車室内の三角マイクロフォンアレイへの ヴァーチャルマイクロフォン技術の適用*

☆瀬川 華子, 髙橋 理希, 李 莉, 陣在 遼河, 牧野 昭二, 山田 武志（筑波大学）

1 はじめに

近年、情報通信技術 (ICT) は急速に発展しており、これを用いた車室内におけるコミュニケーション補助システムの開発が検討されている [1][2]。 車室内では、話者の配置や背景雑音によってコミュニケーションが難しくなる。例えば、後部座席の搭乗者は運転席近くの話者からの発話を聞き取りづらい。このような状況において、特定の話者の発話を強調し、他の搭乗者の近くに設置したスピーカーから再生することによって聞き取りやすさを改善することができる。

目的話者の音声を強調する手法の 1 つとして、ビームフォーミングによる音声強調技術がある。車室内においては音声の到来方向が予測しやすいため、ビームフォーミングによる音声強調が効果的であると考えられる。しかし、ビームフォーミングは音源数よりもマイク数が少ない場合（劣決定条件）においては十分な強調性能が得られないという問題点がある。 車室内においては、話者の増加、エンジンやウインカーなどの機械からの雑音などにより、劣決定条件になりやすいと考えられる。1 つの解決策として、マイク数を更に増やすことが考えられるが、マイクロフォンアレイが大型になることや、コストの増大、車室内のほかの設備との兼ね合いなどを考えると、この解決策は現実的とは言えない。

この状況を改善するために、ヴァーチャルマイクロフォン技術が有効である。ヴァーチャルマイクロフォン信号は 2 つの実マイクの信号から生成され、2 つのマイクを結ぶ直線上に配置される。ヴァーチャルマイクロフォンを用いることで、ビームフォーマのチャネル数を増やすことができる。2 マイクでの実験においては、劣決定条件時のビームフォーマの強調性能改善への有用性がすでに示されている [3]。本研究では、3 マイクを用いた音声強調におけるヴァーチャルマイクロフォンの有効性および、車室内コミュニケーションという現実的な状況への適用による有効性を検証する。

本論文では、車室内の三角マイクロフォンアレイによる観測信号からヴァーチャルマイクロフォン信号を生成し、適応ビームフォーミングの性能で評価する。実験には車室内で実測したインパルス応答を用い、車室内コミュニケーションにおけるヴァーチャルマイクロフォンの有効性を検証する [4]。

2 三角マイクロフォンアレイにおけるヴァーチャルマイクの適用

2.1 ヴァーチャルマイクロフォンの内挿

ヴァーチャルマイクロホンは配置される位置により内挿と外挿に分けられ、推論に異なる規準が用いられる。ここでは、ヴァーチャルマイクロフォンの内挿について紹介する [5][6]。ヴァーチャルマイクロフォン技術では 2 つの実マイクロフォンの観測信号 $x_i(\omega, t)$ から、ヴァーチャルマイクロフォン信号 $v(\omega, t, \alpha)$ を生成する。i は生成に用いる実マイクロフォンの識別子

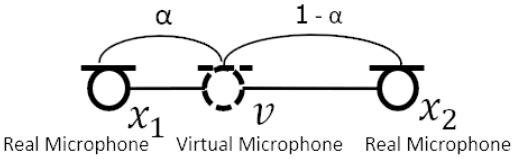


Fig. 1: Arrangement of real and virtual microphones in interpolation technique

$(i = 1, 2)$ であり、ω はそのマイクロフォンにおける周波数ビン、t は時間フレームを表す。$\alpha (0 < \alpha < 1)$ はヴァーチャルマイクロフォンの補間係数であり、実マイクロフォン間の距離を 1 とした時の、1 番目の実マイクとヴァーチャルマイクロフォン間の距離の値である。図 1 に実マイクロフォンとヴァーチャルマイクロフォンの関係を表す。複数の音が異なる方向から到来する時、マイクロフォンの位置と波形の関係は一様ではなく、補間が難しい。そこで、観測信号はスパースであり、1 つの時間周波数ビンでは 1 つの音源のみが支配的であることを仮定する [7]。これによって、複数の音が到来した時であっても各時間周波数ビン内では単一の音とみなすことができ、ヴァーチャルマイクロフォン信号の補間を行うことができる。

ヴァーチャルマイクロフォン技術においては位相と振幅は個別に補間される。実マイクロフォンでの観測信号の位相と振幅は以下のように表される。

$$\phi_i = \angle x_i(\omega, t) = \tan^{-1} \frac{\mathrm{Im}(x_i(\omega, t))}{\mathrm{Re}(x_i(\omega, t))}, \quad (1)$$

$$A_i = |x_i(\omega, t)|. \quad (2)$$

平面波の到来を仮定した時、ヴァーチャルマイクロフォンを内挿した時の位相は線形補間によって次のように表される。

$$\begin{aligned} \phi_{\mathrm{v}} &= \phi_1 + \alpha(\phi_2 - \phi_1) \\ &= (1 - \alpha)\phi_1 + \alpha\phi_2. \end{aligned} \quad (3)$$

ここでは、信号の位相差が π を超えていないことを仮定して補間を行う。

$$|\phi_1 - \phi_2| \leq \pi. \quad (4)$$

ヴァーチャルマイクロフォンの振幅の補間は、多くの条件に依存しており、実際の振幅減衰を忠実にモデル化するのは困難である。そこで、β ダイバージェンスと呼ばれる以下の式を用いて振幅 A_{v} を補間する。

$$A_{\mathrm{v}} = \begin{cases} \exp\left((1 - \alpha)\log A_1 + \alpha \log A_2\right) & (\beta = 1) \\ \left((1 - \alpha)A_1^{\beta - 1} + \alpha A_2^{\beta - 1}\right)^{\frac{1}{\beta - 1}} & (\text{otherwise}). \end{cases} \quad (5)$$

β の値によって、2 つの実マイクロフォン間の振幅を非線形に補間することができる。これにより、ヴァーチャルマイクロフォン信号は以下のように表すことができる。

$$v(\omega, t, \alpha) = A_{\mathrm{v}} \exp(j\phi_{\mathrm{v}}). \quad (6)$$

*Applying Virtual Microphones to Triangular Microphone Array in in-Car Communication. by Hanako SEGAWA, Riki TAKAHASHI, Li LI, Ryoga JINZAI, Shoji MAKINO, Takeshi YAMADA

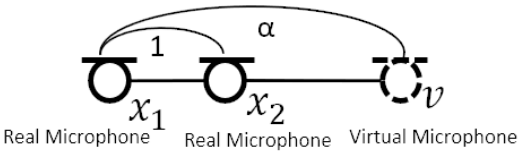


Fig. 2: Arrangement of real and virtual microphones in extrapolation technique

2.2 ヴァーチャルマイクロフォンの外挿

次に、ヴァーチャルマイクロフォンの外挿を紹介する [8]。図 2 に実マイクロフォンとヴァーチャルマイクロフォンの関係を表す。位相の補間に関しては内挿と同じ式を用いることができるが、振幅については前述のとおり補間が困難である。また、外挿の場合、β ダイバージェンスを用いた補間を行うと複素振幅や正の無限大への発散といった現実的でない振幅が生成されてしまうことがあり、これは避けるべきである。そこで、1.5kHz 以下の周波数帯域の信号において、位相と比較して振幅は位置の違いによって生じる差が小さいため [9][10]、ヴァーチャルマイクロフォンの位置に最も近い実マイクロフォンの振幅を外挿したヴァーチャルマイクロフォンの振幅として用いる。

$$A_{\mathrm{v}} = \begin{cases} A_1 & (\alpha < 0) \\ A_2 & (\alpha > 1). \end{cases} \tag{7}$$

これらより、外挿されたヴァーチャルマイクロフォン信号 $v(\omega, t, \alpha)$ は内挿と同様に以下のようにして表すことができる。

$$v(\omega, t, \alpha) = A_{\mathrm{v}} \exp\left(j\phi_{\mathrm{v}}\right). \tag{8}$$

2.3 ヴァーチャルマイクロフォンの波長比例配置

続いて、ヴァーチャルマイクロフォンの波長比例配置 (Wavelength-Proportional Arrangement of Virtual Microphones:WPVM) について紹介する [3]。ヴァーチャルマイクロフォンの波長比例配置において、ヴァーチャルマイクの位置を表す係数である α は以下のように算出される。

$$\alpha(\omega) = \frac{\lambda k}{d} = \frac{2\pi c k}{\omega d}. \tag{9}$$

ここで、λ は波長、c は音速、d は実在のマイク間の距離、k は波長係数を示す。WPVM を用いる時、k を適切に設定した場合、マイクロフォンアレイは低周波数帯では大きく、高周波数帯では小さくなり、位相差も十分に得られる。理論上では、$k = 0.5$ において位相差が十分に得られ、良い強調性能が期待できる。一方、$k = 0.25$ では十分な位相差が得られず、$k = 2$ では空間的エリアシングが発生する。この手法は、2 マイクの実験ではすでに強調性能改善に有効であることが示されており、実験によると適切な k は 0.5, もしくは 1 であるとされている [3]。

2.4 音声強調に用いる三角マイクロフォンアレイビームフォーマ

最後に、本実験の音声強調に用いるビームフォーマについて紹介する。本実験において、ヴァーチャルマイクロフォンの性能を maxSNR ビームフォーマのチャネル数を増やすことにより評価する [11] [12]。maxSNR ビームフォーマは、目的音のみの区間と妨害音のみの区間の共分散行列に関する事前情報を必要とし、目的音と妨害音のパワー比が最大となるように空間フィルタを構成する。なお、原理的には、ヴァーチャルマイクロフォンは maxSNR ビームフォーマと同様に他のマイクロフォンアレイ信号技術にも適用できる。

また、比較対象として、ヴァーチャルマイクロフォンを適用しない Minimum Variance Distortionless Response (MVDR) ビームフォーマを用いた音声強調も行う [11][13]。妨害音と目的音に相関がない時、観測信号の分散は目的音と妨害音の分散の和となるため、MVDR ビームフォーマでは、観測信号の分散を最小化することで妨害音の影響を軽減する。

3 評価実験

3.1 実験概要

本実験では、4 音声と車室内で実測したインパルス応答を用いて、三角マイクロフォンアレイにおけるヴァーチャルマイクの外挿と WPVM 配置の強調性能向上への有効性および、波長係数 k や強調話者、ヴァーチャルマイクの設置方向、窓の開閉による有効性への影響を調べた。

3.2 実験条件

本実験では、ATR デジタル音声データベースのセット B に収録されている、全 503 文の音素バランス文の男性 6 話者，女性 4 話者の計 10 話者分のデータを使用した。このデータベースの中から、ランダムに 4 人を選択し、実録の車室内インパルス応答を畳み込むことで観測信号を 20 パターン作成した。実マイクの数は 3 つであり、ヴァーチャルマイクロフォンを追加する実験の場合はヴァーチャルマイクロフォンを 1 つ追加した。

インパルス応答は車室内で録音した時間引き延ばしパルス (TSP) を用いて測定した。音源は運転席、助手席、運転席側後部座席、助手席側後部座席に配置し、マイクロフォンを自動車前方のマップランプに取り付けて TSP 信号を録音した。音源とマイクロフォンの配置を図 3 に、三角マイクロフォンアレイの配置を図 4 に示す。また、他の実験条件を表 1 に示す。

本実験では、ヴァーチャルマイクロフォンを三角マイクロフォンアレイに適用しており、ヴァーチャルマイクロフォンの生成に用いるための 2 つのマイクの組み合わせは、{ (real mic1, real mic2): virtual mic1, (real mic1, real mic3): virtual mic2 } であり、ヴァーチャルマイクロフォンは後部方向に向かって生成した。WPVM における k は { 0.25, 0.5, 1, 2 }, 外挿の α は 10 とした。k は、[3] において試行されていた値であり、α は、[3] において SDR が高かった値である。ヴァーチャルマイクの設置方向を図 5 に示す。

本実験においては、三角マイクロフォンアレイのみを用いた音声強調に MVDR ビームフォーマと maxSNR ビームフォーマを用い、ヴァーチャルマイクロフォンを付け加える場合の音声強調には maxSNR ビームフォーマを用いた。また、運転席の音声と助手席の音声を強調する場合に分けて実験を行い、評価指針として signal-to-distortion ratio (SDR) を用いる。評価結果は、4 話者の組み合わせによる 20 パターンの SDR の平均で示す。maxSNR ビームフォーマと MVDR ビームフォーマにはどちらも目的音のみの区間と妨害音のみの区間を事前情報として与えている。事前情報として与えられている発話は、テストに用いる発

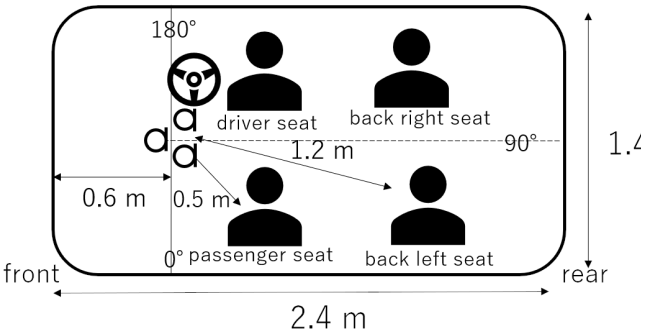


Fig. 3: Sound source and microphone layout in experiment

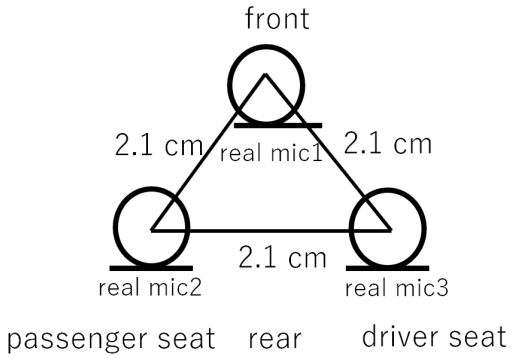


Fig. 4: Arrangement of triangular microphone array

話と同一の話者の別の内容の発話である。

3.3 実験結果と考察

20 パターンの平均で結果を評価する。図 6[1]において音声強調の結果を示す。図 6(a) では窓を閉めた状態、図 6(b) では窓を開けた状態での運転席の話者の音声強調の結果を示す。また、図 6(c)、図 6(d) では同様に助手席の話者の音声強調の結果を示す。また、横線は三角マイクロフォンアレイのみを用いた maxSNR ビームフォーマと MVDR ビームフォーマの SDR を、棒グラフはヴァーチャルマイクを適用した maxSNR ビームフォーマの SDR を示す。

まず、MVDR ビームフォーマと maxSNR ビームフォーマの SDR を比較すると、maxSNR ビームフォーマのほうが MVDR ビームフォーマよりも SDR が高い。これらは劣決定条件下であり、どちらもビームフォーミングが正常に機能していないと考えられる。MVDR ビームフォーマでは特にその影響が大きいため、SDR が低いと考えられる。

次に、三角マイクロフォンアレイのみを用いた maxSNR ビームフォーマにおける SDR と、ヴァーチャルマイクロフォンを付け加えた maxSNR ビームフォーマにおける SDR を比較すると、すべての k と α において、ヴァーチャルマイクロフォンがある場合の SDR がない場合の SDR を上回っている。これは、劣決定条件からヴァーチャルマイクロフォンを追加することで仮想的に決定条件になり、性能が向上したと考えられる。これらの結果から、ヴァーチャルマイクロフォンは劣決定条件において三角マイクロフォンアレイに対しても有効であることがわかる。

次に、WPVM 配置と外挿を用いた場合の SDR を比較する。$k = 0.5, 1$ を選択した時には WPVM 配置の方が外挿よりも SDR が高い傾向がある。反対に、$k = 2$ の場合には、外挿を下回る場合もある。適切な k を用いた WPVM 配置と、三角マイクロフォンアレイのみを用いた maxSNR ビームフォーマを比較すると、SDR が最大で 1.5 dB 上昇している。これらの結果から、適切な k を選択した WPVM 配置は、外挿より高い強調性能を示すことがわかる。

[1]SDR がとる値の関係上、グラフの最小値を (a)、(b)、(c) では 8.5 dB、(d) では 13 dB とした。

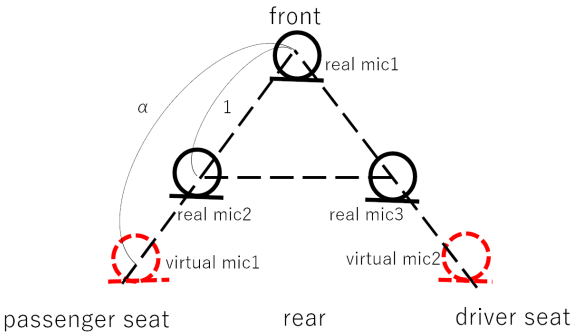


Fig. 5: Arrangement of real and virtual microphones

Table 1: experimental conditions

Sampling rate	8 kHz
Signal-to-noise ratio(SNR)	0 dB
FFT frame length	1024 samples
FFT shift	256 samples
Reverbration time	58 ms

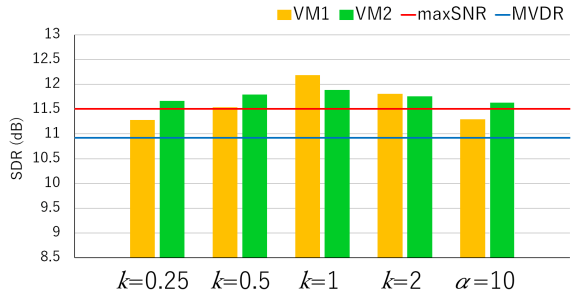
また、設置方向における強調性能の違いを比較すると、全体的には助手席方向に設置したヴァーチャルマイクロフォン（virtual mic1）の方が運転席方向に設置したヴァーチャルマイクロフォン（virtual mic2）の方が SDR が高い傾向があった。一方で、運転席話者の強調においては、virtual mic2 の方が SDR が高い場合がある。この結果から、設置方向による性能の変化は一様に述べられないが、一般には助手席方向にヴァーチャルマイクを設置した方が SDR が高くなる傾向があると考えられる。

次に、窓の開閉による強調性能への影響を比較すると、窓が開いている時の方が、窓が閉まっている時よりもヴァーチャルマイクロフォンの設置による SDR の改善量が大きい。運転席側の話者の音声強調に注目すると、窓が閉まっている時は SDR において最大 0.7 dB 程度の改善がみられているのに対し、窓が開いているには最大 2.2 dB 程度の改善が見られる。これは、ヴァーチャルマイクロフォンが反射がない状態を想定された技術であり、窓を開けることで反射の影響が軽減されているためであると考えられる。

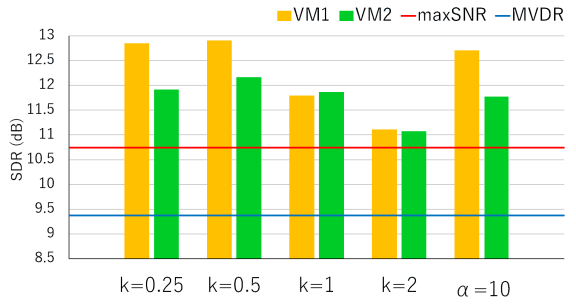
以上の結果は、ヴァーチャルマイクロフォンが三角マイクロフォンアレイにおいても有効であることを示している。

4 結論

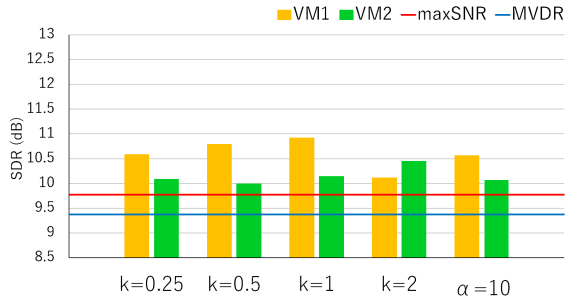
本稿では、ヴァーチャルマイクロフォンの外挿と WPVM 配置を劣決定条件における車室内に設置された三角マイクロフォンアレイでの音声強調に適用した。これらの実験では、音声信号に車室内での実録インパルス応答を畳み込むことで観測信号生成した。ヴァーチャルマイクロフォンを用いていない MVDR ビームフォーマと maxSNR ビームフォーマおよび、ヴァーチャルマイクロフォンを用いた maxSNR ビー



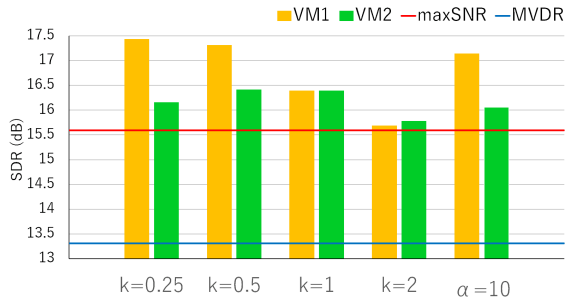
(a) driver's speech enhancement(close window)



(b) driver's speech enhancement(open window)



(c) passenger's speech enhancement(close window)



(d) passenger's speech enhancement(open window)

Fig. 6: results of speech enhancement.

ムフォーマの結果を比較した。実験では、劣決定条件下での車室内の三角マイクロフォンアレイにおいても音声強調性能を改善するためにヴァーチャルマイクロフォンが有効であることがわかり、特に適切な k を用いた WPVM 配置において高い強調性能を示した。

謝辞　本研究は科研費 19H04131，戦略的基盤技術高度化支援事業の助成を受けた。

参考文献

[1] R. Landgraf et al., Gerhard Schmidt. "Can you hear me now? reducing the lombard effect in a driving car using an in-car communication system," In Proceedings of the Speech Prosody, pp. 479–483, 2016.

[2] A. Theiss et al., "Instrumental evaluation of in-car communication systems," In Proc. Speech Communication; 11. ITG Symposium, pp. 1–4, VDE, 2014.

[3] R. Jinzai et al., "Wavelength proportional arrangement of virtual microphones based on interpolation/extrapolation for underdetermined speech enhancement," In Proc. EUSIPCO, pp. 1–5, 2019.

[4] H. Segawa et al., "Applying Virtual Microphones to Triangular Microphone Array in in-Car," in Proc. APSIPA, pp. 421-425, 2020.

[5] H. Katahira et al., "Nonlinear speech enhancement by virtual increase of channels and maximum SNR beamformer," EURASIP Journal on Advances in Signal Processing, Vol. 2016, No. 1, pp. 1–8, 2016.

[6] K. Yamaoka et al., "Performance evaluation of nonlinear speech enhancement based on virtual increase of channels in reverberant environments," In Proc. EUSIPCO, pp. 2324–2328, 2017.

[7] O. Yilmaz et al., "Blind separation of speech mixtures via time-frequency masking," IEEE Transactions on signal processing, Vol. 52, No. 7, pp. 1830–1847, 2004.

[8] R. Jinzai et al., "Microphone position realignment by extrapolation of virtual microphone," In Proc. APSIPA, pp. 367–372, 2018.

[9] B. Moore., "An introduction to the psychology of hearing," 1997.

[10] J. Blauert., "Spatial hearing: the psychophysics of human sound localization," MIT press, 1997.

[11] H.L van Trees., "Optimum array processing," 2002.

[12] S. Araki et al., "Blind speech separation in a meeting situation with maximum snr beamformers," In Proc. ICASSP, Vol. 1, pp. 41-44, 2007.

[13] O. Lamont Frost. "An algorithm for linearly constrained adaptive array processing," Proceedings of the IEEE, Vol. 60, No. 8, pp. 926–935, 1972.

[14] E. Vincent et al., "First stereo audio source separation evaluation campaign: data, algorithms and results," In Proc. International Conference on Independent Component Analysis and Signal Separation, pp. 552-559, 2007.

[15] A. Kurematsu et al., "ATR Japanese speech database as a tool of speech recognition and synthesis," Speech communication, vol. 9, no. 4, pp. 357–363, 1990.