



Audio Engineering Society Convention Paper 9760

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

This convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Close Miking Empirical Practice Verification: A Source Separation Approach

Konstantinos Drossos¹, Stylianos Ioannis Mimitakis², Andreas Floros³, Tuomas Virtanen¹, and Gerald Schuller^{2,4}

¹*Audio Research Group, Dept. of Signal Processing, Tampere University of Technology, Tampere, Finland.*

²*Fraunhofer IDMT, Ilmenau, Germany*

³*Lab of Audiovisual Signal Processing, Dept. of Audiovisual Arts, School of Music and Audiovisual Arts, Ionian University, Corfu, Greece.*

⁴*Technical University of Ilmenau, Ilmenau, Germany.*

Correspondence should be addressed to Stylianos Ioannis Mimitakis (mis@idmt.fraunhofer.de)

ABSTRACT

Close miking represents a widely employed practice of placing a microphone very near to the sound source in order to capture more direct sound and minimize any pickup of ambient sound, including other, concurrently active sources. It is used by the audio engineering community for decades for audio recording, based on a number of empirical rules that were evolved during the recording practice itself. But can this empirical knowledge and close miking practice be systematically verified? In this work we aim to address this question based on an analytic methodology that employs techniques and metrics originating from the sound source separation evaluation field. In particular, we apply a quantitative analysis of the source separation capabilities of the close miking technique. The analysis is applied on a recording dataset obtained at multiple positions of a typical musical hall, multiple distances between the microphone and the sound source multiple microphone types and multiple level differences between the sound source and the ambient acoustic component. For all the above cases we calculate the Source to Interference Ratio (SIR) metric. The results obtained clearly demonstrate an optimum close-miking performance that matches the current empirical knowledge of professional audio recording.

1 Introduction

Capturing sound through electro-acoustic transducers is one of the fundamental tasks in audio engineering. In practice, although audio recording is not restrained by particular specifications, there are applications where certain restrictions apply, for example when capturing an audio source output in the presence

of other active audio sources (e.g. at a live music performance, or in a live recording session). In these cases, the presence of the latter sound sources introduces ambient noise, which is added to the ambient noise of the recording space (if any).

A widely-employed technique that is used for capturing an audio source that is simultaneously active

with other sources is commonly known as close miking [1]. It defines the microphone's placement close to the sound source and nearly in all cases it is used as a rule of thumb [2]. With this specific placement of the microphone, the captured audio tends to contain more energy from the targeted source than from the surrounding ones. Hence, close miking effectively functions as a mechanical source separation method that aims to separate the signal of the targeted source from the mixture of the sound field that is created by all concurrently active sound sources. The suggested distance for the microphone placement roughly spans from 0.03 to 1 meter away from the targeted source [1], balancing the trade off between affecting the timbre of the targeted source and the pickup of unwanted sources. Although this technique is widely-used, according to authors' best of knowledge, there is no previous study that systematically verified the above microphone distance or evaluated its effect on the resulting captured audio in terms of source separation.

The field of source separation is not recent. It regards the estimation of individual signal components, denoted as sources, from their observed mixtures, and there are numerous published works focusing on this paradigm [3]. Source separation has been utilized in many applications spanning from audio signal processing, e.g. for audio up-mixing [4], stereo image enhancement [5, 6], harmonic-percussive separation [7], source modeling [8] and singing voice/solo separation [9, 10], to neurological studies, for separating different electrical sources during physiological signals measurements [11], and satellite images, e.g. for detecting the actual morphology of the ground [12].

For evaluating source separation techniques, a couple of strategies have been proposed. More specifically, in [13] a set of metrics are presented that can assess the extracted information from the mixture taking into account the produced artifacts (i.e. deformations induced by the separation algorithm, such as musical noise), noise (energy perturbations that does not correspond to the extracted source nor the interfering ones) and interference (a deformation of unwanted sources contributing to extracted information). Focusing on modeling and measuring the interference of unwanted sources subject to a targeted one, the notion of disjointness orthogonality is introduced in [14]. Assuming that non-interfering sources are completely orthogonal to each other in a signal domain, i.e. the short-time Fourier transform (STFT), the degree of overlap that

the sources might have can be estimated providing an intuitive estimation of the total interference [14].

Since close miking aims at separating the audio signal of the targeted source from the mixture of the total sound field that is created by all the active sound sources, it can be considered as a source separation technique and its effect to realistic scenarios can be evaluated by the above mentioned strategies. In this work we try to evaluate the close miking technique under the above perspective. We employ the aforementioned method for source separation evaluation based on the orthogonality assumption [14], and assess the effect of distance, targeted sound source sound pressure level, interfering noise sound pressure level, angle of the microphone with respect to the central axis of the targeted sound source, and different types of microphones lobes by means of *signal to interference ratio* (SIR), essentially objectifying the choice of microphone placement. For that cause, we conducted a series of measurements in a reverberant room, i.e. an empty theater, with two sound sources and a sound level meter for calibrating the reproduction levels.

The rest of the paper is organized as follows. Section 2 provides an overview of the existing literature that focuses on close miking, along with the presentation of the appropriate metrics and their computation. Section 3 outlines the methodology followed for the performed measurements, while Section 4 contains the obtained results. Finally, Section 5 holds the discussion of the results and Section 6 concludes the paper and proposes future works.

2 Existing work

2.1 Close miking technique

Close miking is rather based on empirical knowledge and a set of general guidelines that define the location and distance of the microphone from the sound source [1]. Existing studies are particularly focusing on two different aspects. The first considers the varying spectral information and perceived timbre of the music sound sources. The second regards the inspection of the close miking technique from a signal processing point of view and its relation to room acoustics.

Focusing on the first aspect, the author in [15] employs recordings of a variety of musical instruments

and human voice. The recordings are performed using different microphone placement distances, ranging from 0.03 to 1 meter. The recorded signals are transformed into the frequency domain and compared with the emanation patterns of each sound source. As an outcome, different equalization techniques are proposed depending on the placement of the recording microphone. Following the same approach, a work more centered to human voice is presented in [16], which is focusing on the distance of the placement of the microphone. The effect of the close miking technique is evaluated through the spectral content and perceptual evaluation (assessed through listening tests) of the recorded signal. Finally, in [17] an assessment of the ambience reflections transmitted to the recording device and of course the perceptual timbre is presented. Standard equipment is used for recording and extracting the spectral component of electroacoustic devices (e.g. electric guitar amplifier). Different microphone–source distances are used and, in addition, various angle orientations of the microphone with respect to the central axis of the sound source (i.e. $[15^\circ, 30^\circ, 45^\circ, 60^\circ, 90^\circ]$) are also considered.

Differentiating from the above frequency-based studies, the work in [2] evaluates close miking from a different signal processing perspective. In particular, this work aims to validate the close miking technique by examining the effect of the excitation of the surrounding acoustic space. To do so, sound sources are recorded in various distances and the recorded signals are subjectively assessed for their perceptual suppression of the reverberation effect. Nonetheless, all the literature described above relies on the empirical knowledge of the relative distance between the microphone and the sound source. A quantified answer regarding the definition of this distance range is still not being proposed.

2.2 Computation of SIR

For the evaluation of the source separation capabilities of the close miking technique, we employed the Signal to Interference Ratio (SIR) metric. Usually, this metric is used in the evaluation of the source separation task and indicates the energy ratio between a signal, separated from mixture of signals, and the interference from the mixture that is apparent in the separated signal.

More formally, let \mathbf{x} be a vector denoting a single-channel (*monaural*) mixture consisting of 2 additive

sources expressed as vectors \mathbf{s} and \mathbf{u} . Given that each source is known beforehand, the degree of overlap that the targeted source \mathbf{s} and the interfering \mathbf{u} have, can be computed yielding the objective measure of SIR.

To do so, an analysis operator \mathcal{T} is applied to each source (targeted and interfering one) as follows:

$$S(m, k) = \mathcal{T}(\mathbf{s}), \quad (1)$$

$$N(m, k) = \mathcal{T}(\mathbf{u}), \quad (2)$$

where \mathcal{T} corresponds to the STFT analysis operation using the parameters proposed by a standard source separation evaluation (SSE) scheme [18], and m, k denote the time-frames and frequency bins (sub-bands), respectively.

The computation of *SIR*, given a pair of sources, is formulated similarly to [14] as

$$SIR = 10 \log_{10} \left(\frac{\|M(m, k)S(m, k)\|^2}{\|M(m, k)N(m, k)\|^2} \right), \quad (3)$$

where $M(m, k)$ is a time-frequency filtering operation, i.e. *time-frequency masking*, derived from Eq. 4.

$$M(m, k) = \begin{cases} 1, & \text{if } |S(m, k)| \geq |N(m, k)| \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

In the above equations, $|\cdot|$ refers to the modulus, i.e. the *magnitude*, for the time-frequency representation of each source and $\|\cdot\|$ to the ℓ_1 norm. The values of *SIR* will approach $+\infty$ when the magnitude of the acquired signal $S(m, k)$, for each time-frame and frequency sub-band, will be superior to the interfering one. On the other hand, when the values approach $-\infty$, then the interfering source completely dominates over their mixture. Essentially, this means that a straightforward assessment of how well a method describes or estimates the targeted signal \mathbf{x} , in presence of outliers, can be acquired.

3 Experimental procedure

The experimental procedure of the work at hand is separated in two tasks: i) recording of the individual signals, and b) the computation of *SIR* subject to each recording of a pair of sources. The former was utilized in a municipal theater, located in Lixouri, Kefalonia, (Ionian islands, Greece), before the disastrous earthquakes in the Autumn of 2014 and resulted into the

Table 1: List of the equipment used for audio recordings

Apparatus	Model	Apparatus	Model
SLM	B&K 2250 Type A	Mic. A	Shure SM57, dynamic, cardioid
Laptop	Macbook Pro 15"	Mic. B	Behringer ECM8000, condenser, measurements, omni-directional
Recording software	Digidesign ProTools M-Powered 8	Musical instrument amplifier	Behringer V-Tone GMX212
Digital sound card	M-Audio Fast Track Ultra	Loudspeaker	Electrovoice SX300

formation of the audio dataset employed by the SSE task. The latter was implemented by utilizing the signal model described in Section 2.2.

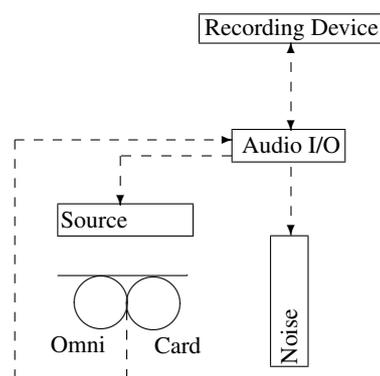
The aim of the first task is to provide a comprehensive set of recorded material containing the source signal, the noise signal, and the mixture of both. Each recorded waveform is characterized by: a) the distance between the microphone and the signal source, b) the type of microphone, and c) the sound pressure level (SPL) of the actual source and the noise source. Various combinations of the above factors were considered in the particular task. On the other hand, the second task involves the evaluation of the performance of close miking as a source separation method. The expected outcome is to determine the effective limits and the relations between the key factors mentioned above, subject to an objective measure. In the following sections the above tasks will be presented in detail.

3.1 Recordings procedure

The audio recordings were performed using a musical amplifier, one loudspeaker, one laptop with recording software and a digital sound card, two microphones (one dynamic and one condenser/measurements), and one Sound Level Meter (SLM). One microphone was omni-directional, while the the other had a cardioid

lobe. The full list of all equipment parts is provided in Table 1.

Close miking aims at diminishing the addition of the noise in the final audio mixture. The prime element that affects the efficacy of this technique is the distance between the microphone and the sound source. But since, on one hand, the distance between the microphone and the actual sound source can result into an attenuation of the SPL and, on the other hand, different sound sources in a real-world scenario are likely to exhibit varying SPL, the question of the effect of SPL in the close miking technique is also raised. Finally, various receiving patterns of microphones are utilized in a recording session. These affect the effective SPL recorded by the microphone and thus different microphone lobes are possible to portray divergent results in close miking. In addition, there are references in the utilization of an angle between the central axes of the microphone and the sound source in order to achieve improved attenuation of the receiving noise from the microphone.

**Fig. 1:** The set-up of the measurements

In order to allow the investigation of the source–microphone distance, the source’s SPL, microphone’s lobe and microphone–sound source angle’s effect in the particular technique, the experimental set-up presented in Figure 1 was performed: 2 audio sources, 1 laptop, 1 digital sound card and two microphones were utilized for the recording. For the angle case, the cardioid microphone was used. The details of each component are listed in Table 1. Thus, the loudspeaker served as the noise source, the musical instrument amplifier as the targeted sound source and the other are self–explanatory with respect to their utilization in the experimental process. All apparatuses employed are

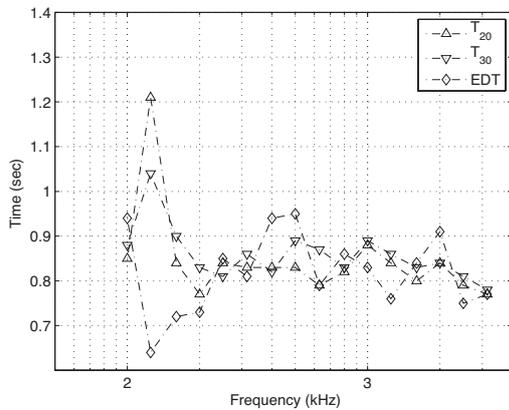


Fig. 2: T_{20} , T_{30} and EDT measured in the stage of the theater at which the recordings took place

rather common to music performances, a case where close miking technique is thoroughly met. As source signal, a clean guitar riff (repetitive tonal sound from electric guitar, amplified through the corresponding amplifier) was employed without applying any distorting sound effects. In order to introduce ambient noise, a pink noise generator was activated. Each signal had a time length of 15 seconds.

The recording process consisted of two phases. The first realized the reverberation measurement of the recording room, while the second included the actual recordings. Regarding the latter case, different recordings were considered with a) the omni-directional lobe microphone, b) the cardioid microphone with its central axis aligned with the central axis of the sound source and c) the cardioid microphone placed with an angle of 45 degrees relative to the sound source's central axis. The reverberation measurement was implemented with the use of the SLM at six different positions in the stage of the theater, forming a hexagon. For all positions the T_{20} , T_{30} and Early Decay Time (EDT) values were obtained. The results are illustrated in Figure 2.

Clearly, the theater stage can be considered fairly reverberant, especially in the region of $2.2kHz$. This fact allows our investigation to be performed in a rather un-favored environment; thus it can provide results that correspond to cases where close miking would be favored in order to eliminate capturing of audio signals emerging from all noise sources, including ambient noise. Regarding the second phase of the recordings process, 12 different sound source-microphone

Table 2: Sound source - microphone distances used in the first phase of the recording procedure

Index i_d	Distance (m)	Index	Distance (m)
01	0.03	07	0.21
02	0.06	08	0.24
03	0.09	09	0.27
04	0.12	10	0.30
05	0.15	11	0.65
06	0.20	12	1.00

Table 3: SPL values used for the recordings procedure

Index	SPL, ref $P_{ref} = 2 \times 10^{-4}$
Sound Source SPL (SPL_S)	
$SPL_S[1]$	100 dB SPL
$SPL_S[2]$	97 dB SPL
$SPL_S[3]$	94 dB SPL
Noise Source SPL (SPL_N)	
$SPL_N[1]$	100 dB SPL
$SPL_N[2]$	97 dB SPL
$SPL_N[3]$	94 dB SPL
$SPL_N[4]$	91 dB SPL
$SPL_N[5]$	88 dB SPL

distances were employed, with index $i_d \in [1, 12]$ and ranging from 0.03 to 1 meter. The third recording type (with the cardioid microphone placed with an angle of 45 degrees relative to the sound source's central axis) included 10 additional distances, ranging from 0.03 to 0.30 meters, marked as $i_d \in [1, 10]$. For clarity, these values are summarized in Table 2. Up to 0.3 meters the distance increment step equals to 0.03 meters. Above that limit, it becomes 0.35 meters. The reason for that is the apparent evidence in the existing literature, that above 0.3 meters close miking technique suffers from leakage and interference, when the sound pressure level of unwanted sources is high, contrary to the desired source [1, 2]. Also, in the second phase the utilized distances are those with index $i_d \leq 10$. Moreover, different SPL values were employed for both sound and noise sources and for all microphone lobe's cases. For the former sound source, a set of 2 different SPL, $SPL_S[i_S]$, $i_S \in [1, 2]$, values was used, whereas for the latter a set of 5, $SPL_N[i_N]$, $i_N \in [1, 5]$. These information is listed in Table 3.

The different SPLs for the sound and noise source have a variation step of 3 dB SPL, since this difference corresponds to two times the acoustic energy. Also, there

are 3 different SPL_S : one that can be considered as high, one as medium and one as low. In conjunction with the SPL_N , these values allow the investigation of the different SPL effect. More specifically, each SPL_S was used with every SPL_N , i.e. for SPL_S all SPL_N were utilized for the noise source and the same stands for $SPL_S[2]$. Thus, for $SPL_S[1]$ it can be seen that the selected SPL_N span in the dynamic range of equal SPL to 1/24 times lower (for the case of $SPL_S[1]$ and $SPL_N[5]$). In the case of $SPL_S[2]$, the dynamic range of SPL corresponds to double acoustic energy emerging from the noise source as well as the same, half, one quarter and one eighth acoustic energy for the noise. Regarding the $SPL_S[3]$ it can be seen that the selected SPL for the noise source corresponds to quadruple, double, equal, half and one quarter acoustic energy emerging from the noise source. All SPLs were calculated in terms of L_{eq} , with a time length average equal to the time length of both the sound and noise source signal (i.e. 15 seconds).

The actual recordings were performed for each microphone and for all SPL_S , SPL_N and appropriate sound source-microphone distances. In particular, if $D[i_d]$ are the different distances as presented in Table 2, $M_t, t \in [1, 2]$ are the different microphone lobes with M_1 to be the omni-directional and M_2 to be the cardioid lobe, $Ang[i_{ang}], i_{ang} \in [1, 3]$ the angle between the microphone's and source's central axis, with $Ang[1] = 0^\circ$, $Ang[2] = 30^\circ$ and $Ang[3] = 45^\circ$, since the effect is minimal for lower angle variations [17], and $SPL_S[i_S]$ and $SPL_N[i_N]$ the different SPLs for the sound and noise source respectively, then the following recording sets, R_i , were created:

$$R_1 = \{D[i_d], M_1, Ang[1], SPL_N[i_N]\} \quad (5)$$

$$R_2 = \{D[i_d], M_1, Ang[1], SPL_S[i_S]\} \quad (6)$$

$$R_3 = \{D[i_d], M_2, Ang[1], SPL_N[i_N]\} \quad (7)$$

$$R_4 = \{D[i_d], M_2, Ang[1], SPL_S[i_S]\} \quad (8)$$

$$R_5 = \{D[i'_d], M_2, Ang[2], SPL_N[i_N]\} \quad (9)$$

$$R_6 = \{D[i'_d], M_2, Ang[2], SPL_S[i_S]\} \quad (10)$$

$$R_7 = \{D[i'_d], M_2, Ang[3], SPL_N[i_N]\} \quad (11)$$

$$R_8 = \{D[i'_d], M_2, Ang[3], SPL_S[i_S]\} \quad (12)$$

where $i_d \in [1, 12]$, $i'_d \in [1, 5]$, $i_S \in [1, 3]$ and $i_N \in [1, 5]$. Must be noted that in the cases where a recording contains both SPL_S and SPL_N , these two were physically

apparent and recorded at the same time. The calibration of the SPL for each sound source (SPL_S and SPL_N) was performed with the SLM, at the point of the recording microphone, for each source-microphone distance separately, and without any other source active.

The recordings in the overall data set were all time trimmed to 15 seconds in order to contain exactly the produced signals from all cases. The audio data from the 15 seconds long recordings were saved under standard CD quality, i.e. sampling frequency equal to 44.1kHz and 16 bit sample length, using the typical wave file format. The latter audio files were utilized by the SSE process presented immediately next, organized in the sets R'_1 to R'_8 , in accordance to Equations 5 to 12.

3.2 Source separation evaluation

For evaluation purposes, pairs of audio files from the recording sets were utilized as input to the SSE process. Each pair contains two audio files, one containing the noise-free recording (i.e. the desired source is active only; an audio file from recording sets with even index), considered as the estimated source in terms of the SSE process, and the audio file from the recording with the noise source active (i.e. audio file from recording sets with odd index).

The SIR was computed for the recording set pairs: a) R'_1 and R'_2 , b) R'_3 and R'_4 , c) R'_5 and R'_6 , and d) R'_7 and R'_8 . As can be seen from Equations 5 to 12, the recording sets with odd indices contain recordings with the noise source active and recording sets with even indices contain recordings with the desired source active. Also, each of the pairs a) to d), contains recording sets with the same microphone type and the same angle between the microphone and the sound source. Thus, the input for the calculation of the SIR for one recording pair was audio from each of the recording sets in this pair and with the same indices i_d/i'_d , i_S , i_N , and i_{ang} .

4 Results

The results from the above experimental process are organized in 12 figures, corresponding to the different combinations of microphone types, placement angles and the produced SPL. Specifically, in Figure 4 are the results for the cardioid microphone and for zero degrees angle between the microphone and the source.

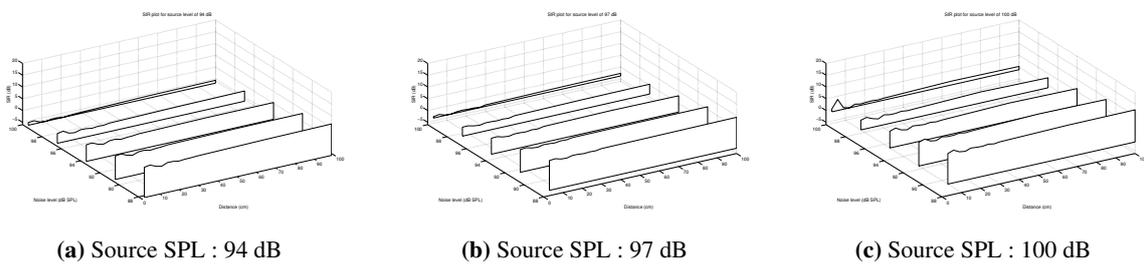


Fig. 3: SIR of omnidirectional microphone over various sound pressure levels of *source*, with respect to distance D and sound pressure level of *noise*.

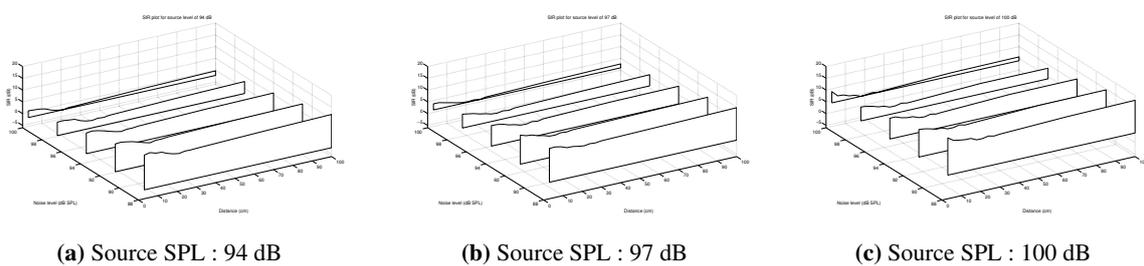


Fig. 4: SIR of cardioid microphone over various sound pressure levels of *source*, with respect to distance D and sound pressure level of *noise*.

In Figure 3 are the results for the omni-directional microphone. In Figure 5 are the results for the cardioid microphone with an angle of 30° between the microphone and the source and in Figure 6 the results for the cardioid microphone and with an angle of 45° between the microphone and the source are shown.

5 Discussion

The results presented in the previous section portray the expected fact that the lower SPL of the noise results in better performance of the close miking technique. Also, a general trend from all figures and sub-figures is that the SPL of the source and the SIR seems to be analogous. This means that the higher the SPL of the source, the higher the SIR. These observations are in accordance with the general purpose and expectations of the close miking technique.

Focusing on Figures 3 and 4, one can see that in all cases the cardioid microphone outperforms the omnidirectional one. The SIR values obtained with the cardioid microphone are almost double of the SIR values obtained with the omnidirectional. In addition, in both cases the maximum performance of the close miking technique seems to be achieved for a 5 cm distance

between the source and the microphone. After that distance, a reduction of the SIR is observed for both cases. For the omni-directional case, the reduction is between 10 and 20 centimeters (cm), while for the cardioid microphone case, the reduction is observed between 20 and 40 cm. Followed by that reduction, the SIR rises up to a limit achieved around 70 cm.

Focusing on Figures 5 and 6, one can also observe better interference reduction (higher SIR values) for all source SPL, distances, and noise SPL when compared to the previous two cases. Additionally, in the same cases, i.e. Figures 5 and 6, there is a maximum of SIR around 12 to 14 cm. This comes in contrast with the previous two cases where the peak was observed below 10 cm. Since for the cases of Figures 5 and 6 we did not perform measurements with distances greater than 15 cm, we cannot conclude if the SIR curves would exhibit the similar behavior as the SIR curves from Figures 3 and 4, i.e. a deep at certain distance followed by a small increase towards a high limit of the SIR.

The SIR values obtained with the placement of the cardioid microphones with an angle are almost three times the values of the SIR that were obtained with the

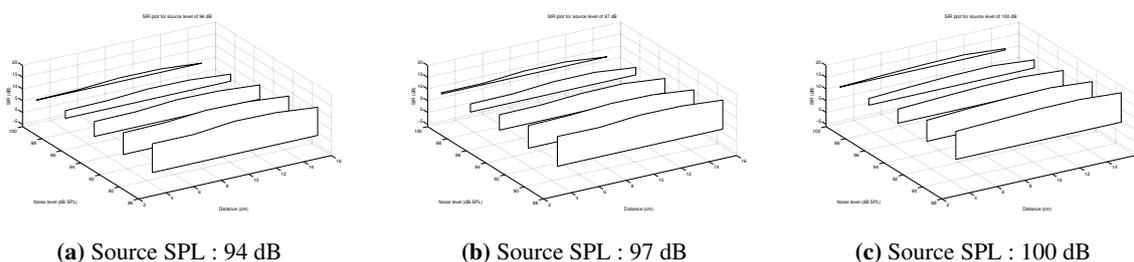


Fig. 5: SIR of cardioid microphone, with an angle of 30° , over various sound pressure levels of *source*, with respect to distance D and sound pressure level of *noise*.

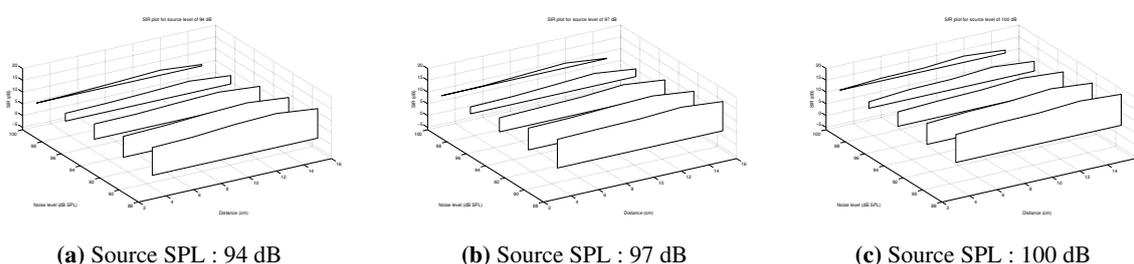


Fig. 6: SIR of cardioid microphone, with an angle of 45° , over various sound pressure levels of *source*, with respect to distance D and sound pressure level of *noise*.

other two cases. This clearly indicates that placing a cardioid microphone with an angle against the central axis of the noise results in better performance of the close miking technique. These values of SIR in the corresponding peaks are almost three times the peak SIR values from the rest two cases of microphone types and angles of placement. This clearly indicates the outperformance of the cardioid microphones placed with an angle versus the cardioid microphone placed without an angle and the omni-directional microphone cases. Finally, between the two different angular placements of the cardioid microphones, there is not any notable difference with the current experimental setup.

6 Conclusions

The work at hand performed a quantitative analysis of the source separation capabilities of the close miking technique. Since this technique is a mechanical source separation method, the present work applies a quantitative analysis of the actual close miking technique. This analysis is performed with two different microphone types, three different angular placements of the microphones, 12 different distances between the microphone and the source, three different source SPL, and, finally, under five different noise SPL values.

The results obtained clearly indicate that the best performance of close miking is achieved when the microphone has a cardioid lobe, placed with an angle of 30 or 40 degrees with respect to the central axis of the source and in distance of around 12 cm.

Feature measurements and studies could, potentially, show the effect of the height of the microphone in the close miking technique. Finally, there would increased interest in a subjective evaluation of the quality of the source separation with close miking with different types of microphones and different angular placements of the microphones with respect to the source.

7 Acknowledgements

The authors would like to thank the Department of Technology of Sound and Musical Instruments, Technological Educational Institute of Ionian Islands, for providing the equipment for the measurements. Part of the research leading to these results has received funding from: i) the European Research Council under the European Unions H2020 Framework Programme through ERC Grant Agreement 637422 EVERSOUND, and ii) the European Union's H2020

Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet.

References

- [1] D. M. Huber and R. E. Runstein, *Modern Recording Techniques*. Burlington, MA: Focal Press, 2005.
- [2] E. K. Kokkinis, E. Georganti, and J. Mourjopoulos, "Statistical properties of the close-microphone responses," in *Proceedings of 132nd Audio Engineering Society Convention*. Budapest, Hungary: Audio Engineering Society, April 2012.
- [3] R. Liu and S. Li, "A review on music source separation," in *Information, Computing and Telecommunication, 2009. YC-ICT '09. IEEE Youth Conference on*, Sept 2009, pp. 343–346.
- [4] D. Fitzgerald, "Upmixing from mono - a source separation approach," in *17th International Conference on Digital Signal Processing (DSP)*. Corfu, Greece: IEEE, July 2011, pp. 1–7.
- [5] A. Floros and N. A. Tatlas, "Spatial enhancement for immersive stereo audio applications," in *2011 17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–7.
- [6] K. Drossos, S. I. Mimilakis, A. Floros, and N. G. Kanellopoulos, "Stereo goes mobile: Spatial enhancement for short-distance loudspeaker setups," in *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Peiraeus, Greece, July 2012, pp. 432 – 435.
- [7] E. Cano, M. Plumbley, and C. Dittmar, "Phase-based harmonic percussive separation," in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, Singapore, September 2014.
- [8] A. Liutkus, Z. Rafii, B. Pardo, D. Fitzgerald, and L. Daudet, "Kernel Spectrogram models for source separation," in *HSCMA*, Nancy, France, May 2014. [Online]. Available: <https://hal.inria.fr/hal-00959384>
- [9] Z. Rafii, Z. Duan, and B. Pardo, "Combining rhythm-based and pitch-based methods for background and melody separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1884–1893, Dec 2014.
- [10] S. I. Mimilakis, E. Cano, J. Abesser, and G. Schuller, "New sonorities for jazz recordings: Separation and mixing using deep neural networks," in *Audio Engineering Society 2nd Workshop on Intelligent Music Production*, London, UK, Sept. 2016.
- [11] S. Choi, A. Cichocki, H.-M. Park, and S.-Y. Lee, "Blind source separation and independent component analysis: A review," *Neural Information Processing - Letters and Reviews*, vol. 6, no. 1, pp. 1–57, Jan 2005.
- [12] M. Lohmari, M. Naceur, and M. Boussema, "A spectral and spatial source separation of multi-spectral images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 12, pp. 3659–3673, Dec 2006.
- [13] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions in Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [14] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.
- [15] B. Bartlett, "Tonal effects of close microphone placement," *Journal of Audio Engineering Society*, vol. 29, no. 10, pp. 726–738, October 1981.
- [16] E. B. Brixen, "Near-field registration of the human voice: Spectral changes due to positions," in *Proceedings of 104th Audio Engineering Society Convention*. Amsterdam, The Netherlands: Audio Engineering Society, May 1998.
- [17] A. Case, "Recording electric guitar - the science and the myth," *Journal of the Audio Engineering Society*, vol. 58, no. 1/2, 2010.
- [18] C. Févotte, R. Gribonval, and E. Vincent, "Bss eval toolbox user guide," IRISA, Rennes, France, Technical Report 1706, April 2005.