音声情報処理特論2　2005年11月22日

ロボット聴覚
―混合音の定位・分離―

奥 乃　博

京都大学　大学院情報学研究科
知能情報学専攻
知能メディア講座　音声メディア分野
http://winnie.kuis.kyoto-u.ac.jp/~okuno/
okuno@i.kyoto-u.ac.jp, okuno@nue.org

1

目次

1. 混合音からの3つの機能
   - 音源定位 (Sound source localization)
   - 音源分離 (Sound source separation)
   - 分離音の認識 (sound recognition)
2. 組み込みシステムの聴覚機能
3. 頭部音響伝達関数
4. 頭部音響伝達関数の近似
   - 聴覚エピポーラ幾何
   - 散乱理論
5. モータ音のキャンセル

2

混合音処理への研究アプローチ

1. 信号処理からのモデル化
   - マイクロフォンアレイ
   - ビームフォーマ（遅延加算型、死角生成型, 適応）
   - 独立成分解析（ICA, independent component analysis）
2. 人の聴覚機能からのモデル化
   - Computational Auditory Scene Analysis（CASA）
     「音環境理解」
   - Missing Feature Theory
     音素修復（auditory induction）
   - Sub-band analysis

3

混合音処理での注意

1. 研究室環境とは異なる実環境
   - interfering sounds　が非定常的、残響
   - 複数の話者の同時発話
   - 背景雑音としてTVやラジオからの音声・音楽
   - 音源移動（移動話者、システムが移動）
2. 理論上の優越は必ずしも実世界での優越
   - ICAかGSS（Geometrical Source Separation）か？
   - 正確な測定がいつも役立つとは限らない（過剰適用）
3. 総合的な能力が重要
4. ロボットの耳に適用すると現実の問題が見えてくる
   - 実時間処理
   - 体の影響 (空間伝達関数に加えて身体伝達関数が発生）
   - モータ音の影響（バーストノイズであり、毎回異なる現象）
5. 試行錯誤によるノウハウの確立を通じたより汎用な技術の確立

4

既存のロボットのマイクロフォンは

QRIO　SDR-4XII

- 7本のマイクロフォン
  内1本は内部雑音除去用
- 音源定位は行う.
- 音源分離は行わず.

ASIMO

- 2本、音源定位のみ.

HRP-2

- 耳はなかった. 8/16本.

11

ロボット聴覚

- ロボット自身の耳で聞く研究は少ない
- 従来の研究
  - マイクは、人間の口元に装着。
  - 単一音源からの入力を想定。
  - モータノイズが無視できるくらい対象音は大きい。
- “Stop-perceive-act” 戦略による処理の簡単化
- ロボット聴覚の機能は、組み込みシステムに音声認識を実現するための重要な一歩
- 情報家電の音声入力・音声コマンダー

12

混合音処理への研究アプローチ

1. 信号処理からのモデル化
   - マイクロフォンアレイ
   - ビームフォーマ（遅延加算型、死角生成型）
   - 独立成分解析（ICA, independent component analysis）
2. 人の聴覚機能からのモデル化
   - Computational Auditory Scene Analysis（CASA）「音環境理解」
   - Missing Feature Theory 音素修復（auditory induction）
   - Sub-band analysis

13

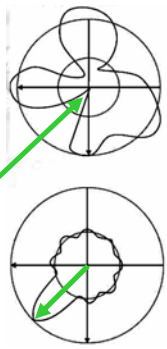
音場の区別

1. Near Field
   - 音源から球面波が届く
   - マイクロフォンの間隔で規定
   - 人の耳だと30cmまで
2. Far Field
   - 音源から平面波として届く
   - 数m程度
3. Reverberation/Reflect Field
   - ambient noise
   - 数m程度以上（アンテナの研究）

15

Beamformer (マイクロフォンアレイ)

- ナルフォーミング(null（beam）forming)
  - 原理「N+1本のマイクロフォンでN個の音響的死角が構成できる」
  - 特定の方向の音に対して逆相で波を重ね合わせて、notch作成
  - 鋭い指向性の形成が可能
- ビームフォーミング (beamforming)
  - 特定の方向の指向性（focus）を強調. 緩やかな指向性.
  - 遅延型加算 (delay-and-sum)
  - 適応型（adaptive）



17

独立成分解析（ICA）

- Independent Component Analysis
- 原理「音源が情報論的に相互独立ならば、N個の音源はN本のマイクロフォンで分離できる」
- 時間領域ICA vs 周波数領域ICA
- Blind Source Separation
- 音源の性質について最小限の仮定
  - 出力の相互情報量を最小化
  - 非ガウス性の最大化（by　中心極限定理）
  - 尤度の最大化（by　最尤推定）
- Beamformer は方向情報が所与
- マイクロフォンに関する幾何学的情報（位置の測定）不用

18

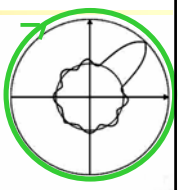
混合音処理への研究アプローチ

1. 信号処理からのモデル化
   - マイクロフォンアレイ
   - ビームフォーマ（遅延加算型、死角生成型）
   - 独立成分解析（ICA, independent component analysis）
   - 音源定位（sound source localization）
     - Steered Beamformer
     - MUSIC法

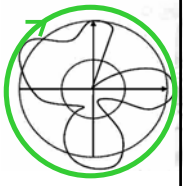
19

Steered Beamformer

- 2D-Steered Beamformer
- Steered Delay-and-Sum Beamformer
- focus (beam) をscan



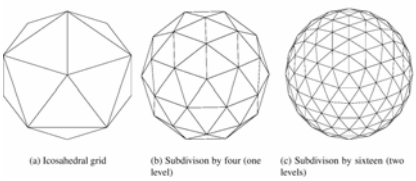
- Steered Null Beamformer
- (beam) をscan



20

Steered Beamformer

- 3D-Steered Beamformer



- 空間分割を粗いものから細かいものへ

21

音源分離（sound source separation）

1. ビームフォーマ
   - Delay-and-Sum（遅延加算）
   - Null former（死角生成）
   - Adaptive （適応）
2. Geometric Source Separation
3. Multi-channel Post-filter
4. 独立成分解析（ICA, independent component analysis）

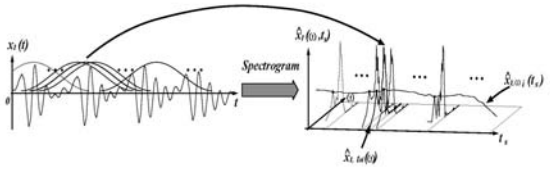
22

Blind Source Separation

1. Sound signal vector $s(t)$ of n components
   $s(t) = (s_1(t), \cdots, s_n(t))^T, \quad t = 0, 1, 2, \ldots$
2. Observed signal vector $x(t)$ by n microphones
   $x(t) = (x_1(t), \cdots, x_n(t))^T, \quad t = 0, 1, 2, \ldots$
3. Sources are mutually independent.
4. $x(t)$ is given by a linear operator A
   $x(t) = As(t) = \left(\sum_k a_{ik} * s_k(t)\right) = \left(\sum_k \sum_{\tau=0}^{\tau_{max}} a_{ik}(\tau) * s_k(t-\tau)\right)$
5. From $x(t)$, find a linear operator B s.t. $y(t) = Bx(t)$, mutually independent $y(t)$, without knowing operator A and the probability distribution of $s(t)$.

On-Line Algorithm Proposed by Murata and Ikeda

1. Human voice is stationary for a period $<$ 30∼40msec
2. Apply Windowed Fourier Transformation with Hamming window of 128 points to obtain spectrogram



3. Apply on-line Independent Component Analysis to each non-symmetric 65 points of frequency components

$$\hat{x}(\omega, t_s) = \widehat{A}(\omega)\hat{s}(\omega, t_s),$$
$$\tilde{u}(\omega, t_s) = \tilde{x}(\omega, t_s) - B(\omega, t_s)\tilde{u}(\omega, t_s)$$
$$\hat{u}(\omega, t_s) = (B(\omega, t_s) + I)^{-1}\tilde{x}(\omega, t_s)$$

4. Learning rule:

$$B(\omega, t_s+\Delta T) = B(\omega, t_s) - \eta(B(\omega, t_s)+I)(diag(\phi(z)z^*) - \phi(z)z^*), \quad z = \tilde{u}(\omega, t_s)$$
$$\hat{v}_\omega(t_s; i) = (B(\omega, t_s) + I)(0, \ldots, \hat{u}_i(\omega, t_s), \ldots, 0)^T.$$

5. Reconstruct separated spectrogram based on the common temporal structure of original source signals. [Assumption] Common AM for the same sound source. Defining an envelope making operator by

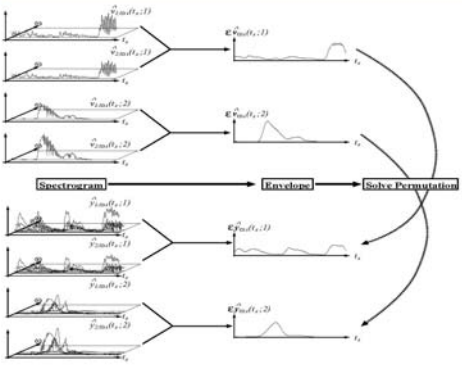
$$\mathcal{E}\hat{v}_\omega(t_s; i) = \frac{1}{M}\sum_{t'_s=t_s-M}^{t_s+M} |\hat{v}_w(t'_s; i)|,$$

Solve permutation based on the correlation of envelopes between

$\mathcal{E}\tilde{v}_\omega(t_s; \sigma_w(i))$, and

$\mathcal{E}\tilde{y}_\omega(t_s; i) = \mathcal{E}\sum_{\omega'}\tilde{v}_{\omega'}(t_s; \sigma_{\omega'}(i))$

Construct Separated Spectrogram



14

ICA(独立成分分析)とは

武田　龍君（奥乃研B4）

- 観測信号のみから信号を分離する
- 源信号の非ガウス性と独立性に着目して分離する
  - 独立性＝結合密度がそれぞれの周辺分布の積に因数分解可能
    - Pxy(X,Y) = Px(X) Py(Y)
  - 音声信号は多くの場合非ガウス性（優ガウス分布）を持ち、ICAが適応できる
- 数学的モデルには瞬時混合モデルや畳み込み混合モデルなどがある

29

ICAの定式化

- 瞬時混合モデル

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

x: 観測信号ベクトル
A: 混合行列
s: 元信号ベクトル

- 畳み込み混合モデル

$$\mathbf{x} = \sum_n \mathbf{A}(t)\mathbf{s}(t-n) \quad \Leftrightarrow \quad \mathbf{x}(\omega) = \mathbf{A}(\omega)\mathbf{s}(\omega)$$

時間領域　　周波数領域

  - 畳み込みモデルの場合、周波数領域に変換すると各周波数ビンでの瞬時混合モデルになる
  - ただし、推定においてAとsの両方を対象とするため、sのパワー、及び順序を決めることはできない

30

ICAによる分離

- 前処理として白色化を行う
  - 白色＝平均値が0・共分散行列が単位行列となる確率ベクトル
  - 次の解探索を簡単にする
- 独立性を評価する関数を定めて勾配法により数値的に解く
  - 独立性の指標にはKL情報量・尖度などが用いられる

31

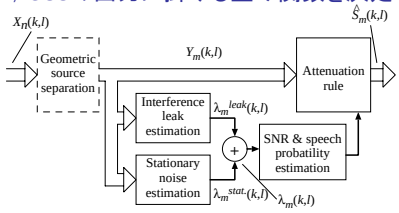
分離結果

- 以下の2つの混合信号をICAで分離
  - Sine波とTriangle波の瞬時混合信号
  - 男性と女性の瞬時混合信号
- 次の段階に分けて、それぞれの分布についてグラフ化
  - 元の信号2つ
  - 混合後の信号2つ
  - 白色化後の信号2つ
  - ICA適応中（3種）の信号2つ
  - ICAで分離後の信号2つ
- 以下の点に注意
  - ICAではスケーリングと分離出力に任意性がある

32

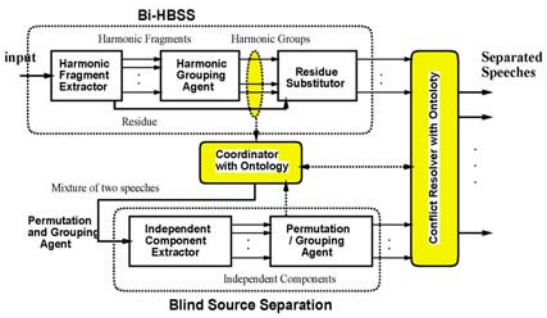
Multi-channel post-filter [Valin,Cohen]

- 幾何学的音源分離（Geometric Source Separation, GSS）によって分離された音を強調する手法
  - 定常性の雑音推定と非定常性の雑音推定を行い，GSSの出力に掛ける重み関数を決定



33

Integration of Bi-HBSS and Blind Source Segregation



混合音処理への研究アプローチ

1. 信号処理からのモデル化
   - マイクロフォンアレイ
   - ビームフォーマ（遅延加算型、死角生成型）
   - 独立成分解析（ICA, independent component analysis）
2. 人の聴覚機能からのモデル化
   - Computational Auditory Scene Analysis（CASA）
     「音環境理解」
   - Missing Feature Theory
     音素修復（auditory induction）
   - Sub-band analysis

36

アクティブオーディションの課題

- active audition: active vision と同じく、マイクロフォンの様々なパラメータを変更・適用して聞く。
- active sensing: 物体を認識するのに、物体を叩いて音を聞く。容器に中身が入っているかを振って調べる。

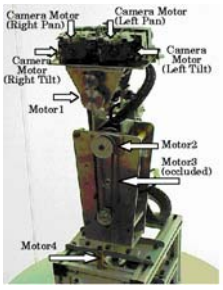
1. 音声に限らない一般的な音の理解
2. 定常雑音に限らない混合音の理解
3. センサ情報統合
   - 信号レベル
   - シンボリックレベル
4. ノイズキャンセル
   - 動作中のモータノイズは避けられない。
   - 一般にマイクはモータの近くにあるため、ノイズは比較的大きな音として収音されてしまう。

37

ヒューマノイド　*SIG*

ソーシャルイン
ターラクション用

- 4 DOFs
- 2 組のマイク
- 1 組のカメラ
- 機能的で美しい外装（デザインとしての研究テーマ）
- AIやセンサフュージョンの実世界への応用を目的



SIG

38

ロボットデザイン

- デザインされたロボットたち
  - J-Star99, SIG, Pino, Posy, SIG2
  - 人間との共生をテーマ
- デザイナー：　松井龍哉
  - Flower Robotics Inc. （Oct.）

39

音源定位に関する特徴量

- 両耳間時間差（Interaural Time Difference）
- 両耳間位相差（Interaural Phase Difference）

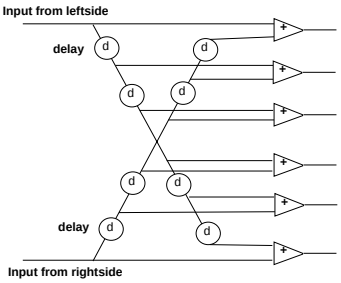
◆両耳間レベル差（Interaural Level Differ.）
◆両耳間振幅差（Interaural Amplitude Differ.）
◆両耳間強度差（Interaural Intensity Differ.）

- これらの特徴と方向情報との対応は？
  ITD, IPD &　ILD, IAD, IID　⇔
  Azimuth & elevation

41

人間の音源定位モデル

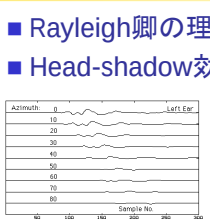
Jeffressモデル
時間差による
モデル化

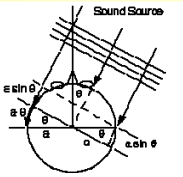


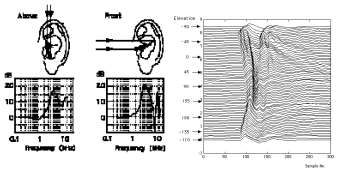
42

頭部音響伝達関数(HRTF)

- Rayleigh卿の理論
- Head-shadow効果







43

外装の音響測定

- 無響室で測定(日東紡音響エンジニアリング)

•四方の壁、天井、床 → 吸音材(グラスウール)
•突起状の形 → 吸音しやすい形状。

125Hz以上の周波数域では、反響が無い部屋

Anechoic room

44

無響室

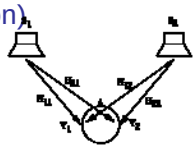
- 272個のマイクロフォン(15度間隔） 直径4.6m、6.7m角
- 防音用耳カバの音源定位への影響
- 残響時間（60dB減衰時間）0.01秒程度

Auditory Localization Facility at Wright-Patterson AFB

45

HRTFの近似

1. 水平方向の近似
   - 頭部の形状
   - 上半身の回折 (diffraction)
   - 肩の反射 (reflection)
2. 垂直方向の近似
   - 耳介（pinnae）の反射
3. クロストークキャンセルステレオ
   - Sweet spot



$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \qquad \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

46

聴覚エピポーラ幾何

HRTF(Head Related Transfer Function、頭部伝達関数)

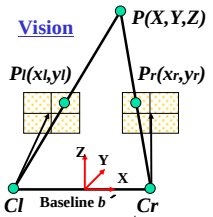
- バイノーラル(両耳聴)の研究でよく使われる
- 環境の変化に敏感(通常は無響室で測定)
- 測定に時間がかかる
- 離散的な関数である

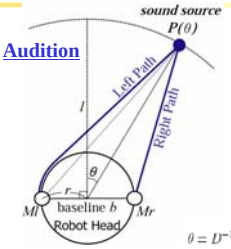
聴覚エピポーラ幾何

- ステレオビジョンで使われるエピポーラ幾何の聴覚への拡張
- 現状では水平方向の音源定位のみ
- 両耳間の位相差から、計算的に方向情報を算出
  ⇒ 測定不要、連続関数
- ステレオビジョンのエピポーラ幾何と情報統合が容易

47

エピポーラ幾何(視覚、聴覚)





$$X = b'(x_l + x_r) / 2(x_l - x_r)$$
$$Y = b'(y_l + y_r) / 2(x_l - x_r)$$
$$Z = b' f / (x_l - x_r)$$

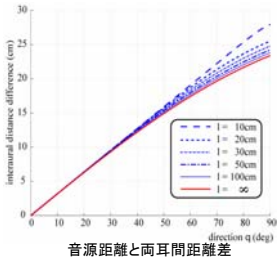
$$\theta = D^{-1}\left(\frac{v}{2\pi f}\Delta\varphi\right)$$

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

音源距離との関係

- 50cm以上離れていれば、距離を無限と仮定することが可能
- 近接学(Proxemics) からも、インタラクションで50cm以上を仮定することは妥当[Hall 66]

$$D(\theta) = \lim_{l \to \infty} D(\theta, l) = r(\theta + \sin\theta)$$



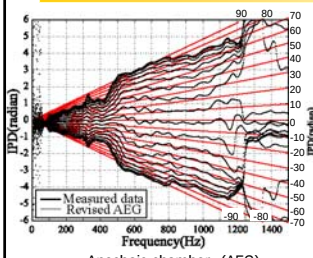
音源距離と両耳間距離差

Interpersonal distance in proxemics

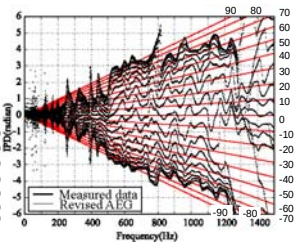
50cm	1m	2m	
intimate	personal	social	public

50

音響特性 (IPD)



Anechoic chamber (AEG)



Non-anechoic normal room (AEG)

The AEG is efficient for sound source localization in an anechoic room.
In a non-anechoic room, it is not enough for robust localization.

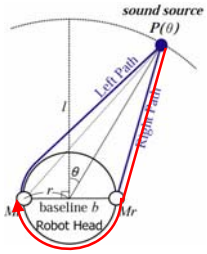
51

頭部の音響モデル

- **頭部伝達関数 (HRTF)**
  - 両耳間位相差(IPD)、両耳間強度差(IID)を取得可能
  - 計測に時間がかかる・離散関数

- **聴覚エピポーラ幾何**
  - 水平方向の定位
  - IPD を計算的に推定可能
  - 高周波、音の回り込みが未考慮

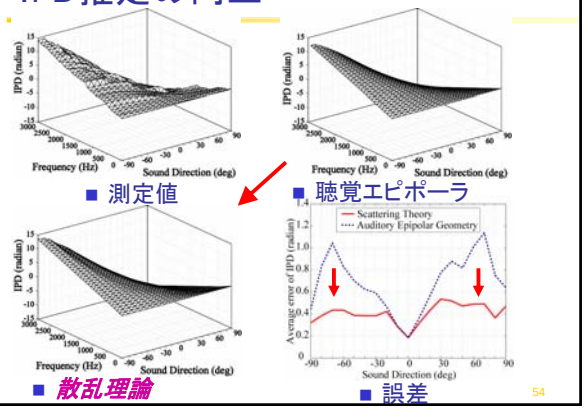
- **散乱理論**
  - 水平方向の定位
  - IPD と IID の計算的な推定



IPD :
$$\Delta\varphi = \frac{2\pi f}{v} \times r\,(\theta + \sin\theta)$$

52

IPD推定の向上



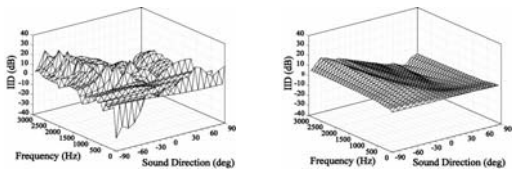
■ 測定値　■ 聴覚エピポーラ

■ *散乱理論*　■ 誤差

54

IID 推定の向上

- 聴覚エピポーラ幾何
  - 大まかな3方向の推定: 正面、右、左
- 散乱理論
  - 方向ごとの計算的な推定



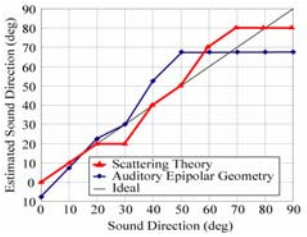
■ 測定値　■ 散乱理論

55

実験1: 音源定位

- 100Hz の調波構造音 (100Hz – 3kHz) の定位

音源定位結果

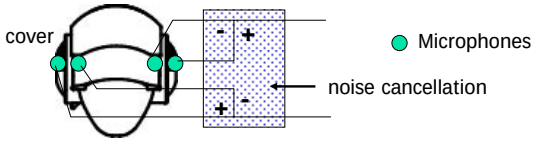


- 50度までは、同程度の精度
- 50度以上になると散乱理論の精度が高い。

56

外装によるノイズキャンセル

- 外装によってロボット内外を区別
- 1組の内部マイクをノイズ集音用に外装の内部に配置
- 1組の外部マイクを外部の音の集音用に外装の外部に配置
- 内部と外部のマイクの差を利用したノイズキャンセル

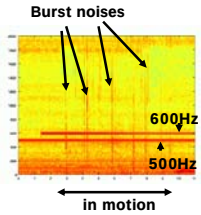


57

SIG ノイズの特徴

バーストノイズ

- 動作中にバーストノイズが発生.
- バーストノイズが特に悪影響を与えている.
- 少なくともバーストノイズのキャンセルは必須.

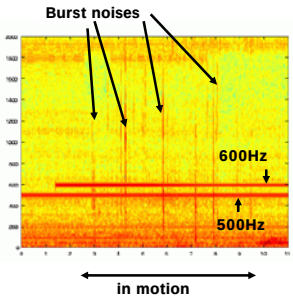


共鳴

- SIG の頭の直径は約 18 cm => 500Hzで λ / 4 に相当
- 外装は、500Hzを中心周波数とした共鳴現象を持っているのでは？

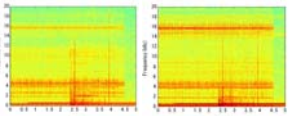
58

スペクトログラム



縦軸：周波数
横軸：時間
色　：パワー(赤強い、青弱い)

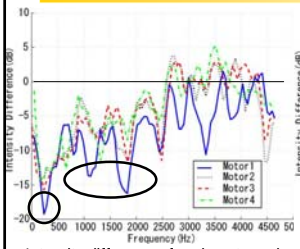
20msの窓長のFFTを7.5msずつ時間的にずらして作成したもの



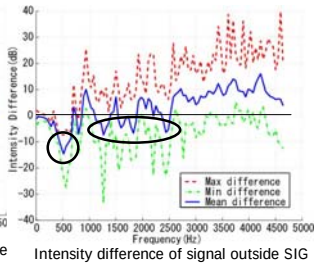
外マイク　内マイク

59

外装の音響効果



Intensity difference of each motor noise



Intensity difference of signal outside SIG

- 500Hz 近辺での共鳴
- それ以外の周波数帯でも同様の現象あり
- 共鳴を考慮せずにノイズキャンセルをすることは困難

61

外装の音響効果を利用したノイズキャンセル

・Heuristics によるバーストノイズキャンセルフィルタ
・音響測定結果をテンプレートとしてバーストノイズ判定に利用

Conditions:
- 内外のマイクの強度差がテンプレートのモータノイズの強度差と近い
- スペクトルの強度とパターンがテンプレートのモータノイズ周波数応答に近い.
- モータが動いている.

上記の3条件を満たした場合にバーストノイズと判定し、キャンセルする。

62

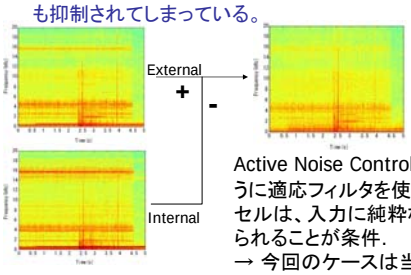
他の方法との比較

- FIR 適応フィルタによるノイズキャンセル（アクティブノイズコントロールなどでよく使われる）
- 外装の音響効果を考慮しない簡単なヒューリスティックによるバーストノイズを対象としたノイズキャンセル法

63

FIR 適応フィルタ

- 100次の FIR(Finite Impulse Response) フィルタ
- バーストノイズが残ってしまっている.
- 外部からの500Hz、600Hz のキャンセルされて欲しくない音も抑制されてしまっている。

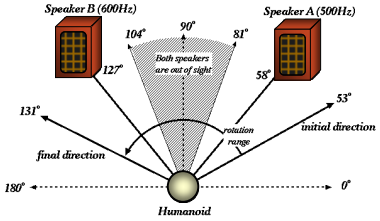


Active Noise Controlに代表されるように適応フィルタを使ったノイズキャンセルは、入力に純粋なノイズだけが得られることが条件.
→ 今回のケースは当てはまらない.

64

実験

通常の部屋で、動作中に、2音源からの混合音の定位

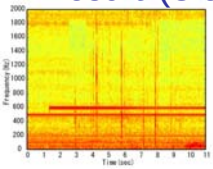


- スピーカ音源
  A(500Hz)、B(600Hz)
- Aの方向からBの方向へターン
- SIG はターンの間、両方の音を聞くことができる
- 斜線内では、両方のスピーカとも見えない。

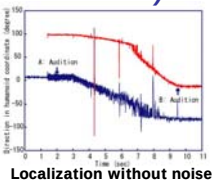
1. 未知の環境での聴覚エピポーラ幾何による定位
2. 外装の音響効果を利用したノイズキャンセル
3. 聴覚と視覚の統合

66

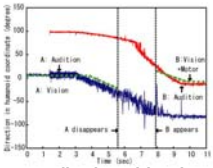
Result (Slow Rotation)



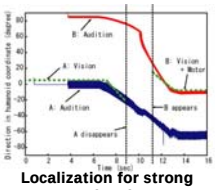
Spectrogram



Localization without noise cancellation



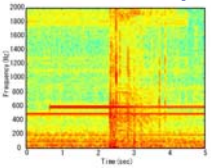
Localization with noise cancellation



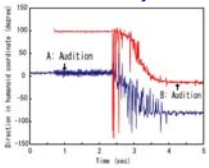
Localization for strong signal

67

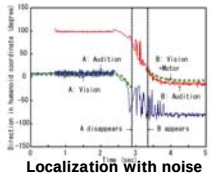
Result (Fast Rotation)



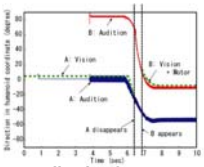
Spectrogram



Localization without noise cancellation



Localization with noise cancellation



Localization for strong signal

68

本日の予定終了

- レポートは3題のうち、2題選択回答

1. 柏野さんの講演（情報学展望）のまとめと感想（5ページ以上）
2. 音源定位・音源分離・分離音認識について2つ異常の技法を詳細に報告（5ページ以上）
3. 未定。

- レポートの締切は12月20日（予定）
- 提出先は10号館レポートボックス

70