第 4 章

動作時のノイズキャンセルと音源定位

2 章で述べたように, 従来のロボットは, そのほとんどが, 対象となる音源数を 1 つに仮定していることに加え, 自分自身が出すノイズを避けるために, 一旦, 静止してから音を聴く "*stop-perceive-act*" 原理に従う, 対象音源の音量をノイズが無視できるくらい大きな音と仮定する, あるいは, 話者の口元に備えたマイクロホンを使って, 部屋の音響環境やロボット自身が作り出すノイズによる影響を避けるといった手法を使っている. つまり, 動きながら音を聞くためには, まず自分自身の作り出すノイズを抑制もしくはキャンセルする必要がある. 本章では, このノイズキャンセルについてロボットの外装を用いる方法を考案し, 工学的な実装を通じてこれを評価する.

4.1 外装を利用したノイズキャンセル

アクティブオーディションを実現するため, 図 4.1 に示すような実験的なシステムを構築した. 入力は, 複数の音源からの混合音とし, これを 4 本のマイクロホンで, 同期を取りながら収音する. 各チャネルに対し, 短時間高速フーリエ変換 (*Fast Fourier Transform, FFT*) による周波数解析を行った後, 内外のマイクの強度差を利用して内部音抑制を行う. さらに, ピーク抽出, 音源分離, 音源定位を行うことによって, 音響ストリームを分離する. また, 分離した任意の音響ストリームに注意を傾け, その音源方向を向くことができる. 以下にシステムの内部音抑制機能と音響ストリーム分離機能を説明する. なお, 周波数解析に関する検討は, 本研究では, ピークの抽出法と密接に関わるため, 次章の 5.2 節で行う.

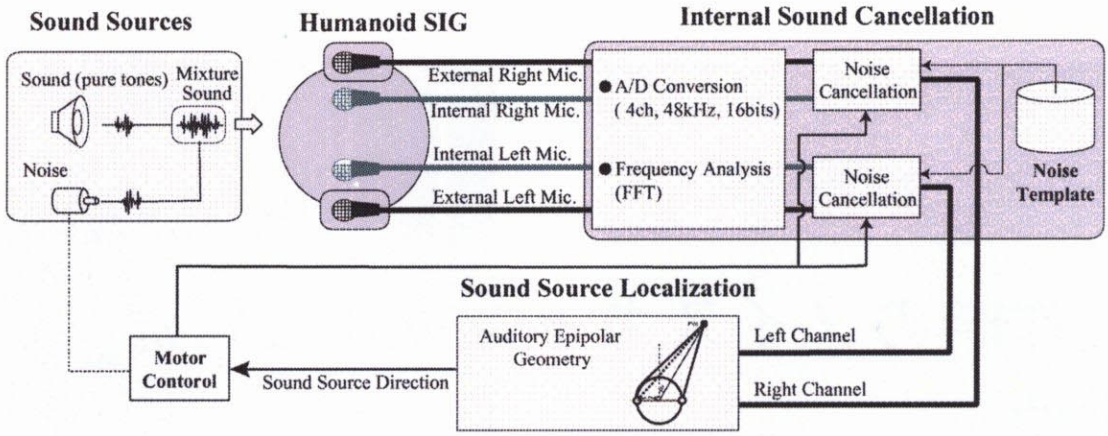
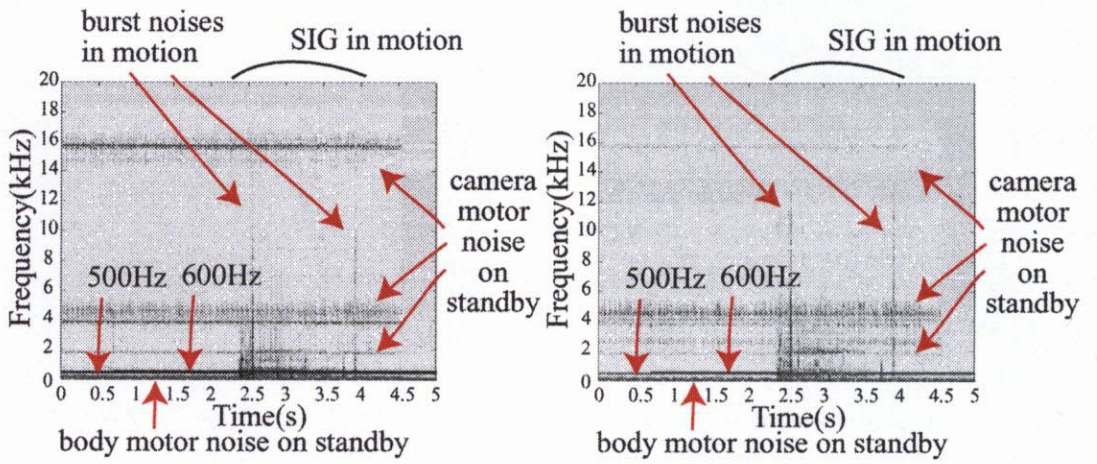


図 4.1: 内部音キャンセルシステムの構成図

4.1.1 内部音の特徴

まず, 内部音の特徴を把握するために, *SIG* 動作中に収音される音響信号を解析した.

図 4.2 は, *SIG* の内外のマイクロホンを用いて 48 kHz, 16 bit でサンプリングし, 4096 点の FFT で 20 ms 毎に周波数解析した例である. 図 4.2 の状況では, 500 Hz と 600 Hz の静的な純音が外部音源から発せられており, 2 秒過ぎから 4 秒過ぎにかけて *SIG* が, 胴体の回転 (Motor 4) を行っている. 各グラフは, 周波数, 時間, パワーの 3 軸からなるスペクトログラムであり, 色の濃い部分がパワーが高いことを示している.



a) 内部マイクロホンのスペクトログラム b) 外部マイクロホンのスペクトログラム

図 4.2: 動作時の *SIG* のモータノイズ例

図 4.2 から, カメラのスタンバイノイズ, 動作時のモータのバーストノイズ, モータのス

タンバイノイズなど様々なノイズが存在していることがわかる. そこで, カメラ, モータのノイズを個別に 10 回ずつ計測し, 実際に, どの程度大きなノイズが出ているかを解析した. 解析の結果, *SIG* の内部音は, 次のような特性を有していることが判明した.

- カメラのモータ動作音は無視できるレベルであるが, カメラの定常ノイズは, 左右それぞれ, 3.7 dB ある. また, ノイズは比較的高い周波数帯域に限られている.
- ロボット駆動モータは, 定常時には 5.6 dB 程度であり, 比較的低い周波数帯域に限られている.
- モータの動作時のノイズは平均で 23 dB であるが, ストッパー, ケーブルや外装の摩擦, カバーの結合部分の軋みなどの原因で, さらに大きなパワーを持ったバースト的なノイズが発生し, 広域に渡って大きな影響を与えている.

スタンバイ時のノイズについては, 周波数帯域制限を行うことでも対応可能であろう, しかし, 一番処理に影響を及ぼす音はモータの動作時のバーストノイズであり, 少なくともバーストノイズについては, 何らかの抑制が必要であることがわかる.

4.1.2 一般的なノイズ抑制法の適用

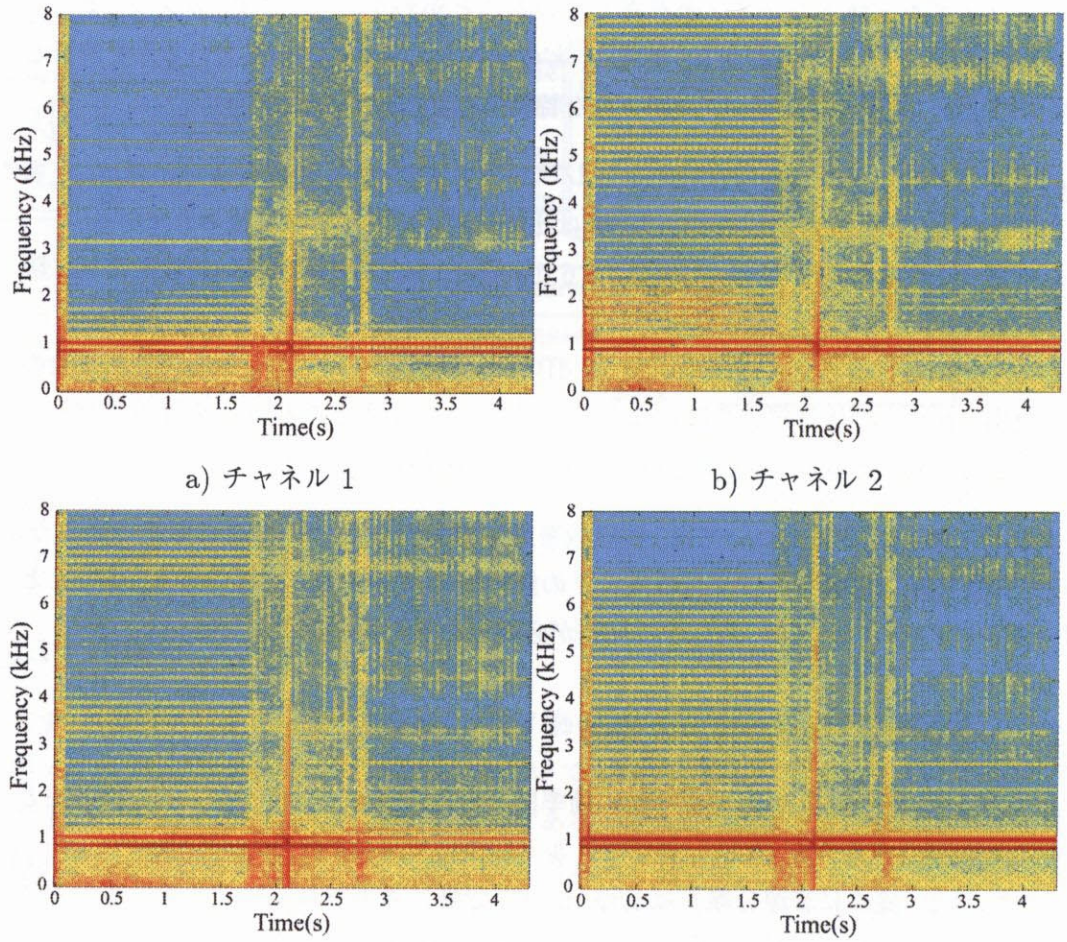
どのような方法が *SIG* の内部音抑制に効率的なのだろうか. 一般的なノイズ抑制法として知られる独立成分分析 (ICA), 適応フィルタについて検討を行った.

ICA は, 近年, 盛んに音源分離をはじめ, 様々な分野に適用されている信号分離の手法である [78, 95]. ICA は統計的に独立な信号が混合した観測信号から, その混合割合や方向などの情報がない場合でも, 元の独立な信号を推定することが可能である.

そこで, 奥乃, 池田らの ICA による分離プログラムを利用して, 4 本のマイクロホンの入力, つまり, 4 チャネルの ICA をノイズ問題に適用した. 図 4.3 は, 分離後の各チャネルのスペクトログラムを示している. なおプログラムの制約上の問題から, 16kHz にダウンサンプリングしたものに対して処理を行っている.

図 4.3 では, 分離がうまく行われていないことがわかる. 一般に ICA では, 音源数とマイクロホンの数が同じ場合でないと効果が少ないことに加え, マイクロホン自体が動作して音源方向が変化してしまう場合に, パラメータをインクリメンタルに適応させる必要があり, 効率的な分離は難しい.

SIG の場合では, 途中で動作を行い, マイクロホンの向きが変わってしまうこと, マイクロホン数 (4 本) に比べノイズ源数が多く, かつ, すべてのノイズ源が常に音を発しているわけではないこと, 4 本のマイクロホン間距離が短いことなどの理由により, 効果的なノイズ抑制を行うことが困難であった. また, 使用した ICA プログラムは, 計算量が大きく実時間での動作が難しいことからもロボットへの適用は難しい.



a) チャネル 1　　b) チャネル 2

c) チャネル 3　　d) チャネル 4

図 4.3: ICA によるノイズの抑制

次に, アクティブノイズコントロール (*Active Noise Control, ANC*) で使用される適応フィルタを適用した. ANC は 2 章で述べたように, 騒音と逆位相の音を適応的に発生させることによって騒音を打ち消し, ノイズを抑制する手法である [41, 87]. 適応フィルタは, 音源定位における IPD 算出誤差を極力抑えるため, 線形位相フィルタを用いた. 具体的には, 次数 100 の 有限インパルス応答 (*Finite Impulse Response, FIR*) フィルタを使用し, パラメータの更新は, 最小二乗法の適応アルゴリズムを用いた. 図 4.4 に結果を示す. 図 4.4a) は, 外部マイクからの入力, 図 4.4b) は, 内部マイクからの入力, 図 4.4c) が適応フィルタによるノイズ抑制の結果である. 内部マイク入力を用いてからノイズ部分外部マイク入力から適応的にノイズ部分を抑制し, 500Hz と 600Hz の音を強調することを期待したが, ノイズが抑制されないばかりか, 強調すべき 500Hz と 600Hz の音が抑制されてしまっていることがわかる. この理由として, 適応フィルタは, 純粋にノイズを取得で

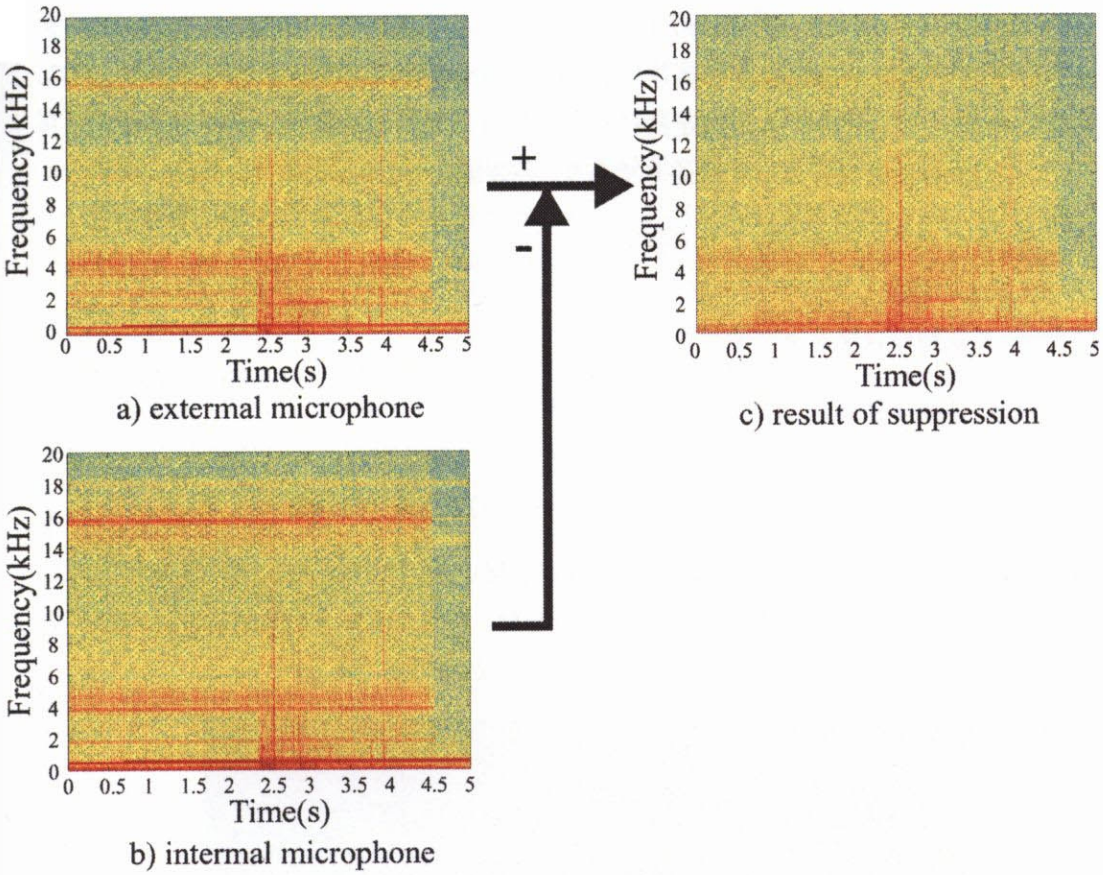


図 4.4: FIR 適応フィルタによるノイズ抑制

きる場合には有効に働くが, 外装がロボット内部を完全に密閉しているわけではないために, ノイズのみを収音することが難しく適応フィルタが有効に働かなかったためと考えられる.

以上より, ICA や適応フィルタといった信号処理の分野でよく使われるノイズ抑制は, この場合には効果的な内部音抑制が難しいことがわかった. さらに, 両者とも分離の結果として得られた信号の位相情報が歪んでしまい, 後処理の音源定位などを考慮すると *SIG* の内部音の抑制には適さなかった.

4.1.3 バーストノイズキャンセルフィルタ

そこで, 汎用的なノイズキャンセルではなく, 音源定位を行うことを念頭に置き, 主に動作時に, 広い周波数帯域に渡り大きな影響を与えるバーストノイズをキャンセルするフィルタをヒューリスティクスに基づいて構築することにした. このフィルタはバーストノイズと見なされるサブバンドをキャンセルし, そのようなサブバンドは以後の処理には用い

ないものとした. サブバンド単位でのフィルタリングを用いたのは, ICA や適応フィルタのように, ノイズ抑制の計算誤差で生じる信号の位相情報の歪みを防ぐためである. 特に, 低周波域では, 小さな歪みでも音源定位に与える影響は大きくなる. バーストノイズを判定するための具体的な条件は以下の通りであり, 各数値は実験的に求めた.

1. 内部音のパワーが外部音よりも強い.
2. 20 以上の連続したサブバンドで一定値 (30 dB) 以上のパワーが観測される.
3. モータ駆動のコマンドが発行されている.

この方法の詳細な評価は, 4.3 節で述べるが, ICA や適応フィルタと比べ, 良好なノイズキャンセル効果が得られている. しかし, この方法では取り除くことができないバーストノイズも存在した. そこで, さらにその原因を探るため, 外装の音響解析を行った.

■外装の音響解析 測定には図 4.5 に示す無響室を用いた. これは, 広さ約 10m 四方の無響室であり, 四方の壁, 床, 天井には突起状の吸音材 (グラスウール) が貼り付けられているため, 125 Hz 以上の音の反響はほぼ 0 とみなすことが可能である[*1].

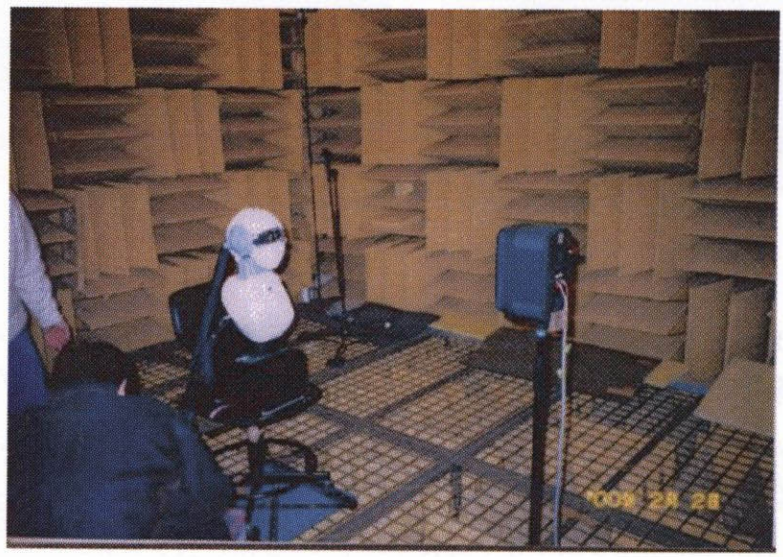


図 4.5: 無響室

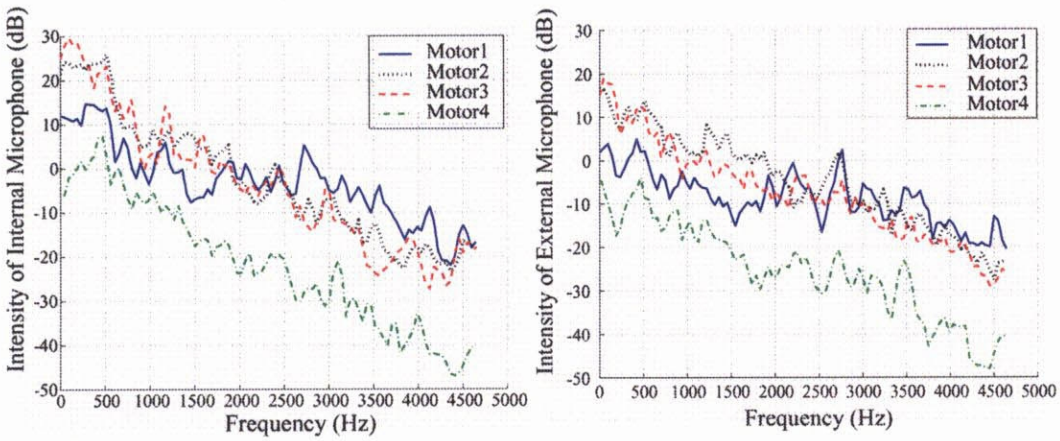
具体的な測定項目は以下の通りである.

1. 内外のマイクロホンを使用して, 各モータノイズの周波数応答を計測した. 各モータを秒速 14.9° で稼動域である ±45° の範囲で動かした. 測定は 3 度行い, 平均を算出した.

[*1] 日東紡音響エンジニアリング株式会社所有 http://www.noe.co.jp/

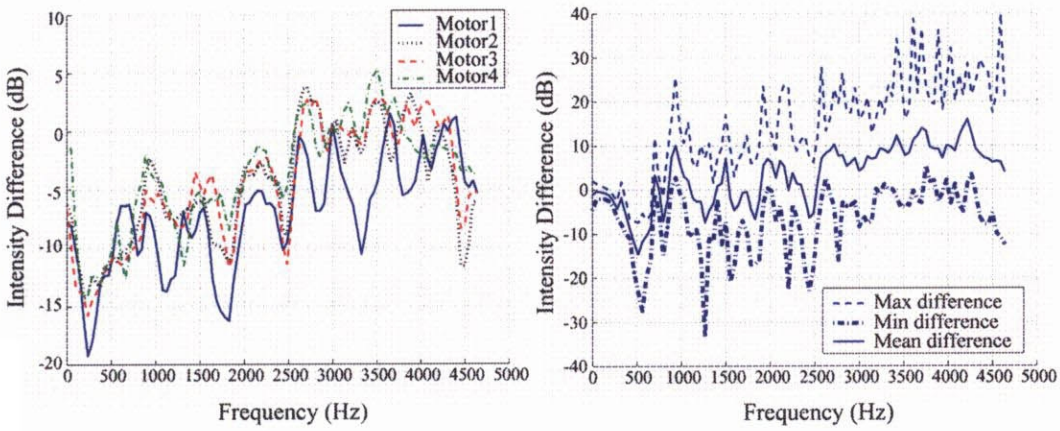
2. 内外のマイクロホンの強度差を計測した．図 4.7a) は各モータノイズの強度差を示している．モータの動作条件は周波数応答の測定と同じである．図の縦軸は，外部マイクロホンに対する内部マイクロホンの強度を dB で示している．図 4.7b) は，外部の音源から音響信号を収音した場合の結果である．具体的な音源としては，全周波数帯域での周波数応答を調べるため，インパルス音源を使用した．インパルス応答は，水平角が 0°, ±45°, ±90°, 180° の 6 点，仰角が 0° と 30° の 2 点の組み合わせで，12 点で測定した．

1 の測定結果を図 4.6a), b) に示す．また，2 の測定結果を図 4.7a), b) に示す．



a) 内部マイクロホン　　b) 外部マイクロホン

図 4.6: モータノイズの周波数応答



a) モータノイズ　　b) 外部音源

図 4.7: 内外のマイクロホンの強度差

1. 図 4.6a), b) に示されるように，モータノイズは広帯域に及んでおり，内部マイクロ

ホンで 30 dB 以下程度である. 外部マイクでは, 20 dB 以下である.

2. 図 4.6a) より, モータノイズは 2.5 kHz 以上では, 内部マイクロホンの方が外部マイクロホンより強いパワーで収音されていることがわかる. これは, 外界からの音が外装で遮断されるので, 内部マイクロホンでモータノイズを収音することが容易になっていることを示している.
3. 2 kHz 以下では, 外部マイクのほうが内部マイクより収音信号が強くなることがある. 特に, 図 4.7b) のように 700Hz 以下ではこの傾向が顕著である. これは, 外装内の共鳴現象によるものである. 外装の内部の径は 約 18cm であり, これは, 500Hz の $\lambda/4$ に相当する. つまり, 500Hz を中心に共鳴がおきている. 図 4.7a) にも同様に 500Hz を中心とした共鳴が見られる.
4. 図 4.7a) と図 4.7b) を比較すると, 総じて, 内部音は外部音よりも約 10dB 強く収音されることがわかる. 従って, 外装による音響的な遮蔽効果は約 10dB であるといえる.

■外装の音響効果を考慮したノイズキャンセル 外装の音響測定結果を利用して, バーストノイズキャンセルフィルタを改良した. まず, 音響測定結果をテンプレートとしてシステムに格納する. 具体的には, 内部および外部のマイクロホンで収音した各モータノイズのパワースペクトルを格納する. これらのテンプレートを利用して, バースト的なモータノイズが発声しているかどうかを判断する. バーストノイズが発生した場合, マイクロホンの位置が比較的モータに近いため, ノイズのパワーは非常に強くなる. このため, 多少の外部音が存在する場合でも, 収音信号のパワースペクトルがテンプレートに似ていると判断されれば, その時刻にバーストノイズが発生しているとみなすことができる. そこで, 下記に示す条件が満たされた場合, そのサブバンドはノイズであると判断する.

1. 内部と外部のマイクロホンの強度差がモータノイズの強度差に近い. つまり, 下記の条件を満たすサブバンドが, 0 Hz – 3000 Hz までに全体の 75% 以上あること. ここで, S_{in}, S_{out} は, 内部マイクロホン, 外部マイクロホンで収音した入力であり, I_{min}, I_{max} は, 図 4.7a) の強度差の最小値, 最大値を示す. また, i はサブバンド番号である.
$$I_{min}[i] \leq S_{in}[i] - S_{out}[i] \leq I_{max}[i]$$
2. 強度とパターンが計測したモータノイズの周波数応答に近い. つまり, 下記の条件を満たすサブバンドが, 0 Hz – 3000 Hz までに全体の 40% 以上あること. ここで, N_{in}, N_{out} は, それぞれ, 図 4.6a), b) におけるモータノイズのテンプレートパターンを示す. なお, 本研究では, Motor 4 のみを対象としている.
$$|S_{in}[i] - N_{in}[i]| < 2\,dB$$

$$|S_{out}[i] - N_{out}[i]| < 2\,dB$$

3. モータが駆動中である.

バーストノイズのキャンセルは, S/N 比の高い音響信号を抽出することが目的ではなく, 安定した音源定位を行うことが目的であるため, 評価については, 4.3 節において, 音源定位と共に述べるものとする. 次節では, 音源定位法について述べる.

4.2 未知環境における音源定位

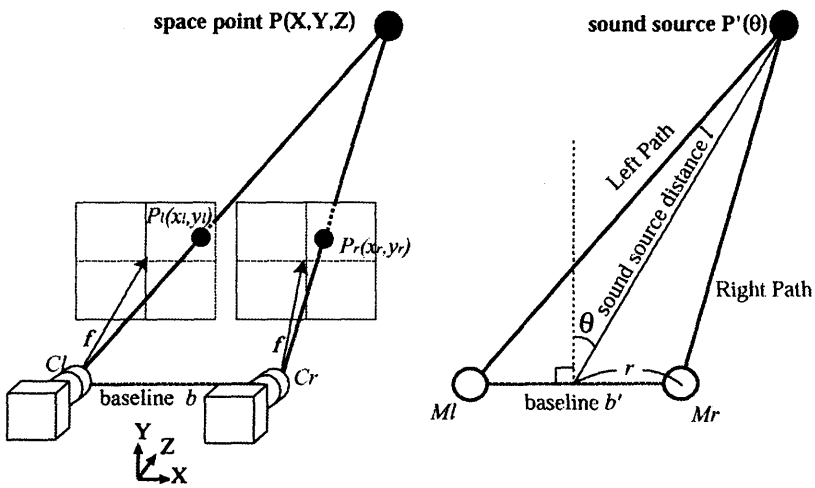
音源定位問題は, ロボット頭部に 2 本のマイクを設置するという前提で行う. 一般に両耳聴の研究では, 音源定位問題に対し, 頭部伝達関数 (HRTF) を用いることが多い. HRTF は一般に, 無響室において様々な方向から, インパルス応答を測定して得られる測定関数であるが, ロボットでは, HRTF を利用した場合, 以下の 2 つの問題が生じる. このため, ロボットにおける音源定位では, HRTF を使用しない音源定位の手法が必要である.

1. **HRTF に基づく音源定位**. 通常, HRTF は無響室で計測される. しかし, 無響室で測定した HRTF を実環境の音源定位に適用すると, 部屋の反響の影響により, 著しく定位精度が低下する. さらに, 実環境では, 頭部の形状によって生じる純粋な HRTF と部屋の反響などによる伝達関数を区別して計測することは難しく, 部屋の反響まで考慮した広義の HRTF が必要である. ところが, このような HRTF の測定は時間を要するばかりでなく, 部屋に新しい家具が設置されたり, ロボットの位置や向きが変わったりするなど, 環境が変わるたびに再測定が必要となる.
2. **HRTF に基づく音源追跡**. HRTF は測定関数であるため測定点は離散値にならざるを得ない. そのため, HRTF による音源定位を利用して, 連続的に動く音源の追跡を実現することは難しい. また, 音響ストリーム分離という観点からみても連続的な定位情報が得られないことは致命的である.

4.2.1 聴覚エピポーラ幾何による音源定位

HRTF に依存しない音源定位方法として, 聴覚用のエピポーラ幾何を用いる. これは, ステレオビジョンの基本的な定位法であるエピポーラ幾何を音源定位に適用したものである. 以下に, 聴覚エピポーラ幾何について詳細に述べる.

■視覚と聴覚のエピポーラ幾何 図 4.8a) に示すような同一の焦点距離を持ち, 光軸が並行でレンズ面が同一平面状にある 2 台のカメラを使った単純なステレオカメラを考える. 空間上の点 $P(X, Y, Z)$ の左右のカメラに対するそれぞれの投影面上の座標を $P_l(x_l, y_l)$,



C_l, C_r: カメラの中心, M_l, M_r: マイクの中心

図 4.8: 視覚と聴覚のエピポーラ幾何

$P_r(x_r, y_r)$ とすると, P の座標は,

$$X = \frac{b(x_l + x_r)}{2(x_l - x_r)}, Y = \frac{b(y_l + y_r)}{2(x_l - x_r)}, Z = \frac{bf}{x_l - x_r}$$

として得られる [34, 98]. ここで f はカメラレンズの焦点距離, b はベースラインを示している. (P, C_l, C_r) によって構成されるエピポーラ面と投影面の交線であるエピポーラ線上に, 左右の画像の対応点が存在するという性質を利用して, ステレオマッチングを行う. *SIG* では, $y_l = y_r$ という関係が保たれるため, マッチングは比較的容易である.

図 4.8b) に示すように, 聴覚でも 2 本のマイクロホンを利用して, マイク間の距離差を利用した定位が可能である. 実際には, 収音したデータに対し, 周波数解析を行ったのち, 時間差 (距離差) に対応する左右の対応する周波数の位相差情報を利用して定位を行う. 同時刻における左右のチャネルのスペクトルをそれぞれ S_l, S_r とすると, IPD は,

$$\triangle\varphi = \arctan\left(\frac{\Im[S_r(f)]}{\Re[S_r(f)]}\right) - \arctan\left(\frac{\Im[S_l(f)]}{\Re[S_l(f)]}\right) \tag{4.1}$$

により得られる. ここで f は周波数, $\Re[S]$, $\Im[S]$ は, それぞれスペクトル S の実数部と虚数部を示す. 得られた IPD より, 音源方向 θ は以下の式により計算することができる.

$$\sin\theta = \frac{v}{2\pi fb}\triangle\varphi \tag{4.2}$$

なお, v は音速であり, 340 m/sec としている.

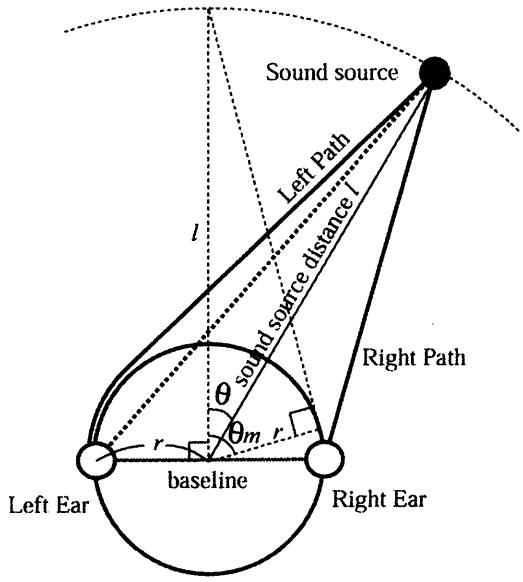


図 4.9: 頭部形状を考慮した聴覚エピポーラ幾何

■頭部形状を考慮した聴覚エピポーラ幾何 実際には, ロボットの場合, 頭部形状に応じた音の回り込み現象が発生する. そこで, ロボットに適用する聴覚エピポーラ幾何では, 図 4.9 に示すように, *SIG* の頭部を球体であると仮定して頭部の回りこみを考慮する. 図 4.9 では, 音源から左右のマイクへの距離差 (両耳間距離差) Δd は,

$$\Delta d = \frac{v}{2\pi f}\Delta\varphi \tag{4.3}$$

と表すことができるので, *SIG* の頭部を球体とみなし, 頭部形状による回折を考慮すれば, 式 (4.4)–(4.6) に示すように Δd は, θ と l の関数 D として表すことができる.

$$D(\theta,l) = \begin{cases} r\left(\frac{\pi}{2} - \theta - \theta_m\right) + \delta(\theta+\pi, l) \\ \qquad\qquad \left(0 \le \theta + \frac{\pi}{2} < \theta_m\right) \\ -2r\theta \\ \qquad\qquad \left(|\theta| \le \frac{\pi}{2} - \theta_m\right) \\ -r\left(\frac{\pi}{2} + \theta - \theta_m\right) - \delta(\theta, l) \\ \qquad\qquad \left(0 \le \frac{\pi}{2} - \theta < \theta_m\right) \end{cases} \tag{4.4}$$

$$\delta(\theta,l) = \sqrt{l^2 - r^2} - \sqrt{l^2 + r^2 - 2rl\sin\theta}, \tag{4.5}$$

$$\theta_m = \arccos\frac{r}{l}. \tag{4.6}$$

ここで, 式 (4.4)–(4.6) で, θ, l を変化させた時の D の振舞いをシミュレーションした結

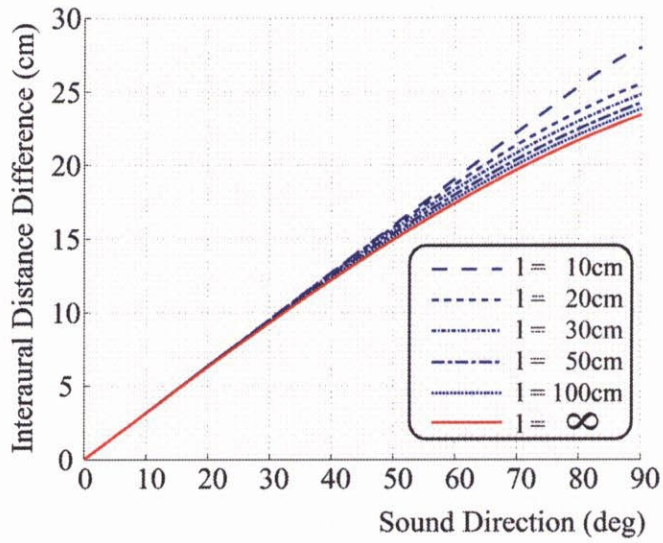


図 4.10: IPD と音源距離の関係

果を図 4.10 に示す. θ が大きくなるほど l の影響は大きくなるが, 50 cm 以上では l による影響は小さくなる. そこで, l を無限大と仮定すると D は θ のみの関数として以下のように定義される.

$$\begin{aligned} D(\theta) &= \lim_{l\to\infty} D(\theta, l) \\ &= r\,(\theta + \sin\theta)\,. \end{aligned} \tag{4.7}$$

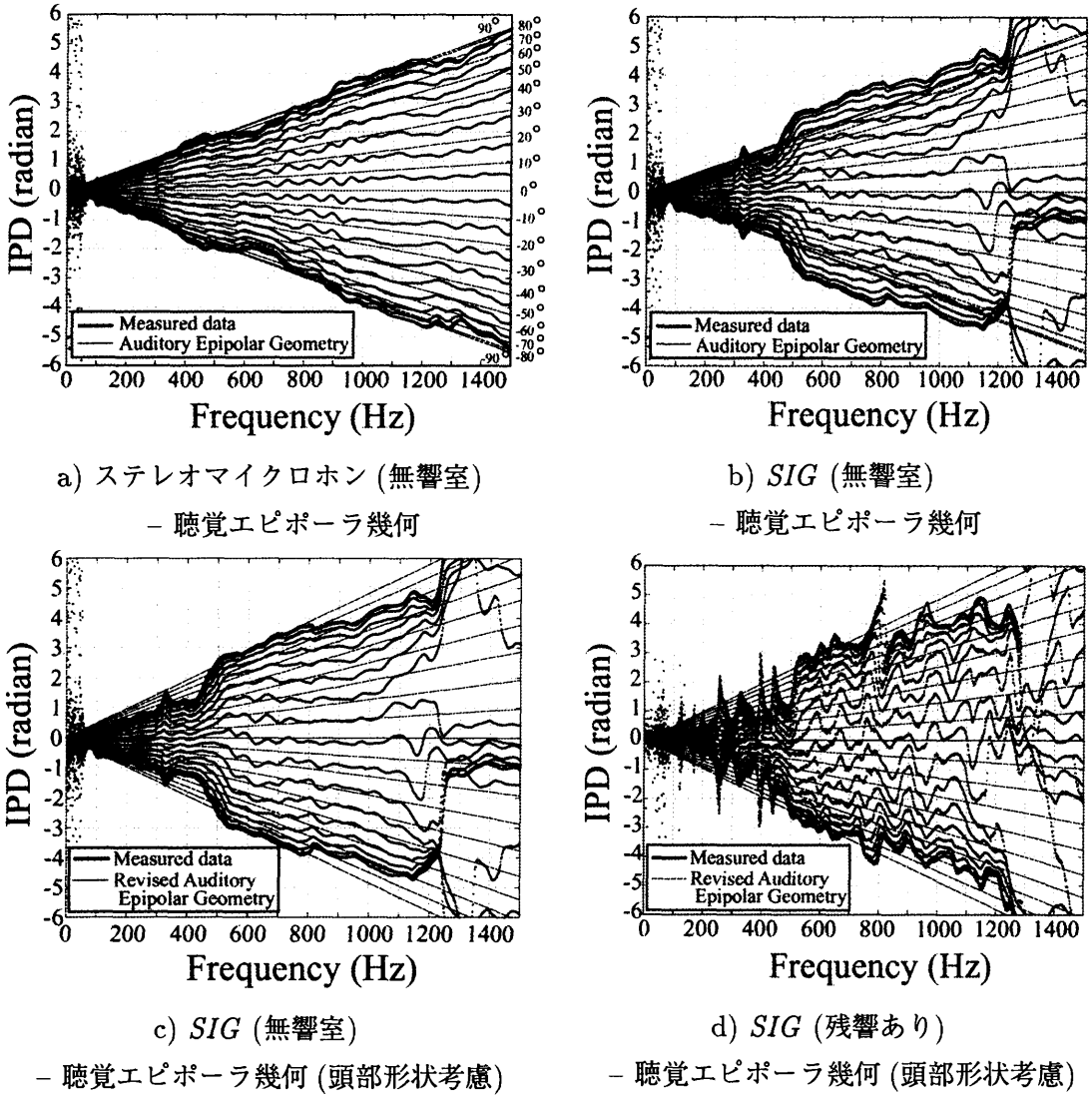
近接学 (*Proxemics*) の観点からも 50cm 以下は "intimate distance" と呼ばれ [40], 人間とのインタラクションなど実際の利用にあたり, l が無限大という仮定をおくことは妥当であろう. 最終的に, 音源方向 θ は, 式 (4.3), (4.7) を用いて, 以下の式により求めることができる.

$$\theta = D^{-1}\left(\frac{v}{2\pi f}\Delta\varphi\right). \tag{4.8}$$

音源を視覚用のエピポーラ幾何によって定位できれば, 図 4.8 において, $P = P'$ となる. この場合, *SIG* では, カメラとマイクのベースラインは並行であるため, ビジョンによる定位結果 P は, 容易に音源方向 θ に変換できる. 実際には, 本研究では, もう一歩進めて, ストリームを利用した視聴覚の統合を行っている. この詳細については 6 章で述べる.

■実環境における聴覚エピポーラ幾何の課題 では, 聴覚エピポーラ幾何は, 実環境での観測結果とどの程度一致するのであろうか. 実環境では, 一般に IPD と IID の値は以下の 3 つの要素に大きく影響を受ける.

1. 音源と左右の耳までの距離の差



a) ステレオマイクロホン (無響室) – 聴覚エピポーラ幾何

b) *SIG* (無響室) – 聴覚エピポーラ幾何

c) *SIG* (無響室) – 聴覚エピポーラ幾何 (頭部形状考慮)

d) *SIG* (残響あり) – 聴覚エピポーラ幾何 (頭部形状考慮)

図 4.11: 聴覚エピポーラ幾何による IPD 推定値と IPD 測定値の対応

2. ロボットの頭部や胴体の反響
3. 部屋の音響環境

そこで, 実際に, これらの影響を調べるため, *SIG* の音響環境の測定を行った. さらに, 境界要素法 (*Boundary Element Method, BEM*) を用いて音響環境のシュミレーションを行うツール SYSNOISE[*2] を併せて使用し, 影響を与えている要素の原因を解析した.

測定は, *SIG* の正中面から水平角で ±90° の範囲で 10° ごとにインパルス応答を計測した. 仰角については, *SIG* の両耳のマイクロホンを結ぶラインと同じ高さに固定した.

[*2] Computational Vibro-Acoustics software, Copyright LMS International 1999.

また, 比較のために空間上に *SIG* の両耳間距離と同じ距離を隔てたステレオマイクロホンについても同様のインパルス応答の計測を行った.

図 4.11a) の太線は, *SIG* を用いず, ステレオマイクロホンで, 音響測定を行った結果である. この場合, 外装の影響を考える必要はない. 図中で, 凡例に AEG と示されている細線は, 聴覚エピポーラ幾何を示す式 (4.2) によって, 推定された IPD を示している. 聴覚エピポーラ幾何はステレオマイクロホンの測定結果との対応がよいことがわかる.

図 4.11b) は, *SIG* のマイクロホンを用いた音響測定結果と, 式 (4.2) によって, 推定された IPD の関係を示している. 300Hz 以上の周波数帯域では, 式 (4.2) は測定結果との対応が悪くなっていることがわかる. これは, HRTF つまり, *SIG* の頭部や胴体の反響の影響である. また, 1200 Hz では, 推定がうまく働かないことがわかる. これは, *SIG* の両耳間距離は 18cm であることに起因し, この周波数帯域より高い周波数では, 1 周期回り込みによって曖昧性が解消できなくなるためである.

図 4.11c) は, *SIG* の音響測定結果と, ロボットの頭部形状を考慮した聴覚エピポーラ幾何である式 (4.8) で推定した IPD の対応を示している. 図 4.11b) に比べ明らかに対応関係が良好になっていることがわかる. つまり, 式 (4.8) を用いることで外装の影響の問題が, ある程度解決されている.

図 4.11d) は, 無響室でなく, 反響のある部屋で計測した *SIG* の音響測定結果と式 (4.8) による推定結果の対応を示している. この部屋は, 約 $3m \times 3m \times 2m$ の部屋で, 部屋の壁, 天井に吸音材を設置してあり, 音響的には若干 *dead* になっている (残響時間 0.2 – 0.3 秒程度). また, 部屋の天井には, 6 つの照明が備わっている. このような部屋でも, 音響効果により, 計測した IPD が全体的に歪んでいることがわかる.

次に, SYSNOISE を用いて, 部屋の音響効果の測定を行った. 図 4.12 は, 30° の方向からのインパルス応答を測定することによって得られた IPD と IID を示している. また, “SYSNOISE (no floor)” とラベリングされた IPD と IID は, *SIG* 頭部の 3 次元メッシュデータを用いてシミュレーションした結果である. シミュレーションの結果では, 300 Hz と 400 Hz の間にピークがあることがわかる. このピークは, *SIG* を用いた計測結果にも同様のピークが見られることから, *SIG* の頭部に起因していることがわかる.

また, *SIG* の頭部の 1 m 下に広さが無限の床があることを仮定して IPD と IID のシミュレーションを行った. 床がない条件でシミュレーションを行った結果と比較し, 余分なピークがいくつか認められる. このように, 簡単な床を仮定しただけで, IPD や IID の振る舞いが大きく変わることがわかる.

以上, 音源と左右の耳までの距離の差とロボットの頭部や胴体の反響に関しては, 聴覚エピポーラ幾何は十分適用可能であることを示した. HRTF を利用した定位を行うことも可能ではあるが, 実環境では, その都度, 測定が必要であり, 実用的とはいえない.

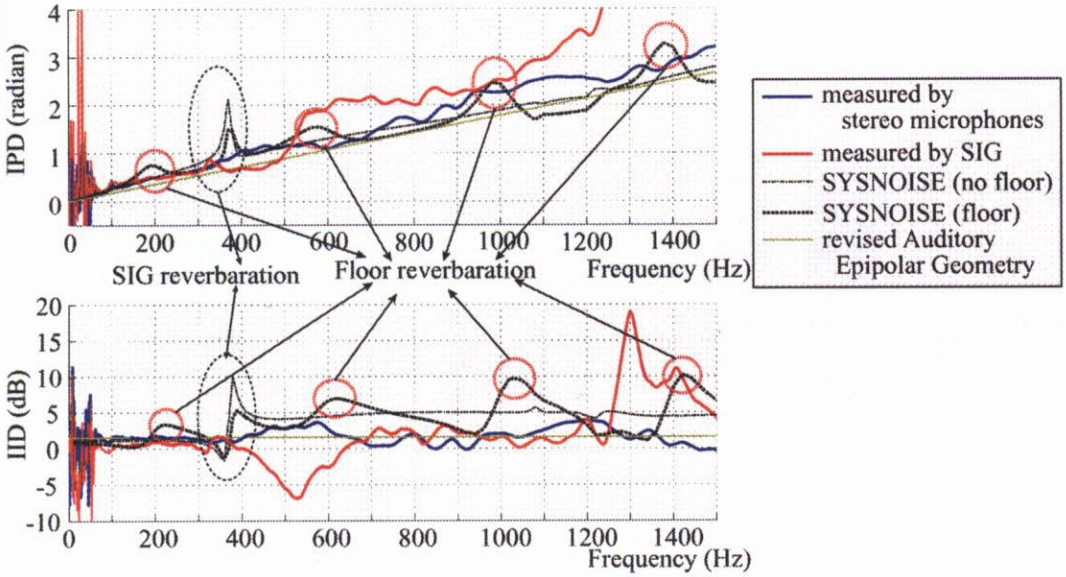


図 4.12: IPD と IID の測定値とシミュレーションによる値の対応 (30°)

部屋の音響環境については, 部屋ごとに, また, 同じ部屋でも位置によって異なるため, 正確な IPD と IID を予測することは難しい. また, 境界要素法などのシミュレーションでは, 時間がかかる上に十分な精度を望むことができない. この問題に対する最も有効な解決は, 複数の聴覚的手がかりを統合してロバスト性を高めることであろう. そのような音源定位については 5 章に述べる.

本章では, ノイズキャンセルの有効性を音源定位によって示すため, あまり反響の影響が強くない 500Hz, 600Hz の純音の定位を行い, ノイズキャンセル, アクティブな動作, 聴覚エピポーラ幾何, それぞれの有効性を示す.

4.3 *SIG* 動作時の音源定位実験

アクティブオーディションシステムにおいて, 視覚, 聴覚, ポテンショメータから得られる *SIG* の方向情報が欠如している場合や, *SIG* が動作中の場合であっても, 方向情報を補完して音源定位を行い, 曖昧性を解決できることを示す. また, アクティブオーディションの有効性を示すため, 動作によって知覚が向上できることも併せて示す.

4.3.1 実験のシナリオ

音源には, 500 Hz の純音を発生するスピーカ A (音源 A), 600 Hz の純音を発生するスピーカ B (音源 B) の 2 つのスピーカ (B&W Nautilus 805) を使用する. スピーカが設

置されている部屋は, 交通量の多い道路に面しており, 日常的なノイズに絶えず晒されている 一般的な住居用アパートの一室 (約 10 平方メートル) である.

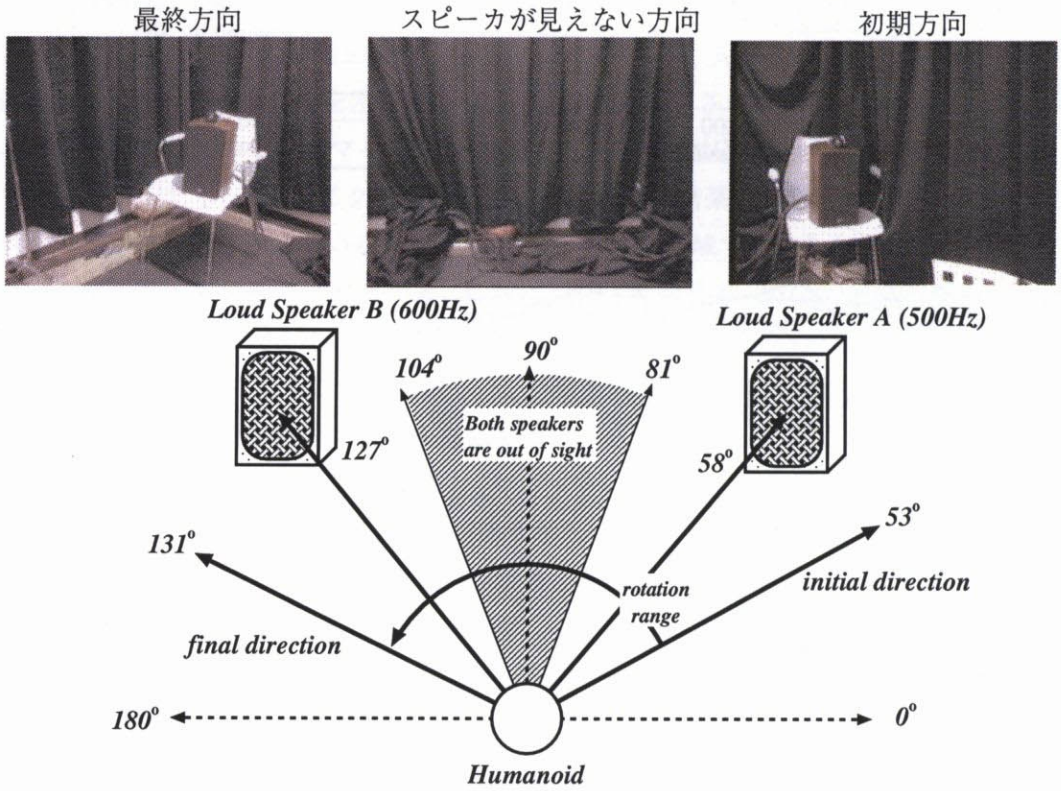


図 4.13: 動作時の純音定位実験

図 4.13 に示すように, 初期状態では, *SIG* は 53° の方向を向いており, 音源 *A*, *B* はそれぞれ *SIG* の 5°, 69° 左に置かれている. また, *SIG* と各音源との距離は約 210 cm である. *SIG* の視野は水平方向で 45° であるため, 初期状態では, 69° 左に置かれている *B* は視野外となり見ることはできない. 実験のシナリオは以下の通りである.

1. 最初に *A* が *SIG* の 5° 左で音を発する.
2. *SIG* は, 抽出した画像と音の方向が同じであることから, 両者を同一音源に由来するものとして, アソシエーションする.
3. *A* が音を発した後, しばらくして *B* が音を発する. この時, *B* は *SIG* の視野外にあり, 音源方向は聴覚情報によってのみ得られるため, *SIG* は *B* に関するアソシエーションを行なわない.
4. *SIG* は, 視野外の音源 *B* の方向を向く.
5. 回転中, *SIG* の向きが 81° から 104° の間は, 両方のスピーカが視野外となるため,

SIG は聴覚情報のみで音源方向を推定する.

6. *SIG* が 104° まで回転すると, *B* が視野内に入るため, *B* を新たな物体として認識し, *B* に関して音と画像のアソシエーションを行なう.

評価には 4 種類のベンチマークを用いた. これらは, *SIG* の回転速度が速い場合 (秒速 68.8°), 遅い場合 (秒速 14.9°), 内部ノイズと同程度の弱い音響信号を入力した場合, および, 内部ノイズを無視できる程度の強い (+20dB) 音響信号を入力した場合の組み合わせである. なお, 本実験では, 視聴覚の定位は実時間では行っていない. *SIG* の動作中に, 予め, 音響データと画像データを同期を取って収録し, 収録データに対してノイズキャンセルおよび視覚, 聴覚による定位の処理を行った. なお, 視覚の定位における *SIG* の左右画像のマッチングには, Lourens らのコーナ抽出アルゴリズムを利用している [71].

4.3.2 定位実験結果

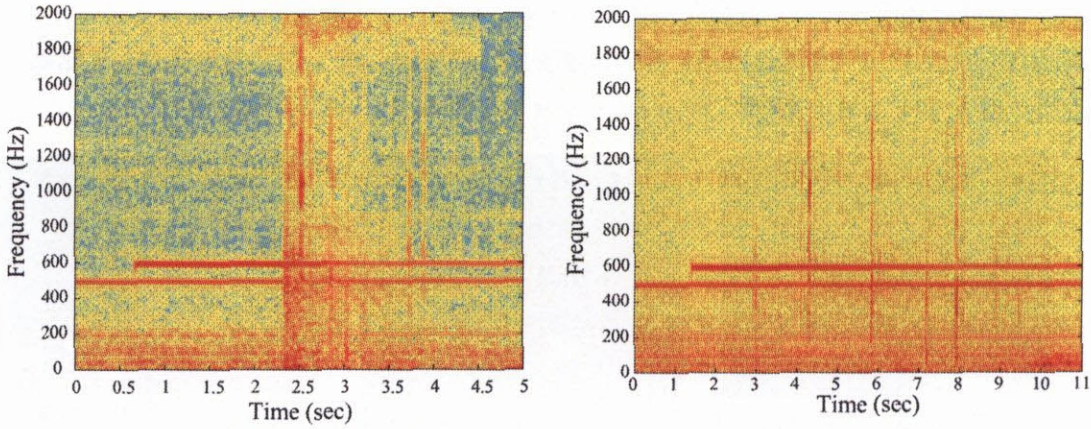
SIG の回転速度が速い場合のスペクトログラムを図 4.14a), 遅い場合を図 4.14b) に示す. これらの場合にモータノイズをキャンセルしないで音源定位を行った結果を, 図 4.15 に, キャンセルして音源定位を行った結果を, 図 4.16 に示す. 図 4.18 は, 入力音響信号が強い場合にノイズキャンセルを行い, 定位した結果である. また, 表 4.1 は, 各ノイズキャンセル手法を用いたときの定位誤差, 誤差の分散, 誤差の最大値・最小値を示している.

図 4.15 – 4.18 では, X 軸が時間, Y 軸が音源方向であるが, 音源方向は *SIG* 座標系でプロットされており, 図中の 0° は *SIG* のその時刻における正面方向, 負の値は右方向, 正の値は左方向を示す. また, 図 4.16, 4.18 における点線は視覚による定位結果を示す.

実験の観測結果は以下の通りである. 図 4.15a) では, 動作時の 2.5–3.5 秒付近で定位結果に深刻な影響が見られる. 図 4.15b) でも, 4.3, 6.0, 7.2, 8.0 秒付近で同様の現象が確認できる. 両者とも, 動作開始前の最初の 2 秒間の定位結果からは, 入力音のパワーの強弱に関係なく音源定位は安定している. 図 4.16 では, 4.3, 6.0 秒付近のバーストノイズがキャンセルできていることがわかる. 図 4.17 では, 目立ったバーストノイズがすべてキャンセルされ, 動作時でも正確な音源定位が達成されている. 図 4.18 では, 信号音が強いため, 動作時でも正確な音源定位が達成されている.

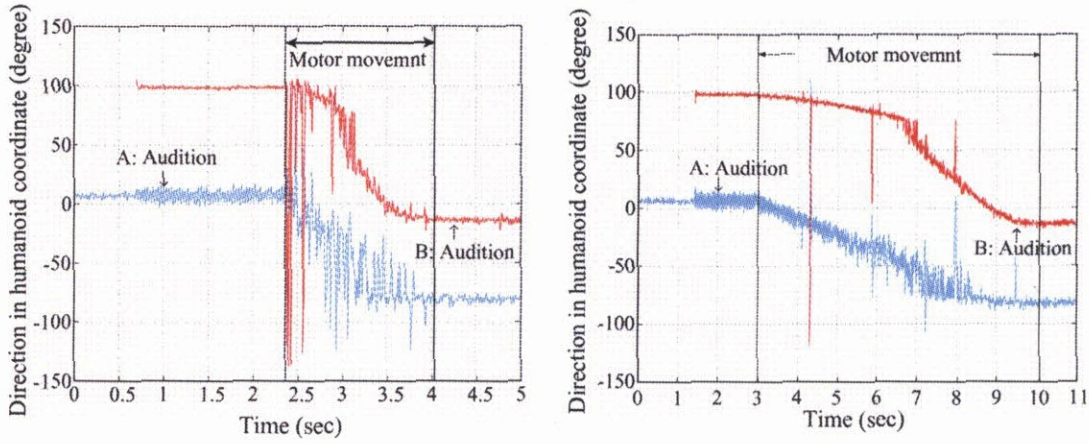
これらの定位結果より, まず, ノイズが少ない静止時には, 聴覚エピポーラ幾何は効果的であるが, 動作時には, 定位を不可能にするようなパワーの大きいモータノイズが発生するため, 聴覚エピポーラ幾何は有効に働いていないことがわかる. 実際にスペクトログラムでは, 動作時に広帯域のモータノイズが観測されている.

提案したノイズキャンセルは, 動作が遅い場合には, バーストノイズをきれいに取り除いている. 動作が速い場合でも聴覚用エピポーラ幾何で音源定位を行うことができる程度



a) 回転が速い場合　　b) 回転が遅い場合

図 4.14: 入力信号のスペクトログラム



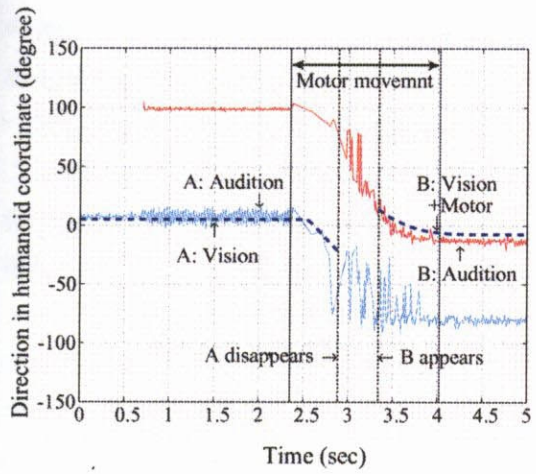
a) 回転が速い場合　　b) 回転が遅い場合

図 4.15: ノイズキャンセルを行わない場合の定位

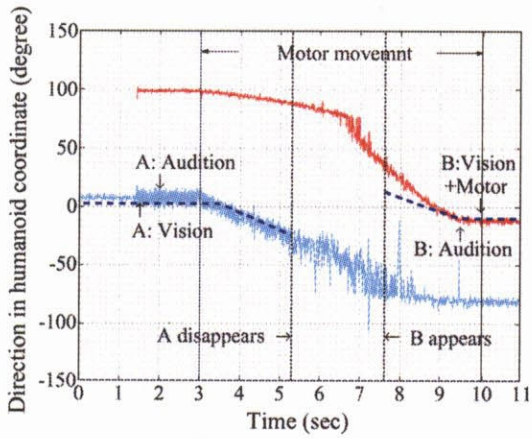
に, ノイズの影響を軽減している. また, 入力信号が強い場合は, より正確な定位を達成している.

視覚による定位結果は, 聴覚による定位の正確さを指示している. 音源 B が視野内に入った際に B に関するアソシエーションが行われることにより, 視覚情報とモータ情報からの補完が可能になり, 定位精度の向上を図ることが可能になることがわかる.

次に, 音源がロボットの正面方向に近くなるに従い, 定位誤差が小さくなることがわかる. 実際, 動作前の状態で正面方向に近い音源 A の定位誤差はほぼ 0° であるのに対し, 遠い音源 B の定位誤差は約 30° である. 音源 B の方向を向くに従い, 定位誤差は小さくなり, 動作終了時には, ほぼ 0° となる. これは, 音源方向を向くという動作が定位精度を向上させることを示しており, アクティブオーディションの有効性を指示している. 一般に,

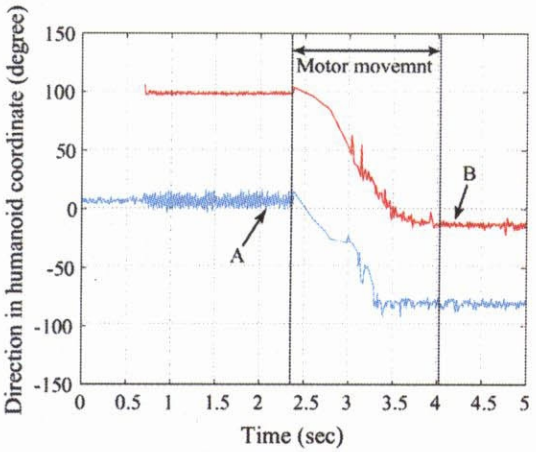


a) 回転が速い場合

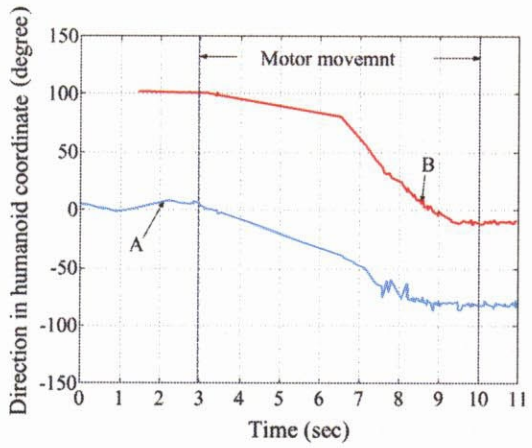


b) 回転が遅い場合

図 4.16: 簡単なバーストノイズキャンセルを用いた定位

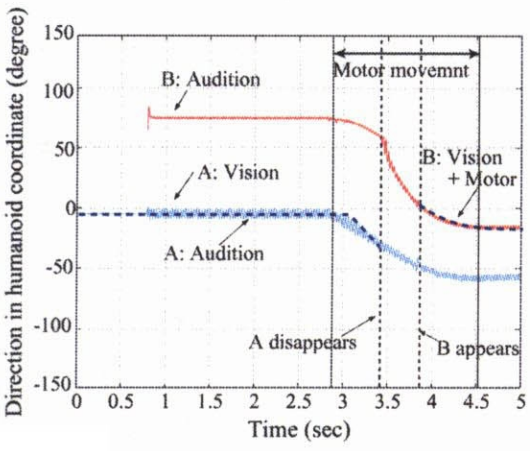


a) 回転が速い場合

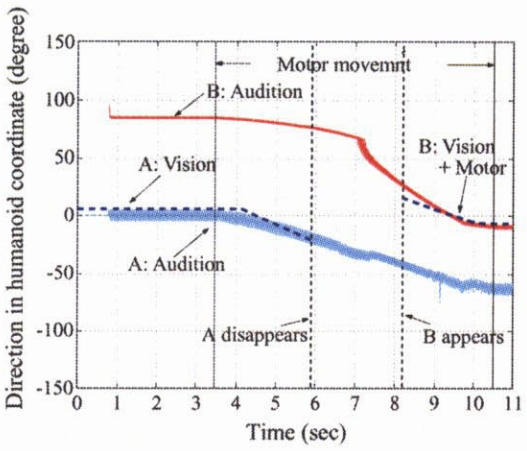


b) 回転が遅い場合

図 4.17: 外装の音響効果の測定結果を利用したバーストノイズキャンセルを用いた定位



a) 回転が速い場合



b) 回転が遅い場合

図 4.18: パワーの強い音響信号を用いた場合の定位 (50dB)

パワーの小さい音や未知環境での定位など, 利用できる情報や信頼できる情報が少ない場合には, 定位誤差が大きくなる. このような場合には, アクティブオーディションの効果はより大きくなる. 特に純音のように他の倍音成分からの情報を得られない場合は, 部屋の音響環境の影響を受けやすいので, アクティブな動作は定位向上に不可欠といえる. 純音定位は一見簡単に思われるが, Hornbostel によれば, 一般に人では純音定位は調波構造を伴う音よりも定位が難しいという知見も得られている.

表 4.1: 各ノイズキャンセル手法ごとの定位誤差・誤差分散・誤差の最大最小値

	without cancellation	with simple cancellation	with cancellation by cover acoustics
平均誤差	10.55°	10.38°	18.76°
誤差の分散	229.64	166.27	117.70
誤差の最小値	-52.44°	-26.98°	-17.18°
誤差の最大値	258.92°	35.43°	30.31°

表 4.1 からは, 各ノイズキャンセル手法の効果がよくわかる. 定位誤差は, 10° から, 20° 程度である. 音響効果を利用した手法で, 定位誤差が大きくなっているが, これは, 定位精度が悪化したわけではない. 全体的に, 定位結果は実際の音源方向より大きくなっており, 本実験では, バーストノイズは, 定位結果を小さく見せるように定位結果に影響したため, バーストノイズをキャンセルしない方が誤差平均が小さくなっているのである. つまり, 音響効果を利用した手法の方が, エピポーラ幾何による定位能力をよりよく表しているといえる. 誤差の分散においては, ノイズキャンセルを行うことによって定位が安定してきていることがわかる. 特に, 音響効果を利用した方法では, ノイズキャンセルを行わない場合と比べ, 分散が半分程度になっている. 誤差の最大値・最小値は, バーストノイズのキャンセルの効果がよくわかる. ノイズキャンセルを行わない場合は, 定位がバーストノイズによって, 250° 以上も悪くなることがあったことがわかる. これに対し, 簡単なノイズキャンセルを行う方法では, 35° 程度に, さらに, 音響効果を利用した方法では, 30° 程度まで, 最大誤差が減少していることがわかる.

4.4 まとめ

アクティブオーディションによるアクティブな動作は, 知覚をロバストにし, インタラクションを豊かにする有効な手段である反面, モータノイズが聴覚処理を困難にするという側面を持っている.

この問題に対して, 本章で説明したノイズキャンセルを伴う音源定位システムは, 外装

の音響効果と聴覚エピポーラ幾何を利用することにより, 動作時のノイズをキャンセルし, ある程度の反響がある部屋での純音定位を可能とした.

SIG の外装は, アクティブな動作を有効利用するためのノイズキャンセルに有効であることを示した. 外装の利用については, ロボットが音響的な身体性を獲得できるように外装の内外にマイクを設置すること, および, 外装内部での共鳴に対応するために外装の音響測定を行うことが重要であることを示した. そして, 動作時に特に問題となるバーストノイズを検出し, ノイズの影響が大きい部分を以後の処理で用いないようにするフィルタを構築した. これにより, ANC, ICA やビームフォーミングといった従来の手法では難しかった, 位相情報を保ったノイズキャンセルを実現し, "stop-perceive-act" という制約を緩和できることを示した.

また, 聴覚エピポーラ幾何は, 計算的手法で IPD を導出することが可能であり, 実際に測定結果によく一致すること, 実環境での連続的な音源定位を可能とすることを示した. この手法は, 一般に両耳聴の音源定位に利用されている HRTF を利用しないため, HRTF の抱えている実環境・未知環境への適用が難しいという問題を避られることを示した.

さらに, 動作時に同時に 2 つの純音を定位するという実験を通じて, 音源方向を向くというアクティブな動作は, 聴覚処理による音源定位の精度を向上させること, および視覚による定位情報を利用すれば, より正確な定位が期待できることを示し, アクティブオーディションの有効性を示すことができた.

実環境では, 通りの自動車や実験室の外からの人の声, ロボット内部の音など複数の音源が存在しており, これらは, 一般には純音ではない. そのため, 音声など調波構造を有する音を考慮する必要がある. また, 純音の場合, 位相差情報が部屋の影響を受けやすく, ロバスト性という面では問題が残る. このような場合の音源定位については 5 章で述べる.