The 2010 Signal Separation Evaluation Campaign (SiSEC2010): Audio Source Separation

Shoko Araki[1], Alexey Ozerov[2], Vikrham Gowreesunker[3], Hiroshi Sawada[1], Fabian Theis[4], Guido Nolte[5], Dominik Lutter[4], and Ngoc Q.K. Duong[2]

[1] NTT Communication Science Labs., NTT Corporation, Japan
[2] INRIA, Centre Inria Rennes - Bretagne Atlantique, France
[3] DSPS R&D Center, Texas Instruments Inc., USA
[4] IBIS, Helmholtz Zentrum München, Germany
[5] Fraunhofer Institute FIRST IDA, Germany

Abstract. This paper introduces the audio part of the 2010 community-based Signal Separation Evaluation Campaign (SiSEC2010). Seven speech and music datasets were contributed, which include datasets recorded in noisy or dynamic environments, in addition to the SiSEC2008 datasets. The source separation problems were split into five 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

SiSEC2010 aims to be a large-scale regular campaign that builds 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) [2]), and the first community-based Signal Separation Evaluation Campaign (SiSEC2008) [3]. The unique aspect of this campaign is that SiSEC is not a competition but a scientific evaluation from which we can draw rigorous scientific conclusions.

This article introduces the audio part of SiSEC2010. In response to the feedback received at SiSEC2008, SiSEC2010 was designed to contain more realistic, and consequently, more challenging datasets, which had not previously been proposed for a large scale evaluation. Such datasets include recordings made in more reverberant rooms, under diffused noise conditions, or under dynamic conditions. We also repeated some of the typical tasks employed in SiSEC2008 (e.g, underdetermined and determined mixtures) with some fresh datasets.

Datasets and tasks are specified in Section 2 and the obtained outcomes 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 detail.

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

V. Vigneron et al. (Eds.): LVA/ICA 2010, LNCS 6365, pp. 114–122, 2010.

2 Specifications

This section describes the datasets, tasks 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. All materials are available at `http://sisec.wiki.irisa.fr/`.

2.1 Datasets

The data consisted of audio signals spanning a range of mixing conditions. The channels $x_i(t)$ $(1 \leq i \leq I)$ of each mixture signal were generally obtained as [3]: $x_i(t) = \sum_{j=1}^{J} s_{ij}^{\mathrm{img}}(t)$, where $s_{ij}^{\mathrm{img}}(t) = \sum_{j=1}^{J} \sum_{\tau} a_{ij}(\tau) s_j(t-\tau)$ is the *spatial image* of source j $(1 \leq j \leq J)$ on channel i, namely the contribution of source i to the mixture in channel j, $s_j(t)$ are source signals, and a_{ij} are mixing gains. Seven distinct datasets were provided for SiSEC2010:

D1 Under-determined speech and music mixtures
This dataset includes dataset D1 from SiSEC2008 [3], and a fresh dataset consisting of 30 stereo mixtures of three to four audio sources of 10 s duration, sampled at 16 kHz. The room reverberation time (RT) for the fresh dataset was 130 ms or 380 ms.

D2 Determined and over-determined speech and music mixtures
For this category, SiSEC2010 had the following four datasets.

D2-1 Determined and over-determined speech and music mixtures
This dataset contains two sets of 2×2 (#sources $\times$ #microphones), 3×3 and 4×4 mixtures. The RT of the recording room was about 500 ms, and microphones were arranged as a linear array with a spacing of approximately 10 cm. This dataset also includes dataset D2 from SiSEC2008 [3], which consists of 21 four-channel recordings of two to four speech or music sources acquired in four different rooms.

D2-2 Robust blind linear/non-linear separation of short two-source-two-microphone recordings
This dataset consists of 36 mixtures of different sources, located at three different positions in two environments. Each recording is 1 second long. The mixtures are six combinations consisting of a target speech signal and a jammer source (male/female speech, sneeze, laugh, glass break or TV sport noise). The data were recorded with directional microphones that were 8 cm apart, in an ordinary living room or study room.

D2-3 Overdetermined speech and music mixtures for human-robot interaction
These data include 27 recordings of three sources, which were recorded with five microphones attached to a dummy head. We consider three different microphone configurations, in three rooms: an anechoic laboratory room, a fully equipped office and a cafeteria. The source signals include male and female speech and music.

D2-4 Determined convolutive mixtures under dynamic conditions
This dataset consists of two kinds of scenarios. One comprises the short (1-2 seconds) mixtures of two sources obtained with a stereo microphone (**D2-4i**). The other is a sequence of audio mixtures obtained by the random combination of source locations and utterances (**D2-4ii**). Here, up to two sources are active at the same time. The components are generated by convolving random utterances with measured impulse responses, which were measured in a real room (RT≈700-800 ms). The microphone spacings were 2, 6, and 10 cm.

D3 Professionally produced music recordings
This dataset contains five stereo music signals sampled at 44.1 kHz, which include the same data as these in D4 used in SiSEC2008 [3], and new recordings for SiSEC2010. In addition to 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.

D4 Source separation in the presence of real-world background noise
These data consist of 80 multichannel speech mixtures in the presence of several kinds of real-world diffused noise. Noise signals were recorded in three different real-world noise environments (in a subway car, cafeterias, and squares), and each noise signal was recorded at two different microphone positions, the center or corner of the environment. There were one or three sources. Two types of microphone arrays, a stereo or a 4-element uniform linear array, were employed.

Datasets D1, D3 and D4 include both test and development 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.

2.2 Tasks

We specified the following five tasks:

T1 Source counting
T2 Mixing system estimation
T3 Source signal estimation
T4 Source spatial image estimation
T5 Source DOA estimation

These tasks consist of finding, respectively: (T1) the number of sources J, (T2) the mixing gains a_{ij} or the discrete Fourier transform $a_{ij}(\nu)$ of the mixing filters, (T3) the source signals $s_j(t)$, (T4) the spatial images $s_{ij}^{\mathrm{img}}(t)$ of the sources for all channels i, and (T5) the direction of arrival (DOA) of each source. Participants were asked to submit the results of their systems for T3 and/or T4, and optionally for T1 and/or T2 and/or T5.

Two oracle systems were also considered for benchmarking task T4: ideal binary masking over a short-time Fourier transform (STFT) [4] (O1) and over a cochleagram [5] (O2). These systems require true source spatial images and provide upper performance bounds for binary masking-based systems.

Table 1. Average performance for tasks T3 or T4 for instantaneous dataset D1. Figures relate to T4 when the ISR is reported and to T3 otherwise.

dataset	Test								Test2							
System	[9]		[10][2]		O1		O2		[9]		O1		O2			
SDR OPS	13.5	60.7	7.7	36.2	10.5	55.0	8.1	39.3	12.8	53.2	9.3	40.5	8.4	33.9		
ISR TPS	24.0	75.2		58.3	20.0	68.1	14.4	49.4	22.8	70.6	17.0	55.5	14.4	46.4		
SIR IPS	20.5	77.1	18.2	67.8	21.6	79.3	17.4	71.4	19.8	70.7	18.7	65.6	17.0	61.8		
SAR APS	14.8	73.2	8.4	36.9	11.5	60.3	9.1	45.5	14.1	64.8	10.2	49.6	9.6	35.7		

Table 2. Average performance for task T4 for convolutive dataset D1. Correlation between the classical (SDRs etc.) and the auditory-motivated measures (OPS etc.) was 0.6-0.68.

System	Test:RT=130ms				Test:RT=250ms				Test2:RT=130ms				Test2:RT=380ms			
	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
[9][2]	2.6	7.3	4.8	7.5	1.4	5.6	2.5	7.4	2.9	8.6	6.3	9.2	0.6	5.0	0.5	6.8
	32.2	55.2	50.4	59.7	25.7	46.2	41.0	52.5	34.8	58.4	54.2	60.6	15.8	38.3	27.2	41.1
[11]	5.3	10.2	9.9	7.1	3.9	8.5	7.3	6.9	3.6	8.4	6.5	7.1	2.1	5.9	3.3	6.0
	25.7	53.8	49.8	40.4	21.5	50.1	43.7	35.6	20.5	45.2	37.2	40.0	11.6	36.2	22.9	26.2
[12][2,3]	3.1	6.4	3.8	7.9	1.9	5.4	1.9	8.3	2.5	6.2	3.4	10.1	0.2	3.4	-0.8	8.2
	30.3	55.2	44.4	64.6	21.2	45.3	29.8	59.8	27.9	52.1	36.3	76.5	14.3	32.3	14.5	68.5
[13][3]	2.7	6.2	4.1	8.3	1.7	5.5	2.7	8.8	3.2	6.7	4.3	9.1	0.9	4.0	-0.6	7.4
	33.5	55.0	45.1	78.0	27.5	48.3	36.2	77.0	32.9	56.9	49.0	61.7	11.7	31.3	13.8	50.0
O1	9.7	18.3	19.9	10.2	8.7	16.2	19.4	10.4	10.2	18.1	19.6	11.0	9.2	16.8	18.5	9.9
	52.4	68.9	79.5	57.2	50.6	63.8	78.3	56.4	52.0	59.8	77.9	59.2	41.6	50.0	72.5	48.8
O2	6.9	12.1	16.5	7.6	6.6	11.5	16.0	7.9	7.3	13.6	16.0	7.9	6.2	11.8	14.4	6.5
	34.7	48.2	69.3	39.6	31.8	41.9	65.1	37.9	27.5	41.0	61.3	32.2	18.2	28.2	49.6	20.7

2.3 Evaluation Criteria

Task T2 was designed to be evaluated by using the mixing error ratio (MER), which was proposed in SiSEC2008 [3]. However, because we had no entrant for task T2, we skipped the T2 evaluation.

Tasks T3 and T4 were evaluated via the criteria in the BSS_EVAL [6,2], termed the signal to distortion ratio (SDR), source image to spatial distortion ratio (ISR), signal to interference ratio (SIR) and signal to artifacts ratio (SAR).

In addition, new auditory-motivated objective measures [7] were used to assess the quality of the estimated signals for T3 and T4 in the stereo cases. Four performance measures akin to SDR, ISR, SIR and SAR are given: overall perceptual score (OPS), target-related perceptual score (TPS), interference-related perceptual score (IPS) and artifact-related perceptual score (APS). Here, a new method for estimating the distortion components is employed based on a gammatone filterbank, and the salience of the target, interference and artifact distortions are calculated by using the PEMO-Q measure [8]. It has been confirmed

[2] The system is an extended version of the reference.

[3] Figure computed by averaging over an incomplete set of mixtures or sources.

that these auditory-motivated measures improve the correlation to subjective measures compared to classical SDR, ISR, SIR and SAR [7]. The new measures are expressed in terms of a figure between 0 and 100 (not in dB).

Task T5 was evaluated from the absolute difference between the true and estimated DOAs.

3 Results

A total of 38 submissions received for the audio tasks. Their average performance values are given in Tables 1 to 8. The details and all the results are available at `http://sisec.wiki.irisa.fr/`. It should be noted that the presented values are the absolute values, not the improvements from the values for mixtures. Although a close analysis of each table is beyond the scope of this paper, what we observed was following. We can obtain good results for instantaneous/anechoic mixtures (e.g., Tables 1 and 5), however, the separation of reverberant mixtures still remains challenging, especially in underdetermined scenarios (e.g., Table 2). The realistic datasets (D2-2, D2-4 and D4) attracted a relatively large number of participants as shown in Tables 4, 6 and 8.

Some tasks were evaluated with the new auditory-motivated objective measures (e.g., Tables 1, 2, 7 and 8). Sometimes, the performance tendency was the opposite of that with BSS_EVAL. More detailed investigations are required, however, it seems that binary mask based methods tend to achieve a poorer grade than non-binary mask approaches.

Table 3. Average SIR for task T3 over dataset D2-1

System	Cushioned rooms			Office/lab rooms			Conference room			New office room		
	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$
[14]	8.0[3]	6.8[3]	4.5[3]							11.6	7.8	9.0
[15]	3.2	0.9	0.1	9.0	2.0	-2.0	4.3	-0.3	-3.1	5.2	2.6	1.5
[15][2]	4.2	4.0	4.4	6.4	4.4	-1.2	4.9	0.9	-2.2	9.3	6.0	4.7
[16]										13.7	10.4	10.0

Table 4. Average performance for dataset D2-2. SIR* and SDR* in [17] for linear systems were also evaluated according to the dataset providers' proposal.

System	speech+speech					speech+(sneeze or laugh)					speech+(glass or TV noise)				
	SIR*	SDR*	SDR	SIR	SAR	SIR*	SDR*	SDR	SIR	SAR	SIR*	SDR*	SDR	SIR	SAR
[18]	8.5	5.0	3.7	8.7	18.6	6.1	3.3	1.9	6.8	16.7	9.4	5.3	4.7	9.8	19.1
[19]	8.1	0.8	-0.9	8.4	23.2	6.3	0.0	-1.8	7.3	27.2	11.3	0.6	-1.2	11.3	18.4
[20]	7.9	6.4	5.9	9.4	9.7	7.4	5.0	3.5	8.6	7.8	8.1	5.6	5.6	9.5	9.3
[21]	11.5	8.2	7.7	12.2	18.8	10.6	6.2	4.0	11.3	17.0	9.9	6.4	6.1	10.7	15.5
[21][2]			3.1	16.3	10.1			2.7	14.0	10.3			2.6	12.8	8.1
[16]			2.0	5.5	10.5			4.5	9.4	12.3			0.4	2.5	8.0
[14]			0.6	6.2	5.0			0.5	7.0	3.7			-0.6	5.5	4.3

Table 5. Average SIR for dataset D2-3 for three microphone configurations C1-C3

	Anechoic			Cafeteria			Office		
System	C1	C2	C3	C1	C2	C3	C1	C2	C3
[16]	24.6	23.4	22.8	9.2	15.0	12.9	11.9	12.7	13.3
[14]	5.5	13.6	12.5	3.7	4.8	5.7	2.5	3.6	4.3

Table 6. Average performance for dataset D2-4. For D2-4ii, the evaluation was performed only considering the segments where two sources overlap.

dataset D2-4i								dataset D2-4ii			
System	SDR	SIR	SAR	System	SDR	SIR	SAR	System	SDR	SIR	SAR
[22]	3.2	9.9	5.2	[23]	3.2	8.1	6.1	[23]	2.5	8.6	5.0
[22][2]	4.2	11.7	5.8	[21]	5.6	11.3	7.6	[21]	4.6	10.7	6.8
[22][2]	3.7	10.6	5.5	[21][2]	6.2	13.8	7.4	[24]	4.5	9.9	7.2
								[25][2]	2.1	9.2	3.9

Table 7. Average performance for dataset D3

System	Development data								Test data							
	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS	SDR	ISR	SIR	SAR	OPS	TPS	IPS	APS
[26]	2.1	6.5	7.9	2.9	26.6	49.9	51.5	43.7	1.3	6.7	4.7	0.9	31.7	49.0	49.9	46.3
[27][2,3]	0.9	11.2	∞	-0.8	24.9	56.4	89.6	6.0	-1.1	7.5	∞	-2.1	21.3	39.0	89.6	3.7
[28]	3.3	8.0	7.2	5.3	17.3	45.0	27.4	48.1	2.2	6.2	6.2	3.9	18.7	40.9	30.4	45.4
O1	6.5	18.3	20.0	6.7	31.8	64.3	70.0	30.7	5.8	20.8	21.2	5.6	24.9	53.6	62.0	23.2
O2	3.5	8.4	17.0	3.3	29.2	47.5	63.8	30.2	2.9	8.9	14.9	1.7	20.9	34.2	53.0	21.8

Table 8. Average performance for source separation of dataset D4. Figures relate to T4 when the ISR is reported and to T3 otherwise. [29], [14], [30] and [31] also addressed task T5. Results can be found at `http://sisec.wiki.irisa.fr/`.

System	2-ch×1-src				2-ch×3-srcs				4-ch×1-src				4-ch×3-srcs			
	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
[13]	2.7	16.1	4.4	11.9	1.3	7.0	1.9	7.7	3.0	15.7	4.0	12.8	2.6	7.7	4.3	9.1
	36.3	69.9	45.2	88.3	18.6	42.3	23.7	63.4								
[29]	-9.3	-7.2	14.1	8.0					-8.1	-4.3	12.2	8.1				
	33.5	49.9	55.5	53.8												
[14]	3.2		∞	3.2	0.5		5.2	5.0	7.4		∞	7.4	4.8		13.7	6.3
[30][2]	-12.2	-7.7	6.2	8.8												
	19.3	35.2	34.8	40.6												
[31][2]	-8.6	-3.2	7.5	5.4												
	16.3	39.4	35.2	28.3												
[16]	2.5	16.0	4.0	12.8					7.5	17.1	10.4	14.3	3.5	13.8	5.3	11.2
	38.4	73.4	51.2	78.4												

4 Conclusion

This paper presented the specifications and results of SiSEC2010. We hope that SiSEC2010 will provide a common platform for the source separation research field. This time, we welcomed all proposed tasks and datasets and finally we had seven datasets. This increase in the number of active participants is the product of a series of evaluation campaigns. On the other hand, we think that seven datasets may have been too large to collect sufficient number of participants for each task, and to evaluate all the submissions in detail. We may need a framework for pre-selecting the task/dataset proposals. In addition, perhaps it is time to reorganize the specifications and datasets in a series of SiSECs for prospective evaluation campaigns and future source separation research. We invite all willing participants to join a continuous collaborative discussion on the future of source separation evaluation.

Acknowledgments. We thank all the participants, as well as B. Lösch, Z. Koldovsky, P. Tichavsky, M. Durkovic, M. Kleinsteuber, M. Rothbucher, H. Shen, F. Nesta, O. Le Blouch, N. Ito, L.C. Parra, M. Dyrholm, K.E. Hild II, M. Vinyes Raso and J. Woodruff, for sharing their datasets and evaluation codes. Our special thanks go to V. Emiya, who helped us to evaluate all the stereo submissions with the new auditory-motivated objective measures.

References

1. Cooke, M.P., Hershey, J., Rennie, S.: Monaural speech separation and recognition challenge. Computer Speech and Language 24, 1–15 (2010)
2. Vincent, E., Sawada, H., Bofill, P., Makino, S., Rosca, J.: First stereo audio source separation evaluation campaign: Data, algorithms and results. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) ICA 2007. LNCS, vol. 4666, pp. 552–559. Springer, Heidelberg (2007)
3. Vincent, E., Araki, S., Bofill, P.: The signal separation evaluation campaign: A community-based approach to large-scale evaluation. In: Adali, T., Jutten, C., Romano, J.M.T., Barros, A.K. (eds.) ICA 2009. LNCS, vol. 5441, pp. 734–741. Springer, Heidelberg (2009)
4. Vincent, E., Gribonval, R., Plumbley, M.D.: Oracle estimators for the benchmarking of source separation algorithms. Signal Processing 87(8), 1933–1950 (2007)
5. Wang, D.L.: On ideal binary mask as the computational goal of auditory scene analysis. In: Speech Separation by Humans and Machines. Springer, Heidelberg (2005)
6. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. IEEE Trans. on Audio, Speech and Language Processing 14(4), 1462–1469 (2006)
7. 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 (submitted)
8. Huber, R., Kollmeier, B.: PEMO-Q – a new method for objective audio quality assessment using a model of auditory perception. IEEE Trans. on Audio, Speech, and Language Processing 14(6), 1902–1911 (2006)

9. Ozerov, A., Vincent, E., Bimbot, F.: A general modular framework for audio source separation. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 33–40. Springer, Heidelberg (2010)
10. Rickard, S.: The DUET blind source separation algorithm. In: Blind Speech Separation, Springer, Heidelberg (2007)
11. Sawada, H., Araki, S., Makino, S.: Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. IEEE Trans. on Audio, Speech and Language Processing (in press)
12. Arberet, S., Ozerov, A., Duong, N.Q.K., Vincent, E., Gribonval, R., Bimbot, F., Vandergheynst, P.: Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation. In: Proc. ISSPA (2010)
13. Duong, N.Q.K., Vincent, E., Gribonval, R.: Under-determined reverberant audio source separation using local observed covariance and auditory-motivated time-frequency representation. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 73–80. Springer, Heidelberg (2010)
14. Vu, D.H.T., Haeb-Umbach, R.: Blind speech separation employing directional statistics in an expectation maximization framework. In: Proc. ICASSP, pp. 241–244 (2010)
15. Ono, N., Miyabe, S.: Auxiliary-function-based independent component analysis for super-gaussian sources. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 165–172. Springer, Heidelberg (2010)
16. Sawada, H., Araki, S., Makino, S.: Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS. In: Proc. ISCAS, pp. 3247–3250 (2007)
17. Schobben, D., Torkkola, K., Smaragdis, P.: Evaluation of blind signal separation methods. In: Proc. ICA, pp. 261–266 (1999)
18. Koldovsky, Z., Tichavsky, P., Malek, J.: Time-domain blind audio source separation method producing separating filters of generalized feedforward structure. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 17–24. Springer, Heidelberg (2010)
19. Koldovsky, Z., Tichavsky, P., Malek, J.: Subband blind audio source separation using a time-domain algorithm and tree-structured QMF filter bank. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 25–32. Springer, Heidelberg (2010)
20. Loesch, B., Yang, B.: Blind source separation based on time-frequency sparseness in the presence of spatial aliasing. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 1–9. Springer, Heidelberg (2010)
21. Nesta, F., Svaizer, P., Omologo, M.: Convolutive BSS of short mixtures by ICA recursively regularized across frequencies. IEEE Trans. on Audio, Speech, and Language Processing 14 (2006)
22. Yoshioka, T., Nakatani, T., Miyoshi, M., Okuno, H.G.: Blind separation and dereverberation of speech mixtures by joint optimization. IEEE Trans. on Audio, Speech, and Language Processing (2010) (in press)
23. Málek, J., Koldovský, Z., Tichavský, P.: Adaptive time-domain blind separation of speech signals. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 9–16. Springer, Heidelberg (2010)
24. Loesch, B., Yang, B.: Adaptive segmentation and separation of determined convolutive mixtures under dynamic conditions. In: Vigneron, V. (ed.) LVA/ICA 2010. LNCS, vol. 6365, pp. 41–48. Springer, Heidelberg (2010)
25. Araki, S., Sawada, H., Makino, S.: Blind speech separation in a meeting situation. In: Proc. ICASSP, pp. 41–45 (2007)

26. Ozerov, A., Fevotte, C.: Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE Trans. on Audio, Speech and Language Processing 18(3), 550–563 (2010)
27. Bonada, J., Loscos, A., Vinyes Raso, M.: Demixing commercial music productions via human-assisted time-frequency masking. In: Proc. AES (2006)
28. Spiertz, M., Gnann, V.: Unsupervised note clustering for multichannel blind source separation. In: Proc. LVA/ICA (2010) (submitted)
29. Even, J., Saruwatari, H., Shikano, K., Takatani, T.: Speech enhancement in presence of diffuse background noise: Why using blind signal extraction? In: Proc. ICASSP, pp. 4770–4773 (2010)
30. Okamoto, R., Takahashi, Y., Saruwatari, H., Shikano, K.: MMSE STSA estimator with nonstationary noise estimation based on ICA for high-quality speech enhancement. In: Proc. ICASSP, pp. 4778–4781 (2010)
31. Takahashi, Y., Takatani, T., Osako, K., Saruwatari, H., Shikano, K.: Blind spatial subtraction array for speech enhancement in noisy environment. IEEE Trans. on Audio, Speech and Language Processing 17(4), 650–664 (2009)