Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

6th International conference on Intelligent Human Computer Interaction, IHCI 2014

NUI for an Artificial Simulation of an Interactive Sound Source Enhancement to Restore Spatial Listening Experience

Susanne Rosenthal, Thomas Gross, Navya Amin, Marvin Offiah, Markus Borschbach

University of Applied Sciences, FHDW, Faculty of Computer Science, Chair of Optimized Systems, Hauptstr. 2, Bergisch Gladbach, Germany

Abstract

A self-determined, autonomous and socially integrated life is nearly impossible for hearing-impaired persons of all ages. Hearing damage affects 16% of the European and 11.5% of the US population, but only every fifth is wearing a hearing aid as they are not able to restore the biological listening experience and allow no interactive influence on it. Thus, a natural user interface is presented as an artificial simulation for an interactive enhancement of the user-preferred sound source. Therefore, established blind source separation algorithms are ported on a smartphone with Android operating system and their performance is evaluated regarding efficiency and robustness.

Peer-review under responsibility of the Scientific Committee of IHCI 2014.

Keywords: NUI; interactive sound enhancement; hearing-impaired; BSS; ICA; binary masking; instantaneous mixtures; smartphone

1. Introduction

The reduction of human hearing usually results in communication difficulties that are responsible for lack of self-confidence thus resulting in social isolation. Only every fifth of hearing-impaired persons that benefit from a hearing aid is using one, according to the National Institute of Deafness and Other Communication Disorders (NIDCD)[1]. The unpopularity of hearing aids is caused due to their inefficiency to mimic the biological hearing action: In general, digital hearing aids mimic the biological action by improving the speech clarity, enhancing particular signals and reducing insignificant noise. The performance of digital hearing aids is quite successful in standard everyday situations like at home or a calm communication between only a few speakers with voices of different loudness. Other situations like the sound effects in public places, restaurants, theaters or concerts or at social events is challenging because of the complexity of the sound sources and the impossibility of restoring the spatial hearing. Hearing aids only provide one particular algorithm setting for every sound situation. It has to decide in every complex sound situation the most important source signal for the affected user that has to be enhanced successfully and the unimportant sources which have to be reduced. The hearing-impaired person has no opportunity to influence the enhancement process of the hearing aid by an interactive selection of the preferred sound or speech source for enhancement.
The smartphone technology is already integrated into people's life and further provides the possibility to be connected with hearing aids for an improvement of the natural hearing action. Hearing aid manufacturers nowadays offer smart-

E-mail address: Markus.Borschbach@fhdw.de

1877-0509
Peer-review under responsibility of the Scientific Committee of IHCI 2014.

phone applications to configure particular hearing aid settings of their devices. The first application is presented by ReSound for iPhones. A first smartphone application for a particular hearing aid type is termed ReSound Control[2] and allows users to configure settings like volume, treble and bass. Further, ReSound publishes the application ReSound Smart[3] that provides the functionality to remember personal settings of different locations set by the user before. In the case where the user enters his everyday restaurant, pushing the button 'Restaurant' resets the saved optimal settings from his previous visit to the restaurant. The University of Essex offers iPhone applications to turn a smartphone into a more sophisticated hearing aid. A setup procedure of this application helps the user to identify his type of hearing loss and the user has the opportunity to select one setting out of six that works best for him. Furthermore, each of these six settings has four additional options to fine-tune the sub-settings. Compared to traditional hearing aids the applications BioAid and the more sophisticated version AUD1[4] provide 24 setting options to enhance the hearing action. The manufacturer Siemens presents an application termed miniTek-App[5] for their devices. This application gives the possibility to fine-tune the user-specific settings. Moreover, it allows the user to make an audio connection of the hearing aids to other audio sources like a television via Bluetooth, transmitter or cable line. A further description of these applications is published in[6].

In this paper, we describe a Natural User Interface (NUI) for a smartphone as an application for the artificial simulation of an interactive speech or sound source enhancement as an possibility to restore spatial hearing. Different state-of-the-art BSS algorithms have been ported on a smartphone with Android operating system to separate the identified sound sources of an acoustic scenario. The NUI is termed 'SmartNavi' and the user has the opportunity to concentrate on one or even more preferred sound or speech sources. This allows the users to orientate themselves more confidently in complex sound situations like social events or outdoor activities and preserves an autonomous lifestyle as well as social integration of hearing-impaired persons. Furthermore, the performance of the BSS algorithms are evaluated according to effectiveness and robustness with the aim of stating recommendations for the algorithm improvement.

2. Methodology

2.1. Background

The main application field of BSS is the separation of speech signals[7]. Blind Source Separation (BSS) is further an active research field due to its potential. However, source separation has some limitations especially in the case of real-world applications: Usually, the number of source signals have to be known in advance and the number of sources is not allowed to exceed the number of sensors. Theoretically, BSS recovers N unknown source signals from M recorded mixtures. These mixtures are linear superpositions of the source signals and are categorized into instantaneous and convolutive mixtures. The matrix-vector description of an instantaneous mixture is given by:

$$\mathbf{y}(n) = \mathbf{A}\mathbf{x}(n) + \mathbf{r}(n), \tag{1}$$

where $\mathbf{x}(n) = [x_1(n), x_2(n), ..., x_N(n)]^T$ is vector of mutually N independent source signals, $\mathbf{y}(n)=[y_1(n), y_2(n), ..., y_M(n)]^T$ are the M recorded source signals, $\mathbf{A} = [a_{i,j}]$ is the MxN mixing matrix, $\mathbf{r}(n)$ the additional noise vector and n the time index. A widely used method to retrieve the original signals is the Independent Component Analysis (ICA)[8,9]. ICA estimates the original source signals $\mathbf{x}(n)$ by finding an inverse matrix $\mathbf{W}$ of the mixing matrix $\mathbf{A}$ ($W = A^{-1}$) so that $\mathbf{x}(n) = \mathbf{W}\mathbf{y}(n)$. A convolutive mixing model with N sources $x_i(t)$ ($i = 1, ..., N$) and M microphones that yield the mixed signals $y_j(t)$ ($j = 1, ..., M$) is given by

$$y_j(t) = \sum_{i=1}^{N} \sum_{l=0}^{K} h_{ji}(t) x_i(t - l), \tag{2}$$

where $h_{ji}(t)$ denotes the impulse respond from source i to microphone j. These filters are of a finite length $K < \infty$. BSS algorithms have been especially evolved to separate convolutive mixtures, which play an important role in the field of hearing aids as reverberations are constant phenomena in daily situations ([10] among others). The separation in the case of convolutive mixtures is usually either performed in the time-domain or in the frequency-domain[11]: The frequency-domain methods transform the signals to frequency-domain by applying Fourier transformation. Since time-domain convolutive mixtures are considered as instantaneous mixtures at every frequency point in the frequency-domain, the algorithm for instantaneous mixtures is applied to separate signals at every frequency point.

Several methods have been evolved for instantaneous mixtures [12 13] and convolutive mixtures [14]. The main drawback of the most methods for instantaneous and convolutive mixtures is that the number of source signals is not allowed to exceed the number of microphones $M \geq N$ [14]. More precisely, if $N > M$, the matrix $\mathbf{A}$ is not invertible and the independent components cannot be recovered exactly [8].

2.2. *ICA methods for instantaneous and convolutive mixtures*

The goal of the BSS approaches is to estimate single signal sources from an acoustic overlapped listening experience by only using the mixed signals observed at each input channel. This estimation is performed blind which means that no information is given about the processing on each source, like the location or the active time. Many algorithms exist for BSS. Independent Component Analysis (ICA) is the major statistic method for work on the BSS problem. The goal of ICA is to separate mixed instantaneous signals from the input channel $y(n)$ into their independent sources $x(n)$ (eq. (1)). Although, the acoustic scenario that occurs most frequently in everyday life is a convolutive environment with reverberation and echo. A lot of research has been done so far to extend ICA/BSS techniques to successfully separate convolutive mixtures (eq. (2)). An overview of these techniques is given in [10]. These techniques are divided into time-domain, frequency-domain BSS techniques and a combination of both [10]. In the case of time-domain techniques, the ICA methods are applied directly on the convolutive mixtures, but the ICA methods are more sophisticated for these mixtures and therefore of a higher computational complexity. In the case of frequency-domains techniques, the complex-valued ICA methods for instantaneous mixtures are directly applied on each frequency bin. One of these methods with reduced computational complexity is the FastICA [15]. FastICA is a fast fixed-point algorithm for ICA and projection pursuit. The fixed-point scheme replaces the gradient descent technique of the traditional ICA. The FastICA algorithm considers the input data matrix to be a linear combination of non-Gaussian variables (independent components). FastICA measures the non-Gaussianity using approximations to the negentropy. First, a pre-processing step is done, by centering and whitening the input data. The next step is the component extraction. The algorithm iteratively finds a single independent component by maximizing the non-Gaussianity. To estimate more than one independent component, the algorithm must be repeated until all multiple components are extracted. The algorithm is described more detailed in [16] and [15]. The FastICA tool is available as free MATLAB program [17]. Another method is the information approximation approach (InfoMax) [18]. This is a self-organized learning system that maximizes the information transfer in a network of non-linear units. The network separates statistically independent components. It has been successfully applied on cocktail party problems up to ten speakers. It is further able to cancel echoes and reverberations in speech signals. InfoMax is also freely available as MATLAB tool [19]. The Joint Approximate Diagonalization of Eigenmatrices (JADE) is also an established method that replaces the estimation of the distribution of the independent sources of other ICA methods by an optimization of cumulative approximations of data. The advantage of this method is that it does not require a gradient descendant and consequently has no convergence problems. The disadvantage is that the computation storage of the cumulative matrices is high. JADE is also available as MATLAB tool [19]. In the case of frequency-domain techniques, the algorithms FastICA, JADE and InfoMax are directly applied as the calculation in (3). In [20], an over-complete BSS method for separating up to six mixed speech signals of anechoic environments is proposed. It combines the standard ICA algorithm with the binary time-frequency masking and is an established method in the case of under-determined separation problems $N \geq M(= 2)$. This method is further termed ICA-BM. The algorithm starts by a two-input-two-output ICA algorithm including a scaling. The re-scaled output signals after the ICA methods are transformed into the frequency domain via Short-Time Fourier Transformation (STFT). After that the binary masks are determined for each time-frequency unit comprising a threshold parameter. The two masks are applied then to the signals and are reconstructed in the time-domain by inverse STFT. If only one source signal is detected in the mask output, it is saved. In the case of more than one signal the process repeats with the ICA algorithm and masking. The algorithm is described in more detail in [20]. According to [20], this algorithm also separates sources under reverberant conditions and is available as a MATLAB tool.

For our benchmark test, the presented BSS tools - FastICA, ICA-BM, InfoMax and JADE - are implemented on the smartphone to separate instantaneous mixtures, which are further described in the chapter 'Experiments'.

3. System architecture

The actual digital hearing aids are complex and compact embedded systems including sophisticated hearing aid algorithms. The extension of such a complex and compact system is very restricted and usually allows no interactivity

with the user. The smartphone technology otherwise provides a wide field to deliver advanced features for such intelligent systems, typically independent of the hardware component.
This paper presents a cost-utility analysis of related work smartphone applications according to the used smartphone technology and potential with the main goal of extending hearing aids with advanced and interactive features as described above[21]. According to this examination, the selected BSS techniques are implemented on a smartphone with Android as the operating system. The smartphone application is evolved using the Java programming language and the Android Development Tools (ADT in the version 23.0.0)[22] as a plugin for the Eclipse IDE. GNU Octave in the version 3.8.1[23] is used to port and execute the BSS algorithms - available as MATLAB tools - on the smartphone. Octave is an open source tool and in the first place intended for numerical computations. As the Octave language is similar to the MATLAB language, this predestinates Octave to be implemented within the application and to execute the MATLAB scripts of the BSS algorithms. This architecture allows a simple future BSS algorithm optimization or adaption without complex system architecture modifications.

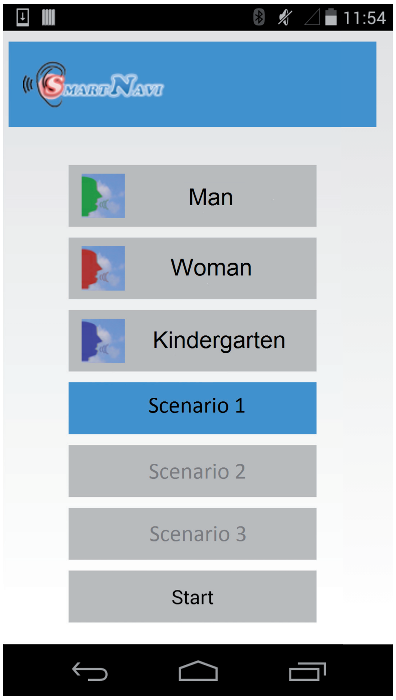

Fig. 1. The surface of the smartphone application.

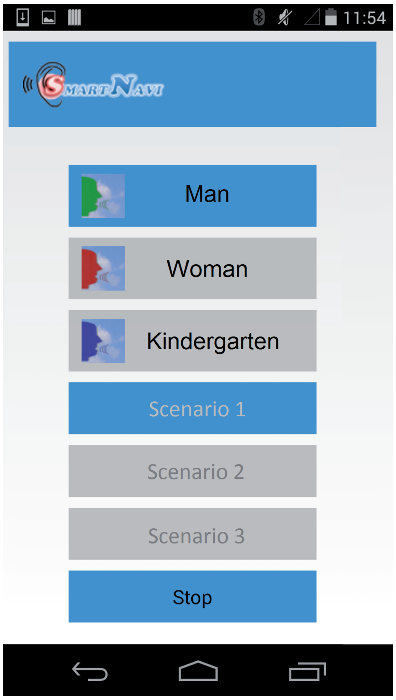

Fig. 2. The simulation of Scenario 1 is running with one speaker selected 'Man'.

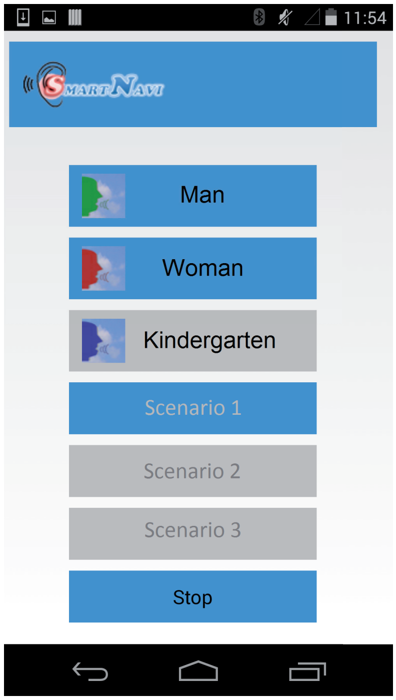

Fig. 3. The simulation of Scenario 1 is running with two signal sources selected: 'Man' and 'Woman'.

3.1. Interface Implementation

The functionalities of the application are arranged on one surface (Figure 1). This surface shows three (initially disabled) buttons in the upper part according to the number of identied sound sources of a scenario (labeled Source 1, Source 2 and Source 3). The scenario itself is first selected and automatically highlighted in further scenario buttons below the sound source buttons (Figure 1). This also enables and highlights a start button at the bottom of the screen. Additionally, it changes the labels of the sound source buttons to proper source names (e.g. Man, Woman, Glass etc.) , so the user is able to identify in advance what kind of content to expect during later playback. To activate the playback of the selected scenario, the user has to press the Start button. This also enables the sound source buttons without highlighting them, indicating that the unseparated mixture of all sound sources is being played. It also changes the label of the start button to Stop to indicate the change of its functionality. During playback, the user is able to select (and thereby also highlight) one (Figure 2) or more (Figure 3.) of the identied and separated sources according to his preference for enhancing (while all other non-selected sources are silenced). Likewise, the user is able to unselect the button again (thereby also removing the highlight) to remove that sound source from playback. At any given moment, only the selected (i.e. highlighted) sound sources are played, while all others remain silent, except in a state where all sound source buttons are deselected (which defaults to the whole mix being played). A state of all source buttons selected plays three separated sources simultaneously, while a state of no source buttons selected plays only the unseparated mix. All this is an important feature to act on the actual situation according to the preferred sound source or to ensure the hazard recognition while concentrating on the enhanced source. The Stop button ends the playback of the scenario and the app returns to the initial view.

4. Experiments

Three categories are stated for the test-scenarios to evaluate the performance and to evaluate the algorithms according to effectiveness and robustness. The first category is stated as the least challenging one: Sound situations with two or more speakers without noises and in non-reverberant environments. A good performance for this category allows hearing-impaired persons to concentrate on one speech source that prevents misunderstandings and finally a decrease of self-worth and social isolation. This first category is denoted by the term 'One Source-Type situation' (OST). The second category is a more sophisticated one: Situations with a combination of speech sources, music and/or other sound sources like radio or television. For hearing-impaired persons, these situations are difficult, as they require a permanent listening effort on one of these sources, if indeed possible. This category is characterized by the term 'Multiple Source-Type situation without Noise' (MSTwN). The third category is the most sophisticated one, but an everyday situation: Situations with the combination of speech sources, music and stationary or fluctuating noises characterized by the term 'Multiple Source-Type with Noises' (MSTN). A good performance of the system ensures an improved spatial hearing that is the most challenging issue until today[24].
FastICA, ICA-BM, InfoMax and JADE are used as BSS to separate instantaneous test-scenarios into their different sound sources. Fourteen test-scenarios as shown in Figure 4 have been produced mixing two or three sound sources like speech, music, children screaming, bar, sounds from the street and the siren of the police van to receive test-scenarios for each category OST, MSTwN and MSTN. The entries 'ID' denote the identification indexes of the used single source audio files. The single source audio files are taken from the Signal Separation Evaluation Campaign (SICEC) website[25] with a length of 10 seconds each and a sampling rate of 16000kHz. These single audio files are mixed together by multiplying the single source file with a random matrix. The first four samples belong to the first category OST, the samples five and six belong to the second category MSTwN and the remaining eight samples (Sample 7 to 14) belong to the third category MSTN. In order to keep the problem simple for the separation on the smartphone, we consider the case $N = M$ meaning the number of acoustic source signals N is equal to the number of microphones M. The input for the BSS algorithms are two or three mixed audio files, theoretically each file for one microphone. These two mixtures are different in the way that they are produced with different random matrices. Another simplification is that the audio files are instantaneous mixed sound sources. The FastICA, InfoMax and JADE

Sample	*Description*	Speech	Speech	Speech	Music	Traffic	Bar	Siren	Children
1	Speech 2 Sources	ID1	ID3						
2	Speech 2 Sources	ID1	ID2						
3	Speech 3 Sources	ID1	ID2	ID3					
4	Speech 3 Sources	ID3	ID4	ID1					
5	Speech & Music	ID1			ID5				
6	Speech & Music	ID1	ID4		ID6				
7	Speech & Children	ID1							ID7
8	Speech & Children	ID1	ID3						ID7
9	Speech & Bar	ID1					ID8		
10	Speech & Bar	ID1	ID3				ID8		
11	Speech & Street	ID1				ID9			
12	Speech & Street	ID1	ID3			ID9			
13	Speech & Siren	ID1						ID10	
14	Speech & Siren	ID1	ID3					ID10	

Fig. 4. List of samples used for the separation

are applied using default parameters. In case of ICA-BM, two parameters are changed based on the enhancement of the quality of the separated mixtures: Window function: The window function used is 'hamming window' which is

used for calculating the non-overlap region between the two signals during the binary mask calculation. This window function is used because a number of previous test runs showed that this parameter setting is faster and the separation quality better than the others. 'rectwin' is another possible choice for the window function, but - as mentioned in the MATLAB help - this choice is equal to a configuration with no window selection. Another parameter of the binary masking iteration of ICA-BM is the parameter termed 'Stopthresholdini', which is the threshold for considering the separated signal of the final separated source. This parameter is increased from the default values 3000 to 6000, because - once more - previous test runs showed for some mixtures comprising three signal sources only a detection of two signals. Increasing this threshold for considering the separation as a separated source enabled the algorithm to continue separating when the value was below this value thus giving three separations for mixtures with three signal sources. However, the disadvantage is the runtime that is higher with this setting.
The separated output signals of the four BSS algorithms are evaluated by an objective assessment using the MATLAB tool BSS-Eval that comprises 3 metrics, which are measured in decibel (dB): Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Signal to Artifact Ratio (SAR). A detailed description of these objective metrics is given in[26]. SIR is a measure for the separation performance, SDR measures the signal quality and SAR quantifies the degree of the artifact removal. The separated sources are compared to their respective original sources. However, due to quality enhancement during separation the separated sources have a higher range compared to their respective source signals. This makes them incomparable for the objective quality assessment. Thus, it becomes necessary to normalize the separated signals such that the range of the separated signals is brought to the range of its respective original signal before quality evaluation so that they are comparable. This is shown in Figure 5, 6, 7: Figure 5 shows the plot of an original source file, Figure 6 of the separated audio and Figure 7 where the range of separated audio is brought to the level of its respective original source range. However, evaluation of the separated sources using

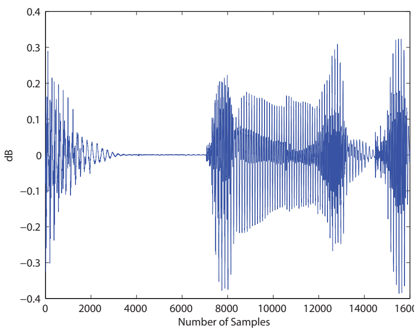

Fig. 5. Source audio plot

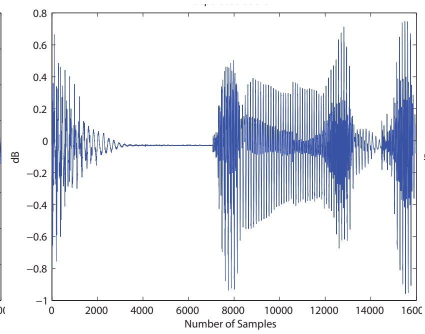

Fig. 6. Separated audio plot

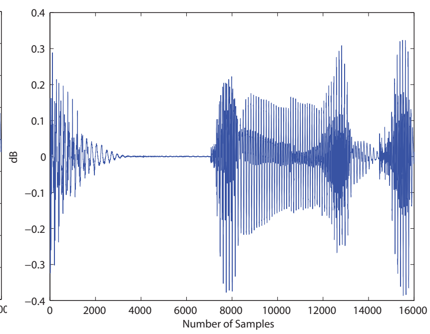

Fig. 7. Normalized Separated audio plot according to source range

only objective assessment does not infer on the basis of human perception. Thus, subjective assessment is performed using a small number of participants. The metrics used for subjective assessment are based on the following criteria: Q1 = Score for global quality compared to source for each separated signal; Q2 = Score for preservation of target signal in each separated signal; Q3 = Score for suppression of other sources in each separated signal; Q4 = Score for absence of additional noise in each separated signal. The scoring ranges from 1 to 5 where, 1 = Worst (Impossible to understand), 2 = Poor (Annoying and unclear), 3 = Fair (Slightly annoying), 4 = Good (Perceptible but not annoying), 5 = Excellent (Almost like the source). These subjective metrics are based on[27].

5. Conclusion and future work

The separation of the mixed signals is done using four algorithms: FastICA, ICA-BM, JADE and InfoMax. The samples used belong to three categories; samples one to four belong to OST which are speech mixtures only; the samples five and six belong to MSTwN which are speech and music mixtures; the samples seven to fourteen are speech mixtures, music and other sounds from the restaurant, children in school, siren and noise belonging to MSTN. The results of Objective and Subjective assessment are shown in Table 1 and 2 respectively. The objective assessment performed using the three metrics SDR, SIR and SAR as explained in the Experiment section show that the average

Table 1. Benchmarking results of the objective assessment

		Binary Mask			**FastICA**			**JADE**			**Infomax**		
Sample No.	*No. of Source*	$\overline{SDR}$	$\overline{SIR}$	$\overline{SAR}$	$\overline{SDR}$	$\overline{SIR}$	$\overline{SAR}$	$\overline{SDR}$	$\overline{SIR}$	$\overline{SAR}$	$\overline{SDR}$	$\overline{SIR}$	$\overline{SAR}$
1	2	14.3	31.1	14.4	12.2	14.9	15.8	10.1	12.9	13.8	4.3	19.9	4.5
2	2	10.2	20.5	10.7	23.4	24.6	30.3	23.2	24.3	29.9	4	24.6	4.1
3	3	8.4	17.2	9.1	17.0	18.0	24.6	17.4	20.0	22.1	4.4	19.3	4.6
4	3	6.6	23.7	6.7	8.5	11.5	12.1	15.7	18.7	19.6	4.3	19.6	4.5
5	2	7.9	25.4	8.1	27.4	33.1	29.6	7.8	19.1	8.2	5.3	27.2	5.4
6	3	11.1	19.7	11.8	23.1	26.8	28.8	25.1	28.8	31.4	4.1	23.1	4.1
7	2	13.4	22.6	13.9	18.4	34.3	18.6	33.3	45.4	34.8	3.8	28.3	3.8
8	3	7.2	19.7	7.9	24.2	26.2	30.5	24.8	26.7	31.2	4.3	21.3	4.5
9	2	8.9	22.1	9.3	11.9	32.9	12.1	11.8	32.8	12.1	4.5	25.6	4.5
10	3	10.7	23.4	11.1	9.7	12.2	14	16.1	18.3	20.6	4.8	20.2	4.9
11	2	4.4	16.2	5.6	33.8	38.9	36.5	24.4	30.9	27	4.7	28.3	4.8
12	3	1.9	18.6	2	10.1	12.5	14.3	15.9	18.3	20.2	4.9	20.5	5.0
13	2	20.5	30.3	20.9	4.2	19.2	4.5	4.2	19.2	4.5	4.4	27.6	4.5
14	3	13.1	23.4	13.8	10.6	13.1	15.3	10.5	13.0	15.2	4.7	19.2	4.9
Average		**9.9**	**22.4**	**10.4**	**16.8**	**22.7**	**20.5**	**17.2**	**23.5**	**20.8**	**4.5**	**23.2**	**4.6**

Table 2. Results of the subjective assessment

	Binary Mask				**FastICA**				**JADE**				**Infomax**			
Sample No.	$\overline{Q1}$	$\overline{Q2}$	$\overline{Q3}$	$\overline{Q4}$	$\overline{Q1}$	$\overline{Q2}$	$\overline{Q3}$	$\overline{Q4}$	$\overline{Q1}$	$\overline{Q2}$	$\overline{Q3}$	$\overline{Q4}$	$\overline{Q1}$	$\overline{Q2}$	$\overline{Q3}$	$\overline{Q4}$
1	4	4.4	4	3.6	4	4.4	4	3.6	4.6	4.6	4.3	4.6	3.7	4.4	3.2	4
2	4.4	4.5	4.2	4.1	4.4	4.5	4.2	4.1	4.6	4.5	4	4.4	3.8	4.1	3.3	4.3
3	4	4.5	3.7	3.7	3.9	4.4	3.7	3.72	4.5	4.6	4.1	4	3.7	4.3	3.1	4.1
4	4	4.4	3.8	4	4	4.4	3.8	4	4.4	4.7	4.4	4.4	4	4.3	3.6	4.3
5	4.2	4.7	4.1	4.2	4.2	4.7	4.1	4.2	4.4	4.7	4.1	4.1	4.2	4.3	4	4.4
6	4.4	4.6	4.2	4.2	4.3	4.6	4.2	4.2	4.5	4.7	4.4	4.2	4.4	4.4	4.1	4.4
7	4.6	4.6	4.3	4.3	4.6	4.6	4.3	4.3	4.4	4.7	4.3	4.1	4.3	4.45	4.1	4.3
8	4.3	4.4	3.8	4	4.3	4.4	3.8	4	4.6	4.5	4	4.1	4.1	4.4	4.1	4.3
9	4.2	4.3	4.3	4	4.2	4.3	4.3	4	4.4	4.3	4.2	4.2	4.1	4.6	4.3	4.3
10	4.2	4.3	3.9	3.8	4.2	4.3	3.9	3.8	4.3	4.4	4.2	3.5	4.1	4.5	4	3.6
11	4	4.3	4	3.9	3.9	4.3	3.9	3.9	4.1	4.3	4.2	3.8	4.3	4.5	4.4	3.9
12	4.2	4.3	4.1	4	4.3	4.3	4.1	3.9	4.4	4.4	4.1	4.3	4.1	4.5	4	4.2
13	4.5	4.5	4.5	4.2	4.5	4.5	4.5	4.2	4.5	4.4	4.4	4.1	4.3	4.6	4.5	4.2
14	4.3	4.5	4.4	4	4.3	4.5	4.4	4.1	4.5	4.5	4.6	3.9	4.1	4.5	4	4
Average	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**5**	**4**	**4**	**4**	**4**	**4**	**4**

score for the JADE is higher than FastICA with an improvement in 1 dB approximately, but much higher than ICA-BM and InfoMax. The category-wise evaluation shows that the interference of the other sources and noise in less with ICA-BM than with the InfoMax, where JADE and FastICA follow for OST which include mixtures of speech only. However, for MSTwN which includes mixtures of speech and music, FastICA outperforms all the other algorithms. Considering the third category, JADE is robust against interference of noise, music, speech etc.. The degree of removal of artifacts and distortions is seen the best for all three categories with JADE, and then with FastICA, ICA-BM and InfoMax show poor separation performance in existence of music, noise and other distortions. The effectiveness of the BSS algorithms is measured by the runtime in seconds. The average run-time of JADE over all samples is the fastest, followed by FastICA and InfoMax which are of the same level nearly 3 times slower than JADE. ICA-BM is the slowest algorithm with 15 times slower than JADE and 5 times slower than FastICA and ICA-BM. Further, a subjective assessment is performed to infer on the results of objective assessment. The metrics for subjective assessment are also explained in detail in the Experiment section. Table 2 shows that the results of the subjective assessment comply with the objective assessment. The average global quality of the separation and the suppression of other sources for all the categories are good with all the algorithms. However, the preservation of the target signal is best with Jade than with Fast-ICA and with ICA-BM and InfoMax which follow. The absence of additional noise

is better for Jade and FastICA when compared to ICA-BM and InfoMax. In case of ICA-BM, there always existed a musical noise in the separated signals except for the mixtures containing two speech sources.
The proposed smartphone application separates only artificial benchmark scenarios for simulation. For future work, the real-time sound source separation - an acceptable delay of the separation is allowed - as well as an improvement of JADE or other state-of-the-art source separation algorithms for convolutive mixtures - which are actually under benchmarking - are in our focus. Another topic for future work is the improvement and development of the NUI. A clearly structured Graphical User Interface (GUI) is essential for the usability especially for aged people. The GUI will be ergonomically optimized to satisfy usability standards. Furthermore, the NUI will be improved in its capacity as orientation guide by recognizing known speech signals (for example voices of the family members) and providing accessible photos or symbols of the telephone directory of the respective person.

Acknowledgements

This work as a part of the SMARTNAVI-project is funded by the German Federal Ministry of Education and Research (BMBF). The financial project organization is directed by the VDI/VDE Berlin.

References

1. National Institute on Deafness and other Communication Disorders., http://www.nidcd.nih.gow/health/statistics/Pages/quick.aspx (Accessed: June 2014).
2. ReSound Control, GN ReSound, Apr. 2014, http://www.gnresound.de/Services/ReSoundControlApp (Accessed: Mai 2014).
3. ReSound Smart, GN ReSound, Apr. 2014, http://www.gnresound.de/Services/Smartapp (Accessed: Mai 2014).
4. BioAid beta, AUD1, Dr. Nick Clark, Aug. 2013, http://bioaid.org.uk/ (Accessed: Mai 2014).
5. miniTek Remote App, Siemens Audiology Solutions, Apr. 2014, http://hearing.siemens.com/de/de/products/wireless/minitek-app/minitek-app.html (Accessed: Mai 2014).
6. E. M. Picou, App-titude: Smartphones as Remote Controls-for Hearing Aids, ASHA Leader 19 (2014) 38–39.
7. J. Common, C. Jutten, Handbook of Blind Source Separation, Academic Press, April 2010.
8. A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, 2001.
9. N. Das, A. Routray, P. Dash, ICA Methods for Blind Source Separation of Instantenous Mixtures: A Case Study, Neural Information Processing - Letters and Reviews 11 (11) (2007) 225–246.
10. S. Makino, H. Sawada, R. Mukai, S. Araki, Blind Source Separation of Convolutive Mixtures of Speech in Frequency Domain., IEICE TRans. Fundamentals E88-A (7) (2005) 1640–1655.
11. L. Ma, H. Li, An Algorithm based on Nonlinear PCA and Regulation for Blind Source Separation of Convolutive Mixtures., In Proc. of the International Conference on Life System Modeling, and Simulation, LSMS 2007 (2007) 1–9.
12. S. Amari, A. Cichoki, Adaptive Blind Signal Processing - Neutral Network Approaches, Prroc. IEEE 86 (10) (1998) 2062–2048.
13. J. Cardoso, A. Souloumiac, Blind Beamforming for non-Goaussian Signals, Proc. Ints. Elect. Eng., Radar Signal Process 140 (6) (1993) 362–370.
14. M. Pedersen, J. Larsen, U. Kjems, L. Parra, Convolutive Blind Source Separation Methods, Springer Handbook of Speech Processing (2008) 1065–1094.
15. A. Hyvarinen, E. Oja, Independent component analysis: Algorithms and applications, Neural Networks 13 (4) (2000) 411–430.
16. A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634.
17. The FastICA package for MATLAB, Aalto University, Department of Information and Computer Science, http://research.ics.aalto.fi/ica/fastica (Accessed: April 2014).
18. A. Bell, T. Sejnowski, An Information Maximization Approach to Blind Separation and Blind De-convolution., Neural Computation 7 (6) (1995) 1129–1159.
19. ICA - Independent Component Analysis, http://www.nlpca.org/ica_independent_component_analysis.html (2014).
20. M. S. Pedersen, D. Wang, J. Larsen, U. Kjems, Two-Microphone Separation of Speech Mixtures, IEEE Transactions on Neural networks 19 (3) (2008) 475–492.
21. M. Offiah, S. Rosenthal, M. Borschbach, Assessing the Utility of Mobile Application Support for or as Replacement of Hearing Aids., In Proc. of FNC Workshop.
22. Android NDK, https://developer.android.com/tools/sdk/ndk/index.html (2014).
23. GNU Octave, http://www.gnu.org/software/octave/ (Accessed: Mai 2014).
24. Binaural Hearing Aid - Spatial Hearing for all. In German: Binaurale Hörgeräte - räumliches Hören für alle., http://www.deutscher-zukunftspreis. de/de/nominierte/2012/team-1/ (Accessed: June 2014).
25. SISEC - Signal Separation Evaluation Campaign, https://sisec.wiki.irisa.fr/ (Accessed: April 2014).
26. E. Vincent, R. Gribonval, C. Fevotte, Performance Measurement in Blind Audio Source Separation, IEEE Transactions on Audio, Speech, and Language Processing 14 (4) (2006) 1462–1469.
27. V. Emiya, E. Vincent, N. Harlander, V. Hohmann, Multi-criteria subjective and objective evaluation of audio source separation, IEEE Transactions on Audio, Speech, and Language Processing 9 (7) (2010) 2046–2057.