

研究課題別評価

1 研究課題名:超分散マイク・スピーカーによる複数の音焦点形成

2 研究者氏名:加賀美 聡

3 研究の狙い:

本研究は広い範囲で動き回る人に離れたところから選択的に高品質に音を授受する手法を確立することを目的とする。対象とする人の頭部周辺に「スポット状」の高感度・高音圧分布を作り出して S/N 比の高い集音と音声伝送とを実現し、たとえその人が動いてもオンラインで追従させることが可能なシステムの実現を目指す。言わば「聞き耳をたてる」ような形で音声を集音し、「耳元で語りかける」ような形でその場所だけ音を聴かせる技術を確立する。

音の焦点を絞るためには、マイクロフォンアレイやスピーカーアレイを用いて特定の場所で位相が揃うように各素子を制御することにより望遠マイクや望遠スピーカーを構成可能である。電波望遠鏡や開口合成レーダーなども同様の原理に依っている。しかし可聴音の周波数帯域は 1000 倍オーダーで広がっていること、波長が長いために装置の規模が大きくなること、などからこれまでは S/N 比の高いシステムの実現が困難であった。音場シミュレーションでは十分な S/N 比を得るためには数十以上の素子が必要であるとの結論が得られた。そこで従来はアナログロジック回路やDSPにより構成された固定焦点の望遠マイクアレイなどの研究が行われている。

しかし近年、計算機が劇的に高速化し、リアルタイムOSのソフトウェア技術の進歩と相まって、例えば CD クオリティーの音を発生させる数十マイクロ秒の間に十分な入出力と計算能力を備えるようになってきた。また同時に広帯域のネットワーク機能をローカルに用いることができるようになった。このため本研究で提案する数百～千のオーダーのマイク・スピーカーネットワークを構築し、多焦点を形成することが可能になってきた。

このように距離の異なる素子からその場で位相を揃えることにより生成された焦点で聞こえる音は、球面上に広がる一般の音と異なり、その場だけで急激に音圧が上昇するため、その場所から離れるとほとんど聞こえない。そのために対象となる人の頭のなかに直接音がするような印象を与える。またマイクアレイでは、全ての入力を保存することにより、その環境内の全ての場所に「後から」焦点を当てることができるようになる。この手法を用いると三次元の環境+時間軸の四次元の音場を全て記録できることになり、例えば話者がどこにいるかと言った音源位置計測、あるいは任意の場所では音がどのように聞こえていたかを再現する任意の場所での音再生などが可能になる。

そこで本研究では 10 以上の焦点に対して S/N 比を 20dB 以上得るために数百オーダーのシステムを実現し、多焦点形成アルゴリズムの研究を行った。

4 研究成果:

以下の 3 つのサブテーマの研究開発を行った。

1) 多焦点形成・取得アルゴリズムの研究

- 2) 超分散マイク・スピーカーアレイシステムの設計・開発
- 3) 会議記録システムの開発

4.1 多焦点形成・取得アルゴリズムの研究

本研究では主たる方法として古くから知られている遅延和音焦点形成法(Delayed-Sum Beam Forming Method, DSBF)を用いている. この手法は各音素子の配置が既知であるとして, 対象となる位置から各音素子までの距離を同一となるように時間方向にずらすことにより, 対象位置で感度を高くするものである. 図1に手法の原理を示す. 図2は 4x4[m]の部屋の四面の壁面に合計 1024 個の音素子(スピーカーやマイク)を配置し, (3.0, 3.0, 3.0) の位置へ音焦点を形成したときの 1kHz での感度分布を示す. 音素子の間隔は 27cm である. この図で示すように, DSBF だけで直径約 50cm のエリアで, 周辺に対して 20dB の音圧比が得られている. このことから多数の音素子を準備し, 効果的に配置すれば高感度の音焦点が形成できることがわかる.

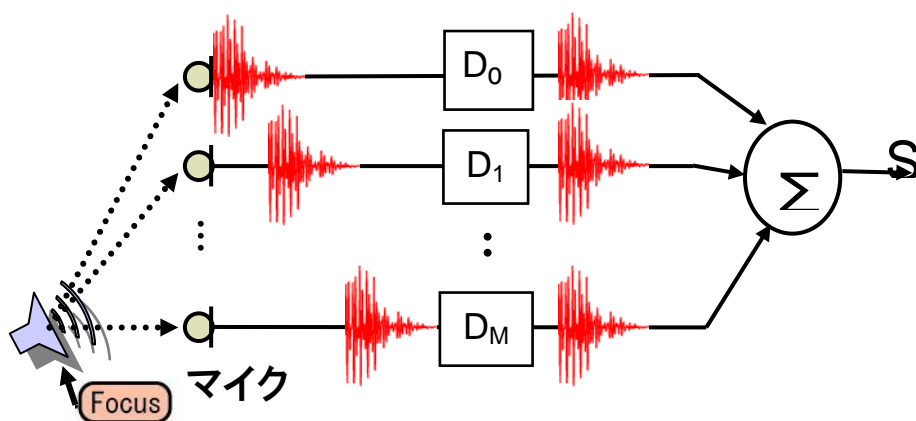


図 1 遅延和音焦点形成法(DSBF)の原理

つぎに人間の音声は周波数領域では比較的まばらであるという知見を利用し, 同時にはある狭いバンド幅(本研究では通常は 2Hz)では単一の音焦点だけがその音に関係していると仮定する. この性質を利用した手法として青木らにより提案された周波数帯域選択手法(Frequency Band Selection)がある. 本研究では多数の音素子から DSBF 法で任意の方角に望遠マイク(あるいはスピーカー)となることができる. そこでさまざまな方角に聞き耳をたて, 音焦点とされた場所での音を周波数領域で細かく見ることにより, 音焦点の音を決定する.

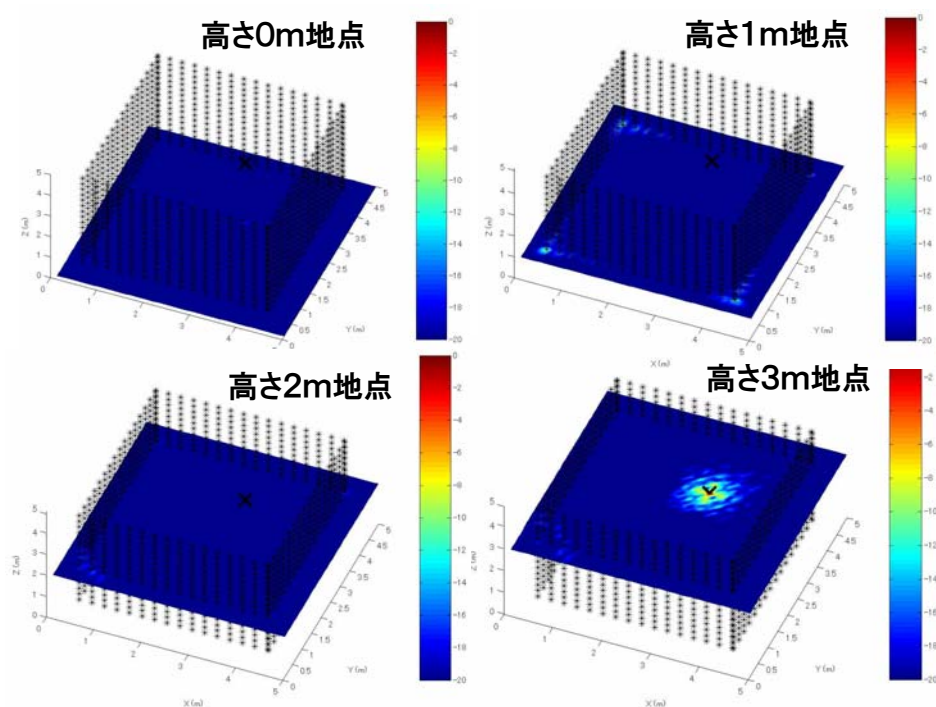


図 2 4面1024素子 1kHz での(3.0, 3.0, 3.0)への音焦点形成シミュレーション

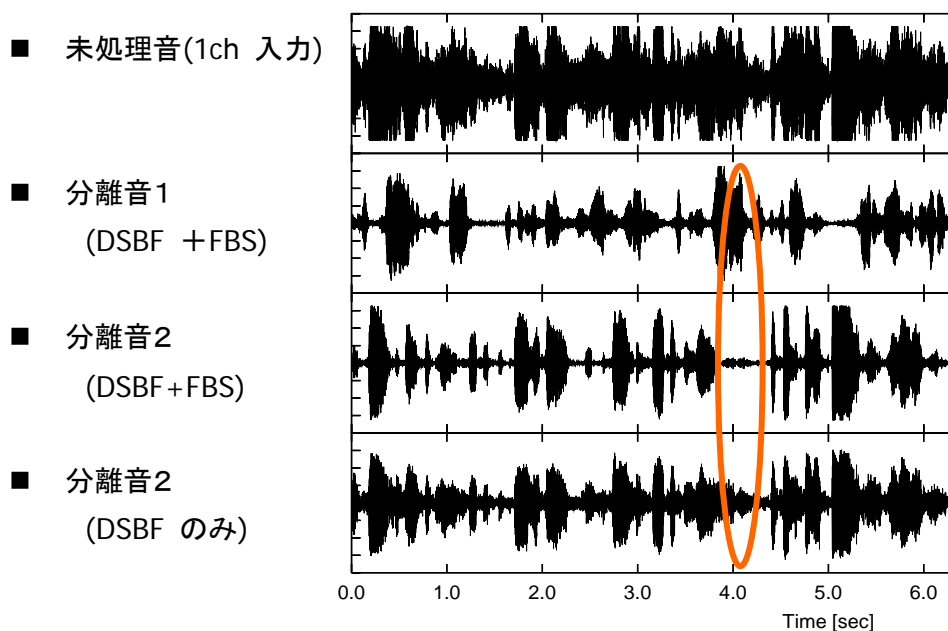


図 3 32ch マイクアレイによる音源分離における各手法の結果

図3に開発した 32ch マイクアレイを例にとって、これらの手法の結果を示す. 実験ではマイクアレイを中心に 1.5m 離れた3方向に、スピーカー(音楽), 男性話者, 女性話者が同時に音を出している際の音源分離結果を示す. このときシステムは音源の位置や数に関する前提知識は持っていない. 図

3の最上段は一つのマイクから得られた入力であるが、すべての音が混ざって聞こえることから、このような波形が得られている。2段目および3段目は提案手法(DSBF+FSB)により得られた音源分離の結果である。また4段目は DSBF のみを用いて音源2を分離したものであるが、丸く示した領域では音源1の音が混ざってしまっている。DSBF 手法は音源位置で音を強めあう手法であるが、それ以外の方向へ感度を持たないわけではないので、聞き耳を立てている方角以外から強い音が入ってくると、その音も重なってしまう。

4. 2 超分散マイク・スピーカーアレイシステムの設計・開発

前述のシミュレーション手法を用いて、音素子の最適配置を検討し、マイク、スピーカーアレイのシステムを構築した。はじめに効果を検証するために、図4左に示すような 128ch のスピーカーを用いて平面システムを構築した。シミュレーションによって得られた音圧分布を図4右に示す。このようにトンネル状の高音圧領域が形成される。この現象は実機でも精度よく再現されており、本研究の実現性が期待できることがわかった。そこで次に図5に示すような二面、四面のシステムを構築した。また密なシステムを空間に疎に配置するシステムを検証する目的で、図6に示すような実験ハウスの天井に 8ch のマイクシステムを合計 16 ユニット設置した。

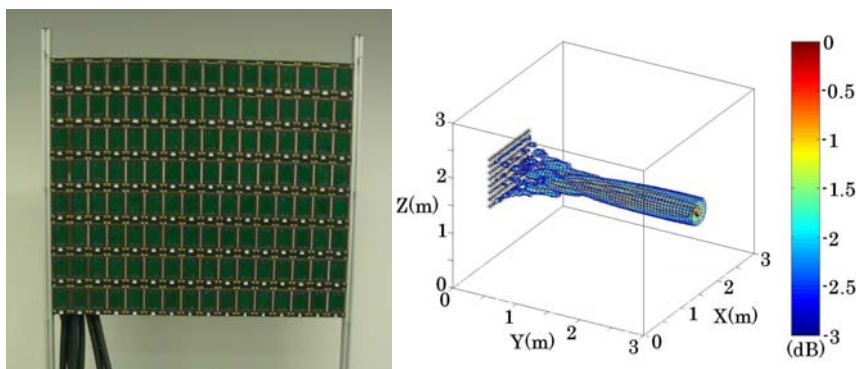


図 4 128ch 平面スピーカーシステム(左), 2kHz 時の音圧分布シミュレーション(右)



図 5 384ch 二面スピーカーシステム, 480ch 四面マイク・スピーカーシステム



図 6 天井設置型 8ch マイクシステム合計 16 ユニット

4.3 会議記録システムの開発

アルゴリズムの有効性を示すために、会議記録システムをアプリケーションとして、卓上やロボット搭載型のシステムを開発した。図7に開発したシステム Ver1, Ver2 を示す。また図8に最終成果である Ver3 を示す。図8の下段に 1, 1.4, 2kHz 時の感度分布シミュレーションを示す。サイドローブのない望遠マイクが実現できている。

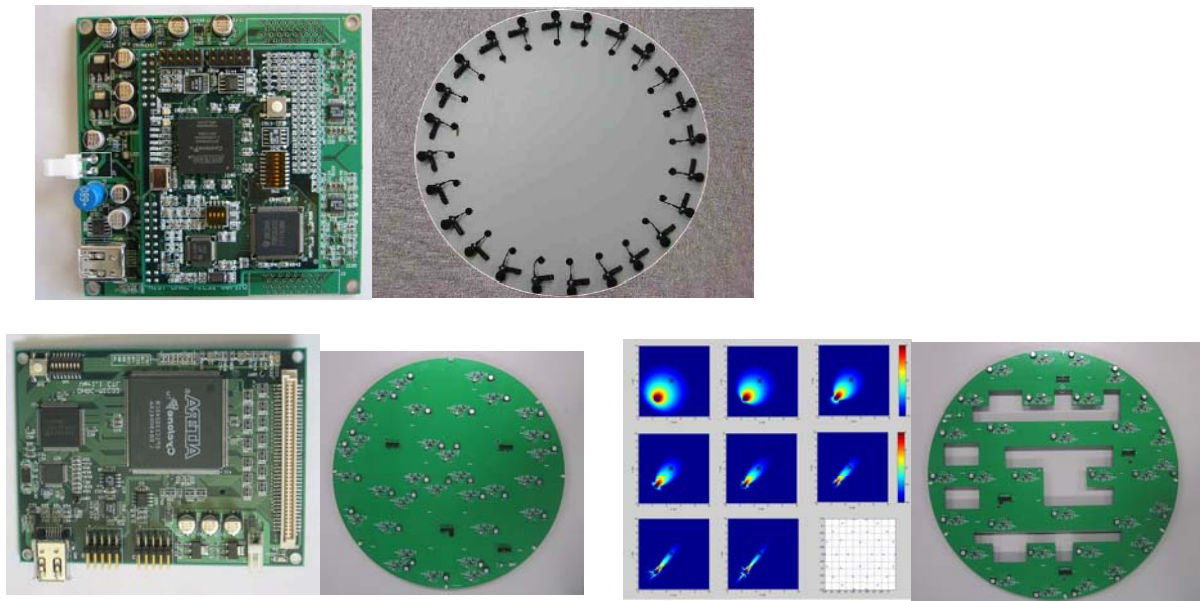


図 7 32ch マイクアレイ Ver.1(上段), Ver.2(下段)

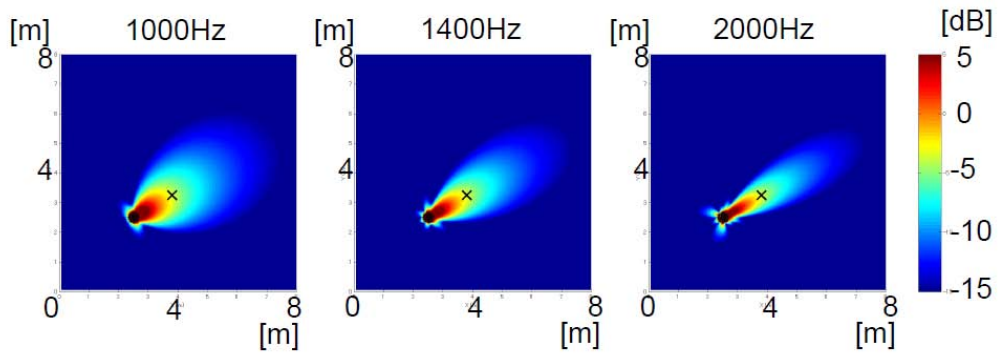
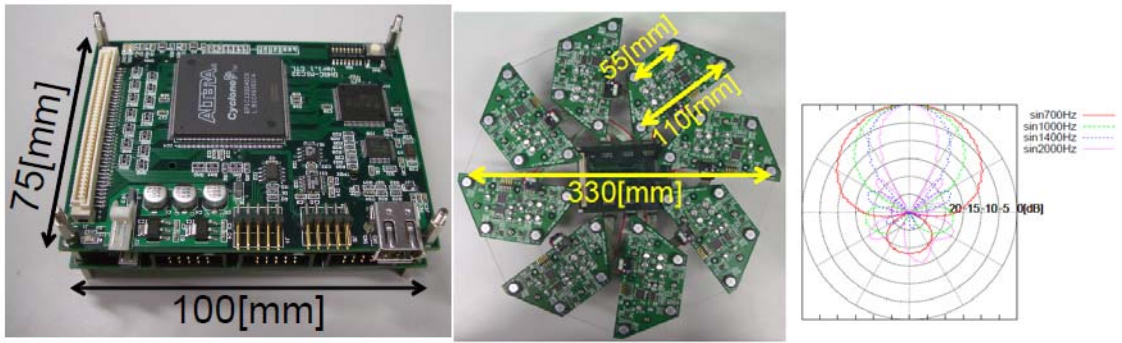


図 8 32ch マイクアレイ Ver.3

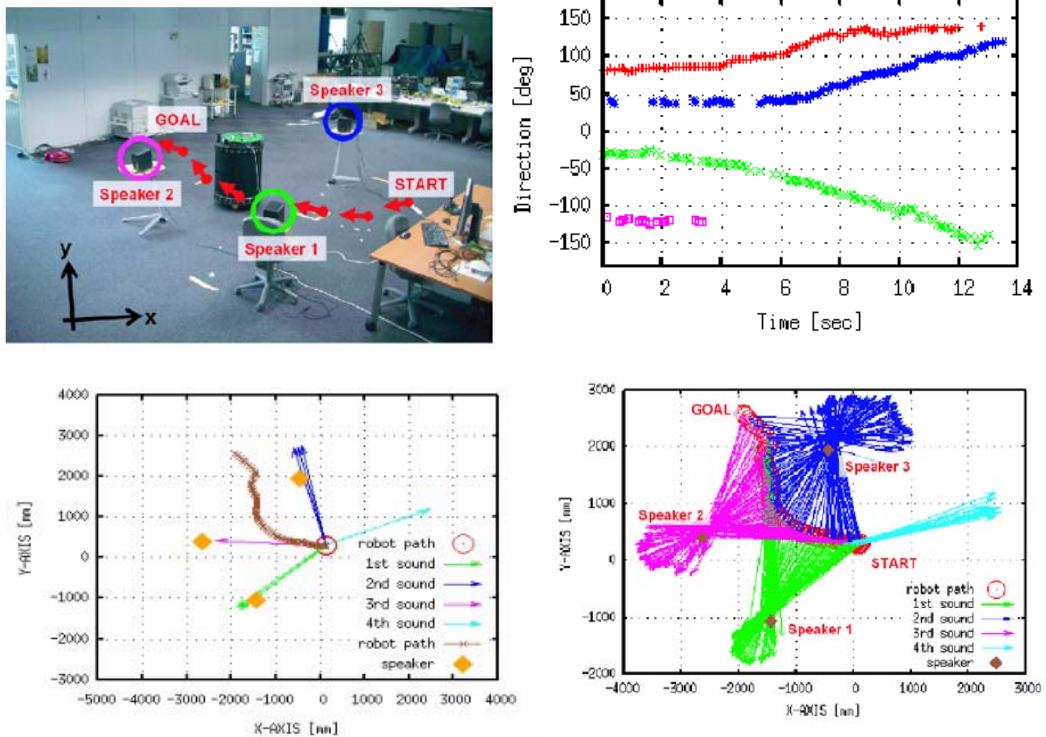


図 9 移動中の音源定位, 音源地図の作成

開発したシステムを用いて, 図9に示すように世界で始めて移動ロボットで移動中に音源を定位, 分離することにより, 音源の地図を作成することに成功した. 実験の配置を左上に, 定位した音源の

方向を右上に示す。ロボットの移動はデッドレコニングで知り、そこに短時間の移動ごとに定位した方角から三角測量することにより(左下)、音源位置を推定し、それらの結果を RANSAC により統計的に推定し結果、右下の図に示すように音源位置をロバストに推定できた。

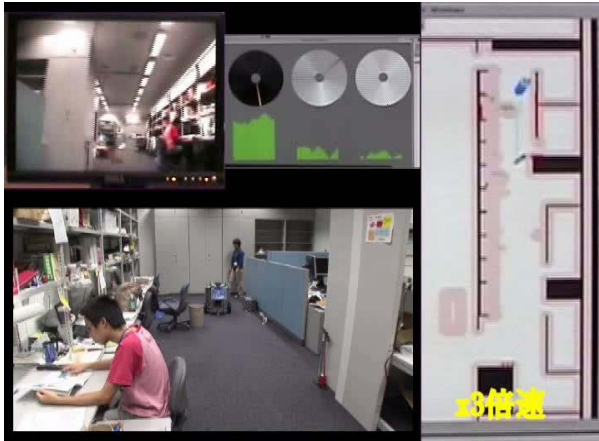


図 10 移動ロボットの遠隔操作における音源定位インターフェース

また図10に示すように移動ロボットを遠隔操縦するタスクにおいて、開発したマイクアレイを搭載し、操作者に音源方向を提示することにより360度方向の情報を容易に知ることができるシステムを開発した。

4.4 今後の展望

多数の場所に音を発生させ、逆に集音することが出来れば、例えば会議において話者の発言を口元にマイクがあるように高感度に記録したり、同時通訳のソフトウェアを通して特定の人に変換された言葉を伝達したりすることが可能になる。また離れた二人の人があたかも隣通しにいるかのように声を張り上げずに会話をすることや、特定の人にだけプロンプタのように情報を伝達することが可能になる。コンファレンス会場や大教室においてフロアからの質問をそこに焦点を形成することにより増幅することも可能になる。

音を発生させる技術は、例えば家庭における臨場感に溢れる劇場システムを、周りには聞こえない状態で構築を可能にしたり、バーチャルなヘッドホンを実現することを可能にする。また美術館などの展示において展示物の場所でのみ説明を聞かせるシステムを構築できる。公共の場において不審な動きや危険な行動を取っている人に対しても、現場に行かずにその場遠くから注意や警告を与えることができる。

また離れた場所の音を集音する機能は、ロボットが人の話し声や足音などからその位置を特定する人を検出する機能に用いることが出来る。セキュリティ上の意義も大きい。警備室のモニタ画面で異常を検知したら、現場に行かずにまずその場所の音を聴いてみる事が出来る。また演劇や舞台においてマイクを装着する必要がなくなる。また例えばコンサートを分散マイクで記録することにより、

あたかも演奏家の中にいるかのような音場を再現するなど、四次元的な音場を記録することが可能になる。そのほか、現状では音声認識システムは口元にマイクを設置しなければ認識率を上げられないが、本システムにより仮想的に口元にマイクがあるように音声を取得することが出来るようになる。

5 自己評価:

本研究により、環境中に多数配置した音素子の位相を制御し、環境中の複数の場所に焦点を合わせて音を取得したり、複数の焦点に音を発生させたりする多焦点形成について、アルゴリズム、システム構成法、アプリケーションの3つの側面から研究を行い、成果を得た。提案した位相差の重ね合わせ、周波数帯域選択手法により、効果的に音焦点が形成できることが、シミュレーションからも実際に構築したシステムからも確認できた。これにより当初目標としてきた音圧比 20dB を達成し、複数の人がしゃべっている状態で、それぞれの人の位置と発話内容を取り出す、といったことが実現された。このようなシステムはマイク・スピーカー両面で応用分野が広く、さまざまなアプリケーションが期待される。スピーカーやマイクといった音素子、およびシステムの構成要素そのものは低価格であるために近い将来の実用化を期している。

一方、研究を進めるうちに、音の波長のダイナミックレンジが広いために、周波数に依存して形成される焦点の形状が異なることがわかってきた。そのために特に複数の音源が近いときに、周波数によって分離される音に影響を与えてしまうという知見が得られた。この点に関しては、より精密な焦点形成手法、人間の知覚のモデル化の二点からあらたに研究を進めている。

6 研究総括の見解:

千個程度配置した音素子の位相を制御し、複数の場所に焦点を合わせて音を取得したり、複数の焦点に音を発生させたりする多焦点形成の手法について研究し、その実現のために種々の技術を提案し、実環境で実証する成果を挙げた。このシステムは特定の人にだけに音情報を提供する、特定の話者の音声のみを記録する、などが可能になり家庭、会議の場、美術館など多くの人が集まるさまざまな場面においていろいろな応用が考えられるものと思われる。システムの完成度が高く、また装置が比較的 low 価格で構成できると思われるので早期の実用化を期待している。

7 主な論文等:

論文

(1) **Virtual Earphone: Integration of Beam Forming by Speaker Array and Real-time Visual Face Tracking**

H. Mizoguchi, T. Kanamori, S. Kagami, K. Hiraoka, M. Tanaka, T. Shigehara

Key Engineering Material, Vol. 243-244, pp. 117-122, 2003

(2) **Microphone Array for 2D Sound Localization and Capture**

Satoshi Kagami, Hiroshi Mizoguchi, Yuuki Tamai, Takeo Kanade

Proceedings of the 2004 IEEE International Conference on Robotics & Automation(ICRA04), pp. 703--708. , Apr. , 2004

(3) **Detecting and Segmenting Sound Sources by Using Microphone Array**

Satoshi Kagami, Yuki Tamai, Hiroshi Mizoguchi, Koichi Nishiwaki, Hirochika Inoue

Proceedings of 2004 IEEE-RAS/RSJ International Conference on Humanoid Robots(Humanoids2004), pp. 67paper(CD-ROM). , Nov. , 2004

(4) **Home Robot Service by Ceiling Ultrasonic Locator and Microphone Array**

Satoshi Kagami, Simon Thompson, Yoshifumi Nishida, Tadashi Enomoto, Toshihiro Matsui

Proceedings of 2006 IEEE International Conference on Robotics and Automation (ICRA2006) , Orlando, Florida, pp. 3171--3176. , May. , 2006

(5) **Combining Ubiquitous and On-Board Audio Sensing for Human-Robot Interaction**

Simon Thompson, Satoshi Kagami, Yoko Sasaki, Yoshifumi Nishida, Hiroshi Mizoguchi, Tadashi Enomoto.

Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, March 2007 (to appear).

受賞

- (1) 「ヒューマノイドの知的行動の研究」により 平成 18 年度文部科学大臣表彰・若手科学者賞を受賞.
2006 年 4 月.

招待講演

(1) **バイオメカニクスで進歩するヒューマノイドロボット**

加賀美聡

日本バイオメカニクス学会 特別講演, 鹿屋, 2004 年 9 月.

(2) **Autonomous and Interactive Behaviors of Humanoid Robots**

Satoshi Kagami

Fifth Japan-America Frontiers of Engineering Symposium, San-Jose, USA, November, 2005.