PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Binaural reproduction capability for multiple off-axis listeners based on the 3-channel optimal source distribution principle

Motoki YAIRI[1], Takashi TAKEUCHI[2], Keith HOLLAND[3],
Dylan G. MORGAN[4] and Laurence HAINES[5]

[1] KAJIMA Technical Research Institute, Japan

[2] OPSODIS Limited, University of Southampton, ISVR, United Kingdom

[3-5] University of Southampton, ISVR, United Kingdom

ABSTRACT

This paper studies the capability of Optimal Source Distribution (OPSODIS) numerically and experimentally for simultaneously synthesizing virtual auditory space to multiple listeners. The OPSODIS is a binaural synthesis method over loudspeakers which utilizes the idea of a pair of conceptual monopole transducers whose azimuthal location varies continuously as a function of frequency. The authors have revealed so far that this makes it possible to provide the same binaural signals not only to the on-axis target listener but also to multiple off-axis listeners. The OPSODIS can be augmented to three channels by adding a center channel to the left/right configuration described above. The three-channel system is potentially superior to the two-channel one as for the on-axis target listener's sweet spot. In this paper, whether the sweet spot of off-axis listeners can also be enhanced by the three-channel system is discussed. Numerical simulations under free-field conditions and experiments in a listening room are carried out. It is proved that the three-channel OPSODIS can also provide the additional controlled regions for the off-axis listeners, which are more robust than those given by the two-channel one.

Keywords: Binaural synthesis, Optimal source distribution, Three-channel system, Off-axis listeners

1. INTRODUCTION

Since M. R. Schroeder et al.[1] had developed it as for studying the room acoustics, 3D sound reproduction technologies based on the binaural synthesis over loudspeakers have been studied by many researchers so far. [2-4]

The system inversion involved in the binaural synthesis gives rise to a number of problems such as a loss of dynamic range, and a lack of robustness to room reflections and individual differences in the head related transfer functions. In order to overcome those fundamental problems, the principle of the optimal source distribution (OPSODIS) have been proposed, which utilizes the idea of a pair of conceptual monopole sources whose azimuthal location varies continuously as a function of frequency.[5] This has a minimum configuration of two-channel system (left and right). The simple signal processing and loudspeaker arrangement coming from the principle of the OPSODIS also overcomes another problem in the conventional binaural syntheses in that the desired virtual auditory space can be synthesized only for one on-axis target listener whose inverse matrix are estimated. The authors have revealed that the principle makes it possible to provide the independent control of the binaural signals not only to the on-axis target listener but also to multiple off-axis listeners.[6]

The OPSODIS can be augmented to the three-channel system by adding another source in the median plane to the left/right configuration of the two-channel system.[7] The three-channel system is potentially superior to the two-channel one regarding the various kinds of errors described above

[1] yairi@kajima.com
[2] tt@isvr.soton.ac.uk
[3] krh@isvr.soton.ac.uk
[4] dylangmorgan@icloud.com
[5] laurencehaines96@gmail.com

including the deviations in listening positions of the on-axis target listener.

In this paper, based on the above findings, whether the deviations in listening positions of the multiple off-axis listeners can also be enhanced by the three-channel system is discussed. Numerical simulations under free-field conditions and experiments in a listening room are carried out. It is proved that the three-channel system can also provide the additional controlled regions for the multiple off-axis listeners, which are more robust to the deviations in listening positions than those given by the two-channel one especially at higher frequencies.

2. PRINCIPLE OF THE OPTIMAL SOURCE DISTRIBUTION

2.1 Potentials of inverse filter matrices

A fundamental principle for the binaural synthesis over loudspeakers with system inversion is illustrated in Fig. 1, which are also known as crosstalk cancelation. We consider two- and three-channel symmetric systems in which the geometrical relations between sources and receivers (a listener's both ears) illustrated in Fig. 2. The two-channel system is configured by only the left and right channels, and another one in the median plane is added to the left/right configuration when it comes to the three-channel system.

The plant matrix (a matrix of transfer functions between sources and receivers) for each system can be expressed as follows: in the case of the two-channel system,

$$\mathbf{C} = \begin{bmatrix} 1 & e^{-jk_0 \Delta r \sin\theta} \\ e^{-jk_0 \Delta r \sin\theta} & 1 \end{bmatrix}, \tag{1}$$

and in the case of the three-channel system,

$$\mathbf{C} = \begin{bmatrix} 1 & ge^{-j\frac{1}{2}k_0 \Delta r \sin\theta} & e^{-jk_0 \Delta r \sin\theta} \\ e^{-jk_0 \Delta r \sin\theta} & ge^{-j\frac{1}{2}k_0 \Delta r \sin\theta} & 1 \end{bmatrix}, \tag{2}$$

where time dependence is assume with the wavenumber $k_0 = \omega/c_0$, with the angular frequency ω and the sound speed in the air c_0. The plant matrices are normalized by the sound pressure at the left ear, and g is the relative sensitivity of the center channel with respect to those of the left and right. In this model, we assume monopole sources and receivers under the free-field condition also not including the effect of the head related transfer functions (HRTFs) in order to understand the physics underlying the inverse filter matrices.

A desired inverse filter matrix $\mathbf{H}$ in order to realize the independent control at the two receivers can be derived from the condition satisfying the following relation:

$$\mathbf{I} = \mathbf{CH} \quad , \tag{3}$$

where $\mathbf{I}$ is the identity matrix. The maximum amplification of the source strengths required for the arbitrary binaural signal input at each frequency can be found from the 2-norm of $\mathbf{H}$ (denoted as $||\mathbf{H}||$) which is the largest of the singular values of $\mathbf{H}$ where these singular values are denoted by σ_i and σ_o:

$$\|\mathbf{H}\| = \max(\sigma_i , \sigma_o) \quad , \tag{4}$$

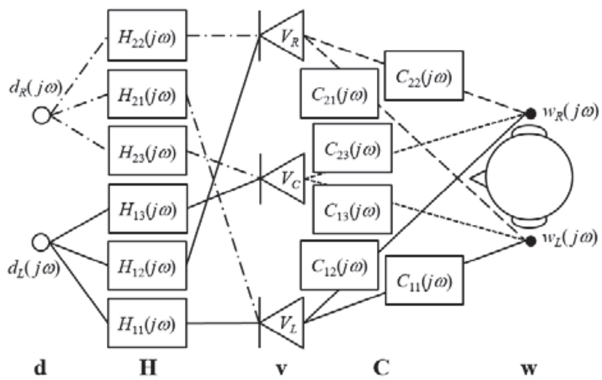

Figure 1 – Block diagram for the binaural synthesis with system inversion.

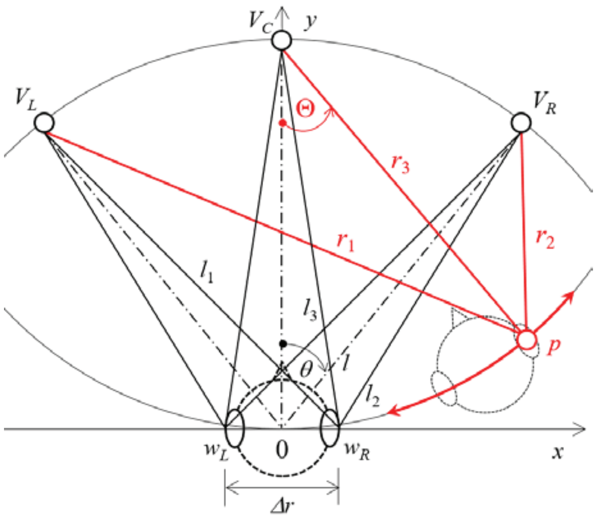

Figure 2 – Geometry of a two/three-source two-receiver system under investigation.

where σ_i and σ_o correspond to the amplification factor of the in-phase and out-of-phase components of the desired binaural signals. The singular values σ_i and σ_o can theoretically derived from the singular value decomposition and expressed as follows: in the case of the two-channel system,

$$\begin{cases} \sigma_i = \dfrac{1}{\sqrt{(1+e^{-jk_0\Delta r\sin\theta})(1+e^{jk_0\Delta r\sin\theta})}} \\ \sigma_o = \dfrac{1}{\sqrt{(1-e^{-jk_0\Delta r\sin\theta})(1-e^{jk_0\Delta r\sin\theta})}} \end{cases}, \tag{5}$$

and in the case of the three-channel system,

$$\begin{cases} \sigma_i = \dfrac{1}{\sqrt{(1+e^{-jk_0\Delta r\sin\theta})(1+e^{jk_0\Delta r\sin\theta})+2g^2}} \\ \sigma_o = \dfrac{1}{\sqrt{(1-e^{-jk_0\Delta r\sin\theta})(1-e^{jk_0\Delta r\sin\theta})}} \end{cases}. \tag{6}$$

These as a function of $k_0\Delta r\sin\theta$ are illustrated in Fig. 3. The singular values change periodically and have peaks and valleys where k_0 and θ satisfy the following relation with the integer number n.

$$k_0\Delta r\sin\theta = \frac{n\pi}{2} \quad , \tag{7}$$

This means that we can find the optimal source positions where the systems are required least effort to reproduce the in-phase and out-of-phase components of desired signals with particular frequencies.

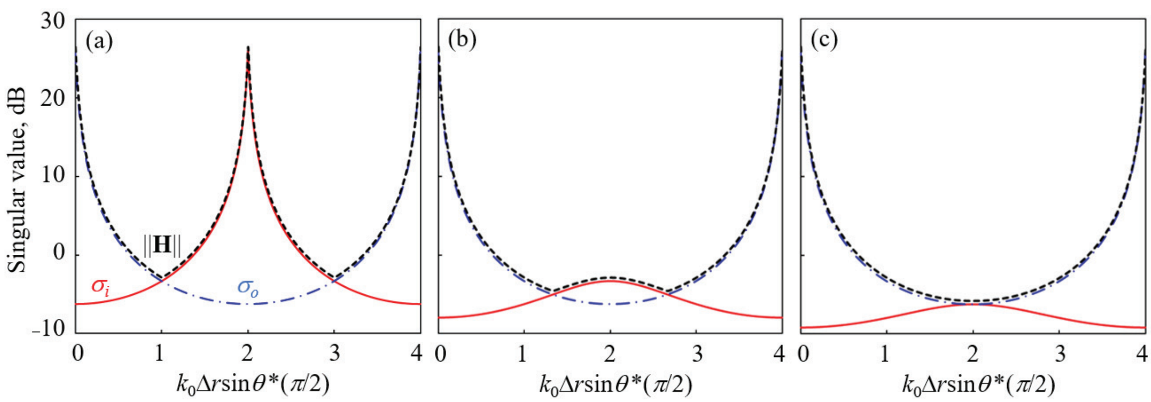

Figure 3 – Norm and singular values of the inverse matrix **H** as a function of $k_0\Delta r\sin\theta$: (a) $g = 0$ (two-channel), (b) $g = 1$ (three-channel) and (c) $g = \sqrt{2}$ (three-channel).

2.2 Difference between the two- and three-channel systems

The fundamental principle of the optimal source distribution (OPSODIS)[5] is based on the idea of a pair of conceptual monopole sources (left and right) whose azimuthal location varies continuously as a function of frequency. This idea came from focusing on the behavior of ||**H**|| in Fig. 3.

The two-channel OPSODIS system essentially utilizes the frequency-azimuth relationship where the two singular values σ_i and σ_o are balanced at n = 1 ($\Delta r\sin\theta$ = quarter wavelength) to minimize the value of ||**H**|| in Fig. 3 (a). Thus

$$k_0 \Delta r \sin\theta = \frac{\pi}{2}, (n=1). \tag{8}$$

In this case, the expression for the inverse filter matrix **H** becomes

$$\mathbf{H} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ -j & 1 \end{bmatrix}. \tag{9}$$

This implies that independent control of the two signals is nearly achieved just by addition of the desired signals with a $\pi/2$ relative phase shift between them. The conceptual monopole sources to satisfy the principle of the two-channel OPSODIS is illustrated in Fig. 4. In practice, a monopole source whose position varies continuously as a function of frequency is not easily available. However, it is possible to realize a practical system based on this principle by discretizing the source position as illustrated in the same figure (discretized OPSODIS).

On the other hand, in the case of the three-channel system,[7] the balance between the two singular values σ_o and σ_i can be changed independently by changing the relative sensitivity g because of having extra number of source than the number of receivers (ears) (i.e. the mathematically under-determined case). Since the useful frequency-azimuth relationship where the two singular values become equal to each other ($\sigma_o = \sigma_i$) varies from a third wavelength to two third wavelength in Eq. (7) as shown in Fig 3 (b) and (c), the three-channel OPSODIS can also utilize the frequency-azimuth relationship where the singular value of the out-of-phase components σ_o alone is minimized at n = 2 ($\Delta r\sin\theta$ = half wavelength):

$$k_0 \Delta r \sin\theta = \pi, \quad (n=2) \tag{10}$$

In this case, the following relationship should be satisfied to minimize the value of ||**H**||.

$$\max(\sigma_i) = \frac{1}{\sqrt{2g^2}} \le \min(\sigma_o) = \frac{1}{2}. \tag{11}$$

Therefore, it is most desirable to derive the inverse system without excessive amplification when the sensitivity of the center channel is greater than the left and right channel by the factor of √2, and that is shown in Fig. 3 (c). In this case, the expression for the inverse filter matrix **H** obtained from the minimum norm solution becomes

$$\mathbf{H} = \frac{1}{4}\begin{bmatrix} 1 & -1 \\ \sqrt{2}j & \sqrt{2}j \\ -1 & 1 \end{bmatrix}. \tag{12}$$

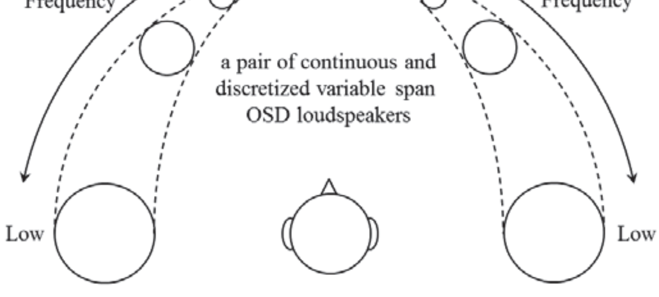

Figure 4 – Conceptual monopole sources to satisfy the principle of the optimal source distribution.

As a result, this implies the center channel source contributing to in-phase components σ_i should have twice the source strength to counter the two sources predominantly contribution to out-of-phase components σ_o. The conceptual monopole sources to satisfy the principle of the three-channel OPSODIS can be illustrated in a form of adding a third source in the median plane in Fig. 4.

Comparing the two- and three-channel systems, the point is that the shape of $||\mathbf{H}||$ around the ideal frequency-azimuth relationship are different from each other. In the case of the three-channel system, it forms a simple U shaped valley around $n = 2$ rather than a V shaped valley around $n = 1$ in the case of the two channel system. When the sensitivity g is varied, the shape of the valley changes slightly but it remains largely U shape with minor dents and humps. Various kinds of errors in realistic conditions coming from the discretization of the principle curve, the deviations in listening positions and individual differences in HRTFs etc. can be shown as the deviations from the bottom of those valleys to the left and right in Fig. 3. Therefore, the three-channel system is more robust to the errors than the two-channel system due to the fact that the valley is U shaped rather than V shaped.

3. CONSIDERATIONS FOR OFF-AXIS LISTENERS

3.1 Analysis

The principle of the OPSODIS can essentially provide the independent control of the binaural signals not only for the on-axis listener but also for multiple off-axis listeners. The augmented three-channel system is potentially superior to the two-channel case regarding the deviations in listening positions of the on-axis target listener because of its higher robustness as described in the previous section. In this section, we discuss whether the deviations in listening positions of off-axis listeners can also be enhanced by the three-channel system.

We now consider the inverse system to realize 1 (0 dB) at the right ear and 0 ($-\infty$ dB) at the left ear of the on-axis target listener in Fig. 2. The sound pressure p reproduced by two- and three-source systems including the inverse system are shown in the Cartesian coordinate system as

$$p = A\frac{e^{-jk_0r_1}}{r_1} + B\frac{e^{-jk_0r_2}}{r_2} + C\frac{e^{-jk_0r_3}}{r_3}, \tag{13}$$

where

$$\begin{cases} r_1^2 = (l\sin\theta - x)^2 + (l\cos\theta - y)^2 \\ r_2^2 = (-l\sin\theta - x)^2 + (l\cos\theta - y)^2 \\ r_3^2 = (0 - x)^2 + (l - y)^2 \end{cases}, \tag{14}$$

where x, y and Θ are in the following geometrical relationship:

$$\begin{cases} x = l\sin\Theta \\ y = l(1-\cos\Theta) \end{cases}. \tag{15}$$

In this model, we can discuss what interference state happens all in the region. In the case of the two-channel system, the amplitudes A, B and C can be expressed as

$$\begin{bmatrix} A & B & C \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{j}{2} & 0 \end{bmatrix}, \tag{16}$$

and in the case of the three-channel system, those can be expressed as

$$\begin{bmatrix} A & B & C \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & \frac{j}{2} \end{bmatrix}. \tag{17}$$

The source direction θ in Eq. (14) is related to the frequency satisfying Eq. (8) in the case of the two-channel system and Eq. (10) in the case of the three-channel system, respectively.

3.2 Numerical simulations and discussion

Numerical simulations are calculated by Eq. (13) and the sound pressure are normalized by that from a monopole source, throughout. Considering a realistic near-feild condition, the distance l between the soures and the target on-axis lisntner is set to be 2 m.

Contour plots of sound pressure level as a function of Θ and frequency are shown in Fig. 5. In the case of the two-channel system, lots of vertical stripes can be seen in the figure. This means that the additional controlled regions where perfect crosstalk cancellation across the entire frequency range above the low frequency limit is obtained for the off-axis listeners are periodically repeated at both sides of the on-axis listener, and which have turned out to be separating distance between ears Δr ($\approx$ 7.2° under this calculated condition).[6] The similar interference states happen in the case of the three-channel system. It is found that the vertical stripes in the three-channel system is much more distinguished than those in the two-channel system. This means that the higher robustness to listening positions of the three-channel system happens not only for the on-axis listener but also for the multiple off-axis listeners.

The physically maximum source span $2\theta = 180°$ gives the lowest frequency limit f_l associated with the principle of the OPSODIS as

$$f_l \geq \frac{nc_0}{4\Delta r}. \tag{18}$$

Not that a smaller value of the integer number n gives a lower f_l so that the two-channel system (n = 1) has an advantage rather than the three-channel system (n = 2) in terms of a lower frequency limit.

Contour plots of sound pressure level with respect to x-y plane of the listening space are shown in Fig. 6. The sound pressure is averaged from f_l to 20 kHz. In the case of the two-channel system, interference fringes of spreading radially are only observed. In the case of the three-channel system, however, in addition to those, interference fringes of spreading concentrically from the source in the median plane are also observed. They can affect the robustness against the back and forth displacement in the three-channel system and are subject of future investigation.

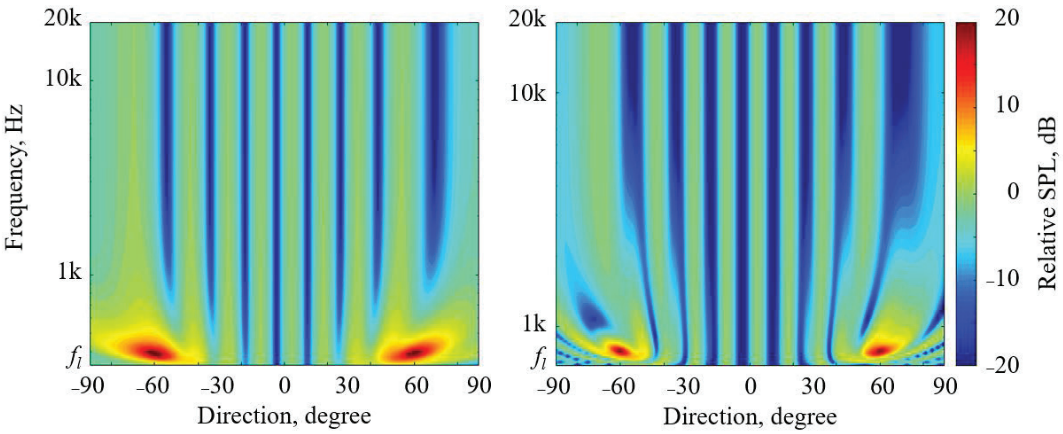

Figure 5 – Contour plots of sound pressure level of the two-channel (left) and the three-channel (right) systems as functions of receiver direction and frequency.

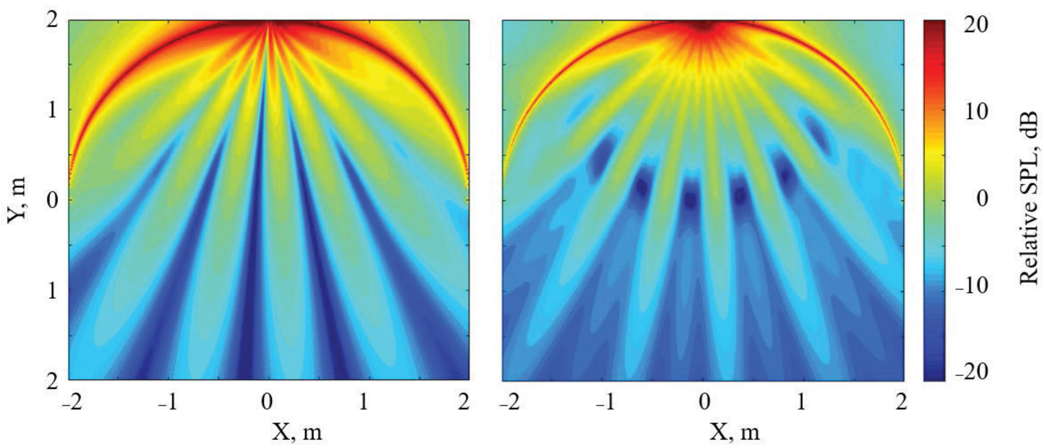

Figure 6 – Contour plots of sound pressure level of the two-channel (left) and the three-channel (right) systems in the X-Y plane.

4. EXPERIMENTS

4.1 Procedures

In order to validate the theory, an experiment using discretized OPSODIS loudspeakers with multiway-uint.[6, 8, 9] The left and right channels are split into six frequency bands according to the theory of the two- and three-channel systems, respectively. Binaural impulse responses at the ears of the Neumann KU100 dummy head were measured. The sound pressure was controlled to be 1 (0 dB) at the right ear as well as 0 ($-\infty$ dB) at the left ear by using the inverse system. Starting from the on-axis position, the dummy head was moved from the origin to around −600 and −1000 mm to the left/right-hand sides. To emulate a realistic listening environment, the experiment was carried out in a listening room with some degree of reflections.

Using the binaural impulse response data obtained in the experiment, the binaural crosstalk cancelation (*CTC*) was derived according to the following equation:

$$CTC = 10\log\frac{|F(\omega)_r|^2}{|F(\omega)_l|^2}, \tag{19}$$

where $F(\omega)_r$ and $F(\omega)_l$ are the Fourier transform of the binaural impulse response at the right and left ear, respectively.

4.2 Results and discussion

The acquired *CTC* data is best represented in a color plot form as a function of off-axis position and frequency. Figure 7 shows results from the two-channel (3-way) and three-channel (4-way) systems. Figure 8 shows a comparison between the two-channel and three-channel systems through discretized OPSODIS loudspeakers with 6-way.

The results in Fig. 7 and Fig. 8 (a) show data on the left/right of the central listening position when only measurements were taken on the left-hand side. The right-hand data in the figure is purely a symmetric re-plotting of the left-hand data, to ease the reading of the data.

From the result presented in both figures, it can be clearly seen that second and third listening positions exist in the system: approximately 350 and 700 mm off-axis, respectively. The *CTC* at the second and third listening positions are less complete along the frequency range than at the center. This is caused by the coarseness of discretization and finer discretization would help create a more consistent performance for the off-axis positions.

Comparing the results from the two- and three-channel systems, the vertical stripes in the three-channel system turned out to be more distinguish than those in the two-channel system at higher frequencies. In particular, the existence of the third listening position in the three-channel system is much clearer than that in the two-channel system.

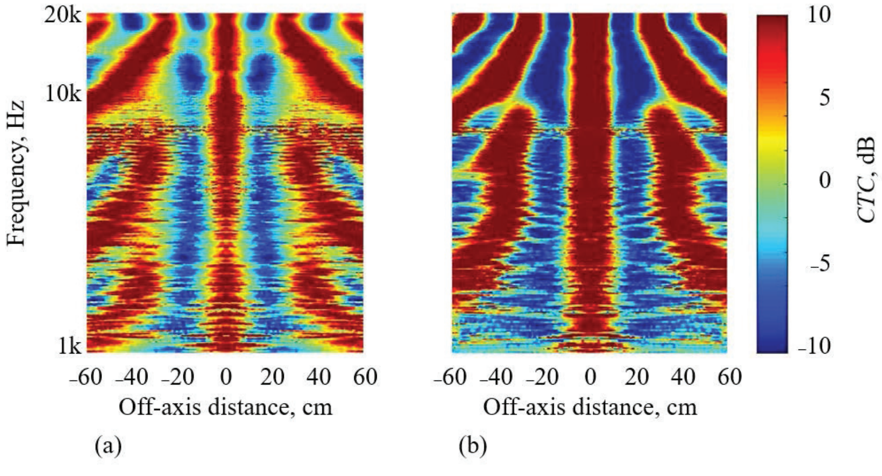

Figure 7 – *CTC* of the OPSODIS system as a function of off-axis position and frequency with the *CTC* amplitude on color axis: (a) two-channel system (3-way) and (b) three-channel system (4-way).

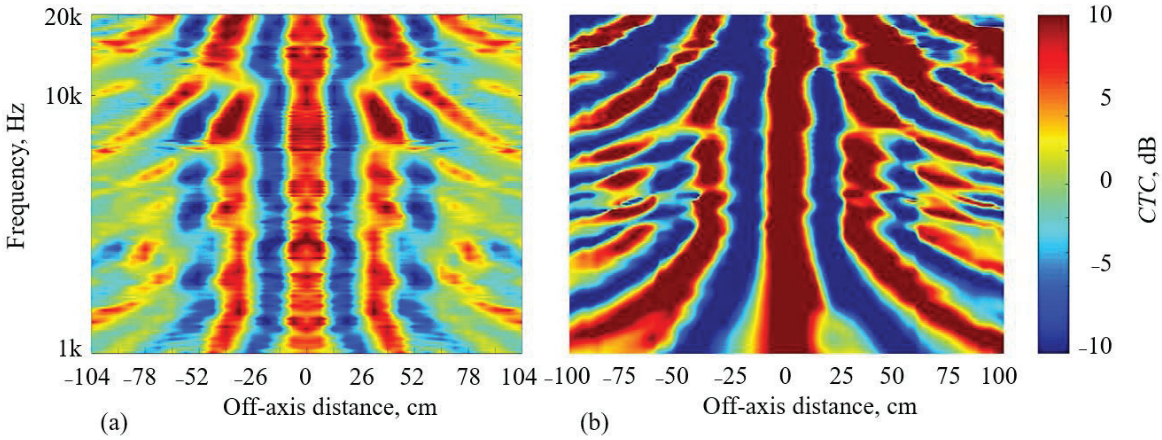

Figure 8 – *CTC* of the OPSODIS system as a function of off-axis position and frequency with the *CTC* amplitude on color axis: (a) two-channel system (6-way) and (b) three-channel system (6-way).

5. CONCLUSIONS

In order to realize virtual auditory space to multiple listeners using binaural synthesis over loudspeakers, we focused on the principle of the optimal source distribution. This principle utilizes the idea of a pair of conceptual monopole sources whose azimuthal location varies continuously as a function of frequency, and enable to provide the independent control of the binaural signals not only for the on-axis target listener but also multiple off-axis listeners. The principle can also be augmented to the three-channel system by adding another source in the median plane to the original left/right configuration of the two-channel system. The three-channel system is potentially superior to the two-channel one as for the robustness to the deviations in listening positions of the on-axis target listener. In this paper, whether the deviations in listening positions of the multiple off-axis listeners can also be enhanced by the three-channel system is discussed. Numerical simulations under free-field conditions and experiments in a listening room are carried out. It is proved that the three-channel system can also provide the additional controlled regions for the off-axis listeners, which are more robust to the deviations in listening positions than those given by the two-channel one especially at higher frequencies.

REFERENCES

1 M. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Int. conv. Record*, 7, 150-155 (1963).

2 M. Morimoto and Y. Ando, "On the simulation of sound localization, *J. Acoust. Soc. Jpn*. (E) 1, 167-174 (1980).

3 O. Kirkeby, P. A. Nelson and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *J. Acoust. Soc. Am.*, 104, 1973-1981 (1998).

4 Eric Hamdan, Andrew Allen, Raphael Melgar, and Peter Otto, "Bridging near and far acoustical fields: a hybrid systems approach to improved dimensionality in multi-listener spaces", 2016 AES International Conference on Sound Field Control, 2016. 7

5 T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *J. Acoust. Soc. Am.*, 112, 2786-2797 (2002).

6 M. Yairi, T. Takeuchi and K. Holland, "Binaural reproduction capability for multiple off-axis listeners based on the optimal source distribution principle," 2019 AES International Conference on Immersive and Interactive Audio, 2019. 3

7 T. Takeuchi and P. A. Nelson, "Extension of the Optimal Source Distribution for binaural sound reproduction", Acta Acustica united with Acustica, 94(6), 981-987, 2008

8 D. G. Morgan, T. Takeuchi and K. Holland. Off-axis cross-talk cancellation of 2-channel and 3- channel OPSODIS sound bars. Institute of Acoustics, Vol. 38. Pt 2. 2016.

9 L. A. T. Haines, T. Takeuchi and K. Holland. Investigating multiple off-axis listening positions of an OPSODIS sound bar. Institute of Acoustics, Vol. 39. Pt 1. 2017.