

音源定位における能動耳介での動作の影響について

On effect of active pinnae motion to sound source localization

尾堂航, 公文誠

Wataru ODO, Makoto KUMON

熊本大学大学院自然科学研究科

Graduate School of Science and Technology, Kumamoto University

wataru@as.mech.kumamoto-u.ac.jp

Abstract

本研究では能動的に動作可能な反射板 (能動耳介) を備えた二つのマイクロホンを用いたロボットシステムにおいて, 能動耳介の動作が音源定位に与える影響を考える. 能動耳介は反射板の向きを変えて伝達特性を変化させ, 音源定位性能の向上を目指すシステムであるが, 能動耳介の動作時にエゴノイズが生じる. これまでこのエゴノイズが音源定位に与える影響は知られていなかった. 本研究では能動耳介の動作時におけるエゴノイズが音源定位に与える影響を明らかにする.

1 はじめに

私たち人間は, 外部の情報を得るために五感と呼ばれる感覚を用いて様々な情報を得ている. 特に聴覚においては, 左右の耳で聞こえた音の遅れや音の大きさの違いなどを脳で判断し, それを音源の方向として認識した結果, 方位角, 仰伏角, 距離などの情報を得ることができる. ロボットが私たちの身近で活動するためには, 人間の普段の生活環境を認識する必要がある. 従って人間のように音の情報によって周囲の環境を認識するシステムがロボットにとって求められる. ロボットにも音情報から音源の位置・方向を認識する音源定位と呼ばれる聴覚機能と, これを用いた環境を認識する能力が重要となる.

ロボットにおける音源定位については3つ以上のマイクロホンを用いるマイクロホンアレイによる研究が行われている. 例えば MUSIC 法やビームフォーマなどが良く知られた手法である [大賀 07]. マイクロホンの数が増えると音源定位が容易に行えるようになるが装置の大型化やマイクロホンを配置する空間の確保, 計算量の増加などシステムが複雑になるという問題点が挙げられる.

一方, 人間や動物は二つの耳で音の到来方向を認識することができ, 両耳間時間差 (Interaural Time Dif-

ference, ITD), 両耳間位相差 (Interaural Phase Difference, IPD), 両耳間レベル差 (Intersural Level Difference, ILD) などの音響特徴量を用いて音源定位を行うとされている [Garas 00]. ここから, ロボットも人間や動物と同様に音源定位に二つの耳を用いることが考えられる. このような二つのマイクロホンを用いたロボットの研究の一つに章ら [章 08] が提示する音源推定のための特徴量として ILD を用いた手法がある. これは事前に学習した特徴量の分布との相関を求めることで仰伏角や方位角を推定するものである. また西野ら [西野 07] はバイノーラル聴覚信号を用いて ILD を特徴量として単一ガウス分布で近似した音源方向モデルを構築することで仰伏角や方位角の推定を行っている. ところで人間や猫は頭部を動かすことによって音源定位能が向上することが知られている. 例えば, 猫頭部を固定すると水平方向の音源定位における分解能が劣化することが報告されており [Populin 98] 身体動作を伴う能動的な音源定位が重要であることが示唆されている. ロボットによる音源定位についても同様に能動的な作用を考えることは有用と考えられ, 金らは人工可動耳介を用いて追従動作を行うことで音源定位能力を向上させている. [金 12] また, 野田は確率的パラメータ推定を行い耳介の動作の下で正中面の音源定位を行っている. [野田 12] しかし, このような耳介の動作を行うには, モータ等で耳介を駆動することになりマイク近くで駆動騒音 (以下エゴノイズと呼ぶ) が生じることになり, 耳介動作そのものが音を用いた環境認識での妨げになる可能性がある. しかし, これまでの研究で能動耳介のエゴノイズが音源定位に与える影響について明らかになっていないため, 本研究では能動耳介の動作時におけるエゴノイズが音源定位に与える影響について調べることとした. 具体的には音声信号, 動作パターンの違いなどを考慮して数パターンの音収録を行い音源の方向を推定することで調査を行った.

2 能動耳介

本研究で用いる耳介を Fig.1, Fig.2 に示す。耳介は幅 70mm, 奥行き 40mm, 高さ 80mm となっており, マイクは耳介中央の奥行き 10mm の部分に配置してある。耳介を可動させるためのモータを含めると奥行きが 85mm となる。骨材となる部分はアクリル棒を用い, 反射板となる部分は厚さ 1mm のアルミニウム板である。また, マイク正面に幅 30mm, 厚さ 4mm, 高さ 30mm の耳珠に相当する板を取り付けている。これは耳介による収録音への影響を際立たせるためである [本田 85]。耳介を円弧状の基台に取り付け, 頭部にのせたものを Fig.3, Fig.4 に示す。



Figure 1: 正面図



Figure 2: 上面図



Figure 3: 正面図



Figure 4: 側面図

耳介には RC サーボ 4ch が搭載されており, 耳介一つにつき 2ch を用いる。アクリルによる骨材をモータで押し引きすることで上下方向の動きを可能にし, 左右の回転に関しては, 土台自体をモータにより回転させる (Fig.5)。それぞれのサーボはマイコンによって制御され, マイコンへの指令値は Bluetooth 接続で外部の PC から行う構成になっている。Fig.5 から分かるようにマイク近くにモータが取り付けられており, 動作時には騒音を生じる。

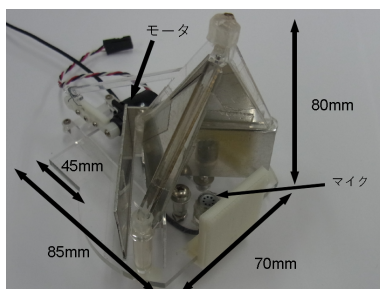


Figure 5: システム図

3 能動耳介動作時のエゴノイズ

3.1 エゴノイズの特徴

能動耳介の動作時に発生するエゴノイズを計測するため耳介を起こした姿勢と伏せた姿勢を繰り返す, 次の2つを動作パターンとして設定した。

1. 耳介の一方のみを動作させる移動量が少ない動作パターン
2. 両耳介の可動範囲全体を移動する動作パターン

また, この動作を行いながら白色雑音または音楽をマイクにて収録した。受聴した音信号は, サンプリング周波数 44.1kHz で録音し, 以下では FFT 長 1024 点を 1 フレームとして処理をし, 周波数領域での信号を考える。

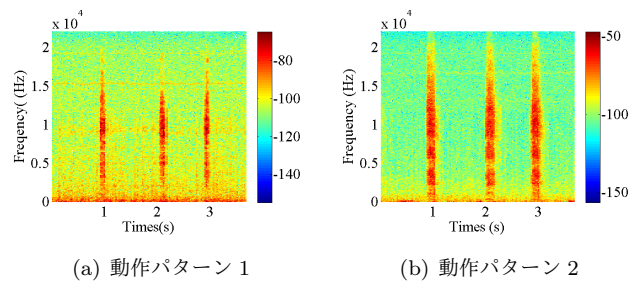


Figure 6: エゴノイズのみのスペクトログラム

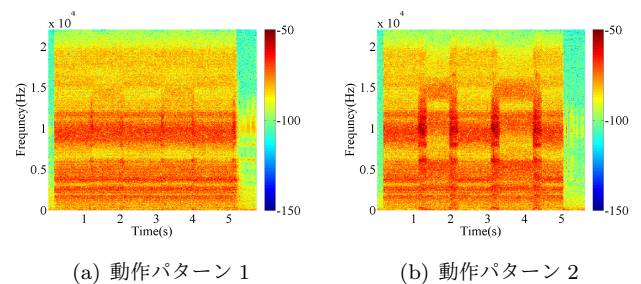


Figure 7: 白色雑音を対象音とした時のスペクトログラム

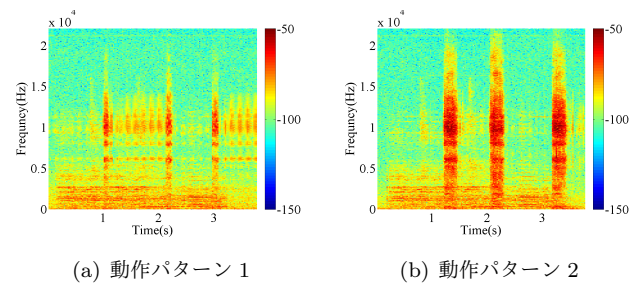


Figure 8: 音楽を対象音とした時のスペクトログラム

Fig. 6 はエゴノイズのみを収録したスペクトログラムであり, 能動耳介から発生するエゴノイズは広い周波数帯域で瞬間的にパワーをもつ信号であることがわかる。さらに耳介の移動量によってノイズのパワーは異なるが, 帯域

や時間区間はほぼ共通しているため能動耳介のエゴノイズは耳介の移動量に依存しないと言える。Fig. 8 と Fig. 7 からエゴノイズのパワーは提示した対象音信号に比較して大きいことがわかる。これはエゴノイズの発生源である能動耳介のサーボモータが耳介の後方に設置してあり能動耳介の集音部に近いたためであると考えられる。

3.2 エゴノイズ区間の検出

能動耳介のエゴノイズが短い時間に大きなパワーをもつ信号であることが分かったので、雑音に影響される時間区間の検出は収録した音信号のパワーに注目すれば可能であると考えた。フレーム k において周波数 f の信号 $s_k(f)$ のパワーを $P_k(f)$ と書けば

$$P_k(f) = 10 \log_{10} s_k^*(f) s_k(f) \quad (1)$$

と表せる。エゴノイズが全ての周波数帯域に現れることから $P_k(f)$ の合計を S_k とし

$$S_k = \sum_{f \in F} P_k(f) \quad (2)$$

を考える。ここで F は周波数点の全体集合を示す。適当な閾値 α の下で

$$S_k > \alpha \quad (3)$$

となるフレーム k はエゴノイズが生じていると判断すると思われる。ノイズ源であるモータがマイクに非常に近く、対象とする音信号に比べ、ノイズのパワーは顕著に大きいため容易に α を決定できる点は強調したい。

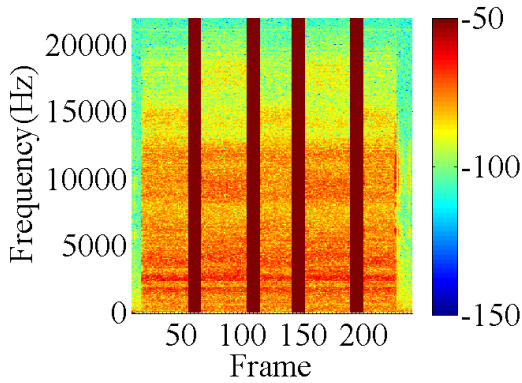


Figure 9: エゴノイズ除去後

実際に Fig. 8(b) に上記の検出法を適用し、ノイズ区間を除去した結果を Fig. 9 に示す。適切に機能していることが分かる。

4 両耳間レベル差と両耳間位相差を用いた音源定位

エゴノイズの影響を音源定位の観点から調べるため、まず定位法について説明する。

4.1 音源定位に用いる音響特徴量

周囲の環境による暗騒音や残響、ロボットの身体による反射や回折によりロボットが受聴する信号は原信号とは異なるものとなる。ある位置 \mathbf{X} より得られる周波数を ω とし、耳介の姿勢を u とする。左右のマイクロホンの伝達関数のうちロボットの身体によるものは $H_l(\mathbf{X}, u; \omega), H_r(\mathbf{X}, u; \omega)$ 、環境によるものは $H_{le}(\mathbf{X}; \omega), H_{re}(\mathbf{X}; \omega)$ と表すことができる。原信号 s_O 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 $s_l(\mathbf{X}, u; \omega), s_r(\mathbf{X}, u; \omega)$ とすれば

$$s_l(\mathbf{X}, u; \omega) = H_l(\mathbf{X}, u; \omega) H_{le}(\mathbf{X}; \omega) s_O(\omega) \quad (4)$$

$$s_r(\mathbf{X}, u; \omega) = H_r(\mathbf{X}, u; \omega) H_{re}(\mathbf{X}; \omega) s_O(\omega) \quad (5)$$

の関係がある。もし環境の影響が $H_{le} \equiv H_{re}$ であれば両耳間レベル差 (Interaural Level Difference, ILD) を z_{ILD} と表すと

$$z_{ILD} \equiv 20 \log |H_l(\mathbf{X}, u; \omega)| - 20 \log |H_r(\mathbf{X}, u; \omega)| \quad (6)$$

と近似でき z_{ILD} が u, ω の関数として $z_{ILD}(u, \omega)$ となり、ロボットの身体のみ影響で特徴づけられる。

両耳間位相差は両耳間時間差と似た特徴量である。両耳間レベル差と同様に左右のマイクロホンの伝達関数を考えると原信号 s_O に対して以下のような両耳間位相差 (Interaural Phase Difference) z_{IPD} を得ることができる。

$$z_{IPD} \equiv \angle H_l(\mathbf{X}, u; \omega) - \angle H_r(\mathbf{X}, u; \omega) \quad (7)$$

4.2 音源の尤度

音響特徴量の逆変換から直接、音源の位置 (以下 ${}^t\mathbf{X}$ と書く) を得ることは難しいため、事前に周波数成分を十分に含んだ試験信号を与え、位置と対応づけた音響特徴量を事前に計測し、あらかじめ測定された ILD, IPD のデータセットを規範として、受聴信号と照らし合わせることで定位を行う。ロボットの姿勢を u 、周波数領域での ILD, IPD を表したベクトルをそれぞれ z_{ILD}, z_{IPD} と表し音源定位を行うのに有効な周波数帯域を識別するインデックスベクトルを z_{ACT} と表す。またこれらのベクトルの周波数 ω に対する要素を $z_{ILD}(\omega)$ などと書くとする。ロボットが受聴する信号 $s_l(\mathbf{X}, u; \omega), s_r(\mathbf{X}, u; \omega)$ について、下記のように z_{ACT} を定める。

$$z_{ACT} = g(s_l(\mathbf{X}, u; \omega), s_r(\mathbf{X}, u; \omega)) \quad (8)$$

ここで g はインデックス関数であり、

$$g(x, y) = \begin{cases} 1 & \epsilon < x, \epsilon < y \\ 0 & (\text{上記以外}) \end{cases} \quad (9)$$

また ϵ は正の定数である。この操作によって ϵ より小さい値の受聴信号 $s_l(\mathbf{X}, u; \omega), s_r(\mathbf{X}, u; \omega)$ を除外することが

できる。これらの観測などの情報の下での位置 \mathbf{X} に音源の存在する ILD, IPD からの尤度 $l_{\text{ILD}}(\mathbf{X}|u, \mathbf{z}_{\text{ILD}}, \mathbf{z}_{\text{ACT}})$, $l_{\text{IPD}}(\mathbf{X}|u, \mathbf{z}_{\text{IPD}}, \mathbf{z}_{\text{ACT}})$ を

$$l_{\text{ILD}}(\mathbf{X}|u, \mathbf{z}_{\text{ILD}}, \mathbf{z}_{\text{ACT}}) = \frac{\sum_{i=1}^N \mathbf{z}_{\text{ACT}} \exp \{-10(\mathbf{z}_{\text{ILD}}(\omega_i) - \mathbf{z}_{\text{ILD}}^d(\mathbf{X}, u, \omega_i))^2\}}{\sum_{i=1}^N \mathbf{z}_{\text{ACT}}} \quad (10)$$

$$l_{\text{IPD}}(\mathbf{X}|u, \mathbf{z}_{\text{IPD}}, \mathbf{z}_{\text{ACT}}) = \frac{\sum_{i=1}^N \mathbf{z}_{\text{ACT}} \exp \{-10(1 - \cos(\mathbf{z}_{\text{IPD}}(\omega_i) - \mathbf{z}_{\text{IPD}}^d(\mathbf{X}, u, \omega_i)))\}}{\sum_{i=1}^N \mathbf{z}_{\text{ACT}}} \quad (11)$$

と定め、全観測からの音源位置の結合し尤度を $l(\mathbf{X}|u, \mathbf{z}_{\text{ILD}}, \mathbf{z}_{\text{IPD}}, \mathbf{z}_{\text{ACT}})$ として表せば、

$$l(\mathbf{X}|u, \mathbf{z}_{\text{ILD}}, \mathbf{z}_{\text{IPD}}, \mathbf{z}_{\text{ACT}}) = l_{\text{ILD}}(\mathbf{X}|u, \mathbf{z}_{\text{ILD}}, \mathbf{z}_{\text{ACT}}) l_{\text{IPD}}(\mathbf{X}|u, \mathbf{z}_{\text{IPD}}, \mathbf{z}_{\text{ACT}}) \quad (12)$$

のように与えるものとする。以下、この尤度に基づいて音源位置 ${}^t\mathbf{X}$ を推定する。

4.3 ベイズ推定による音源定位

時刻 k でのロボットの姿勢を u_k , 得られた観測量を z_k , 音源方向を ${}^t\mathbf{X}_k$ とする。また、時刻 1 から k までの量を $u_{1:k}$ のように添字で表すことにする。耳介の傾きは u_k の関数として表現できるとする。

時刻 1 から k までの制御量 $u_{1:k}$ と観測量 $z_{1:k}$ が得られ時刻 k における音源方向 ${}^t\mathbf{X}_k$ の存在についての信念が $B({}^t\mathbf{X}_k|z_{1:k}, u_{1:k})$ となっていたとする。 ${}^t\mathbf{X}_{k-1}$ が与えられたとするとモーションモデル h によって

$$B({}^t\mathbf{X}_k|z_{1:k-1}, u_{1:k}) = \int_{{}^t\mathbf{X}_{k-1}} h({}^t\mathbf{X}_k|{}^t\mathbf{X}_{k-1}, u_k) B({}^t\mathbf{X}_{k-1}|z_{1:k-1}, u_{1:k-1}) d{}^t\mathbf{X}_{k-1} \quad (13)$$

となる。本研究ではモーションモデルとして

$$h({}^t\mathbf{X}_{k+1}|{}^t\mathbf{X}_k) = \exp(-2({}^t\mathbf{X}_{k+1} - {}^t\mathbf{X}_k)^2) \quad (14)$$

を用いる。今、制御量 u_{k+1} の操作の下で、新しい観測 z_{k+1} が得られたとするとベイズの推定法 [Thrun 05] から $k+1$ における音源位置に関する信念分布は

$$B({}^t\mathbf{X}_{k+1}|z_{1:k+1}, u_{1:k+1}) = \frac{l(z_{k+1}|{}^t\mathbf{X}_{k+1}, u_{1:k}, u_{k+1}) B({}^t\mathbf{X}_{k+1}|z_{1:k}, u_{1:k})}{\int_{\mathbf{X}} l(z_{k+1}|\mathbf{X}, u_{1:k}, u_{k+1}) B(\mathbf{X}|z_{1:k}, u_{1:k}) d\mathbf{X}} \quad (15)$$

と漸的に求められる。これから、初期時刻から $k+1$ まで式 (15) を観測ごとに繰り返し適用することで音源の存在位置に関する分布を求めることができる。また音源位置の最尤推定値を ${}^t\hat{\mathbf{X}}_{k+1}$ とし次式で求める。

$${}^t\hat{\mathbf{X}}_{k+1} = \arg \max B({}^t\mathbf{X}_{k+1}|z_{1:k+1}, u_{1:k+1}) \quad (16)$$

5 音源定位におけるエゴノイズの影響

上述の定位法で音源位置を推定する際、能動耳介の動作に伴うエゴノイズが与える影響について、実際の実験データを基に考察する。

5.1 収録実験

収録は Fig.10 に示す奥行き 6.0m, 幅 5.9m, 高さ 2.5m の居室で行った。音源として図の奥に見えるようなスピーカーを用いている。

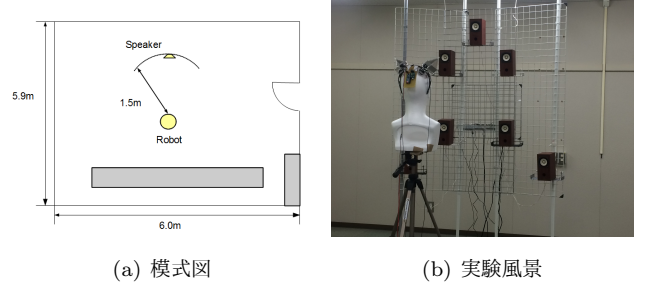
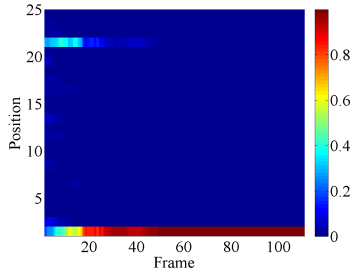


Figure 10: 実験環境

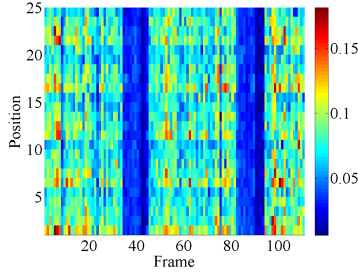
能動耳介を人形頭部に取り付け能動耳介の正面方向 1.5m, 正面高さ 1.5m を基準としロボットから見て方位角方向, 仰伏角方向ともに $-20^\circ \sim 20^\circ$ の 10° 刻みで計 25 点から白色雑音, 音楽をそれぞれ印加し計測を行い, さらにそれぞれの音源位置で耳介の姿勢を変化させながら収録を行った。姿勢は耳介を起こした姿勢と伏せた姿勢とを繰り返す動作パターンとした。

5.2 音源定位結果

音源が左下 (位置番号 1) にある時に収録したデータのスペクトログラムを Fig.11 に示し, このときの音源定位を行った結果を Fig.11 に示し縦軸は推定した音源位置のインデックス番号, 横軸は時間区間を示しており色によって信念 B を表している。



(a) 音源の存在確率



(b) 尤度

Figure 11: 定位結果

この Fig.11(a) 例では最初の 35 フレームまでにエゴノイズがなかったため、エゴノイズが生じた段階で、すでによく定位が行えており、その後のエゴノイズの影響があっても正解音源位置の検出が可能であることがわかる。しかし Fig.11(b) の 40 フレーム付近、90 フレーム付近の時間区間は尤度の値が他の時間区間と比べて低い値を示し、明確な音源方向は得られていない。これらの時間区間はエゴノイズが発生した時間区間と一致することから、事前分布の状態によってはエゴノイズが音源定位に悪影響を与えている可能性がある。そこで、これらのフレーム付近として、エゴノイズ前の 5 フレームとノイズ後の 20 フレームに着目し、

1. エゴノイズ区間の前後フレームとエゴノイズ区間を用いて音源定位を行う場合
2. エゴノイズ区間では式 (13), 式 (14) に替えて、一切更新を行わない次のモデル

$$p({}^t\mathbf{X}_{k+1}|z_{1:k+1}, u_{1:k+1}) = p({}^t\mathbf{X}_k|z_{1:k}, u_{1:k}) \quad (17)$$

を適用し、エゴノイズ区間の前後の時間区間では式 (13), 式 (14) 用いて音源定位を行う場合

3. 2 と同様にエゴノイズ区間をスキップするがエゴノイズ区間中にも音源の存在確率分布の更新にモーションモデル式 (14) を用いて音源定位を行う場合

の 3 つについて検討することとした。定位性能の評価には理想的な信念の分布と推定で得られた分布間の KL 情報

Table 1: KL 情報量の変化

	1 の場合	2 の場合	3 の場合
KL 情報量の平均値	3.17	2.77	3.11

量を用いた。KL 情報量は

$$D_{\text{KL}}(R||Q) = \sum_i R(i) \log_2 \frac{R(i)}{Q(i)} \quad (18)$$

と定義され、本研究では R に正解位置の確率が 1 でそれ以外が 0 となる確率分布、 Q に音源の信念分布 B を与えた。

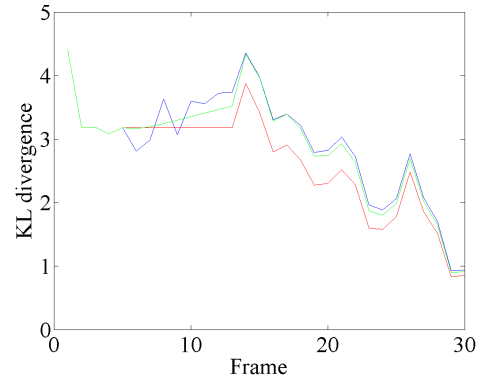


Figure 12: KL 情報量

この場合 D_{KL} が小さい程性能が良いことを示す。Fig.12 に正解音源位置 1 番での結果を示す。横軸が時間区間、縦軸が KL 情報量、青は 1 の場合、赤は 2 の場合、緑は 3 の場合を示している。5 フレームより後のエゴノイズ区間では 1,3 の場合は KL 情報量が増加し定位性能が低下しているが 2 の場合はエゴノイズ区間をスキップしているため KL 情報量が増加せず 1,3 の場合よりも良好な定位性能を維持している。このような評価を 25 点の異なる音源位置に対して実験し、その KL 情報量の平均を Tab.1 にまとめた。Tab.1 よりエゴノイズ区間を音源定位に用いない 2 の場合が KL 情報量が最も小さく音源定位性能が良い。このことはエゴノイズ区間での観測が信念の分布を理想的な分布から遠ざけていることを示しているが能動耳介の音源定位性能を下げていると言える。

6 まとめ

能動耳介動作時のエゴノイズが音源定位性能に与える影響を調べた。エゴノイズ区間を音源定位の際に取り除くことで KL 情報量の増加を防ぐことができるため、エゴノイズが音源定位性能を下げるのが能動耳介システムでも確認できたが他の音源や動作パターンについても検証を行うことが必要であると考えている。

参考文献

- [大賀 07] 大賀 寿郎, 山崎 芳男, 金田 豊: 音響システムとデジタル処理, コロナ社, 2007.
- [Garas 00] Garas J.: Adaptive 3D Sound Systems, Kluwer, 2000.
- [章 08] 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推定, 日本機械学会論文集 C 編, Vol.74-739, pp. 642-649, 2008.
- [西野 07] 西野 隆典, 井上直哉, 伊藤克亘, 武田一哉: 両耳間音圧差の包絡を用いたガウス分布に基づく音源方向推定, 日本音響学会学会誌 C 編, Vol.63 no.1, pp. 3-12, 2007.
- [Populin 98] Luis C. Populin, Tom C. T. Yin: Pinna Movements of the Cat during Sound Localization, Journal of Neuroscience, 18(11), pp. 4233-4243, 1998.
- [金 12] 金天海, 中臺一博, 辻野広司: ウェアラブル人工耳介-音追従動作による音源定位能力の向上-, 日本ロボット学会第 30 回記念学術講演会, 3D1-1, 2012.
- [野田 12] 野田佳孝 公文誠: 二つの能動耳介による正中面内の音源方向推定, 第 13 回システムインテグレーション部門講演会 (SI2012), pp1643-1646, 2012
- [本田 85] 本田 学: 耳珠のはたらき, 耳鼻臨床, 78. 増 1, pp789-801, 1985.
- [Thrun 05] S. Thrun, W.Burgard, and D.Fox: Probabilistic Robotics, MIT Press, 2005,