A SURVEY ON SEPARATION METHODS FOR QUALITY ENHANCEMENT OF AFFECTED SIGNALS

Vladimír Sedlák[1], Daniela Ďuračková[1], Roman Záluský[1], Tomáš Kováčik[1], Marcel Černák[1], Frank Schwierz[2]

[1]*Institute of Electronics and Photonics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava*

[2] *Technical University Ilmenau, Ilmenau, Germany*

E-mail: vladimir.sedlak@stuba.sk

Received 30 April 2014; accepted 12 May 2014

1. Problem definition

There are many signal, speech and audio applications where desired signal is corrupted by highly correlated noise sources. Separation such signals from their mixture has often been considered as one of the most challenging research topic in the area of signal enhancement. Acquiring of a signal in sufficient quality has become more and more important because the count of related application is increasing and it covers much more spectrum of tasks. This increase caused that this issue has become more interesting and that the count of techniques suitable for separation of signals is increasing as well. Till now have been developed different approaches and each of them has own advantages, disadvantages, computational demands and lots of other conditions which have to be met. The aim of this paper is investigation of these methods and to show how can be used.

The basic classification of separation methods is based on a count of sources and on a count of sensors. In area of speech processing a microphone represents sensor and a speaker represents source. A problem is called over-determined when a count of sensors is higher than a count of sources. In case when only one source is recorded using multiple sensors this process is denoted as Single Input Multiple Output (SIMO) when there are multiple sources it is Multiple Input Multiple Output (MIMO). On the opposite side, a problem is called under-determined when a count of sources is higher than a count of sensors. A typical example occurs in a meeting room where larger number of speakers shares only a few microphones. A special case in an under-determined are is situation when only one signal is recorded by one sensor, this state is called a single channel signal processing or a single channel signal separation (SCSS).

Several different approaches to SCSS have been proposed in the literature, most of which can be seen as filtering, decomposition and grouping or source modelling approach. In the filtering approach a set of functions are found that transform the mixture to estimates of the sources. In second approach the signal mixture is first decomposed into components which are in the next stage grouped together to form source estimates. In the source modelling approach a statistical model is formulated for each of the sources as well as for the mixing process.

High performance system for separation can play important role in offering robustness and reliability in many practical applications including speech and speaker recognition, hearing aids and speech coding. For example a performance of the speech recognition system may significantly degrade in an adverse noise condition since commonly it is trained from clean signals. Currently the research groups working on speech-separation problem especially focus on topic of how to separate signals from interfering sounds, including other speech [1].

2. Methods overview

Instead of classification which was presented in the previous section the separation methods are often divided based on approaches which they use. There are three main groups: blind methods (BSS), model based methods (MBSS) and source driven or computational auditory scene analysis based method (CASA). This classification is graphically depicted in Fig. 1.

Blind separation methods are focused on separation of a set of source signals from a set of mixed signals without the aid of information about the source signals or the mixing process. There are two main groups of blind methods. One is based on seeking source signals that are minimally correlated or maximally independent in a probabilistic or information-theoretic sense. Second approach is based on imposing structural constraints on the source signals. These constraints try to be some kind of low-complexity constraint, such as sparsity in some basis for the signal space. Well known methods such as independent component analysis (ICA) [2] and independent subspace analysis (ISA) [3] belong to the first group. We presented in our paper [2] the performance of ICA on various types of input signals. The investigation showed that this method can be very effective in case when multiple sensors are available. Second group can be represented by nonnegative matrix factorization (NMF)[4].

Model based methods can be divided into a few stages. First stage represents features extraction and models creation. These models are chosen to capture properties of sources and mixing process to effectively allow the sources to be separated. Common features already used for models are: time waveform, long-spectrum or discrete cosine transform. Mixing process models represent a key part of MBSS and they task is to model the probability of observing the mixture when the sources are given. Reconstruction of separated signals can be done in two ways, by employing an add procedure and by producing mask. Model based separation is often based on statistical models including vector quantization [5], Gaussian mixture models [6] and Hidden Markov models [7].

The CASA-based methods search auditory scenes in the time-frequency domain which are probably to come from the same sources of speech signals by exploiting the characteristics of human auditory system [8]. The CASA-based methods rely on extraction psychoacoustics cues from the given mixed signals and work in two stages: segmentation and grouping. During segmentation stage the input signal is decomposed into time-frequency cells which are specific for one type of input source and in the next step are these cells grouped to find the specified regions where only one speaker is dominant. Whole concept of segmentation and grouping is based on characteristics cues specific for given CASA-based method. The main ones are harmonicity, common onset and offset, amplitude and frequency and temporal or spectral proximity. The most important cue used in CASA is the pitch information that is required to be estimated directly from the mixed signal [9,10].

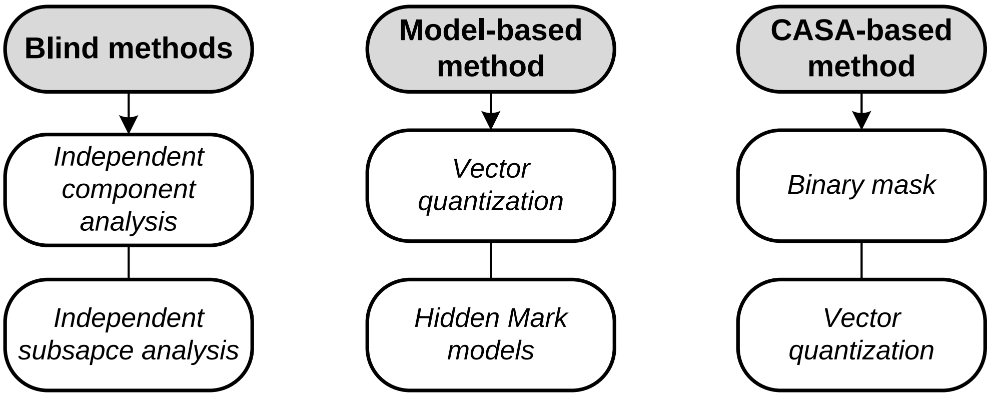

Fig.1:*Classification of the separation methods.*

3. Analysis

This part is devoted to investigation of performance of the selected methods. We have chosen one method from each category (blind methods, model-based methods and CASA-based methods): independent subspace analysis, separation based on vector quantization and separation based on binary mask.

Speech samples were taken from AURORA database which was primary collected to support the use of common material in speech perception and automatic speech recognition studies but it was also used in the different signal processing tasks. Input signal has been mixed together with multi-talker babble at various values of signal to signal ratio (SSR) measured in dB. A necessary condition for development of source separation algorithms is ability to measure the quality of result. In these experiments for this purpose has been used the PESQ. The binary mask has been computed using three different types of signals: target, masker and mixed signal. Input signal has been first divided into 20 ms frames with 50 % overlap and then the fast Fourier transform was applied to transform frames into frequency domain. In the next step the local SNR was compared against the local criteria LC (threshold)

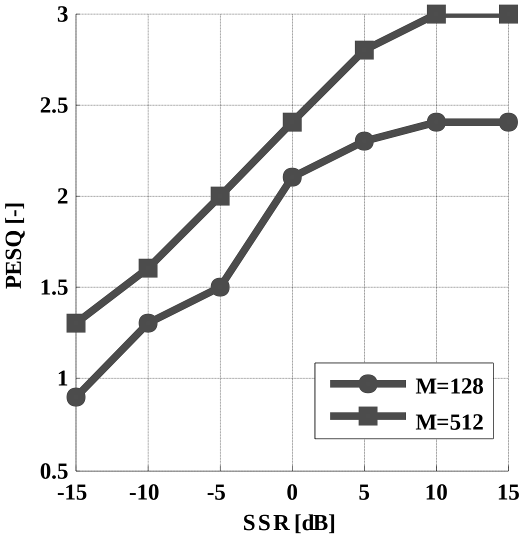

Fig. 2:*Performance of vector quantization for codebook size 128 and 512.*

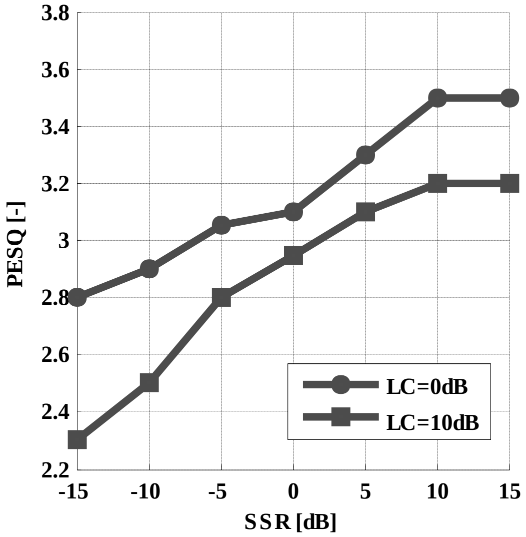

Fig. 3:*Performance of binary mask for local threshold 0dB and 10dB.*

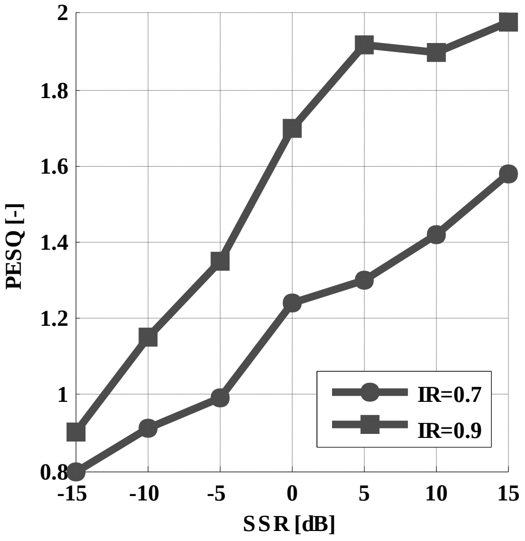

Fig. 4:*Performance of ISA for information ratio 0.7 and 0.9.*

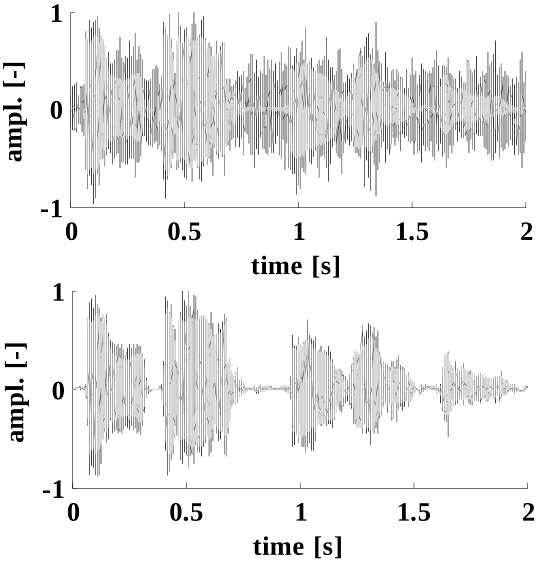

Fig. 5:*Waveforms of input and output signal processed by the binary mask.*

to determine whether to retain this unit or to eliminate it. Achieved results are depicted in the Fig.3 for two different values of threshold. Signals waveforms are depicted in Fig. 5 and represent input (mixed at 0 dB) and original signal, and output signal together with original.

In the vector quantization approach the mixture signal is represented as a sequence of vectors. The first step in VQ procedure is learning of codebook that consists of code-vector for each isolated source. Finally the separation of single or multiple sources can be achieved by finding the best matching code-vector consistent with an observed mixture. Creation of codebooks starts similar as in the previous procedure. Input signal is first divided into frames and then the fast Fourier transformation is applied. Features used for training codebook are chosen from this time-frequency representation of signal. In Fig. 2 are depicted achieved results for codebooks with 128 and 512 elements.

The last evaluated method was independent subspace analysis which was first time used for separation of audio signals in the paper [3]. The method includes two well known blind methods such as ICA and principal component analysis (PCA). The core of method is ICA but PCA is also very important because prepares input signal for ICA. That is the reason why can be typically multichannel method (ICA) used for single channel signal.Achieved results are depicted in Fig. 4 for two different values of information ratio. This parameter is set during principal component analysis and specifies how many features of input signal should be processed by ICA.

4. Conclusion

This paper presents a brief survey on separation methods. Model-based method showed quite good results but its disadvantage is necessity of creation of codebooks and sufficient amount of input samples. Blind method showed the worst results but in contrast to other, does not need any information about the original signals. It relies on an independence of signals what is big advantage. The binary mask showed the best result but it was computed directly from the original and mixed signals.

Acknowledgement

This work is resulting from the project VEGA 1/0987/12 sponsored by Ministry of Education, Slovak Republic.

References:

[1] K. Ananthakrishnan and K. Dogancy: *TENCON*, Singapore, 124 (2009)

[2] V. Sedlak, D. Durackova and R. Zalusky: *ELEKTRO*, Rajecke Teplice, 256 (2012)

[3] M. A. Casey and A. Westner: *International Computer Music Conference*, Berlin, 430 (2000)

[4] P. Smaragdis: *IEEE Transactions on Audio, Speech and Language Processing*, **15**, 1 (2007)

[5] V. Sedlak, D. Durackova, T. Kovacik and R. Zalusky: *Advances in Electronic and Photonic Technologies*, Stary Smokovec, 124 (2013)

[6] H. Wang, Y. Wang, W. Wang and S. Ma: *Conference on Image and Signal Processing*, Shanghai, 240 (2011)

[7] R. J. Weiss and D. P. Ellis: *Computer Speech and Language – Elsevier*, **1**, 16 (2010)

[8] Y. K. Lee and O. W. Kwon: *IEEE Transactions on Consumer Electronics***,** **4**, 2453 (2010)

[9] G. Hu and W. DeLaing: *IEEE Transactions on Audio, Speech and Language Processing*, **8**, 2067 (2010)

[10] W. Chu and A. Alwan: *IEEE Conference on Signal Processing, Communications and Computing*, Xi'an, 24 (2011)