

マイクアレイネットワークを用いた宅内サービス実現可能性の検討

祖田 心平[†] 梶本 真佑[†] 中村 匡秀[†]

和泉慎太郎[†] 川口 博[†] 吉本 雅彦[†]

[†] 神戸大学 〒 657-8531 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]soda@ws.cs.kobe-u.ac.jp, ^{††}{shinsuke,masa-n,shin,yosimoto}@cs.kobe-u.ac.jp,

^{†††}kawapy@godzilla.kobe-u.ac.jp

あらまし ユビキタス技術の進化に伴い、家庭内の設備をネットワークに接続したホームネットワークシステム (HNS) の研究開発が進んでいる。住人の位置に応じて適切なサービスを提供する位置アウェアなサービスは、HNS における重要課題の一つである。従来、複数のマイクアレイをネットワークで接続し、話者の位置推定等に応用する研究が行われている。本研究では、このマイクアレイネットワークの宅内位置アウェアサービスへの適用可能性を、3 つのサービス例を挙げて検討する。

キーワード ホームネットワークシステム、マイクアレイネットワーク、位置アウェア、宅内サービス

A Feasibility Study of Home Services Using a Microphone Array Network

Shimpei SODA[†], Shinsuke MATSUMOTO[†], Masahide NAKAMURA[†],

Shintaro IZUMI[†], Hiroshi KAWAGUCHI[†], and Masahiko YOSHIMOTO[†]

[†] Kobe University Rokkoudai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8531 Japan

E-mail: [†]soda@ws.cs.kobe-u.ac.jp, ^{††}{shinsuke,masa-n,shin,yosimoto}@cs.kobe-u.ac.jp,

^{†††}kawapy@godzilla.kobe-u.ac.jp

Abstract The home network system (HNS), which provides value-added services by orchestrating networked home appliances, equipments and sensors, attracts great attention to realize the next-generation smart home. Implementing the location-aware services within the the HNS context is one of greatest challenges, where the appropriate services are performed based according to the location of the inhabitants. We have been studying the technologies of sound source localization and sound source separation using networked multiple microphone arrays. In this paper, we conduct a feasibility study of applying the microphone array network to the location-aware services within the HNS. Specifically, we first present three kinds of home services using illustrative examples. We then enumerate three kinds of requirements (accuracy requirement, installation requirement, user requirement), which are essential for implementing the location-aware home services using the microphone array network. In a preliminary experiment, we evaluate the accuracy requirement using an actual microphone array. Moreover, we conduct a directivity shape simulation assuming multiple arrays.

Key words home network system, microphone array network, location-aware, home services

1. はじめに

ユーザにマイクの位置を意識させないユビキタスな音声インタフェースとして、複数のマイクロホンと並べたマイクアレイをネットワークで接続し協調動作させるマイクアレイネットワーク [1] の研究が進んでいる。マイクアレイとは、複数のマイクロホン間の音の到達時間差から音源の方向や位置を推定す

ることができるデバイスである。このマイクアレイを複数協調動作させることにより、より高精度な音源方向の推定が可能となる。また、各マイクアレイが取得した音声を集約することでサーバまでの伝送量を削減し、信号対雑音比 (SNR) を改善しつつマイクアレイの大規模化に付随する問題を解決できるといったメリットも持つ。しかし、従来のマイクアレイ研究の多くは、実機を用いた動作検証 [2] や伝送量削減のための改善 [3]

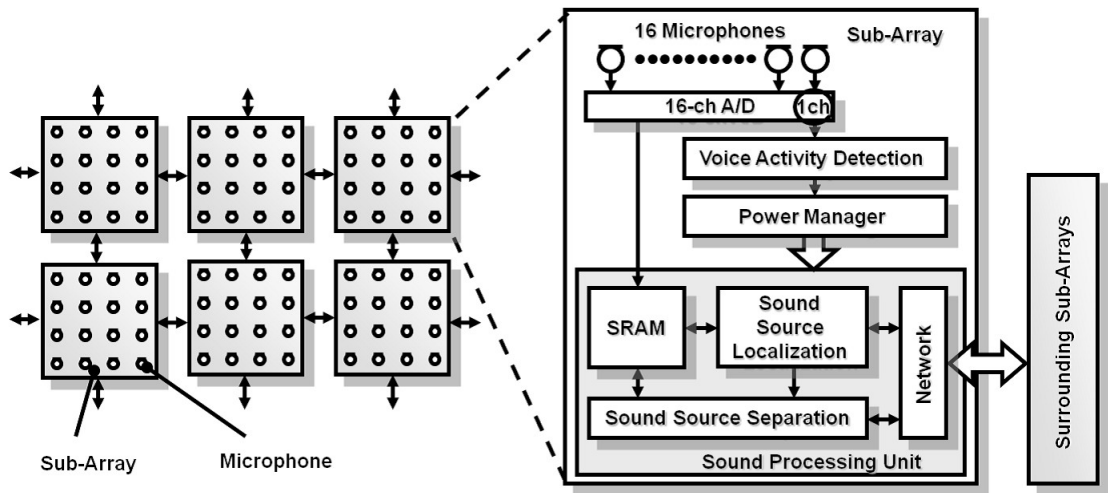


図1 マイクレイネットワーク

といった基礎研究であり、実用的なサービスへの適用や応用に関する研究は少ない。

一方、我々の研究グループでは、宅内の家電機器やセンサをネットワークに接続し、様々な付加価値サービスを実現するホームネットワークシステム（以下 HNS）の研究・開発を行っている。この HNS に代表されるような次世代の宅内環境では、音声による家電操作 [4] や、外部環境のコンテキストに応じた家電機器の自律操作 [5] などの付加価値サービスが提供されることが考えられる。

本研究ではマイクレイネットワークの応用方法の一つとして、HNS などの次世代スマートホーム内での位置アウェアサービスへの適用可能性について検討する。位置アウェアサービスは、発話者の位置や方向を外部コンテキストの一つとして利用することで、より便利で快適な付加価値サービスを提供する。具体的には、話者の方向から「誰による」音声機器操作かを推定し、そのユーザの好みに応じた機器操作を自動的に行うといったサービスが考えられる。従来の位置アウェアサービスの多くは、GPS や WiFi 信号による屋外での位置推定を利用したものである [6] が、マイクレイネットワークを利用することで、宅内やオフィス内といった狭い環境下での高精度な位置推定が期待できる。また、ユーザは発信器等のデバイスを持つ必要が無く、低負荷でコピキタスな位置アウェアサービスが実現可能である。

以降では、3 つの具体的な位置アウェアサービスを題材とし、各サービスが要求するマイクレイネットワークへの精度要件と設置要件、およびユーザ要件について整理する。さらに、実際の HNS 上にマイクレイネットワークを設置した際の音源方向推定についての精度評価実験に加え、複数のマイクレイを用いた場合の指向性形状の変化のシミュレーションを行う。最後に、これらの実験により得られた結果に基づいて、3 つの位置アウェアサービスの実現可能性について議論する。

2. 準備

2.1 マイクレイネットワーク

マイクレイとは複数のマイクロホンの音到達時間差を利用することで、音の到来方向の推定や指向性の制御を行うことができるデバイスのことである。また方向の推定のみならず、反射や残響の影響を抑制し、特定の音声を分離・抽出することにより、信号対雑音比（SNR）の改善も可能である。これらの効果は、使用するマイクの数が多くなるほど向上することが知られている。しかし、取得した生情報をサーバへ送る際の伝送量やマイクレイ全体の電力消費が増大するという理由から、実用化されているマイクレイのマイク数は数個から十数個に留まっている。

図1にマイクレイネットワークの構成図を示す。マイクレイネットワークでは、16chの小規模なマイクレイ（サブアレイ）を多数配置し、サブアレイ間で有線ネットワークを構築することによって大規模なマイクレイ音声処理システムを実現する。各サブアレイが行う処理は主に以下の3つである。

- (1) 発話の有無を検出する発話推定
- (2) 音源の位置を推定する音源定位
- (3) 特定位置の音を抽出する音源分離

これらの処理の結果として得られた SNR の高い音声データをネットワーク上で集約することで、さらに SNR が高められた音声データが得られる。

2.2 ホームネットワークシステム

ホームネットワークシステムは、照明やテレビなど家庭内における様々な家電機器と、温度計や湿度計などセンサをネットワークに接続することで構築される。HNS 内における機器はユーザや外部エージェントがネットワーク越しに制御できるように、制御 API を備えている。HNS は次世代のスマートホーム技術として期待されており、音声による機器操作インタフェースや、外部環境やユーザの状況といったコンテキストに応じた機器の自律制御などの付加価値サービスを提供可能である。我々の研究室で開発している CS27-HNS [7] は、サービス

指向アーキテクチャ(SOA) [8] の考えを取り入れ、すべての制御 API を、機種や実行環境に依存しない標準的な Web サービスとして公開している。

3. マイクアレイネットワークを用いた宅内サービス

本研究では、マイクアレイネットワークと HNS を組み合わせた宅内サービスとして、以下の 3 つのサービスを考える。

- ・コンテキスト推定サービス
- ・ヒューマンインタフェースサービス
- ・ライフログサービス

本章では、これらの 3 つのサービスの内容を具体的に説明する。

3.1 コンテキスト推定サービス

マイクアレイネットワークの発話推定や話者位置推定の機能は、オフィスや会議室といった多人数が集まる空間において、議論の盛り上がり度や人と人との繋がり等のコンテキストを推定するサービスに応用できる。

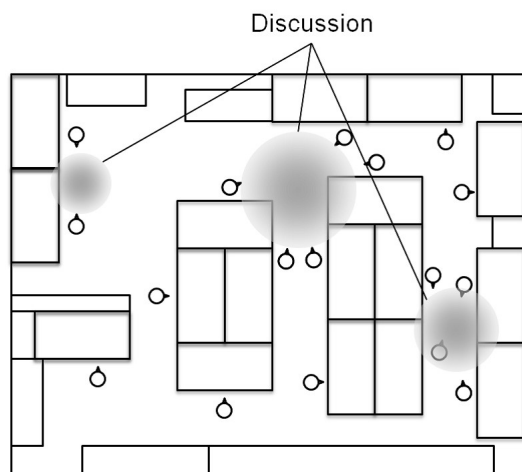


図 2 議論盛り上がり検出サービス

例えば、議論の盛り上がりを自動検出するサービスを考える。空間内で話が盛り上がっている所があった場合、マイクアレイネットワークはその方向に音源が存在すると判別する。オフィスや会議室など空間内におけるユーザの座標がある程度固定されている環境であれば、音源の位置から誰と誰が頻繁に会話しているかといったことも推定できる (図 2)。マイクアレイネットワークは複数音源を推定することもできるので、盛り上がり箇所が複数あっても対応可能である。ユーザの位置が固定されていない環境で人同士の繋がりを判別するには、音声情報から誰の声か推定する処理が必要となる。

3.2 ヒューマンインタフェースサービス

音声インタフェースでの音声利用では、単一マイクを口元に近付けて集音する手法が一般的だが、日常生活の中では常にマイクに向かって話す煩わしさが問題となる。マイクアレイネットワークを用いることで、マイクが存在を意識しないユビキタスな音声インタフェースが実現する。

従来、音声入力を用いた HNS の統合操作インタフェースの

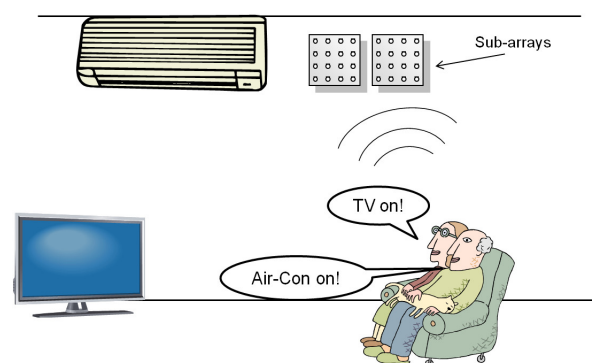


図 3 ボイスリモコンサービス

研究が行われている [4]。この操作インタフェースを用いたボイスリモコンサービスをサービス例として考える。マイクアレイネットワークでは、推定した音源の位置に指向性を向け、周りの雑音を除去して SNR の高い音声を集音する。よって、リビングや寝室だけでなくキッチンなどの雑音環境下でも音声だけで HNS 上の家電製品を一括で操作できる (図 3)。

3.3 ライフログサービス

Sound	Date	Location
	2011/9/20/15:20:45	(320, 256, 55)
	2011/9/20/15:21:03	(238, 448, 129)
	2011/9/20/15:21:54	(171, 66, 60)
	2011/9/20/15:22:36	(194, 71, 95)
	2011/9/20/15:22:58	(101, 321, 57)
	2011/9/20/15:23:42	(250, 227, 133)
	2011/9/20/15:24:04	(97, 146, 87)

(a)

(b)

図 4 (b) は、カメラとスピークの連携を示しています。カメラは円形で、スピークは長方形で、両者は互いに接続されています。

図 4 ライフログサービス (a) 発言記録, (b) カメラとの連携

マイクアレイネットワークで取得した音声と位置情報をライフログとして蓄積していけば、いつ、どこで、何をしゃべったかといった発言記録を作成できる (図 4(a))。Web カメラなどと連携すれば、推定した音源方向に自動でカメラを向けることで、発言者の音声と映像を繋げた議事録が作成される (図 4(b))。録音された議事録を見返すことで、参加者の発言回数や発言内容から意欲評価にも繋がられる。マイクアレイネットワークではサーバに送られる音声は高々 1ch だが、複数 ch を録音することができれば、後から指定した座標付近の音声を抽出できる。これは、監視カメラの映像から怪しい箇所を特定し、音声だけを抽出するといった防犯サービスにも有効である。音声データを常に蓄積していくとデータ量が膨大になるという問題が生じるが、マイクアレイネットワークの発話推定機能を用いて録音期間を制御することで、無駄なデータ取得を抑えることも可能である。

4. 宅内サービス実現のためのマイクアレイネットワーク要件

本章では、マイクアレイネットワークを用いたサービスを

考える時に必要となる要件を定義し、3章で挙げたそれぞれのサービスの実現可能性を検討する。

4.1 品質属性

マイクアレイネットワークを用いたサービスの品質を決める要件として、精度要件、設置要件、ユーザ要件の3つの要件を考える。精度要件は、サービスを実現するために必要な話者位置推定の精度や音源分離の性能を示す。設置要件は、サービスを使用する環境やサブアレイを設置する数といった制限を示す。ユーザ要件は、サービスを使用するユーザの人数や求められる行動等を示す。

4.2 コンテキスト推定サービスに対する要件

(精度要件) 盛り上がっている箇所を一つの音源とみなせば、各ユーザの座標を推定できるほどの精度が無くてもサービスを実現可能である。ただし、盛り上がっている箇所は1箇所とは限らないので、複数の音源を同時に推定しなければならない。また、このサービスは音源の位置のみを利用するため、音源分離の精度は要求されない。

(設置要件) サービスを使用する場所はオフィスや会議室が想定される。部屋の至る所で会話が起これるため、部屋全体をカバーできる数のサブアレイを設置する必要がある。

(ユーザ要件) 話の盛り上がり方を推定するため、空間内に多人数がいることが前提となる。また、人同士の繋がりを推定する際、音源の位置から誰が話しているかを推定するため、空間内でのユーザの動きが少ない環境が適している。

4.3 ヒューマンインタフェースに対する要件

(精度要件) 取得した音声を認識に使用するため、高精度な音源分離や残響除去等の処理が要求される。音源分離で高い精度を得るためには音源の位置に正確に指向性を向ける必要があるため、話者位置推定の精度も必要である。

(設置要件) このサービスは、HNSが実装されたスマートホームでの使用を想定している。一般的に、ユーザが家電製品を操作しようとする場所は、玄関やテレビの前等ある程度限定されるため、すべての範囲をカバーする必要は無い。

(ユーザ要件) 音声インタフェースと連携するため、同時にサービスを使用するのは高々1人である。ユーザの行動制限は特に無い。

4.4 ライフログサービスに対する要件

(精度要件) 蓄積するログのデータ容量を削減するには、発話の有無を検知して録音システムを制御しなければならない。音声会議システムでカメラと連携する場合は、話者の方向に正確にカメラを向けるために高い方向推定の精度が必要となる。取得した音声情報の音質は、解析等に使用しないのであれば人が聞き取れる程度で問題無い。

(設置要件) このサービスは、家庭内、オフィス、会議室等场所を問わず使用できるが、全ての会話を記録するには、部屋全体をカバーしなければならない。座標を指定して音声抽出を行うには、複数chを録音するために各サブアレイが音声を蓄積するストレージを持つ等の改良が必要となる。

(ユーザ要件) 同時に使用できるユーザ数に制限はないが、ユーザごとにログをクラスタリングするには音声や位置情報が

らユーザを特定する必要がある。カメラとの連携を行う場合はユーザの動きは少ないほうが良い。

5. 評価実験・シミュレーション

本章では、精度要件調査の一環として行った単一サブアレイを用いた3次元音源方向推定の精度評価実験と、設置要件調査の一環として行った複数サブアレイを用いた場合の指向性形状シミュレーションについて述べる。

5.1 精度評価実験

部屋の天井に16chサブアレイを釣り下げ、床から1mの位置で500Hzのsin波を再生する。30回推定した結果(θ , ϕ)の平均値と実際の角度を比較する。サブアレイを設置する場所はCS27研究室の天井(高さ2.44m)、音源の位置は $-50 \leq (\theta, \phi) \leq 50$ [deg]の計41箇所とする(図5)。

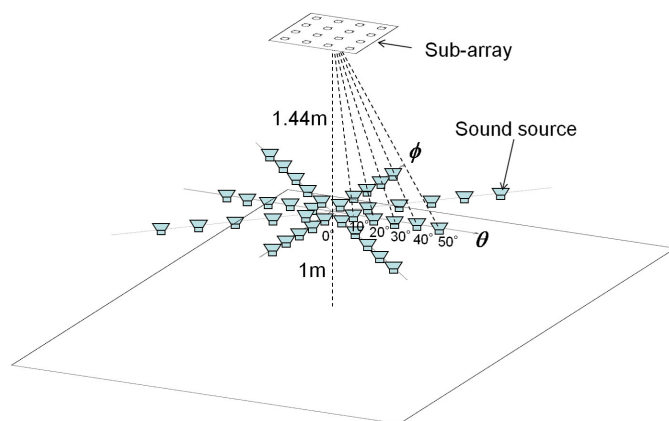


図5 精度評価実験

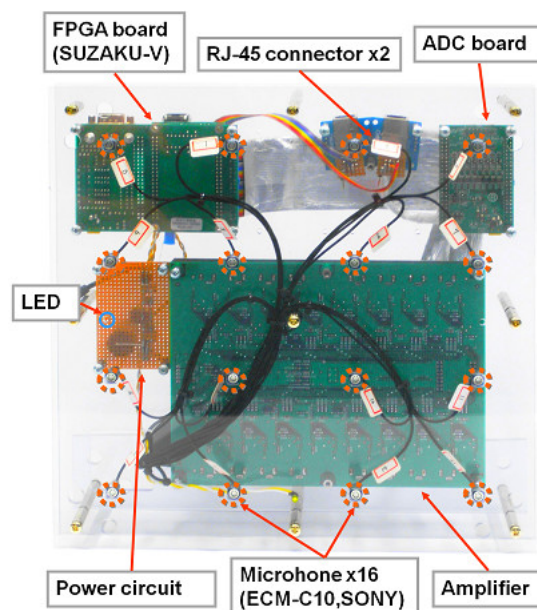


図6 サブアレイ

図6に使用したサブアレイを示す。16個のマイクは格子状に配置されており、内側の4chを方向推定に使用する。音声信号はアンプとADボードを通してFPGA内に実装された音源方

向推定モジュールに入力される．方向推定は 10° 刻みで行い，アルゴリズムには MUSIC 法 [9] を用いる．一度の推定にかかる時間は約 1.5 秒である．

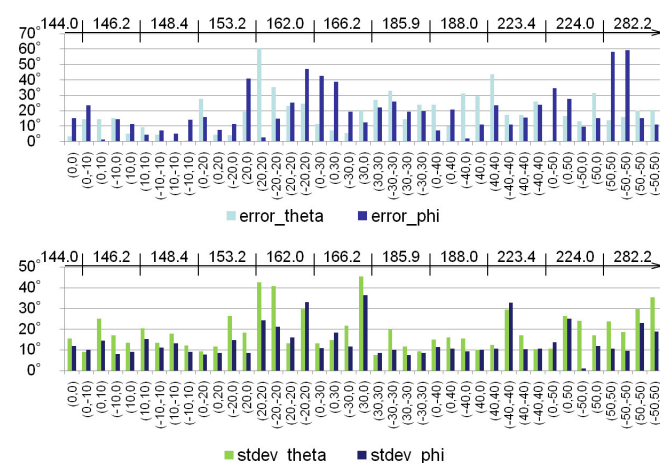


図 7 実験結果

図 7 に実験結果を示す．上のグラフが各座標 (θ , ϕ) 毎の誤差，下のグラフが標準偏差を表している．横軸はサブアレイと音源の距離 [cm] で，座標は左からサブアレイに近い順で並んでいる．グラフより，サブアレイから 148.4cm 以内の範囲は (0, -10) を除いて 20° 以内の誤差であることがわかる．153.2cm 以上では場所によって誤差が大きく変化し， θ か ϕ のどちらかが大きくずれていることが多い．また，185.9 ~ 188.0cm の方が 162 ~ 166.2cm よりも誤差や標準偏差が小さいように，距離が離れるほど精度が落ちるといった傾向は見られない．これらの原因としては，残響音や反射音の影響が考えられる．FPGA の容量の都合上，除去のためのフィルター等は未実装で，現状では 3 章で挙げた 3 つのサービスを実現する精度は得られていない．一般家庭やオフィスでは，机やソファといった様々な家具で音声が反射されることが考えられるため，反射・残響除去フィルターの実装は精度向上には必要不可欠だろう．

5.2 指向性形状シミュレーション

複数の 16ch サブアレイを用いた場合の，配置によるの全体の指向性形状の変化をシミュレーションにより調査した．一般的に，男声の周波数が 300 ~ 550Hz，女声の周波数が 400 ~ 700Hz と言われているため，求める指向性は周波数が 300Hz，500Hz，700Hz の場合及び，300Hz から 1000Hz まで 100Hz 刻みで計算した際の平均値とする．

図 8 に 4 つのサブアレイを用いた場合の結果を示す．青丸がマイク，緑丸が音源の位置を示しており，すべてのサブアレイが音源方向に指向性を向けていると仮定する．(a),(b) は 4 つ全てを天井に配置した場合，(c),(d) は壁と天井に 2 つずつ配置した場合である．天井のサブアレイの間隔は (a),(c) が 50cm，(b),(d) が 100cm とする．

複数サブアレイが同時に指向性を形成した時，周波数に応じた干渉縞が発生し，音源付近以外にもゲインが下らない位置ができる．また，周波数が高くなるほど指向性が鋭くなり，音源定位の結果が少しでもずれると強調できなくなる．(a) ~ (d)

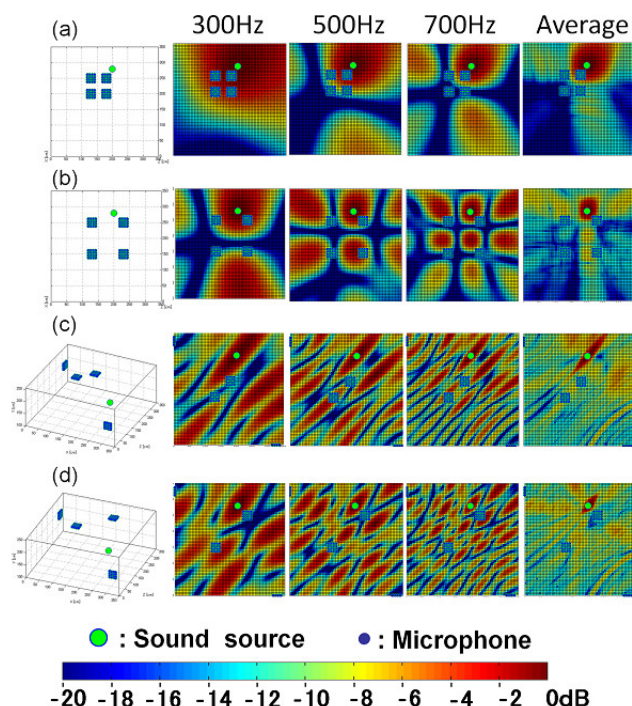


図 8 4 つのサブアレイを用いた場合の指向性形状

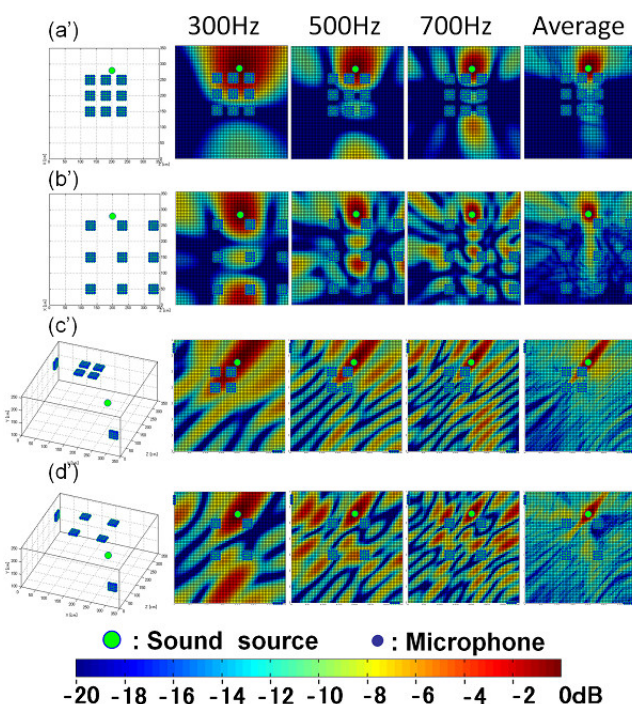


図 9 サブアレイを増やした場合の指向性形状

より，干渉縞はサブアレイの間隔が狭くなるほど粗くなり，指向性は鈍くなることがわかる．天井と壁に分けて配置した場合は干渉縞の形状が歪になり，音源付近以外のゲインが高い部分が増加している．

図 9 にサブアレイ数を増やした場合の指向性形状を示す．(a')(b') はそれぞれ (a)(b) についてサブアレイ数を 4 から 9 に増やした場合，(c')(d') はそれぞれ (c)(d) について天井のサブアレイ数を 2 から 4 に増やした場合である．サブアレイ数を増

やしたことにより、音源付近以外のゲインが全体的に下がっているのがわかる。

以上より、複数サブアレイを用いた場合に生じる指向性の干渉縞への対策としては、多数のサブアレイを密に配置することが有効であることがわかった。ヒューマンインタフェースサービスやライフログサービスといった音源分離が重要となるサービスでは、干渉縞の影響を少なくするためにこれらの対策が必要となる。

6. おわりに

本研究では、マイクアレイネットワークの応用方法の一つとして、HNS などの次世代スマートホーム内での位置アウェアサービスへの適用可能性について検討した。3 つの具体的な位置アウェアサービスに対して、各サービスが要求するマイクアレイネットワークへの精度要件と設置要件、およびユーザ要件について整理した。さらに、実際の HNS 上にマイクアレイネットワークを設置した際の音源方向推定についての精度評価実験に加え、複数のマイクアレイを用いた場合の指向性形状の変化のシミュレーションを行った。

7. 謝辞

本研究の一部は、株式会社半導体理工学研究センター (STARC)、科学技術研究費 (基盤研究 B 23300009, 若手研究 B 21700077, 研究活動スタート支援 22800042)、および、ひょうご科学技術協会の助成を受けて行われている。

文 献

- [1] T. Takagi, H. Noguchi, K. Kugata, M. Yoshimoto, and H. Kawaguchi, "Microphone array network for ubiquitous sound acquisition," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1474–1477, 2010.
- [2] 祖田心平, 久賀田耕史, 高木智也, 和泉慎太郎, 野口紘希, 吉本雅彦, 川口 博, "分散処理を用いた超低消費電力ネットワーク型マイクロホンアレイの研究," 日本音響学会 2010 年秋季研究発表会, pp.585–588, 2010.
- [3] S. Izumi, H. Noguchi, T. Takagi, K. Kugata, S.S. and M. Yoshimoto, and H. Kawaguchi, "Data aggregation protocol for multiple sound sources acquisition with microphone array network," 20th International Conference on Computer Communications and Networks (ICCCN), pp.1–6, 2011.
- [4] 松原典行, 江上公一, 井垣 宏, "暗黙的なユーザ要求を抽出・推定するホームネットワークのための対話型音声インターフェース," 電子情報通信学会技術研究報告, 第 109 巻, pp.61–66, 2010.
- [5] 松尾周平, 井垣 宏, 中村匡秀, "エンドユーザーによるセンサー駆動サービスの構築支援環境の提案," 電子情報通信学会 OIS 研究会, 第 OIS2008 巻, pp.043–048, 2009.
- [6] S. Burigat and L. Chittaro, "Location-aware visualization of vrml models in gps-based mobile guides," Proceedings of the tenth international conference on 3D Web technology, pp.57–64, Web3D '05, New York, NY, USA, ACM, 2005. <http://doi.acm.org/10.1145/1050491.1050499>
- [7] M. Nakamura, A. Tanaka, H. Igaki, H. Tamada, and K. Matsumoto, "Constructing home network systems and integrated services using legacy home appliances and web services," International Journal of Web Services Research, vol.5, no.1, pp.82–98, 2008.
- [8] M.P. Papazoglou and D. Georgakopoulos, "Service-oriented computing," Communication of the ACM, vol.46, no.10, pp.25–28, 2003.
- [9] R. Schmidt, "Multiple emitter location and signal parameter estimation," Antennas and Propagation, IEEE Transactions on, vol.34, pp.276–280, 1986.