音響シミュレーションを用いた発話方向推定

○鈴木淑正　荒井隆行 (上智大学)　鶴秀生 (日東紡音響エンジニアリング)
中島弘史　中臺 一博 (HRI-JP)

Sound Source Orientation Estimation using Wave Acoustic Numerical Simulation

*Toshimasa SUZUKI, Takayuki ARAI (Sophia University), Hideo Tsuru (Nittobo Acoustic Engineering Co., Ltd.), Hirofumi NAKAJIMA, Kazuhiro NAKADAI(Honda Research Institute Japan Co., Ltd.)

Abstract— This paper addresses speaker orientation estimation using an acoustical simulation. In human-robot (or human-computer) interaction in multi-party situation, a robot should understand the orientation of each speaker, because the robot needs to understand which speaker speaks to the robot. We have developed beamforming-based speaker orientation estimation. However, this system requires a lot of time-consuming measurements of transfer functions, and thus it has low applicability. In this paper, to solve this issue, we propose the use of an acoustical simulation technique based on the wave theory to obtain transfer functions instead of measuring them for all possible combinations of positions and orientations. We performed the experiments to evaluate the system precision. The experimental results showed that our proposed speaker orientation estimation system achieves an enough estimation precision for human-machine interactions.

Key Words: Sound source orientation, Acoustical simulation, Beamforming, Human-robot interaction

1. はじめに

ロボットが人間社会に溶け込み役割を果たすためには，人間と同じような音声によるコミュニケーションが必要である．このため，ロボット上で人の聴覚と同様の機能を実現するロボット聴覚の研究が進められている．これまでの研究では，主に音源定位，音源分離，音声認識の３つ聴覚機能を中心に進められており，音源（発話者）の向き推定に関しては着目されていなかった．しかし，人・ロボット間のスムーズなコミュニケーションのためには，発話者の向きを推定する機能 (発話方向推定) も重要である．例えば複数人の集団の中で，ロボットが話者の発話方向を推定できれば，自身が話しかけられているのかどうかを判断することが可能になり，人間と同じようなコミュニケーションが実現できる．発話方向推定のために，中島らは伝達関数を音源の向きによって変化する関数に拡張することにより，位置だけでなく音源の向きに対しても焦点を形成するビームフォーミングを設計できることを示した [1]．この手法は大量の伝達関数をデータベースとして必要とするため，伝達関数を測定するためのコストの大きさが課題として挙げられる．そこで我々は音響シミュレーションにより，必要となる伝達関数を算出し，本手法の実用性を高めることを目標とした．音源の向きを含んだ伝達関数をシミュレーションにより算出する際に，音源の指向性の再現が課題となる．従来の音響シミュレーションにおいては，音源は無指向性の点音源である．そのため我々は音源形状のモデル化による指向性の再現について研究してきた [2]．本稿では，発話方向推定の技術として我々が用いた手法と波動音響シミュレーションの適用手法について述べる．また，実際に波動音響シミュレーションを用いた発話方向推定を行い，その評価検証について述べる．

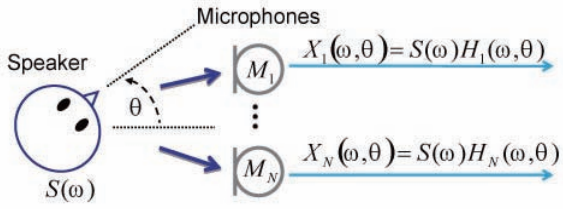


Fig.1 Model of wave propagation including sound source orientation.

2. 発話方向推定

音源の向きを推定する手法として, 音源の向きに拡張したビームフォーミングを用いた．ビームフォーミングは空間的な指向性を形成する技術であり, 系の伝達関数を用いて, 音源の位置を推定する際に一般的に用いられる．中島らは伝達関数を音源の向きによって変化する関数に拡張することにより, 位置だけでなく音源の向きに対しても焦点を形成するビームフォーミングを設計できることを示した [1]．本章では, その手法のアルゴリズムと実用上での課題を説明する．

2·1 音源の向きへ拡張した伝達関数

Fig. 1 に N 素子のマイクロホンアレイを用いた, 音源の向きを含んだ音声信号の伝播モデルを示す．$S(\omega)$ は発話音声の周波数応答, M_k は k 番目のマイクロホン, $H_k(\omega,\theta)$ は話者が θ 方向を向いている時の話者からマイクロホンへの伝達関数である．ここでは話者の位置は既知であるとした．マイクロホン M_k での受音信号

X_k は

$$X_k(\omega) = S(\omega)H_k(\omega) \tag{1}$$

と表される．伝達関数はベクトル表記で

$$\boldsymbol{h}(\omega,\theta) = [|H_1(\omega,\theta)|,\cdots,|H_N(\omega,\theta)|]^T \tag{2}$$

と表される．ここで, T は転置を示す．$\boldsymbol{h}(\omega,\theta)$ は複素成分（振幅成分と位相成分）のうち, 絶対値をとることで, 振幅成分のみを抽出した．理由は高周波帯域において位相成分が系の変動（例えば話者の口の高さや位置）に敏感で変化しやすく, 推定精度を下げる原因となるからである．

2·2 伝達関数データベース

伝達関数データベースは, 各向きの音源から各マイクロホンまでの伝達関数ベクトルをまとめたものであり

$$\boldsymbol{h}_0(\omega,\theta) = \frac{\boldsymbol{h}(\omega,\theta)}{\sqrt{\boldsymbol{h}(\omega,\theta)^H\boldsymbol{h}(\omega,\theta)}} = \frac{\boldsymbol{h}(\omega,\theta)}{|\boldsymbol{h}(\omega,\theta)|} \tag{3}$$

と表される．ここで, H は複素共役転置を示す．方向推定においては, 伝達関数ベクトル $\boldsymbol{h}_0(\omega,\theta)$ の向きのみが必要となる．そのため各周波数および各方向で正規化した, 出力機器の特性を含まない伝達関数データベースとなる．

2·3 伝達関数データベースを用いた発話方向推定

話者が θ_s(未知) 方向を向いて発話した際の受音信号の振幅成分は, ベクトル表記で

$$\boldsymbol{X}(\omega,\theta_s) = [|X_1(\omega,\theta_s)|,\cdots,|X_N(\omega,\theta_s)|]^T \tag{4}$$

と表され, 正規化すると

$$\boldsymbol{X}_0(\omega,\theta_s) = \frac{S(\omega)\boldsymbol{h}(\omega,\theta_s)}{|S(\omega)\boldsymbol{h}(\omega,\theta_s)|} = \frac{S(\omega)}{|S(\omega)|}\boldsymbol{h}_0(\omega,\theta_s) \tag{5}$$

となる．ここで Eqs. (3) と (5) の内積の絶対値は

$$\begin{aligned} C_\omega(\omega,\theta) &= |\boldsymbol{h}_0(\omega,\theta)^H\boldsymbol{X}_0(\omega,\theta_s)| \\ &= \left|\frac{S(\omega)}{|S(\omega)|}\boldsymbol{h}_0(\omega,\theta_s)\right| \\ &= |\boldsymbol{h}_0(\omega,\theta)^H\boldsymbol{h}_0(\omega,\theta_s)| \end{aligned} \tag{6}$$

となる．ここで, $C_\omega(\omega,\theta)$ は伝達関数における θ と θ_s の類似度関数を表す．ω を含まない類似度関数 $C(\theta)$ は

$$C(\theta) = \sum_\omega w(\omega,t)C_\omega(\omega,\theta) \tag{7}$$

と表される．ここで, $w(\omega,t)$ は非発話区間と雑音区間をマスキングするための重み付け関数である．この関数は HRLE (histogram-based recursive level estimation) マスキングを参考に算出した．詳しくは参考文献 [4] を参考にされたい．以上により得られた類似度関数 $C(\theta)$ が最大となるときの θ が, 推定方向 $\hat{\theta}$ として得られる．

2·4 実用化における課題

データベースとなる伝達関数は, 発話者の位置ごとに既知である必要がある．そのため, 実環境で本手法を利用するためには, 対象となる空間の伝達関数が大量に必要となる．伝達関数は通常実測により算出されるが, いくつかの課題がある．例えば, 手動で測定する場合は多大な時間と工数が必要となり, また測定精度が安定しない．自動測定装置を利用した測定でも, 装置が設置できる場所が限られたり, 費用がかかるなどの点が大きな課題となり，実用的でない場合がある．この問題を解決するためにシミュレーションによる伝達関数の算出が考えられる．シミュレーションにおいては, 音源の位置や向きを変えることは, パラメータを一部変更するだけの簡単な作業である．そのため実環境の再現さえ出来れば, 伝達関数のデータベース作成は, 実測に比べれば現実的な作業となり, 発話方向推定が実用的なアプリケーションとして実現できる．次章において, 本研究の目的のために我々が行ったシミュレーション手法について説明する．

3. 音響シミュレーションによる音源の指向性の再現

3·1 使用したシミュレータ

本研究では数値計算ソフト COMFIDA ver.2.03（日東紡音響エンジニアリング社製）を使用した．本ソフトは時間領域差分法を用いたシミュレータである．音響シミュレーションには幾何的手法と波動的手法がある．幾何的手法は音を線として扱う簡素な古典力学的な手法で, 音の波動性を考慮していない．そのため, 位相干渉, 回折, 固有モードといった現象を精確に再現できない．一方, 波動的手法は波動音響理論に基づいたアルゴリズムを持つため, 理論的には波動の性質を忠実に再現できる．代表的な波動的手法として, 時間領域差分法と有限要素法, 境界要素法がある．中でも時間領域差分法は物性値, 空間の再現精度, 計算メモリ, 計算時間の点でバランスが良い．本研究では比較的大きな空間の再現を行う．また, データベース作成のため, 計算が実用的な時間で行われる必要がある．さらには高精度な実環境の伝達関数を算出するためには, 物性値の設定に柔軟さが必要である．これらの目的を考慮し, 我々は時間領域差分法を利用した本シミュレータを用いた．

また時間領域差分法において, 空間は格子状に離散化されるが, 計算精度を高めるためには空間格子を小さくする必要がある．そこで本シミュレータはコンパクト差分法 [6] を適用することにより, 比較的大きな空間格子でも高精度な計算が可能となっている．

3·2 数値計算の精度評価

本シミュレータの計算精度の検証を行った．時間領域差分法において, 計算精度は波長 (λ) に対する空間格子 (ΔL) の比率に依存する．$\lambda/\Delta L$ が十分に大きくないと数値分散という現象が発生し, 波の伝播において振

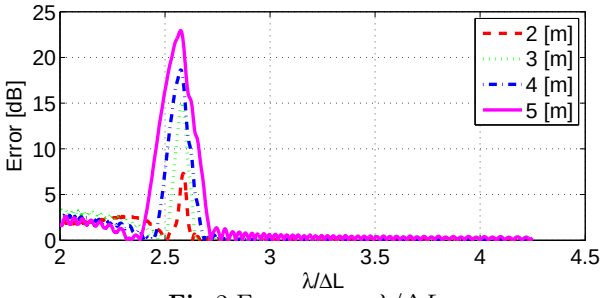


Fig.2 Error versus $\lambda/\Delta L$

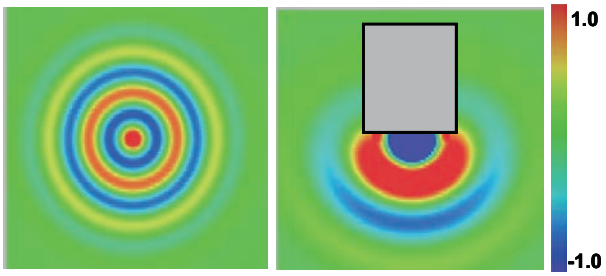


Fig.3 Sound pressure map.

幅と位相に大きな計算誤差が生じる．また，この計算誤差は伝播するほど大きくなる．Fig. 2 に本シミュレータにおける振幅誤差と $\lambda/\Delta L$ の関係を示す．振幅誤差は伝播距離 1 m を基準として，伝播距離 2, 3, 4, 5 m における距離減衰を理論計算と比較した．Fig. 2 より，数値分散による計算誤差は $\lambda/\Delta L$ の値に比例して小さくなると同時に，2.6 付近で局所的に増大している．つまり $\lambda/\Delta L$ が 3 以上の場合，計算誤差が比較的低く，また安定していることが分かる．

3·3 音源の指向性の再現

シミュレーションによる指向性の情報を含んだ伝達関数の算出のために，音源のモデル化による指向性の再現手法の説明をする．従来の手法では指向性を持たない点音源であり，音は同心円状に伝播する (Fig. 3，左)．それに対し，我々は適当な形状を点音源に付加することで指向性が作り出せることに着目した (Fig. 3，右)．指向性を再現する手法はいくつかあるが，本提案法は比較的簡単に指向性を作り出せる点にメリットがあり，今回のような目的に適していると考えた．また，波動理論に従う波動的手法を用いているので，周波数による回折の違いも再現できる点も大きなメリットである．

4. 音響シミュレーションの発話方向推定への適用

4·1 実験環境

部屋の大きさは 7.0 m × 4.0 m × 3.2 m であり，残響時間は約 230 ms である．Fig. 4 に実験室の様子を示す．室内にはキッチン台がある．マイクロフォンは壁面とキッチン上に計 96ch 取り付けられている．マイクロフォンの配置を Fig. 5 に示す．ここで音源の位置は部屋の中心とした．

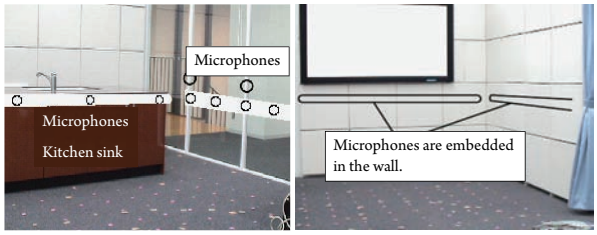


Fig.4 Experimental room.

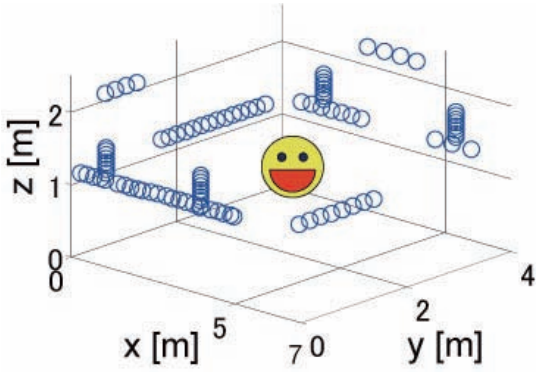


Fig.5 Microphone positions.

4·2 入力信号

実験室において，女性話者が音源位置で，90 °ごと 4 方向に，“a, i, u, e, o”と発話した音声を収録した．また 発話方向推定システムの動作検証のために，伝達関数の元のデータであるインパルス応答と白色雑音を畳み込みにより合成した信号を用意した．

4·3 伝達関数の算出

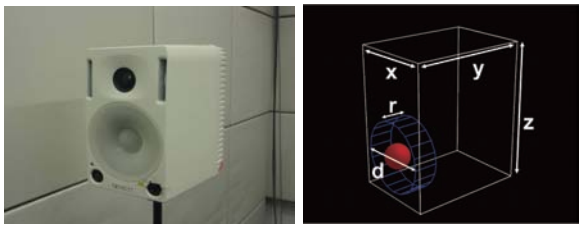
4·3.1 実測

スピーカ（GENELEC 1029A）を音源とし，96ch のマイクロフォンでインパルス応答を測定した．音源信号には信号長 2^{14} の TSP 信号を用いた．サンプリング周波数は 16 kHz とした．インパルス応答はスピーカを 45 °ごとに回転させ，計 24 方向のデータを測定した．

4·3.2 音響シミュレーション

実測による伝達関数の算出を再現する．Fig. 7 に示すように，実測で行ったインパルス応答の測定環境をシミュレーションで再現した．緑の球は評価点つまり，マイクロフォンである．

スピーカの指向性の再現手法は前章と同様である．Fig. 6(a) に再現の対象となるスピーカである GENELEC 1029A を示す．Fig. 6(b) にシミュレーションで再現したスピーカのモデルを示す．剛壁の立方体に対して振動板に対応する面に円柱状の窪みを作ったモデルであり，スピーカの構造，吸音率，平面振動などは無視した簡素なデザインである．スピーカの指向性の再現に関して，詳しくは参考文献 [2] を参考にされたい．



(a) GENELEC 1029A. (b)Loudspeaker model.
Fig.6 Loudspeaker.

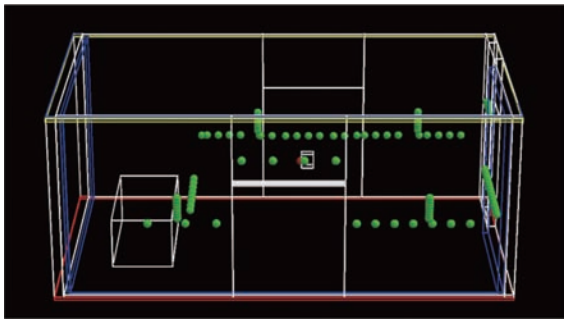


Fig.7 Simulated experimental room.

Fig. 6(b) におけるスピーカモデルのパラメータは, [x, y, z, r, d] = [0.14, 0.20, 0.24, 0, 0] とした. ここで, 各パラメータの単位は [m] である. 点音源の位置は固定し, 付加する立方体を 45 °ごとに回転させ, 実測の様に計 24 方向の音源の回転を再現した.

実験室は比較的大きいため, 空間格子間隔を局所的に細かくする手法 [5] を用いた. 空間格子間隔は, スピーカモデルや評価点の周りでは 0.02 m と細かく設定し, それ以外の領域では 0.08 m と粗い格子間隔に設定した. 音源は 500–2500 Hz においてフラットな周波数特性を持つパルス信号を用いた. 前章に示したとおり, 計算精度を下げないために, $\lambda/\Delta L$ を 3 より大きくしたい. そのための周波数の帯域の最大値は 1400 Hz である. 計算されたインパルス応答を 16 kHz にダウンサンプリングし, 実測と同様に伝達関数を算出した.

4·4 結果

入力信号は, 合成信号 (NOISE), 音声信号 (SPEECH) の 2 種類である. 伝達関数は実測ベース (PRA) とシミュレーションベース (SIM) の 2 種類である. また, 評価する周波数は, シミュレーションにおける音源の周波数帯域 (500–2500 Hz) と, 計算精度が安定している周波数帯域 (500–1400 Hz) の場合で分けた. つまり全部で 8 つの条件で発話方向の推定誤差を評価した. 各条件において, 各音源方向 (0 °, 90 °, 180 °, 270 °) の推定誤差の平均をとった. 結果を Fig. 8 に示す. 入力信号が合成信号の場合, 実測・シミュレーションどちらの場合でも推定誤差は 13 °以下であり, 方向推定が正しく行われていることがわかる. 入力信号が音声の場合, 実測ベースは 10 °程度に対してシミュレーションベースは 25 °程度と, 比較的高い精度であった. シミュレーションにおいては, 周波数帯域が変わっても推定精度があまり変わらないことから, 計算精度よりも空間の再現精度やスピーカの指向性の再現精度に, 推定精度向上のキューがあると考えられる.

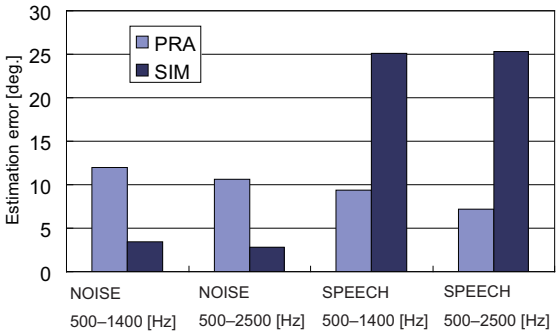


Fig.8 Orientation estimation errors

5. まとめ

本稿では波動音響シミュレーションを音源の向きに拡張した伝達関数によるビームフォーミング法を用いた発話方向推定に適用した. 音源方向を含んだ伝達関数を算出するために音源形状のモデル化により, スピーカの指向性の再現を行った. 実環境において発話方向推定を行った. 伝達関数は実測と音響シミュレーションにより算出した. 実験の結果より, シミュレーションを適用することにより, 誤差 25 °程度の推定精度が達成された. この結果は話者の発話方向を前後左右のレベルで推定が可能であることを示しており, 人・ロボット間のスムーズなコミュニケーションの実現のためには, 我々が用いた手法が十分に適用できることが分かった. また, さらに推定精度を高めるためには, 計算精度よりも空間や指向性の再現精度に課題があることも分かった. 今後の課題は, 中島らによる本推定手法の実時間処理化 [1] に適用することである.

参考文献

[1] H. Nakajima *et al.*, "Real-time sound source orientation estimation using a 96 channel microphone array," IROS-2009, 676–683, 2009.
[2] 鈴木淑正 他, "波動音響シミュレータによる指向性の精度検証," 信学技報, 109 (100), 109–114, 2009.
[3] V. Rogijen *et al.*, "Acoustic source number estimation using support vector machine and its application to source localization/separation system," IEICE EA, 102 (249), 25–30, 2002.
[4] H. Nakajima *et al.*, "An easily-configurable robot audition system using histogram-based recursive level estimation," IROS-2010 (to be appeared).
[5] H. Tsuru and R. Iwatsu, "Accurate numerical prediction of acoustic wave propagation," Int. J. Adapt. Control Signal Process., 24, 128–141, 2010.
[6] S. K. Lele, "Compact finite difference scheme with spectral-like resolution," J. Comput. Phys, 103, 16–42, 1992.