

研究種目：基盤研究（C）
 研究期間：2007～2008
 課題番号：19500164
 研究課題名（和文） 重み付きウィナー利得を用いた音源方向推定処理に基づく聴覚処理モデル
 研究課題名（英文） AUDITORY PROCESSING MODEL BASED ON DIRECTION OF ARRIVAL ESTIMATION USING THE WEIGHTED WIENER GAIN
 研究代表者
 永田 仁史（NAGATA YOSHIFUMI）
 岩手大学・工学部・准教授
 研究者番号：40301030

研究成果の概要：本研究グループで提案している重み付きウィナー利得に基づく方向推定法に関し、両耳受音信号からの複数音源方向推定における逐次処理を用いた精度向上法を開発し、性能を評価した。この結果は、国際雑誌 IEEE Transactions on Audio, Speech and Language Processing (2009.1)に採録された。また、ヒト頭部の正中面上の音源定位精度を測定し、本提案方向推定処理との性能比較を行った結果、提案処理の方が高い精度であったことから、ロボット聴覚の処理としては問題ないものの、聴覚処理モデルとしてはさらに検討が必要であることがわかった。

交付額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|---------|-----------|
| 2007年度 | 900,000 | 270,000 | 1,170,000 |
| 2008年度 | 900,000 | 270,000 | 1,170,000 |
| 年度 | | | |
| 年度 | | | |
| 年度 | | | |
| 総計 | 1,800,000 | 540,000 | 2,340,000 |

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音源方向推定、頭部伝達関、数ロボット聴覚、重み付きウィナー利得

1. 研究開始当初の背景

近年開発が盛んなヒト型のロボットにおいては、自律動作と人とのコミュニケーション能力向上のため、周囲の状況を認識するセンサ能力が重要であると考えられている。とくに、聴覚による状況認識能力については、到来音方向の感知が状況認識に大きく関わっていると考えられ、また、「The advantage of knowing where to listen」, (J. Acoust. Soc. Am, pp3804-3815, Dec., (2007))において報告されているように、音源方向の情報が音声内容の把握に役立っていることが実

験的に確かめられている。

このような背景から、本研究グループは、ロボット聴覚において有用な音源方向推定処理に関する研究を進めており、ヒトが2つの耳からの信号を用いて種々の重要な情報を抽出できることに習い、2個の指向性マイクロホンの出力を用いた重み付きウィナー利得による方向推定法を開発した (IEEE Trans. Audio, Speech & Language Process., Vol. 15, No. 2, pp. 416--429, Feb. 2007)。この方法は、相関ベースの方法でありながら、高雑音下で複数の到来音声に対し、方位角一

仰角の2次元分解能があることなど、大幅に従来法を上回る性能を持つ。この方法は、マイクロホンの指向特性を人間の頭部伝達関数に置き換えても、他の従来法に比べて大幅に高い推定精度であるが、複数音源環境における性能は音源1個の場合に比べて大幅に低いものであった。

2. 研究の目的

そこで、この研究においては、まず、重み付きウィナー利得に基づく方向推定法において、頭部伝達関数を経た両耳受聴音の場合の複数音源の推定精度向上法を開発、評価する。ヒトが数個の音声方向を同時に定位できることから、その能力に匹敵する性能となることを目指す。次に、未だにはっきりとは解明されていない聴覚の正中面上の方向定位処理について、本手法と人間の聴覚処理との関連を探り、従来、状況によって異なった定位のキューが提案されていたのに対してこれらを統一的に解釈できるような音源定位のモデルを提案する。以上が本研究の目的である。

3. 研究の方法

まず、(1) 重み付きウィナー利得に基づく音源方向推定法において、両耳受聴音からの2次元(方位-仰角)の複数音源推定の際の精度向上のため、逐次アルゴリズムによる性能向上法について、計算機シミュレーションにより性能を評価する。

次に、(2) 音源定位に関する新しい処理モデルの検討のため、聴覚における音声の到来方向推定の精度を測定し、本提案の方向推定法の精度との一致具合を調べる。

聴覚の方向同定精度測定は、被験者による主観評価によって行う。本研究の場合、(A) 被験者の HRTF を用いて到来音を模擬し、ヘッドホンから模擬した音を提示する方法と、(B) 無響室において実音源から音を放射する方法とがある。(A)の方法は、無響室や実音源の設置が不要であるが、被験者本人の HRTF を用い、十分な定位精度があることを確保する必要がある。HRTF については、H17, 18 年度に頭部モデルを作成して測定した4人分のデータを用いる。一方、(B)の方法は直接的に信頼性が高いが、無響室を長時間使用すること、音源を移動させる、あるいは、多数設置するなどの装置が必要であるなど、施設を持たない場合、実験実施は困難であるが、(A)の方法がうまくゆかない場合は、(B)により行う。

4. 研究成果

(1) 重み付きウィナー利得に基づいた音源方向推定法の高精度化

提案法の高精度化のため、逐次的な方法による性能向上法を開発し、評価した。ここで述べる逐次処理とは、まず、音源が1個であ

ることを仮定して方向推定し、求めた1個の音源に起因する周波数成分を減衰させた上で、さらに音源方向推定する、という処理を繰り返すものである。どの方法にも使えるわけではなく、複数の音源の影響が重畳した空間スペクトルにおいて、最大ピークの方向が音源の一つと一致する割合 (Maximum Peak Correct Rate (MPC)) が大きくなければ適用できない。そこで、いくつかの主な方向推定手法について、MPC を計算したのが下の図1の結果である。

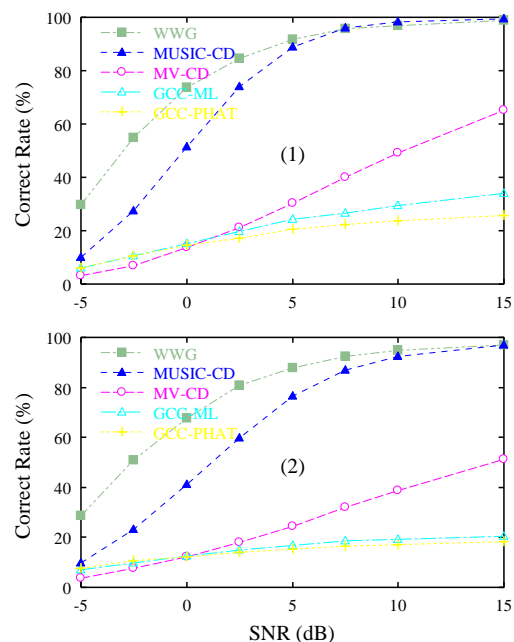


図1. 初期スペクトルの最大ピークが音源方向の一つと一致する割合

図1において、WWGは重み付きウィナー利得、MUSIC-CDはコヒーレンス検出付きのMUSIC法、MV-CDは、同検出付きの最小分散法、GCC-MLは最尤推定重み付きの一般化相互相関関数(GCC)、GCC-PHATは白色化重み付きのGCCである。図からわかるように、WWGは最もMPCが高く、2番目はMUSICであり、その他の方法は、これら2つの方法に比べて大幅に精度が低い。従って、本方法が逐次法に向いていることがMPCによって確認できた。

次に、逐次法による方向推定性能の評価結果を図2に示す。図2において、パネル(1)、(2)、(3)は、各々、音源を1個、2個、3個と仮定した場合の結果であり、WWG-ISAと記されているのが提案逐次法である。この図を見ると、音源1個の場合はWWGはMUSIC-CDと同等の性能であるが、音源が2個になると、WWGの逐次法であるWWG-ISAの性能が最も高くなり、他の方法との性能差が大幅に開く。音源3個の場合に至っては、他の方法ではほとんど推定できないことがわかる。

さらに、正中面上（仰角の1次元）に限定した場合の音源方向推定精度について検討した結果を図3に示す。図3は、正中面上に2個の音源が仰角5度だけ離れて存在する場合に、音源の仰角を変えながら検出率を求めた結果である。この図からわかるように、WWG-ISAは最も性能が高く、前方では仰角によらず90%程度の検出率である。

(2) 聴覚の定位精度測定と聴覚処理モデル

まず、聴覚の定位精度測定のため、研究方法で述べた(A)のヘッドホン受聴による方法を試みた。H17, 18年度に頭部モデルを作成して測定した頭部伝達関数を用いて両耳受聴音を模擬したが、前方の定位精度が十分ではなく、耳道の伝達関数やヘッドホン特性等、様々な補正を試みた改善しなかった。また、受音位置をいくつか変えて伝達関数を測定し直したが、やはり、改善しなかった。そこで、(B)の方法により、スピーカから提示することとし、研究室内に簡易な無響室を作成して実験した。

音源は1個で提示信号は、260Hz~6kHzの音声とし、正中面上の仰角をランダムに変えながら被検者に信号を提示した。被検者は5人であり、実音源、回答方向、ともに5度きざみである。この結果を図4と図5に示す。結果は、回答方向の頻度に応じて濃淡で示してある。図4は頭を固定した場合、図5は上下の頭の動きを許した場合の結果である。

これらの図からは、音の再生仰角と回答した仰角に相関があるが、実際の角度よりも回答した角度の方が絶対値が小さいことがわかる。平均の正解率は、許容誤差を10%とした場合に、頭を固定した場合で56%、動きを許した場合で約61%の正解率であった。提示信号の帯域制限がない場合と比べれば低い正解率であるが、仰角定位のスペクトルキューがあることとされている高い周波数帯域が提示信号から除去されているにもかかわらず、おおよその方向が知覚できることを示しており、伝達関数の微小な左右差を知覚に使用している可能性があることが示唆される。

一方、図3からわかるように、WWG-ISAの場合は2音源のときに許容誤差5度でほぼ90%の正解率であることから、WWG-ISAは聴覚の定位精度をはるかに上回ってしまっていることがわかる。この結果だけからは、GCCのようなもっと簡単な方向推定処理の方が聴覚処理のモデルとして妥当のようにも考えられるが、図2からわかるように、GCCは複数音源環境における2次元の推定ではほとんど性能が出ないのに対し、聴覚は数個の音声を定位できることから、この考えは否定される。ここで考慮すべきなのは、シミュレーションで検討した方法が詳細な頭部伝達関数の情報をすべて使っている一方で、聴覚処理の方は、伝達関数に関しては概

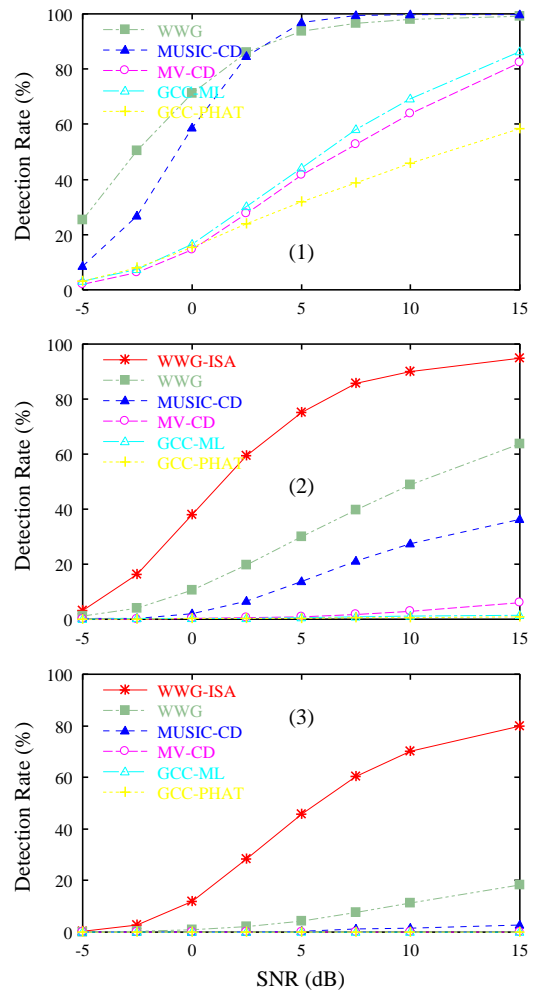


図2. 音源方向検出率

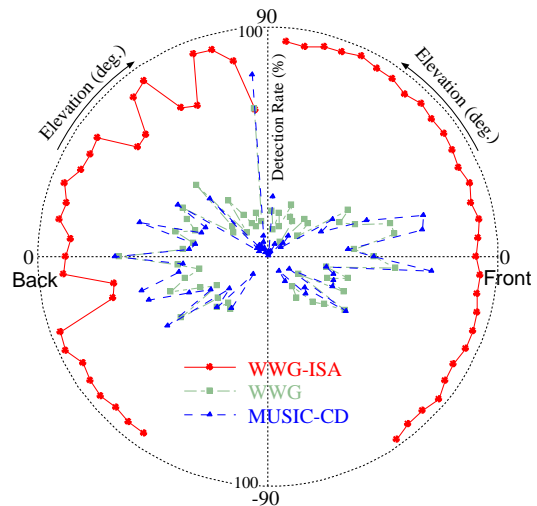


図3. 正中面において仰角が5度離れた音源2個の検出率 (S/N=10dB)

略の値しか持っていないであろうと考えられることである。従って、頭部伝達関数を簡略化して情報量を低下させてゆけば、

WWG-ISA の精度を聴覚の定位精度に近づけることはできるものと考えられる。頭部伝達関数上のどのような情報を残すのが妥当であるのかについても本研究において検討する予定であったが、当初予定していたヘッドホンを使った音源定位の精度が聴覚の定位精度測定には不十分であり、これに代えて無響室の作成が必要となったことにより、これに関しては検討の余裕がなく、今後の課題として残された。

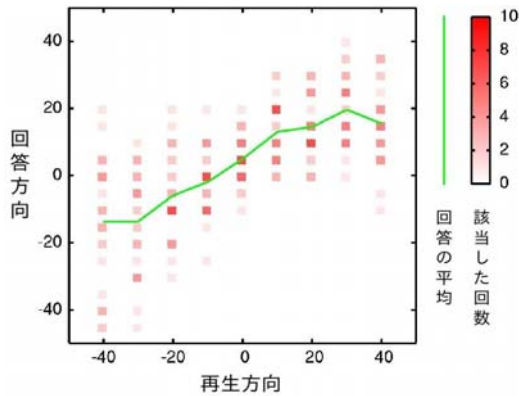


図4. 正中面上における仰角の音源定位結果（頭を固定した場合）

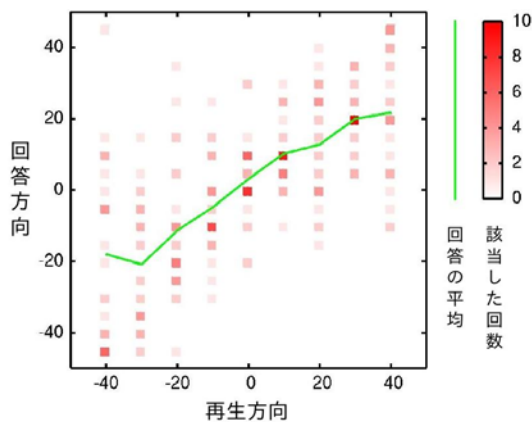


図5. 正中面上における仰角の音源定位結果（頭の動きを許した場合）

5. 主な発表論文等

〔雑誌論文〕（計 1 件）

① Yoshifumi Nagata, Satoshi Iwasaki, Takahiko Hariyama, Toyota Fujioka, Tomita Obara, Takayuki Wakatake and Masato Abe : Binaural Localization Based on Weighted Wiener Gain Improved by Incremental Source Attenuation
IEEE Trans. on Audio, Speech, and Language Processing, Vol. 19, No. 1, pp. 52--65, Jan. 2009

〔学会発表〕（計 1 件）

①両耳受聴音からの音源方向推定の精度向上に関する一検討、若竹孝幸、永田仁史、岩崎聡、針山孝彦、藤岡豊太、安倍正人、電子情報通信学会技術報告、EA2007-4 pp. 17--22 (2007. 4)

6. 研究組織

(1) 研究代表者

永田 仁史 (NAGATA YOSHIFUMI)
岩手大学・工学部・准教授
研究者番号：40301030

(2) 研究分担者

安倍 正人 (ABE MASATO)
岩手大学・工学部・教授
研究者番号：00159443

藤岡 豊太 (FUJIOKA TOYOTA)
岩手大学・工学部・助教
研究者番号：60292174