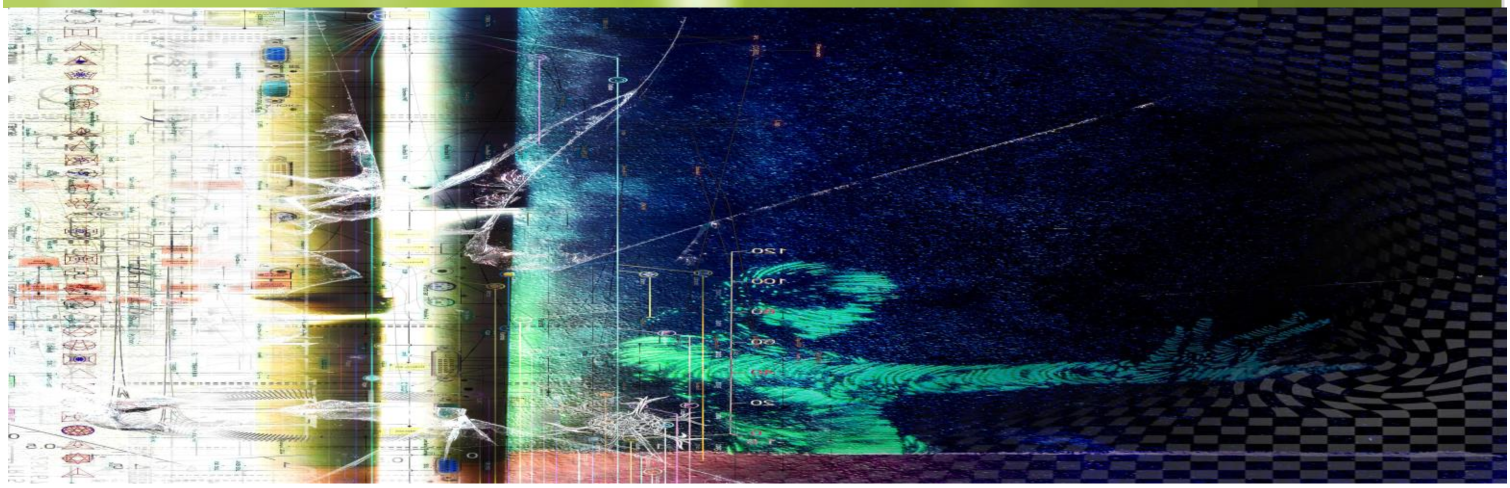
INNOVATION PROJECT

SIMULATION OF REALISTIC BACKGROUND NOISE USING MULTIPLE LOUDSPEAKERS

Technical University of Denmark: Juan David Gil Corrales, Marton Marschall, Torsten Dau

Brüel & Kjær: Wookeun Song, Claus Blaabjerg

GN Netcom: Michael Hoby Andersen, Søren W. Christensen



Algorithms for speech enhancement in modern telecommunications devices, such as noise and echo cancelation, make use of the acoustic background noise to improve the sound quality of the transmitted and received signals. During the development, improvement and evaluation of these algorithms, the devices are exposed to a reproduced sound that needs to be realistic, in order to guarantee that the performance of the device under laboratory conditions is as similar as possible to the performance in real use.

Different methods for reproduction of realistic sound exist, some of them require a lot of loudspeakers, and some others cannot reproduce the full characteristics of the sound field in the frequency range of interest. In this project, a new standardized method, based on inverse filters and matrix formulation of the crosstalk cancellation problem, was studied and compared to Higher-Order Ambisonics and the traditional four-loudspeaker-based method.

Technical University of Denmark

Brüel & Kjær

GN Netcom

COLOPHON

Publisher

Danish Sound Innovation Network
Technical University of Denmark
Richard Petersens Plads, Building 321,
2800 Kongens Lyngby, Denmark
+45 45253411
www.danishsound.org

November 2014

About the publication

This publication and possible comments and discussions can be downloaded from www.danishsound.org.

The content of this publication reflects the authors' point of view and not necessarily the view of the Danish Sound Innovation Network as such.

About the network

Danish Sound Innovation Network is an innovation network funded by the Danish Agency for Science, Technology and Innovation. The Network is hosted by the Technical University of Denmark and is headed by Director, Associate Professor, PhD Jan Larsen.

Danish Sound is the facilitator of the national ecosystem for SOUND, creates value for all parts in the value chain and contributes to growth and wealth in Denmark.

Network membership is free of charge and open for all. Registration at www.danishsound.org

This publication is the result of an innovation project, an instrument to strengthen the cooperation between knowledge institutions and private companies. The primary goal is to promote innovation by combining accessible/existing research and technologies with creative uses in order to facilitate the creation of new products, services or experiences. Innovation projects are mainly short term feasibility studies conducted on a pre-competitive level.

PREFACE

This publication is the result of an innovation project entitled “Simulation of realistic background noise using multiple loudspeakers”. The project is financed by the Danish Sound Innovation Network through a grant from the Danish Agency for Science, Technology and Innovation. The project is completed in the period 01.Sep.2014 – 30.Nov.2014 and managed by the Technical University of Denmark (DTU); the project manager is Torsten Dau. Additional project participants are: Juan David Gil Corrales, Marton Marschall (DTU), Wookeun Song, Claus Blaabjerg (Brüel & Kjær), Michael Hoby Andersen, and Søren W. Christensen (GN Netcom).

INTRODUCTION

BACKGROUND

Telecommunications devices are becoming more versatile instruments as new technologies are incorporated. The need of manufacturers to cover the different ways that customers make use of the devices necessitates adaptability and multipurpose functionality (e.g. headsets that connect wirelessly with multiple devices and can be used in the car, in the office, or in the streets).

Improved user experience requires that telecommunications devices provide high quality sound regardless of the background noise conditions. Techniques to enhance the sound quality include algorithms for noise and echo cancellation, in the transmitter or the receiver side. These algorithms are naturally developed in the laboratory, and the conditions in which they are tested must be as realistic as possible to be able to anticipate their performance in real use.

Reproduction of realistic background noise using multiple loudspeakers is one tool manufacturers can apply to evaluate the sound quality of these algorithms. Devices are exposed to a virtual sound environment and the features of the device are tested. The sound captured by the device has to be as similar as possible to the one it would capture out of the lab; therefore, any objective or subjective test, on the processed audio signals by the device, will accurately represent its actual performance.

Existing sound-field synthesis techniques, such as Wave-Field Synthesis or Higher-Order Ambisonics that allow accurate reproduction of arbitrary sound fields require a large number of loudspeakers, which in some cases reaches an amount that prevents easy implementation, or even standardization. However, a recent publication of a standardized method has shown good performance in reproducing the characteristics of magnitude and phase of the sound field in a wide frequency range with a limited number of loudspeakers. In this study, this method was tested and compared with existing techniques.

OBJECTIVES

The general goal of this project is to find an optimal setup with limited number of loudspeakers for reproduction of background noise for testing telecommunications devices. Specific objectives are:

- To record reference and reproduced background noise programs in different background noise conditions, as well as at different recording positions; which can be used for further investigations in multichannel reproduction of sound for testing telecommunications devices.

- To implement three methods for reproduction of sound using multiple loudspeakers: Higher-Order Ambisonics (HOA), four loudspeaker-based method (ETSI EG 202 396-1), and matrix inversion method (ETSI TS 103 224).
- To evaluate the performance of the reproduction of sound with different methods, by means of objective measures, using a headset.

IMPACT/EFFECT

The outcome of this project serves as a guideline for future implementations in the telecommunications industry. One of the implemented methods was recently standardized; therefore, this project helps the companies involved, to be up to date with the new requirements of the industry. The recordings and procedures established during the project, serve as input for future research projects, in which subjective quality of the reproduced sound can be assessed. The results of this project contribute indirectly to the development of better telecommunications devices, and consequently to improve user's experience.

METHODS AND RESULTS

RECORDING OF THE REFERENCE SOUND

As will be described later, the three methods implemented for this project use different methodologies to record the reference sound and subsequently calculate the loudspeaker signals. Therefore, to be able to compare the reproduced sounds, a controlled reference sound environment was emulated in two rooms, a classroom and a meeting room, both located at DTU in Lyngby, Denmark. Figure 1 shows a picture of the classroom.

Figure 1. A reference background noise scene using 6 loudspeakers arbitrarily placed in the classroom.

Six loudspeakers were placed at arbitrary positions in the two rooms. Three programs of speech material, two of music, and one of pink noise were played back through these loudspeakers[1]. The gains of the loudspeakers were adjusted by playing back speech, and measuring roughly 60dBA at 1m.

SOUND REPRODUCTION METHODS

Three methods for reproduction of sound using multiple loudspeakers were implemented: Higher-Order Ambisonics [1], a four-loudspeaker-based method [2], and the newly standardized method using inversion

[1] In this report, the results are shown for pink noise only; however, similar ideas can be concluded for other programs.

of a matrix of transfer functions measured in the reproduction room [3]. The main constraint when implementing these methods was the use of maximum 8 loudspeakers. The reproduction room selected is located in the headquarters of GN Netcom in Ballerup, Denmark; see Figure 2.

Figure 2. Reproduction room at GN Netcom.

A *JABRA SUPREME UC* headset was provided by GN Netcom, which was a modified model enabling access to the three built-in microphones without passing through the built-in digital signal processor of the headset; see Figure 3. These three microphone signals were used for evaluation of the reproduction with the three methods.

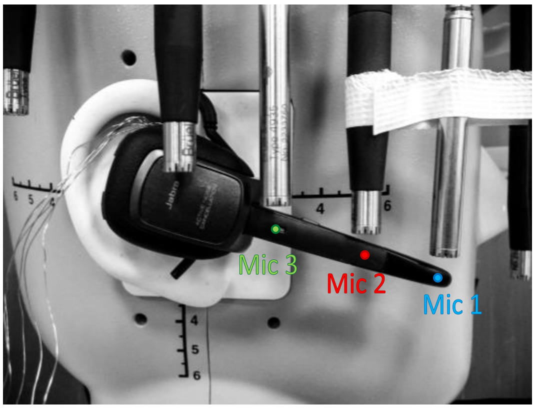


Figure 3. The JABRA SUPREME UC headset and the external microphone positions of the designed microphone array matching the approximate location of the built-in microphones of the headset.

HIGHER-ORDER AMBISONICS (HOA)

Higher-Order Ambisonics uses multiple loudspeakers to synthesize a sound field. The loudspeaker signals are determined by the spherical harmonics expansion of the reference sound. A spherical microphone array is used to capture these spherical harmonic components, and depending on the dimensions of the sphere, and the distribution of the microphones, certain directionality and resolution can be achieved.

The procedure to synthesize the sound field starts by encoding the directional information of the three-dimensional sound field. In the case of First-Order Ambisonics, this information can be captured by means of an omni-directional microphone and three figure-of-eight microphones, ideally located at the origin of the coordinate system. For higher orders, the spherical harmonics are recorded using a spherical microphone array like the one shown in Figure 4. In this project, 16 microphones located on the equator of the sphere were used, allowing to capture spherical harmonics up to third-order in the horizontal plane [1].

Figure 4. Spherical microphone array.

The decoding process consists of calculating the loudspeaker signals based on the measured spherical harmonics. In this project they were calculated for a regular loudspeaker layout composed of eight loudspeakers in the horizontal plane only.

The objective measures used to evaluate the reproduction of sound with the methods implemented are the magnitude spectrum difference in 1/3rd octave bands, the coherence, and the phase of the cross-spectrum between the reproduced and the reference sounds. The results of reproduction of pink noise using HOA, measured at the device are shown in Figure 5.

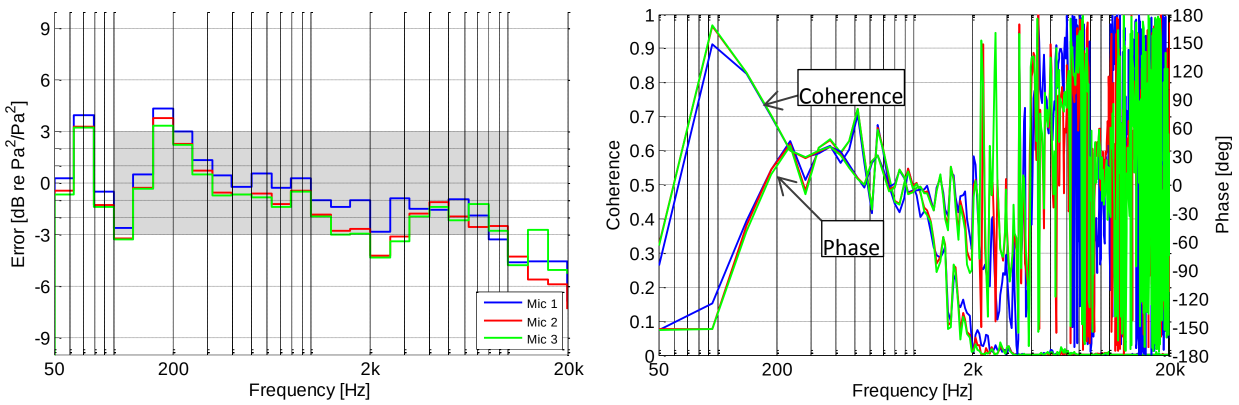


Figure 5. Magnitude spectrum difference, coherence and phase difference between reproduced sound and reference sound at the built-in microphones of the headset. Reproduction: Higher-Order Ambisonics.

FOUR-LOUDSPEAKER-BASED METHOD: ETSI EG 202 396-1

One standardized method for reproduction of sound uses four loudspeakers and it is based on binaurally recorded sounds using Head-And-Torso-Simulator (HATS) [2]. A calibration procedure is followed to derive equalization filters for each loudspeaker, and then use them to calculate the loudspeaker signals.

The location of the loudspeakers in the reproduction room should follow a square shape, with the HATS in the center. The first loudspeaker is placed at 45 degrees with respect to the median plane, and the rest, in the remaining corners of the square. The calibration procedure starts with a separate equalization of each of the loudspeakers using pink noise.

Then, the calculated equalization filters are used to filter a realistic background noise, and feed it to two loudspeakers at a time (front-right/rear-right, and front-left/rear-left). The equalization filter is adjusted to obtain reproduction errors within ±3dB. The same procedure is repeated again, but now the signal is fed to all four loudspeakers at the same time. In this step, and the previous one, one has to take into account that each time one loudspeaker is added, the sound pressure increases by 3dB, therefore the gain of the loudspeaker signals should be adjusted to maintain the sound pressure level. The influence of crosstalk is not taken into account with this method.

Figure 6 shows the errors of the reproduction of pink noise at the headset´s microphones when the reproduction of sound has followed the equalization procedure previously described.

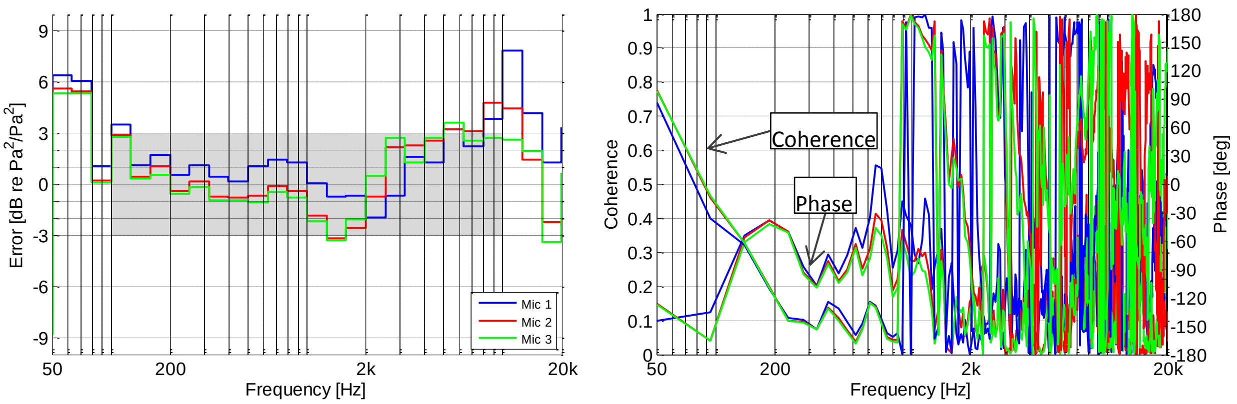


Figure 6. Magnitude spectrum difference, coherence and phase difference between reproduced sound and reference sound at the built-in microphones of the headset. Reproduction: Four-loudspeakers-based method.

MATRIX INVERSION METHOD: ETSI TS 103 224

A recently published standard [3] describes a procedure for reproduction of sound based on previous investigations by Kirkeby et al. [4] on a matrix formulation of the crosstalk cancellation problem, and deconvolution using regularization.

The reference sound for this method is recorded with a specially designed microphone array like the one shown in Figure 7. The purpose of this technique is to reproduce sound at particular positions in space, defined by the location of the microphones of the designed microphone array. This array has been designed to cover approximate location of modern telecommunications devices [3].

The designed microphone array is composed by eight target microphones (1-8) for which the reproduction of sound was optimized. Four additional validation microphones (v1-v4) have been added in order to monitor the response of the reproduced sound at places that were not optimized (i.e. when the microphone array is changed for the telecommunications device, the optimization positions might not match the position of the microphones of the device).

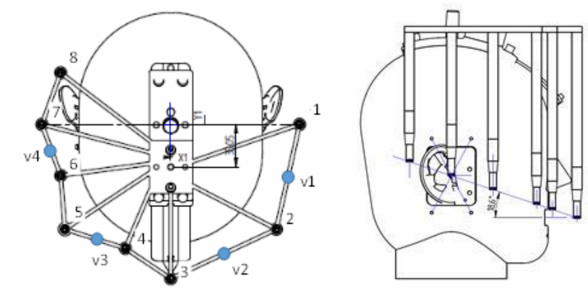


Figure 7. Sketch of the designed microphone array.

The optimization of the sound field in the reproduction room is done by measuring the transfer functions from every loudspeaker to every microphone of the designed array, represented by the matrix $\mathbf{C}$. The location of the loudspeakers is not important because the solution of the problem considers the electrical signals only as presented in Figure 8.

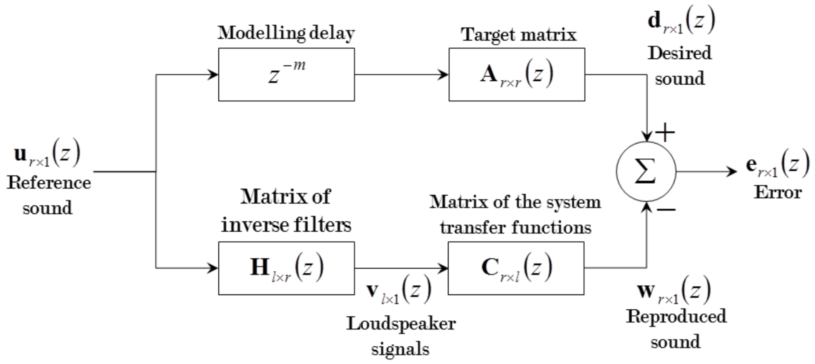


Figure 8. Block diagram form of the multichannel sound reproduction problem using the matrix inverse method, [4].

A set of inverse filters $\mathbf{H}$ can be calculated, based on the matrix of measured transfer functions $\mathbf{C}$, using Equation (1); where $\mathbf{A}$ and $\mathbf{I}$ are identity matrices; and β is the regularization parameter, which changes the solution from minimizing the squared error only to minimize the energy of the loudspeaker signals only, when it is changed from zero to infinity, respectively.

$$\mathbf{H}(z) = \frac{\mathbf{C}(z^{-1})^T \mathbf{A}(z)}{\mathbf{C}(z^{-1})^T \mathbf{C}(z) + \beta\mathbf{I}} \qquad (1)$$

In this project two methodologies to determine the regularization parameter were used: one basic setting in which the regularization was constant for all frequencies (this value was selected to $\beta = 10^{\frac{-15}{20}} = 0.1778$ after preliminary tests). The other methodology is the complete description of the Technical Specification ETSI TS 103 224 [3], which consists of a preprocessing of the impulse response of the room using a time-variant low-pass filter; the calculation of the inverse filter using sub-band division; and, an automated procedure to find the best regularization parameter based on the energy of simulated loudspeaker signals. The second methodology, named in this report as 'complete settings', uses a magnitude correction of the inverse filters based on the averaged reproduction error at the target and validation positions, obtained when reproducing pink noise.

Ideally, for this method, the error at the target positions, which are part of the matrix inversion, should be zero across the whole frequency range. For that reason, an additional optimization condition in which only three microphones placed close to the device's microphones were taken into account to evaluate a best case scenario; see Figure 3.

Due to the response of the loudspeakers at low and high frequencies, the non-minimum phase characteristics of the room impulse response, and the regularization applied, the response will not be perfect and some audible artifacts can arise.

In Figure 9, Figure 10 and Figure 11, the results of the reproduction of pink noise using the matrix inverse method are presented. Figure 9 shows the basic setting, in which the inverse filters were calculated using constant regularization for all frequencies taking as target positions the microphones 1-8 of the designed microphone array. It can be seen that the magnitude spectrum difference has been kept in the range of ±3dB for frequencies between 100Hz and 16kHz, and the coherence is higher than 90% from 100Hz to 2kHz. When the complete settings are taken into account for the optimization of the sound field, using again microphones 1-8, the magnitude spectrum difference at the microphones of the device is improved especially at low frequencies, but the coherence, and the phase difference is affected by the extra processing applied, see Figure 10.

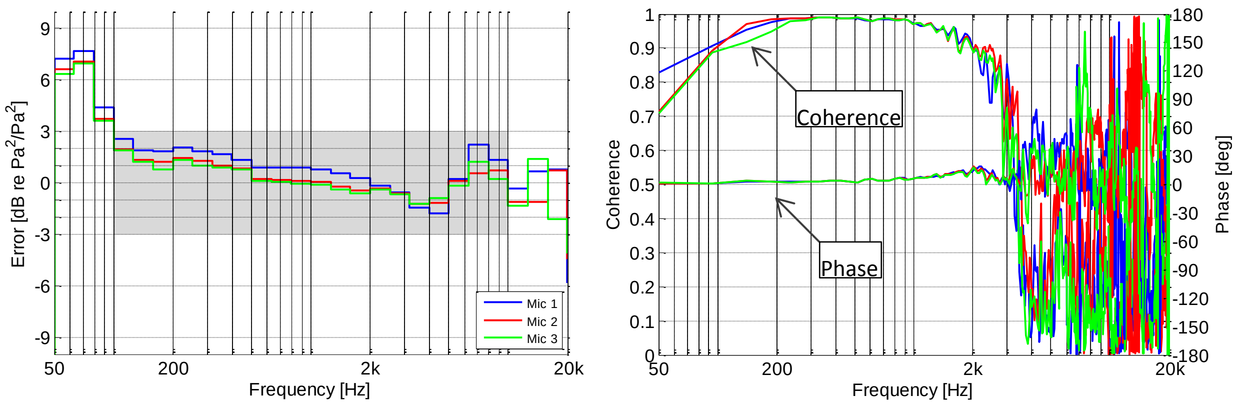


Figure 9. Magnitude spectrum difference, coherence and phase difference between reproduced sound and reference sound at the built-in microphones of the headset. Reproduction: Matrix inversion method, with basic settings.

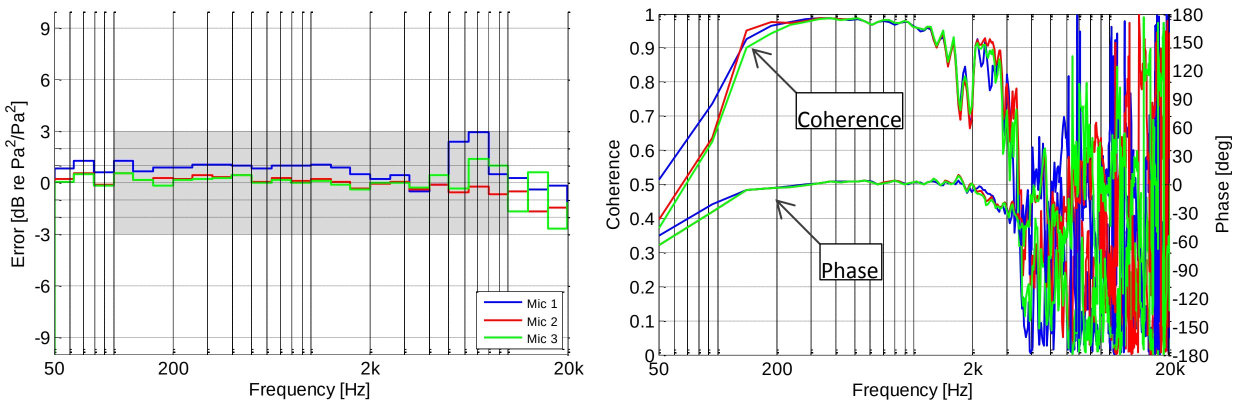


Figure 10. Magnitude spectrum difference, coherence and phase difference between reproduced sound and reference sound at the built-in microphones of the headset. Reproduction: Matrix inversion method, with complete settings.

In Figure 11, the same results are shown when the optimization is made at the microphones close to the headset's built-in microphones. The settings used were the basic ones as in Figure 9, but in this case, given the proximity of the target positions to the evaluation points, the coherence has been improved up to 3kHz, and except for one microphone, the magnitude spectrum difference remains in the error region of ±3dB.

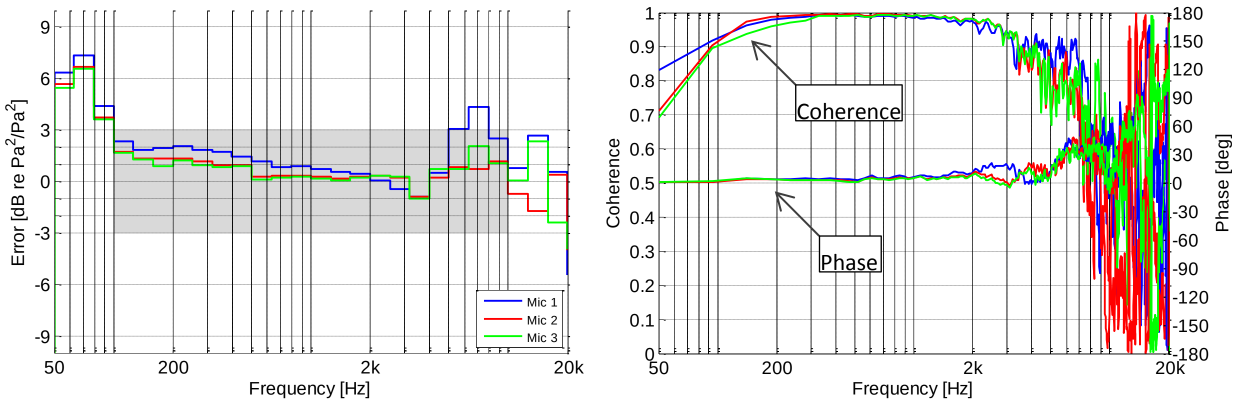


Figure 11. Magnitude spectrum difference, coherence and phase difference between reproduced sound and reference sound at the built-in microphones of the headset. Reproduction: Matrix inversion method, with basic settings; Optimization at positions close to the headset's microphones.

CONCLUSIONS

In this project, three methods for reproduction of sound were compared in the context of testing telecommunications devices. Each method relies on somewhat different assumptions to optimize the sound field. In terms of results, the magnitude spectrum differences were similar across methods, but the phase difference and coherence showed large variations.

Sound reproduction using HOA and the four-loudspeaker-based method showed acceptable performance in the frequency range of interest from 100Hz to 10kHz. Considering that the number of loudspeakers used for HOA was low, and that the four-loudspeaker-based method did not consider crosstalk cancellation, the errors obtained in terms of magnitude spectrum difference can be considered satisfactory.

The estimation of the coherence and phase difference between the reference and reproduced sounds was difficult for HOA and the four-loudspeaker-based method, because the precise delay between the two signals cannot be determined accurately. The graphs obtained were calculated using the lag corresponding to the maximum of the cross-correlation between the two signals, representing the best estimate possible.

The matrix inversion method was revealed to be the best method for reproducing sound for testing telecommunications devices. The magnitude spectrum differences stayed within ±3dB in the frequency range from 100Hz to 16kHz, and the coherence was higher than 90% for frequencies between 100Hz and 2kHz. Applying a magnitude correction of the inverse filters based on the measured errors in all microphone positions of the designed microphone array ("complete settings", as described in the Technical Specification ETSI TS 103 224 [3]), provided an improvement of the magnitude spectrum difference, but led to lower coherence at low frequencies.

A framework to perform recordings in reference and reproduction environments was implemented and tested during this project, which will facilitate future work in this area. The application of the new standard (ETSI TS 103 224) also raised questions and generated feedback about its implementation and descriptions.

It should be noted that these conclusions are based solely on the objective measures considered, and provide a first step in the comparison of background noise reproduction methods. The current evaluation might be sufficient for applications where only an accurate reproduction of the time-average sound pressure level is needed. However, perceptual artifacts are especially critical for applications where listening tests are involved; the filtering methods applied, particularly in the matrix inversion method, are known to cause time-domain artifacts, such as temporal smearing and pre-ringing. As these artifacts might be audible or otherwise affect sound quality adversely, it is important to consider perceptual effects as a next step. The importance of considering perceptual outcome measures was also demonstrated in a previous work [5] on which the current study is based.

REFERENCES

[1] T. Weller, S. Favrot, and J. Buchholz, "Application of a circular 2D hard-sphere microphone array for higher-order Ambisonics auralization," *Forum Acusticum 2011*, no. c, pp. 2269–2274, 2011.

[2] European Telecommunication Standards Institute, "Speech processing, transmission and quality aspects - Part 1: Background noise simulation technique and background noise database," *ETSI EG 202 396-1*, vol. 4, pp. 1–58, 2011.

[3] European Telecommunication Standards Institute, "A sound field reproduction method for terminal testing including background noise database," *ETSI TS 103 224*, vol. 1, pp. 1–36, 2014.

[4] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–194, Mar. 1998.

[5] J. D. Gil C., "Reproduction of realistic background noise for testing telecommunications devices," Technical University of Denmark, 2014.