Audio Engineering Society

Conference Paper 41

Presented at the International Conference on Spatial & Immersive Audio,

2023 August 23-25, Huddersfield, UK

Localisation in virtual choirs: outcomes of simplified binaural rendering

Kajornsak Kittimathaveenan and Sten Ternström

KTH Royal Institute of Technology, Stockholm, Sweden

Correspondence should be addressed to Kajornsak Kittimathaveenan (kkit@kth.se)

ABSTRACT

A virtual choir would find several uses in choral pedagogy and research, but it would need a relatively small computational footprint for wide uptake. On the premise that very accurate localisation might not be needed for virtual rendering of the character of the sound inside an ensemble of singers, a localisation test was conducted using binaural stimuli created using a simplified approach, with parametrically controlled delays and variable low-pass filters (historically known as a 'shuffler' circuit) instead of head-related impulse responses. The direct sound from a monophonic anechoic recording of a soprano was processed (1) by sending it to a reverb algorithm for making a room-acoustic diffuse field with unchanging properties, (2) with a second-order low-pass filter with a cut-off frequency descending to 3 kHz for sources from behind, (3) with second-order low-pass head-shading filters with an angle-dependent cut-off frequency for the left/right lateral shadings of the head, and (4) with the gain of the direct sound being inversely proportional to virtual distance. The recorded singer was modelled as always facing the listener; no frequency-dependent directivity was implemented. Binaural stimuli corresponding to 24 different singer positions (8 angles and 3 distances) were synthesized. 30 participants heard the stimuli in randomized order, and indicated the perceived location of the singer on polar plot response sheets, with categories to indicate the possible responses. The listeners' discrimination of the distance categories 0.5, 1 and 2 meters (1 correct out of 3 possible) was good, at about 80% correct. Discrimination of the angle of incidence, in 45-degree categories (1 correct out of 8 possible) was fair, at 47% correct. Angle errors were mostly on the 'cone of confusion' (back-front symmetry), suggesting that the back-front cue was not very salient. The correct back-front responses (about 50%) dominated only somewhat over the incorrect ones (about 38%). In an ongoing follow-up study, multi-singer scenarios will be tested, and a more detailed yet still parametric filtering scheme will be explored.

1 Introduction

Choir researchers could use virtual acoustics to provide a realistic experience of standing inside a choir, as an experimental tool [1–4]. Other uses could include rehearsing at home, or in a virtual rendering of a concert venue. This requires making a credible rather than exact binaural representation of multiple sources in a room, and also rendering the singer's own voice faithfully even through headphones. Here we revisit the first of these problems.

With choral sounds, the combined sound field from all singers exhibits what is often called the 'chorus' or 'ensemble' effect, meaning that individual voices

blend together and create a decorrelated sound field, where localising individual singers can be difficult and perhaps not even desirable. In [4], the authors submit that, since choir singers typically do not move much, a "system intended for use by choral ensembles may not require motion tracking to maintain immersion." In [5], the authors note that "Reproduction based on a small number of parameters may be advantageous when the complexity of the sound field cannot be captured by the recording array. In this case, estimating a small number of perceptually important parameters may be more useful than attempting to capture the full complexity of the sound field."

Could it then suffice with a simplified parametric simulation of the interaural time and level differences (ITD, ILD), and front-back filtering, as derived from the literature? Baseline data was obtained on using such binaural stimuli for localisation, in a listening test.

2 Rendering

The source-to-listener gain was modelled by the inverse of the distance. The angle of incidence θ was modelled using an ITD of 0.001*sin(θ) s; where $-\pi \leq \theta \leq \pi$ rad and θ=0 signified straight ahead; and second-order low-pass Butterworth shading filters. Lateral shading filter cut-offs were $10 \pm 6 \cdot \sin(\theta)$ kHz, while for front-back, an additional LP cut-off was scaled linearly from 20 to 3 kHz for $\pi/2 \leq \text{abs}(\theta) < \pi$. No broadband ILD change with θ was applied. These filter characteristics were derived from graphs in Lee et al. [6].

The source was an anechoic recording of a soprano performing the first few bars of a bespoke composition. Using a small SuperCollider program [7], 24 different stimuli were generated, with the two factors distance and angle: { 0.5 | 1 | 2 m } × { -135 | -90 | -45 | 0 | 45 | 90 | 135 | 180 ° }. The room acoustic was simulated using the GVerb library function, with settings chosen to give an auditory impression of a medium-sized rehearsal hall.

3 Listening test

30 participants aged 19-21, gender balanced, who had taken several 'technical ear training' courses, were recruited from years 2-4 of a music engineering programme. Participants wore headphones in a recording studio and heard an unaccompanied soprano virtually placed at random positions around the listener. The possible distances and angles were chosen to correspond approximately to the choral formations 'close' (0.5 m), 'lateral' (1 m) and 'circumambient' (2 m), Daugherty [8].

Listeners marked the perceived location of each stimulus sound on a response sheet of polar plots with the categorical distances and angles of incidence (see Figure 1). Responses could be equal to the stimulus, or with a discrepancy in angle, in distance, or in both. There were two groups of 15 participants. Before the test, Group A got to hear six of the sounds, with left-right variation only. Group B got to hear all 24 stimuli, and thus knew better what to expect. Each group did two identical trials of 24 randomized stimuli with a two-minute break in between, for a total of 1440 stimulus assessments (15 persons × 2 groups × 2 trials × 24 stimuli). After the second trial, the participants also filled in a survey with four open questions on how they had experienced the listening test.

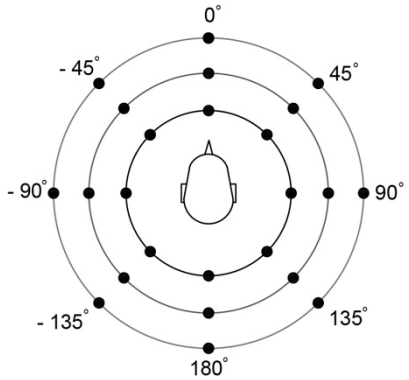

Figure 1. Listening test response sheets each contained six of these figures, for a total of 48 stimuli per particpant. Participants pencilled the perceived location.

4 Results

Discrimination of distance was quite accurate, with Group A reaching 78% correct (Figure 2) and Group B 85% correct (Figure 3). These matrices contain a lot of information; so Figure 4 is provided as a guide to their interpretation. Both groups improved somewhat in Trial 2 compared to Trial 1. The correct angle was judged in 42% of the stimuli by Group A and 53% by Group B. Here the 'cone of confusion' plays in, with most angle errors being due to front-back confusion.

To assess separately the front-back discrimination, Figure 5 shows the responses only to the stimuli with a front or back component, excluding the stimuli with

the source placed at ±90°. Regardless of distance, participants discriminated correctly between front and back with a probability of about 4/8 in both trials, the chance probability being 3/8. The front-back discrimination was slightly better in Group B.

In addition, the results of the survey indicated that one-third of the participants thought the overall binaural model worked reasonably well, while the rest of the group provided specific feedback on which positions sounded reasonable and which did not. The cone of confusion was often mentioned as a source of discontent, which also agrees with the results of the listening test.

5 Discussion

Externalisation was not explicitly evaluated, but we noted informally that it was only sporadic. The good results for discrimination of distance are probably thanks to the simulated direct/diffuse ratio and to the gain changing with distance. While the simple back-shading filtering of the three rear locations did contribute a little to localisation, it would not be sufficient as a reliable cue for front-back localisation. In a follow-up study, we will investigate in which aspects this would be important for choral realism, and if it is, explore more elaborate parametric filtering schemes.

		N-135	N180	N135	N90	N45	N0	N-45	N-90	M-135	M180	M135	M90	M45	M0	M-45	M-90	F-135	F180	F135	F90	F45	F0	F-45	F-90
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-135	1	10						8	8						1			1							2
N180	2	3	7				1	9	2		3	3				1		1							
N135	3		1	4	12	10							1		1							1			
N90	4			8	17	1							1		1							1	1		
N45	5			7	11	7						2	1	1								1			
N0	6		5				17	2		1	1				2	1							1		
N-45	7	5						8	12							1	3								1
N-90	8	2						4	21							1		1						1	
M-135	9	2	1					1		10	1					1	4	3	1					5	1
M180	10	2	2							2	8	1		1	8			1	4				1		
M135	11											6	7	13	1							3			
M90	12			1	1	2						4	11	7								3	1		
M45	13			1	1							6	8	10								2	2		
M0	14			1				1		1	6		1	1	11	5							2	1	
M-45	15	1						2	2	2						8	12	1							2
M-90	16	2						1	1	10						6	7							2	1
F-135	17	1							1	2						1	1	13	1					6	4
F180	18		2												2			1	18	1			5	1	
F135	19				1		1						3	1						6	12	6			
F90	20				1	2							2						2	10	7	6			
F45	21			2							1			1	5					5	6	10			
F0	22		1				1				3				4	2			10				7	2	
F-45	23	1								2					1	2	2	5					1	10	6
F-90	24	1								3					1	3	1	6	1					3	11

Figure 2. Response confusion matrix for Group A. Each square shows the number of responses. N – Near, M – Medium, F – Far. Angles are in degrees, < 0 to the left, 0 straight ahead, > 0 to the right.

		N-135	N180	N135	N90	N45	N0	N-45	N-90	M-135	M180	M135	M90	M45	M0	M-45	M-90	F-135	F180	F135	F90	F45	F0	F-45	F-90
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
N-135	1	24		1						1						1	2	1							
N180	2		24					1	1					1					1						
N135	3			8	15	8							1												
N90	4			8	15	7			1				1												
N45	5			9	10	8			1			1	3												
N0	6	1	2			2	22	3		1				1											
N-45	7	6						6	17						2	1									
N-90	8	6						8	16	1						1									
M-135	9	2	1							17	1				2	2		4						2	1
M180	10	1	2								18			5				1	3				2		
M135	11					1						6	9	12						1	3				
M90	12			2	2							11	11	4							1	1			
M45	13			2	2	1						12	8	5		1					1				
M0	14						3	1		1	4			1	21			1							
M-45	15	1								6				1	1	7	13							2	1
M-90	16	2						4	1	4						4	14							1	2
F-135	17									2	1							16	1					6	6
F180	18										1				1			1	22	1			6		
F135	19									1		6		3						6	8	8			
F90	20											1	2	1					1	9	14	4			
F45	21											1	2	3					1	5	10	9			1
F0	22		1												4			1	6		1	1	18		
F-45	23									2						3	4	4	1				1	8	9
F-90	24									3						1	2	13						6	7

Figure 3. Response confusion matrix for Group B. N – Near, M – Medium, F – Far. Angles are in degrees, < 0 to the left, 0 straight ahead, > 0 to the right.

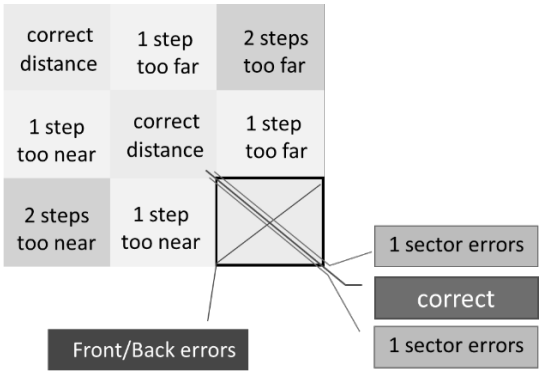

Figure 4. Guide to interpreting the confusion matrices in Figures 2 and 3.

6 Conclusions

In the current study, a binaural model rendering a soprano recording was tested with regard to its ability to provide localisation cues. The localisation test was conducted using a greatly simplified approach with parametrically controlled delays and variable low-pass filters. The discrimination of distance had an accuracy of 82%. The discrimination of front/back stimuli was somewhat better than chance, at about 4/8 rather than 3/8. The SuperCollider script and stimulus files used for this study are available from the authors on request.

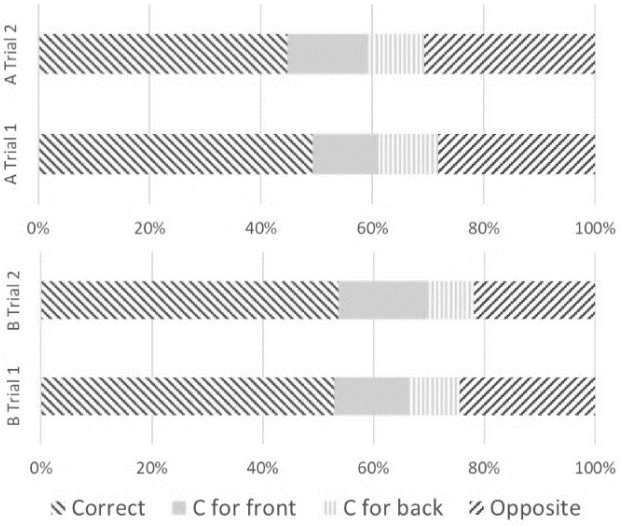

Figure 5. Front-back discrimination of front and back stimuli, excluding side stimuli. Group B (lower) scored somewhat higher. The 'C' (center) bars tally the incorrect votes for side stimuli.

In ongoing follow-up studies, parametric filtering schemes will be evaluated and/or developed to reduce localization errors, and multi-singer scenarios will be tested. The goal is to enable choir researchers to use virtual acoustics as an experimental tool to provide a realistic experience of standing inside a choir.

Acknowledgements

The anechoic recording used as the stimulus was kindly provided for this experiment by Helena Daffern of the Audio Lab of the University of York. The GUI of the listening test was supported by Rathachai Chawuthai of the School of Engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL). The authors are grateful to the volunteer listeners from the Institute of Music, Science and Engineering, KMITL for their participation in the listening test.

References

[1] Libeaux, A., Lentz, T., Houben, D., and Kob, M., "Voice assessment in choir singers using a virtual choir environment," *19th International Congress on Acoustics 2007*, pp. 1819–1824 (2007).

[2] Brereton, J. S., Murphy, D. T., and Howard, D. M., "Singing in Space(s): Singing performance in real and virtual acoustic environments —Singers' evaluation, performance analysis and listeners' perception," [Doctoral Thesis, University of York]. (2014).

[3] Dedousis, G., Andreopoulou, A., and Georgaki, A., "The impact of room acoustics on choristers' performance: from rehearsal space to concert hall," in *Proc. Stockholm Music Acoustics Conference 2023*, Stockholm 14-15 June, (2023).

[4] Eley, N., Mullins, S., Stitt, P., and Katz, B. F. G., "Virtual Notre-Dame: preliminary results of real-time auralization with choir members," *Intl Conf 3D Audio (I3DA)*, (2021).

[5] Rafaely, B., Tourbabin, V., Habets, E., Ben-Hur, Z., Lee, H., Gamper, H., Arbel, L., Birnie, L., Abhayapala, T., and Samarasinghe, P., "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes – review and challenges," *Acta Acustica*, vol. 6, 47 (2022).

[6] Lee, G. T., Choi, S. M., Ko, B. Y., and Park, Y. H., "HRTF measurement for accurate sound localization cues" (2022). Available: https://arxiv.org/abs/2203.03166v2

[7] *SuperCollider 3.13.0*. (2023), McCartney, J. Available: http://supercollider.github.io/

[8] Daugherty, J. F., "Spacing, formation, and choral sound: preferences and perceptions of auditors and choristers," *Journal of Research in Music Education*, 47(3), pp. 224–238 (1999).