AN EXPERIMENTAL SURVEY ON NON-NEGATIVE MATRIX FACTORIZATION FOR SEPARATION OF SIGNALS

Vladimír Sedlák[1], Daniela Ďuračková[1], Roman Záluský[1]and Tomáš Kováčik[1]

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

E-mail: vladimir.sedlak@stuba.sk

Received 30 April 2015; accepted 11 May 2015

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 verify a performance of the non-negative matrix factorization as an effective framework for separation of speech signals.

2. Methods analysis

Source separation methods can usually be classified as blind and non-blind methods based on characteristic of underlying mixtures. In the blind source separation, the completely unknown sources are separated without the use of any other information besides the mixture; the NMF belongs between these methods. In NMF, the non negative constraint leads to the parts based representation of the input mixture which helps to develop structural constraints on the source signals. NMF does not require the independence assumption, and is not restricted to data lengths, and also the bases functions are not ranked like using Independent Component Analysis.

Most NMF algorithms focus on minimizing the cost function such as Kullback-Leibler divergence or squared Frobenius norm or Itakura Saito Divergence etc multiplicative or additive updates which are very good described in [1-3].

3. Experiments

All experiments described below are focused on verification of a presented method for separation of signals which have been affected by different kind of noise, transmission conditions and overall quality of samples.For this purpose a mathematical model of the testing room has been created and all experiments have been affected by its properties.A room acoustics may significantly affect the overall intelligibility of the produced speech. The problem is degradation of the desired signal caused by the acoustics channel within and enclosed space. Because the microphone cannot always be located near the producer of desired signal, the received signal is commonly affected by reverberation introduced by multi-path propagation of this sound to microphone. The received signal generally consists of direct sound, reflections that arrive shortly after the direct sound (early reverberation), and

reflections that arrive after the early reverberation (commonly called late reverberation). Overall room acoustics can be described by room impulse response. This response can be used to generate reverberant (received) speech by convolving anechoic speech with the room impulse response (RIR). Response can be measured directly in testing room or can be simulated. For all experiments in this paper has been used RIR generator developed by Habets [4]. Required input parameters are: sampling frequency, room size, coordinates of the receiver, coordinates of the source and the reflections coefficients.

A model of testing room and RIR as function of time are depicted in Fig. 1 where the room dimensions are 5×4×3 meters, source coordinates 1×2×1 meters, receiver coordinates 4×2×1 meters and reverberation time 0.25 seconds.

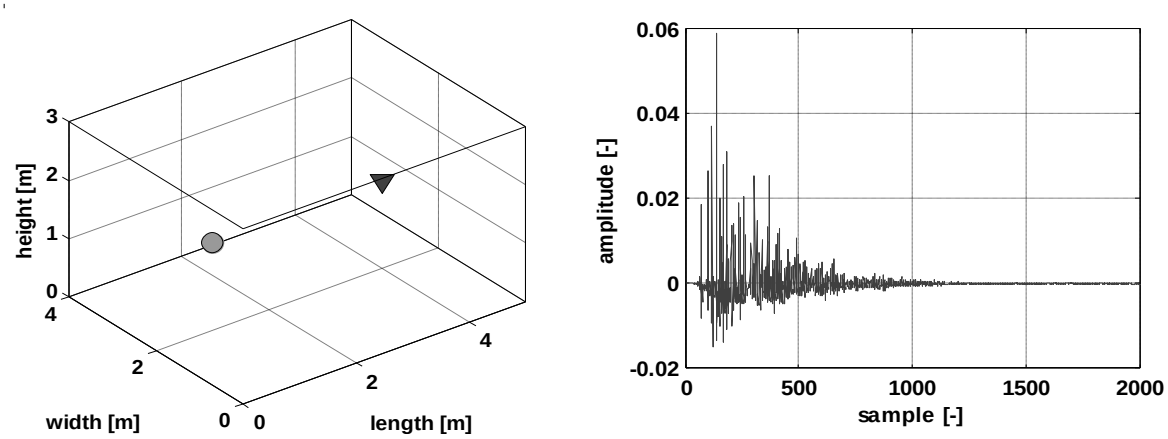

Fig.1: *Model of testing room and its impulse response.*

A necessary condition for development of an algorithm for source separation is ability to measure the perfomance or quality of result. In general, the separation quality can be measured by comparing separated signals with reference sources (objective methods) or by listenning to the separated signals (subjective methods). Subjective methods are considered to be more accurate than objective methods but they are more time consuming and required sufficient amount of listeners.

In these experiments the intelligibility have been measured by PESQ [5] and STOI [6]. The PESQ measure is recommended by ITU-T P.862 for speech quality assesment for 3.2 kHz handset telephony and narrow-band speech codec. The STOI measure shows good corellation with the subjective quality assesments metrics. As reference signal has been used clean signal without reflections.

Speech samples were taken from database presented in [7]. This database 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. Sentences are simple, syntactically identical phrases such as “place red at C one now”.

Presented method is based on the time-frequency representation of signals that is the reason why the signals were divided into 20 ms frames with 50 % overlap between frames in the first step. In the next step the fast Fourier transform (FFT) was applied to transform frames into frequency domain. An output from this transformation is complex signal and because NMF supports only non-negative values as an input, only absolute value of spectrum has been processed. Information about the phase of input signal has been stored for the latest step of procedure which is signal reconstruction because this approach is much simpler than direct reconstruction of phase. Finally NMF has been applied on this pre-processed signal and has returned matrix of bases and gains. The bases have been manually clustered in to ground and together with gains and stored phase have been used for reconstruction of signals. All

used masking are summarized in the Tab. 1, them value have been constant or have been varied dependently onanalyze.

First experiment has been focused on efficiency of proposed method for separation of signals which have been affected by different types of masking signals. SNR of input signal was varied from -20dB to 20dB with step 5dB and achieved results are presented in Fig. 2.

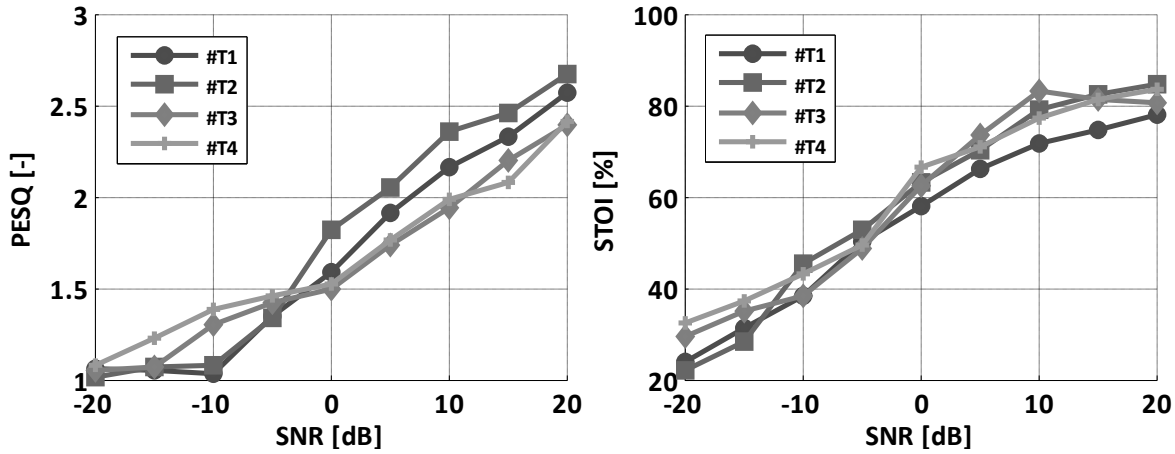

Fig.2*PESQ and STOI scores as a function of SNR of input signal.*

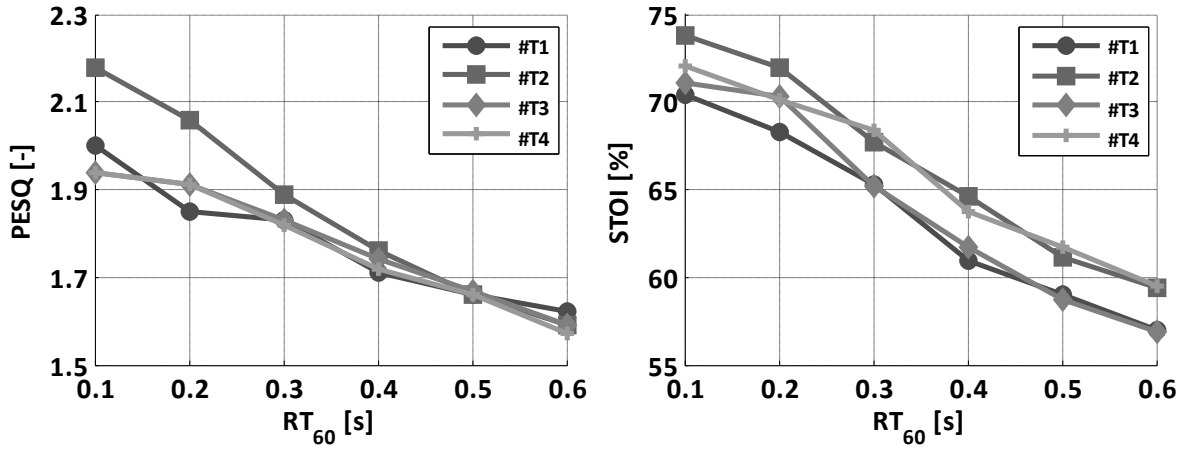

Fig. 3 *PESQ and STOI scores as a function of reverberation time.*

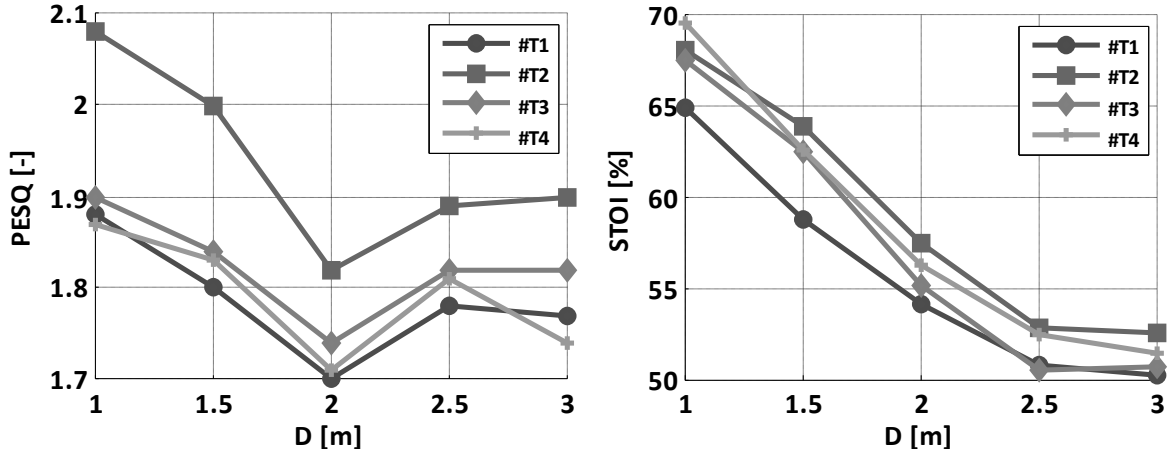

Fig. 4 *PESQ and STOI scores as a function of distance between source and receiver.*

Second experiment investigated an effect of reverberation time on an intelligibility of separated speech. At the beginning target signal was masked by masking signal at SNR 10 dB (this value was the same for all presented experiments) and then the RIR was applied to achieve the input signal corrupted by reflections. The reverberation time RT_{60} varied from 0.1 to 0.6 seconds and the distance between source and microphone was fixed to one and three meters. The results of this simulation are depicted in the Fig. 3. In the next analyze a distance was varied from 0.5 to 3 meters and RT_{60} was fixed to 0.2s. Results are depicted in Fig. 4.

Achieved results have shown expected values because it is obvious that if the input conditions are worse, the global efficiency of separation will be also worse. The goal of these experiments has not been only a verification of these expectations but also evaluate the effect of room acoustics with different types of masking signals. For this purpose the achieved results have been presented in form of PESQ and STOI and can be compared with different kind of analysis and methods.

Tab.1. *Masking signals for NMF*

Id	Type
#T1	Babble
#T2	Speech of another speaker
#T3	Noise at station
#T4	Noise in the car

4. Conclusion

The goal of this article has been to show how can be non-negative matrix factorization used for separation of speech signals. Have been presented how the global quality of input signal and transmissions conditions can affect efficiency of separation. The model of testing room and its acoustic features have been also included in to analysis to make experiments more realistic although the result have been obtained only from simulations. With same reason the test samples have been chosen to cover the widest possible range of realistic conditions. Quality of input signals have been affected by different kinds of masking signal, reverberation time and distance between source and receiver.

Acknowledgement

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

References:

[1] M. Nandakumar and E. Bijoy: *Internal Journal on Computer Applications*, **100**, 1 (2014).

[2] B. Wang, Q. Mary, M. Plumbley and Q. Mary: *Proceeding of UK Digital Music Network Conference,* London, 301 (2005)

[3] Y. Liu and J. Yao: *Journal of Computer Information Systems,* **10**, 5723 (2014)

[4] E. Habets: Room Impulse Response Generator, Available: http://home.tiscali.nl/ehabets/rir_generator.html, January 2015

[5] N. Shiran and I. Shallom: *Proceeding of Workshop on Quality of Multimedia Experience,* San Diego, 157 (2009).

[6] C. Taal, R. Hendriks, R. Heusdens and J. Jensen: *IEEE Transaction on Audio, Speech and Language Processing,* **19,** 2125 (2011).

[7] M. Cooke and J. Barker: *The Journal of the Acoustical Society of America*, **120**, 2421 (2010).