HAL
open science

The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio source separation -

Shoko Araki, Francesco Nesta, Emmanuel Vincent, Zbynek Koldovsky, Guido Nolte, Andreas Ziehe, Alexis Benichoux

To cite this version:

Shoko Araki, Francesco Nesta, Emmanuel Vincent, Zbynek Koldovsky, Guido Nolte, et al.. The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio source separation -. 10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA), Mar 2012, Tel Aviv, Israel. pp.414-422. hal-00655394

HAL Id: hal-00655394

https://inria.hal.science/hal-00655394

Submitted on 28 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio source separation -

Shoko Araki[1], Francesco Nesta[2], Emmanuel Vincent[3], Zbynek Koldovsky[4], Guido Nolte[5], Andreas Ziehe[5], and Alexis Benichoux[3]

[1] NTT Communication Science Labs., NTT Corporation, Japan
[2] Fondazione Bruno Kessler - Irst, Center of Information Technology, Italy
[3] INRIA, Centre Inria Rennes - Bretagne Atlantique, France
[4] Technical University of Liberec, Czech Republic
[5] Fraunhofer Institute FIRST IDA, Germany

Abstract. This paper summarizes the audio part of the 2011 community-based Signal Separation Evaluation Campaign (SiSEC2011). Four speech and music datasets were contributed, including datasets recorded in noisy or dynamic environments and a subset of the SiSEC2010 datasets. The participants addressed one or more tasks out of four source separation tasks, and the results for each task were evaluated using different objective performance criteria. We provide an overview of the audio datasets, tasks and criteria. We also report the results achieved with the submitted systems, and discuss organization strategies for future campaigns.

1 Introduction

The Signal Separation Evaluation Campaign (SiSEC) is a regular campaign focused on the evaluation of methods for signal separation. It was built on the experience of previous evaluation campaigns (e.g., the MLSP'05 Data Analysis Competition[1], the PASCAL Speech Separation Challenge [1], and the Stereo Audio Source Separation Evaluation Campaign (SASSEC)) and has been organized since 2008 [2]. SiSEC is not a competition but a community-based scientific evaluation whose aspects are publicly defined. A call for participation precedes the evaluation and aims to define datasets, tasks and evaluation criteria.

This article describes the audio part of SiSEC 2011. In response to the feedback received at SiSEC2008 and SiSEC2010, previous datasets were reorganized as follows:

1. datasets sharing similar scenarios were merged in order to remove some redundancies (e.g. the 2-channel 1-source dataset of the "Source separation in the presence of real-world background noise" task of SiSEC2010 was merged with a new dataset from the PASCAL CHiME Challenge [3]).
2. tasks with little participation in the previous campaign were excluded;

[1] `http://mlsp2005.conwiz.dk/index.php@id=30.html`

3. unrealistic data was removed (e.g. the synthetic mixtures of the "Underdetermined speech and music separation" task of SiSEC2010 were eliminated and fresh real-world data was provided).

In general the new campaign was designed so as to better match with real-world scenarios. We believe this data could be of high potential interest for many audio applications in the future years. Specifically, the new datasets embody more realistic features such as a) more reverberant rooms b) real-world diffuse or rapidly varying noise c) source movements.

Datasets and tasks are specified in Section 2 and the obtained outcomes are summarized in Section 3. Due to the variety of the submissions, we focus on the general outcomes of the campaign and ask readers to refer to `http://sisec.wiki.irisa.fr/` for further details.

2 Specifications

This section describes the tasks, datasets and evaluation criteria, which were specified in a collaborative fashion. A few initial specifications were first suggested by the organizers. Potential participants were then invited to provide feedback and contribute additional specifications through the wiki or the mailing list.

2.1 Tasks

For each dataset, audio mixtures spanning a variety of mixing conditions are provided. The channels $x_i(t)$ $(1 \leq i \leq I)$ of each mixture signal were generally obtained as $x_i(t) = \sum_{j=1}^{J} s_{ij}^{\mathrm{img}}(t)$, where $s_{ij}^{\mathrm{img}}(t)$ is the *spatial image* of source j $(1 \leq j \leq J)$ on channel i [2]. For point sources, $s_{ij}^{\mathrm{img}}(t) = \sum_{\tau} a_{ij}(t-\tau, \tau) s_j(t-\tau)$ where $s_j(t)$ are the source signals and $a_{ij}(t, \tau)$ the (possibly time-varying) mixing filters. For these mixtures, we specified the following four tasks:

T1 Source counting
T2 Source signal estimation
T3 Source spatial image estimation
T4 Source DOA estimation

These tasks consist in finding, respectively: (T1) the number of sources J, (T2) the source signals $s_j(t)$, (T3) the spatial images $s_{ij}^{\mathrm{img}}(t)$ of the sources for all channels i, and (T4) the direction of arrival (DOA) of each source. Participants were asked to submit the results of their systems for T2 and/or T3, and optionally for T1 and/or T4.

Two oracle systems were also considered for benchmarking task T3: ideal binary masking over a short-time Fourier transform (STFT) [4] (O1) and over a cochleagram [5] (O2). These systems require the true source spatial images and provide upper bounds on the performance of binary masking-based systems.

2.2 Datasets

Four distinct datasets were provided for SiSEC2011:

D1 Under-determined speech and music mixtures
This dataset includes the stereo dataset D1 from SiSEC2010 [6], and a fresh dataset containing ten 3-channel mixtures of four audio sources of 10 s duration, sampled at 16 kHz. For 3-channel data we used a linear microphone array. The room reverberation time (RT) for the fresh dataset was 130 ms or 380 ms. Instantaneous mixtures are also included. Tasks T1, T2 and T3 are considered.

D2 Determined convolutive mixtures under dynamic conditions
This dataset consists of two kinds of scenarios: (1) random source activity of multiple sources in multiple static locations, and (2) a source continuously moving and overlapped with a source in a fixed or random location. The former aims to simulate a meeting scenario, where multiple talkers utter from fixed locations and their activity is unknown. The latter was specifically designed to evaluate systems able to handle dynamic variations of the mixing parameters. Due to the challenging reverberation conditions, datasets with different difficulty levels were provided (i.e. varying the source-array distance and the angular direction of simultaneously active sources). In the mixtures, two speakers are simultaneously active at most. In these datasets 4-channel mixtures are provided, and participants can decide whether using all the available channels or only a subset of them. The recordings were obtained in a real room of size ($6 \times 5 \times 4$ m) with an estimated RT of 700 ms. For both the datasets the signals were recorded by a uniform linear array of four (directional) microphones with a different spacing (of about 2 cm, 8 cm, and 18 cm) and sampled at 16 kHz. T2 and T3 are considered for this dataset.

D3 Professionally produced music recordings
According to many positive requests from the community, we decided to repeat this dataset in SiSEC2011. This dataset contains stereo music signals sampled at 44.1 kHz, including those of the dataset D3 from SiSEC2010 [6]. In addition to the 20-second snips to be separated, full-length recordings are provided as well. The mixtures were created by sound engineers, and the ways of mixing and the mixing effects applied are unknown. Task T3 is imposed on this dataset.

D4 Two-channel mixtures of speech and real-world background noise
This task aims to evaluate source separation and denoising techniques in the context of speech enhancement by merging two datasets: the dataset D3 from SiSEC2010 [6] and the CHiME corpus [3]. Both datasets consist of two-channel mixtures of one speech source and real-world background noise sampled at 16 kHz. In both datasets, the spatial image of the background noise was recorded in real-world environments: a subway car, a cafeteria, or a square for the former, and a British family living room for the latter. Tasks T2, T3 and T4 are evaluated for this dataset.

All datasets include both test and development data, and the CHiME corpus in D4 also includes training data. The true source signals and source positions underlying the test data were hidden to the participants, while they were provided for the development data. The true number of speech/music sources was always available.

2.3 Evaluation criteria

Tasks T2 and T3 were evaluated via the criteria in the BSS Eval toolbox termed signal to distortion ratio (SDR), source image to spatial distortion ratio (ISR), signal to interference ratio (SIR) and signal to artifacts ratio (SAR) [7,2]. In addition, version 2.0 of the PEASS toolbox [8,9] was used to assess the perceptual quality of the estimated signals for stereo data according to four performance measures akin to SDR, ISR, SIR and SAR: overall perceptual score (OPS), target-related perceptual score (TPS), interference-related perceptual score (IPS) and artifact-related perceptual score (APS).

Task T4 was evaluated by the absolute difference between the true and estimated DOAs.

3 Results

Despite the challenging specifications of each dataset, a remarkable participation was obtained. A total of 32 submissions were received from 18 different research centers. Many participants were involved in SiSEC for the first time, revealing a positive enlargement of the community. Tables 1 to 5 summarize the average performance obtained over the submitted algorithms. The algorithm details and all the results are available at `http://sisec2011.wiki.irisa.fr/tiki-index.php`. It should be noted that the presented values are the absolute values, not the improvements from the values for mixtures.

By comparison with the previous SiSEC, an unexpected high participation was observed for dataset D3. This trend seems to be in line with the recent increasing interest in NMF-based techniques, which have shown to marry well with the task of music recordings separation. The traditional dataset D1 has attracted a satisfactory amount of new participants, although the performance improvement seems to be still limited by the amount of reverberation. The datasets D2 and D4, aimed to simulate more realistic real-world scenarios, have attracted a sufficient but yet limited number of participants, probably due to the intrinsic difficulty of the data. Furthermore, the proposed algorithms do not seem to be equivalently effective in all the scenarios, which reveals that the acoustic source separation is still an open problem for real-world applications.

Note that a close analysis of each table is beyond the scope of this paper and a more detailed investigations will be discussed at the LVA/ICA 2012 conference.

[2] The system details can be found at the SiSEC2011 wiki.
[3] Figure computed by averaging over an incomplete set of mixtures.
[4] The same algorithm as [15] without the Wiener-Filter post-processing.
[5] The values for "2mic." are from SiSEC2010 submissions.
[6] Algorithm derived from the weighted Natural Gradient in [15].
[7] The same algorithm as S2 with additional Binary Masking post-processing.
[8] The same algorithm as S4 with additional TF post-processing.
[9] The same algorithm as S6 with additional Wiener-Filter like post-processing.
[10] The same algorithm as S5 with different parameter settings.

Table 1. Average performance for task T2 or T3 for instantaneous dataset D1. 2 mic: average over test & test2 datasets, 3 mic: average over test3 dataset.

System	2 mic, 3 speech				2 mic, 3 music				2 mic, 4 speech				3 mic, 4 speech			
	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR
	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS
S1 [10]	13.4	25.7	21.2	14.5	16.6	27.0	23.1	20.5	8.9	17.2	15.4	9.7	-	-	-	-
	43.9	55.4	61.0	58.6	52.3	58.9	66.6	55.5	42.4	65.1	62.2	47.0	-	-	-	-
S2 [11]	7.9	-	13.6	9.7	6.9	-	12.2	10.2	3.0	-	8.0	5.8	11.7	-	19.1	12.6
	43.2	-	61.7	25.6	40.0	-	62.7	10.6	29.8	-	46.7	10.7	39.7	-	64.0	37.7
S3 [2]	8.8	-	19.8	9.4	5.9	-	13.9	8.5	5.8	-	16.4	6.7	8.0	-	20.5	8.4
	38.5	-	75.3	10.4	35.7	-	68.9	16.2	35.7	-	65.7	12.1	38.6	-	75.5	9.2
O1	10.8	20.1	21.7	11.1	10.4	18.0	18.8	12.5	9.1	17.6	20.0	9.3	-	-	-	-
	38.9	61.8	70.5	37.7	33.3	48.5	64.8	34.2	27.1	57.7	71.8	21.9	-	-	-	-
O2	8.5	15.7	17.4	9.1	9.0	14.1	18.1	11.3	7.5	13.7	16.4	8.1	-	-	-	-
	24.0	29.8	72.4	20.0	30.4	28.3	69.5	21.6	22.0	20.9	70.8	13.1	-	-	-	-

Table 2. Average performance for task T3 for convolutive dataset D1. 2 mic: average over test & test2 datasets, 3 mic: average over test3 dataset. The values are averaged over all the reverberation time.

System	2 mic, 3 speech				2 mic, 3 music				2 mic, 4 speech				3 mic, 4 speech			
	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR
	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS	OPS	TPS	IPS	APS
S1 [10]	3.4	8.2	6.4	7.8	2.1	7.2	4.4	10.0	2.0	6.1	3.8	5.5	-	-	-	-
	27.9	47.6	38.5	55.9	21.8	33.3	29.7	39.1	31.2	46.9	39.2	48.9	-	-	-	-
S2 [12][3]	1.8	4.1	2.2	4.3	-	-	-	-	1.1	3.3	0.1	2.8	1.6	3.4	1.8	3.4
	21.4	33.9	43.8	38.8	-	-	-	-	19.9	27.4	40.5	35.8	20.1	33.6	53.6	34.0
S3 [13]	5.3	9.3	7.7	10.0	-	-	-	-	-	-	-	-	-	-	-	-
	26.9	51.5	35.3	62.1	-	-	-	-	-	-	-	-	-	-	-	-
S4 [14]	4.3	9.0	6.9	8.8	0.2	4.8	0.6	7.1	1.4	4.7	1.4	6.2	1.2	2.9	2.4	5.6
	25.4	49.9	38.1	56.3	19.7	28.9	19.6	42.0	27.2	41.7	28.4	50.6	29.7	58.8	59.3	30.0
S5 [15][4]	5.4	8.9	8.9	9.1	2.8	6.8	5.0	8.8	3.3	6.3	5.6	6.3	-	-	-	-
	34.4	59.8	52.2	57.7	27.3	43.8	37.8	43.1	35.0	58.3	47.9	49.2	-	-	-	-
S6 [15]	6.1	10.9	10.5	9.1	3.0	7.6	5.4	8.9	3.6	7.4	6.9	6.5	-	-	-	-
	38.3	58.8	53.7	55.0	26.5	39.7	38.0	42.0	35.1	56.0	49.5	48.7	-	-	-	-
S7 [16][5]	5.8	10.8	10.3	8.2	1.7	6.3	3.0	6.7	3.2	7.3	5.9	5.6	5.3	10.0	9.9	7.5
	37.2	61.9	52.3	51.4	22.4	35.9	32.6	38.8	30.3	54.6	48.2	42.6	31.1	63.1	61.6	34.4
O1	10.2	18.7	20.2	10.7	9.9	16.9	18.0	11.0	8.7	16.4	18.5	9.1	-	-	-	-
	43.4	63.1	69.9	45.1	36.0	52.7	64.2	40.6	36.2	63.1	71.9	34.3	-	-	-	-
O2	7.6	13.8	16.8	8.2	7.1	12.3	14.5	8.3	6.1	11.3	15.1	6.3	-	-	-	-
	26.7	41.9	72.7	23.8	25.9	21.8	69.3	19.0	23.7	37.8	72.4	18.8	-	-	-	-

4 Conclusion

This paper presented the specifications of SiSEC2011 and summarized the performance obtained over all the submissions. This time, in accordance with discussions at previous SiSECs, we carefully selected the datasets and tasks in a collaborative fashion. Ultimately, four datasets and tasks were provided which attracted many submissions from 18 research institutions.

Despite some open challenges which still do not allow us to provide an unambiguous evaluation of all the submissions, we hope that SiSEC2011 will continue to represent a common platform for sharing new ideas and perspectives in the source separation research field. We believe SiSEC2011 data could be of high potential interest for many audio applications and encourage the community to use it as a reference for future evaluations.

Following the experience maturated till this campaign, new criteria seem needed for better evaluating more realistic scenarios, such as source separation

Table 3. Average performance for dataset D2, "random source activity of multiple sources in multiple static locations" (top) and "a continuously moving active source overlapped with a source in a fixed or random location" (bottom). All the signals are evaluated as source signal and spatial source signal estimates. For more details see `http://www.irisa.fr/metiss/SiSEC11/dynamic/main.html`.

System	Source signal estimation						Spatial image estimation							
	SDR	SIR	SAR	OPS	IPS	APS	SDRi	SIRi	SARi	ISRi	OPSi	TPSi	IPSi	APSi
S1 [17]	3.5	9.2	7.0	30.5	69.5	11.9	2.0	6.0	7.5	3.0	29.5	30.3	67.2	27.3
S2 [15,18][6]	3.7	6.2	9.3	35.4	53.2	20.9	2.6	4.1	12.1	4.4	33.1	48.7	51.4	41.3
S3 [15,18][7]	3.5	7.3	7.5	31.8	63.0	7.1	2.3	5.2	10.1	3.6	29.6	41.1	61.6	32.6
S4 [19]	2.2	6.6	6.1	28.5	66.9	4.1	2.1	4.5	7.0	3.6	27.7	41.7	66.5	24.6
S5 [19][8]	2.3	7.4	5.9	26.9	70.6	3.0	1.9	4.8	6.9	3.3	25.9	32.1	70.3	21.1
S6 [20,21]	1.8	7.3	5.1	26.9	71.9	1.6	1.5	5.4	5.7	2.3	26.4	31.1	71.8	23.5
S7 [20,21][9]	3.1	10.6	5.3	27.1	72.6	1.9	1.2	6.7	6.2	1.7	27.0	20.4	72.3	22.5

System	Source signal estimation						Spatial image estimation							
	SDR	SIR	SAR	OPS	IPS	APS	SDRi	SIRi	SARi	ISRi	OPSi	TPSi	IPSi	APSi
S1 [17]	2.5	7.7	6.2	30.9	66.5	6.2	1.3	6.3	8.0	1.9	29.0	31.3	65.1	30.8
S2 [15,18][6]	4.2	7.1	9.1	36.2	55.3	21.3	4.3	5.5	12.8	7.0	33.9	59.5	53.4	40.8
S3 [15,18][7]	4.0	8.5	7.3	32.2	65.9	6.7	4.0	6.9	10.9	6.1	30.2	53.8	64.2	30.7
S4 [19]	3.3	10.5	5.3	26.0	77.1	1.4	2.5	8.0	7.1	3.8	26.9	39.9	76.8	18.1
S5 [19][8]	3.5	11.0	5.4	25.7	78.2	1.3	2.5	8.5	7.2	3.8	27.0	36.5	78.1	18.2
S6 [20,21]	2.4	8.5	5.1	28.1	71.3	2.6	2.0	7.2	6.7	3.0	27.2	36.7	70.8	24.6
S7 [20,21][9]	3.8	12.9	5.4	28.4	72.1	2.9	1.7	9.0	7.6	2.2	28.0	23.8	70.3	24.7

Table 4. Average performance for T2/T3 for testset of D3. The results only for the vocal and drum tracks, which most of the submissions addressed, are summarized. S4, S6 and S8 addressed all the specified tracks. Complete results can be found at SiSEC2011 wiki.

System	Vocal								Drums							
	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS
S1 [22]	3.8	6.2	Inf	3.1	22.4	28.8	59.0	30.8	-	-	-	-	-	-	-	-
S2 [23]	4.5	6.8	Inf	3.8	26.6	29.3	62.7	29.5	-	-	-	-	-	-	-	-
S3 [24]	-2.7	-0.8	Inf	-7.2	22.5	5.0	64.6	10.0	-	-	-	-	-	-	-	-
S4 [25]	-5.5	-1.3	7.0	3.6	15.7	15.7	27.6	15.3	-7.1	-2.9	2.9	2.7	23.3	23.2	50.4	12.8
S5 [26]	2.4	8.5	Inf	0.1	25.2	15.9	70.5	11.6	-0.2	2.2	5.9	-5.4	23.6	44.0	67.8	2.1
S6 [10]	3.1	8.1	7.7	3.7	24.4	37.1	20.8	54.3	2.0	4.3	2.9	2.1	29.3	54.6	28.9	50.4
S7 [27] [3]	4.1	10.7	6.3	7.3	41.6	74.5	61.2	40.0	-	-	-	-	-	-	-	-
S8 [2]	3.0	7.7	9.0	2.4	19.4	28.7	55.2	31.9	1.7	2.1	11.6	1.1	20.7	19.9	58.7	9.0
O1	6.2	22.1	22.3	6.2	28.4	69.1	69.1	16.6	6.3	24.6	23.2	6.2	25.7	73.7	74.5	2.8
O2	4.7	17.0	16.3	4.7	23.6	38.1	61.9	14.8	1.4	2.7	17.3	0.4	18.1	32.2	69.4	4.6

involving dereverberation or tracking of time-varying mixing conditions. Furthermore, it would be worthwhile to investigate on new objective evaluation criteria more related to the separation filter accuracy rather than to the quality of the signals itself, with the hope of minimizing the presence of outliers. With this regard, we invite all willing participants to join a continuous collaborative discussion on the future of source separation evaluation.

Acknowledgments: We thank all the participants and data providers. We also thank to all the researchers who gave their opinions on the wiki and the mailing-list. Special thanks go to Mr. Antoine Liutkus, who proposed to repeat the dataset D3, and encouraged researches of the field to join in SiSEC2011.

Table 5. Average performance for task T2/T3 for test dataset D4. Outdoor and indoor indicates the recordings 2ch-1src in the dataset D3 from SiSEC2010 [6] and the CHiME corpus [3], respectively. Performance of S1 are evaluated on the source signal estimates (i.e. 'src' files), while the remaining systems are evaluated on the spatial source image estimates (i.e. 'sim' files).

System	Outdoor								Indoor							
	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS
S1[2]	-	-	-	-	-	-	-	-	1.8	-	6.1	6.2	-	-	-	-
S2[2]	-1.8	13.3	-0.7	16.0	11.2	46.7	35.8	81.3	-	-	-	-	-	-	-	-
S3[28]	8.8	13.4	15.1	14.4	14.6	50.8	39.3	80.0	-1.7	8.0	0.7	14.1	20.9	50.3	30.0	69.3
S4[28][2]	6.1	13.1	13.4	10.9	43.8	59.8	58.2	57.6	-	-	-	-	-	-	-	-
S5[29,15]	3.5	16.6	6.4	12.2	33.4	59.0	57.5	70.0	6.0	7.3	16.5	11.0	37.3	43.5	68.7	38.9
S6[29,15][10]	3.4	17.6	5.8	12.8	29.6	58.2	55.2	73.3	8.0	11.0	14.7	12.0	38.5	55.3	65.0	49.9
S7[10]	-	-	-	-	-	-	-	-	5.4	7.3	14.0	11.7	35.2	62.2	49.9	51.0
S8[2]	4.0	7.0	8.8	7.6	36.5	51.2	63.4	41.8	-	-	-	-	-	-	-	-
baseline [30]	2.4	8.9	7.2	8.7	22.2	49.9	47.6	64.3	1.7	3.5	5.2	8.6	29.3	34.8	44.1	37.5
O1	15.8	27.1	24.3	16.9	51.3	65.9	75.6	45.5	14.5	20.9	22.7	16.6	53.5	67.2	73.6	57.0

References

1. Cooke, M.P., Hershey, J., Rennie, S.: Monaural speech separation and recognition challenge. Computer Speech and Language **24** (2010) 1–15
2. Vincent, E., Araki, S., Theis, F.J., Nolte, G., Bofill, P., Sawada, H., Ozerov, A., Gowreesunker, B.V., Lutter, D., Duong, N.Q.K.: The Signal Separation Evaluation Campaign (2007–2010): Achievements and remaining challenges. Signal Processing (to appear)
3. Christensen, H., Barker, J., Ma, N., Green, P.: The CHiME corpus: a resource and a challenge for computational hearing in multisource environments. In: Proc. Interspeech. (2010) 1918–1921
4. Vincent, E., Gribonval, R., Plumbley, M.D.: Oracle estimators for the benchmarking of source separation algorithms. Signal Processing **87**(8) (2007) 1933–1950
5. Wang, D.L.: On ideal binary mask as the computational goal of auditory scene analysis. In: Speech Separation by Humans and Machines. Springer (2005)
6. Araki, S., Ozerov, A., Gowreesunker, V., Sawada, H., Theis, F., Nolte, G., Lutter, D., Duong, N.Q.K.: The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation. In: Proc. LVA/ICA. (2010) 114–122
7. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. IEEE Trans. on Audio, Speech and Language Processing **14**(4) (2006) 1462–1469
8. Emiya, V., Vincent, E., Harlander, N., Hohmann, V.: Subjective and objective quality assessment of audio source separation. IEEE Trans. on Audio, Speech and Language Processing **19**(7) (2011) 2046–2057
9. Vincent, E.: Improved perceptual metrics for the evaluation of audio source separation. In: Proc. LVA/ICA. (2012) (to appear).
10. Ozerov, A., Vincent, E., Bimbot, F.: A general flexible framework for the handling of prior information in audio source separation. IEEE Trans. on Audio, Speech and Language Processing **PP**(99) (2011) 1
11. Makkiabadi, B., Sanei, S., Marshall, D.: A k-subspace based tensor factorization approach for under-determined blind identification. In: Proc. ASILOMAR 2010. (2010)
12. Hirasawa, Y., Yasuraoka, N., Takahashi, T., Ogata, T., Okuno, H.G.: A GMM sound source model for blind speech separation in under-determined conditions. In: Proc. LVA/ICA 2012. (to appear).

13. Iso, K., Araki, S., Makino, S., Nakatani, T., Sawada, H., Yamada, T., Nakamura, A.: Blind source separation of mixed speech in a high reverberation environment. In: Proc. HSCMA2011. (2011) 36–39
14. Cho, J., Choi, J., Yoo, C.D.: Underdetermined convolutive blind source separation using a novel mixing matrix estimation and MMSE-based source estimation. In: Proc. MLSP2011. (2011)
15. Nesta, F., Omologo, M.: Convolutive underdetermined source separation through weighted interleaved ICA and spatio-temporal correlation. In: Proc. LVA/ICA 2012. (to appear).
16. Sawada, H., Araki, S., Makino, S.: A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures. In: Proc. WASPAA. (2007) 139–142
17. Malek, J., Koldovsky, Z., Tichavsky, P.: Semi-blind source separation based on ICA and overlapped speech detection. In: Proc. LVA/ICA 2012. (to appear).
18. Nesta, F., Omologo, M.: Generalized state coherence transform for multidimensional TDOA estimation of multiple sources. Audio, Speech, and Language Processing, IEEE Transactions on **20**(1) (2012) 246 –260
19. Loesch, B., Yang, B.: Blind source separation based on time-frequency sparseness in the presence of spatial aliasing. In: Proc. LVA/ICA. (2010) 1–8
20. Loesch, B., Yang, B.: Adaptive segmentation and separation of determined convolutive mixtures under dynamic conditions. In: Proc. LVA/ICA. (2010) 41–48
21. Loesch, B., Nesta, F., Yang, B.: On the robustness of the multidimensional state coherence transform for solving the permutation problem of frequency-domain ICA. In: Proc. ICASSP. (2010) 225 –228
22. Durrieu, J.L., David, B., Richard, G.: A musically motivated mid-level representation for pitch estimation and musical audio source separation. IEEE Journal of Selected Topics on Signal Processing **5**(6) (2011) 1180–1191
23. Durrieu, J.L., Thiran, J.P.: Musical audio source separation based on user-selected F0 track. In: Proc. LVA/ICA 2012. (to appear).
24. Cano, E., Dittmar, C., Schuller, G.: Interaction of phase, magnitude and location of harmonic components in the perceived quality of extracted solo signals. In: Proc. AES. (2011)
25. Spiertz, M., Gnann, V.: Note clustering based on 2D source-filter modeling for underdetermined blind source separation. In: Proc. AES. (2011)
26. Marxer, R., Janer, J.: A Tikhonov regularization method for spectrum decomposition in low latency audio source separation. In: Proc. ICASSP2012. (to appear).
27. Sawada, H., Kameoka, H., Araki, S., Ueda, N.: Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization. In: Proc. ICASSP2012. (to appear).
28. Mustiere, F., Bolic, M., Bouchard, M.: Real-world particle filtering-based speech enhancement. In: Proc. CIP. (2010) 75 –80
29. Nesta, F., Matassoni, M.: Robust automatic speech recognition through on-line semi-blind source extraction. In: Proc. CHIME. (2011)
30. Blandin, C., Ozerov, A., Vincent, E.: Multi-source TDOA estimation in reverberant audio using angular spectra and clustering. Signal Processing (to appear)