様式 C-19

科学研究費補助金研究成果報告書

平成　21　年　10　月　13　日現在

研究種目：若手研究（B）
研究期間：　2007　～　2008
課題番号：19700186
研究課題名（和文）　音源方向推定および音源分離機能を有するロボットの擬似聴覚機能に関する研究
研究課題名（英文）　A study of DOA estimation and speech segregation using auditory model for humanoid robot

研究代表者
中島　栄俊　（NAKASHIMA HIDETOSHI）
熊本電波工業高等専門学校　電子制御工学科　准教授
研究者番号：00353350

研究成果の概要：　本研究では，人間聴覚をモデル化した周波数領域両耳聴モデル（FDBM）を用い，ロボット聴覚機能の一部である音源方向推定機能，音源分離機能を構築した．また，同時に実環境下において変動する伝達関数を，ロボット頭部を回転させることにより擬似的に推定し，FDBM 内部のデータベースを更新することによって音源分離性能の改善を試みた．この結果，FDBM のデータベースを更新することにより分離信号のＳＮＲを改善することができた．

交付額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	1,800,000	0	1,800,000
2008 年度	700,000	210,000	910,000
年度			
年度			
年度			
総　計	2,500,000	210,000	2,710,000

研究分野：総合領域
科研費の分科・細目：知覚情報処理・知的ロボティクス
キーワード：音源分離，HRTF，聴覚モデル，音源方向推定，データベース

１．研究開始当初の背景

ロボット研究が盛んに行われる昨今，その形態は自律型へと進化している．この自律型ロボットにおいてはその行動を決定する思考メカニズムが重要な役割りをなしており，多くの場合，各種センサから得られる様々な信号を利用している．こうしたセンサ類のうち，人間と会話によるコミュニケーションを行う上で音響センサは必要不可欠の存在である．しかしながらロボットに人間の聴覚と同等の機能を持たせるにはいくつかの解決すべき問題が存在する．実環境においてロボットはコミュニケーションを図っている特定人物の音声だけではなく周囲の人間の音声，環境雑音，騒音等を入力信号として捉える．この為，特定人物の音声を認識するにはその音声信号とそれ以外の不要な信号を分離する必要がある．また特定音源がどの方向から到来しているのかを知ることも重要な問題である．

これに対し，我々人間は雑音環境下においても目的音のみを注意して聞き取ることができる．カクテルパーティ効果として知られるこの聴覚機能を利用した音源分離手法が Bodden によって提案されているが，我々はこのアルゴリズムを高速化，高性能化した周波数領域両耳聴モデル(Frequency Domain Binaural Model : FDBM)を提案しその有効性

を示してきた．

本研究では，この FDBM をロボット聴覚機能の一部として用い，観測信号から特定音声信号のみを分離し，人間とのコミュニケーションで非常に重要となる音声認識精度を向上させるロボット聴覚システムを構築する．さらにこのシステムでは音源の方向推定を行い特定話者方向にロボット頭部を回転させより人間に近い動作が行えるロボットの開発を行う．このロボット頭部の回転は動作が人間に近いだけでなく，音源分離性能，音声認識性能および音源方向推定性能を向上させることができる極めて重要な機能となる．

２．研究の目的

本研究においての目的は反射等の存在する実環境下において音源分離・音声認識を行うことのできるロボットの開発である．本研究で構築するシステムの全体像を図１に示す．

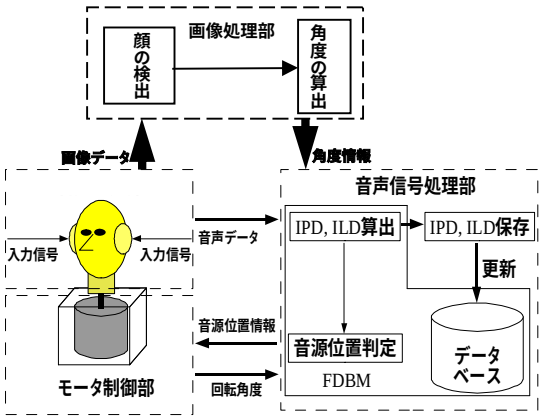

図１ システム全体図

（１）実時間音源分離と音声認識

FDBM はこれまで主としてオフラインで動作させていたが，これをオンライン動作にし，ロボット頭部で観測した信号の音源分離および音声認識を行うシステムを構築する．FDBM は演算量が究めて少なくその実時間動作には DSP を必要としない． なお，（３）ステレオカメラによる音源方向推定についても実時間処理を行う．

（２）ダミーヘッドの頭部回転によるデータベース更新

FDBM の音源方向推定性能および音源分離性能は FDBM のもつデータベースの精度に依存する．即ち，データベース内に予め登録している周波数－方向別の両耳間位相差(interaural phase difference : IPD)と両耳間レベル差(interaural level difference : ILD）マップが FDBM を動作させる環境下における IPD，ILD と一致すれば方向推定性能および分離性能が向上する．これまでの研究ではデータベースの IPD，ILD は無響室で測定された HRTF から求めていた．しかしながら実環境下においては壁面，床面等からの反射が発生することから，無響室における IPD, ILD と実環境下における IPD, ILD が一致しない．このため，音源方向推定性能および音源分離性能が著しく低下する．この問題を解決するためには FDBM を動作させる環境ごとにデータベース内の IPD，ILD を更新させる必要がある．

本研究ではダミーヘッドの頭部を回転させることにより，瞬時 ILD を複数観測させデータベースを更新することを試みる．またデータベース更新による音源分離性能を音声認識率により定量的に評価する．

（３）ステレオカメラによる音源方向推定

（２）におけるデータベース更新はダミーヘッド頭部を 10°ごとに回転させ，－90°～＋90°（ダミーヘッド正面を 0°とする）方向の ILD を求め，データベースを更新している．この際，音源方向は既知であるとしているが，実際には音源方向の推定は難しく，また一方で音源方向を正しく推定しておく必要がある．

本研究ではダミーヘッド頭部にステレオカメラを搭載し，カメラで取得したステレオ画像から人間の顔を検出させ話者方向推定を試みる．これにより，反射の多い環境や雑音環境下においても話者位置を正しく推定することができ，データベースの更新に利用することができる．

（４）実環境下におけるデータベース更新

（２）におけるデータベース更新をより現実的にするために，ダミーヘッドの頭部回転とステレオカメラを用いた実環境下におけるデータベース更新を試みる．（２）では頭部を 10°ごとに回転させ全方向の瞬時 ILD を計測していたが，頭部の回転角度を限定し，短時間でデータベースを更新させる．なお，観測していない角度に関しては観測した瞬時 ILD をもとにデータ補間を行う．なお，データベース更新におけるデータ補間の評価は分離信号の SNR を用いて行う．

３．研究の方法

（１）実時間音源分離と音声認識

FDBM は演算量が非常に少なく，汎用の PC でも実時間処理が可能であることから実時間音源分離処理と音声認識処理は汎用 PC で行う．ただし，PC の処理能力等を考え，分離処理と認識処理は別々の PC で行う．

まず音声信号が入力されると音源分離処理を PC1 側で行う．PC1 側で分離された音声は TCP/IP を利用したパケット通信で PC2 側

に伝送される．伝送された信号を随時 PC2 側において音声認識処理を行う．

この開発においては上記処理と同時に PC1 側でステレオカメラ画像も所得し，話者の位置推定を行うことが可能であるようにする．また，必要に応じてダミーヘッド頭部を回転させる制御を PC1 側で行うこととする．

（2）ダミーヘッドの頭部回転によるデータベース更新

データベースの更新の有効性を確認するために，実環境下において音源位置を固定し，ダミーヘッドの頭部を 10°ごとに回転させ，それぞれの角度で得られた瞬時 ILD を用いてデータベースを更新させる．更新には下式を用いる．

$$D(n,\theta) = D(n,\theta)\cdot(1-\alpha) + D'(n,\theta)\cdot\alpha$$

ここで n は周波数インデックス，θ は更新する角度，$D(n,\theta)$ および $D'(n,\theta)$ はそれぞれ更新前のデータベース，観測信号から得られる瞬時 ILD とする．また α は忘却係数である．この更新されたデータベースを用いて音源分離を行い，分離信号を用いた音声認識率でデータベース更新の有効性を確認する．また同時に SNR を用いた客観的評価値によるデータベースの有効性の検証も行う．

（3）ステレオカメラによる音源方向推定

データベース更新時における音源方向の推定としてステレオカメラを用いる．このステレオカメラで取得したステレオ画像から話者の方向を推定する．

まず，取得した 2 枚の画像からそれぞれに Haar-Like 特徴量を利用した顔検出を行う．顔位置は矩形として出力されるため，この矩形の重心画素位置を求め，話者の方向を求める．一方、一般にカメラ画像は画像の中心から離れるにしたがって歪むことが知られている．これはレンズ歪みが原因であるが，この歪みにより方向推定性能が低下する．これに対し，歪み補正を行う処理を施して方向推定精度を改善させる．

（4）実環境下におけるデータベース更新

実環境下におけるデータベース更新は以下の手順で行う．なお，データベース更新時における音源数は 1 とする．

i) 話者が発声
ii) ステレオカメラによる話者方向推定
iii) 音声信号から瞬時 ILD を計算，平均化
iv) 頭部回転（θ°）
v) 再度，瞬時 ILD を計算
vi) 必要に応じて iv),v)を繰り返し
vii) 得られた複数の瞬時 ILD より他方向の瞬時 ILD を補間
viii) データベース更新

これらの処理により更新されたデータベースの評価は分離信号の SNR を用いて行う．なお，瞬時 ILD 補間時における観測値の数（即ち，v），vi)の繰り返し回数）がデータベース更新にどのような影響を与えるかについても検証する．

4．研究成果

（1）実時間音源分離と音声認識

実時間音源分離と音声認識においてはその開発が主たる目的であるため，全体としての定量的な評価は行っていない．しかし，システムを個別にみると FDBM の分離処理における遅延が 30ms 程度，音声認識におけるパケット伝送およびバッファリング処理に 500ms 程度の遅延が発生する．また，画像処理による顔検出と話者方向推定を行う場合には 200ms 程度の遅延が発生する．従って，システムを総合的に動作させる場合においては全体で 1 秒程度の遅延量となってしまう．

今回，構築するシステムにおいては 1 秒程度の遅延は大きな問題にならないが，場合によっては注意する必要がある．

（2）ダミーヘッドの頭部回転によるデータベース更新

図 2 にデータベースの更新を行った際の ILD の値を示す．この図において横軸は周波数，縦軸は ILD 値を示す．図中の Dn はデータベースの初期値，Dm が真値であり，D が初期値から更新された値である．この図からわかるように，データベースを更新することによって，データベースが真値に近づいている．

図 3 にはこれら 3 種類のデータベースを用いて音源分離を行った音声信号の音声認識装置を用いて評価した結果を示す．この図における縦軸は認識率である．この図からわかるように，音声認識率においても更新したデータベースを用いた場合は真値を用いた場合の認識率に近く，データベース更新の有効性が伺える．

また，図 4 はデータベースを更新した際の分離信号を SNR により定量的な評価を行った結果である．この実験では目的信号を-30°に固定し，妨害音を-20°から＋30°まで変化させた際の SNR の変化である．図中の Dn は更新前のデータベースを用いた結果，D は更新後のデータベースを用いた結果を表す．なお，図の横軸は妨害音位置，縦軸は SNR 値であり，目的音と妨害音の SNR は 0dB に設定している．この図からもわかるように，分離信号の精度はデータベース更新により平均で 1dB 程度改善されている．これらのことからデータベース更新が音源分離精度改善に効果があることが分かった．

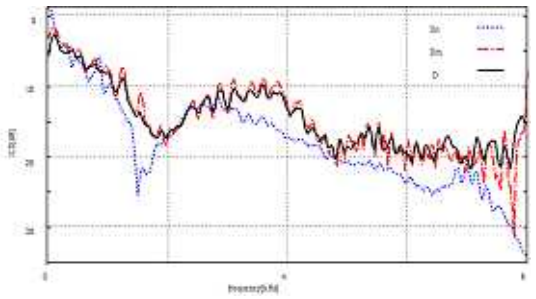

図 2 データベース更新における ILD の変化

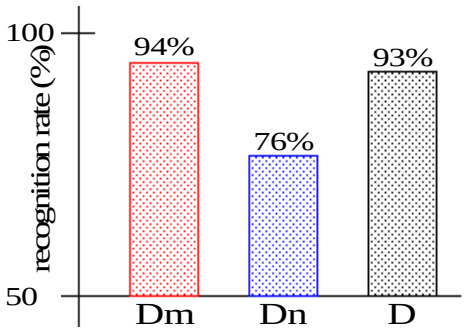

図 3 データベース更新による分離信号の音声認識率変化

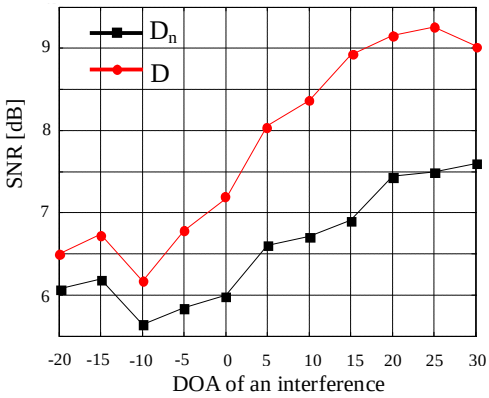

図 4 データベース更新における分離信号の SNR 値

（3）ステレオカメラによる音源方向推定

実環境での正確な話者の位置情報を取得するため，顔の角度推定実験を行った．測定にはステレオカメラ(TD-BD-SCAMv2)を使用し，実験の際は，これをダミーヘッドの両目に埋め込んだ. 検出させる顔として，人間と同じ大きさ程のマネキン(頭部のみ)を用意した．なお，カメラから取り込む画像は VGA(640×480)サイズである．実験では，カメラの正面を 0°として，話者（マネキン）が－30 ° ～+30°まで 5°おきに移動したときの推定角度と実角度との誤差を測定した．データは，それぞれの実角度において 5 回分の誤差を測定し，実角度毎に推定角度との平均誤差を算出した．カメラの初期位置は，マネキンを実角度 0°に置いたときに，推定角度が 0°を検出する位置に調整した. また, マネキンとカメラの距離を 1 m とし，それぞれの高さが同じになるよう配置した.

測定結果グラフを図 5 に示す. 横軸の実角度が側方にいく程，縦軸の平均誤差が大きくなっていることが分かる．この原因として，カメラのレンズの歪みにより，マネキンの位置が外側に膨張している画像を取得してしまい，これを画像処理していることが挙げられる．マネキンの正確な方向を得るために，最小二乗近似を用いた角度の補正を行い, 再度測定した. 補正前に比べ, 補正後は実角度との平均誤差が全て 1°未満に収まっており，概ね正確な位置推定が行われていることが分かる．このことからステレオカメラによる話者方向推定が有効であることが分かった.

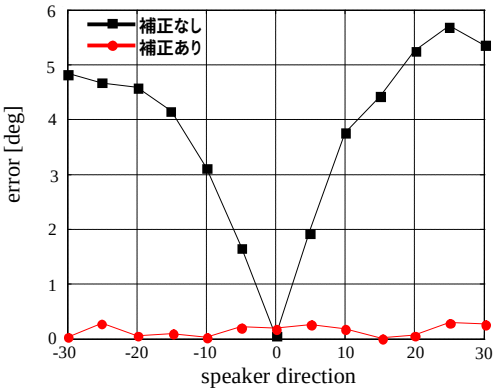

図 5 ステレオカメラによる話者方向推定結果

（4）実環境下におけるデータベース更新

ダミーヘッドの頭部回転とステレオカメラによる話者方向推定を組み合わせたデータベース更新アルゴリズムの検証を行った. この実験ではデータベースを更新させる際の瞬時 ILD の補間方法として，2 方向からの補間，3 方向からの補間，4 方向からの補間の 3 種類を行い，これら補間の違いによる音源分離精度の差についても検証した．なお，補間に用いた方向および得られたデータベースは表 1 のとおりである.

図 6 に瞬時 ILD の補間によるデータベース更新を行った際の分離信号 SNR を示す．図中の D はデータベースの初期値を用いた結果，Dm は補間ではなく，全方向の瞬時 ILD を用いてデータベースを更新した結果である．なお，この実験においてはデータベース更新の後，妨害音を-30 ° に固定し，目的音を-20 ° から +30 ° まで 5 ° おきに変化させ，音源分離を行

った．この図からデータベースを更新させたことにより分離信号の性能が改善していることがわかる．特に，データベース更新前は10°における分離性能が極めて低かったのに対し，これがデータベースにより改善されている．ただし，データベース更新における瞬時 ILD の補間では 3 点以上の瞬時 ILD を観測することが望ましい．

以上のことから FDBM におけるデータベース更新により，実環境下においての分離精度を改善することが可能となった．

表 1　瞬時 ILD 補間における観測方向とデータベース

補間に使う瞬時 ILD の方向	データベース
30°，0°	D_2
30°，−15°，0°	D_3
30°，−20°，−10°，0°	D_4

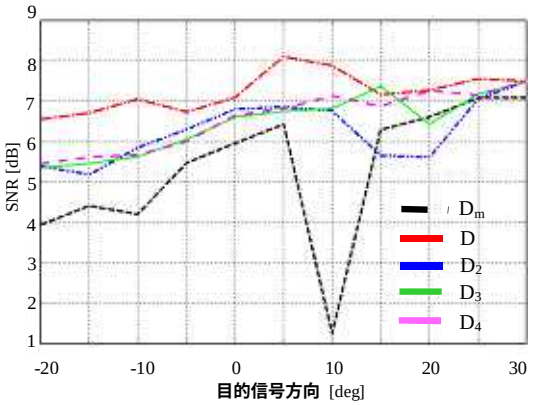

図 6　瞬時 ILD の補間によるデータベース更新と分離信号 SNR

5．主な発表論文等

（研究代表者，研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　1　件）

① 中島栄俊，常田貴史，脇坂龍，加茂田浩史，“周波数領域両耳聴モデルを用いたロボット聴覚実現の試み－頭部回転による音源分離性能の改善－”，日本高専学会，Vol.13, No.4，32—35，2008　査読無

〔学会発表〕（計　3　件）

① RYO WAKISAKA, TAKAAKI ISHIBASHI, HIDETOSHI NAKASHIMA, “Database adaptation for FDBM using stereo images and instantaneous ILD,” IWPASH2009, Nov. 2009 (To be Published)

② 脇坂龍，常田貴史，中島栄俊，石橋孝　苣木禎史，宇佐川毅，“画像情報を利用した FDBM のデータベース更新による音源分離性能改善の試み”，日本音響学会講演論文集（春），CD-ROM, 2009

③脇坂龍, 加茂田浩史, 中島栄俊, 苣木禎史, 宇佐川毅, “頭部回転制御による FDBM のデータベース更新に関する検討”，日本音響学会講演論文集（春），CD-ROM，2008

6．研究組織

(1)研究代表者

中島　栄俊（NAKASHIMA HIDETOSHI）
熊本電波工業高等専門学校・電子制御工学科・准教授
研究者番号：00353350

(2)研究分担者

（　　）

研究者番号：

(3)連携研究者

（　　）

研究者番号：

(3)研究協力者

脇坂　龍（WAKISAKA RYO）
熊本電波工業高等専門学校・専攻科・制御情報システム工学専攻
研究者番号：なし