microphone assumption is valid) and for a maximum of 60 cm (where the close-microphone assumption marginally holds). Furthermore two different sets of sources in two different setups (see Fig. 7) will be examined, in order to look at how sources with different spectra interact. The results are summarized in Fig. 9 for setup A (six sources) and Fig. 10 for setup B (five sources), both for $RT_{60} = 0.5$ s.

When the overestimation effect is not taken into account, the method provides increased leakage suppression (indicated by the SIR measure) at the cost of increased distortion as shown mainly by SDR and PEAQ. On the other hand, when the ideal
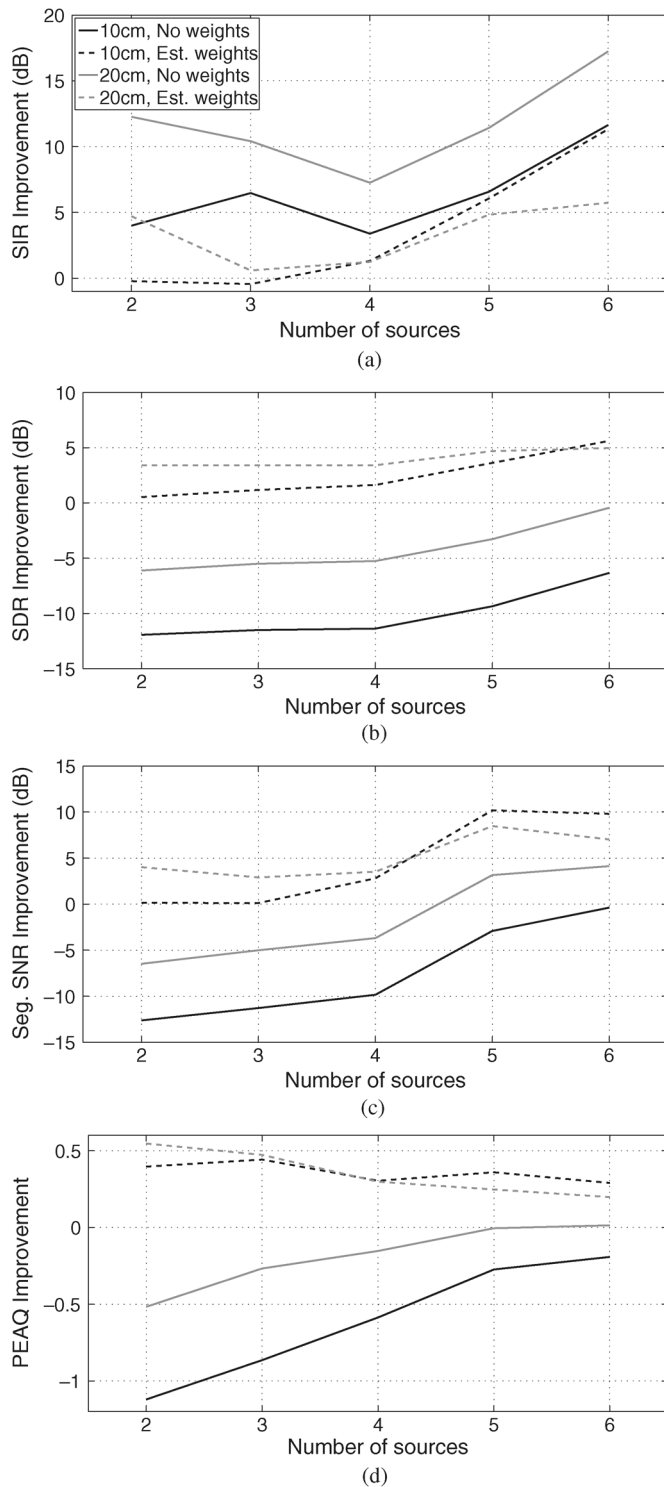
Fig. 8. Average performance for setup (A) with $RT_{60} = 0.8$ s and variable number of sources for source–microphone distances of 10 cm and 20 cm. Performance is shown for case with no and estimated weighting coefficients.
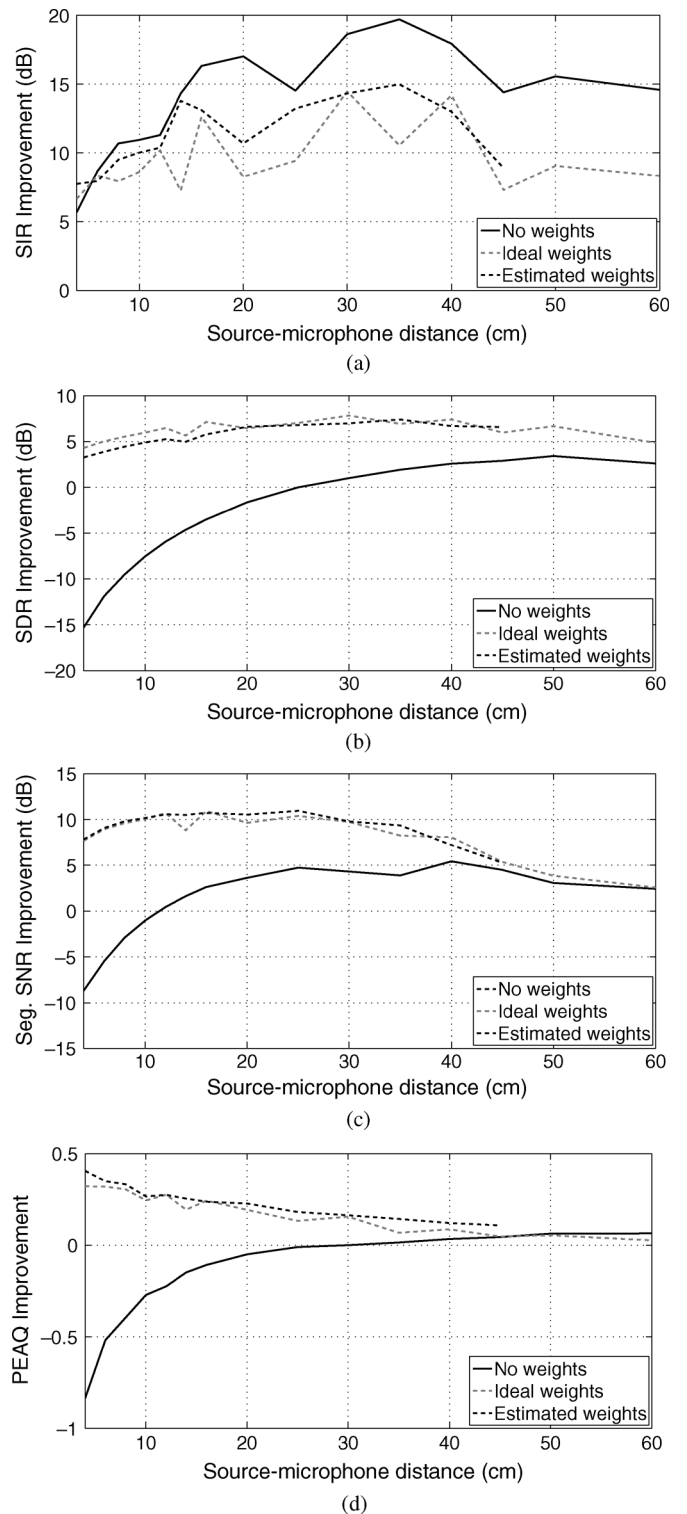

Fig. 9. Average performance for setup A for variable source–microphone distance for $RT_{60} = 0.5$ s and six sources. The case where no weighting coefficients ($\alpha_{mi} = 1$) are used is presented, along with the case of ideal weighting coefficients derived from the RIRs and the estimated ones via the method of Section III-C.

values of the weighting coefficients are used, derived from the ratio of the RIRs maxima, the performance in terms of SIR is somewhat compromised but the overall quality of the processed signal is improved, as the appropriate amount of leakage is suppressed. Note however, that there is a minimum of 5-dB improvement in SIR for all cases. What is more interesting is that the estimated coefficients provide almost the same performance

with the ideal ones, limited only by the solo detection method performance with respect to source–microphone distance. Another interesting point to note is that as the source–microphone distance increases, the effect of the weighting coefficients is less
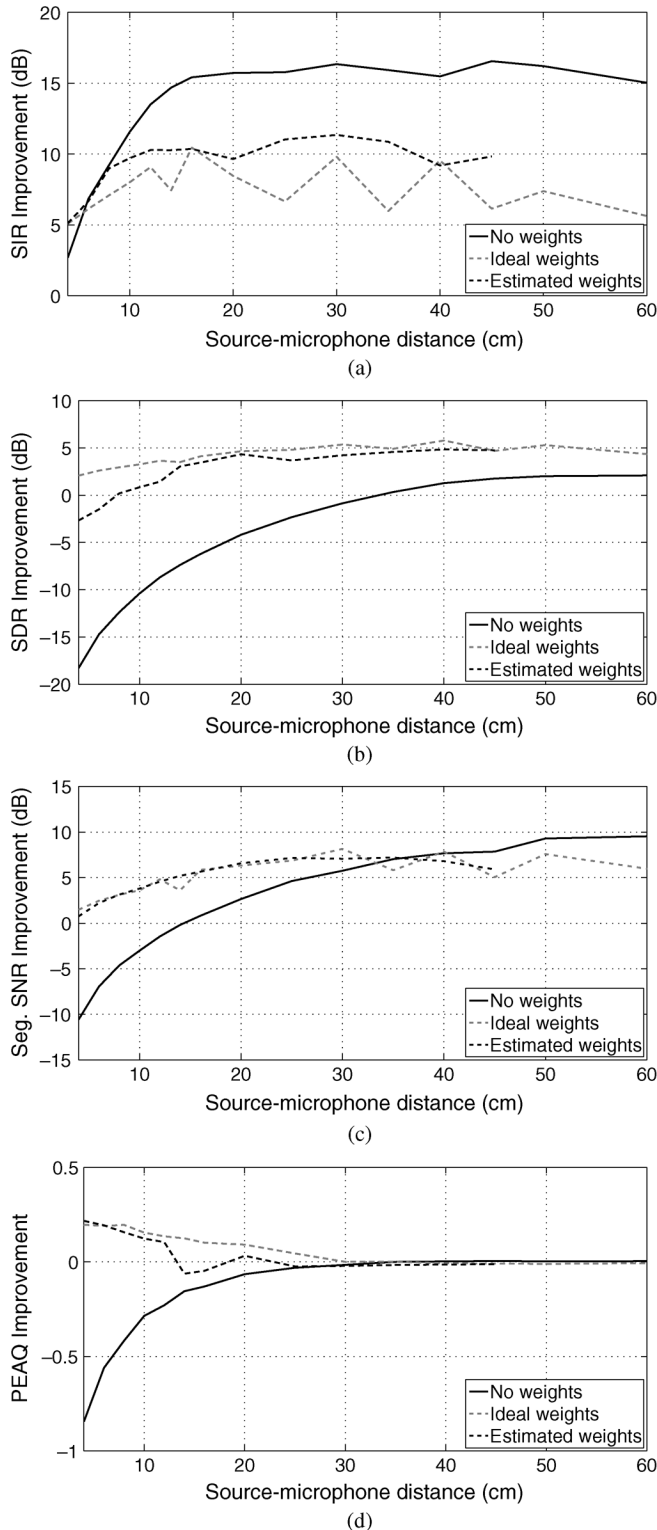
Fig. 10. Average performance for setup (B) for variable source–microphone distance for $RT_{60} = 0.5$ s and six sources. The case where no weighting coefficients ($\alpha_{mi} = 1$) are used is presented, along with the case of ideal weighting coefficients derived from the RIRs and the estimated ones via the method of Section III-C.

prominent mainly in terms of segmental SNR and PEAQ, becoming negligible for long distances. Hence, the limitation of the solo detection method does not significantly hinder the performance, since it works quite well for short distances where the overestimation effect is most evident.

Overall, the performance for both setups is similar and follows the same trends; however, the performance for setup (B) is a bit lower. This is probably due to the presence of the electric guitar, which is heavily distorted and has a strong spectral fingerprint that biases PSD estimations and hence the calculated Wiener filter.

C. Effect of Reverberation Time

The effect of reverberation time on the performance of the proposed method is assessed here (Figs. 11 and 12). In general, the performance is not significantly affected and the trends remain the same. When the estimated weighting coefficients are used, the performance is even more insensitive to reverberation time changes. Note however that the performance of the solo detection method drops rapidly for $RT_{60} = 1.2$ s and hence provides estimations of the weighting coefficients up to 16 cm.

The results presented here further support the argument that source interference for close-microphone applications is strongly dependent on room size and the proximity of reflective surfaces to the sources and microphones [1]. While in [1] the performance decreased for shorter reverberation times, here the performance is consistent for the reverberation times examined. The difference between previous and current setups is that here reverberation time changes for a constant room size and geometry while in the real recordings of previous work shorter reverberation times resulted from smaller room sizes. Hence, in order to fully assess the effect of room acoustics on source interaction for close-microphone applications as well as the performance of leakage suppression and separation methods, more acoustic parameters should be examined, besides reverberation time.

D. Effect of PSD-WE

Fig. 13 shows the performance of the proposed method with and without the PSD-WE for setup B with five sources and $RT_{60} = 1.2$ s. The weighting coefficients are not used here since the PSD-WE is mainly employed for longer source–microphone distances where the PSD estimations are more susceptible to interference. The results indicate a performance improvement for longer distances and support the reasoning behind the use of PSD-WE. It should also be noted that for short distances PSD-WE does not affect performance while SDR and PEAQ suggest that the distortion introduced by PSD-WE is minimal.

E. Effect of STFT Frame Length

An important part of the PSD estimation method as presented in Section III-D is the identification of active frequency bins. It is then reasonable to assume that a better frequency resolution might produce more accurate PSD estimations. The performance of the proposed method was examined for different STFT frame lengths and the results are summarized in Fig. 14, where it is shown that the effect of the frame length on performance is minimal.

F. Performance Comparison With BSS Methods

Despite the limitations mentioned in Section I, it is useful to examine whether BSS methods may provide some im-
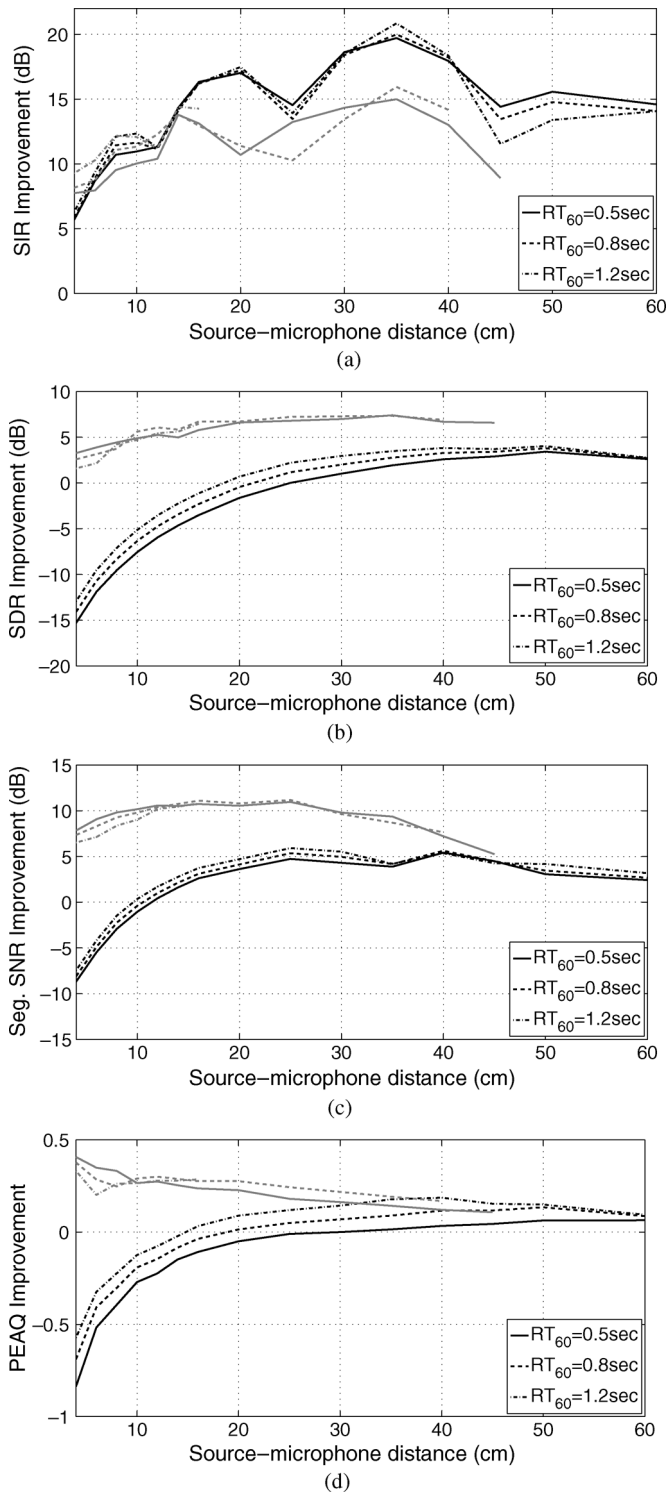
Fig. 11.   Average performance for setup (A) for variable source–microphone distance, increasing reverberation times and six sources. The case where no weighting coefficients are used is presented (black lines), along with the case of estimated weighting coefficients (gray lines).

Fig. 12.   Average performance for setup (B) for variable source–microphone distance, increasing reverberation times and six sources. The case where no weighting coefficients are used is presented (black lines), along with the case of estimated weighting coefficients (gray lines).

provement, especially for longer source–microphone distances, where the close-microphone assumption is less valid. The performance of (a) the proposed method, (b) a BSS method based on non-stationarity (PS) [32], and (c) one based on multichannel blind deconvolution (ZC) [33] are shown in Fig. 15, with respect to typical BSS performance measures.
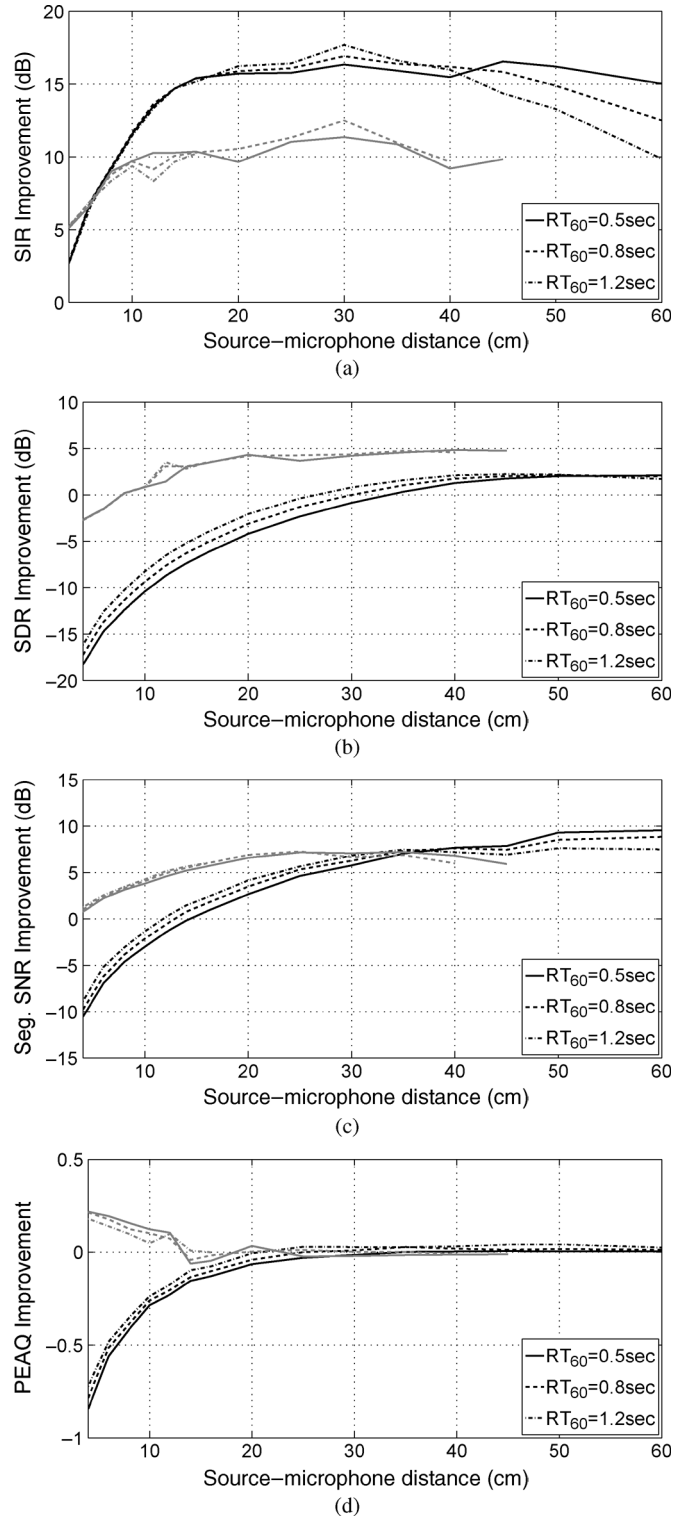
In terms of SIR, both BSS methods perform similarly providing increasing separation for increasing source–microphone distances, although the proposed method achieves significantly higher improvement. For SDR, method ZC has the same or slightly lower performance as the proposed method without weighting coefficients, while PS is a bit better even for short
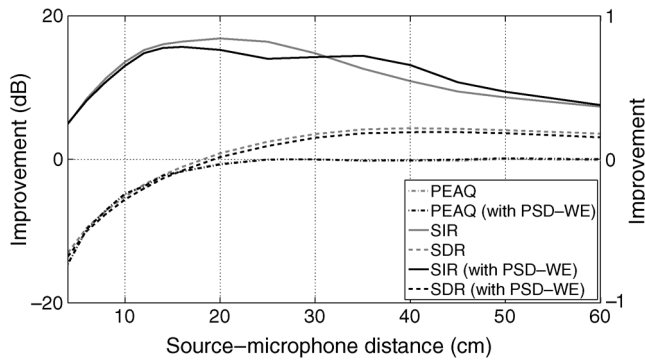
Fig. 13. Average performance with and without the use of PSD-WE for setup (B) and $RT_{60} = 1.2$ s. No weighting coefficients were used.
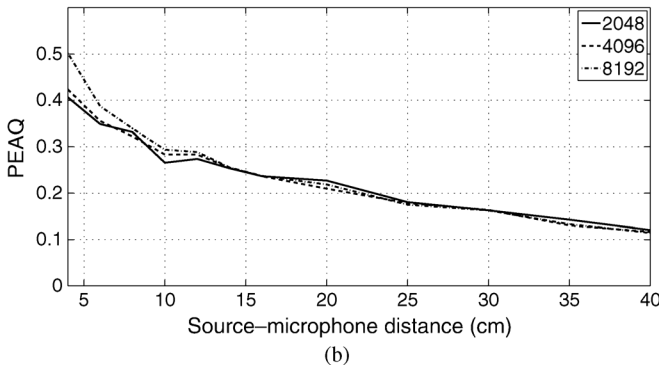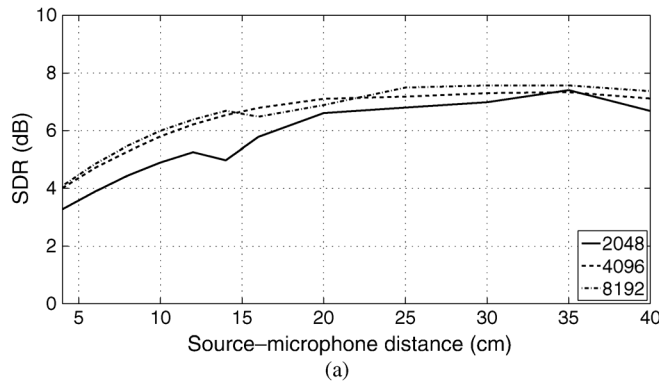

Fig. 14. Average performance for different frame lengths for setup (A), $RT_{60} = 0.5$ s and six sources. The estimated weighting coefficients are used.

distances. However, the proposed method still outperforms these two when the estimated coefficients are used. Overall, the proposed method is quite effective for short distances combining adequate suppression and output signal quality while the BSS methods seem to perform better for longer distances, however without achieving the same amount of suppression.

V. CONCLUSION

Here a method for the suppression of microphone leakage in close-microphone applications was proposed based on an extended Wiener filter, that takes into account a multichannel noise term. A PSD estimation method is introduced based on the identification of dominant frequency bins, i.e., regions of the microphone and output PSDs that are approximately the same with that of the original source signal. A simple way to estimate the leakage PSDs was also presented, based on a set of weighting coefficients which are estimated during time intervals where only one source is active. The results presented in Section IV
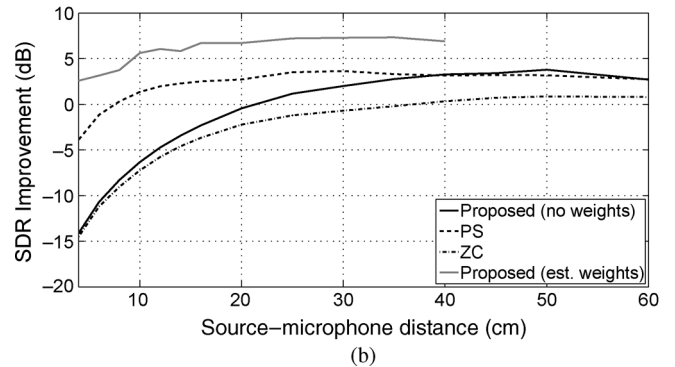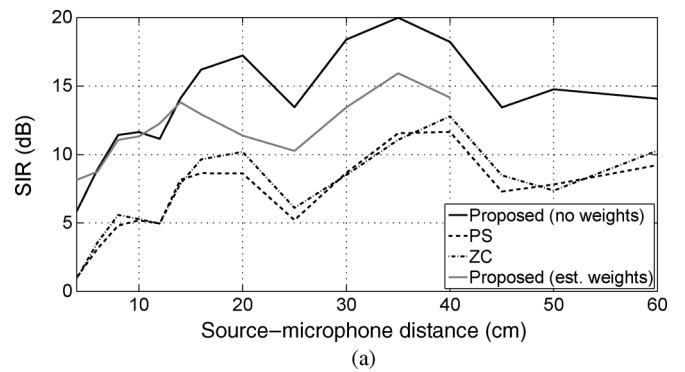

Fig. 15. Comparison between the proposed and BSS methods, using setup (A) with six sources and $RT_{60} = 0.8$ s.

justify the suitability of the noise suppression framework for the problem of microphone leakage.

From the results presented, the proposed method exhibits a consistent performance for various number of sources, different source spectral properties and source–microphone distances, while it was also shown that changes in reverberation time without respective changes in room geometry do not affect performance. Taking into account the overestimation effect enables the method to adequately suppress leakage while retaining output signal quality. The lower performance provided for setup (B) indicates that the PSD estimation method presented here is susceptible to bias from strongly interfering sources with high energy spread across the entire spectrum, such as the electric guitar.

Future work should focus on including the full effect of the leakage responses $h_{mi}(k)$, instead of a scalar gain, on the estimation of leakage PSDs employing blind identification methods, which should improve the overall noise term estimation and further reduce distortion in the output signal. Moreover, a perceptually driven control of the amount of suppression applied, as has been suggested in speech enhancement applications, could minimize audible distortion and maximize the perceived leakage reduction.

REFERENCES

[1] E. K. Kokkinis and J. Mourjopoulos, "Unmixing acoustic sources in real reverberant environments for close-microphone applications," *J. Audio Eng. Soc.*, vol. 58, no. 11, pp. 1–10, Nov. 2010.

[2] E. K. Kokkinis and J. Mourjopoulos, "Identification of a room impulse response using a close-microphone reference signal," in *Proc. Audio Eng. Soc. Conv. 128*, May 2010.

[3] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 9, no. 10, pp. 2009–2025, Oct. 1998.

[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[5] F. Abrard and Y. Deville, "Blind separation of dependent sources using the TIme-frequency ratio of mixtures approach," in *Proc. Int. Symp. Signal Process. and Its Applicat. (ISSPA)*, 2003, pp. 81–84.

[6] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. 5, pp. 2737–2740.

[7] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind source separation with different sensor spacing and filter length for each frequency range," in *Proc. 12th IEEE Workshop Neural Netw. Signal Process.*, 2002, pp. 465–474.

[8] N. Mitianoudis and M. E. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 489–497, Sep. 2003.

[9] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[10] P. Batalheiro, M. Petraglia, and D. Haddad, "Online subband blind source separation for convolutive mixtures using a uniform filter bank with critical sampling," in *Independent Component Analysis and Signal Separation*, ser. Lecture Notes in Computer Science, T. Adali, C. Jutten, J. Romano, and A. Barros, Eds. Berlin, Heidelberg, Germany: Springer , 2009, vol. 5441, pp. 211–218.

[11] M. J. Terrell and J. D. Reiss, "Automatic monitor mixing for live musical performance," *J. Audio Eng. Soc.*, vol. 57, no. 11, pp. 927–936, 2009.

[12] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[13] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.

[14] H. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'95)*, May 1995, vol. 1, pp. 153–156.

[15] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[16] M. Marzinzik and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 2, pp. 109–118, Feb. 2002.

[17] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.

[18] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, 2004.

[19] T. V. den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, Jan. 2009.

[20] D. M. Howard and J. A. S. Angus, *Acoustics and Psychoacoustics*, 4th ed. Waltham, MA: Focal Press, 2009.

[21] U. Heute, "Noise reduction," in *Topics in Acoustic Echo and Noise Control*, ser. Signals and Communication Technology, E. Hänsler and G. Schmidt, Eds. Berlin, Heidelberg, Germany: Springer, 2006, pp. 325–384.

[22] E. J. Diethorn, "Subband noise reduction methods for speech enhancement," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Norwell, MA: Kluwer, 2004.

[23] E. K. Kokkinis, J. Reiss, and J. Mourjopoulos, "Detection of 'solo intervals' in multiple microphone multiple source audio applications," in *Proc. Audio Eng. Soc. Conv. 130*, May 2010.

[24] E. A. Lehmann and A. M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1429–1439, Aug. 2010.

[25] B. Fox, A. Sabin, B. Pardo, and A. Zopf, "Modeling perceptual similarity of audio signals for blind source separation evaluation," in *Proc. 7th Int. Conf. Ind. Compon. Anal. Signal Separat.*, Sep. 2007.

[26] J. Kornycky, B. Gunel, and A. Kondoz, "Comparison of subjective and objective evaluation methods for audio source separation," in *Proc. Acoustics*, Jun. 2008.

[27] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[28] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[29] C. Févotte, R. Gribonval, and E. Vincent, "BSS_EVAL Toolbox User Guide," IRISA, Rennes, France, Tech. Rep. 1706, Apr. 2005 [Online]. Available: http://bass-db.gforge.inria.fr/bss_eval/

[30] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, "PEAQ—The ITU standard for objective measurement of perceived audio quality," *J. Audio Eng. Soc.*, vol. 48, no. 1/2, pp. 3–29, 2000.

[31] E. Benjamin, "Evaluating digital audio artifacts with PEAQ," in *Proc. Audio Eng. Soc. Conv. 113*, Oct. 2002.

[32] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[33] K. Zhang and L.-W. Chan, "Convolutive blind source separation by efficient blind deconvolution and minimal filter distortion," *Neurocomputing*, vol. 73, no. 13–15, pp. 2580–2588, 2010.

Elias K. Kokkinis received the diploma degree from the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece, in 2008. He is currently pursuing the Ph.D. degree in the Audio and Acoustic Technology Group, Department of Electrical and Computer Engineering, University of Patras, supervised by Prof. J. Mourjopoulos.

From October 2010 to January 2011, he was visiting the Center for Digital Music, Queen Mary University of London. He has been working as a Sound Engineer for concerts and studios since 2005.

His research interests include single- and multichannel audio signal processing and enhancement, identification of acoustic systems, and intelligent audio applications.

Joshua D. Reiss received the Ph.D. degree in physics from the Georgia Institute of Technology, Atlanta, specializing in analysis of nonlinear systems.

He is a Senior Lecturer with the Centre for Digital Music, Queen Mary University of London, London, U.K. He made the transition to audio and musical signal processing through his work on sigma delta modulators, which led to patents and a nomination for a best paper award from the IEEE. He has investigated multichannel and real-time audio signal processing, time scaling and pitch shifting techniques, polyphonic music transcription, loudspeaker design, automatic mixing for live sound, and digital audio effects. His primary focus of research, which ties together many of the above topics, is on the use of state-of-the-art signal processing techniques for professional sound engineering.

John Mourjopoulos (M'90) received the B.Sc. degree in engineering from Coventry University, Coventry, U.K., in 1978 and the M.Sc. and Ph.D. degrees from the Institute of Sound and Vibration Research (ISVR), Southampton University, Southampton, U.K., in 1980 and 1985, respectively.

Since 1986, he has been with the Electrical and Computing Engineering Department, University of Patras, Patras, Greece, where he is now Professor of Electroacoustics and Digital Audio Technology and head of the Audio and Acoustic Technology Group of the Wire Communications Laboratory. In 2000, he was a Visiting Professor at the Institute for Communication Acoustics, Ruhr-University Bochum, Bochum, Germany. He has authored and presented more that 100 papers in international journals and conferences. He has worked in national and European projects, has organized seminars and short courses, has served in the organizing committees and as session chairman in many conferences, and has contributed to the development of digital audio devices. His research covers many aspects of digital processing of audio and acoustic signals, especially focusing on room acoustics equalization. He has worked on perceptually motivated models for such applications, as well as for speech and audio signal enhancement. His recent research also covers aspects of the all-digital audio chain, the direct acoustic transduction of digital audio streams, and WLAN audio and amplification.

Prof. Mourjopoulos was awarded the Fellowship of the Audio Engineering Society (AES) in 2006. He is a member of the AES (currently serving as section vice-chairman) and of the Hellenic Institute of Acoustics being currently its vice-president.