Preserving sound source radiation-characteristics in network-based musical performances

Markus Noisternig[1], Franz Zotter[2], Rémi Desmonet[1], Winfried Ritsch[2],
[1] *IRCAM – CNRS UMR STMS, 75004 Paris, France, Email: {markus.noisternig, remi.desmonet}@ircam.fr*
[2] *Institute of Electronic Music and Acoustics, KUG, 8010 Graz, Austria, Email: {zotter, ritsch}@iem.at*

Introduction

When reproducing only the temporal and spectral characteristics of a sound source, but not its spatially varying radiation patterns, the perception of timbre at the listener's position can be both inaccurate and unrealistic. This is especially true in reverberant environments such as concert halls. Preserving the radiation characteristics of a sound source has therefore become a challenging task in virtual audio environments. This article discusses technologies applied to a networked audio live concert in which the sound radiation characteristics of a solo instrument have been transmitted over the Internet in order to enhance the listening experience. The sound field was captured with a surrounding 64-channel spherical microphone array at the IEM in Graz, decomposed into spherical harmonics and streamed with lowest possible latency over the Internet to IRCAM in Paris. At IRCAM the solo instrument was reproduced with a linear 128-loudspeaker wave field synthesis (WFS) array. The mapping of radiation patterns encoded in spherical harmonic domain to circular harmonic domain is outlined with respect to WFS rendering.

Capture and encoding of sound radiation patterns

Weinreich and Arnold [1] have shown that the expansion of the sound pressure field measured on two concentric spheres into spherical harmonics allows to determine the forward and backward propagation of the sound waves independently. In this study, a single surrounding 64-channel spherical microphone array was used to derive a spatiotemporal *fingerprint* of the recorded instrument, assuming that the incoming sound wave is negligible.

The following subsections give a brief overview on how to determine the spatiotemporal wave spectrum from discrete observations on the sphere or a great circle [2, 3]. It is further shown how the microphone array signals $\boldsymbol{x}(t)$ are encoded before streaming them over the Internet.

Circular harmonics expansion. Any arbitrary and square-integrable signal $x(\varphi, t)$ on a great circle of a sphere $\mathbb{S}^2$ at azimuth angle φ, which represents the radiated sound pressure field, can be expanded into an infinite sum of circular harmonics weighted with the corresponding expansion coefficients $\xi_m(t)$

$$x(\varphi, t) = \sum_{m=-\infty}^{\infty} \xi_m(t)\, \Phi_m(\varphi). \qquad (1)$$

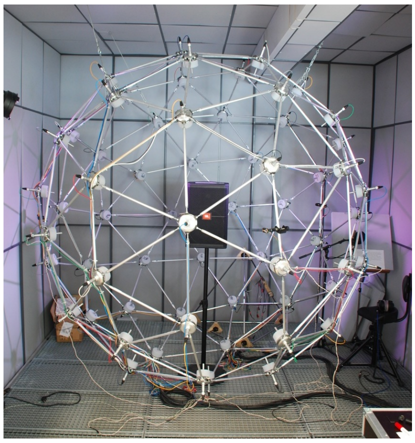

Figure 1: Surrounding 64-channel spherical microphone array in the semi-anechoic chamber at the IEM.

Real-valued circular harmonics can be defined as (cp. [4])

$$\Phi_m(\varphi) = \sqrt{\frac{2-\delta_m}{4\pi}} \begin{cases} \cos(m\varphi), & \text{for } m \geq 0, \\ \sin(|m|\varphi), & \text{for } m < 0. \end{cases} \qquad (2)$$

where δ_m is the Dirac delta function. The coefficients $\xi_m(t)$ are not directly accessible from observations with a spherical microphone array. However, circular harmonics are associated with spherical harmonics and the expansion coefficients can be thus determined from the given arrangement of microphones.

Spherical harmonics expansion. On a sphere with the azimuth angle φ and the zenith angle ϑ, the sound radiation can be expanded into spherical harmonics

$$x(\varphi, \vartheta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \chi_n^m(t)\, Y_n^m(\varphi, \vartheta). \qquad (3)$$

The real-valued spherical harmonics are defined as

$$Y_n^m(\varphi, \vartheta) = N_n^{|m|}\, P_n^{|m|}(\cos(\vartheta))\, \Phi_m(\varphi), \qquad (4)$$

$$\text{with } N_n^{|m|} = \sqrt{\frac{(2n+1)}{2} \frac{(n-|m|)!}{(n+|m|)!}}.$$

and the expansion coefficients $\chi_n^m(t)$ can be determined by forward harmonic transform

$$\chi_n^m(r_0, t) = \int_{\mathbb{S}^2} x(r_0, \theta, \phi) Y_n^m(\varphi, \vartheta)^* d\Omega, \qquad (5)$$

with the rotation invariant measure $d\Omega = \sin\theta d\theta d\phi$.

Circular re-expansion of the spherical wave spectrum in the horizontal-only plane. On the equator $\vartheta = \pi/2$, i.e. the horizontal plane, a spherical function $x(\varphi, \pi/2, t)$ solely depends on the azimuth angle φ. Eq. (1) can be re-arranged to obtain the circular harmonic expansion coefficients

$$x(\varphi) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \underbrace{\chi_n^m(t)\; N_n^{|m|}\; P_n^{|m|}(0)}_{\xi_m}\; \Phi_m(\varphi),$$

$$\xi_m(t) = \sum_{n=|m|}^{\infty} \chi_n^m(t)\; N_n^{|m|}\; P_n^{|m|}(0). \tag{6}$$

To convert the spherical harmonic signals $\boldsymbol{\chi}_{\mathrm{N}}(t)$ to circular harmonic signals $\boldsymbol{\xi}_{\mathrm{N}}(t)$, the terms $N_n^{|m|} P_n^{|m|}(0)$ from Eq. (6) are combined in a conversion matrix $\boldsymbol{C}_{\mathrm{N}}$. Thus the desired signals can be derived as

$$\boldsymbol{\xi}_{\mathrm{N}} = \boldsymbol{C}_{\mathrm{N}}\; \boldsymbol{Y}_{\mathrm{N}}^{-1}\; \boldsymbol{x}(t). \tag{7}$$

The rows of matrix $\boldsymbol{Y}_{\mathrm{N}}$ contain the truncated spherical harmonics of orders $n \leq \mathrm{N}$ evaluated at each discrete microphone position.

Transmission of sound source radiation characteristics over the Internet

The real-time exchange of high-quality audio data over the Internet requires high speed backbones. Live music interaction requires short round trip times: the ensemble performance threshold (EPL) should be < 25 ms [5]. The signal delays originate from network and sound card buffers, encoder/decoder, and the transmission speed in a fiber optic cable (approx. $0.7 \times$ the speed of light). Audio signal dropouts and artifacts are noticeable and may be annoying, thus stable network connections with low error and loss rates and no breaks in continuity are mandatory. The national research and education networks ACOnet (Austria) and RENATER (France) provide fast and reliable high-bandwidth networks and backbones, which have been connected via the pan-European network for research and education (GEANT). Network speed tests showed a stable connection without any packet loss at transmission speeds of 1 Gbps. Due to the stable network connections the User Datagram Protocol (UDP) could be used; it does not perform hand-shaking dialogues and thus further reduces network latency. Audio streaming was done using SoundWIRE's JackTrip software running on Linux. JackTrip is designed for low-latency uncompressed bi-directional multi-channel audio streaming over IP networks. Video streaming was done using low-latency uncompressed DVTS (Digital Video Transport System). The video and audio streams were synchronized on the client side and the overall latency was in the order of 20 ms.

Playback of directional sound sources with wave field synthesis arrays

Wave field synthesis (WFS) is a spatial sound field reproduction technique that is principally based on the Huygens-Fresnel principle. It aims to authentically reproduce any given sound field over an extended listening area. This can be only achieved below an upper frequency limit, which is mainly determined by the spacing of the loudspeakers. The WFS approach was originally limited to omnidirectional sound sources. Corteel [6] proposed the use of a subset of circular harmonics to synthesize virtual sound sources with adjustable radiation characteristics. The order-truncated elementary directivity functions are implemented by time domain filters. A linear loudspeaker array only allows to reproduce the horizontal dependencies of the sound field and source radiation characteristics. The sound field outside the horizontal plane is a combination of the reproduced field and the radiation characteristics of the loudspeakers.

The network streamed solo instrument performance was reproduced in the variable acoustics concert hall at IRCAM using a linear 128-loudspeaker WFS array. The array was installed behind the stage at 2m height from the stage floor below a video projection screen. A focal sound source was synthesized on stage, in front of the array, preserving the radiation characteristics of the recorded instrument. The limitation to the horizontal plane was transferred to the encoder, cf. Eq. 7, thus reducing the required transmission channels. The received audio channels were directly matrixed to the corresponding elementary directivity functions without the need for a special decoder thus providing lowest possible latencies.

Acknowledgment

This project was supported by the EACEA Culture Project of the European Commission. The authors are grateful to Sascha Hois, Thomas Eibinger for rehearsing and performing the instrumental piece, and to Matthias Frank, Laurent Ghys, Arnaud Gomes-do-Val, and Iohannes Zmölnig for help with experimental procedures.

References

[1] Weinreich, G., and Arnold, E. B.: Method for measuring acoustic radiation fields. J. Acoust. Soc. Am. 68:2 (1980), 404–411

[2] Williams, E. G.: Fourier Acoustics. Academic Press, London (1999)

[3] Noisternig, M., Zotter, F., and Katz, B. F. G.: Reconstructing sound source directivity in virtual acoustic environments. In Principles and Applications of Spatial Hearing, World Scientific Publishing (2011)

[4] Zotter, F.: Analysis and synthesis of sound-radiation with spherical arrays. PhD thesis, Univ. of Music and Performing Arts, Graz (2009)

[5] Schuett, N.: The effect of latency on ensemble performance. Technical report, CCRMA Dept. of Music, Stanford (2002)

[6] Corteel, E.: Synthesis of directional sound sources using wave field synthesis, possibilities, and limitations. EURASIP J. Applied Sig. Proc. 11 (2007), 188–202