Dynamic voice directivity in room acoustic auralizations

Barteld NJ Postma, Brian FG Katz

Audio & Acoustic Group, LIMSI, CNRS, Université Paris-Saclay, Orsay, France. Email: {first.lastname}@limsi.fr

Introduction

The use of room acoustic auralizations has increased due to the improving computing power available and the quality of numerical modelling software. In such auralizations, it is often possible to prescribe the directivity of an acoustic source in order to better represent the way in which a given acoustic source excites the room. However, such directivities are static, being defined according to source excitation as a function of frequency for the numerical simulation. While sources such as pianos vary little over the course of playing, it is known that voice directivity varies, sometimes considerably, due to both phoneme dependent radiation patterns [1] linked to changes in mouth geometry and dynamic orientation.

Studies by Rindel and Otondo [2, 3] proposed to achieve the inclusion of dynamic vocal/instrumental directivity through the usage of multi-channel source directivity auralization[1]. This method employs anechoic multi-channel recordings. The radiation sphere source is divided into segments representing each microphone position. The room impulse response (RIR) is then calculated for each segment and convolved with the corresponding microphone channel of the anechoic recording. Convolutions of each channel are then down-mixed to create a multi-channel source directivity auralization. This source representation follows changes in direction, movement, asymmetry, and orientation of the recorded source, unlike simple single channel source representations. Multi-channel source directivity auralizations were subjectively compared to a static directivity source type. The geometrical acoustics (GA) software ODEON was employed to create auralizations of an anechoic clarinet recordings convolved with 2, 5, and 10 channels and a single channel with a static clarinet directivity. Fig. 1 depicts the multi-channel sources which were combined without overlap to represent the spherical recording area around the musician. A listening test compared these auralizations in terms of *perceived spaciousness of sound in the room* and *perceived naturalness of timbre of the clarinet*. Results of that study indicated that the 10-channel representation was judged significantly less spacious than the three other source representations. Additionally, the test subjects significantly preferred the 10-channel auralization over the other in terms of *perceived naturalness*.

Vigeant et al. [4] compared 1–, 4–, and 13–channel source directivity auralizations by means of a subjective listening test. The multi-channel source directivity representations and employed GA software were the same as the previously mentioned studies. The first phase of the test compared the different source representations for a violin, trombone, and flute in terms of *realism* and *source size*. Subjects rated the 13-channel auralization significantly more realistic than the other two. No significant trend was found regarding *source size*. In the second phase, the effect of orientation (facing the audience and facing 180° from the audience) of the 4-channel and 13-channel auralization on *Clarity* were studied. The results indicated that the 13-channel auralization was perceived clearer when the source faced the audience. No significant difference regarding *Clarity* was observed when the sources faced 180° from the audience.

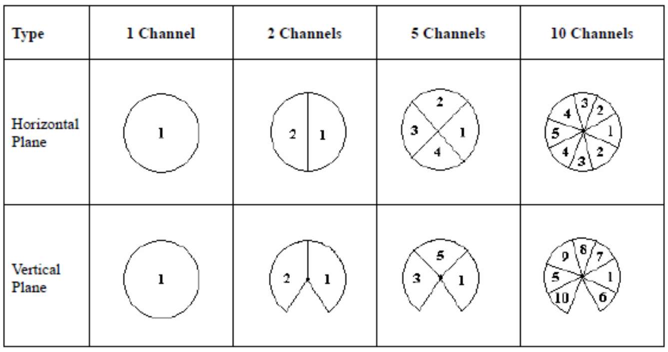

Figure 1: partial sources used for multi-channel source directivity auralizations (from [2]).

In contrast to previous studies, the final goal of this project is to employ multi-channel source directivity for the inclusion of dynamic source directivity and orientation using a single channel anechoic recording. Advantages are a better representation of source directivity, simulations need to be run only once even when the selected instrument is adjusted, and source directivity can be adjusted post-simulation in real-time. A first step towards this goal is taken in this study, by perceptually examining the usage of a newly established source decomposition. Where previous studies employed segmented directivity approaches, the current study investigates multi-channel source decomposition using an overlapping beamforming approach, described in Sec. 2. In order to validate this multi-channel source directivity, this source was placed in a GA model based on the Théâtre de l'Athénée, created and calibrated according to [5]. The resulting auralizations were compared by means of a subjective listening test to auralizations exploring static directivities. The setup and results of this test are described in Sec. 3. The inclusion of a single channel anechoic recording into the multi-channel source directivity application is beyond the scope of the current study.

[1] The original paper coined this application multi-channel auralization. In order to prevent confusion with distributed sources multi-channel auralizations, this article will employ the term multi-channel source directivity auralization.

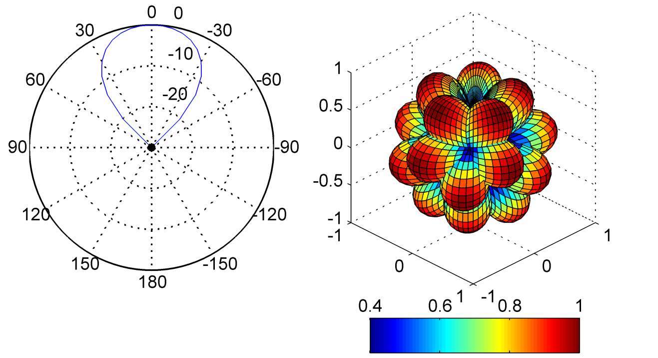

Figure 2: (left): 2D polar plot of a single beam (dB scale), and (right): superposition of the 20 3D beam patterns (lineair scale) to show orientations.

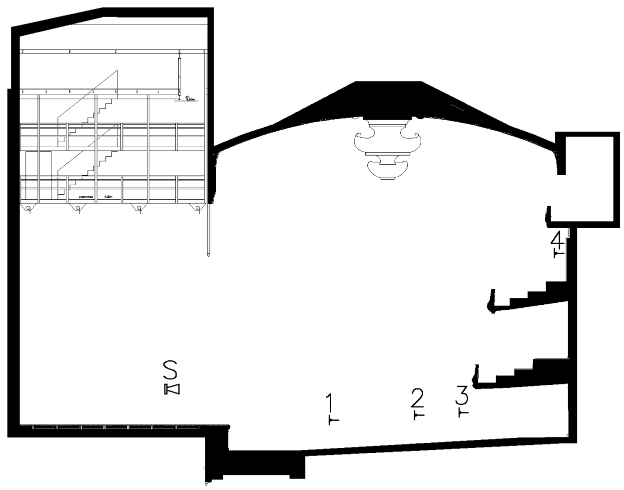

Figure 3: Section of the Théâtre de l'Athénée depicting the source position on stage and the 4 receiver positions (1, 2, 3 first floor, and 4: third floor)

Creation of the employed auralizations

Based on the microphone configuration of an anechoic recording a beam pattern was established with slightly overlapping segments. A multi-channel source which approximated this beam pattern as well as sources with an omni-directional, static singer, and static loudspeaker directivity were positioned in a GA model of the Théâtre de l'Athénée. The resulting RIRs were employed to create auralizations.

Anechoic Recordings

Anechoic vocal recordings were made in an anechoic chamber using 20 microphones geometrically positioned at the verticies of a dodecahedron [6, 7]. The singer's mouth was situated at the center of the array. The selected extract for this study was a female soprano singing *Abendempfindung*, by W.A. Mozart. The singer was instructed to remain with her head in the same position and keep the same orientation during the recording.

Beam pattern

A beam forming design approach was used to subdivide the sphere. The beams were designed to have minimal overlap while having an equal gain sum for all sections in order to approximate an omni-directional pattern. The following control points were employed:

- 0° - No attenuation
- 21° - (center of the rib between two microphones) designed to sum 2 beams to 0 dB
- 42° (center of the pentagon) to sum 5 beams to 0 dB
- 180° - Maximum attenuation

The 2D beam was produced (see Fig. 2) using a spline interpolation in 5° steps. It was rotated around its symmetry axis to create the 3D beam. The 20 instances of the beam pattern were aimed at one of the 20 microphone positions of the anechoic recordings. The result of the summation produces an omni-directional sphere, with a variation in the directivity pattern of ±0.2 dB.

GA model

Using CATT-Acoustic (v.9.0.c, TUCT v1.1a) [8], the established 20 sources were positioned in a GA model of the Théâtre de l'Athénée, a 570-seat theater with a reverberation time of approx. 1.5 s (500-1000 Hz). Simulations were run with 400,000 rays using *Algorithm 2: Longer calculation, detailed auralization*, suitable for the chosen venue.

As a baseline validation of the multi-channel source directivity auralization, the mix of the channels RIRs (reconstructed RIR) should be perceptibly equal to the RIR resulting from a simulation with an omni-directional source (omni-directional RIR). Therefore, in addition to the designed multi-channel source simulations were performed with an omni-directional source. In order to be able to compare the multi-channel source directivity auralization with static sources, simulations were also carried out using sources with static singer [9], and static loudspeaker [10] directivities.

All sources were positioned on the center of the stage. 4 binaural receiver positions were simulated on the center axis of the theater at various positions (see Fig. 3). Post-simulation, the reconstructed, omni-directional, static singer, and static loudspeaker RIRs were convolved with a single-channel recording of the selected extract. Finally, the 20 channels were convolved with the corresponding 20 channels of the anechoic recording and summed. This resulted in five binaural auralizations per receiver position. RMS of the convolutions was used for normalization.

Listening test

The resulting auralizations were compared by means of a subjective listening test. This section first describes the setup of this test after which the results are discussed.

Test setup

The test was setup as a randomized experiment with five variants corresponding to the source directivity-types. Binaural auralizations were compared per receiver position 1, 2, 3, and 4. Additionally, one iteration was repeated in order to monitor the repeatability of the test (receiver position 3). Participants were initially presented one training iteration with the test administrator present in order to ensure they understood the task (receiver position 2), resulting in six iterations. The training session results were not tabulated in the presented results.

For each iteration, participants compared and rated the five auralizations in terms of *Plausibility*, *Clarity*, *Distance*, *Apparent Source Width (ASW)*, and *Listener Envelopment (LEV)* on a discrete scale ranging from 1 ('least ...') to 7 ('most ...'). Participants were forced to use the 2 extreme scale values at least once per attribute. They were allowed to give auralizations the same rating. Presentation order of the receiver position and correspondence to source directivity-type were randomized. This protocol is similar to [4], which employed similar acoustic attributes and a 7 point scale. However, the current study employed two additional acoustical attributes, the auralizations were compared during the same iteration, and participants were forced to use the extreme scale values.

28 participants (mean age: 35.3 SD: 12.6) who all reported normal hearing took part in the study. They were selected to have experience with either room acoustics or vocal/instrumental performances as it was hypothesized that experienced listeners would perform significantly better than untrained listeners [11]. 15 participants took the test in an isolation booth at LIMSI (ambient noise level < 30 dBA), 10 in an isolation booth at the *Institut Jean le Rond d'Alembert* (LAM) (ambient noise level < 30 dBA), and 3 participants in a quiet office at the *Institut National d'Histoire de l'Art* (THALIM) (ambient noise level $= 31$ dBA). Participants were given written instructions before commencing the test which explained the task, described the attribute definitions, and illustrated the software usage. Participants were able to listen to the auralizations as many times as desired. Auralizations were presented via headphones (Sennheiser model HD 650) at an RMS level of 80 dBA.

Results

Initial attention is given to the repeatability of the responses, determined from the absolute difference between the repeated iteration condition. The mean difference between repetitions for each attribute on the 7 *pt* scale across participants was: *Plausibilty*$=$ 1.9, *Clarity*$=$ 2.1, *Distance*$=$ 1.4, *ASW*$=$ 1.7, and *LEV*$=$ 1.5).

In order to validate the multi-channel auralization the omni-directional and reconstructed omni-directional auralization are compared first. One could employ a one-way analysis of variance (ANOVA) with an $\alpha = 0.05$ level. This found significant differences for the attributes

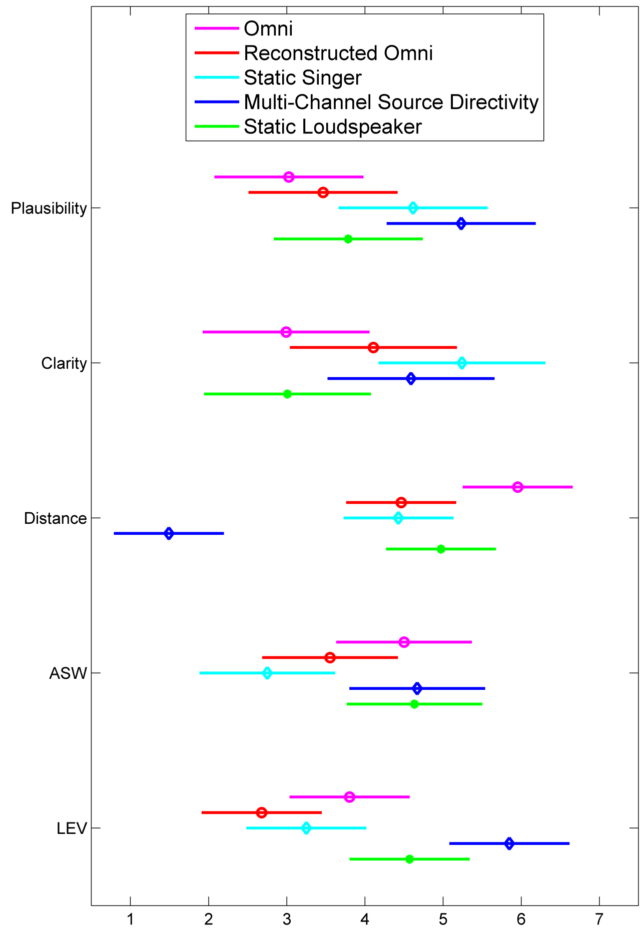

Figure 4: Mean ratings of the combined positions for the omni-directional, reconstructed omni-directional, static singer, multi-channel auralization, and static loudspeaker auralizations regarding the tested attributes. Overlapping comparison intervals (repeatability interval) indicate lack of perceived difference.

Clarity ($F = 17.14$, $p < 10^{-2}$), *Distance* ($F = 40.37$, $p < 10^{-2}$), *ASW* ($F = 10.14$, $p < 10^{-2}$), and *LEV* ($F = 17.92$, $p < 10^{-2}$). However, as the mean differences between the repeated iteration conditions were rather large, this study opts to employ individual attribute repeatability mean values as tolerance ranges to estimate whether an acoustic attribute differed perceptually between auralization types. Fig. 4 shows that the reconstructed omni-directional auralization was perceived slightly closer than the omni-directional auralization, however the remaining attributes were perceived similarly.

Subsequently, the multi-channel directivity source auralizations are compared to the static loudspeaker and singer auralizations. For completeness, the one-way ANOVA results are presented in Table 1. Using the mean difference between repetitions, it can be seen that the multi-channel source directivity auralizations were perceived significantly closer than the static loudspeaker and singer auralizations as well as wider and more enveloping

Table 1: One-way ANOVA F and p-value results comparing either static loudspeaker or static singer to the multi-channel auralization.

Acoustical Attribute	sta. loudspeaker vs. multi-channel		sta. singer multi-channel	
	F	p-value	F	p-value
Plausibility	25.08	$< 10^{-2}$	4.18	0.04
Clarity	28.17	$< 10^{-2}$	4.40	0.03
Distance	282.87	$< 10^{-2}$	175.45	$< 10^{-2}$
ASW	0.01	0.90	43.39	$< 10^{-2}$
LEV	23.41	$< 10^{-2}$	98.57	$< 10^{-2}$

than the static singer auralizations (see Fig. 4).

Conclusion

The purpose of this study was to take a first step towards the inclusion of dynamic source directivity in auralizations employing only a single channel anechoic recordings using the multi-channel source directivity application. Therefore, the use of a specifically defined multi-channel source directivity auralization was explored. An overlapping beam pattern was established which correctly reproduced an omni-directional pattern. A multi-channel source based on this beam pattern was positioned in a GA model of the Théâtre de l'Athénée. The resulting multi-channel auralization was compared by means of a subjective listening test to auralizations with a reconstructed omni-directional, omni-directional, static singer, and loudspeaker directivity for the acoustical attributes *Plausibility*, *Clarity*, *Distance*, *ASW*, and *LEV*.

Attention was given to the justification of the application of the multi-channel auralization. As the listening test identified one perceptual and four statistical differences between omni-directional and reconstructed omni-directional auralizations, additional studies are underway to understand the reason for these differences as they can create audible artifacts.

For this reason one needs to be cautious when drawing conclusions from the current results of the multi-channel source directivity auralization and auralization based on static directivity types. Therefore, the perceptual tolerance range was chosen to compare results. These results indicated that there are perceptual differences between the presented multi-channel source directivity application and static singer source auralizations in terms of perceived *Distance*, *ASW*, and *LEV*.

As the singer in the anechoic recording was instructed to keep her head in the same orientation, it can be concluded that the inclusion of phoneme dependent directivity leads to perceptibly different auralizations than those based on static directivity source types. This notion justifies further endeavours towards creating a dynamic source directivity employing single channel anechoic recordings. This entails the reconstruction of directivity patterns from multi-channel anechoic recordings, comparison of the dynamically reconstructed directivity patterns to direct multi-channel auralizations, and finally examination of dynamic source orientation variations.

Acknowledgement

The authors would like to thank Tapio Lokki and Jukka Pätynen of Aalto University for assisting with the anechoic vocal recordings as well as LAM and THALIM for their help in hosting the listening test. The authors would specifically like to thank all participants of the listening test for their time. This work was funded in part by the ECHO project (ANR-13-CULT-0004, echo-projet.limsi.fr). Partners include THALIM/ARIAS-CNRS, *Bibliothèque nationale de France* (BnF), and LIMSI-CNRS.

References

[1] Katz, B.F.G, Prezat, F. and d'Alessandro, C.: Human voice phoneme directivity pattern measurements. 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan Honolulu, Hawaï, 3359, 2006.

[2] Rindel, J.H., Otondo, F. and Christensen, C.L.: Sound source representation for auralization. In Proc. Int Symp on Rm Acoust: design and science, Hyogo, Japan, pp. 1–8, 2004.

[3] Otondo, F. and Rindel, J.H.: A new method for the radiation representation of musical instruments in auralizations. Acta Acustica 91(5) pp. 902-906, 2005.

[4] Vigeant, M.C., Wang, L.M., and Rindel, J.H.: Objective and subjective evaluations of the multi-channel auralization technique as applied to solo instruments. Applied Acoustics 72, pp. 311-323, 2011.

[5] Postma, B.N.J. and Katz, B.F.G.: Creation and calibration method of virtual acoustic models for historic auralizations. Virtual Reality, vol. 19, pp. 161-180, 2015.

[6] Lokki, T., Pätynen, J., and Pulkki, V.: Recording of Anechoic Symphony Music. In Proc. Acoustics '08, Paris, pp. 6431-6436, 2008.

[7] Pätynen, J., Katz, B.F.G. and Lokki, T.: Investigations on the balloon as an impulse source. J Acoust Soc Am vol. 129 no. 1, pp. EL27-EL33, 2011.

[8] Dalenbäck, B CATT-A v9: User's Manual CATT-Acoustic v9. CATT, Gothenburg (Sweden), 2011.

[9] Marshall, A.H. and Meyer, J.: Directivity and Auditory impression of Singers . Acustica, vol. 58, p. 130-140, 1985.

[10] Choueiri, E., Genelec 8351A directivity., URL: `https://www.princeton.edu/3D3A/Directivity/Genelec\%208351A/images/Plots/`, accessed: 2016-01-19, 2015.

[11] Olive, S.: Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study. J Audio Eng Soc vol. 51, no. 9, pp. 806-825, 2003.