

平成 21 年 5 月 29 日現在

研究種目：基盤研究（C）
 研究期間：2006～2008
 課題番号：18500135
 研究課題名（和文） 周波数領域両耳聴モデルを利用した視聴覚連携による空間情報抽出システムの開発
 研究課題名（英文） Development of special information retrieving system based on the collaboration of visual and acoustical information using frequency domain binaural model
 研究代表者 宇佐川 毅（USAGAWA TSUYOSHI）
 熊本大学・大学院自然科学研究科・教授
 研究者番号：30160229

研究成果の概要：

これまで開発した周波数領域両耳聴モデルを発展させ、聴覚的空間情報の抽出機能を向上するとともに、視覚的空間情報と連携することで、両空間情報を補完的に統合することにより、これまでにない高精度な空間情報の抽出と、それに基づく高品質な情報抽出機能を実現することを目的として研究した。その結果、両耳聴モデルにより抽出した音響情報に基づく音源の方向情報と画像でのターゲット方向との連携が可能であることを示し、セキュリティカメラへの応用の可能性についてもその動作特性を含め明らかにした。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	1,700,000	0	1,700,000
2007年度	800,000	240,000	1,040,000
2008年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,600,000	570,000	4,170,000

研究分野：総合領域

科研費の分科・細目：知覚情報処理・知能ロボット

キーワード：情報センシング、両耳聴モデル、空間情報抽出、視聴覚連携

1. 研究開始当初の背景

音声および音響情報に含まれる空間情報は、視覚により得られる空間情報と連携することにより、極めて有用性の高い情報となる。例えば会議等で議論中、音声に関する空間的情報から、視野外の発話者の存在を検出し、視野の移動と発話音声の到来方向に関する情報を統合することで発言者を同定するという処理は、人間にとっては極めて日常的な動作である。しかしながら、遠隔会議システムにおける撮影・收音システムやヒューマンノ

イド・ロボットなどのセンサーシステムでは、聴覚により得られる空間情報と視覚による空間情報の有機的な連携が、十分にできているとは言えない。より高度なヒューマンインターフェースの構築や、知的情報処理のための高機能センサーの実現には、人間の持つ柔軟性を模倣しそれを発展させた形で利活用することは重要で、空間把握や方向情報に基づく情報の選択的抽出を実現するには、聴覚的空間情報と視覚的空間情報の連携は必要不可欠である。

従来、画像情報処理および音響情報処理分野では、それぞれ独立した形での空間情報の抽出やその利活用の研究が主であった。しかも、高臨場感情報通信のための視聴覚情報センサーや、ヒューマノイドロボットでの空間把握の分野での応用を想定した場合、視覚的空間情報の活用に比べ、聴覚的空間情報はまだ十分に利用されていないのが現状である。一方、人間は、空間情報として、視覚により得る情報が大きな意味を持つことは当然としても、聴覚により得られる情報を有機的に連携させることで、極めて柔軟で高い空間把握能力を発揮する。このことは、例えば、強制的に聴覚情報を遮断した状況で混雑した歩道の移動や車道の横断を想定した場合、視野外の情報を収集により多くの注意を払う必要が生じ、場合によっては危険や不安を感じるなどからも推測される。このことは、注意の向けられていない視野周辺や視野外からの情報を、聴覚における空間情報が相補的に補っていることを示している。

これまで、人間の両耳聴能機能をモデル化することで、複数の方向から同時に到来する信号をそれぞれの到来方向を推定するとともに、各々を分離抽出することが可能な周波数領域両耳聴モデル (Frequency Domain Binaural Model: FDBM) を開発してきた。この周波数領域両耳聴モデルを活用することで、人間のもつ両耳聴に基づく音源方向推定と画像情報とを連携させることが可能であると想定された。

2. 研究の目的

本研究は、この周波数両耳聴モデルのもつ聴覚的空間情報の抽出機能の精度を向上させるとともに、視覚的空間情報と連携させ空間的情報を補完的に統合することにより、これまでにない高精度な空間情報の抽出と、それに基づく高品質な情報抽出機能を実現することを目的とする。

3. 研究の方法

本研究は、(1) 視覚的空間情報と聴覚的空間情報の相補的連携、(2) 視覚領域での顔画像を中心とした空間情報抽出、(3) 聴覚的空間情報の抽出精度向上と分離音響信号の高品質化について3つのサブテーマからなり、全体を統合するサブテーマ(1)と、二つの基盤のサブテーマ(2)(3)という相互関係をもち、それぞれについて次のような方法で実施した。

(1) 視覚的空間情報と聴覚的空間情報の相補的連携

聴覚的空間情報と視覚的空間情報を統合するため、他の2つのサブテーマで得られた結果を踏まえ、統合した両耳聴モデルを計算機システム上に構築し、提案モデルの妥当性

の評価を行った。また、最終年度は、スタンダードアーロン型の実時間システムを模した単独のラップトップ PC 上にシステムを構築した。構築した視聴覚による空間情報の統合機能を有する両耳聴モデルを、

ヒューマノイドロボットの頭部としての視聴覚センサー機能

視聴覚情報を統合した監視システム

としての応用を想定した評価を行う。その際、については、実環境下での音声認識システムでの評価や、特定音源に対応する分離抽出音響信号の客観的品質評価を行った。また、については、遠隔地での監視者の動きに対応した形でセンサーを移動させる場合や、自動的に発音体を検出し、これに注目した画像抽出および音響信号抽出という視点からの基本的性能を評価した。

(2) 視覚領域での上半身を中心とした空間情報抽出

既存の画像処理アルゴリズムおよび色相に基づく顔エリア抽出アルゴリズムを活用して、視覚的空間情報における「音源候補」領域の抽出を行う。一方、人間以外の音源については、聴覚的な空間情報を画像情報に重畳させた形での情報を提供することとし、セキュリティ分野での利用を想定し、この段階での画像切りだしは行わないこととする。

(3) 聴覚的空間情報の抽出精度向上と分離音響信号の高品質化

聴覚的空間情報の抽出精度、特に残響場における抽出精度を向上させるために、“先行音効果”として知られる聴覚現象をモデル化を試み、最終的には両耳聴モデルに組み込むことを目指した。具体的には、周波数帯域毎の時間差分から、初期到来音に含まれる成分を抽出することで、これをモデル化することを試みた。さらに、特定方向からの音響信号を分離抽出した際の品質について客観的な指標について検討し、品質改善のための指標を明確にすることで、分離音響信号の品質の向上を目指した。

4. 研究成果

本研究では、すでに開発した周波数領域両耳聴モデルを発展させ、聴覚的空間情報の抽出機能を向上するとともに、視覚的空間情報と連携させることで、両空間情報を補完的に統合することにより、これまでにない高精度な空間情報の抽出と、それに基づく高品質な情報抽出機能を実現することを目的とし、研究を行い次のような成果を得た。

(1) 視覚的空間情報との有機的な連携について研究を進めた。特に視覚情報との連携を、単一のCPUで実時間システムをラップトップ上に実装した。

具体的に、セキュリティカメラを想定して、音響センサーによる音源方向推定を行うシステムを開発し、その性能評価を行った。図1に、実装したセキュリティカメラとカバーに装着したマイクロホンの配置を示す。図2は、2マイクロホンと3マイクロホンでの音源方向推定精度の違いについて検討した結果である。

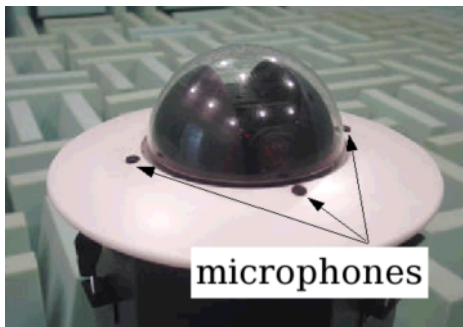


図1 カメラカバーに装着した3本のマイクロホン

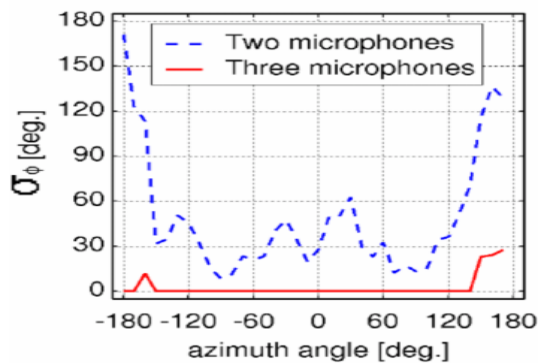


図2 マイクロホン2本(青の点線)と3本(赤の実線)を利用した場合の方向推定誤差。3本を利用すると、背面を除くほとんどの方向角で誤差が10度以下に収まる。

次に、エレベータホールを模擬した空間の天井面にセキュリティカメラを装着した場合についての、音源方向の推定精度について検討を行った。図3に、想定した環境ならびに音源等の配置を示し、図4に(-60°, +30°)に配置した音源の2次元での方向推定精度の検討結果を示す。

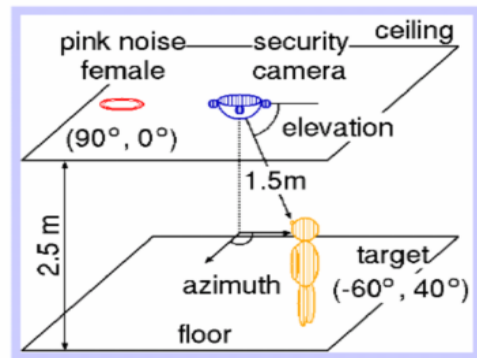


図3 エレベータホールでの実証実験の配置図：検出のターゲットは人間を模したフローア上の音源(-60°, +40°)から男声音声を提示、妨害音はPAシステムを想定して天井(90°, 0°)から提示。

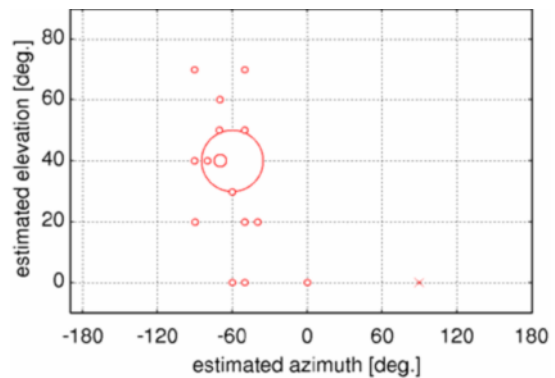


図4 音源方向の推定精度(SNR=20dB, 妨害音：女声音声の場合)観測フレーム数で69%が真の方向を推定。10°以内の誤差を許容すると81%。

(2) 視覚情報から複数の発話者が存在する場合についての顔領域の検出を試み、「音源候補」領域の抽出を実装した。また、人間以外の音源からの音響事象については画像上に音響信号の到来方向を画像に重畳して示す形で実装した。

カメラの回転運動にともなる空間情報の時間差分情報活用や、移動する音源に対して聴覚的空間情報と視覚的空間情報が相補的に統合された情報の時間変化を統合することによる検出精度の向上についても検討した。

(1)で実装した実時間処理による人物の追跡実験のビデオ映像の一部を、図5に示す。左の図は、実験風景の全体図であり、音響センサーを備えたセキュリティカメラは左下の黒いドーム内にあり、カメラ画像は奥のほうの人物を捕らえている。画像処理により、人物の上半身が認識されていることが、同図左のカメラ画像において赤いにより示さ

れている。デモビデオでは、カメラ画像内に存在しな別の人物（左図手間の人物）が発話することで、音響センサーが音源方向を推定し、カメラを自動的にその方向に向け、画像処理により人物の上半身を検出し、図6となる。



図5 ビデオ映像による人間の検出と3本のマイクロホンを利用した音源抽出を連携させた実験風景（ビデオ）の例：画角外の音源の追跡をさせたデモ



図6 手前の人物が発話したことで、カメラの方向が変化し、発話した人物を検出した状態

(3) 聴覚的空間情報の抽出精度を向上するために、先行音効果のモデル化を試みた。特に、衝撃性の音の立ち上がり検出を行うことで、ある程度残響の影響を低減できることを確かめた。しかしながら、人間のもつ残響抑制能力に比較すると、相当な開きがあり今後さらに研究を進める必要がある。

次に、方向を特定した音源からの音響信号を分離した場合の、分離信号の品質についても検討を行った。品質の検討に際しては、セキュリティカメラに限定せず、周波数領域両耳聴モデルとしての分離音声の品質を、従来のSNRやコヒーレンスという物理指標のみでなく、聴感特性を踏まえた音声品質尺度の一つであるPESQを用いて検討した。

その結果、周波数領域両耳聴モデルでの分離信号は、左右両チャンネルの品質を同一にするように機能することが明らかとなるとともに、PESQでの評価が聴取結果との対応が比較的よい事が、予備的な主観評価であるが確認された。

5. 主な発表論文等

〔雑誌論文〕(計4件)

1) Yoshifumi Chisaki, Tsuyoshi Usagawa, "Real-time processing using the frequency domain binaural model," *Applied Acoustics*, 査読有, 68(8), 923-938, 2007.2.20

2) Yoshifumi Chisaki, Sho Kwano, Kyoko Nagata, Kotaro Matsuo, Hidetoshi Nakashima, Tsuyoshi Usagawa, "Azimuthal and elevation localization of two sound sources using interaural phase and level differences," *Acoustical Science & Technology*, 査読有, 29(2), 139-148, 2008.3.1

3) Yoshifumi Chisaki, Ryoji Kawano, Tsuyoshi Usagawa, "On bit rate reduction of Inter-Channel Communication for a Binaural Hearing Assistance System," *IEICE Trans. Fundamentals*, 査読有, E91-A(8), 2041-2044, 2008.8.1

4) Yoshifumi Chisaki, Masahiro Naganishi, Tsuyoshi Usagawa, "Azimuthal and elevation localization using inter-channel phase and level differences for a hemispheric object," *IEICE Trans. Fundamentals*, 査読有, E91-A(10), 3059-3062, 2008.10.1

〔学会発表〕(計25件)

1) Tsuyoshi Usagawa, Minako Tomita, Sharifah Binti Saon, Yoshifumi Chisaki, "An Objective Evaluation of Segregated Speech Signal By Means of Frequency Domain Binaural Model," 日本音響学会聴覚研究会資料, 2008.12.6, 唐津

2) 河野 亮詞, 菅木 禎史, 宇佐川 毅, 両耳補聴システムのための周波数領域両耳聴モデルを用いた音源分離 - 時間-周波数フィルタリングによるミュージカルノイズの低減 - , 電子情報通信学会技術研究報告, 2008.11.21, 大分

3) 富田 美奈子, シャリファ サウン, 菅木 禎史, 宇佐川 毅, "周波数領域両耳聴モデルによる音源分離性能における定量評価 - 音源方向への依存性についての検討 -", 電子情報通信学会技術研究報告, 2008.11.21, 大分

4) Tomohisa Mashima, Takeshi Yosemoto,

Kousuke Matsuo, Anik Amalia, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Utterance Training System for Japanese Language -Experimental Results at ITS-", The 6th Kumamoto University Forum, 2008.11.5, スラバヤ市インドネシア

5) Taisuke Katayama, Masahiro Naganishi, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Microphone array system equipped on a security camera -Detection of sound event-", The 6th Kumamoto University Forum, 2008.11.5, スラバヤ市インドネシア

6) Masahiro Naganishi, Taisuke Katayama, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Estimation of sound source direction using a direct sound component extracted from three microphones mounted on a security camera cover," InterNoise 2008, 2008.10.27, 上海

7) Minako Tomita, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Quantitative evaluation of sound source segregation performance by means of the frequency domain binaural model," InterNoise 2008, 2008.10.28, 上海

8) 長西 将弘, 片山 泰輔, 菘木 禎史, 宇佐川 毅, 音源の方向推定機能を有する監視カメラシステムの構築に関する検討 - 設置環境での反射音による方向推定性能への影響 -, 日本音響学会 2008 年秋季研究発表会講演論文集, 2008.9.11, 福岡

9) 片山 泰輔, 長西 将弘, 菘木 禎史, 宇佐川 毅, 音源の方向推定機能を有する監視カメラシステムの構築に関する検討 - 残響場での観測信号の立ち上がり検出 -, 電気関係学会九州支部連合大会講演論文集, 2008.9.24, 大分

10) 河野 亮詞, 菘木 禎史, 宇佐川 毅, "両耳補聴システムにおける音声コーデックを用いた両耳間通信の検討", 電子情報通信学会技術研究報告, 2008.6.20, 本庄

11) 中島 栄俊, 脇坂 龍, 加茂田 浩史, 菘木 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 2008 年春季研究発表会, 2008.3.17, 千葉

12) 菘木 禎史, 長西 将弘, 宇佐川 毅, 周波数領域両耳聴モデルを利用した視聴覚連携

による空間情報抽出システム, 第 22 回信号処理シンポジウム電子情報通信学会, 2007.10.12, 仙台

13) 永田 京子, 今村 浩二郎, 中島 栄俊, 菘木 禎史, 宇佐川 毅, 周波数領域両耳聴モデルを用いた音源方向推定の精度に関する検討 - 正面ならびに側方における精度推定, 電子情報通信学会応用音響研究会, 2007.11.15, 熊本

14) 長西 将弘, 荒木 潤一, 菘木 禎史, 宇佐川 毅, 音響情報により捕捉した画角外音源の追跡機能を備えた監視カメラシステムの構築に関する検討, 電子情報通信学会応用音響研究会, 2007.11.15, 熊本

15) 荒木 潤一, 長西 将弘, 菘木 禎史, 宇佐川 毅, 視聴覚情報を用いた画角外音源に対応可能なカメラシステム, 日本音響学会九州支部学生のための研究発表会, 2007.10.27, 福岡

16) Masahiro Naganishi, Toshimichi Takada, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Development of an Automatic Measurement System of Object Related Transfer Function for Small Object," Japan-China Joint Conference on Acoustics 2007, 2007.6.4, 仙台

17) 菘木 禎史, 長西 将弘, 高田 俊亨, 神澤 龍市, 宇佐川 毅, 小型物体のための object related transfer function 自動測定, 電子情報通信学会応用音響研究会, 2007.6.29, 札幌

18) Tsuyoshi Usagawa, Toshimichi Takada and Yoshifumi Chisaki, "Sound source tracking for security camera using sensor array utilizing object-related transfer function," Proc. 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Vol.120, No.5-2, p.3178, 2006.11.28, Honolulu

19) Yoshifumi Chisaki, Sho Kawano and Tsuyoshi Usagawa, "Azimuth and elevation estimation for multiple sound sources using two-dimensional frequency domain binaural model," Proc. 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Vol.120, No.5-2, p.3217, 2006.11.28, Honolulu

20) 河野 翔, 中島 栄俊, 菘木 禎史, 宇佐川 毅, 周波数領域両耳聴モデルによる音源

の方位角および仰角の同時推定に関する検討，日本音響学会 2006 年秋季研究発表会講演論文集，pp.479-480，2006.9.15，金沢

21) 高田 俊亨，菘木 禎史，宇佐川 毅，物体の回折を用いた 3 素子アレイによる音源到来方向推定，日本音響学会 2006 年秋季研究発表会講演論文集，pp.533-534，2006.9.15，金沢

22) 長西 将弘，高田 俊亨，菘木 禎史，宇佐川 毅，物体の回折現象を用いた音源到来方向推定のための基礎的検討 -3 素子アレイを用いた特定方位角における仰角推定-，電気関係学会九州支部連合大会，p.50，2006.9.29，宮崎

23) 永田 京子，河野 翔，中島 栄俊，菘木 禎史，宇佐川 毅，周波数領域両耳聴モデルを用いた音源の方位角・仰角推定の精度に関する検討，電気関係学会九州支部連合大会，p.51，2006.9.29，宮崎

24) 河野 翔，中島 栄俊，菘木 禎史，宇佐川 毅，両耳間位相差およびレベル差を用いた複数音源の方位角・仰角推定手法の検討 -音源方向の 2 次元推定による周波数領域両耳聴モデルの拡張-，電子情報通信学会技術研究報告，Vol.106，No.205，EA2006-46，pp.25-30，2006.8.10，仙台

25) Toshimichi Takada, Sho Kawano, Yoshifumi Chisaki, Tsuyoshi Usagawa, "Sound source tracking for security camera based on object related transfer function," Proc. The 9th Western Pacific Acoustics Conference, No.657 (CDROM), 2006.6.26, Seoul

6. 研究組織

(1) 研究代表者

宇佐川 毅 (USAGAWA TSUYOSHI)
熊本大学・大学院自然科学研究科・教授
研究者番号：30160229

(2) 研究分担者

菘木 禎史 (CHISAKI YOSHIFUMI)
熊本大学・大学院自然科学研究科・助教
研究者番号：50284740

(3) 連携研究者

なし