

両耳間レベル差を用いた音源定位における耳孔位置の最適化

Optimization of the ear canal position for sound localization using interaural level difference

木元大輔, 公文誠

Daisuke KIMOTO and Makoto KUMON

熊本大学

Kumamoto University

d.kimoto@ick.mech.kumamoto-u.ac.jp

Abstract

耳介のあるバイノーラル聴覚システムでは、両耳間レベル差を用いた音源定位によって正中面内の音源の上下方向の定位を行うことが可能とされている。この場合、限られたマイクロホン数の制限と、マイクロホンの配置の影響で、両耳間レベル差によっては定位の困難な、または定位精度が著しく低下してしまう方向が生じる問題がある。このような方向はマイクロホン位置や頭部形状などの影響を受けるため、マイクロホンの配置を適切に選ぶことで、定位性能を改善できる可能性がある。

本研究では、マイクロホン周辺に取り付けた耳介の音響特性を考慮して、耳孔位置を最適化することで両耳間レベル差を用いた音源定位性能の向上を図る。実験を通して特に最適化しない耳介を用いた場合と比較し、最適化した耳介を用いた場合、音源定位性能が向上することが確かめられた。

1 はじめに

ロボットが自動で柔軟に適切な対応を行うためには、周辺環境の認識がとても重要である。この周辺環境の認識を行う方法の一つとして、音信号を利用する方法がある。これは、ロボット聴覚として盛んに研究されている[奥乃, 2010]。その中でも、音の到来方向や音源方向を正確に推定することは、将来、ロボットが人と共生する上で必要な能力であり、重要な基礎機能と言える。

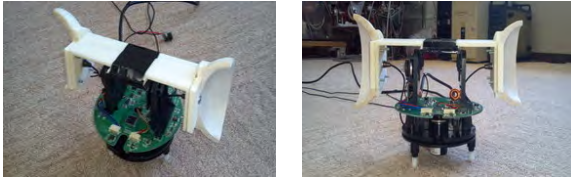
人間や動物は2つの耳のみで現実的な音源定位を実現しており、音源定位のための必要最小限の構成だと考えられる。ロボットでも2つのマイクロホンだけで音源定位を実現することは、聴覚システムの簡素化や音源定位の原理の解明など、興味深い課題を含んでいると言える。

また、人間や動物では、耳に耳介と呼ばれる音の反射・集音を果たす器官が存在している。耳介形状は一般的には複雑なため、耳介の音響特性が音の到来方向に応じて異なることが知られている[E.A.Lopez, 1976]。特に、耳介のゲイン特性が顕著に抑制されている帯域を耳介ノッチ[E.A.G.Shaw, 1968]と呼ぶが、耳介ノッチの周波数が音源方向の関数になっているので、耳介ノッチの検出を行えば音源方向を推定することが可能である。実際に、章ら[章, 2008]は、音源方向の推定のための特徴量として両耳間レベル差 (Interaural Level Difference, ILD) を用いて、事前に学習した特徴量との相関を求めることで仰伏角および方位角を推定している。

本研究では、耳介に埋蔵するマイクロホン位置を適切に選ぶことでこの方法の改善を試みる。このため、音源から直接マイクロホンに収録される直接波、および耳介上で反射し収録される信号(2回反射まで)を考慮した耳介の周波数応答モデルを考え、数値的に最適な耳介を設計することとした。最適化には特にILDが音源方向毎に十分異なるような指標を考え、マイクロホン位置を最適化パラメータとした。なお、本研究では定位においてILDが重要とされる仰伏角方向に着目するため、モデルでの音源は正中面内に位置するものを考える。

2 バイノーラル聴覚ロボット

本研究ではFigure1に示すバイノーラル聴覚ロボットを用いる。2つのマイクロホンには同じ形状の耳介に取り付けられているが、その取り付け位置は後述するように異なることがある。本研究では、左右のマイクロホンで受聴される信号が音源方向やマイクロホン位置によって異なる周波数伝達特性の影響を受けると考え、それを利用することを想定している。



(a) ロボット頭部 (b) ロボット正面

Figure 1: バイノーラル聴覚ロボット

ロボットは Figure2 に示すような方位、仰伏角の 2 方向に動作可能である。使用しているモータが超音波モータのため、ロボットが動作中であっても可聴域でのエゴノイズはほとんど生じない。ロボットを制御する計算機には PC を使用する。PC は姿勢角および角速度の規範値を指定し、シリアル通信によりモータを制御する。また、PC のマイク端子よりマイクロホンが受信した音信号を 44.1[kHz] で左右の 2ch とともにサンプリングする。

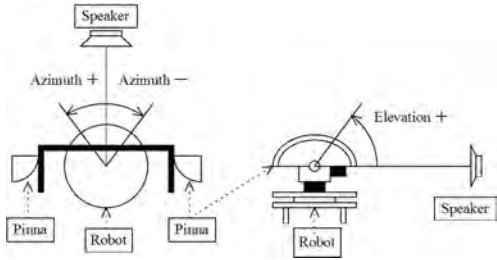


Figure 2: ロボット回転方向

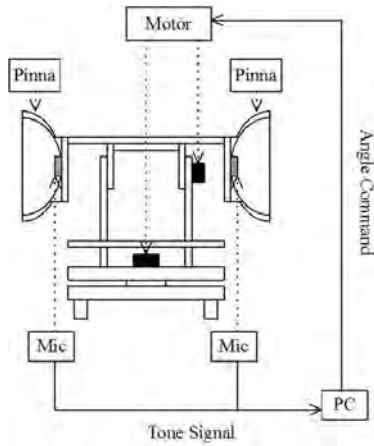


Figure 3: システム

3 両耳間レベル差

環境やロボット身体の影響によりロボットの受聴する音信号は原信号とは異なったものとなる。今、ロボットを基準とした音源までの距離、方位角、仰伏角をそれぞれ r, θ, ϕ とすると左右のマイクロホンへの伝達関数のうちロボッ

ト身体によるものは $H_l(\theta, \phi; \omega), H_r(\theta, \phi; \omega)$ 、環境によるものは $H_{le}(r, \theta, \phi; \omega), H_{re}(r, \theta, \phi; \omega)$ と表すことができる。また、原信号 $s_O(\omega)$ に対してロボットの左右のマイクロホンで受聴する音信号をそれぞれ $s_l(r, \theta, \phi; \omega), s_r(r, \theta, \phi; \omega)$ とすれば Figure4 に示すように

$$\begin{aligned} s_l(r, \theta, \phi; \omega) &= H_l(\theta, \phi; \omega)H_{le}(r, \theta, \phi; \omega)s_O(\omega) \\ s_r(r, \theta, \phi; \omega) &= H_r(\theta, \phi; \omega)H_{re}(r, \theta, \phi; \omega)s_O(\omega) \end{aligned}$$

の関係がある。

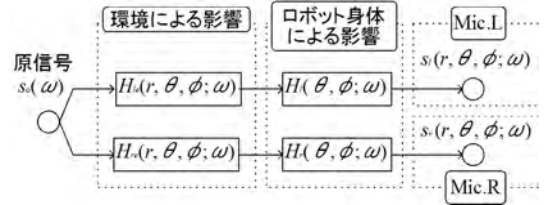


Figure 4: 受聴される音信号

もし環境の影響が $H_{le} \equiv H_{re}$ であれば、両耳間レベル差 (Internal Level Difference, ILD) S は

$$\begin{aligned} S &\equiv 20 \log |s_l(r, \theta, \phi; \omega)| - 20 \log |s_r(r, \theta, \phi; \omega)| \\ &= 20 \log |H_l(\theta, \phi; \omega)H_{le}(r, \theta, \phi; \omega)s_O(\omega)| \\ &\quad - 20 \log |H_r(\theta, \phi; \omega)H_{re}(r, \theta, \phi; \omega)s_O(\omega)| \\ &\equiv 20 \log |H_l(\theta, \phi; \omega)| - 20 \log |H_r(\theta, \phi; \omega)| \end{aligned}$$

と近似でき、ILDS が音源方向 θ, ϕ の関数として $S(\theta, \phi; \omega)$ となり、ロボット身体の影響のみで特徴づけられることになる。特に耳介はマイクロホン近傍にあるため、伝達特性に大きく影響していると考えられ、音源定位の情報を与えると期待される。

4 両耳間レベル差を用いた音源定位法

$ILDS(\theta, \phi; \omega)$ が音源方向によって特徴づけられたものであるので、本研究では対象とする周波数帯域の ILD を特徴量ベクトルの 1 つとする。なお、原信号 s_O に含まれていない (あるいは非常に小さい) 周波数成分については S が正しく求まらない可能性があるため、除外して考える必要がある。このため、適当な正定数 ε に対して

$$f(x, a, b) = \begin{cases} 0 & (\text{if } a < \varepsilon \text{ or } b < \varepsilon) \\ x & (\text{otherwise}) \end{cases}$$

となる関数 f を用いて、特徴量ベクトル S_{ILD} を

$$\begin{aligned} S_{ILD} &= [f(\Delta S(\omega_1), |s_l(\omega_1)|, |s_r(\omega_1)|), \\ &\quad \dots, f(\Delta S(\omega_N), |s_l(\omega_N)|, |s_r(\omega_N)|)]^T \end{aligned}$$

と定める。ここで $\omega_1, \dots, \omega_N$ は対象とする周波数成分である。

本手法では音源方向 θ, ϕ から周波数成分を十分に含んでいる音信号を与え、ILD の特徴ベクトルを事前に計測し、学習ベクトル $S_d(\theta, \phi)$ として保存する。このような推定対象となる方向を $\theta_1, \dots, \theta_{N_\theta}, \phi_1, \dots, \phi_{N_\phi}$ とすると、 $N_\theta \times N_\phi$ 点の特徴ベクトルを学習ベクトルとしてデータベースに保存することになる。

次に方向を推定したい音信号が与えられ、この特徴ベクトル S が得られたとする。この時 S と同じ方向から得られた学習ベクトルは、 S との間で高い相関を示すと考えられる。相関は以下のように内積として表わされる。

$$X_{ILD}(S, S_d) = \frac{\langle S, S_d \rangle}{|S||S_d|} \quad (1)$$

ここで、 X_{ILD} は特徴ベクトル S が与えられたとき θ, ϕ の関数として求まるので $X_{ILD}(S, S_d) = X_{ILD}(S, \theta, \phi)$ となり、音源方向の推定値として X_{ILD} を最大とするものを用いる。

本研究では、そのまま学習ベクトルを使用するのではなく、学習ベクトルデータベースの擬似逆行列を使用する。これにより、単純に学習ベクトルを使用するのに比べ方向推定性能が改善されると期待できる[公文, 2011]。

5 耳介の周波数応答モデル

ここでは、マイクロホン取り付け位置がILDに与える影響を調べるため、耳介の音響特性を与える周波数応答モデルを考える。

5.1 耳介の周波数応答モデル

$H(\mathbf{P}_s, \omega)$ を \mathbf{P}_s に位置する音源からの角周波数 ω の伝達特性(ゲイン)とし、正弦波でのモデル(音源が発している音信号を $\sin \omega t$ と仮定)を考える。

音源から直接マイクロホンに収録される信号(直接波)は、音信号の振幅が距離により減衰するので、マイクロホン位置では以下のように表すことができる。

$$r(\mathbf{P}_s, \omega, t) = \frac{1}{l} \sin(\omega(t - \frac{l}{V_s})) \quad (2)$$

ここで、 l は音源からマイクロホンまでの距離、 V_s は音速を表している。

次に、耳介上で反射し、その後マイクロホンに収録される音信号を考える。反射点までの音信号は直接波と同様に考えると、耳介の反射点での信号は

$$\frac{1}{l_{s1}} \sin(\omega(t - \frac{l_s}{V_s})) \quad (3)$$

となる。ここで l_{s1} は音源位置から反射点までの距離を表す。反射後の信号は反射点を音源位置とする(3)式の音信号と考えられ、すると、反射した音信号はマイクロホン位置において

$$r_{r1}(\mathbf{P}_s, \omega, t) = \frac{\lambda(\omega)}{l_{s1}l_{m1}} \sin(\omega(t - \frac{l_s + l_{m1}}{V_s})) \quad (4)$$

と考えられる。ここで $\lambda(\omega)$ は耳介での反射率、 l_{m1} は反射点からマイクロホンまでの距離を表す。

同様に、耳介や耳介の周りで反射したのち、再び耳介で反射しその後マイクロホンに収録される2回反射を考える。これは1回反射の時と同様に考え

$$r_{r2}(\mathbf{P}_s, \omega, t) = \frac{\lambda^2(\omega)}{l_{s2}l_{e2}l_{m2}} \sin(\omega(t - \frac{l_{s2} + l_{e2} + l_{m2}}{V_s})) \quad (5)$$

のように表せる。ここで l_{s2} は音源位置から反射点1までの距離、 l_{e2} は反射点1から反射点2までの距離、 l_{m2} は反射点2からマイクロホンまでの距離を表す。

これらより、正弦波での応答モデルは以下のように示すことができる。

$$h(\mathbf{P}_s, \omega, t) = r(\mathbf{P}_s, \omega, t) + \iint_S r_{r1}(\mathbf{P}_s, \omega, t) dS + \iint_S r_{r2}(\mathbf{P}_s, \omega, t) dS \quad (6)$$

本研究ではILDに着目しているため、(6)式のゲイン特性のみを考慮すればよい。したがって、(6)式のゲイン特性を $H(\mathbf{P}_s, \omega)$ として用いる。

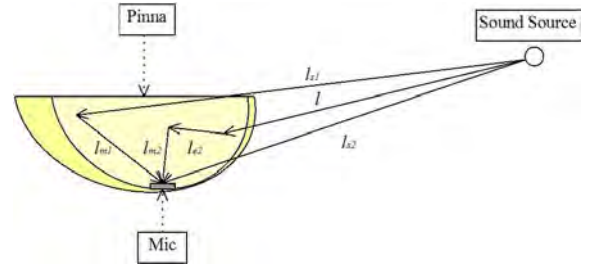


Figure 5: 反射モデル

5.2 モデルの検証

作成した耳介応答モデルの検証を行う。検証に用いた耳介形状は $x = 40[\text{mm}]$, $y = 30[\text{mm}]$, $z = 20[\text{mm}]$ であり、マイクロホン位置は原点の位置とした。

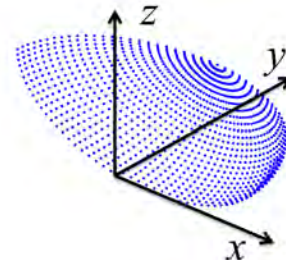


Figure 6: 耳介形状

また、音の収録は $9.4[\text{m}] \times 3.45[\text{m}] \times 2.8[\text{m}]$ の Figure7 に示すような部屋で行った。

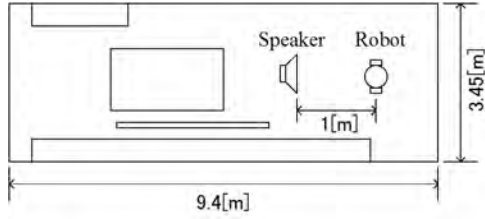


Figure 7: 収録環境

まず、収録を行い計測により求めた ILD 結果を Figure8 に示す。また、ILD をモデルより推定した結果を Figure9 に示す。なお、反射率 $\lambda(\omega)$ は実測した周波数応答に良く合うよう適宜調整を行った。

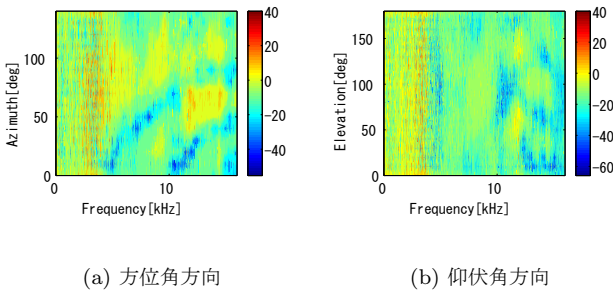


Figure 8: 収録結果より求めた ILD

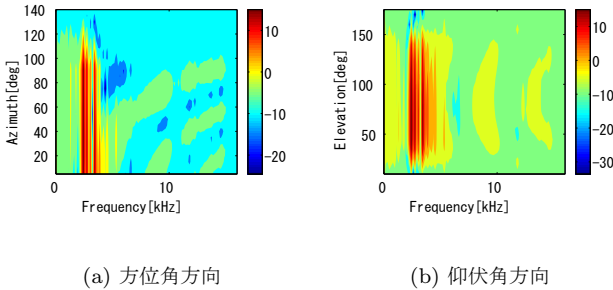


Figure 9: 応答モデルより求めた ILD

計測により求めた ILD とモデルより推定した ILD を比較すると、ノッチの応答に酷似した傾向があることがわかる。

6 耳孔位置の最適化

6.1 耳孔位置の最適化法

左側の耳介の応答を H_l 、右側の耳介の応答を H_r とし、5 節で導入したモデルを用いて ILD を求めることができる。

今、ILD を用いた音源定位では音源方向毎に S が大きく異なっていることが望ましい。このことから、音源 s_i 、 s_j の音源位置を \mathbf{P}_{s_i} 、 \mathbf{P}_{s_j} とすると、それらの音源方向

の ILD 間の内積

$$X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j}) = \frac{\langle S(\mathbf{P}_{s_i}, \omega), S(\mathbf{P}_{s_j}, \omega) \rangle}{|S(\mathbf{P}_{s_i}, \omega)| |S(\mathbf{P}_{s_j}, \omega)|} \quad (7)$$

について、すべての音源位置 \mathbf{P}_{s_i} 、 \mathbf{P}_{s_j} の組み合わせについて考える。ここで、ある値 $\epsilon (< 1)$ よりも小さい $X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j})$ が多い場合に、音源方向によって特性により違いがあると言えるので、本研究では $X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j}) < \epsilon$ となる $X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j})$ の集合を M_ϵ 、また、 M_ϵ の要素数を m_ϵ とし m_ϵ の多いものを適当なマイクロホン位置と考える。

$$M_\epsilon = \{X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j}) | X(\mathbf{P}_{s_i}, \mathbf{P}_{s_j}) < \epsilon\} \quad (8)$$

$$m_\epsilon = |M_\epsilon| \quad (9)$$

なお、本研究では ILD が重要となる仰伏角方向に着目し、モデルでの音源位置を正中面内に限定し、音源の上下方向の変化に対応した ILD を対象に最適化を行った。

6.2 耳孔位置の最適化

応答モデルを用いて耳孔（マイクロホン）位置の最適化を行う。今回は、計算量の観点から最適化を行うマイクロホン位置を Figure11 に示すように 5[mm] 四方のメッシュ上に限定した。また、マイクロホンは耳介表面上にあるとして最適化を行った。

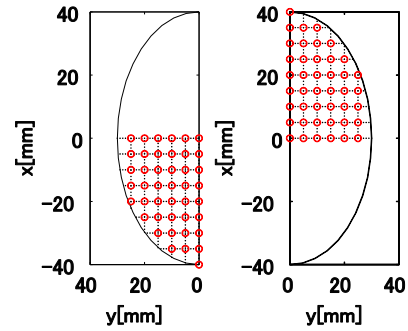


Figure 10: マイクロホン位置

$\epsilon = 0.985$ とし最適化計算を行った結果を Figure11 に示す。Figure11 は m_ϵ の値が大きいものの上位 6 つを示しており、色毎に左右のマイクロホン位置が対応している。

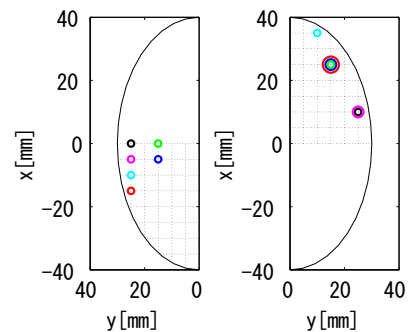


Figure 11: マイクロホン位置推定結果

Figure11 に示す結果から、マイクロホン位置が $x_l = 25[\text{mm}]$, $y_l = 15[\text{mm}]$, $x_r = -15[\text{mm}]$, $y_r = 20[\text{mm}]$ 近傍にあることが望ましいということがわかる。

6.3 最適化結果の検証

6.2節で求めた最適化結果の検証を行う。比較のために、

- Symmetric pinnae - マイクロホン位置が左右とも $x = 0[\text{mm}]$, $y = 0[\text{mm}]$ のもの
- Non-optimal pinnae - マイクロホン位置が $x_l = 15[\text{mm}]$, $y_l = 10[\text{mm}]$, $x_r = -15[\text{mm}]$, $y_l = 10[\text{mm}]$ のもの
- Optimal pinnae - 最適化結果 $x_l = 25[\text{mm}]$, $y_l = 15[\text{mm}]$, $x_r = -15[\text{mm}]$, $y_r = 20[\text{mm}]$ のもの

とする。

実際に、以上の位置にマイクロホンを有する耳介を作成し、音源定位を行った結果を Figure12 に示す。さらに、音源定位性能を比較した結果を Figure13 に示す。図中の縦軸は定位誤差の RMS を表す。

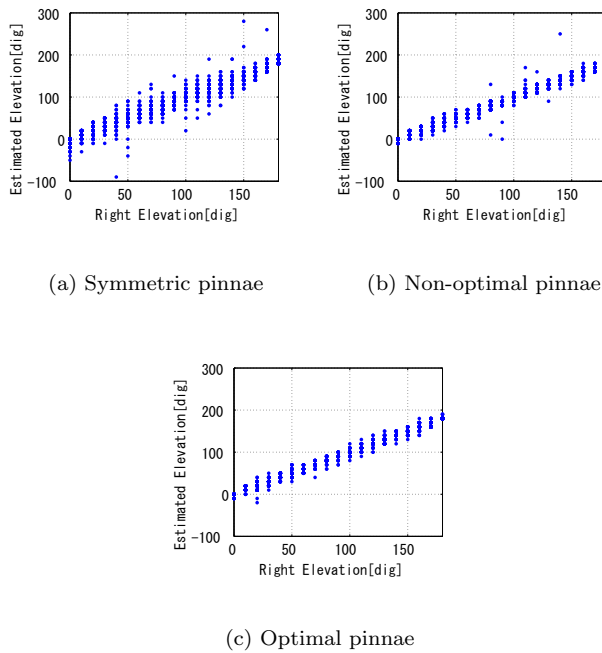


Figure 12: 音源定位結果

Figure13 より、まずマイクロホン位置を単純に変化させた場合でも音源定位の性能に大きな影響があることがわかる。さらに、単純にマイクロホン位置を変化させた Non-optimal pinnae に比べ、最適化結果を用いた Optimal pinnae の方が誤差の値が小さいことから、最適化結果が比較したマイクロホン位置の中でもっとも音源定位性能が良く、Symmetric pinnae と比較すると 55.9%、Non-optimal pinnae と比較すると 21.5%の誤差を削減することができた。このことから、今回導入した最適化法の有効性が言える。

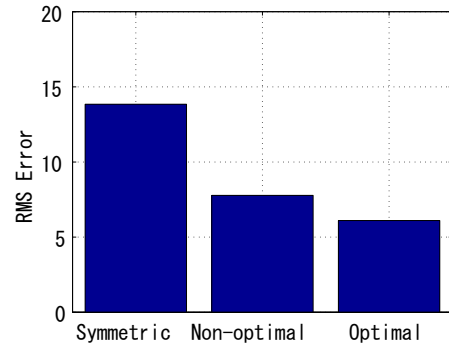


Figure 13: 最適化結果の評価 (仰伏角方向)

7 結言

本研究では、ILD を用いた音源定位において、耳介形状、特に耳孔位置の変更により定位性能の向上を試みた。その結果、耳孔位置の最適化を行うことにより、音源定位性能の向上を確認することができた。

今後は、今回導入した最適化法を耳介形状に適用させ、耳介形状の最適化を行いたい。

参考文献

- [奥乃, 2010] 奥乃 博: ロボット聴覚の現状と展望, 日本ロボット学会誌 Vol.28 No.1, pp.2-5, 2010.
- [章, 2008] 章 忠, 井和章, 三宅 哲夫, 今村 孝, 堀畑 聡: バイノーラルモデルを用いた音源方向定位, 日本機械学会論文集 C 編, Vol.74-739, pp.642-649, 2008.
- [C.Knapp, 1976] C.Knapp, G.Carter: The generalized Correlation Method for Estimation of Time Delay, IEEE Trans.on Acoustic Speech and Signal Process., vol.4, pp.320-327, 1976.
- [E.A.Lopez, 1976] E.A.Lopez, Ray Meddis: A physical model of sound diffraction and reflections in the human concha, J.Acoust. oc.Am.Volume100, Issue 5, pp.3248-3259, 1996.
- [公文, 2011] 公文 誠, 木元 大輔: 耳介を持つバイノーラル聴覚ロボットの音源方向推定の検討, 人工知能学会研究資料 SIG-Challenge-B102-11, 2011.
- [森下, 2009] 森下 巖, 小畑 秀文: 信号処理, コロナ社, 2009.
- [E.A.G.Shaw, 1968] E.A.G.Shaw, R.Teranishi: Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source, J.of Acoust.Soc.Am.44 1), pp.240-249, 1968.