

歩行者追尾型自律移動する超指向性スピーカ・システムの開発

研究代表者	近藤和弘	山形大学大学院理工学研究科	教授
研究分担者	三浦正範	山形大学大学院理工学研究科	研究員

1 はじめに

本研究では、端末を用いなくても音声通信ができるハンズフリー環境を提供するシステムを試作した。利用者の耳元だけに音声再生され、また利用者が話した音声を遠隔位置から集音し、両手を塞がずに、利用者が歩行しながらでも音声通信を行うことができる環境を目指す。そこで、まず利用者の頭部位置を検出し、その耳元でのみ音を再生するため超指向性パラメトリック・スピーカを、自走式ロボット・プラットフォームに搭載し、利用者の耳位置に追従して向きを変えてビーム再生するシステムを試作し、評価した。

また試作システムを用いて、音声会話内容の漏洩を防ぐノイズマスキング・システムを試作し評価した。マスキングしたい特定人物に対してマスキングノイズをピンポイントで再生し、音声会話内容を聞き取りづらいうようにする。同時にマスキングノイズをビーム状に再生することで、周囲の騒音レベルが上昇して環境が悪化することを防ぐ。

2 歩行者を追尾し自律移動する超指向性スピーカ・システムの開発

2-1 提案システムの目標

提案者らは、移動型人追跡指向性スピーカ、すなわち受話者を追尾して音声を提示するシステムを用いた音声通信方法について検討してきた [1-3]。この研究では、Kinect センサーと移動ロボットを用いて、受話者位置を認識してシステム自体が移動することで追尾を行い、指向性が強く音の広がりが小さいパラメトリック・スピーカを用いて、受話者にのみ音声を提示するシステム構成の実現性や性能について検証した。

本研究では、上記先行研究で検証された基本構成要素の実装経験を基に、より高性能な要素をあらたに導入して、先行研究からの性能改善、および実用的な性能の追求を目的に、システムの改良、および評価を行った。

2-2 超指向性スピーカ・システムの構成

受話者追尾音声提示システムの外観を図 1 に、またその制御の流れを図 2 に示す。コンピュータ制御可能なパンチルト雲台に Kinect センサーとパラメトリック・スピーカを一体として設置し、パンチルト雲台により Kinect センサーとパラメトリック・スピーカの仰角、ならびに方位角を制御する。

パラメトリック・スピーカは三菱電機エンジニアリングの超指向性音響システム MSP-50E/MSP-30M を用いた。このシステムは 91 dB (2kHz 出力時) の大きな出力が可能なシステムであるが、放射器 (MSP-50E) の他に変調器 (MSP-30M) が必要であり、特に後者は商用電源を必要としており、50 W の消費電力が定格となっている。

Kinect センサーは Microsoft 社が同社のテレビゲームのセンサーとして開発したものであり、可視光カメラの他、赤外線カメラも備えており、距離画像を取得することができる。本研究ではその性能を大幅に向上させた Kinect for Windows v2 [4-6] を用いた。

パンチルト雲台は FLIR Motion Control Systems 社製の PTU-D46 を用いた。このユニットと PC を RS-232C で接続し、コマンドを送信することで仰角と方位角を制御可能である。このユニットは最大 4 kg の荷重を 60° /s で移動可能である。

Kinect センサーとパラメトリック・スピーカの放射器 (MSP-50E) を一体化してパンチルト雲台に固定し、制御した。パンチルト雲台はカメラ用三脚で床上に固定した。

制御手順は、まず Kinect v2 センサーで認識した人物の中から追尾対象者を検出し、頭部座標を取得する。これには Kinect センサーより得られる可視光画像と、赤外線カメラより得られる距離画像から、Kinect for Windows SDK [4] と OpenCV [8] ライブラリを用いて人物の領域を認識し、これより頭部位置を算出する。Kinect for Windows SDK は Kinect をプログラム制御、ならびにデータ入出力を行うドライバやプログラムを集めた

ライブラリである。一方、OpenCV はコンピュータビジョンで用いる共通の機能をライブラリ化したものである。

頭部位置を検出後、追尾対象者とパラメトリック・スピーカから再生されるビームの有効範囲のズレの量、方向に応じてパンチルト雲台の移動スピード、移動ポジションの設定を行う。以上の処理を繰り返すことで受話者追尾動作を行う。



図 1. 受話者追尾音声提示システム外観

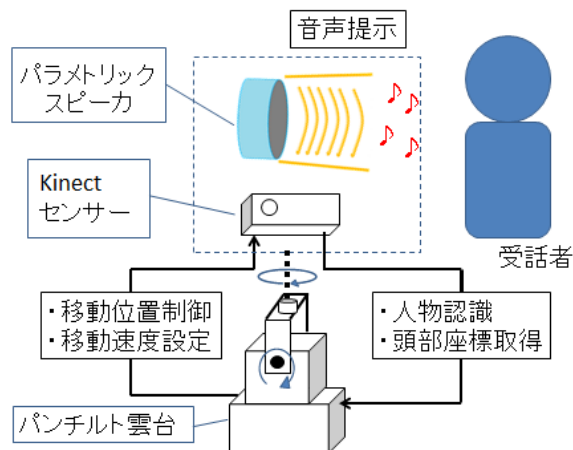


図 2. 受話者追尾制御の流れ

2-3 超指向性スピーカ・システムの性能評価

(1) 追尾性能評価

音声提示システムの追尾精度を確認するため、追尾精度測定実験を行った。図 3 に示す指定経路 (A→B→C→D) 上を被験者が歩行し、その追尾精度を評価した。水平、垂直方向の誤差距離 30cm 以内が可聴領域と想定できるので、これを許容誤差として、この範囲内に追尾可能な期間に対し、全測定時間の比率を追尾率と定義し、これを算出した。なお、許容誤差範囲内に追尾されていることは Kinect で撮影された可視光画像を記録しておき、これを手動で解析して、画像中心部の許容誤差範囲内にあることを視認して判定した。Kinect の撮像画像は Kinect for Windows SDK に付属されている Kinect Studio[9]を用いて録画した。

実験結果を表 1 に示す。本実験の平均追尾率は 82.07% であり、先行研究[1]での追尾であった 77% に比べて精度改善ができたと考えられる。なお、今回の実験では取って歩行速度の統制を行っていないため、追尾精度の個人差の一因となっていることが考えられる。また、受話者とシステムが近づき過ぎると追尾が困難となる場面が見られた。そこで、受話者とシステムとの距離を一定量以上に保つ配置を工夫することで、追尾精度を更に改善できると考えられる。

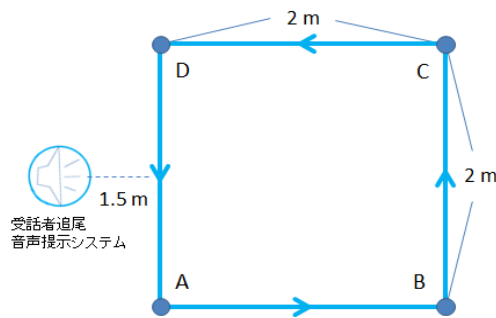


図 3. 追跡精度測定実験で用いた歩行経路

表 1. 追尾精度測定実験結果

被験者	水平平均誤差 X [cm]	垂直平均誤差 y [cm]	追尾率 [%]
1	12.65	9.99	93.49
2	25.18	8.10	74.85
3	20.67	10.17	80.84
4	21.89	9.97	77.97
5	20.24	13.45	80.41
6	13.01	7.79	84.86
全被験者平均	18.94	9.91	82.07

(2) 再生音圧レベルの評価

受話者に音が提示されていること及びパラメトリック・スピーカの有用性を検証するために、音圧測定実験を行った。ここで、図 4 に示すように、音声を提示する対象者を被験者、被験者以外の人物を第三者と定義する。実験に先立ち、システムとの距離が最短となる点で 70 dB の音圧となるように音量を調整した。その後、被験者が指定経路を歩行し、システムで被験者の頭部を追尾し試験音を再生した際に耳に入る音を、イヤホン型小型マイクで、また第三者に聞こえる音を通常のマイクを用いて録音した。被験者は図 4 に示す直線経路上を A→B→A→C と往復した。この時の録音音声から音圧の変化を測定した。また、従来型の音再生システムとしてラウドスピーカを用いて試験音を再生した場合でも同様の実験を行った。

音圧の変化を図 5 に、また、それぞれの平均音圧を表 2 に示す。本システムにより、被験者に対して最大 5 dB の変動はあるが、70 dB 程度で定常的な音声提示ができていることを確認した。また、本システム使用時の第三者位置での音圧が、被験者の音圧より 15 dB、従来のラウドスピーカ使用時の第三者位置の音圧より 12 dB 程度低いことから、パラメトリック・スピーカを用いることで秘匿性の高い音声提示ができていることが分かった。

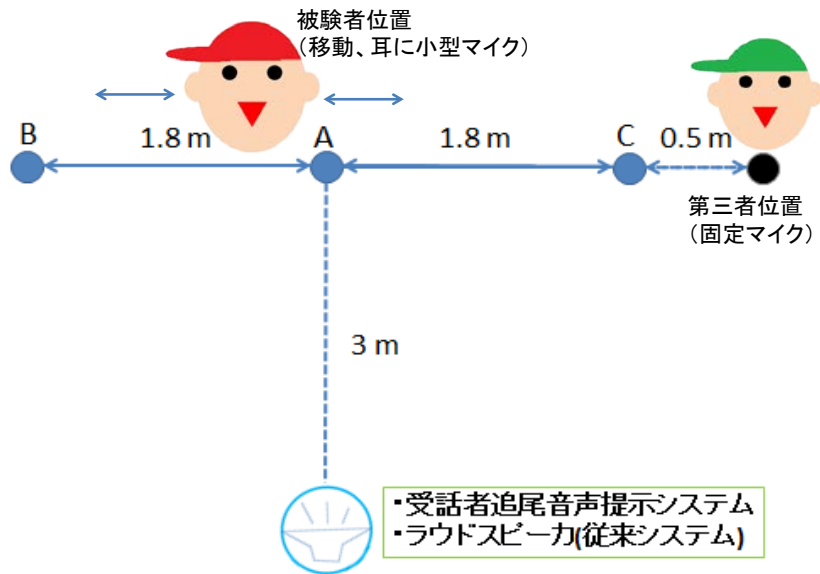


図 4. 音圧測定に用いた音源と観測マイク配置

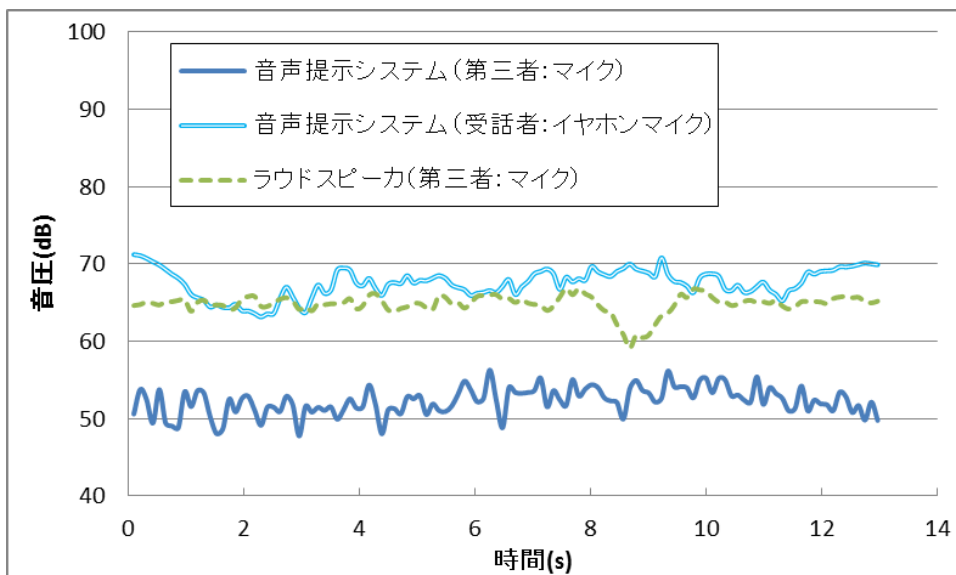


図 5. 音圧測定実験の音圧変化

表 2. 各システムで観測された平均音圧

項目	音声提示システム (受話者位置)	音声提示システム (第三者位置)	ラウドスピーカ (第三者位置)
平均音圧 [dB]	67.54	52.29	64.84

3 超指向性スピーカ・システムのマスキング・システムへの応用

3-1 超指向性スピーカ・システムを用いたマスキング・システムのねらい

情報化社会となった近年では、住所、氏名、年齢はもちろんのこと、勤め先ですら個人情報として悪用される恐れがある。通常、情報交換は会話によって行うことが多いが、この時近くにいる第三者にこの情報を

盗み聞きされてしまう恐れがある。

会話による情報漏えいへの対応策として、音声マスキングが挙げられる。これは会話と同時にマスキングと呼ばれる妨害音を再生し、会話の聞き取りを妨害して情報漏えいを防ぐ方法である。しかし、マスキングは周囲に拡散するため、会話している本人や関係の無い人にまで聞こえてしまい、不快感を与えるという問題がある。この問題を解決する方法として、本研究では音の指向性の強い超指向性スピーカを用いた音声マスキングを利用することを検討した[10-12]。

そこで、2章で提案された自律移動型聴取者追尾システムをマスキング・システムに応用することを検討する。このシステムで超指向性スピーカでマスキングを再生することで、ピンポイント・マスキングが可能となる。この時、対象者の周辺にはマスキングが再生されないため、周囲の騒音レベルが上昇せず、周囲環境への影響を抑えられることが期待できる。

3-2 超指向性スピーカ・システムを用いたマスキング・システムの構成

本研究で用いた自律移動型聴取者追尾システムの概要を図6に示す。2章で用いたシステムに比べて可動性をさらに重視して、全体を移動可能なロボット・プラットフォームの上に搭載して利用した。電源も含めて自律したシステムとするため、小型で変調ユニットを必要としない小型超指向性音響システムを利用した。

ロボット・プラットフォームは iRobot のルンバに似た研究開発用のロボット・プラットフォームである Turtlebot を用いた。ここでは Yujin Robot 社の Turtlebot 実装である Kobuki [13] を用いた。Kobuki は PC から簡単なコマンドを送ることで移動方向および速度を制御できる。

小型超指向性音響システムは TriState 社製のパラメトリックスピーカ・キットを用いた[14]。このキットは 12 V、最大 500mA で駆動できるので、Kobuki 搭載のバッテリーから電源供給が可能である。

仰角、方位角の制御には Eagletron 社製 TrackerPod [15] を用いた。USB から簡単なコマンドで仰角、方位角を制御可能である。電源も USB より供給可能である。パラメトリック・スピーカと Kinect を TrackerPod の上に固定した。

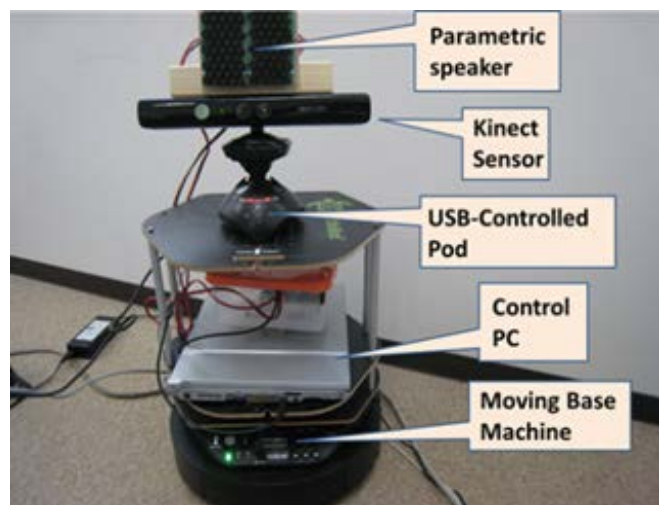


図 6. 超指向性マスキング・システムの概観

このシステムでは、まず Kinect から得た可視光画像と距離画像から対象人物の頭部座標を取得する。取得した頭部座標データを元に Control PC で Kobuki を動かす。パラメトリック・スピーカ (以下 PS) を TrackerPod の上に設置し、Kobuki と、TrackerPod と共用して PS を追尾対象者の頭部に向け続けることで継続して音声を聞かせ続ける。

このシステムを応用し、マスキング対象者を追尾して PS でマスキングを再生することで、マスキング対象者のみをマスキングすることが可能になる。

3-3 超指向性スピーカ・システムのマスクング性能評価

(1) 評価条件

このマスクング・システムを利用した際のマスクング性能の評価は、日本語用了解度試験の一つである Japanese DRT (Diagnostic Rhyme Test) [16]を用いて評価する。

提案したシステムで DRT を実施する際、複数の被験者が同じ経路を同じ速度で何度も歩行する必要があり、実験条件を統一するのは困難であると考えられる。そこで、音響用ダミーヘッド(サザン音響製 SAMAR)を用いてあらかじめ試験用の音声を録音して試験音とし、これをヘッドホンより再生して DRT を行うこととした。

DRT に用いる音声をダミーヘッドで録音するために、ダミーヘッドを図 7 に示す移動経路 1 上を約 5 秒かけて移動させて、その時の提示音と妨害音を、ダミーヘッドの耳介部に設置した小型イヤホンマイクで録音した。この時、提示音はラウドスピーカで、妨害音は超指向性スピーカで音声を再生した。また、ダミーヘッドの耳介部の高さは 1.6 m、提示音用のラウドスピーカの高さも同じく 1.6 m、追尾システムのスピーカ部分は 0.6 m の高さにあり、ダミーヘッドに対して斜め下から妨害音を提示するようになっている。追尾システムはダミーヘッドを追って移動するため、図 7 に示す移動経路 2 に沿って移動する。

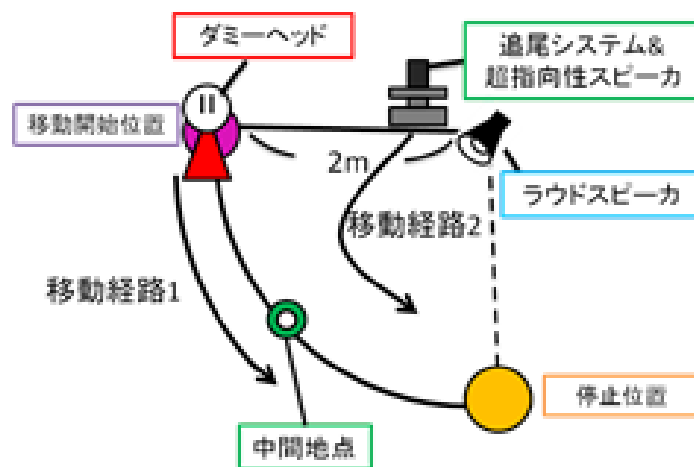


図 7. 超指向性マスクング・システム性能評価に用いる音源と測定点配置

妨害音は追尾システムで追尾しながら超指向性スピーカから提示し、その時の音を録音した。音圧は追尾システムがダミーヘッドを追って停止する位置で、超指向性スピーカの再生音が歪まない 50 dBA 程度になるように調整した。また、ラウドスピーカからも試験音を再生して録音した。この時の音圧は超指向性スピーカに合わせ、中間地点で 50 dBA 程度となるように調整した。

また、従来のマスクング・システムとの比較のために、図 7 に示すラウドスピーカからも妨害音を再生して録音した。提示音と妨害音は別々に録音を行い、後に PC で合成して試験音とした。

合成した試験音はヘッドホンで評価者に提示して DRT を実施した。提示音圧は 60[dBA]とし、試験音と妨害音の音圧比である SNR は 0, -6, -12 の 3 パターンで実験を行った。試験に使用する目標音には女性 1 名の声を用い、妨害音は pink noise, 不特定話者雑音 (バブル雑音), 同性話者 (Multi-Speaker Gender Dependent; MSGD) 雑音 [17, 18] の 3 種類を用いた。MSGD 雑音とは不特定多数の人の声から作った妨害音で、同性話者雑音とは目標音と同じ性別の人のみの声から作った妨害音である [18]。被験者は 20 代男性健聴者 5 名である。

(2) 評価結果

DRT の結果を、妨害音の種類ごとに図 8, 図 9, 図 10 に示す。横軸は SNR, 縦軸は了解度を表わしている。また、図中に 95% 信頼区間も示した。この結果から、自律移動型聴取者追尾マスクング・システムによるマスクングと従来のラウドスピーカを用いたマスクング・システムによるマスクングのマスクング性能には有意差は無かった。また、従来システムと同様に音声了解度が低下し、同等のマスクング効果があることがわかった。この結果はどの妨害音の種類についても同様であった。

以上の結果から、自律移動型聴取者追尾音声提示マスクング・システムでピンポイント・マスクングを行うと、従来のシステムと同等のマスクング性能があることがわかった。

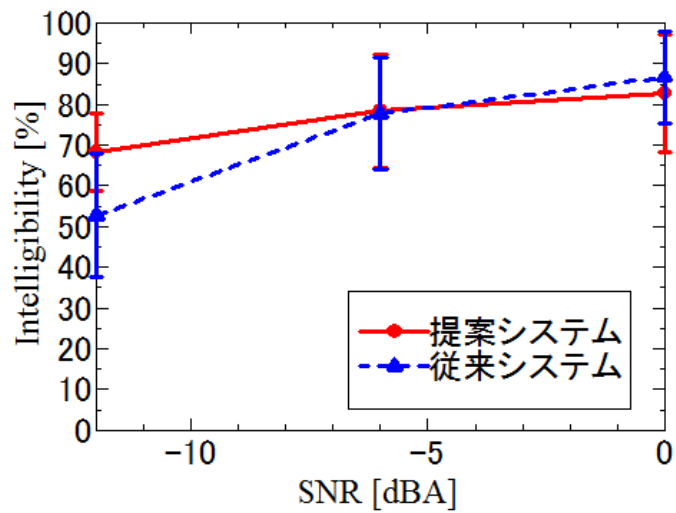


図 8. pink noise を妨害音に用いた場合の音声了解度

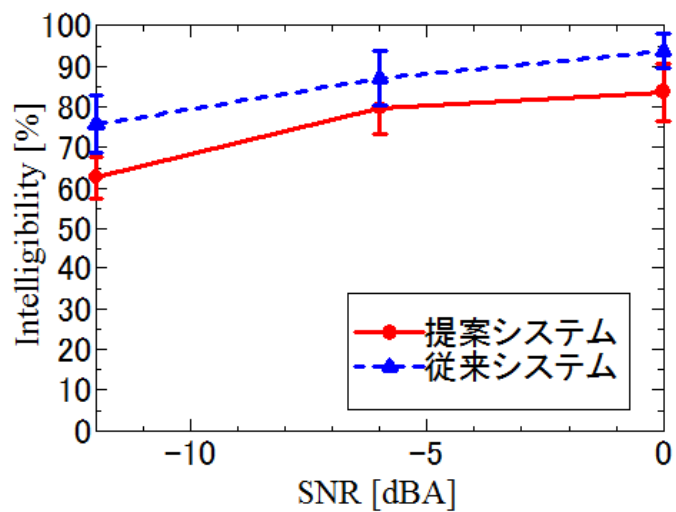


図 9. 不特定話者雑音を妨害音に用いた場合の音声了解度

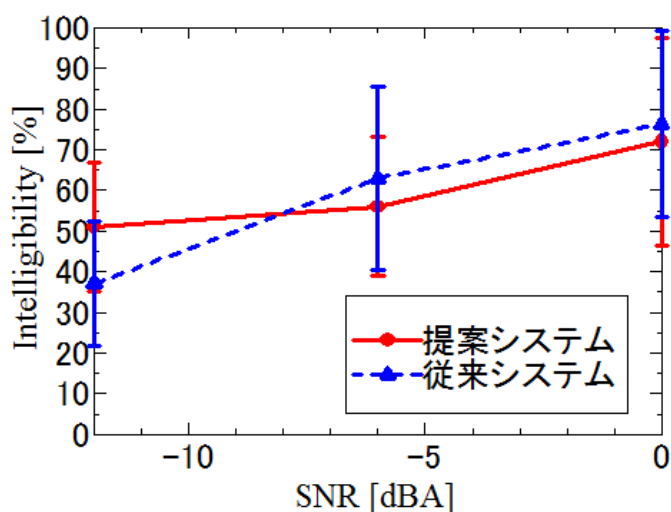


図 10. MSGD 雑音を妨害音に用いた場合の音声了解度

3-4 超指向性マスクング・システムの妨害音の周囲環境への拡散量の評価

提案した超指向性マスクング・システムでは、再生音の指向性が強い超指向性スピーカを用いて音声マスクングを行っている。このため、ラウドスピーカを用いた従来型の音声マスクング・システムに比べ、周囲への妨害音の拡散防止に役立つと考えられる。本項では、実際に提案マスクング・システムを用いてマスクング雑音を再生し、周囲にいる第三者に聞こえる妨害音の音圧を測定し、本システムの妨害音の拡散防止効果を評価する。

(1) 評価条件

超指向性マスクング・システムを使って実際にマスクング対象に妨害音を提示してマスクングを行い、その時のマスクング対象者周囲の各測定点の妨害音の音圧を騒音計で測定した。この時の実験条件を表 3 に示す。

表 3. 妨害音拡散測定の評価条件

項目	条件
妨害音種類	pink noise
妨害音提示音圧	60 [dBA]
スピーカ設置の高さ	0.65 [m]
マスクング対象の高さ	1.60 [m]
水平距離	1.00 [m]
測定範囲	スピーカ正面 3.5 [m] スピーカ左右 各 1.5 [m]
暗騒音	20.8 [dBA]

自律移動型聴取者追尾マスクング・システムのマスクングを行う位置は、マスクング対象に対して追尾を行い停止した位置とする。また、マスクング対象にはダミーヘッドを使用した。この時、胴体部分に対する反射特性を実際の人間に近づけるために、ダミーヘッドにはコートを着用させた。音圧の測定範囲は、追尾システムの正面方向 3.5 [m]、左右方向に各 1.5 [m] の範囲を、0.5 [m] 間隔で計測した。測定した高さは、一般的な直立している人の頭部の高さである 1.6 [m] の高さで測定を行った。この時、マスクング対象にマスクングが行われている時に、その周囲にいる第三者に聞こえる妨害音の大きさを模擬して測定するために、音圧の測定中はマスクング対象のダミーヘッドをそのまま設置して音圧の測定を行った。妨害音には pink noise を使用し、ダミーヘッドの頭部で 60 [dBA] になるように音圧を調整した。

妨害音拡散の比較するために、従来型のマスクング用のラウドスピーカはマスクング対象の頭部を追尾後に停止した超指向性マスクング・システムの超指向性スピーカと同じ位置に設置した。この時、向きはマス

キング対象のダミーヘッド頭部へ向け、音圧もダミーヘッドの頭部で 60 [dBA]になるように調整した。こちらの妨害音の音圧の測定も自律移動型聴取者追尾マスキング・システムと同じようにスピーカの正面方向 3.5 [m]、左右方向に各 1.5 [m]、高さ 1.6 [m]の範囲で、0.5[m] 間隔で計測した。

(2) 評価結果

実験は壁面と床に簡単な音響処理をした簡易無響室を用いて行った。

実験で測定した音圧を図 11 と図 12 に示した。この図はスピーカを原点にして、スピーカの正面 3.5 [m]、左右方向各 1.5 [m]の範囲で測定した音圧を表わしたもので横軸と縦軸は[m]、カラーバーは音圧を [dBA] で表示している。音圧の測定はマスキング対象となるダミーヘッドを設置した状態で測定した。この時、ダミーヘッドは図の横軸を X、縦軸を Y としたとき $(X, Y) = (0, 1.0)$ の座標に位置している。

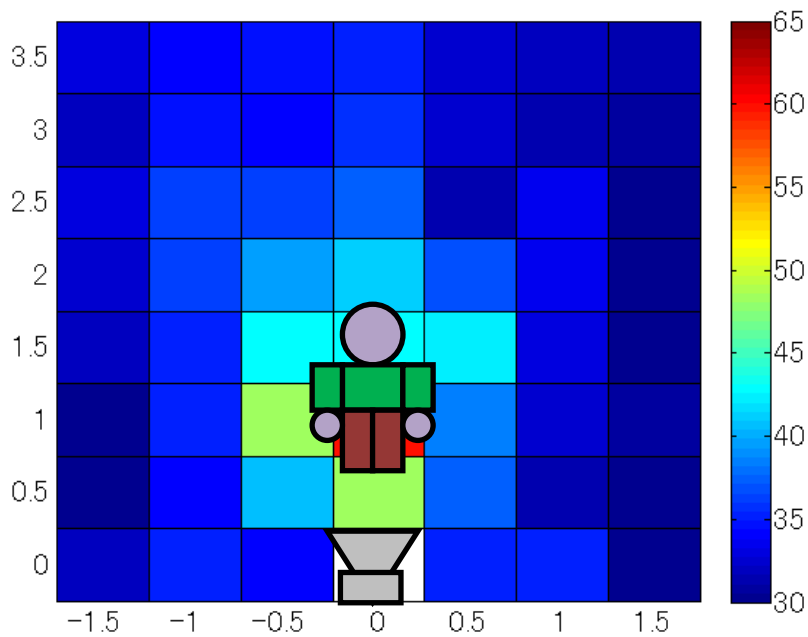


図 11. 超指向性マスキング・システムの妨害音提示時の周囲音圧分布

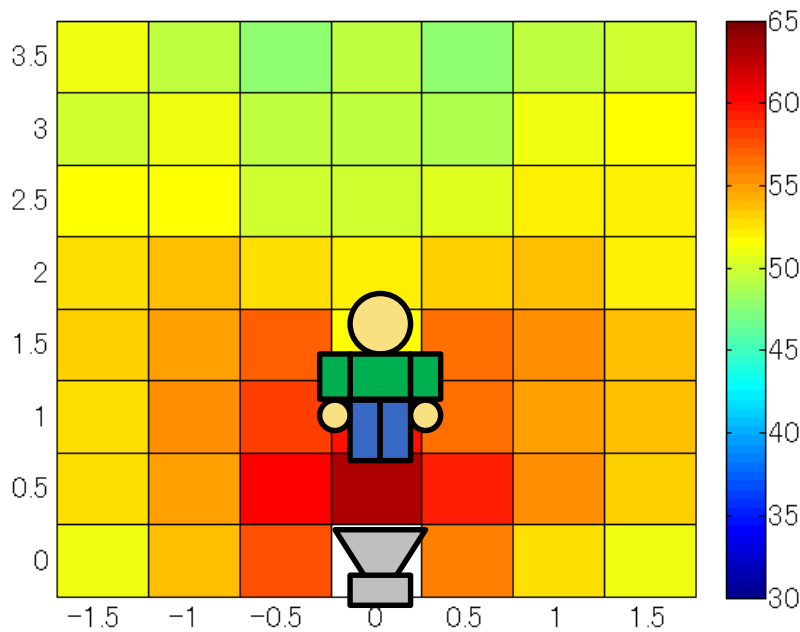


図 12. 超指向性マスキング・システムの妨害音提示時の周囲音圧分布

図 11 に示す超指向性マスキング・システムでは、超指向性スピーカを用いたことによりマスキング対象であるダミーヘッド以外の場所で音圧が約 10 から 20 [dB]程度低下し、多くの場所で音圧が約 30~35 [dBA]に抑えられていることがわかる。また、追尾システムとマスキング対象に挟まれ、マスキングの通り道になっていると考えられる $(X, Y) = (0, 0.5)$ の位置でも音圧が低下していた。これは、音圧の測定位置が直立している人の頭部の高さにあたる 1.6 [m]であるため、音の指向性の強い超指向性スピーカでは妨害音が拡散せずに音圧が上昇しなかったことが原因であると考えられる。

一方で、図 12 で示されるように、ラウドスピーカによる従来のマスキング・システムでは、スピーカを中心に妨害音が広範囲に拡散し、多くの場所で妨害音の音圧は 50 [dBA]以上、もっとも音圧の低い場所でも 45 [dBA]の音圧が観測された。このとき、ダミーヘッドが $(X, Y) = (0, 1.0)$ の位置にあったため、ダミーヘッドの陰になる場所では音が届きづらくなり、今回のような妨害音の広がり方となったと考えられる。

以上の結果から、今回の実験の自律移動型聴取者追尾マスキング・システムを用いたピンポイント・マスキング・システムでは、従来のマスキング・システムに対し妨害音の音圧は約 10~20 [dBA]、最高で 24.9 [dBA]低減できた。このことから、この自律移動型聴取者追尾マスキング・システムでピンポイント・マスキングを行うことで、従来のマスキングシステムに比べ妨害音の拡散を防止することができ、周囲への不快感を防止することに有効であるということがわかった。

【参考文献】

- [1] 柏瀬瞬, 近藤和弘, “移動型人追跡指向性スピーカを用いた音声通信方法の基礎検討,” 日本音響学会 2014 年秋季研究発表会, 2-P-41, 2014.9.
- [2] Shun Kashiwase, Kazuhiro Kondo, “Towards a Parametric Speaker System with Human Head Tracking Beam Control,” Proc. IEEE Global Conference on Consumer Electronics, pp. 22-23, Oct. 2014.
- [3] 柏瀬瞬, 近藤和弘, “自律移動スピーカを用いた音声通信方法の検討,” 日本音響学会 2015 年春季研究発表会, 1-Q-49, 2015.3.
- [4] 中村, 杉浦, 高田, 上田, 「KINECT for Windows SDK プログラミング Kinect for Windows v2 センサー対応版」, 秀和システム, 2015.5.

- [5] 薬師寺、「Kinect v2 プログラミング入門」、インプレス、2014 年 12 月。
- [6] 薬師寺、「Kinect v2 プログラミング実践」、インプレス、2015 年。
- [7] FLIR Pan-Tilt Unit-E46 Models, <http://www.flir.com/mcs/view/?id=63554>
- [8] OpenCV, <http://opencv.jp>
- [9] Kinect Studio, <https://msdn.microsoft.com/ja-jp/library/dn785306.aspx>
- [10] Makoto Ohba, Kazuhiro Kondo, "On an Efficient Masking System Using Human Head Tracking Parametric Speaker System," Proc. IEEE Global Conference on Consumer Electronics, pp. 222-223, Oct. 2015.
- [11] 大場真人、近藤和弘、「超指向性スピーカを応用した歩行者追尾型マスキングシステムの基礎検討」、平成 26 年東北地区若手研究者研究発表会, YS-12-P4、2014.2.
- [12] 大場真人、近藤和弘、「超指向性スピーカを応用した歩行者追尾型マスキングシステムの改良」、日本音響学会 2015 年春季研究発表会, 2-P-27、2015.3.
- [13] Kobuki, <http://kobuki.yujinrobot.com/>
- [14] パラメトリック・スピーカ実験キット、<http://www.tristate.ne.jp/parame.htm>
- [15] TrackerPod, pan/tilt base for webcams, <http://www.trackercam.com/TCamWeb/productdes.htm>
- [16] 近藤, 他, 「二者択一型日本語音声了解度試験方法の検討」, 音響学会誌, 63 (4), pp. 196-205, 2007.4.
- [17] 桜井弘毅, 近藤和弘, 「複数の同性話者音声から合成したマスキング雑音の性能評価」, 平成 26 年東北地区若手研究者研究発表会, YS-12-A5、2014.2.
- [18] Kazuhiro Kondo, Hiroki Sakurai, Shintaro Kashiwada, Tomohiro Komiyama, "Speaker and Gender-Dependent Maskers for Efficient Speech Privacy Protection," Noise Control Eng. J., Vol. 62, No. 6, pp. 411-421, Nov./Dec. 2014.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
自律移動型聴取者追尾システム及びそれを応用したマスキング・システムの性能評価法の検討	日本音響学会秋季研究発表会	2015 年 9 月
On an Efficient Masking System Using Human Head Tracking Parametric System	IEEE Global Conference on Consumer Electronics	2015 年 10 月
超指向性パラメトリック・スピーカを用いたジュワ者追尾音声提示システムの改良	東北地区若手研究者研究発表会	2016 年 3 月
自律移動型聴取者追尾マスキング・システムの了解度評価	日本音響学会春季研究発表会	2016 年 3 月
Efficiency Evaluation of a Super-Directive Masking System That Tracks the Masker Beam Towards the Target Human Head	IEEE Global Conference on Consumer Electronics	2016 年 10 月 (投稿中)