

## 複数の球状マイクロフォンアレイによる音源分離性能の基礎的検討

愛知 秀斗<sup>†</sup> ブラチリウス ビルギリウス<sup>†</sup> トレビーニョ ホルヘ<sup>†</sup>

サルバドル セザル<sup>†</sup> 坂本 修一<sup>†</sup> 鈴木 陽一<sup>†</sup>

吉川 教治<sup>††</sup> 山崎 隆志<sup>††</sup> 樹所 賢一<sup>††</sup>

<sup>†</sup> 東北大学電気通信研究所/大学院情報科学研究科 〒980-8577 宮城県仙台市青葉区片平 2-1-1

<sup>††</sup> リオン株式会社 〒185-8533 東京都国分寺市東元町 3-20-41

E-mail: †{aichi@dc, virgis@ais.riec, jorge@ais.riec, salvador@ais.riec, saka@ais.riec, yoh@riec}.tohoku.ac.jp,  
††{yoshikawa, yamasaki, kidokoro}@rion.ac.jp

あらまし 複数のマイクロフォンアレイを用いることで1つのアレイを用いたときよりも音源分離の精度が向上することが報告されている。本研究では、我々がそのような方法の一つとして先に提案した、2つの球状マイクロフォンアレイを1つのシステムとみなして收音する音源分離手法の性能評価を行った。実験はマイクロフォンの校正と音源分離性能の評価の順に行った。まず校正用スピーカから発した TSP 信号によるマイクロフォン出力から、マイクロフォンごとの校正フィルタを作成した。その後、目的音源と2つの妨害音源を用い、目的音源と2つのアレイとの角度を変化させて、各音源から白色雑音を出力したときの音源分離性能の測定、比較を行った。その結果、60°のときに、30°と比べて良好な音源分離を行うことができた。

キーワード 球状マイクロフォンアレイ, 音源分離, ビームフォーミング, アレイオブアレイズ

### Performance of sound source separation with array of arrays

Hideto AICHI<sup>†</sup>, Virgilijus BRACIULIS<sup>†</sup>, Jorge TREVINO<sup>†</sup>,

Cesar SALVADOR<sup>†</sup>, Shuichi SAKAMOTO<sup>†</sup>, Yôiti SUZUKI<sup>†</sup>,

Kyoji YOSIKAWA<sup>††</sup>, Takashi YAMASAKI<sup>††</sup>, and Kenichi KIDOKORO<sup>††</sup>

<sup>†</sup> Research Institute of Electrical Communication and Graduate School of Information Sciences, Tohoku University 2-1-1 Katahira, Aoba-ku, Sendai-shi, Miyagi, 980-8577 Japan

<sup>††</sup> RION Co., Ltd. 3-20-41 Higashimotomachi, Kokubunji-shi, Tokyo, 185-8533 Japan

E-mail: †{aichi@dc, virgis@ais.riec, jorge@ais.riec, salvador@ais.riec, saka@ais.riec, yoh@riec}.tohoku.ac.jp,  
††{yoshikawa, yamasaki, kidokoro}@rion.ac.jp

**Abstract** Recent studies in sound source separation using multiple microphones show that an array-of-arrays recording system can outperform individual arrays. The present study considers a previously proposed algorithm, in which multiple arrays are treated as a unified system. The technique is evaluated with a system comprising two spherical microphone arrays. Microphones calibration is attained through inverse filters derived from actual recordings. A sound source separation experiment shows performance is highest when the arrays lie at a separation of 60° as seen from the target sound source.

**Key words** spherical microphone array, spherical acoustics, sound source separation, beamforming, distributed arrays

## 1. はじめに

我々の身の回りには様々な音が混在している。それらの音には会話や楽曲のように我々が必要とする音だけではなく、不要な音も含まれている。そのため特定の相手が話している内容のみを聞き取りたい、曲の中で特定の楽器の音のみを抽出したいというような状況はしばしば訪れる。マイクロフォンアレイは、特定の位置で発生した音のみを強調し、それ以外の音を弱める処理を柔軟に行うことができることから、この課題の対処に有効である。

マイクロフォンアレイのマイクロフォン配置は、目的や設置環境に応じて、直線状アレイ、平面アレイ、包囲型アレイなど様々である [1, 2]。空間音響情報の取得や解析には、マイクロフォンを剛球の表面に配置した球状マイクロフォンアレイがよく用いられる [3]。その理由としては、球の形の対称性からアレイ周りで起こる音響散乱を球面調和解析等を用いて解析的に表せること [4]、球が人間の頭の形に近いことから音響散乱の大局的特性が類似していること [3-5] などが挙げられる [6]。これらの特性を生かした形で、モーダルビームフォーミングと呼ばれる球状マイクロフォンアレイを用いた音源分離技術が開発されている [7]。この技術を用いることで、どの方向にも均一なビームを向けての音源分離が可能となる。

しかし、球状アレイを1つだけのみを用いた場合、必要とする音(目的音)と不要な音(妨害音)がアレイからみていずれも同じ方向の場合に音源分離性能がほとんど期待できないという問題点を抱えている。このような場合にも良好な音源分離性能を実現するため、著者らの研究グループでは、複数のマイクロフォンアレイを一つのシステムと見なす、アレイオブアレイズ技術を用いることを提案している [8]。この提案では、方向ではなく領域ごとに良好な音源分離性能を得る信号処理法を示している。

本研究では、上記の提案について、2つの球状マイクロフォンアレイを用いて実際に録音した信号に対し、各アレイと目的音源との角度を変化させながら提案法による音源分離を試みた。本文では、実験で得られた音源分離性能について、計算機シミュレーションによる理論値との比較も含めて報告する。

## 2. 球状マイクロフォンアレイを用いた音源分離手法

### 2.1 球状マイクロフォンアレイ

球状マイクロフォンアレイとは、複数のマイクロフォンを球体の表面に規則的に配置して構成される收音システムである。極座標系を用いることでマイクロフォンアレイを容易にモデル化することができる。極座標系を図1に示す。マイクロフォンアレイの中心を原点とした場合、方位角  $\theta$  と仰角  $\phi$  により原点からの方向を決定することができ、さらに  $r$  を決めることで空間位置を特定することができる。

剛球と仮定したアレイ表面のマイクロフォンにより観測される点音源の音圧は以下の式により与えられる。

$$p_{\text{mic}}(k, \vec{r}_{\text{src}}) = p_{\text{src}}(k) \sum_{n=0}^{\infty} (2n+1) \frac{h_n(k, r_{\text{src}})}{h'_n(kr_{\text{mic}})} P_n \left( \frac{\vec{r}_{\text{mic}} \cdot \vec{r}_{\text{src}}}{|\vec{r}_{\text{mic}}| |\vec{r}_{\text{src}}|} \right) \quad (1)$$

式(1)において、 $p_{\text{mic}}$  はアレイ表面の  $\vec{r}_{\text{mic}}$  に設定されたマイクロフォンによって観測される音圧であり、 $p_{\text{src}}$  は  $\vec{r}_{\text{src}}$  に配置された点音源の音源位置における音圧である。 $k$  は波数を表し、 $h_n$  と  $P_n$  はそれぞれ次数  $n$  の球ハンケル関数とルジャンドル関数で表される。

### 2.2 アレイオブアレイズ技術を用いた音源分離手法 [8]

マイクロフォンアレイを1つだけ使った方法と2つ使った方法を図2で比較する。2つのアレイを用いることによる利点は片方のアレイが、もう一方のアレイが分離できない妨害音を取り除くことができるという点にある。このことにより音源分離の精度は向上し、正確に必要な音を取り出すことができる。

式(1)を各マイクロフォンアレイに適用して音源の音圧を得ることになるが、ここで注意しなければいけないのは、アレイを複数用いるためアレイの位置も考慮しないといけないという点である。アレイの位置を考慮すると伝達関数は以下の式で書き表すことができる。

$$H_{\text{mic}}^{\text{array}}(\vec{r}_{\text{src}}, k) = \sum_{n=0}^{\infty} (2n+1) \frac{h_n[k|\vec{r}_{\text{src}} - \vec{r}_{\text{array}}|]}{h'_n(kR_{\text{array}})} P_n \left( \frac{(\vec{r}_{\text{mic}} - \vec{r}_{\text{array}}) \cdot (\vec{r}_{\text{src}} - \vec{r}_{\text{array}})}{|\vec{r}_{\text{mic}} - \vec{r}_{\text{array}}| |\vec{r}_{\text{src}} - \vec{r}_{\text{array}}|} \right) \quad (2)$$

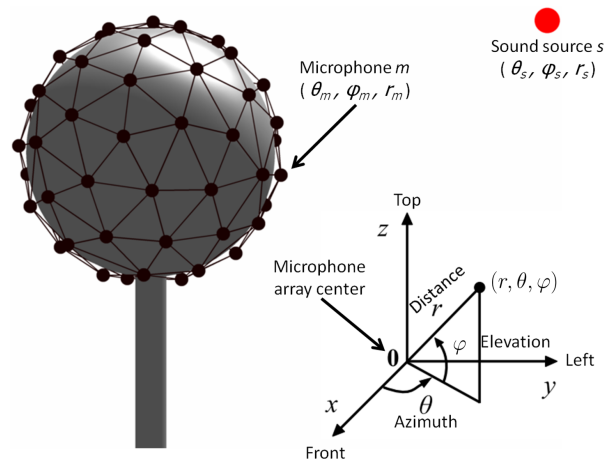


図1 アレイの中心を原点とした極座標系

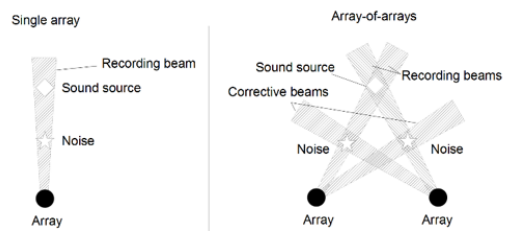


図2 1アレイによるビームフォーミングとアレイオブアレイズ技術によるビームフォーミング

$\vec{r}_{array}$  はアレイの位置を指し、 $R_{array}$  はアレイの半径を表す。  
 $\vec{r}_{src}$  と  $\vec{r}_{mic}$  は元の定義のままである。式 (2) を用いることで必要とする音源の音圧は次式で表される。

$$p_{target}(k) = \begin{bmatrix} W(\vec{r}_1) & W(\vec{r}_2) & \dots & W(\vec{r}_a) \end{bmatrix} \begin{bmatrix} H_1^1(\vec{r}_1, k) & H_1^1(\vec{r}_2, k) & \dots & H_1^1(\vec{r}_a, k) \\ H_2^1(\vec{r}_1, k) & H_2^1(\vec{r}_2, k) & \dots & H_2^1(\vec{r}_a, k) \\ \vdots & \vdots & \ddots & \vdots \\ H_b^1(\vec{r}_1, k) & H_b^1(\vec{r}_2, k) & \dots & H_b^1(\vec{r}_a, k) \\ H_1^2(\vec{r}_1, k) & H_1^2(\vec{r}_2, k) & \dots & H_1^2(\vec{r}_a, k) \\ H_2^2(\vec{r}_1, k) & H_2^2(\vec{r}_2, k) & \dots & H_2^2(\vec{r}_a, k) \\ \vdots & \vdots & \ddots & \vdots \\ H_b^N(\vec{r}_1, k) & H_b^N(\vec{r}_2, k) & \dots & H_b^N(\vec{r}_a, k) \end{bmatrix}^+ \begin{bmatrix} p_1^1(k) \\ p_2^1(k) \\ \vdots \\ p_b^1(k) \\ p_1^2(k) \\ p_2^2(k) \\ \vdots \\ p_b^N(k) \end{bmatrix} \quad (3)$$

式 (3) は  $N$  個のアレイにそれぞれ  $b$  個のマイクロフォンを取り付けた場合で考えられている。アレイ  $i$  の  $j$  番のマイクロフォンの信号は  $p_j^i(k)$  のように表される。伝達関数と録音信号の積に窓関数  $W$  をかける。  $W$  は、ターゲット信号を位置を中心として空間上に広がりを持つ関数であり、抽出すべき信号の範囲を規定するものである。妨害音が密集しているような環境では、急峻な窓関数を設定することで、妨害音を含まずにターゲット信号を抽出することが可能となる。

### 3. 評価実験

#### 3.1 実験装置・信号

音源分離実験には半径 8.5 cm、64 個のマイクロフォンが剛球表面上に規則的に設置された 2 つの球状マイクロフォンアレイを使用した。アレイの写真を図 3 で示す。音源分離性能評価実験ではターゲットとなる目的音、妨害音とも、白色雑音 (音圧レベル: 70 dB) を使用した。また、目的音を発生するために 1 台、妨害音を発生するために 2 台、計 3 台のラウドスピーカを用いた。前者は、TSP 信号 [9] とホワイトノイズが出力され、後者からはホワイトノイズが出力された。

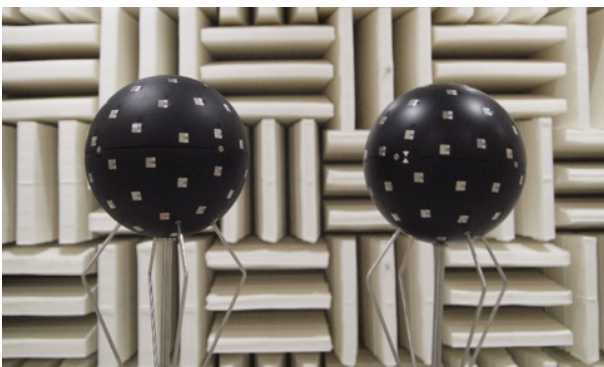


図 3 2 つの 64 チャンネル球状マイクロフォンアレイ

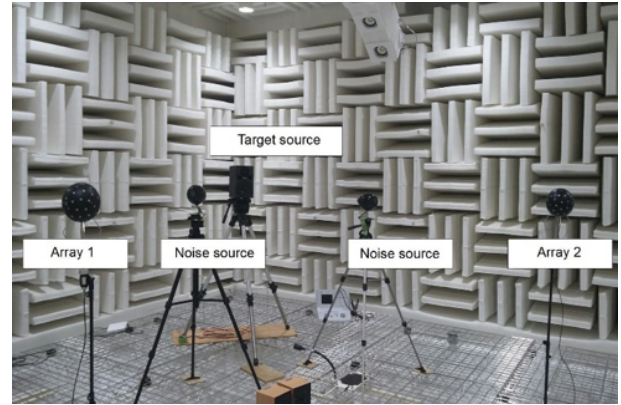


図 4 実験環境

#### 3.2 実験環境

実験は無響室で行われ、図 4 の環境で評価した。今回の実験ではどちらも目的音用スピーカ、妨害音用スピーカ、マイクロフォンアレイはすべて同じ高さに設定した。したがってこれらの相対位置は 2 次元環境で表すことができる。

#### 3.3 システム校正

実験に先立って、実験システム全体の校正を行った。まず、校正用ラウドスピーカを球状マイクロフォンアレイから 1 m 先に置き、サンプリング周波数 48 kHz、音圧レベル 70 dB、512 ポイントの TSP 信号を 1000 回提示して同期加算を行った。得られた信号に対して、逆 TSP 信号をたたみ込むことで、各マ

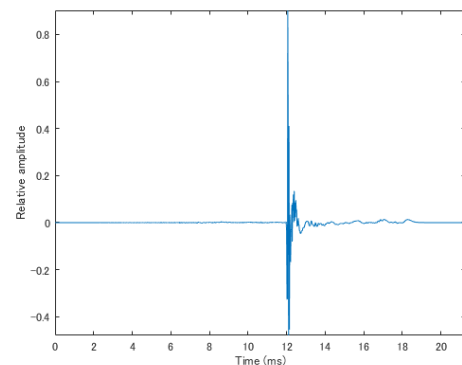


図 5 マイクロフォンで収録インパルス応答

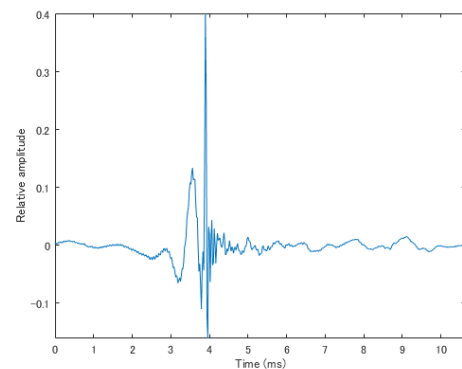


図 6 校正フィルタのインパルス応答

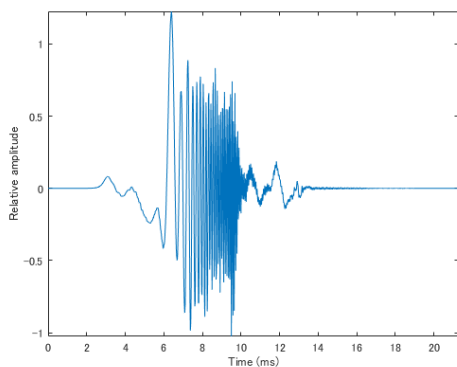


図7 校正フィルタ適用後の時間波形

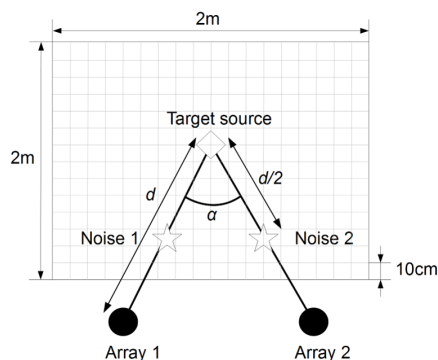


図8 アレイとラウドスピーカの配置位置

イクロフォンのインパルス応答を求めた。

このインパルス応答を、剛球上の該当する点におけるインパルス応答に校正する必要があるが、算出されたインパルス応答には、ラウドスピーカの周波数特性が含まれている。そこで、球状マイクロフォンアレイの中心に校正用マイクロフォン (Bruel & Kjaer Type 4165) を置いてラウドスピーカからの出力を録音し、最小二乗法によりラウドスピーカの特性の逆フィルタを算出した。次に、算出された逆フィルタを球状マイクロフォンアレイで収録された信号に適用した。その上で、120～15000 Hz の帯域において剛球上の該当する点におけるインパルス応答となるように、最小二乗法を用いて逆フィルタを設計した。校正用ラウドスピーカから 1 m 先の位置に設置されたマイクロフォンで収録されたインパルス応答を図 5、最終的に設計された逆フィルタのインパルス応答を図 6 に、マイクロフォンで収録された信号に算出された逆フィルタを適用した時間波形を図 7 に示す。

### 3.4 音源分離性能評価

音源分離性能を評価する実験はラウドスピーカとマイクロフォンアレイを設置して実験を行った。ラウドスピーカとアレイの位置は図 8 のように設定した。目的音源から各アレイまでの距離を  $d$ 、目的音源から妨害音源までの距離は  $d/2$ 、各アレイと目的音源との狭角は  $\alpha$  とした。今回の実験では  $d=2$  m と設定し、 $\alpha=30^\circ, 60^\circ$  に変化させた。

目的音源のみ、2つの妨害音源のみ、および、すべての音源

から白色雑音を出力で測定を行い、それぞれの場合で音源分離性能を評価した。範囲は  $2 \times 2$  m とし、グリッドのサイズは 10 cm とし、2つのアレイの中心位置を原点として座標を設定した。その上で、グリッドの中心位置が  $(0, 1.75)$  となるようにグリッドの位置を合わせた。したがって、 $\alpha$  が  $60^\circ$  の時には、目的音は  $(0, 1.73)$ 、妨害音は  $(\pm 0.5, 0.87)$  に位置し、 $30^\circ$  の時は、目的音は  $(0, 1.9)$ 、妨害音は  $(\pm 0.27, 0.95)$  に位置することになる。各アレイで前述のような座標を設定したうえで、目的音源から放射される音のエネルギー分布を求めた。各点のエネルギー  $E$  は次の式で表される。

$$E = H^+P \quad (4)$$

$H$  は伝達関数、 $P$  は録音信号の音圧を表す。このエネルギーに Tukey フィルタをかけた。Tukey フィルタをかけたエネルギーは次式で表される。

$$E_{\text{filter}} = WE \quad (5)$$

$W$  は窓関数を指している。

### 3.5 結果及び考察

まず、目的音源のみから白色雑音を提示した時について説明する。図 9 は各アレイと目的音源との角度が  $60^\circ$  のとき、図 10 は各アレイと目的音源との角度が  $30^\circ$  のときの Tukey フィルタをかけた音圧分布である。音圧分布図を比較すると、アレイ間の角度が  $60^\circ$  の時の方が、 $30^\circ$  の時に比べ、目的音の位置におけるエネルギーが高くなっているように見て取れる。

次に、2つの妨害音源のみから白色雑音を提示した時について説明する。図 11 は各アレイと目的音源との角度が  $60^\circ$  のとき、図 12 は各アレイと目的音源との角度が  $30^\circ$  のときの Tukey フィルタをかけた音圧分布である。妨害音源間の距離は  $30^\circ$  のときは  $60^\circ$  のときより短い、音圧分布図を見てみると妨害音源間での音圧の下がり方は  $60^\circ$  のときより  $30^\circ$  のときの方が急であることが分かる。

以上の結果から、アレイ間隔が  $60^\circ$  の方が、目的音のエネルギーが顕著に観測され、その一方、妨害音のエネルギーは  $30^\circ$  の時の方が顕著に見られていることから、音源分離性能はアレイ間の角度が  $60^\circ$  の時の方が高いことが想定される。

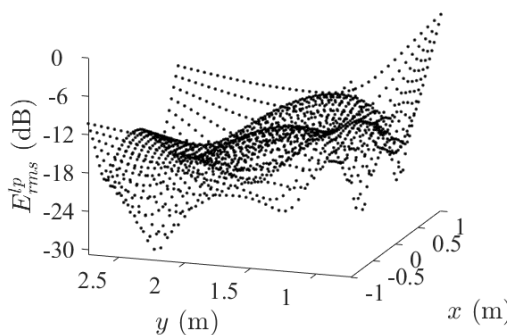


図9 目的音源の音圧分布 ( $60^\circ$ )

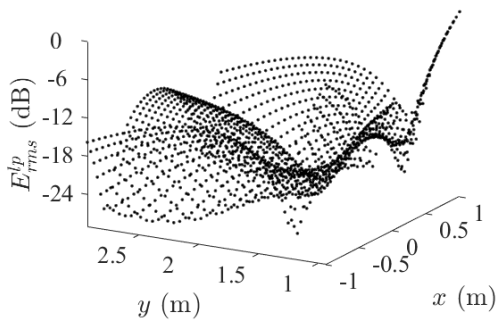


図 10 目的音源の音圧分布 (30°)

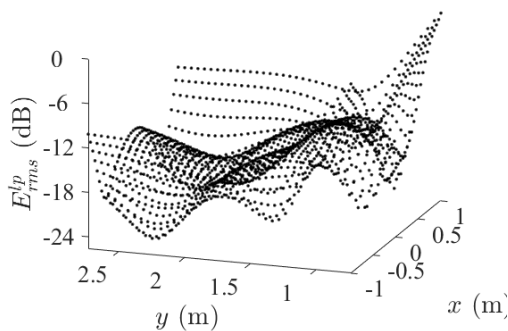


図 11 妨害音源の音圧分布 (60°)

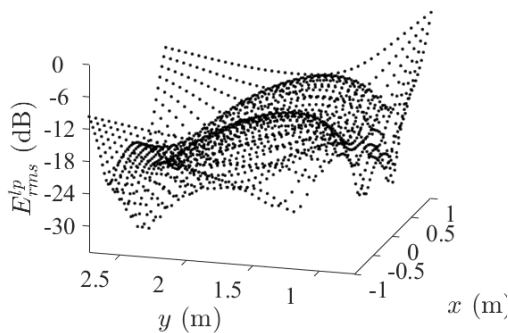


図 12 妨害音源の音圧分布 (30°)

#### 4. ま と め

本報告では、各アレイと目標音源の狭角を変化させたときの音源分離性能の測定、比較を行った。狭角が30°と60°の場合の音圧分布を比較してみたところ、目的音のエネルギーは60°の方が、妨害音のエネルギーは30°の方が顕著であることが観測されたため、60°の方が30°よりも音源分離性能が高いと想定される。

今後は、フィルタとなる窓関数を変化させた場合やアレイ上のマイクロフォンの数を増やした場合の音源分離性能評価を行っていく予定である。

謝辞 本研究の一部は、文部科学省研究費基盤研究

(A)(No.16H01736)の補助による。

#### 文 献

- [1] D. N. Zotkin, R. Duraiswami. and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays", IEEE Trans. On Audio, Speech and Lang. Proc., vol.18, no.1, pp.2-16, 1964.
- [2] S. Holmes, "Circular harmonics beamforming with spheroidal baffles", Proc. Int. Congr. On Acoust. Soc. Am., POMA 19, no.055077, pp.1-9, 2013.
- [3] R. O. Duda, and W. L. Martens, "Range dependence of the response of a spherical head model", J. Acoust. Soc. Am., vol.104, no.5, pp.3048-3058, 1998.
- [4] S.Sakamoto, S. Hongo, T.Okamoto, Y.Iwaya and Y.Suzuki, "Sound-space recording and binaural presentation system based on a 252 ch microphone array", Acoust. Sci. and Tech., vol.36, no.6, pp.516-526, 2015.
- [5] S.Sakamoto, S. Hongo, R.Kadoi and Y.Suzuki, "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones", Proc. 2nd Int. Symp. On Univ. Comm., pp.429-434, 2008.
- [6] B.Rafaely, "Analysis and design of spherical microphone arrays", IEEE Trans. on Speech and Audio Proc., vol.13, no.1, pp.135-143, 2005.
- [7] A. M. Torres, M. Cobos. B.Pueo and J. Lopez, "Robust acoustic source localization based on modal beamforming and time-frequency processing using circular microphone arrays", J. Acoust. Soc. Am., vol.132, no.3, pp.1511-1520, 2012.
- [8] T. Jorge, S. Sakamoto, and Y. Suzuki, "Separation of spatially-segregated sound sources using multiple spherical microphone arrays", 日本音響学会秋季研究発表会講演論文集, no.1-Q-44, pp787-790, 2015.
- [9] N. Aoshima, "Computer-generated pulse signal applied for sound measurement", J. Acoust. Soc. Am. Vol.69, No.5, pp.1484-1488,1981.