

# 球状マイクロホンアレイ信号のバイノーラル再生について

皆川孝志

2020 年 8 月 25 日

本資料では球状マイクロホンアレイ信号をバイノーラル再生する手法についての近年の動向を紹介する。

## 1 はじめに

従来のバイノーラル録音ではダミーヘッドを用いて行うものが多いが，ダミーヘッドを用いて行うバイノーラル録音ではある一方向に関する信号しか再生することができない [1]．従来手法として，原音場から得られた球面調和展開係数をデコードし，仮想音源からの信号にに対して頭部伝達関数を畳み込み，それらを重ね合わせてバイノーラル信号を得る手法 [7, 8] (virtual loudspeaker approach) も存在する．しかし，この手法は仮想音源をデコードし，さらに頭部伝達関数を畳み込む必要があることから計算コストが高い．またバイノーラル信号の質が原音場の観測点と頭部伝達関数の観測点の位置に依存しやすいという問題もある．そこで [10] で提案された球面調和領域にて IACC を算出する理論を基礎とした，原音場の信号および頭部伝達関数の球面調和展開を利用し，音場の任意の方向におけるバイノーラル信号を再生する手法が提案されている [1, 3, 4, 9, 12, 13, 14]．

本資料では球状マイクロホンアレイで録音した原音場における信号をバイノーラル再生する最近手法について，その概要を紹介する．

## 2 理論

以下，[1] を参考に球状マイクロホンアレイ信号および頭部伝達関数からバイノーラル信号を得るための理論について述べる．

### 2.1 球面調和関数

波動方程式

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 \quad (1)$$

に支配される音場について音圧  $p$  が球座標系で  $p(r, \theta, \varphi, t) = R(r)\Theta(\theta)\Phi(\varphi)T(t)$  の形に変数分離できると

仮定し、両辺の時間-周波数に関するフーリエ変換をとると式 1 から

$$\begin{aligned}\frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + (k^2 r^2 - n(n+1)) R &= 0 \\ \frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \left( n(n+1) - \frac{m^2}{\sin^2 \theta} \right) \Theta &= 0 \\ \frac{d^2 \Phi}{d\varphi^2} + m^2 \Phi &= 0\end{aligned}$$

が得られる。これらの微分方程式の解は動径方向、角度方向についてそれぞれ得られる。動径方向については解として以下の第 1 種および第 2 種球ベッセル関数によって与えられる。

$$\begin{aligned}j_n(kr) &= \sqrt{\frac{\pi}{2kr}} J_{n+1/2}(kr) \\ y_n(kr) &= \sqrt{\frac{\pi}{2kr}} Y_{n+1/2}(kr)\end{aligned}$$

ここで  $J, Y$  はそれぞれ第 1 種および第 2 種ベッセル関数である。角度成分に関しては、球面調和関数  $Y_n^m(\theta, \varphi)$  が用いられるが、その具体的な式は文献によって異なる。代用的なものとして以下の 3 つの式が用いられている。[1] では式 2 は複素非対称型 (complex asymmetric), 式 3 は複素対象型 (complex symmetric), 式 4 は実数型 (real convention) と呼んでいる。

$$Y_n^m = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \theta) e^{im\varphi} \quad (2)$$

$$Y_n^m = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos \theta) e^{im\varphi} \quad (3)$$

$$Y_n^m = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos \theta) \cdot \begin{cases} \sqrt{2} \cos(m\phi), & m > 0 \\ 1, & m = 0 \\ \sqrt{2} \sin(m\phi), & m < 0 \end{cases} \quad (4)$$

$$(5)$$

これらの球面調和関数はすべて正規直交系であり、単位球面  $S$ 、微小角度要素  $d\Omega$  について以下が成り立つ。

$$\int_S Y_n^m (Y_{n'}^{m'})^* d\Omega = \delta_{n,n'} \delta_{m,m'}$$

また、球面調和関数の複素共役は複素非対称型では

$$(Y_n^m)^* = (-1)^m Y_n^{-m}$$

となり、複素対象型では

$$(Y_n^m)^* = Y_n^{-m}$$

となる。自明であるが、実数型の球面調和関数とその複素共役は等しい値をとる。

## 2.2 球面調和関数展開

波動方程式を満たす任意の関数  $f$  は球面調和関数の線形結合で表せる．すなわち展開係数  $P_{nm}$  について以下が成り立つ．

$$f = \sum_{n=0}^{\infty} \sum_{m=-n}^n P_{nm} Y_n^m \quad (6)$$

展開係数  $P_{nm}$  を得るには両辺の  $Y_n^m$  との内積を取り，以下の積分を実行すればよい．

$$P_{nm} = \int_S f(\Omega, \omega) Y_n^m(\Omega)^* d\Omega \quad (7)$$

この積分により得られる  $P_{nm}$  の組は設定した球面  $S$  によって異なり，異なる  $P_{nm}$  の組同士の関係は波動方程式の動径方向の解によって定義される． $S$  が開球の場合，異なる半径  $r, r_0$  を持つ球面上の積分で得られた展開係数  $P_{nm}(r), P_{nm}(r_0)$  間の関係は

$$P_{nm}(r) = \frac{j_n(kr)}{j_n(kr_0)} P_{nm}(r_0)$$

となる [6]． $S$  が剛球の場合や  $f$  が音圧を表さない場合（カージオイド型の指向性をもつマイクロホンで収録した信号など）は異なる関係式で表される．

式 7 における球面  $S$  の特性の違いを吸収するフィルタとして，動径フィルタ (radial filter)  $d_n$  が導入される．開球面については，

$$d_n = \frac{1}{4\pi i^n j_n(kr_0)}$$

被展開関数  $f$  が音圧と音圧勾配の組み合わせである場合は

$$d_n = \frac{1}{2\pi i^n (j_n(kr_0) - i j_n'(kr_0))}$$

剛球面については

$$d_n = \frac{1}{4\pi i^n \left( j_n(kr_0) - \frac{j_n'(kr_0)}{h_n^{(2)}(kr_0)} h_n^{(2)}(kr_0) \right)}$$

となる．なお， $h_n^{(2)}$  は第 2 種球ハンケル関数である．

## 2.3 平面波分解

波動方程式の解の 1 つの形として平面波がある．平面波の和も波動方程式を満たすため，波動方程式を満たす任意の関数は平面波の線形結合で表せると考えられる．いま，連続面を考えると波動方程式を満たす関数  $f$  は以下のように書ける．

$$f(\omega, \Omega) = \int_S D(\omega, \Omega') P(\Omega', \Omega) dS \quad (8)$$

ここで， $P$  は角度  $\Omega$  で観測される入射角  $\Omega'$  の大きさ 1 の平面波を表し， $D$  は角周波数  $\omega$  における入射角  $\Omega'$  の大きさを表す．平面波  $P$  は開球面において以下のように球面調和展開できる．

$$P(\Omega', \Omega) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^n i^n j_n(kr) Y_n^m(\Omega) Y_n^{m*}(\Omega') \quad (9)$$

重み係数  $D$  を

$$D(\omega, \Omega') = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{1}{4\pi i^n j_n(kr)} P_{nm}(\omega) Y_n^m(\Omega') \quad (10)$$

のように与え、式 9 と式 10 を式 8 に代入すれば、結局

$$f(\omega, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n P_{nm}(\omega) Y_n^m(\Omega)$$

となり、式 6 と同じ式が得られる。 $D$  は  $f$  について各入射方向の平面波が与える寄与を表し、 $D$  を得る過程を平面波分解 (plane wave decomposition) と呼ぶ。 $f$  が開球面上にない場合、動径フィルタ  $d_n$  を用いて、

$$D(\omega, \Omega') = \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n(kr) P_{nm}(\omega) Y_n^m(\Omega')$$

のように一般化して表せる。

## 2.4 頭部伝達関数

頭部伝達関数はある方向から到来する平面波に対する耳の位置での特性とみなすことができる。すなわち人間が聴取する耳の位置での信号  $S^{l,r}$  は平面波分解  $D$  を用いて

$$S^{l,r} = \int_S H^{l,r}(\omega, \Omega) D(\omega, \Omega) d\Omega \quad (11)$$

で表せる。ここで  $D$  は球面上で定義される関数であり、その複素共役を球面調和展開すると展開係数  $(D^*)_{nm}$  は

$$\begin{aligned} (D^*)_{nm} &= \int_S D^*(\Omega) Y_n^m(\Omega)^* d\Omega \\ &= \int_S \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} d_n^*(kr) P_{n'm'}^* Y_{n'}^{m'}(\Omega)^* Y_n^m(\Omega)^* d\Omega \\ &= \int_S \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} d_n^*(kr) P_{n'm'}^* a_{m'} Y_{n'}^{-m'}(\Omega) Y_n^m(\Omega)^* d\Omega \end{aligned}$$

となる。ここで  $a_{m'}$  は用いる球面調和関数  $Y_n^m$  の形に依存する係数であり、複素非対称型を用いる場合  $a_{m'} = (-1)^{m'}$  となり、複素対象型を用いる場合は  $a_{m'} = 1$  となる。実数型を用いる場合は  $(D^*)_{nm}$  を求める過程自体が不要となる。球面調和関数の正規直交性より、

$$(D^*)_{nm} = d_n^*(kr) P_{n(-m)}^* a_m$$

を得る。式 11 に平面波分解の展開式および頭部伝達関数の球面調和展開を代入し、球面調和関数の正規直交性を再度利用すれば

$$\begin{aligned} S^{l,r} &= \int_S D(\omega, \Omega) H^{l,r}(\omega, \Omega) d\Omega \\ &= \int_S \left( \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n^*(kr) P_{n(-m)}^* Y_n^m(\Omega) a_m \right)^* \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} H_{n'm'}^{l,r} Y_{n'}^{m'}(\Omega) d\Omega \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n a_m P_{n(-m)} H_{nm}^{l,r} \end{aligned} \quad (12)$$

が得られる。したがって原音場の球面調和展開係数と頭部伝達関数の球面調和展開係数を計算することにより、対応するバイノーラル信号が得られることが分かる。球面調和関数に実数型を用いる場合は、球面調和関数の正規直交性を利用する際に複素共役を用いる必要がないため  $D$  の複素共役の展開係数を計算しなくてよく、この場合  $S^{l,r}$  は

$$\begin{aligned} S^{l,r} &= \int_S D(\omega, \Omega) H^{l,r}(\omega, \Omega) d\Omega \\ &= \int_S \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n P_{nm}(\omega) Y_n^m(\Omega') \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} H_{n'm'}^{l,r} Y_{n'}^{m'}(\Omega) d\Omega \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n P_{nm} H_{nm} \end{aligned} \quad (13)$$

として得られる。

バイノーラル信号の水平面上での回転は式 12 の展開係数に複素指数を掛けることによって実現できる。すなわち、音場に対して頭部を時計回りに角度  $\alpha$  だけ回転した場合のバイノーラル信号は

$$S^{l,r}(\alpha) = \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n a_m P_{n(-m)} H_{nm} e^{-jm\alpha} \quad (14)$$

として得られる。なお、式 14 導出は量子力学などで用いる Wigner の D 行列を用いており [4] 本資料の範囲を大きく超えるため省略する。

## 2.5 離散化

ここでは上記の連続信号での議論を、コンピュータ上でバイノーラル信号を得るために離散化する過程を示す。

球面調和展開係数を  $N$  次で打ち切ると、式 6 は以下のように書ける。

$$f(\Omega) = \sum_{n=0}^N \sum_{m=-n}^n P_{nm} Y_n^m(\Omega)$$

さらに、音場内に観測点を  $l$  個設定すれば上式は以下のように行列形式で表現できる。

$$\mathbf{Y}\mathbf{P} = \mathbf{F}$$

ここで、 $\mathbf{Y}$  は各測定点での角度および次数における球面調和関数の値を格納した  $l \times (N+1)^2$  行列、 $\mathbf{P}$  は球面調和展開係数を格納した  $(N+1)^2 \times 1$  ベクトル、そして  $\mathbf{F}$  は各観測点における音場を表す値（音圧に限らない）を格納する  $l \times 1$  行列である。

この連立方程式は  $\mathbf{P}$  を未知数としたとき、 $l > (N+1)^2$  の場合優決定系（overdetermined）となる。この場合、一般的には  $\mathbf{P}$  の代わりに  $|\mathbf{F} - \mathbf{Y}\hat{\mathbf{P}}|^2$  を最小にする最小二乗解  $\hat{\mathbf{P}}$  を求める。 $\hat{\mathbf{P}}$  は  $\mathbf{Y}$  の Moore-Penrose 擬似逆行列  $\mathbf{Y}^\dagger$  を用いて

$$\hat{\mathbf{P}} = \mathbf{Y}^\dagger \mathbf{F}$$

として求めることができる。観測点以外の音場内の点での再現性能を高めるため、Tikonov の正則化がしばしば用いられる。正則化を行う場合、最小化する目的関数は  $|\mathbf{F} - \mathbf{Y}\hat{\mathbf{P}}|^2 + \lambda|\mathbf{D}\hat{\mathbf{P}}|^2$  となる。ここで、 $\mathbf{D}$  は [5]

で提案されている，対角要素が  $d_{n,n} = 1 + n(n+1)$  となるような対角行列である． $\mathbf{D}$  をこのように定めることで高次の係数に低次の係数に比べて大きいペナルティを与えることになり，補間性能が向上することが報告されている [5]．

この最小化問題の解は以下のように与えられる．

$$\hat{\mathbf{P}} = (\mathbf{Y}^* \mathbf{Y} + \lambda^2 \mathbf{D})^{-1} \mathbf{Y}^* \mathbf{F}$$

同様に，頭部伝達関数に関しても球面調和展開係数  $H_{nm}^{l,r}$  を計算すれば，式 12 を用いてバイノーラル信号を計算することができる．

また [12] や [14] ではバイノーラル信号を  $\hat{b}_{l,r}(\omega) = \mathbf{w}_{l,r}^H(\omega) \mathbf{a}(\omega)$  のように，原音場の展開係数ベクトル  $\mathbf{a}(\omega)$  とバイノーラル変換ベクトル  $\mathbf{w}_{l,r}$  の内積の形で書かれている．

### 3 課題および改善手法

式 12 が示すように，球面調和展開を利用したバイノーラル信号の生成では原音場と頭部伝達関数の球面調和展開次数を一致させる必要がある．一般に，頭部伝達関数の測定点の数は片耳だけでも数百に及ぶが，原音場の収録においてそれほどの数のマイクロホンを用いたマイクロホンアレイを用意するのは非現実的であり，それに合わせて頭部伝達関数の展開次数も減らす必要がある．したがって，頭部伝達関数の展開次数の削減により定位性能や広がり感の低下 [2]，および高域での減衰が生じる [2, 3] ([3] の Fig.1 を参照)．

本章では上記の問題点の改善手法の概説を行う．

#### 3.1 Spherical Head Filter Method [3, 13]

上記の高域の減衰が生じる理由を述べるために，頭部を剛球で近似することを考える．以下，[3] での説明を参考にした．

半径  $r_0$  の剛球面上のある点  $\Omega$  に任意の方向  $\Omega_0$  から到来する単位振幅を持つ平面波による音圧は

$$p(kr_0, \Omega, \Omega_0) = \sum_{n=0}^N \sum_{m=-n}^n b_n(kr_0) [Y_n^m(\Omega_0)]^* Y_n^m(\Omega) \quad (15)$$

で与えられる．ここで  $b_n(kr_0)$  は放射関数 (radial function) であり，剛球では以下のように与えられる．

$$b_n(kr_0) = 4\pi i^n \left[ j_n(kr_0) - \frac{j_n(kr_0)}{h_n'(kr_0)} h_n(kr_0) \right]$$

拡散音場を仮定すれば，球面上のある点の音圧は球面上の音圧の平均で与えられると考えられる．したがって球面上のある点  $\Omega$  での音圧は

$$\bar{p}(kr_0, \Omega) = \sqrt{\frac{1}{4\pi} \int_{\Omega_0 \in S^2} |p(kr_0, \Omega, \Omega_0)|^2 d\Omega_0}$$

式 15 を代入して整理すると，以下のように点  $\Omega$  での音圧が球面調和展開次数  $N$  の関数として得られる．

$$\bar{p}(kr_0)|_N = \frac{1}{4\pi} \sqrt{\sum_{n=0}^N (2n+1) |b_n(kr_0)|^2} \quad (16)$$

図1に式16を次数  $N$  ごとにプロットしたものを示す．球面調和展開次数  $N$  が低くなるほど，高域の減衰が著しくなることがわかる．

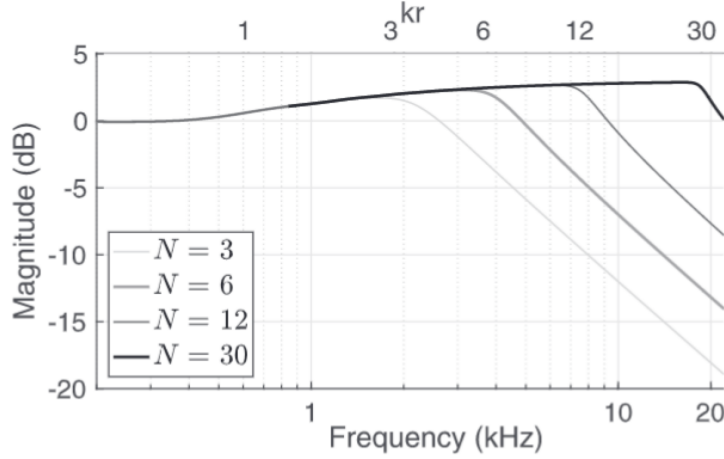


図1 球面調和展開次数  $N$  と剛球上の音圧（式16）の関係（[3]のFig.2より引用）

[13]ではこの高域での減衰を補正するためのフィルタが提案されている．具体的には，低い展開次数  $N$  によって生成されたバイノーラル信号を補正するためのフィルタは以下のように与えられる．

$$H(k)|_{N \rightarrow N_h} = \frac{\overline{p(kr)_{N_h}}}{p(kr)_N}$$

### 3.2 High-Frequency Time-Aligned Binaural Decoding (TAC)[12, 14]

[14]で示されているように，音場の高周波成分のエネルギーは球面調和展開したときの高次の展開係数に含まれている．その様子を図2に示す．図2はある頭部伝達関数データセットを球面調和展開したときの展開次数ごとの正規化エネルギーをプロットしており，低次の展開係数では高周波成分のエネルギーを含みきれておらず，高域のエネルギーを十分に再現するには展開次数を30まで上げる必要があることがわかる．

二重理論（duplex theory）によれば，ヒトが両耳間時間差（ITD）をほとんど低域でのみ音源定位の手がかりとして用いている [11]．そこで Zaunschirm et al.[14]は高域での各周波数の位相を均一化し，必要な展開次数を削減する手法を提案した．位相均一化処理は以下のように定式化される．

$$h_{l,r}(\Omega_p, \omega) = h^{l,r}(\Omega_p, \omega) A_p^{l,r}(\omega)$$

オールパスフィルタ  $A_p^{l,r}(\omega)$  の周波数特性は

$$A_p^{l,r}(\omega) = \begin{cases} 1 & \text{for } \omega < \omega_c \\ e^{-i(\omega - \omega_c)\tau_p^{l,r}} & \text{for } \omega \geq \omega_c \end{cases}$$

$\omega_c = 2\pi f_c$  はカットオフ角周波数， $\tau_p^{l,r}$  は半径  $r_H$  の頭部の中心に対する点  $p$ （耳の位置）での信号の位相差である．

$$\tau_p^r = \cos(\theta_p) \sin(\phi_p) r_H c^{-1}, \quad \tau_p^l = -\tau_p^r$$

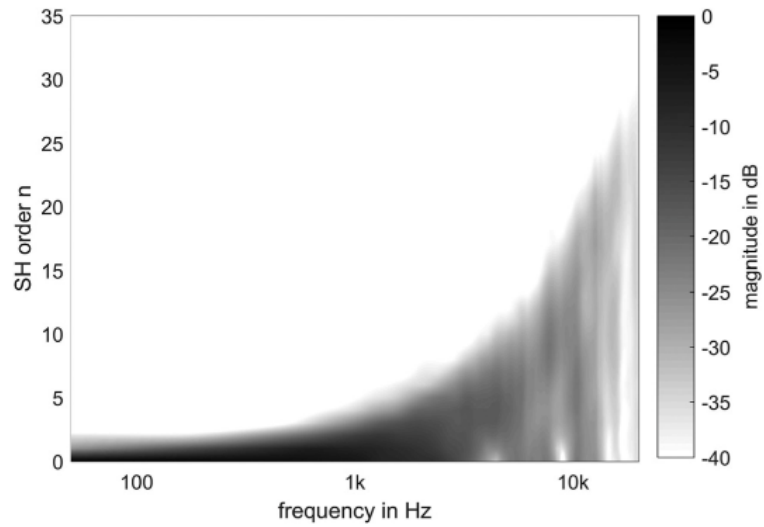


図 2 ある頭部伝達関数データセットを球面調和展開したときの展開次数ごとの正規化エネルギーをプロットしたグラフ ([14] の Fig.1 より引用)

図 3 に TAC 適応後の頭部伝達関数の展開次数ごとのエネルギー分布を示す．図 2 と比較すると，高域でのエネルギーが低次元の展開次数に集中しており，展開次数を 15 ほど取れば TAC 適応後の頭部伝達関数の高域のエネルギーを再現できることがわかる．

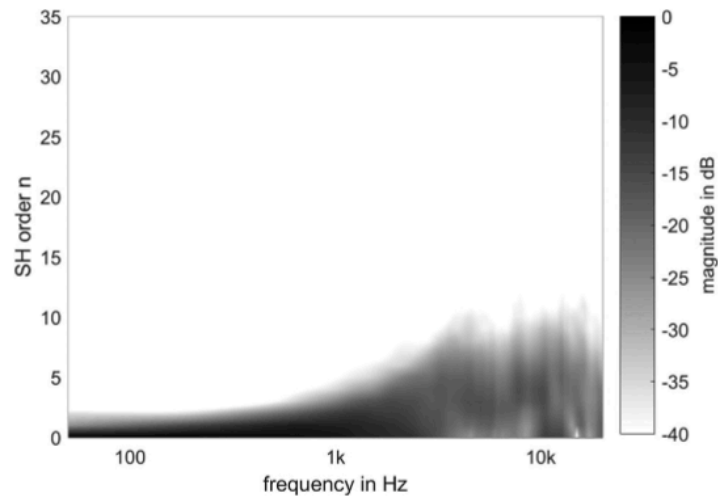


図 3 カットオフ周波数 1.5kHz で TAC を適応した頭部伝達関数データセットに対する展開次数ごとの正規化エネルギーをプロットしたグラフ ([14] の Fig.4 より引用)

[12] によれば，カットオフ周波数は 2kHz ほどあれば主観的にも十分な結果が得られる．



## 参考文献

- [1] Carl Andersson. Headphone auralization of acoustic spaces recorded with spherical microphone arrays. Master's thesis, Chalmers University of Technology.
- [2] Amir Avni, Jens Ahrens, Matthias Geier, Sascha Spors, Hagen Wierstorf, and Boaz Rafaely. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *J. Acoust. Soc. Am.*, 133(5):2711–2721, May 2013.
- [3] Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, and Boaz Rafaely. Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *J. Acoust. Soc. Am.*, 141(6):4087, June 2017.
- [4] Benjamin Bernschütz. *Microphone arrays and sound field decomposition for dynamic binaural recording*. PhD thesis, 2016.
- [5] R Duraiswaini, D N Zotkin, and N A Gumerov. Interpolation and range extrapolation of HRTFs [head related transfer functions]. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages iv–iv, May 2004.
- [6] E.G. Williams. *Fourier acoustics – Sound radiation and nearfield acoustical holography*. Academic Press, 1999.
- [7] Jean-Marc Jot, Veronique Larcher, and Jean-Marie Pernaux. A comparative study of 3-D audio encoding and rendering techniques. Audio Engineering Society, March 1999.
- [8] Markus Noisternig, Alois Sontacchi, Thomas Musil, and Robert Holdrich. A 3D ambisonic based binaural sound reproduction system. In *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Audio Engineering Society, June 2003.
- [9] Makoto Otani, Haruki Shigetani, Masataka Mitsuishi, and Ryo Matsuda. Binaural ambisonics: Its optimization and applications for auralization. *Acoust. Sci. Technol.*, 41(1):142–150, 2020.
- [10] Boaz Rafaely and Amir Avni. Interaural cross correlation in a sound field represented by spherical harmonics. *J. Acoust. Soc. Am.*, 127(2):823–828, February 2010.
- [11] A. Roginska and P. Geluso. *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Audio Engineering Society Presents. Taylor & Francis, 2017. ISBN 9781317480105. URL <https://books.google.co.jp/books?id=elk6DwAAQBAJ>.
- [12] Christian Schörkhuber, Markus Zaunschirm, and Robert Höldrich. Binaural rendering of ambisonic signals via magnitude least squares. In *Proceedings of the DAGA*, volume 44, pages 339–342, 2018.
- [13] Jonathan Sheaffer, Shahar Villeval, and Boaz Rafaely. Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction. April 2014.
- [14] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich. Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. *J. Acoust. Soc. Am.*, 143(6):3616, June 2018.