Audio/Video Supervised Independent Vector Analysis Through Multimodal Pilot Dependent Components

Francesco Nesta*, Saeed Mosayyebpour*, Zbyněk Koldovský†, Karel Paleček†
*Conexant Systems, 1901 Main Street, Irvine, CA (USA). E-mail: {francesco.nesta, saeed.mosayyebpour}@conexant.com
†Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic. E-mail: {zbynek.koldovsky, karel.palecek}@tul.cz

***Abstract*—Independent Vector Analysis is a powerful tool for estimating the broadband acoustic transfer function between multiple sources and the microphones in the frequency domain. In this work, we consider an extended IVA model which adopts the concept of pilot dependent signals. Without imposing any constraint on the de-mixing system, pilot signals depending on the target source are injected into the model enforcing the permutation of outputs to be consistent over time. A neural network trained on acoustic data and a lip motion detection are jointly used to produce a multimodal pilot signal dependent on the target source. It is shown through experimental results that this structure allows the enhancement of a predefined target source in very difficult and ambiguous scenarios.**

***Index Terms*—independent vector analysis; source separation; independent component analysis; speech enhancement; multimodal processing;**

I. Introduction

Independent Vector Analysis (IVA) is a popular tool for unsupervised multichannel source separation [1]. Its virtues are related to the ability to avoid the permutation problem of traditional narrow-band frequency-domain methods for source separation [2], [3]. Differently from Independent Component Analysis (ICA), IVA uses a multivariate source model in order to jointly estimate the separated components in each frequency. The multivariate model allows to bypass the need for additional permutation solver algorithms, which often rely on prior assumptions on the geometrical interpretation of the mixing system [4], [5].

On-line implementations [6] and several other extensions have been proposed in [7]. Nevertheless, despite its potential, IVA is still not widely used in commercial applications such as VoIP and ASR preprocessing. Indeed, the effectiveness of IVA is intrinsically limited by the core paradigm of unsupervised source separation: The nature of the sources of interest is not explicitly defined. Therefore, although the internal permutation problem is solved by the multivariate model, the external order of the recovered sources cannot be guaranteed. The same output can contain portions of different source signals at different time instants, especially when the mixing conditions are not static.

To overcome this issue, geometrical constraints have been employed by imposing that a given output signal is associated to a source having known angular position [8]. However, these constraints cannot work well if the source and the noise are located at similar angles or when the target source position cannot be uniquely defined. Furthermore, these constraints make IVA similar to adaptive beamforming [9] or to geometrically constrained ICA [10], thus limiting the potential of multivariate modeling [11].

To mitigate the mentioned IVA ambiguities but without imposing any geometrical constraint, in [12] we proposed to modify the multivariate model by injecting pilot signals that are mutually dependent with the sources of interest. The pilot signals were defined to be proportional to the posterior probabilities to observe each source, given an observed wide-band spatial or spectral feature. Inspired by [11], in this work we further extend the model by considering posteriors derived from multimodal signals:

- A neural network is trained using extensive prior acoustic data such that it produces a pilot signal that contains posteriors of the target source dominance in the observed mixture.
- Another pilot signal is derived based on the lip detection in a video recorded together with audio. This allows to disambiguate the separation in cases where the target as well as the interference are speech signals.

Experimental evaluations are carried out to confirm the effectiveness of the supervised structure in separating speech from noise sources, in difficult ambiguous scenarios, e.g. when the target and the noise are both speech sources and are located at a similar angular direction.

II. Supervised IVA

N source signals are assumed to be recorded by an array of M microphones. Let S_n^k and X_m^k be the STFT coefficients obtained for the kth frequency bin, the nth source and the mth mixture signal, respectively. Let $\mathbf{S}^k = [S_1^k \cdots S_N^k]^T$ and $\mathbf{X}^k = [X_1^k \cdots X_M^k]^T$. The mixing model is

$$\mathbf{X}^k = \mathbf{H}^k \mathbf{S}^k + \mathbf{N}^k, \tag{1}$$

where $\mathbf{N}^k = [N_1^k, \cdots, N_M^k]^T$ is the vector of background noise and interference signals, and $\mathbf{H}^k$ indicates the mixing matrix for the kth frequency bin. Assuming $N = M$, the objective of IVA is to estimate a set of de-mixing matrices

[1]The work of Zbyněk Koldovský was supported by The Czech Science Foundation through Project No. 17-00902S

ISBN 978-0-9928626-7-1 © EURASIP 2017

$\mathbf{W}^k = \{W^k_{nm}\}$, $k = 1, \dots, K$, where K is the number of the frequency bins. The de-mixing matrices jointly recover independent multidimensional sources $\mathbf{Y}_n = [Y^1_n, \cdots, Y^K_n]$, $n = 1, \dots, N$, where

$$Y^k_n = \sum_{m=1}^{M} W^k_{nm} X^k_m, \tag{2}$$

up to a scaling ambiguity, which can be subsequently resolved by applying the Minimal Distortion Principle [13] to each matrix $\mathbf{W}^k$.

A typical way to model the sources is with a multivariate spherical super-Gaussian distribution[1] defined as [1]

$$\mathbf{a}_n = [a^1_n, \dots, a^K_n]^T, \quad f(\mathbf{a}_n) = \alpha \exp\left(-\sqrt{\sum_{k=1}^{K} |a^k_n|^2}\right). \tag{3}$$

In the supervised IVA (S-IVA) [12], the multivariate model (3) is extended by injecting additional "pilot" dependent components. In this work we consider Q pilots for each source, $P^1_n, \dots, P^Q_n$, which will be related to different modalities:

$$\widetilde{\mathbf{a}}_n = [a^1_n, \dots, a^K_n, \gamma_1 P^1_n, \dots, \gamma_q P^Q_n],$$
$$f(\widetilde{\mathbf{a}}_n) = \alpha \mathrm{exp}\left(-\sqrt{\sum_{k=1}^{K} |a^k_n|^2 + \sum_{q=1}^{Q} \gamma_q^2 |P^q_n|^2}\right), \tag{4}$$

where γ_q is an hyper-parameter controlling the influence of each pilot. To obtain the update rule, the Maximum Likelihood (ML) [1] approach is used by considering the cost function

$$\mathcal{L} = \sum_{k=1}^{K} \log|\det \mathbf{W}^k| + \sum_{n=1}^{N} E[\log f(\widetilde{\mathbf{Y}}_n)], \tag{5}$$

where $\widetilde{\mathbf{Y}}_n = [Y^1_n, \dots, Y^K_n, \gamma_1 P^1_n, \dots, \gamma_Q P^Q_n]$ denotes the extended observation output vector. The expectation $E[\cdot]$ is approximated with the time average over the frames. Then, by taking the derivatives of (5) with respect to W^k_{nm} and applying the natural gradient modification to maximize (5), we obtain the update rule

$$\Delta W^k_{nm} = (I_{nm} - E[\phi^k(\mathbf{Y}_n)(Y^k_m)^*])W^k_{nm}, \tag{6}$$
$$\mathbf{W}^k_{new} = \mathbf{W}^k_{old} + \eta \Delta \mathbf{W}^k,$$

where η is the adaptation rate, I_{nm} indicates the nmth element of the identity matrix, and the nonlinearities $\phi^k(\cdot)$, $k = 1, \dots, K$ are the score functions related to the density (4), namely,

$$\phi^k(\widetilde{\mathbf{Y}}_n) = \frac{Y^k_n}{\sqrt{\sum_{j=1}^{K} |Y^j_n|^2 + \sum_{q=1}^{Q} \gamma_q^2 |P^q_n|^2}}. \tag{7}$$

As the pilot components do not depend on W^k_{nm}, the second sum in (7) remains constant during the optimization. This way, any IVA algorithm can be modified to its supervised version.

[1]This simplified multivariate model is obtained by assuming that all the source components are zero mean and uncorrelated [1].

III. Definition of pilot signals

The proposed method can be related to a previous work in [14] where a user-guided source activity was used to supervise the IVA adaptation. However, the formulation of S-IVA is far more general as it can naturally include many supervising modalities, through the definition of multiple pilot signals. In this work, the pilot signals are derived from audio and video information.

As we use the spherical Laplacian model in (4), the pilots are assumed complex-valued zero-mean signals, uncorrelated to the frequency bins but with a dependent time-varying variance. Therefore, only the variance of the pilots has to be defined. By indicating with a^l_n and b^l_n the posteriors of source activity derived from the audio and video modalities (with $a^l_n,\ b^l_n \in [0, 1]$), the pilot signal variance is defined as

$$\begin{aligned} c(l) &= E\left[\frac{1}{N}\sum_n \sum_{k=1}^{K} |X^k_n(l)|^2\right], \qquad (8)\\ |P^1_n(l)|^2 &= (a^l_n)^2 \times c(l), \quad |P^2_n(l)|^2 = (b^l_n)^2 \times c(l), \end{aligned}$$

where l is the time frame index, $E[\cdot]$ indicates the expectation, which is approximated as a smooth time-average and the term $c(l)$ rescales the pilot to a dynamic range proportional to the sum of the frequency components[2].

A. Derivation of a^l_n through an Acoustic Neural Network

A neural network (NN), trained to solve a regression problem, is used to predict the source activity posteriors a^l_n. Namely, the network is trained to estimate the power ratio between the true target speech and the noisy mixture. Any machine learning method for regression can be used, such as recurrent neural networks (see, e.g., [15]) but we found that a naive multilayer feed forward NN, often named deep NN (DNN), is sufficiently accurate to produce a useful prediction.

Let S^{kl}_{nd} be the klth time-frequency representation of the *d*th signal corresponding to the *n*th source, included in the training set D_n. Example mixtures for the training are obtained as

$$X^{kl}_d = \sum_n S^{kl}_{nd}. \tag{9}$$

The DCT and the logarithm is applied to $|X^{kl}_d|$ to define the transformed features $\widehat{X}^{\widehat{k}l}_d = \mathrm{DCT}[\ln(|X^{kl}_d|)]$, where $\widehat{k}$ is the index of the DCT coefficient. For the frame l, the input layer is defined as

$$\mathbf{v}^l_d = [\widehat{\mathbf{X}}^{l-1}_d, \widehat{\mathbf{X}}^l_d, \widehat{\mathbf{X}}^{l+1}_d], \tag{10}$$

where $\widehat{\mathbf{X}}^l_d = [\widehat{X}^{1l}_d, \cdots, \widehat{X}^{\widehat{K}l}_d]$, $\widehat{K} < K$ (i.e. only the first $\widehat{K}$ DCT coefficients are used). Two hidden layers of 256 neurons are used with the hyperbolic tangent as the activation function. The softmax function is used in the output layer, which has dimension N. Each output represents a dominance-related feature for the nth source. For the dth mixture at the

[2]Here we assume that a proper care is adopted to scale the demixing matrices in order to keep their norm within a limited range.

ISBN 978-0-9928626-7-1 © EURASIP 2017

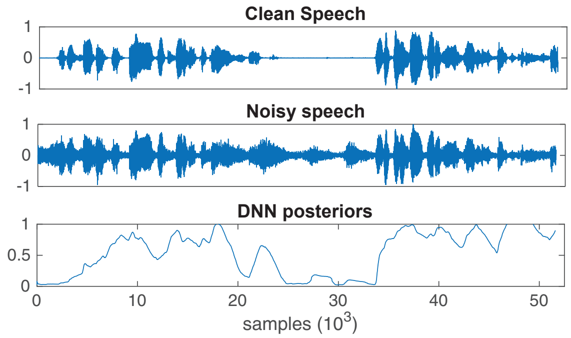

Fig. 1. Examples of DNN posteriors prediction

lth frame, the training output labels are defined as

$$\mathbf{g}_d^l = [g_{1d}^l, \cdots, g_{Nd}^l], \quad g_{nd}^l = \frac{\sum_k |S_{nd}^{kl}|^2}{\sum_{ki} |S_{id}^{kl}|^2}. \tag{11}$$

In this work we focus on the scenario where the source of interest is "speech" while any other non-speech acoustic event is considered as "noise" (i.e. the noise can be considered as composed by multiple sources). For the training of the DNN, a large set of 100k mixtures was generated by randomly combining noise examples with speech sentences in the TIMIT database. Noises were collected from different sources and the dataset was designed to balance the amount of noises belonging to different categories. Noise signals selected did not contain any speech, as the scope of the network is only to discriminate between speech and noise. Two datasets of 10k mixtures were generated for both cross-validation and testing.

After training, the output prediction for the nth source at the lth frame, indicated as a_n^l, is obtained through the feed-forward propagation of the input vectors $\mathbf{v}^l$ computed on the test recordings.

Figure 1 shows an example of the DNN output for a given test recording used in the experimental evaluation. We want to highlight that, although in this work our target is a speech source, S-IVA can also be applied to separate other type of acoustic sources, e.g. musical sources, as long as the DNN can discriminate them from their time-frequency representation.

B. Derivation of b_n^l from lip-motion detection

In scenarios where the target source and the noise are acoustically similar, additional modalities can be used to disambiguate the definition of the target source. Here, we use the video signal synced with the audio recording to extract the lip motion of a main target speaker.

In order to track the movement of the speakers' lips, a set of 68 facial landmarks is extracted from each frame of the video using the Ensemble of Regression Trees algorithm [16]. A subset of 8 landmarks describing the inner lip region then defines a polygon whose area r_i approximately corresponds to the mouth opening in the ith frame. Then, the mean m_i and variance v_i of 21 consecutive values $r_{i-10}, \ldots, r_{i+10}$ are computed and normalized to the range $\langle 0, 1 \rangle$. The posteriors b_n^l are produced by a logistic regression classification of the (m_i, v_i) feature pair for each video frame. To this end, all frames of the video sequence were manually labeled as either speech or non-speech and then split into train and validation sets in a $70:30$ ratio.

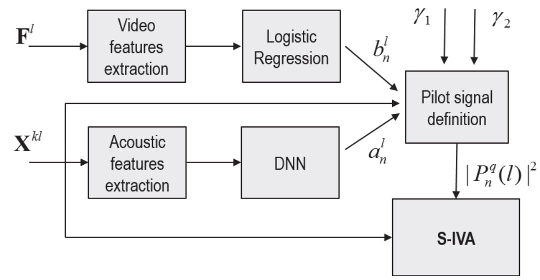

Fig. 2. Block diagram for the multimodal supervised S-IVA

Since the audio and video stream were captured at different rates, resampling was applied in order to produce a signal consistent with the time-frequency representation used by IVA.

IV. Experimental Evaluation

We conduct experiments with $M = 2$. An on-line S-IVA implementation is realized through updating the de-mixing matrices at each frame l according to

$$\begin{aligned} \mathbf{Y}^k(l) &= \mathbf{W}^k(l)\mathbf{X}^k(l), \\ \phi^k[\widetilde{\mathbf{Y}}_n(l)] &= \frac{Y_n^k(l)}{\sqrt{\sum_{j=1}^K |Y_n^j(l)|^2 + \sum_{q=1}^Q \gamma_q^2 |P_n^q(l)|^2}}, \quad (12) \\ \Delta W_{nm}^k(l) &= (I_{nm} - \phi^k[\widetilde{\mathbf{Y}}_n(l)] Y_m^k(l)^*) W_{nm}^k(l), \\ \mathbf{W}^k(l+1) &= \mathbf{W}^k(l) + \eta \Delta \mathbf{W}^k(l). \quad (13) \end{aligned}$$

The scaling normalization is applied to each bin to stabilize the convergence as in [17]. The signal mixtures are transformed in their corresponding time-frequency representation through Short-time Fourier Transform with the Hanning window of 4096 points with 75% overlap. After separations, the images of the target source at each microphone are recovered through MDP, and signals are transformed back to the time-domain using overlap-add. Two different experimental evaluations were carried out:

- Test1: separation with pilots based on acoustic features only;
- Test2: separation based on audio/video combined acoustic features.

The block diagram of the supervised IVA is depicted in Fig. 2.

A. Test1: Separation of speech from noise

In this experiment, the video information was not available. Thus, $|P_n^2(l)|^2$ in (13) was set to 0, for each n and l.

Recordings were made with two microphones with mutual distance of 0.2 m. Signals were recorded at $f_s = 16$ kHz in a room of size $5 \times 5 \times 2.5$ m with $T_{60} = 300$ ms. Partially diffuse noise was simulated according to the 3Quest standard

ISBN 978-0-9928626-7-1

by playing multichannel signals through 4 loudspeakers consisting of different types of real-world noise such as cafeteria, road noise, train station, etc. The target speaker was recorded at the distance of 2 m from the center of the microphones at different angles. Note, this can be considered an underdetermined scenario, as the noise was generated by playing partially uncorrelated signals through multiple loudspeakers.

In order to validate the robustness of the proposed approach, a dataset of 100 mixtures was generated by combining speech signals (speakers at random angles) with randomly selected noise examples (not included in the training set of the DNN model). Performance were evaluated by computing both the Noise-to-Speech ratio improvement (NSRi) at the noise output, and the Signal-to-Distortion ratio improvement (SDRi) at the speech output. Indeed, it should be noted that the scenario is highly underdetermined and a complete good speech extraction system should make use of both speech and noise estimates [18][19]. Fig. 3 shows the performance averaged over the test recordings, comparing standard IVA with S-IVA with γ_1 tuned for the best SDRi ($\gamma_1 = 24$). It is seen that S-IVA consistently improves the average performance compared to standard IVA, i.e. when $\gamma_1 = 0$, as the source order for IVA is not guaranteed to be consistent over all test samples.

In a second experiment we evaluate the robustness of S-IVA to an inaccuracy in the VAD. To simulate errors in the DNN prediction, artificial noise was added as

$$\widetilde{a}_n^l = (1-\beta)a_n^l + \beta \times \mathtt{rand}(1) \tag{14}$$

Fig. 4 shows the performance with varying β, demonstrating the robustness of S-IVA to noisy pilot signals.

B. Test2: Separation of target speech from noise speech

In this experiment, we consider S-IVA endowed by a multimodal audio/video pilot signal. A target speaker was recorded live in a front of a commercial laptop while simulating a

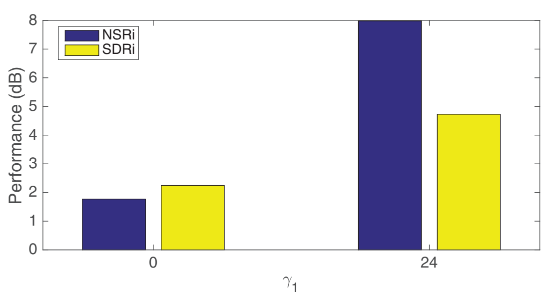

Fig. 3. Performance of IVA (i.e. $\gamma_1 = 0$) and S-IVA (tuned with $\gamma_1 = 24$)

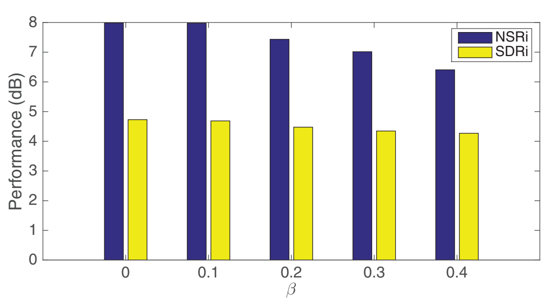

Fig. 4. Robustness of S-IVA versus noise in the pilot signal.

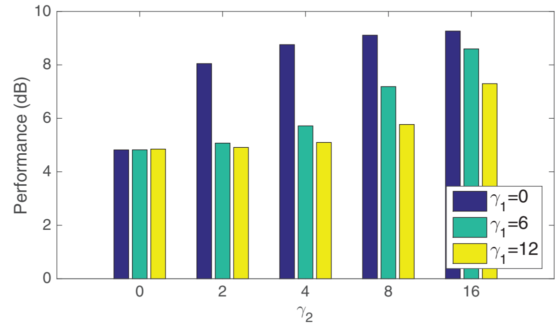

Fig. 5. SNRi performance for the multimodal S-IVA when the noise is speech.

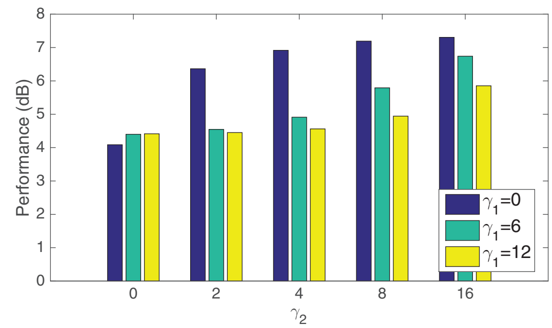

Fig. 6. SDRi performance for the multimodal S-IVA when the noise is speech.

VoIP conversation. Noise was generated by recording a TV, located in the back of the laptop, at a distance of about 2 meters. Although the speaker position is known in advance, it is worth noting that applying spatial constraints as in [8] would not be effective in these conditions. In fact, the angular positions of the target and of the noise sources are very close to each other. For a more detailed analysis of this aspect, see the experimental evaluation in [12].

A multimodal audio/video pilot signal is generated as in (8) and the performance were evaluated by varying both the parameter γ_1 and γ_2.

In a first experiment, we consider a recording where the TV noise contains only spoken news. This scenario is very difficult for the acoustic DNN as it cannot discriminate between the target and noise speech. Figures 5 and 6 show the SNRi and SDRi performance averaged over the target and noise source. It is straightforward that the pilot based on the acoustic DNN prediction is not reliable as by increasing γ_1 the performance degrades. On the other hand, the video information undoubtely provides a robust supervision as both SDRi and SNRi increase with γ_2.

In a second experiment, we consider a recording where the TV noise contained a mix of speech and music. From Figures 7 and 8 it can be seen that the acoustic DNN prediction is more effective in these noise conditions as the presence of non-speech related events, helps S-IVA to converge to the correct source order. Interestingly, this experiment shows that the best performance is obtained when combining both audio and video information together. Indeed, while the lip-detection accuracy

ISBN 978-0-9928626-7-1

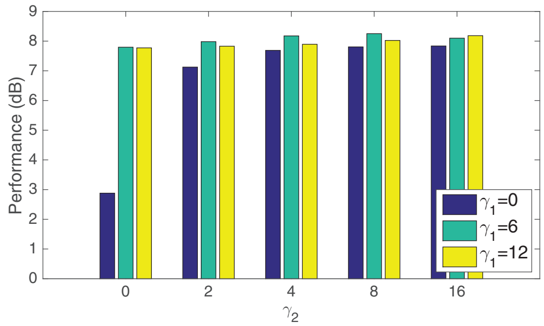

Fig. 7. SNRi performance for the multimodal S-IVA when the noise is a mix of speech and music.

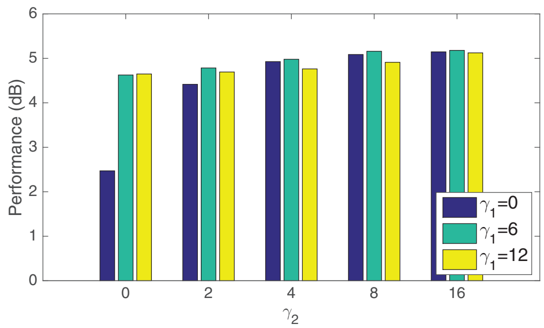

Fig. 8. SDRi performance for the multimodal S-IVA when the noise is a mix of speech and music.

should not be sensitive to the presence of acoustic noise, other disturbances could make it less reliable. For example, false detections are produced by movements of the lips that happens also when no speech is produced. This might also suggest that more effective multimodal formulations can be defined in alternative to (8), in order to better reflect the statistical correlation of the errors produced by each modality.

V. Conclusions

In this work we have presented a supervised extension of Independent Vector Analysis. A pilot signal is injected in the multivariate model to steer the estimation toward the extraction of a specific wanted source. A multimodal pilot signal was defined combining both audio and video information. A deep neural network was used to produce time-varying posteriors of source dominance in order to discriminate speech from acoustic noise events. A lip motion detection was used to distinguish between the activity of the desired speaker from that of interfering speech. It is shown that, without explicit constraints to the demixing system, it is possible to have a consistent enhancement of a specific target source in difficult scenarios, such as in far-field, in underdetermined conditions and when sources propagate from a similar direction.

It was shown that when S-IVA is supervised by the DNN-based pilot signal, good performance can be obtained if the noise does not contain any speech. On the other hand, when the noise is a speech source, the performance obtained with a video-based pilot signal clearly outperforms the acoustic supervision. Nevertheless, it was also observed that in mixed noise conditions the best performance was obtained by combining audio and video modalities together. This result suggests that further work is required to design multimodal formulations more effective than a naive weighted combination of each single modality. Furthermore, future work might also explore the use of EEG-based pilot signals, to realize effective biofeedback source enhancement methods [20].

References

[1] I. Lee, T. Kim, and T.-W. Lee, "Independent vector analysis for convolutive blind speech separation," in *Blind Speech Separation*. Springer, Sep. 2007.

[2] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, pp. 530–538, Sep. 2004.

[3] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, and T. Morita, "Real-time implementation of two-stage blind source separation combining simo-ica and binary masking," in *IWAENC*, 2005, pp. 229–232.

[4] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, "Blind source separation combining independent component analysis and beamforming," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 1135–1146, 2003.

[5] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Proceedings of ISSPA*, vol. 2, Jul. 2003, pp. 411–414.

[6] T. Kim, "Real-time independent vector analysis for convolutive blind source separation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1431–1438, 2010.

[7] J. Harris, B. Rivet, S. M. Naqvi, J. A. Chambers, and C. Jutten, "Real-time independent vector analysis with student's t source prior for convolutive speech mixtures," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 1856–1860.

[8] A. H. Khan, M. Taseska, and E. A. P. Habets, *A Geometrically Constrained Independent Vector Analysis Algorithm for Online Source Extraction*. Cham: Springer International Publishing, 2015, pp. 396–403.

[9] M. Brandstein and D. Ward, *Microphone Arrays*. Springer Verlag, 2001.

[10] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Transaction on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[11] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[12] F. Nesta and Z. Koldovský, "Supervised independent vector analysis through pilot dependent components," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017.

[13] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proceedings of International Symposium on ICA and Blind Signal Separation*, San Diego, CA, USA, Dec. 2001.

[14] T. Ono, N. Ono, and S. Sagayama, "User-guided independent vector analysis with source activity tuning," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 2417–2420.

[15] F. Weninger, J. R. Hershey, J. L. Roux, and B. Schuller, "Discriminatively trained recurrent neural networks for single-channel speech separation," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, Dec 2014, pp. 577–581.

[16] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, 2014, pp. 1867–1874.

[17] S. Douglas and M. Gupta, "Scaled natural gradient algorithms for instantaneous and convolutive blind source separation," in *Proceedings of ICASSP*, vol. II, Apr. 2007, pp. 637–640.

[18] F. Nesta and M. Matassoni, "Blind source extraction for robust speech recognition in multisource noisy environments," *Comput. Speech Lang.*, vol. 27, no. 3, pp. 703–725, May 2013.

[19] F. Nesta, T. Thormundsson, and Z. Koldovský, "On-line multichannel estimation of source spectral dominance," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2015, pp. 404–412.

[20] N. Das, S. Van Eyndhoven, T. Francart, and A. Bertrand, "Adaptive attention-driven speech enhancement for eeg-informed hearing prostheses," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 77–80.

ISBN 978-0-9928626-7-1 © EURASIP 2017