

複素対数補間を用いたヴァーチャル多素子化マイクロホンアレーの 周波数依存素子配置最適化*

☆片平拓希 (筑波大), 小野順貴 (NII/総研大),
宮部滋樹, 山田武志, 牧野昭二 (筑波大)

1 はじめに

マイクロホンアレー信号処理は、複数のセンサ (マイクロホン) により観測される音声信号から、様々な処理を行うものであり、空間上の音源の位置を推定する音源定位、複数の音声信号が混合された観測信号から個々の音源を推定・分離する音源分離など、目的に応じて様々な手法が研究されている。

我々はこれまでに、実際にはマイクロホンが存在しない位置での観測信号を、実際の観測信号の非線形補間で推定し、疑似的にマイクロホンアレーを多素子化する手法を提案した [1]。これにより、線形アレー信号処理の枠組を劣決定条件に拡張し、音声強調性能に改善が見られることを確認した。

我々が提案したマイクロホンアレーのヴァーチャル多素子化は、単に素子数を疑似的に増加させるだけでなく、様々な配置を柔軟かつ容易に設計できるといった利点が考えられる。また、周波数ごとに独立した処理であるため、周波数ごとに異なるヴァーチャル素子配置の設計も理論上可能である。本稿では、複素対数補間による観測信号のヴァーチャル多素子化を応用し、周波数ごとに異なる配置を持つヴァーチャルマイクロホンアレーの設計を試みる。具体的には、評価値 SDR (Signal-to-Distortion Ratio) を基準とし、SN 比最大化ビームフォーマにおいて、周波数ごとにどのような配置が最適となるかを調査し、また、音声強調性能がどの程度向上し得るかを検証する。

2 複素対数補間によるヴァーチャル多素子化

本節では複素対数補間によるマイクロホンアレーのヴァーチャル多素子化のアルゴリズム [1] について述べる。ここでは、2 素子のマイクロホンアレーによる観測信号 $\mathbf{x}(\omega, t) = [x_1(\omega, t) \ x_2(\omega, t)]^T$ を、疑似的に多素子化し、多チャンネルの観測信号として扱うことを考える。なお、 x_j が j 番目の素子による観測信号、 ω が周波数、 t が時間フレーム番号であり、本稿では信号を短時間フーリエ変換による時間周波数領域で表すものとする。ヴァーチャル素子による観測値 v は観測信号を非線形写像 f のドメイン上で補間する事により求められる。1 番目のマイクロホンの位置を 0、2 番目のマイクロホンの位置を 1 としたとき、2 つのマイクロホンを結んだ直線上における座標 α でのヴァーチャル観測値を次のように表す。

$$v(x_1, x_2, \alpha) = f^{-1}\{(1 - \alpha)f(x_1) + \alpha f(x_2)\} \quad (1)$$

ここでは、補間を行うドメインとして観測値の複素対数を選び、ヴァーチャル素子での観測値を以下のよう表す。

$$v(x_1, x_2, \alpha) = \exp\{(1 - \alpha)\log|x_1| + \alpha\log|x_2|\} + j\{(1 - \alpha)\angle x_1 + \alpha\angle x_2\} \quad (2)$$

3 周波数依存配置ヴァーチャル多素子化マイクロホンアレー

本節では、前節で述べたヴァーチャル多素子化アレーを、周波数依存素子配置に拡張する。これまで我々が提案したヴァーチャル多素子化アレーでは、すべての周波数にわたって統一した配置を用いている。しかし、ヴァーチャル素子による観測値は周波数ごとに独立して計算されるため、周波数ごとに異なるヴァーチャル素子配置を考慮して拡張することは容易である。周波数ごとにヴァーチャル多素子化された観測信号を以下のように表す。

$$\hat{\mathbf{x}}(\omega, t) = [\hat{x}_1(\omega, t), \dots, \hat{x}_{\widehat{M}(\omega)}(\omega, t)]^T \quad (3)$$

ここで、 $\widehat{M}(\omega)$ は周波数 ω における多素子化後の素子数である。 $\hat{x}_j(\omega, t)$ は j 番目のヴァーチャル素子の観測値であり、周波数 ω における j 番目の素子の位置を $\alpha_j(\omega)$ として以下のように表される。

$$\hat{x}_j(\omega, t) = v(x_1(\omega, t), x_2(\omega, t), \alpha_j(\omega)) \quad (4)$$

なお、 $\alpha_j(\omega) = 0$ のとき $\hat{x}_j(\omega, t) = x_1(\omega, t)$ 、 $\alpha_j(\omega) = 1$ のとき $\hat{x}_j(\omega, t) = x_2(\omega, t)$ となり、それぞれ実素子による観測値と同一になる。また、 $0 < \alpha_j < 1$ のときヴァーチャル素子の内挿、 $\alpha_j < 0$ および $1 < \alpha_j$ のときヴァーチャル素子の外挿となる。これまで、ヴァーチャル素子の内挿のみを扱ってきたが、本稿では外挿を含めた様々な素子配置を扱うものとする。

4 実験

4.1 SN 比最大化ビームフォーマ

本実験では、提案手法を SN 比最大化ビームフォーマ [2, 3] に適用し、性能の検証を行う。SN 比最大化ビームフォーマでは、時間周波数領域のマルチチャンネルフィルタ $\mathbf{w}(\omega)$ により観測信号 $\mathbf{x}(\omega, t)$ 中の非目的音・雑音を抑圧し、目的信号のみを強調した信号 $y(\omega, t)$ を得る。

$$y(\omega, t) = \mathbf{w}(\omega) \mathbf{x}(\omega, t) \quad (5)$$

フィルタ $\mathbf{w}(\omega)$ は、目的音のみの区間、および非目的音・雑音のみの区間の観測信号を事前情報とし、信号の SN 比 $\lambda(\omega) = \frac{\mathbf{w}(\omega)\mathbf{R}_T(\omega)\mathbf{w}^H(\omega)}{\mathbf{w}(\omega)\mathbf{R}_I(\omega)\mathbf{w}^H(\omega)}$ を最大化するように設計される。ここで、 \mathbf{R}_T は目的信号区間、 \mathbf{R}_I は非目的信号区間それぞれの共分散行列である。 $\lambda(\omega)$ を最大化する $\mathbf{w}(\omega)$ は、以下の一般化固有値問題の最大固有値に対応する固有ベクトルに相当する。

$$\mathbf{w}(\omega)\mathbf{R}_T(\omega) = \lambda(\omega)\mathbf{w}(\omega)\mathbf{R}_I(\omega) \quad (6)$$

4.2 実験条件

図 1 に示す条件で音声強調の実験を行った。実マイクロホンは 2 素子とした。目的音は正面を 0° として、時計回りに $+10^\circ$ の方向とし、雑音は -90° から 90° まで 20° 刻みの方向のうち目的音方向以外の 9 方

*Constructing virtual microphone array via complex logarithmic function with subband-optimized layout. by Hiroki KATAHIRA (University of Tsukuba), Nobutaka ONO (National Institute of Informatics / The Graduate University for Advanced Studies), Shigeki MIYABE, Takeshi YAMADA, Shoji MAKINO (University of Tsukuba)

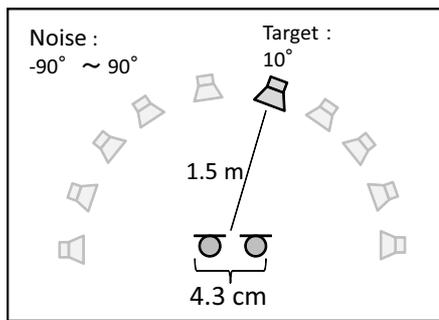


Fig. 1 Experimental condition

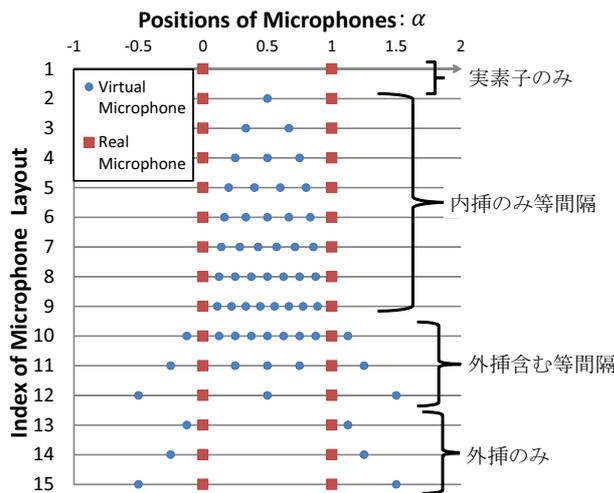


Fig. 2 List of microphone layouts

向とした。このように少ない素子のアレイで多数の方向から到来する雑音を抑圧する処理は、一般に大変困難であることに注意する。雑音、目的音ともに日本語の発話音声(重点領域研究「音声言語」・試験研究「音声DB」連続音声データベース(PASL-DSR)[4])を使用した。残響時間は130msの環境で測定したインパルス応答を各音声に畳み込み、シミュレーションによって観測信号を作成した。SN比は0dBとなるように調整した。また、観測信号のうち、目的音のみの区間と非目的音のみの区間それぞれ5分間を用いてビームフォーマの学習を行った。20秒の混合音声に適用し性能評価を行った。なお、ヴァーチャル素子数を増やした際、共分散行列のランク落ちによりビームフォーマ出力に歪が発生する[1]。このランク落ちによる歪を低減するため、ここでは長めの学習区間を用いたことを注記する。サンプリング周波数は16kHz、短時間フーリエ変換のフレーム長が1024サンプル、フレームシフト幅が256サンプルとなっている。

本実験では、周波数ごとの評価値を基準に周波数依存配置の選択を行った。まず、図2に示した各配置を用いて実際に音声強調を行い、周波数ごとにSDRを算出した。続いて、各周波数について最も高いSDRを示した配置を、その周波数での最適配置として選択し、各周波数における最適配置について分析を行った。また、帯域ごとに最適配置を用いた周波数依存配置マイクロホンアレイにより音声強調を行い、性能の向上を確認した。

4.3 結果と考察

図2の各配置、および周波数ごとに最適配置を用いた場合のSDRを図3に示す。全周波数で統一した配置を用いた場合、実素子のみの配置(配置1)に比べSDRは1dB程度の向上にとどまるものの、周波

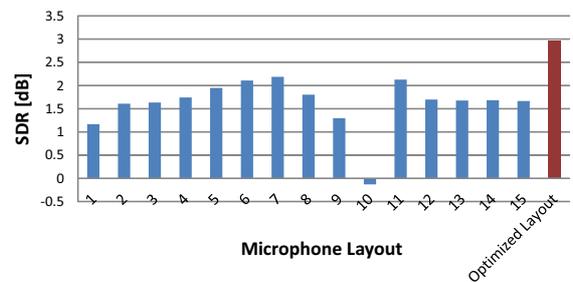


Fig. 3 SDR for each microphone layout

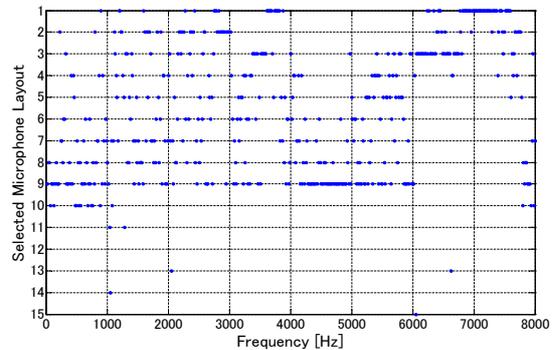


Fig. 4 Optimized microphone layout for each frequency

数ごとに最適な配置を選択することにより、上昇幅は2dB程度まで増加する。また、各周波数で選択された最適配置を図4に示す。一般に低域では広範囲に素子を外挿した全長の長い配置が有利であると考えられるが、このような配置が選択される帯域は殆どなく、このような配置が選択される帯域は殆どなく、広範囲に渡る素子の外挿は難しいものと考えられる。低域では比較的多数のヴァーチャル素子を用いた配置が選ばれるのに対して、約5~6kHz以上の高域においては、徐々に少ないヴァーチャル素子の配置が多く選ばれるようになってきている。実素子間隔4.3cmの条件では約4kHz以上の帯域で、空間的エイリアシングにより補間が正しく行われなかった場合があり、この帯域でのヴァーチャル素子の性能が低下していることが考えられる。

5 まとめ

本稿では、複素対数補間による観測信号のヴァーチャル多素子化を応用し、周波数ごとに異なる配置を持つヴァーチャルマイクロホンアレイについて検証を行った。またSN比最大化ビームフォーマによる音声強調性能を検証したところ、SDRが上昇し、性能の向上が確認できた。今回設計した最適配置は、特定の音源方向、雑音環境、素子間隔で最適化したものであるが、今後は音源位置や雑音条件に依存しない周波数依存配置の設計を目指し、さらに詳細な分析を進める予定である。

参考文献

- [1] 片平, 他, “複素対数補間によるヴァーチャル観測に基づく劣決定条件での音声強調” 音講論(春), 741-744, 2013.
- [2] Van Trees, ed., “Optimum Array Processing,” Wiley, 2002.
- [3] 荒木, 他, “話者分類とSN比最大化ビームフォーマに基づく会議音声強調,” 音講論(春), 571-572, 2007.
- [4] 板橋, “文部省「重点領域研究」による音声データベース,” 日本音響学会誌, 48(12), 894-898, 1992.