Meyer, Julie; Lokki, Tapio; Ahrens, Jens

Identification of Virtual Receiver Array Geometries that Minimize Audibility of Numerical Dispersion in Binaural Auralizations of Finite Difference Time Domain Simulations

Published in:
149th Audio Engineering Society Convention 2020

Published: 22/10/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Meyer, J., Lokki, T., & Ahrens, J. (2020). Identification of Virtual Receiver Array Geometries that Minimize Audibility of Numerical Dispersion in Binaural Auralizations of Finite Difference Time Domain Simulations. In *149th Audio Engineering Society Convention 2020* Curran Associates Inc.. https://www.aes.org/e-lib/browse.cfm?elib=20929

Audio Engineering Society

Convention Paper 10392

Presented at the 149th Convention
Online, 2020 October 27-30

Identification of Virtual Receiver Array Geometries that Minimize Audibility of Numerical Dispersion in Binaural Auralizations of Finite Difference Time Domain Simulations

Julie Meyer[1], Tapio Lokki[2], and Jens Ahrens[3]

[1]*Aalto Acoustics Lab, Department of Computer Science, Aalto University, Espoo, Finland*
[2]*Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[3]*Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden*

Correspondence should be addressed to Julie Meyer (julie.meyer@aalto.fi)

ABSTRACT

This paper presents a perceptual evaluation of numerical dispersion in free-field headphone-based head-tracked binaural auralizations of finite difference time domain (FDTD) simulations. The simulated pressure, captured by virtual volumetric receiver arrays, is used to perform a spherical harmonics decomposition of the sound field and generate binaural signals. These binaural signals are compared perceptually to dispersion error-free binaural signals in a listening experiment designed using a duo-trio paradigm. The aim of the present work is to identify the size and density of the receiver array minimizing the audibility of numerical dispersion in the generated binaural signals. The spherical harmonics order was chosen to be 12 for the spatial decomposition. The overall reconstruction error, defined as the absolute value of the difference between the dispersion error-free and FDTD-simulated left-ear magnitude spectrum, was used as an objective metric to measure the spectral differences between the two signals. The listening experiment results show that this error does not correlate with the discrimination rates of the subjects. These results therefore suggest that this error does not suffice to describe the perceptual aspects introduced by numerical dispersion in the free-field dynamic binaural auralizations presented in the listening experiment. The results also show that increasing the receiver density for a fixed array size does not necessarily render numerical dispersion inaudible in the auralizations. Five out of 27 volumetric arrays led to FDTD-simulated binaural auralizations indistinguishable from the dispersion error-free binaural auralizations.

1 Introduction

In the context of finite difference time domain (FDTD) simulations, there exist mainly three approaches to generate binaural signals. A recently developed spherical harmonic spatial encoding process [1], formulated and integrated directly in the FDTD scheme, could in principle allow for auralization of FDTD simulations. Another method consists of incorporating the morphology of the listener's head directly in the FDTD grid [2, 3, 4, 5]. The main disadvantage of this method is that highly detailed scans of the listener's head are to be faithfully represented in the FDTD grid, which implies running simulations at a high computational cost. Although the listener's head could be approximated with simple geometric models such as a rigid sphere [5, 6], these models do not include the pinna which has a significant influence on the head-related transfer function (HRTF) at frequencies where the wavelength is short compared to the size of the pinna [7]. A third method consists of embedding an array of receivers in the FDTD grid and combining the

decomposed grid data with free-field HRTFs [8, 9]. However this approach, formulated in the spherical harmonics domain, poses the question of the array performance in the spatial decomposition. In particular, it is well known that the number of receivers as well as the size of the array will limit the accuracy of the spatial decomposition [10]. Nevertheless, it is possible to take advantage of the FDTD spatial grid by filling a portion of volume with receivers, and thus construct an array free of scattering containing a large number of receivers. Previous studies focused on assessing the performance of different volumetric spherical arrays fitted to FDTD spatial grids in terms of numerical robustness and spatial aliasing [8, 9] considering a single plane-wave impinging the array. However, to the present authors' knowledge, no study investigated how numerical dispersion propagates through the spatial decomposition for different arrays varying in size and density of receivers in a perceptual point of view. Note that numerical dispersion is a non-physical phenomenon whereas the spatial decomposition assumes physical signals. The present work aims at filling this gap by identifying the volumetric array size and density of receivers that produces binaural signals with inaudible numerical dispersion.

Another advantage of using the third aforementioned approach is that head rotations in the azimuthal plane can be calculated with a Wigner-D function which simplifies into a single complex exponential as a function of the azimuth angle [11, 12]. This is particularly useful when a large set of head rotations is desired (e.g. for dynamic reproduction with head tracking). Contrariwise, the second described approach could be cumbersome as it would necessitate to run several computationally demanding simulations to calculate the HRTFs for different head orientations.

The remaining sections of this paper are organized as follows. Section 2 gives a brief overview of the numerical dispersion and the binaural processing formulated in the spherical harmonics domain. Section 3 outlines the analysis performed prior to performing the perceptual evaluation. Section 4 describes the perceptual evaluation, and Section 5 presents the obtained results before concluding on the paper in Section 6.

2 Theoretical Background

2.1 Numerical Dispersion

Numerical dispersion is one of the sources of error in FDTD simulations [13, 14]. It is due to the approximation of the second-order partial derivatives governing the acoustic wave equation (Eq. (1)) by finite-difference operators.

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right), \qquad (1)$$

where p is the acoustic pressure and c is the speed of sound, taken to be 344 m/s throughout this work. Such discretization of the wave equation in time and space, using the standard rectilinear (SRL) scheme, leads to

$$\delta_t^2 p_{l,m,i}^n = \lambda^2 (\delta_x^2 + \delta_y^2 + \delta_z^2) p_{l,m,i}^n, \qquad (2)$$

where $\lambda = cT/X$ is the Courant number, T is the time step and X is the grid spacing. $p_{l,m,i}^n \equiv p(x,y,z,t)|_{x=lX,y=mX,z=iX,t=nT}$ is the update variable, n denotes the time index and l, m, and i are the spatial indices in the x-, y-, and z-direction, respectively. δ_t^2, δ_x^2, δ_y^2, δ_z^2 are the second-order derivative centered finite-difference operators as given in [15], for example.

The pressure for a plane-wave traveling in the negative x-direction (as in Fig. 2) can be expressed as Eq. (3) in the continuous time-space domain, which translates into Eq. (4) in the discrete time-space domain.

$$p(x,t) = Ae^{i(\omega t + k_x x)} \qquad (3)$$

$$p_{l,m,i} = Ae^{i(\omega nT + \hat{k}_\mathrm{x} lX)}, \qquad (4)$$

where A denotes the amplitude, ω is the angular frequency, k is continuous-time wavenumber, and $\hat{k}_\mathrm{x}$=$\hat{k}\cos\alpha\cos\beta$ is the numerical wavenumber component. α and β denote the azimuth and elevation angles of the plane-wave propagation direction, respectively.

Assuming plane-wave solutions for the wave equation and considering only the axial direction propagation as in Eq. (4), the dispersion relation is obtained from Eq. (2) and can be expressed as follows [16]

$$\sin^2\left(\frac{\omega T}{2}\right) = \lambda^2 \sin^2\left(\frac{\hat{k}_\mathrm{x} X}{2}\right). \qquad (5)$$

Solving the dispersion relation Eq. (5) for $\hat{k}_\mathrm{x}$, the relative phase velocity v_p, which is defined as the ratio between the numerical wave speed $\hat{c}$ and the real sound wave propagation velocity c, can be expressed as in Eq. (6). It is commonly used as a means to quantify the extend of dispersion error [17, 15].

$$v_p = \frac{\hat{c}}{c} = \frac{\omega}{c\hat{k}} = \frac{\omega T}{2\lambda \arcsin\left(\frac{1}{\lambda}\sin\left(\frac{\omega T}{2}\right)\right)}. \qquad (6)$$

2.2 Binaural Processing

The pressure signals at the FDTD grid nodes $p(\mathbf{r},t)$ can be approximated in the frequency domain by [9]

$$p(\mathbf{r},\omega) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \breve{S}_n^m(\omega) \underbrace{j_n(kr) Y_n^m(\theta,\phi)}_{\mathbf{B}}, \quad (7)$$

where N is the maximum spherical harmonics order, $\breve{S}_n^m(\omega)$ are the spherical harmonics expansion coefficients for a plane-wave, $j_\mathrm{n}(.)$ is the nth order spherical Bessel function, and $Y_n^m(\theta,\phi)$ are the spherical harmonics basis functions of degree n and order m as defined in [9] with the colatitude θ and the azimuth ϕ angles of each of the array receivers with respect to the center of the array. A spherical harmonics order of $N = 12$ was chosen as it was shown in [12, 18] that $N \geq 8$ leads to nearly perfectly authentic auralization. That way, an audible impact of the auralization procedure itself could be excluded in principle.

The spherical harmonics expansion coefficients for a plane-wave with propagation direction (colatitude θ_{pw}, azimuth ϕ_{pw}) are given by [10]

$$\breve{S}_n^m(\omega) = 4\pi i^n Y_n^m(\theta_{pw},\phi_{pw})^*, \quad (8)$$

where $Y_n^m(\theta_{pw},\phi_{pw})^*$ denotes the complex conjugate of $Y_n^m(\theta_{pw},\phi_{pw})$.

From the pressure signals captured at the receiver array nodes, it is possible to numerically retrieve $\breve{S}_n^m(\omega)$ by calculating the Moore–Penrose inverse of the matrix $\mathbf{B}$ from Eq. (7) if $p(\mathbf{r},\omega)$ is known inside a densely sampled volume [9]. The numerical operation of retrieving the coefficients $\breve{S}_n^m(\omega)$ from the captured pressure at any point of the sound field will be hereafter referred as *spatial decomposition*. Soft-limited radial filters, as defined in [9], with a maximum amplification of 60 dB were also used in the *spatial decomposition* since radial filtering was shown to further enhance the numerical robustness of volumetric arrays [8, 9]. Finally, combining the obtained $\breve{S}_n^m(\omega)$ with free-field HRTFs transformed in the spherical harmonics domain as in Eqs. (9) and (10), the left- $p(\omega,\alpha)^l$ and right-ear $p(\omega,\alpha)^r$ pressure signals in the frequency domain can be obtained [9].

$$p(\omega,\alpha)^l = \sum_{n=0}^{N} \sum_{m=-n}^{n} (-1)^m \breve{S}_n^{-m}(\omega)^* H_{nm}^l(\omega) e^{-im\alpha} \quad (9)$$

$$p(\omega,\alpha)^r = \sum_{n=0}^{N} \sum_{m=-n}^{n} (-1)^m \breve{S}_n^{-m}(\omega)^* H_{nm}^r(\omega) e^{-im\alpha}, \quad (10)$$

where α (in radians) is the azimuth angle of the head orientation, which was evaluated from 0° to 359° with 1° increments. $H_{nm}^l(\omega)$ and $H_{nm}^r(\omega)$ are the spherical harmonics expansion coefficients of the free-field HRTFs for the left- and right-ear, respectively, taken from the publicly available KU-100 database from the University of Cologne [19].

3 Preliminary Analysis

3.1 Volumetric Arrays

The performance of several volumetric receiver arrays in the *spatial decomposition* was first evaluated independently of the presence of numerical dispersion. For that aim, the coefficients $\breve{S}_n^m(\omega)$ given in Eq. (8) were inserted into Eqs. (9) and (10) to first compute reference binaural signals for a virtual plane wave. These reference binaural signals were then compared with binaural signals computed using a plane-wave formulation (as in Eq. (3)) at the receiver array points undergoing the *spatial decomposition*. The comparison was performed for full volumetric cubical (FVC) and spherical (FVS) arrays of different size (side length and diameter varying from 0.1 m to 1 m, respectively) and receiver densities. The spacing between the receivers was equal along the Cartesian dimensions and in the range from 4.8 mm to 125 mm. Both time and frequency domain left-ear signals were inspected visually as a first sanity check. Furthermore, the comparison was done by computing the overall reconstruction error $\varepsilon_{prelim.}$ defined in Eq. (11) representing the absolute value of the difference between the reference and the plane-wave-formulated left-ear magnitude spectrum averaged across the frequency bandwidth limited up to 12 kHz.

$$\varepsilon_{prelim.} = \frac{1}{K} \sum_{k=1}^{K} \left| 20\log_{10}|p_{ref,k}^l| - 20\log_{10}|p_{pw,k}^l| \right|, \quad (11)$$

where K is the total number of frequency bins, $p_{ref,k}^l$ and $p_{pw,k}^l$ are the k^{th} frequency bin of the reference and the plane-wave-formulated left-ear pressure spectrum, respectively.

Additionally, one of the present paper's authors took part in an informal listening session to identify if any differences could be perceived between the reference binaural signals and these generated from the decomposed receiver array data. The informal listening session took place in a quiet environment and consisted in listening through headphones (Sennheiser HD 600) to the binaural signals convolved with an anechoic recording of castanets (duration 7 s), the spectrum of which had most of

its energy below 12 kHz. Fig. 1 shows the overall reconstruction error $\varepsilon_{prelim.}$ for the different full volumetric arrays tested.

The results from these comparisons show that no audible difference was identified when the overall reconstruction error $\varepsilon_{prelim.}$ was below or equal to 2.0 dB for both FVC and FVS arrays. These results were used as a basis to choose which array size and receiver array grid spacing should be investigated in the perceptual evaluation. In particular, it was chosen to focus on parameter sets where no audible difference was perceived for neither of the two arrays. The number of volumetric arrays to investigate was further reduced by only considering volumetric cubical (VC) and spherical (VS) arrays of 0.1 m in size. This ensured that the maximum receiver array density could be reached in the FDTD simulations with a limited, but high, number of receivers.

3.2 Array Nodes Fitted to the FDTD Grid

To avoid spatial interpolation and therefore be compatible with the FDTD spatial period ($X = 3.8$ mm) and given the fact that the number of receivers constituting the receiver array was changed so as to vary the receiver density, the size of the receiver arrays fitted to the FDTD grid varied within the range [-31.6,+14] mm from the target size of 100 mm. Similarly, the distance between the center of the arrays and the source deviated by 0.6 mm from the target distance of 2900 mm. In order to ensure that these deviations in size did not have an influence on the computed binaural signals, the same overall reconstruction error $\varepsilon_{prelim.}$ as previously described in Section 3 was computed using the receiver arrays reported in Table 1. The results are reported in Table 2 and show that the overall reconstruction error $\varepsilon_{prelim.}$ averaged across VC arrays was 0.9 dB (maximum 1.2 dB) and 1.3 dB (maximum 1.9 dB) across the VS arrays. Additionally, the same informal listening session as previously described in Section 3.1 was performed. During this informal listening session, no audible difference could be perceived between the reference binaural signals computed with the theoretical expansion coefficients defined in Eq. (8) and these generated from the *spatial decomposition* of the receiver array data expressed with a plane-wave formulation.

4 Perceptual Evaluation

While the previous section discussed the accuracy of the *spatial decomposition* yielded from several volumetric receiver arrays in the absence of numerical dispersion in the pressure signals, the present section deals with the comparison of binaural signals computed from dispersion error-free impulse responses, generated using the same plane-wave formulation as mentioned in previous sections, with binaural signals computed from FDTD-simulated impulse responses (which contain dispersion).

Table 1: Characteristics of the VC and VS arrays fitted on the FDTD spatial grid. The symbol - indicates that the receiver array could not be fitted on the FDTD grid.

#	Array size (cm)	Array grid spacing (mm)	Number of receivers	
			VC	VS
1	10.26	11.4	1000	-
2	11.40	11.4	1331	515
3	8.36	7.6	1728	-
4	9.12	7.6	2197	925
5	9.88	7.6	2744	-
6	10.64	7.6	3375	1419
7	11.40	7.6	4096	-
8	12.16	7.6	4913	2109
9	12.92	7.6	5832	-
10	6.84	3.8	6859	3071
11	7.22	3.8	8000	-
12	7.60	3.8	9261	4169
13	7.98	3.8	10648	-
14	8.36	3.8	12167	5575
15	8.74	3.8	13824	-
16	9.12	3.8	15625	7153
17	9.50	3.8	17576	-
18	9.88	3.8	19683	9171

4.1 Stimuli

Impulse responses between the receiver array nodes and the source, located at an axial distance of 2.9 m from the center of the array, were simulated in a free-field using an FDTD solver for room acoustics [20] running with three GPUs (Tesla P100). The 3D SRL scheme was used and the choice of the propagation direction for the direct path was made such that the worst-case propagation direction of the scheme (axial direction) was studied, thus ensuring to evaluate the higher bound of dispersion error content. The temporal sample rate of the simulations was adjusted such that the relative phase velocity, defined in Eq. (6), was 2 % at a cutoff frequency chosen to be 12 kHz. This corresponds to a temporal sample rate of 156796 Hz, and a spatial grid resolution of X = 3.8 mm at the stability limit of the Courant number of the SRL scheme, that is $\lambda = 1/\sqrt{3}$. Since a closed volume is required to run

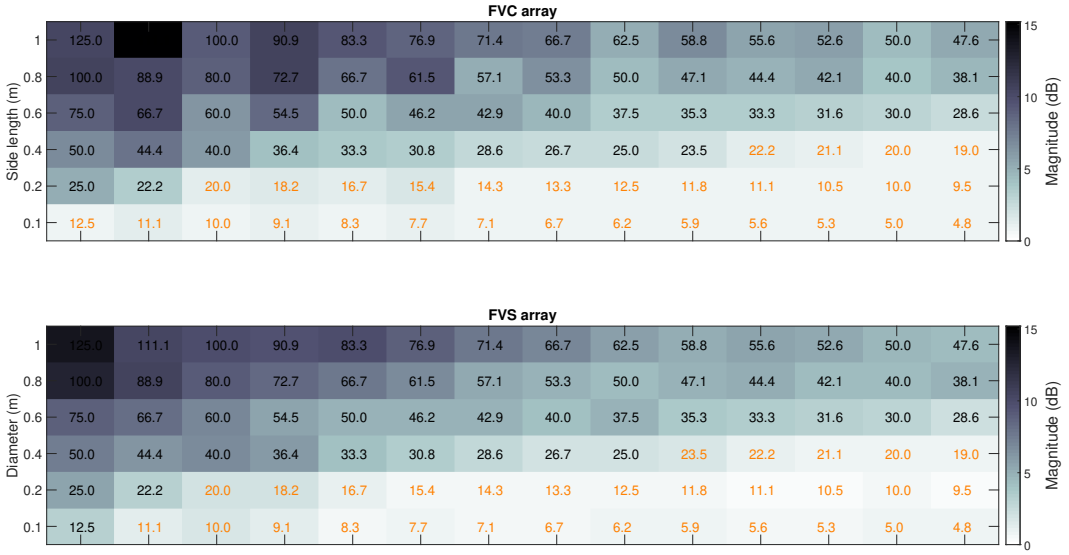

Fig. 1: Overall reconstruction error $\varepsilon_{prelim.}$. The numbers on the plot represent the spacing between the receivers in the array (in mm). The colored orange numbers represent cases where no audible difference was perceived in the informal listening.

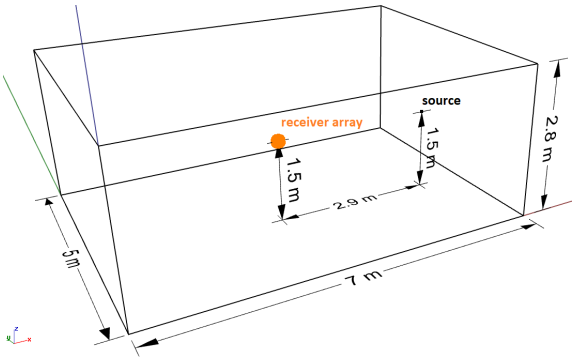

Fig. 2: 3D model of the box and its dimensions essentially illustrating the simulation setup. The FDTD simulation was stopped before any reflections from the boundaries reached the receivers.

FDTD simulations, a box of dimension 7 m $\times$ 5 m $\times$ 2.8 m was modeled with rigid boundaries. However, the simulations were run for a duration of 11.5 ms to ensure that the incoming sound wave passed through all the arrays' receivers while excluding the capture of the first order reflections from the box surfaces, so that the impulse responses were simulated in a free-field. A soft source with a discrete delta sequence was used as the source excitation signal in the simulations. The 3D model of the box as well as the simulation setup are shown in Fig. 2.

The FDTD-simulated impulse responses were then low-pass filtered using a 200th-order finite impulse response filter with a cutoff frequency of 12 kHz using a Chebyshev window (with 100 dB of relative side lobe attenuation). In order to fairly compare the dispersion error-free binaural signals with the FDTD-simulated binaural signals, the same low-pass filter was applied to the impulse responses generated using the plane-wave formulation. Prior to performing the *spatial decomposition*, the impulse responses were resampled at a sampling rate of 48 kHz to match with the sample rate of the HRTF dataset used after the *spatial decomposition*. Each FDTD-simulated impulse response was also normalized in amplitude by a ratio of the root mean square value of the signal by the dispersion error-free impulse response.

After the *spatial decomposition* and convolution with the free-field HRTFs, appropriate minimum-phase headphone compensation filters were applied to both the dispersion error-free and FDTD-simulated binaural signals for all head rotation angles. Resampling of the normalized binaural signals at a sample rate of 44.1 kHz was applied to match with the castanet excerpt sample rate used in the auralizations. The volume of the stimuli was fixed such that the peak sound pressure level of the unprocessed anechoic recording of the castanet was 70 dB(A). This sound pressure level was measured by placing the microphone of the sound level meter (RadioShack 33-2055 Digital Sound Level Meter) at the entrance of the left-ear cushion and manually adjusting the output level to 70 dB(A) using a fast time weighting.

Similarly to the preliminary analysis and prior to performing the perceptual evaluation, the overall

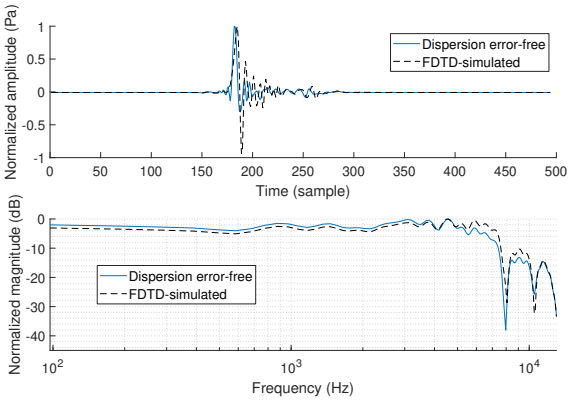

Fig. 3: Dispersion error-free and FDTD-simulated left-ear signals for ($\theta = \frac{\pi}{2}$, $\phi = 0$) generated using VC #3 in the time domain (top) and in the frequency domain (bottom). Note that the time sample for the direct sound does not correspond to a 2.9 m propagation distance. That is because a circular shift was applied to the signal.

reconstruction error $\varepsilon_{percept.eval.}$ was calculated for the arrays reported in Table 1 so as to provide an objective measure of difference between the dispersion error-free and the FDTD-simulated binaural signals.

$$\varepsilon_{percept.eval.} = \frac{1}{K}\sum_{k=1}^{K}\Big|20\log_{10}|p^{l}_{pw,k}| - 20\log_{10}|p^{l}_{FDTD,k}|\Big|, \quad (12)$$

where K is the total number of frequency bins, $p^{l}_{ref,k}$ and $p^{l}_{FDTD,k}$ are the k^{th} frequency bin of the plane-wave-formulated and the FDTD-simulated left-ear pressure spectrum, respectively.

Fig. 3 shows an example of a plane-wave-formulated and an FDTD-simulated left-ear signal in both time and frequency domains, generated by decomposing the data from the volumetric array VC #3, prior to convolution with the castanet excerpt. It can be seen that the time domain response of the simulated left-ear signal is more spread over the time than the dispersion error-free signal which is expected and in line with results from previous studies on numerical dispersion [21, 16]. As for the frequency domain response of the two signals, larger differences can be observed in the higher frequencies, which is in line with the fact that numerical dispersion increases with frequency. Note that the magnitude of each left-ear pressure spectrum signal was normalized by its maximum value in Fig. 3.

Table 2: Overall reconstruction error $\varepsilon_{prelim.}$ and $\varepsilon_{percept.eval.}$ of the left-ear signals from the preliminary analysis and the perceptual evaluation, respectively. The symbol - indicates that the receiver array could not be fitted on the FDTD grid.

#	$\varepsilon_{prelim.}$ (dB)		$\varepsilon_{percept.eval.}$ (dB)	
	VC	VS	VC	VS
1	0.9	-	1.2	-
2	1.0	0.9	0.8	1.3
3	0.8	-	2.3	-
4	0.8	1.5	1.8	1.9
5	0.9	-	1.5	-
6	0.9	1.5	1.1	1.2
7	0.9	-	0.9	-
8	0.8	1.4	0.6	0.5
9	0.8	-	0.4	-
10	1.2	1.5	4.3	3.4
11	1.0	-	3.7	-
12	1.0	1.9	3.3	4.1
13	0.9	-	2.9	-
14	0.9	1.3	2.7	2.2
15	0.9	-	2.3	-
16	0.9	1.0	2.0	1.4
17	1.0	-	1.8	-
18	1.0	0.9	1.7	1.3

4.2 Experimental Setup

The listening experiment took place in a quiet room with a measured background noise level below 28 dB using a low-noise measuring system (G.R.A.S. Type 40 HF 1"). The stimuli were presented over circumaural open headphones (Sennheiser HD 600) using an audio interface (Motu UltraLite mk3 Hybrid) connected to a laptop computer running the SoundScape Renderer software (SSR) [22, 23] in the binaural room synthesis mode. The listeners' head rotations in the azimuthal plane were tracked using a tracking system (NaturalPoint Inc. OptiTrack V100 and Tracking Tools software, version 2.5.3, 2012) composed of six infrared cameras surrounding the listening space. The Tracking Tools software was running on a separate computer and broadcasting the data from the head tracking using a Virtual-Reality Peripheral Network streaming engine. The overall latency of the experimental setup was of less than 7 ms decomposed in the following way: 2 ms for the head tracking system latency, 4 ms for the network latency (average round-trip times measured by running a ping test between the two computers), 0.73 ms for the SSR latency (2 frames of buffer with a buffer frame size set to 16

samples at a sample rate of 44.1 kHz).

Prior to participating in the listening experiment, subjects were provided with written instructions describing the task to perform. The instructions also contained a question to be addressed at the end of the experiment asking the subjects to give the attributes they focused on to perform the task. After reading the instructions, the subjects completed a training phase to familiarize themselves with the task as well as with the user interface, which was designed using Matlab. The training phase consisted of three randomly chosen triads to evaluate during which feedback for correct/incorrect answer was provided.

Nineteen subjects (18 males), excluding the authors, participated in the listening experiment. The participants aged from 24 to 41 years (average = 31 years, standard deviation = 6 years). Eighteen subjects had self-reported normal hearing (no audiogram was conducted to confirm it). One participant reported having tinnitus but also said to be accustomed to it to such an extent that it was not problematic to perform the discrimination task, thus that participant's data was not excluded from the analysis. All subjects except one had previous experience in participating in listening experiments. The listening experiment lasted 19 minutes in average (standard deviation = 6 minutes).

4.2.1 Listening Test Method

The duo-trio paradigm [24] was adopted as it is useful for determining whether a sensory difference exists between two samples when no attribute is specified to the subjects. The experimental design of such paradigm in the context of audio consists in presenting simultaneously three sound samples to the subjects, one of which is labelled as the reference, and two other sound samples, one of which is also the reference and the other differs from the reference. The task is to identify which of the two other sound samples corresponds to the labelled reference. The listening experiment contained a total of 57 triads to evaluate. Each volumetric array reported in Table 1 was evaluated twice (27×2 triads), and three randomly chosen volumetric arrays were evaluated once in the training phase (3×1 triads). The reference always consisted of the binaural signals computed from dispersion error-free impulse responses whereas the other sound sample was the binaural signals containing numerical dispersion and generated from the same volumetric array as the reference. The presentation order of the conditions was randomized across subjects and conditions to ensure independence of the subjects' responses and to avoid learning effects.

4.2.2 Statistical Analysis

In order to determine if the subjects could discriminate between the plane-wave-formulated and the FDTD-simulated dynamic binaural auralizations, a threshold for significance of correct answers from the listening experiment results was established based on a one-tailed z-test. The significance level denoted α was chosen to be of 1 %, thus fixing the critical z value at 2.33 for a one-tailed test. Since each subject had to evaluate each volumetric array twice, the probability of correct decision by chance was $(1/2)^2$ because only two successive correct answers for each subject were taken into account. The z-score can be calculated from the following formula [25]

$$z = \frac{X - np - 0.5}{\sqrt{npq}}, \tag{13}$$

where X is the minimum number of correct responses to reach statistical significance, $n = 19$ is number of judges, p is probability of correct decision by chance = $(1/2)^2$, $q = 1 - p$. Solving Eq. (13) for X sets the threshold for significance of the results at $X = 10$ correct answers (i.e. $X = (10/19) \times 100 = 52.6\%$ of correct answers) for the chosen significance level α of 1 %.

5 Results

As can be seen from Fig. 4, the dynamic binaural auralizations containing numerical dispersion could not be distinguished from those that did not contain the dispersion error for only five (VC #11, #16, #17, #18, VS #18) out of 27 volumetric arrays. The overall reconstruction error $\varepsilon_{percept.exp}$ for the arrays for which subjects could not discriminate the reference is below or equal to 2.0 dB, except for VC #11 where $\varepsilon_{percept.exp}$ = 3.7 dB. Besides these five arrays, 13 other volumetric arrays have $\varepsilon_{percept.exp}$ values below 2.0 dB. Since the overall reconstruction error and the perceptual results from the listening experiment are inconsistent, it can be concluded that this objective metric does not suffice to depict the perceptual differences between the plane-wave-formulated and the FDTD-simulated binaural auralizations.

It can be observed from the results that increasing the number of receivers for arrays of the same size led to discrimination rates dropping below the threshold for significance. That was observed for arrays VC #4 and VC #5, with respective array size 9.12 cm and 9.88 cm, which were perceived as different from the reference in the listening experiment when the array grid spacing was 7.6 mm. Reducing the array grid spacing by a factor of two, thus increasing the receiver array density for VC #4 and

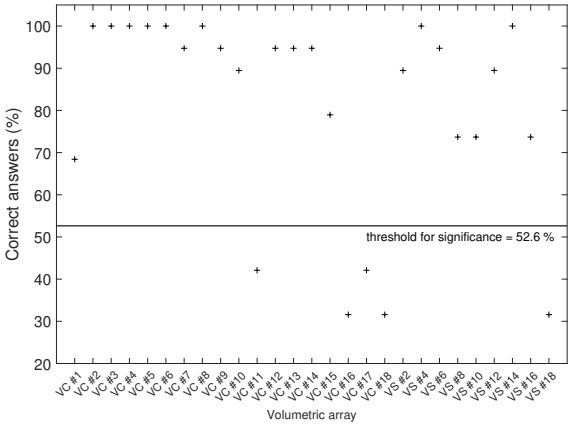

Fig. 4: Discrimination rates from the listening experiment.

VC #5 arrays led to arrays VC #16 and VC #18 which could not be distinguished from the reference in the listening experiment. However, unlike the two previously mentioned cases, both VS #4 and VS #16, corresponding to an array grid spacing of 7.6 mm and 3.8 mm, respectively, were perceived as different than the reference. It can therefore be concluded that increasing the receiver density for the same array size does not necessarily render numerical dispersion inaudible in the auralizations. Since increasing the number of receivers has shown to improve the performance of volumetric arrays [10], this conclusion is limited to the arrays reported in Table 1. It is likely that further increasing the total number of receivers for VS #16, e.g. by fitting the array nodes onto another FDTD spatial grid, would lead to different results.

Subjects also commented on how often they used the head rotations. The majority indicated that they rotated their head a few times during the listening experiment while the remaining subjects indicated that they never did. Since using the head rotations was encouraged in the written instructions, these comments suggest that rotating the head in the azimuthal plane did not help the participants to perform the discrimination task. As for the attributes they used to perform the task, the subjects indicated coloration ($\times$ 9), loudness ($\times$ 5), source width ($\times$ 4), localization ($\times$ 3), reverberation ($\times$ 2).

5.1 Discussion

The condition number of the matrix **B** from Eq. (7) as well as the aliasing error were computed in order to understand where the difference between the results from the perceptual evaluation and the overall reconstruction error values comes from. The interested reader is referred to [11, 9] for the detailed definition of these two parameters. These parameters are shown in Figs. 5 and 6, in which the thicker colored lines correspond to the volumetric arrays for which the FDTD-simulated binaural auralizations could not be distinguished from the reference binaural auralizations in the listening experiment. As can be seen on Fig. 5, the condition number of the matrix **B** is decreasing as a function of frequency, which is expected. The condition number $\kappa(\mathbf{B})$ seems comparable for all volumetric arrays, especially at frequencies above 1 kHz. Below 1 kHz, more differences can be observed and in the case of the volumetric arrays that led to inaudible difference between the reference and the FDTD-simulated binaural auralizations, the condition numbers are not the smallest. Whether $\kappa(\mathbf{B})$ plays a big role in the observed listening experiment results is therefore inconclusive and further investigation would be required. However, as can be seen in Fig. 6, the aliasing errors ε are relatively different across volumetric arrays. Moreover, the values corresponding to the volumetric arrays for which the FDTD-simulated binaural auralizations could not be distinguished from the reference binaural auralizations seem concentrated in the bottom region of the graph (VC # 11, # 16, # 17, # 18, and VS # 18, cf. solid colored lines in Fig. 6), which shows consistency with the listening test results. However, quantitative conclusions cannot be drawn from this parameter as there are volumetric arrays with a lower aliasing error than VC # 11, # 16, # 17, # 18 that could be distinguished from the reference.

It is worth pointing out that the listening test method that was used is very critical in that it allows to detect smallest differences as the task consists in finding the sample that is exactly the same as the reference. This is reflected in the high discrimination rates observed for most of the volumetric arrays tested, since even a tiny difference can be detected with the duo-trio paradigm. That being said, it is likely that using another sound sample than the castanet excerpt for the auralizations as well as a different listening test method, the results would be different.

6 Conclusion

This paper presented a listening experiment aiming at identifying which volumetric array size and receiver density minimizes the audibility of numerical dispersion in free-field headphone-based head-tracked binaural auralizations of FDTD simulations. First, dispersion error-free as well as FDTD-simulated binaural signals were generated by performing a *spatial decomposition* of the data captured

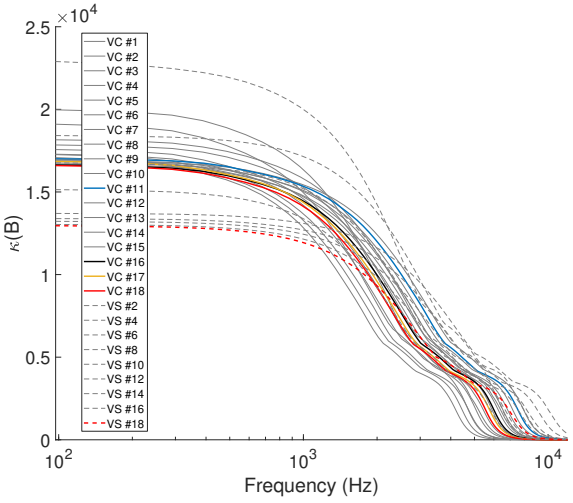

Fig. 5: Condition number $\kappa(\mathbf{B})$ of the matrix $\mathbf{B}$ evaluating the numerical robustness of the array design which limits the accuracy of the *spatial decomposition* at low frequencies.

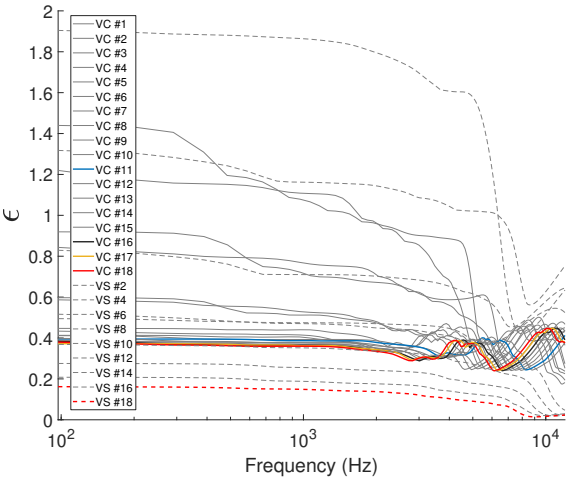

Fig. 6: Aliasing error ϵ quantifying the amount of spatial aliasing which limits the accuracy of the *spatial decomposition* at high frequencies. Colored lines denote configurations that were indistinguishable from the dispersion-free reference.

by different volumetric cubical (VC) and spherical (VS) arrays varying in size and receiver density. Second, the binaural signals were convolved with a castanet sound sample and incorporated in a listening experiment designed with a duo-trio paradigm, in which the reference consisted of the dispersion error-free auralizations. Head rotations in the azimuthal plane were tracked to allow for dynamic reproduction. The listening experiment results show that the discrimination rates are not consistent with the overall reconstruction error used as an objective metric to compare the reference and FDTD-simulated binaural signals. The results also demonstrate that increasing the number of receivers for a fixed array size does not necessarily reduce the audibility of the differences between the reference and the FDTD-simulated binaural auralizations.

However, for five out of 27 volumetric arrays, the FDTD-simulated dynamic binaural auralizations were not distinguishable from the reference. These cases could serve as a basis to perform other dynamic binaural auralizations, e.g. where room reflections could be included.

7 Acknowledgments

The authors would like to thank Christoph Hold for integrating the head tracking system used in the present work in the SSR software. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 721536.

References

[1] Bilbao, S., Politis, A., and Hamilton, B., "Local Time-Domain Spherical Harmonic Spatial Encoding for Wave-Based Acoustic Simulation," *IEEE Signal Processing Letters*, 26(4), pp. 617–621, 2019.

[2] Xiao, T. and Huo Liu, Q., "Finite difference computation of head-related transfer function for human hearing," *The Journal of the Acoustical Society of America*, 113(5), pp. 2434–2441, 2003.

[3] Mokhtari, P., Takemoto, H., Nishimura, R., and Kato, H., "Computer simulation of HRTFs for personalization of 3D audio," in *Universal Communication, 2008. ISUC'08. Second International Symposium on*, pp. 435–440, IEEE, 2008.

[4] Webb, C. J. and Bilbao, S., "Binaural simulations using audio rate FDTD schemes and CUDA," in *Proc. 15th Int. Conf. Digital Audio Effects (DAFx-12)*, 2012.

[5] Sheaffer, J., Webb, C., and Fazenda, B. M., "Modelling binaural receivers in finite difference simulation of room acoustics," in *Proceedings of Meetings on Acoustics ICA2013*, volume 19, p. 015098, ASA, 2013.

[6] Schymura, C., Winter, F., Kolossa, D., and Spors, S., "Binaural sound source localisation and tracking using a dynamic spherical head model," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[7] Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z., "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, 112(5), pp. 2053–2064, 2002.

[8] Sheaffer, J., van Walstijn, M., Rafaely, B., and Kowalczyk, K., "A spherical array approach for simulation of binaural impulse responses using the finite difference time domain method," in *Proc. Forum Acusticum*, 2014.

[9] Sheaffer, J., Van Walstijn, M., Rafaely, B., and Kowalczyk, K., "Binaural reproduction of finite difference simulations using spherical array processing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(12), pp. 2125–2135, 2015.

[10] Rafaely, B., *Fundamentals of spherical array processing*, volume 8, Springer, 2015.

[11] Rafaely, B. and Kleider, M., "Spherical microphone array beam steering using Wigner-D weighting," *IEEE Signal Processing Letters*, 15, pp. 417–420, 2008.

[12] Bernschütz, B., *Microphone arrays and sound field decomposition for dynamic binaural recording*, Ph.D. thesis, Technische Universität Berlin Berlin, Germany, 2016.

[13] Oberkampf, W. L. and Trucano, T. G., "Verification and validation in computational fluid dynamics," *Progress in aerospace sciences*, 38(3), pp. 209–272, 2002.

[14] Eça, L. and Hoekstra, M., "A procedure for the estimation of the numerical uncertainty of CFD calculations based on grid refinement studies," *Journal of Computational Physics*, 262, pp. 104–130, 2014.

[15] Kowalczyk, K. and Van Walstijn, M., "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1), pp. 34–46, 2011.

[16] Saarelma, J., Botts, J., Hamilton, B., and Savioja, L., "Audibility of dispersion error in room acoustic finite-difference time-domain simulation as a function of simulation distance," *The Journal of the Acoustical Society of America*, 139(4), pp. 1822–1832, 2016.

[17] Kowalczyk, K., *Boundary and medium modelling using compact finite difference schemes in simulations of room acoustics for audio and architectural design applications*, Ph.D. thesis, Queen's University Belfast, 2010.

[18] Ahrens, J. and Andersson, C., "Perceptual Evaluation of Headphone Auralization of Rooms Captured with Spherical Microphone Arrays with Respect to Spaciousness and Timbre," *The Journal of the Acoustical Society of America*, 145(4), pp. 2783–2794, 2019, doi: 10.1121/1.5096164.

[19] Bernschütz, B., "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *Proceedings of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA) conference on acoustics*, p. 29, AIA/DAGA, 2013.

[20] Saarelma, J., "Finite-difference time-domain solver for room acoustics using graphics processing units," *Master's thesis, Aalto University, Finland*, 2013.

[21] Southern, A., Murphy, D., Lokki, T., and Savioja, L., "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. Forum Acusticum, Aalborg, Denmark*, pp. 1553–1558, 2011.

[22] Ahrens, J., Geier, M., and Spors, S., "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *Audio Engineering Society Convention 124*, Audio Engineering Society, 2008.

[23] SSR Team, "SoundScape Renderer," [Online] Available: `http://spatialaudio.net/ssr/`, 2019.

[24] ISO, BS, "10399: 2004," *Sensory analysis-Methodology-Duo-trio test*, 2004.

[25] Lawless, H. T. and Heymann, H., *Sensory evaluation of food: principles and practices*, Springer Science & Business Media, 2013.