小特集—マイクロホンアレイの新しい技術展開—

SENZI：球状マイクロホンアレイを用いた3次元音空間収音再生*

坂本修一 (東北大学電気通信研究所)**

43.60.+d; 43.38.Kb; 43.60.Fg

1. はじめに

遠隔地の音空間情報を，音源の位置や場の広がり感などを含め，精度高く収音・再生する技術は，高次臨場感通信の核となる極めて重要な技術である。特に頭部伝達関数 (Head Related Transfer Function：HRTF) を規範とした，ヘッドホンベースの音空間再生方式は，クロストークの影響を考慮しなくてもよいため，単純なシステムで高精度な音空間を提示することが可能な方式として期待されている。これまでにも，ダミーヘッドによるバイノーラル収音・再生技術といった古典的な手法から，聴取者と同じ頭部形状を有する可動式ダミーヘッド *TeleHead* [1, 2]，球状に付置したマイクロホン群と，聴取者頭部に設置した位置センサから構成され，聴取者の頭部運動に応じて収音するマイクロホンを切り替えて聴取者に提示する Motion-Tracked Binaural (MTB) [3, 4] といった様々な収音手法が提案されている。しかしいずれの手法を用いても，とある一点の聴取点における音空間情報を，時間，空間を超えて，聴取者の頭部運動を含めて精度よく再現することができていない。

近年，マイクロホンの小型化，ディジタル化が進み，多チャネルマイクロホンを有するマイクロホンアレイの構築が盛んに行われるようになっており，実際に 32 チャネルのマイクロホンを有する球形アレイが市販されるに至っている (例えば，[5, 6])。特に球状に等密度にマイクロホンを配置した球状マイクロホンアレイは球面調和展開との親和性も高いことから，High-order Ambisonics (HOA) [7, 8] の収音用システムとして使われ始めている。更に，HOA については，多チャネルスピーカアレイを想定したもののほか，ヘッドホンを再生系として想定した binaural Ambisonics [9, 10] も提案されており，いまだ不十分である再現音空間の精度向上を目指すべく，現在も研究が続けられている。

我々は，高精度の音空間収音・再生技術の実現に向け，多数のマイクロホンを設置した球状マイクロホンアレイを用いた音空間収音再生手法 SENZI (Symmetrical object with ENchased ZIllion microphones) を提案してきた [11]。SENZI では，収音に際し，遠隔地あるいは録音とは異なる再生時の聴取者の頭部の動きに合わせ，マイクロホンに入力された信号の加算方法を変化させることにより，聴取者個人の HRTF に合うように信号処理を行うことで，システムを固定したままで聴取者の聴感に合致した音を提示できるという特徴を持つ。すなわち，この手法は，とある音空間を時間，空間を超えて不特定の聴取者に対して同時に提示することが可能な収音手法である。

本報告では，本提案手法のアルゴリズムを説明すると共に，SENZI を 252 チャネルのマイクロホンを有する球状アレイで実現したシステムの概要を紹介する。合わせて，今後の展開について述べる。

2. SENZI の概要 [11]

2.1 全体像の概略

SENZI の概要図を **Fig. 1** に示す。多数のマイクロホンを持った頭部モデルを用いて収音を行い，各マイクロホンに入力された音に基づいて，音がどの方向から到来しているかを聴取者が正確に判断できるよう信号処理し，聴取者に提示する。例えば，図のように頭部モデルを劇場に設置し多数のマイクロホンの入力を適切に信号処理し，遠隔地にいる聴取者に提示することにより，聴取者はあたかも劇場にいるかのような臨場感のある音を聴取することができる。

* SENZI: 3D sound-space recording and reproduction method based on spherical microphone array.

** Shuichi Sakamoto (Research Institute of Electrical Commucanition, Tohoku University, Sendai, 980–8577) e-mail: saka@ais.riec.tohoku.ac.jp

Fig. 1 Concept of SENZI.

SENZI は「収音部」「信号処理部」「再生部」から構成されている。収音部は多数のマイクロホンを配置した球状マイクロホンアレイであり，これらの入力を用いて信号処理を行う。なお，左右方向の頭部運動に対応するため，収音部は軸対称型となる。信号処理部では，マイクロホンアレイからの入力信号から各聴取者の HRTF に基づいて提示すべき音を合成する処理を行う。最後に再生部では，信号処理された音を聴取者に提示する。再生手法としてはヘッドホンを用いたバイノーラル再生手法を想定している。

2.2 アルゴリズムの概要

説明を容易とするため，水平方向のみを考慮した際の具体的な合成手法を述べる。なお，仰角方向も含めた合成も同様の手法により実現可能である。聴取者の片耳における HRTF を $\boldsymbol{H}_{\mathrm{lis}}$ と表す。HRTF は通常周波数の関数として表記されることが多いのに対して，本手法では，方向の関数として HRTF を考える。

n チャネルのマイクロホンを配置した球状マイクロホンアレイを用いて音空間の収音を行った場合を考える。このとき，収音する音空間について，音が到来すると想定した方向の要素数を m，その個々の方向を $\boldsymbol{\Theta}$ とする。ここで，とある周波数 f において，$\boldsymbol{H}_{\mathrm{lis},f}$ は，以下の式のように表すことができる。

$$\begin{aligned}\boldsymbol{H}_{\mathrm{lis},f} &= \begin{bmatrix} H_{\mathrm{lis},f}(\Theta_1) \\ \vdots \\ H_{\mathrm{lis},f}(\Theta_m) \end{bmatrix} \\ &= \begin{bmatrix} H_{1,f}(\Theta_1) & \cdots & H_{n,f}(\Theta_1) \\ \vdots & \ddots & \vdots \\ H_{1,f}(\Theta_m) & \cdots & H_{n,f}(\Theta_m) \end{bmatrix} \begin{bmatrix} z_{1,f} \\ \vdots \\ z_{n,f} \end{bmatrix}\end{aligned} \tag{1}$$

ここで，$H_{i,f}(\Theta_j)$ は周波数 f で j 番目の想定方向 Θ_j から，i 番目のマイクロホンまでの伝達関数，いうなればオブジェクト伝達関数である。$z_{i,f}$ は，周波数 f における i 番目のマイクロホンに対する重み係数を表している。式 (1) を見ると分かるように，各マイクロホンの位置における伝達関数と，音が到来すると想定した方向に依存しない $z_{i,f}$ から聴取者の HRTF を合成することにより，音の到来方向を知る必要なく各聴取者に最適な音情報を提示することが可能となる。

頭部回転を含め 4π 空間すべての方向から到来する音すべてを想定することを考えると，その要素数 m はマイクロホンのチャネル数 n よりも大きくなる場合がほとんどである。この場合，式 (1) は優決定問題となるため，下記のような式として表されることとなる。そこで，残差 ϵ を最小にすべく非線形最小自乗法や擬似逆行列を用いることで，$\hat{\boldsymbol{z}}_f$ を算出する。

$$\begin{aligned}&\boldsymbol{H}_{\mathrm{lis},f} = \boldsymbol{H}_f \cdot \hat{\boldsymbol{z}}_f + \boldsymbol{\epsilon}, \qquad (2)\\ &\boldsymbol{H}_f = \begin{bmatrix} \boldsymbol{H}_{1,f} & \cdots & \boldsymbol{H}_{n,f} \end{bmatrix}, \\ &\boldsymbol{H}_{i,f} = \begin{bmatrix} H_{i,f}(\Theta_1) & \cdots & H_{i,f}(\Theta_m) \end{bmatrix}^{\mathrm{T}}, \\ &\hat{\boldsymbol{z}}_f = \begin{bmatrix} \hat{z}_{1,f} & \cdots & \hat{z}_{n,f} \end{bmatrix}^{\mathrm{T}}\end{aligned}$$

$\hat{z}_{i,f}$ は，周波数 f，マイクロホン i に決定される重み係数で，音の到来方向に依存しない複素数である。なお，反射や残響が存在する環境においては，それぞれの反射音や残響音を Θ_j の方向から到来する音と見立てることで，本手法を用いて音空間を精度高く再現することが可能となる。

以上まとめると，SENZI は，頭部回転，反射音，残響音を含めて音が到来すると想定した方向からの m 個の伝達関数群を，密に配置したより多くの n チャネル（ただし，通常は $m > n$）のマイクロホンの位置におけるオブジェクト伝達関数の複素線形和として求める処理である。

2.3 聴取者の頭部運動への対応

Fig. 2 に示すような状態を考える。まず，頭部運動角を，振り向き方向の頭部運動が起こったときの聴取者の正面方向と 0° とのなす角と定義する。このとき，頭部運動がない状況での聴取者の正面方向を 0° とし，合成対象 HRTF とする耳とは反対方向（Fig. 2 では時計回り）を正の方向とする。一方，音源の位置を示す水平角の座標系は

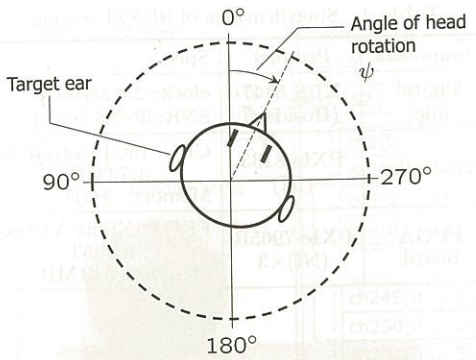

Fig. 2 Assumed scenario of the listener's head rotation.

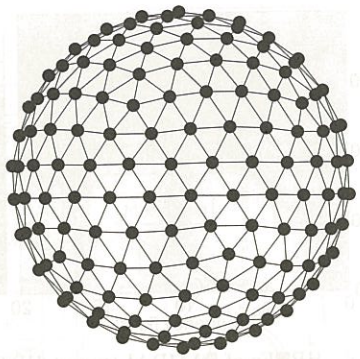

Fig. 3 Microphone arrangement.

今までと同様聴取者の正面方向を 0° とし，合成対象 HRTF とする耳方向回り（Fig. 2 では反時計回り）を正の方向とする。

振り向き方向の頭部運動が起こった場合では，HRTF において水平角方向が変化することになる。つまり，頭部運動角の角度分だけ音源位置を示す座標系における水平角方向にずらした HRTF を，目的対象の HRTF として合成を行えば，頭部運動が起こった場合の合成を行えることになる。具体的に説明する。頭部運動角を ψ とすると，音源の位置を示す座標系における $(360-\psi)°$ が聴取者にとっての正面方向となる。そのため Fig. 2 の座標の 0° に音源がある場合には，聴取者にとっては，ψ の位置に音源があるということに外ならない。すなわち，音源が 0° にあるときには，音源の位置を示す座標系において水平角 ψ に音源が存在する際の HRTF を合成対象の HRTF として用いれば，0° の位置に音源があるという状況を聴取者に提示することができる。このことから，音源が 0° にあるときには聴取者の HRTF の水平角 ψ のデータが対応することになる。従って，音源 0° を聴取者 HRTFψ，音源 5° を聴取者 $\psi+5°$，音源 10° を聴取者 $\psi+10°$，音源 15° を聴取者 $\psi+15°$，…，のように対応を変えることとなる。つまり，合成時に使用する HRTF データを ψ 分だけ全体的にずらせばよいということになる。

垂直方向の頭部運動に対しても同様の処理により，頭部運動を反映させた音空間提示が可能となる。ちなみに，収音部の形状は球状であり，頭部運動に対しては，聴取者と収音部との位置関係が周期的に現れることになる。従って，上記処理はすべての角度で行う必要はない。

3. SENZI により再現される音空間の精度 [12]

前章で示したような手法により再現される音空間の精度を計算機シミュレーションにより検証した。今回考察に使用した球状マイクロホンアレイのマイクロホン配置は **Fig. 3** のとおりである。聴取者の頭部サイズを想定して直径を 17 cm とした球状オブジェクトに，252 チャネルのマイクロホンを設置している。設置したマイクロホンの位置は，正 20 面体の各面を 25 等分割し，すべての頂点が直径 17 cm の球状オブジェクトに内接するように座標変換を行うことで算出した [13]。この操作の結果，各マイクロホンの間隔はおおむね一定となり，約 2 cm であった。これは空間的折り返しひずみの影響が，約 8.5 kHz 以上の周波数で見られることを意味している。

今回設計したマイクロホンアレイの配置図と同様に，直径 17 cm の剛球の表面上に 252 チャネルのマイクロホンを配置した。収音・再現すべき想定到来音として，半径 1.5 m の球面上にあり，Θ_j 方向から到来する音を考えた。各到来音の方向 Θ_j は，マイクロホン配置と同じ要領で各面を 256 等分割し，すべての頂点が半径 1.5 m の球面上に内接するように座標変換を行うことで求めた。算出された方向の要素数は 2,562 となる。

合成対象となるダミーヘッド（SAMRAI，高研）の左耳の HRTF を **Fig. 4** に，合成結果の HRTF を **Fig. 5** に示す。また，SAMRAI の HRTF と合成 HRTF とのスペクトルひずみ（Spectral Distortion：SD）を **Fig. 6** に示す。Fig. 4，Fig. 5 を見ると，空間的な折り返しひずみの影響が見られなくなる 10 kHz 以下のあたりでは，HRTF のピー

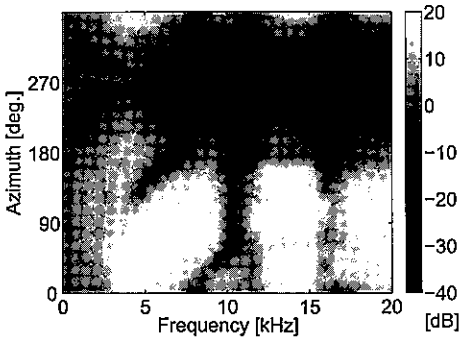

Fig. 4 HRTFs of SAMRAI (target HRTFs).

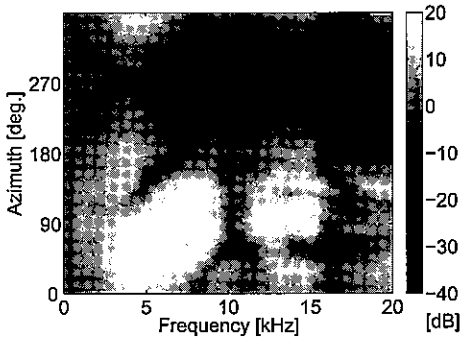

Fig. 5 Transfer functions synthesized by SENZI.

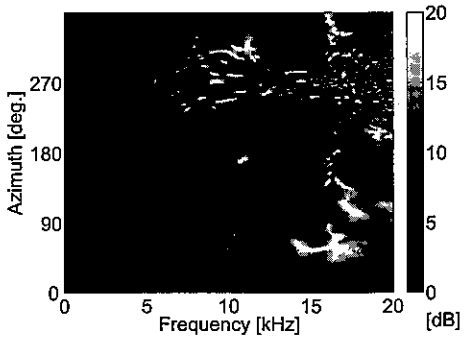

Fig. 6 Spectral distortion between target HRTFs and synthesized HRTFs.

クやディップが精度よく再現できていることが分かる。これらの結果は，今回構築した SENZI を用いることで，音空間の収音が十分可能であることを示している。使用するマイクロホンのチャネル数を増やし，間隔を密にすることで，更なる再現精度の向上が見込まれる。

4. リアルタイム SENZI の構築

前章までの結果に基づき，リアルタイム動作するシステムとして構築した SENZI について説明する。**Table 1** にシステム構築時に使用した機器の概要を示す。また，「収音部」として構築した球状マイクロホンアレイを **Fig. 7** に示す。球状マイクロホンアレイは光造形法を用いエポキシ樹脂により構築した球状のオブジェクトに ECM (Electric Condenser Microphone) マイクロホン (KUS5147，ホシデン) を前章で計算した位置に 252 チャネル配置してる。球状マイクロホンアレイは，直径が 0.2 cm の 5 本の支柱で支えられ，制御用コンピュータに接続される。量子化ビット数が 16 bit，サンプリング周波数が 48 kHz で，252 チャネルの同時収音が可能となっている。構築したシステムの全体像を **Fig. 8** に示す。

現状では頭部回転には未対応であるが，位置センサ等から取得した頭部の位置情報をシステムに入力し，対応する重み係数 $z_{i,f}$ を動的に切り替えることで，聴取者の頭部回転を反映した音空間を提示するように実装する予定である。

Table 1 Specifications of SENZI system.

Component	Product	Spec.
Digital mic.	KUS 5147 (Hosiden)	clock: 2.4 MHz SNR: 58 dB (typ.)
Controller	PXIe-8133 (NI)	CPU: Intel Core i7-820 1.73 GHz Memory: 3 GB
FPGA board	PXIe-7965R (NI)×3	FPGA: Xilinx Virtex-5 SX95T Memory: 512 MB

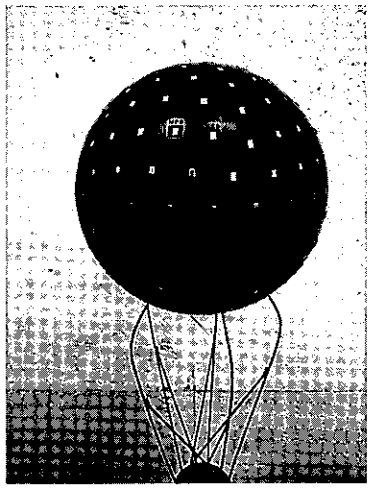

Fig. 7 Photograph of constracted SENZI (AMS-251S, System Keisoku Co. Ltd.).

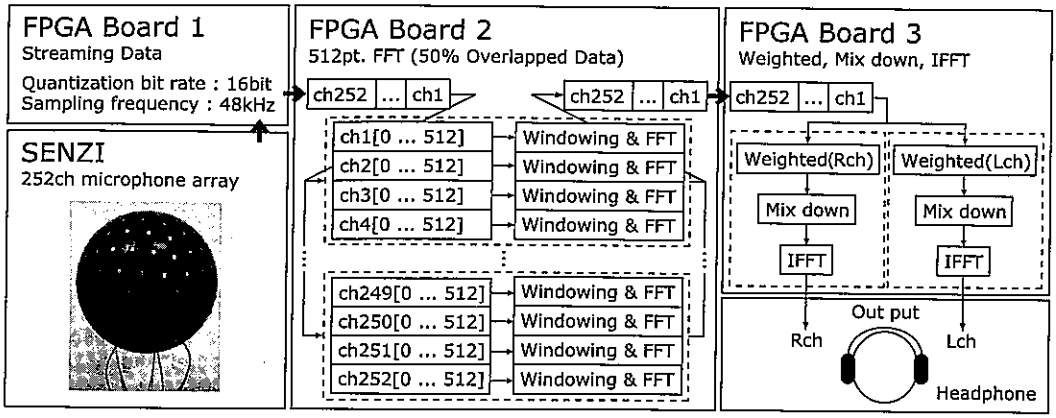

Fig. 8 Implemented SENZI system.

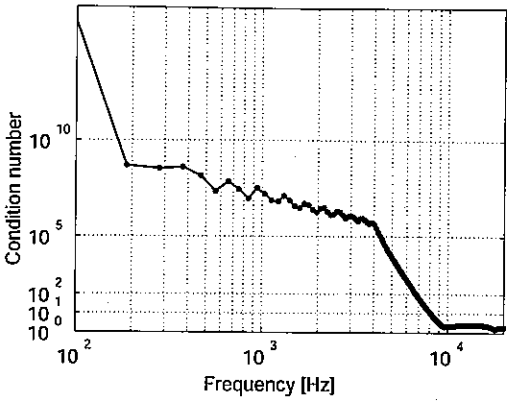

Fig. 9 Condition number of $\boldsymbol{H}_f$ as a function of frequency.

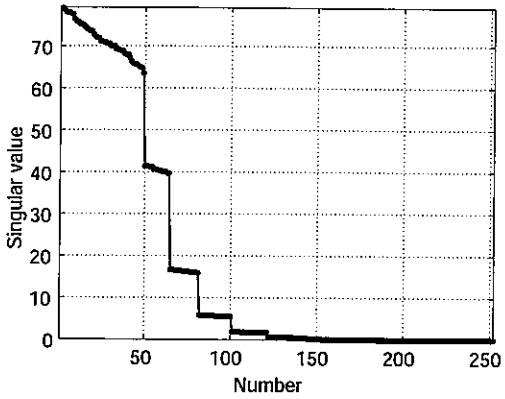

Fig. 10 Distribution of absolute singular value at 4 kHz.

5. リアルタイムSENZIによるにより再現される音空間の精度向上に向けて

音空間を高周波数まで精度高く収音再生するためには，マイクロホンを超密に等密度に配置する必要がある。その一方で，隣接するマイクロホン間の間隔が狭くなると，マイクロホンに入力される信号が似通ってきてしまうため，特に低い周波数において，$\hat{z}_{i,f}$ を求める $\boldsymbol{H}_f$ が悪条件となってしまう。その結果，マイクロホンの内部雑音や，設計時と実際のマイクロホンの位置のズレといった要因が再現される音空間に大きな悪影響を及ぼすことになる。

Fig. 9 に，各周波数における $\boldsymbol{H}_f$ の条件数を示す。この図から，条件数が低周波数領域ほど大きくなっていることが確認できる。特に4kHz以下の領域では，条件数が非常に大きくなっており，再現される音空間が様々な要因により劣化してしまうことになる。更に，**Fig. 10** に，4kHzにおける $\boldsymbol{H}_f$ の特異値の絶対値の分布を示す。図から明らかなように，絶対値が0に近い特異値が多数存在していることが見て取れる。

従って，条件数を基準としてある一定以上の特異値のみを使用するようにして $\hat{z}_{i,f}$ を求めることで，再現される音空間の精度を向上させることが可能となる。これは，低周波数領域に関して，実質的に使用するマイクロホン数を減らすこととなり，使用するマイクロホンの間隔が広がることにより，マイクロホンごとに入力される音信号の差が大きくなることを意味する。

6. 最 後 に

本報告では，我々が提案する球状マイクロホンアレイを用いた3次元音空間収音再生手法SENZI

について述べた。SENZI は極めて単純なアルゴリズムでありながら，個々の聴取者に時間・空間を超えて音空間を高精度に再現することができる手法である。

しかし，実際にシステムとして構築する際にはまだ幾つか課題が残っている。例えば，多チャネルマイクロホンで収音した音をどのように遠隔地にネットワーク配信するかといった点や，再現される音空間の精度を保ちつつ，いかにマイクロホンを減らしていくかといった点が検討項目として考えられる。更には，単に擬似逆行列や非線形最小自乗法で求めている $\hat{z}_{i,f}$ の算出法に関しても，人間の聴覚特性を生かした形で必要とされる精度を重み付けして $\hat{z}_{i,f}$ を求めるといった手法も考えられる。そのためには，構築システムを使った聴感評価が必須であり，今後，重点的に取り組んでいく予定である。

謝　辞

本研究を進めるにあたり，日々ご議論いただく，仙台高等専門学校本郷哲教授，(独) 情報通信研究機構岡本拓磨博士，東北学院大学岩谷幸雄教授，東北大学電気通信研究所鈴木陽一教授に深く感謝する。また，SENZI の精度検証，システム構築において多大なご協力をいただいた，門井涼氏，小玉純一氏，松永純平氏に，感謝する。

HRTF の合成方法に関して信州大学大谷真准教授に貴重なご意見をいただいた。球状マイクロホンアレイを構築するにあたり，ECM マイクロホン (KUS5147) をホシデン株式会社の滋野安広氏からご提供いただいた。

本研究の一部は，戦略的情報通信研究開発推進制度 (SCOPE) No. 082102005，東北大学電気情報系 GCOE プログラム「情報エレクトロニクスシステム教育研究拠点 (CERIES)」，及び，日本学術振興会日中韓フォーサイト事業の補助による。

文　献

[1] I. Toshima, H. Uematsu and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: *TeleHead*," *Acoust. Sci. & Tech.*, 24, 327–329 (2003).
[2] I. Toshima, S. Aoki and T. Hirahara, "Sound localization using an auditory telepresence robot: TeleHead II," *Presence*, 17, 392–404 (2008).
[3] V.R. Algazi, R.O. Duda and D.M. Thompson, "Motion-tracked binaural sound," *J. Audio Eng. Soc.*, 52, 1142–1156 (2004).
[4] V.R. Algazi, R.O. Duda, J.B. Melick and D.M. Thompson, "Customization for personalized rendering of motion-tracked binaural sound," *Proc. 117th AES Convention*, 6225 (2004).
[5] mh acoustics, `http://www.mhacoustics.com/`
[6] Brüel & Kjær, Type 8606, `http://www.bksv.com`
[7] D.H. Cooper and T. Shiga, "Discrete-matrix multichannel stereo," *J. Audio Eng. Soc.*, 20, 346–360 (1972).
[8] R. Nicol and M. Emerit, "3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic," *Proc. AES 16th Int. Conf.*, 16, 436–453 (1999).
[9] M. Noisternig, A. Sontacchi, T. Musil and R. Höldrich, "A 3D ambisonic based binaural sound reproduction system," *AES 24th Int. Conf.: Multichannel Audio, The New Reality*, `http://www.aes.org/e-lib/browse.cfm?elib=12314` (2003).
[10] L.S. Davis, R. Duraiswami, E. Grassi, N.A. Gumerov, Z. Li and D.N. Zotkin, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," *119th Convention of the Audio Engineering Society*, `http://www.aes.org/e-lib/browse.cfm?elib=13369` (2005).
[11] S. Sakamoto, R. Kadoi, S. Hongo and Y. Suzuki, "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," *Proc. 2nd Int. Symp. Universal Communication (ISUC 2008)*, pp. 429–434 (2008).
[12] S. Sakamoto, J. Kodama, S. Hongo, T. Okamoto, Y. Iwaya and Y. Suzuki, "Effects of microphone arrangements on the accuracy of a spherical microphone array (SENZI) in acquiring high-definition 3D sound space information," in *Principles and Applications of Spatial Hearing* (World Scientific, Singapore, 2011), pp. 314–323.
[13] R. Sadourny, A. Arakawa and Y. Mintz, "Integration of the nondivergent barotropic vorticity equation with an icosahedral-hexagonal grid for the sphere," *Mon. Wea. Rev.*, 96, 351–356 (1968).