

# Simulation of realistic background noise using multiple loudspeakers

W. Song<sup>1</sup>, M. Marschall<sup>2</sup>, J.D.G. Corrales<sup>3</sup>

<sup>1</sup>*Brüel & Kjær Sound & Vibration Measurement A/S, Denmark, Email: woo-keun.song@bksv.com*

<sup>2</sup>*Technical University of Denmark, Denmark, Email: mm@elektro.dtu.dk*

<sup>3</sup>*Odeon A/S, Denmark, Email: juandagilc@gmail.com*

## Abstract

Three methods for reproduction of sound using a maximum of eight loudspeakers were investigated in the context of testing telecommunication devices. They are the four-loudspeaker-based method as described in ETSI EG 202 396-1, Higher-Order ambisonics (HOA), and a matrix inversion method. HOA optimizes the reproduced sound at a sweet spot in the center of the array with radius determined by a spherical microphone array, which is used to derive the spherical harmonics decomposition of the reference sound. The four-loudspeaker-based method equalizes the magnitude response at the ears of a head and torso simulator (HATS) for sound reproduction, while the matrix inversion method optimizes the local sound field around a few target positions. The matrix inversion method had two conditions, i.e. with or without the extra processing steps described in ETSI TS 103 224; and three sets of optimization positions were defined, i.e. the ears of the HATS, positions close to a device under test, and standardized positions as described in ETSI TS 103 224. A listening experiment was performed to evaluate the perceived quality of the reproduced sounds at the microphones close to a device under test and at the ears of the HATS. The matrix inversion method performed best when listening to the reproduced sounds at target positions used for sound-field optimization and when listening to the microphones close to the device. HOA resulted in similar perceived quality as the matrix inversion method while a large degree of perceptual degradation was observed using the four-loudspeaker-based method.

## Introduction

Modern telecommunication devices use multiple microphones and advanced signal processing in order to enhance the speech signal in noisy environments. The development and performance evaluation of such devices increasingly necessitates the reproduction of more realistic and spatially accurate background noise scenes. This study focuses on a perceptual evaluation of several methods for generating such background noise scenes.

There exist a number of techniques for the reproduction of recorded sound fields using multiple loudspeakers. Higher-order ambisonics (HOA) [1], for example, decomposes the recorded sound field into a set of basis functions, termed spherical harmonics. Several studies have demonstrated the recording and playback of sound fields using spherical microphone and loudspeaker arrays, applying the principles of HOA. The approach attempts to recreate the recorded sound field by matching the reproduced basis functions to the recorded ones as closely as possible [2][3]. While HOA provides numerous advantages, such as independence of encoding and decoding, practical applications of the technique are limited by the large number of microphones and loudspeakers required to achieve good 3D sound field reconstruction. As a compromise, 2D HOA or mixed-order ambisonics can be applied, where a higher spatial resolution is only maintained in the horizontal plane, at the cost of reduced resolution in the upper and lower hemispheres [4][5]. However, in terms of sound field reconstruction, the upper frequency limit of HOA is still restricted with a reasonable number of loudspeakers (e.g. about 2 kHz with 8 loudspeakers in a circle).

In the context of testing telecommunication devices, the European Telecommunications Standards Institute (ETSI)

published the ETSI EG 202 396-1 recommendation for background noise reproduction using four loudspeakers [6]. The method attempts to reproduce the measured magnitude spectrum at the ears of a Head and Torso Simulator (HATS) with the two left loudspeakers playing the left ear signal and the two right loudspeakers playing the right ear signal. The main drawback of this approach is that the microphone positions for which the sound reproduction is optimized (i.e. the ears of the HATS) can be far away from the microphones on a device under test (typically located closer to the mouth). This becomes critical for modern telecommunication devices that make use of more than a single microphone for speech signal processing. Phase as well as magnitude differences between device microphones need to be accurately reproduced in order for the device to function as intended.

Matrix inversion methods can also be used to optimize the reproduced sound field locally [7]. The approach takes a recording from a set of microphones, and optimizes the reproduced sound field to best match the recording around the microphone positions. It is recommended that the number of loudspeakers be higher than or equal to the number of microphones in order to ensure that the system equation can be inverted stably. Even though the technique optimizes the sound field locally, a recent study [8] showed that it can also be used to reproduce the 3D sound field around a spherical microphone array using a spherical array of loudspeakers. Compared to HOA and the above-mentioned ETSI method, the technique is easy to implement and highly scalable, e.g. even allowing the use of two loudspeakers for cross-talk cancellation.

The newly introduced ETSI TS 103 224 recommendation [9] outlines the matrix inversion method for testing telecommunication devices. The proposed method seems to perform quite well in a given loudspeaker setup. However,

some details of the method were not adequately described, e.g. regarding the selection of the optimal regularization parameter. Therefore, a simpler method using a constant regularization parameter across frequencies is proposed in the current investigation. In addition, the matrix inversion method is also investigated using target positions (i.e. microphone positions optimized for sound reproduction) close to the microphones of the device under test. In summary, the present investigation contrasts the following five methods for the reproduction of background noise:

- 1) ETSI EG 202 396-1 (*etsi*)
- 2) Higher-order ambisonics (*hoa*)
- 3) Matrix inversion method (*knor*)
- 4) ETSI TS 103 224 (*kmod*)
- 5) Matrix inversion method optimized for a specific device (*kbin, kdev*)

A brief introduction to these five methods is given in the following section.

## Theoretical background

### ETSI EG 202 396-1 (*etsi*)

This method uses four loudspeakers to reproduce binaural recordings made with a Head-And-Torso-Simulator (HATS) [6]. An iterative calibration procedure is followed to derive equalization filters for each loudspeaker. The loudspeakers in the reproduction room are located in the corners of a square, with the HATS in the center, such that the first loudspeaker is placed at 45 degrees with respect to the frontal direction.

The calibration procedure starts with a separate equalization of each loudspeaker using pink noise. Then a realistic background noise is played through two loudspeakers at a time (right side, then left side), applying the equalization filters calculated in the previous step. The equalization filters are then adjusted to obtain magnitude errors within  $\pm 3$ dB. In the final step, the signal is fed to all four loudspeakers at the same time, and the equalization is adjusted again if the magnitude errors are larger than  $\pm 3$ dB. In order to maintain the sound pressure level at the microphones of the HATS, the level is decreased by 3 dB each time the number of active loudspeakers is doubled. The influence of crosstalk is not taken into account with this method.

### Higher-order ambisonics (*hoa*)

In HOA, the sound field is described by a set of spherical harmonic components or ambisonic signals, which encapsulate the directional information in the sound field. The recording and playback process involves estimating these ambisonic signals (encoding), and deriving the appropriate loudspeaker signals for a specific loudspeaker array (decoding). For example, in the case of first-order ambisonics, encoding can be accomplished by using an omni-directional microphone and three figure-of-eight microphones [10], ideally located at the center of a potential listener's head, without the presence of the head. For higher orders, ambisonic signals are recorded using a spherical microphone array like the one shown in Figure 2 [5]. In this project, a horizontal-only (2D) formulation of HOA was used, utilizing 16 microphones located on the equator of the

sphere [4]. Similarly, a circular loudspeaker layout was used for playback, composed of 8 loudspeakers in the horizontal plane (see Figure 5). A "mode-matching" decoding procedure was used [3], i.e. the loudspeaker signals were derived by prescribing that the reproduced ambisonic components match those of the reference sound field.

### Matrix inversion method (*knor*)

Kirkeby et al. [7] describe the matrix inversion method in detail; here, only a short summary is given. When playing a set of signals through  $l$  loudspeakers, the resulting signals measured at  $r$  microphone positions can be formulated as

$$\begin{pmatrix} w_1 \\ \vdots \\ w_r \end{pmatrix} = \begin{pmatrix} H_{1,1} & \cdots & H_{1,l} \\ \vdots & \ddots & \vdots \\ H_{r,1} & \cdots & H_{r,l} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_l \end{pmatrix} \quad (1)$$

where  $v$  are the loudspeaker signals,  $w$  are the microphone signals, and  $H_{r,l}$  is the room frequency response (RFR) between the  $l$ -th loudspeaker and the  $r$ -th microphone. The pseudo-inverse of the RFR matrix  $\mathbf{H}$  can be calculated by making use of Tikhonov regularization:

$$\mathbf{C} = (\mathbf{H}^* \mathbf{H} + \beta \mathbf{I})^{-1} \mathbf{H}^* \quad (2)$$

where  $\mathbf{C}$  is the inverse matrix of the RFR matrix  $\mathbf{H}$ ,  $\mathbf{I}$  is the identity matrix, and  $\beta$  is a regularization parameter. The regularization parameter  $\beta$  is introduced to make the matrix inversion more stable. The threshold for the regularization parameter in dB, i.e.  $20 \log_{10}(1/\beta)$ , is defined to quantify the amount of regularization. The matrix inversion procedure is applied for each frequency line separately, and the loudspeaker signals are calculated by the following equation:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_r \end{pmatrix} = \begin{pmatrix} C_{1,1} & \cdots & C_{1,r} \\ \vdots & \ddots & \vdots \\ C_{l,1} & \cdots & C_{l,r} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_r \end{pmatrix} \quad (3)$$

where  $u$  are the reference signals recorded with the same array of microphones, but in a reference sound field.

Instead of performing the matrix multiplication in the frequency domain, the equalization filters  $c_{l,r}(t)$  are calculated by taking the inverse FFT for each element of  $\mathbf{C}$ . Subsequently, the  $r \times l$  convolutions between the reference microphone signals and the equalization filters are computed in the time domain. The driving signal for each loudspeaker is obtained by summing the corresponding set of  $r$  filtered microphone signals. In the present study, magnitude equalization was applied in addition, in order to minimize the resulting magnitude error. The equalization was done across all microphones in the array, including the validation microphones.

The regularization parameter can be optimized for a given loudspeaker setup prior to running actual tests. This can be achieved by cycling through a number of regularization

thresholds and choosing the setting with the lowest magnitude error and highest coherence. The range of thresholds is suggested to be between -20 dB (10) to 40 dB (0.01), based on the experience in the current investigation. A step size of 5 or 10 dB can be selected.

Additional care must be taken when choosing a regularization parameter, as audible artefacts may be produced when a low regularization threshold is used, despite the resulting magnitude errors being relatively small.

### ETSI TS 103 224 (*kmod*)

The basic algorithm of this recommendation is the same as the matrix inversion method (*knor*) described in the previous section. However additional steps are used to simplify the measured room impulse responses, to find the regularization parameters across frequencies automatically, and to compensate the magnitude errors at the microphones in the array mounted around a HATS. A description of the additional processing steps is given in the recommendation [9]. Unfortunately, these additional steps are not adequately described, and major details seem to be missing. Therefore, the implementation used in the current study may not always follow the recommendation as it was intended. While the method aims to reduce distortion in the reproduced sound field, the simplification of the impulse responses by low-pass filtering may lead to incorrect phase information at high frequencies even at the target positions, i.e. the microphones where the sound field is optimized.

### Matrix inversion method optimized for a specific device (*kdev*, *kbin*)

As the matrix inversion method optimizes the sound field locally, it can be expected that performance improves if the target microphones are positioned close to the microphones of the devices under test. When the target microphone positions are further away from the device, the coherence between the reference and reproduced fields may drop significantly at high frequencies, despite the magnitude error being within a reasonable range. A decrease in coherence may indicate potential perceptual differences between the reference and reproduced sounds. To investigate this, in this condition the setup is optimized for a specific device, by either using the three array microphones close to the built-in microphones of the device (*kdev*), or the HATS microphones (*kbin*) as target positions.

## Method

### Subjects

Fifteen normal-hearing listeners participated in the listening tests. All listeners were employees or students at the Technical University of Denmark. The subjects' hearing thresholds were measured using standard pure-tone audiometry, and none of the thresholds were found to exceed 20 dB hearing level. The subjects were also screened for known hearing problems using a questionnaire, and they were paid for their participation. None of the subjects were familiar with the test stimuli prior to the experiment.

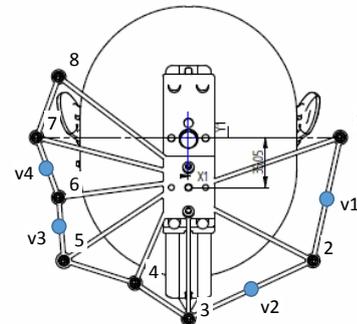
### Apparatus and stimuli

For HOA, a 52-channel spherical microphone array with a radius of 5 cm shown in Figure 1 was used. In this investigation, 16 microphones located on the equator of the sphere were used, which allows recording up to 7<sup>th</sup> order, horizontal-only HOA. However, as a practical reproduction system consisting of 8 loudspeakers was considered, only 3<sup>rd</sup> order HOA was used in the encoding and decoding process.



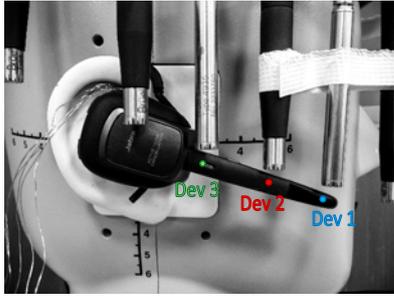
**Figure 1:** The 52-channel microphone array for HOA used in the current study.

A microphone array to be mounted around the head of the HATS was built for the current investigation. The array consists of 8 target microphones and 4 additional validation microphones positioned between the target microphones. The complete arrangement is depicted in Figure 2. The target positions were used to optimize the reproduced sound field, while the validation microphones were only used to compensate the magnitude errors and to evaluate the quality of sound reproduction both objectively and subjectively.



**Figure 2:** The microphone array mounted on a Brüel & Kjær 4128 HATS. Target microphone positions are labeled 1-8, validation positions v1-v4.

A commercially available communication headset, depicted in Figure 3, was used as the device under test. The headset was modified to allow direct access to the signals of the three built-in microphones, bypassing any processing done by the built-in DSP. The three microphones on the array that were closest to the built-in microphones of the device (“Dev1-3”) are the ones labeled “v3”, “Mic 6”, and “v4” in Figure 2. The distance between the built-in microphones and the neighboring array microphones was approximately 0.5 cm. The objective analysis was done using one of the device microphones directly, while for the subjective analysis, position “v3” was used instead to investigate the quality of reproduction at the input of the headset device.



**Figure 3:** A headset and the microphone positions of the designed array matching the approximate location of the built-in microphones of the headset, which are marked as “Dev 1”, “Dev 2”, and “Dev 3”.

To set up the reference scenes (i.e., the scenes to be reproduced) for the present study, 6 loudspeakers were placed in arbitrary locations in a class room at the Technical University of Denmark (DTU) (see Figure 4). The level of each loudspeaker was calibrated such that the A-weighted sound pressure level (SPL) at 1 m distance was approximately 60 dBA with pink noise. The 6 loudspeakers played incoherent pink noise to obtain the reference signals measured at the microphones around the HATS, and this recording was used for the objective performance analysis of the five reproduction methods.

For the listening experiment, a speech sample, as well as a music sample were used. The loudspeakers either played speech from different talkers (speech program), or sounds from different musical instruments (music program). In both cases, the stimulus started with only one loudspeaker playing, with an additional loudspeaker fading in at 2 second intervals, until all loudspeakers were playing.



**Figure 4:** A reference background noise scene using 6 loudspeakers arbitrarily placed in a classroom.

An IEC listening room was used for the sound reproduction setup. The RT60 of the room was approximately 0.4 s. Eight loudspeakers were positioned in a circle with a radius of 2 m, as shown in Figure 5. The height of the loudspeakers’ acoustic center was adjusted to the ear entrance point of the HATS. For  $k_{dev}$ ,  $k_{bin}$ , and  $k_{nor}$ , -15 dB regularization threshold was used for sound reproduction.



**Figure 5:** The loudspeaker setup in the IEC listening room

### Procedure

In order to compare the subjective quality of the reproduction methods, three perceptual attributes, namely *similarity*, *localization*, and *noise*, were evaluated on the two program materials. The stimuli were recorded in the reproduction setup at three sets of positions: microphone position “v3” close to the device (*device*, see Figure 2), the two ears of the HATS (*hats*), and at positions “v1” and “v4” on the microphone array (*array*, see Figure 2).

In the listening test, the sound samples were presented binaurally through Sennheiser HD 650 headphones. The headphone frequency response was equalized by applying the inverse headphone transfer function measured on a HATS. All three recording positions were evaluated for similarity. Localization was evaluated only using the recordings from the HATS, while noise was only tested on the recordings from the array microphones.

A psychophysical method similar to "MULTI Stimulus test with Hidden Reference and Anchor (MUSHRA)" [11] was applied, but the method did not use anchors. Figure 6 displays the user interface presented to the subjects participating in the listening experiment. Subjects were asked to compare ten different sets of stimuli, where each set, forming a single trial, was presented on a separate page. The subjects were asked to rate the sound samples based on one of three attributes, using sliders that were labeled with two adjectives, as shown in Figure 6. Table 1 summarizes the adjectives describing an opposing pair for each attribute. There were 6 trials for similarity (three recording positions and two program materials), and 2 trials each for localization and noise (one recording position and two program materials).

The sequence of trials, as well as the order of the buttons within each trial were created using balanced Latin Square design [12], in order to minimize potential order effects.

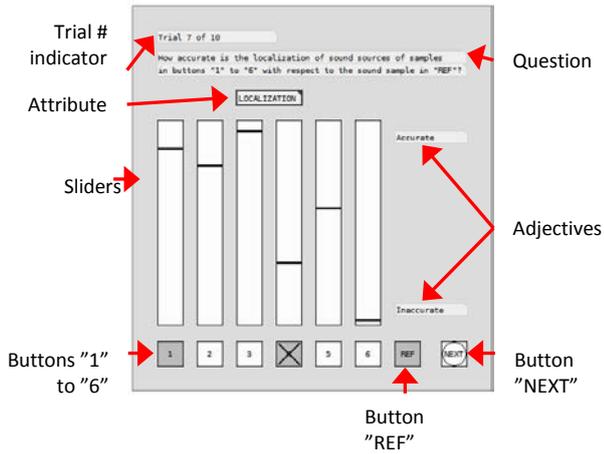


Figure 6: The user interface used in the listening experiment

Attribute	Adjectives	
Similarity	<i>different</i>	<i>similar</i>
Localization	<i>inaccurate</i>	<i>accurate</i>
Noise	<i>equally noisy</i>	<i>noisier</i>

Table 1: Attributes and the corresponding adjectives used in the listening test

The experiment began with a training session, consisting of one trial for each of the three perceptual attributes. The training material for similarity was recorded at the *array* position, while for localization and noise, the HATS recordings were used. The order of the trials in the training session was counter-balanced across subjects.

## Results

### Objective Analysis

The recordings using incoherent pink noise were used to analyze the performance of the different reproduction methods in terms of the magnitude error, as well as the coherence between the reference and reproduced signals. The estimation of the coherence was challenging for HOA and the ETSI EG 202 396-1 method, as the reproduced sound field was not specifically optimized near the device microphones. As a result, the precise delay between the reference and reproduced signals could not be determined accurately. The results shown here are calculated using the lag corresponding to the maximum of the cross-correlation between the two signals. To show the performance at target microphones, the results measured at microphone 6 are shown in Figure 7. The three methods based on matrix inversion (*kdev*, *kmod*, and *knor*) give rise to magnitude errors within  $\pm 3$  dB, while the 4-loudspeaker ETSI EG 202 396-1 (*etsi*) method and HOA result in larger errors. This discrepancy may be caused by the fact that the matrix inversion methods specifically optimize the sound field near the microphones of the array around the HATS and that additional magnitude error correction is applied, while this is not the case for *etsi* and HOA. When inspecting the coherence results, the same tendency can be observed. The sharp fall in coherence above 300 Hz reflects the difficulty in time-aligning the reference and reproduced signals for *etsi* and HOA. It is also noticeable that the performance of ETSI

TS 103 224 (*kmod*) above around 3 kHz degrades further than *knor* or *kdev*. This degradation is caused by the 2 kHz low-pass filtering of the room impulse responses employed in *kmod*.

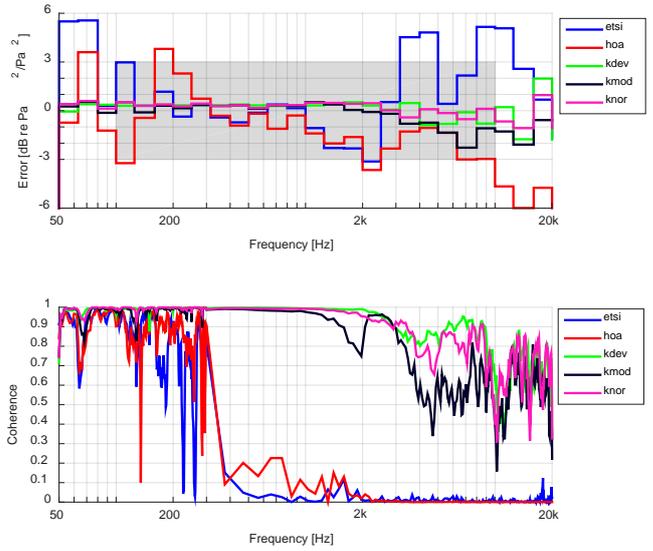


Figure 7: Magnitude error and coherence at microphone 6 across different reproduction methods.

When calculating the errors at a microphone of the headset (“Dev 2”) directly, the results are similar as for the previous position: the matrix inversion methods show the lowest magnitude error and highest coherence (see Figure 8). However, there is a large difference in coherence between *kdev* and the other two matrix inversion methods in this case. This may be attributed to the fact that *kdev* employs an over-determined system equation when deriving the equalization filters for sound reproduction. More details on the objective analysis can be found in [13].

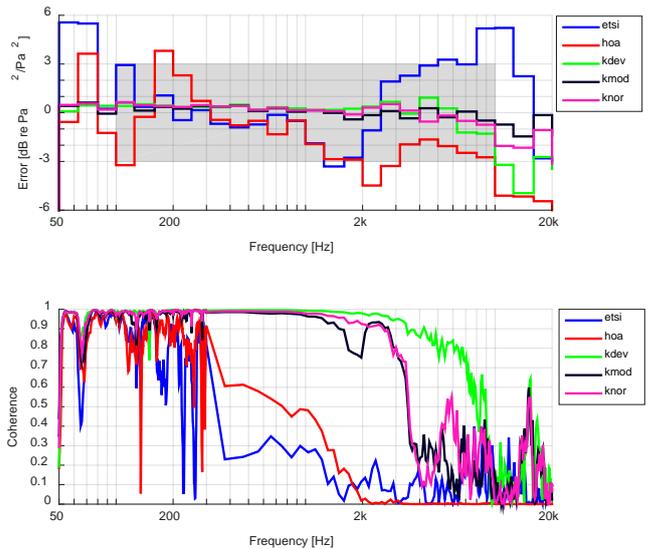


Figure 8: Magnitude error and coherence at microphone “Dev 2” across different reproduction methods.

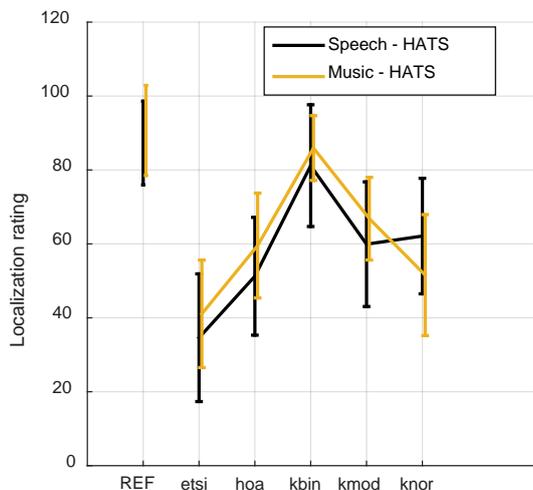
Despite the fact that the objective analysis shows obvious differences between the reproduction methods, it is not clear to what extent the objective measures chosen here reflect a

perceptually relevant difference. Furthermore, the challenges regarding the calculation of the coherence functions with ETSI EG 202 396-1 (*etsi*) and HOA may question the validity of the comparison with the matrix inversion methods. Therefore, a subjective listening test was performed to investigate perceptual differences between the reproduction methods.

### Subjective Analysis

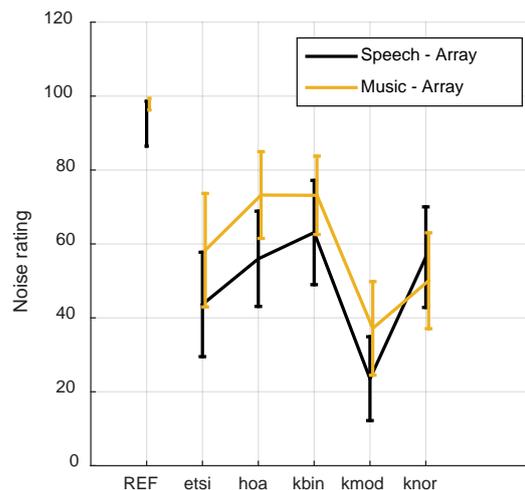
One of the main purposes of using several loudspeakers for sound reproduction is to preserve the spatial perception of the recorded sound field. Spatial perception was tested only using the binaural microphones of the HATS, as the other microphone positions do not provide realistic spatial cues.

Figure 9 shows the results of the subjective localization ratings. It can be seen that the *kbin* method performs the best, and the perceived localization is close to that of the reference, i.e. the binaural signal recorded in the reference sound field. It appears that the optimization of the reproduced sound field with the HATS microphone positions as a target largely maintains the perceived spatial characteristics of the reference field. As expected, *etsi* performs worst, since this method does not correct for the cross-talk between the loudspeakers, and only optimizes the magnitude spectrum of the reproduced field. It is interesting to note that HOA performs as well as *kmod* and *knor*, and which is contrary to the findings of the objective analysis.



**Figure 9:** Mean subjective localization ratings when listening to the HATS microphone signals (*hats* position). 95% CIs shown.

When the subjects are asked to judge noise in the recordings (see Figure 10), HOA and *kbin* seem to perform the best, with *kmod* being reported as most noisy. In general, the matrix inversion methods may be affected by processing artefacts, such as temporal smearing, especially when listening to non-target positions.

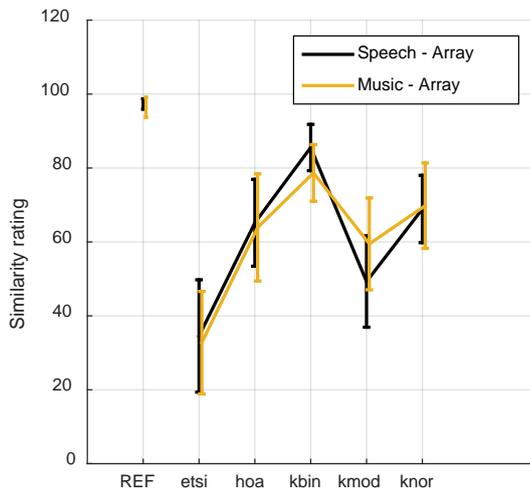


**Figure 10:** Mean subjective noise ratings (lower values mean more noisy) when listening to signals from the array microphones “v1” and “v4” (*array* position). 95% CIs shown.

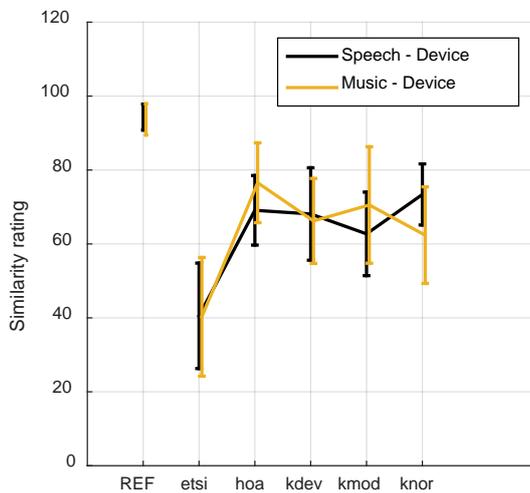
Subjective similarity ratings are shown in Figures 11-13, for three recording positions. In terms of similarity, optimizing the sound field near the listening position, as for the *kbin* condition in Figures 11 and 13, yields the highest ratings, with the exception of the *device* listening position shown in Figure 12. These cases have the advantage of utilizing an over-determined system equation and using either target or close to target positions for listening. Similarly to the previous attribute ratings, the subjects report large differences between the reference and the *etsi* and *kmod* stimuli, especially when listening to the HATS microphone signals. Listeners indicate a relatively high degree of similarity between HOA and the reference stimulus, despite the fact that the method incorporates very little knowledge of the playback environment, i.e., the sound field is not optimized at a specific position, and no room compensation is employed.

The results for the *device* recording position (Figure 12), show that listeners had difficulty in distinguishing the reproduction methods other than *etsi*. These stimuli were from a single microphone presented diotically, which resulted in a loss of most spatial cues, and may explain the different pattern of similarity ratings here.

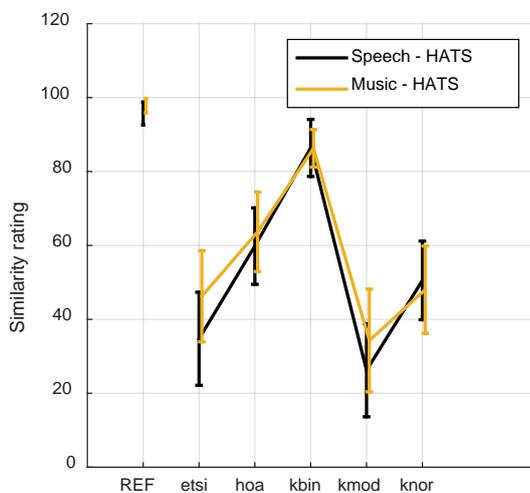
In general, there does not appear to be a significant difference between the two program materials in terms of the attributes tested.



**Figure 11:** Mean subjective similarity ratings when listening to signals from the array microphones “v1” and “v4” (*array* position). 95% CIs shown.



**Figure 12:** Mean subjective similarity ratings when listening to microphone “v3” (*device* position) 95% CIs shown.



**Figure 13:** Mean subjective similarity ratings when listening to the HATS microphone signals (*hats* position). 95% CIs shown.

## Conclusions

Five sound reproduction methods were compared in terms of magnitude error and coherence in the objective analysis, and in terms of the perceptual attributes *localization*, *noise* and *similarity*, in the listening experiment. The matrix inversion method optimized for specific microphone positions, i.e. *kdev/kbin*, performed the best both in the objective as well as in the subjective analysis. These methods utilize an over-determined system equation, i.e. use more loudspeakers than target microphones, and this may be the main reason why they are able to reproduce the reference sound field well near the target microphones.

In general, higher-order ambisonics scored well on the subjective ratings, but showed large errors in the objective analysis. The latter is not surprising, as the playback room was not anechoic, and no room compensation was employed. Since the decoding procedure in HOA is relatively straightforward compared to the other methods, HOA could be useful in cases where subjective rather than objective quality is important.

The additional procedures employed in ETSI TS 103 224 (*kmod*) do not seem to improve the quality of background noise reproduction; rather, degradations were observed both in terms of subjective and objective quality measures. The matrix inversion method with a constant regularization parameter, i.e. *knor*, performed as well as or better than ETSI TS 103 224 (*kmod*) on most measures. However, the implementation of *kmod* used in the current study could not follow the recommendation where the algorithms were not specified in sufficient detail, which may have affected the results for this particular method.

## Acknowledgments

The authors would like to thank Michael Hoby Andersen and Søren W. Christensen at GN Netcom for providing the headset for the investigation.

## References

- [1] J. Daniel, “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un context multimédia.” PhD thesis, University of Paris VI, France, 2000
- [2] S. Moreau, J. Daniel, S. Bertet, “3D sound field recording with higher order ambisonics – objective measurements and validation of spherical microphone,” AES 120th Convention (2006), paper nr. 6857
- [3] M. A. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” J. Audio Eng. Soc., Vol 53., (2005)
- [4] T. Weller, S. Favrot, J. M. Buchholz, “Application of a circular 2D hard-sphere microphone array for higher-order Ambisonics auralization,” Proceedings of Forum Acusticum (2011)

- [5] M. Marschall, S. Favrot, J.M. Buchholz, "Robustness of a mixed-order Ambisonics microphone array for sound field reproduction," AES 132<sup>nd</sup> Convention (2012), paper nr. 8645
- [6] Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database, ETSI EG 202 396-1 (European Telecommunications Standards Institute, Sophia Antipolis, 2006)
- [7] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, H. Hamada, "Local sound field reproduction using digital signal processing", J. Acoust. Soc. Am. Vol. 100. No. 3 (1996), 1584-1593
- [8] P. Minnaar, S. F. Albeck, C. S. Simonsen, B. Søndersted, S. A. D. Oakley, J. Bennedbæk, "Reproducing real-life listening situations in the laboratory for testing hearing aids", the 135th Audio Engineering Society Convention (2013), paper nr. 8951
- [9] Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database, ETSI TS 103 224 (European Telecommunications Standards Institute, Sophia Antipolis, 2014)
- [10] M. A. Gerzon, "Periphony: With-Height Sound Reproduction". J. Audio Eng. Soc Vol. 21 No. 1 (1973), 2-10
- [11] Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems, International Telecommunication Union, 2003
- [12] D. C. Montgomery, "Design and Analysis of Experiments", John Wiley & Sons, INC. 2001
- [13] J.D.G. Corrales, M. Marschall, T. Dau, W. Song, C. Blaabjerg, M. H. Andersen, and S. W. Christensen, "Simulation of realistic background noise using multiple loudspeakers", Danish Sound Innovation Project Report, 2014.