車載向け学習型マイクロホンアレー技術

Microphone Array Technique for Automotive Applications

天田　皇
■ AMADA Tadashi

東芝は，自動車内での音声認識やハンズフリー通話向けのノイズキャンセラとして，マイクロホンアレー技術の研究を行っている。従来型のマイクロホンアレーは車内残響下で性能劣化を起こすことがあったが，今回開発した方法は，対象とする残響空間で事前学習を行うことで，その空間に最適化された指向特性を獲得し，残響に頑健な性能を実現している。実車を用いた音声認識と妨害音抑圧の実験を行い，車内残響下において良好な性能が得られることを確認した。

Toshiba has been developing a microphone array technique for application as a noise canceller for speech recognition and hands-free communication in automotive environments. To cope with performance degradation of conventional microphone arrays in the reverberant environment in a car cabin, we have newly developed a robust method enabling acquisition of directional characteristics by means of offline learning methods and achievement of optimized performance in the target reverberant room. Experiments on speech recognition rates and noise suppression capabilities in real car environments showed that the proposed method achieves successful performance under reverberant conditions in cars.

1　まえがき

マイクロホンアレーは，雑音環境下で目的とする話者の声を抽出する音声強調技術として有効な手法であり，音声認識やハンズフリー通話などへの応用を目的とした研究が盛んである。

その方式は固定型アレーと適応型アレーに大別され，前者は十分な指向性を得るために多くのマイクロホンを必要とするのに対し，後者は妨害音に対し少数のマイクロホンで高い抑圧性能を発揮する。そのため，適応型アレーは低コスト化や小型化が可能であり，車内のような狭い空間にも設置しやすく，実現上有利である。しかし，適応型アレーは残響の影響を受けやすく，条件によっては本来残すべき話者の声が部分的に抑圧されてしまう現象（目的音除去）により，音声にひずみを生じる場合がある。これは，車内などの実環境でマイクロホンアレーを用いる場合には深刻な問題となる。

東芝は，残響下でも目的音除去が起きにくいマイクロホンアレー方式を開発した。開発した方式（以下，提案法と記す）はアレー重みを解析的に求めるのではなく，事前に残響下で学習された重み辞書から動作時に最適な重みを選択することで，残響に対するロバスト性を実現している。

ここでは，提案法の概略と，計算機シミュレーションによる指向特性の検証，実車収録音声に対する音声認識実験の結果について述べる。

2　マイクロホンアレーと残響

目的音除去の原因としては，残響のほかに，マイクロホンアレーにより強調される音源方向（目的音方向）と実際に話者の声が到来する方向との誤差の問題や，マイクロホン素子間の特性変動なども挙げられる。

その中でも残響は，アレーが使用される環境によってもたらされる問題であるため，対策が難しい。残響は空間で音が繰り返し反射することによって生じ，反射波はアレーに対し様々な方向から到来して直接波と干渉する。その結果，本来は目的音方向からだけ到来するはずの話者の声が，様々な方向から到来することになり，観測される到来方向に変動をきたす。適応型アレーは，一般に目的音方向に対する感度が高いため，残響により到来方向がずれると急激に感度が下がり，目的音除去につながる。

先行研究として，音源からマイクロホンまでの伝達関数を利用した方法[1]が挙げられるが，伝達関数が既知であるという前提は，音源位置が完全には定まらない車載用途には向かない。残響を推定し除去する研究も近年行われているが，推定できたとしても，安定な逆フィルタを実時間で構成しなければならないという問題が残る。また，目的音区間で適応フィルタの更新を停止する方法[2]は有効であるが，残響下で妨害信号と目的音を区別することは必ずしも容易ではない。

3　学習型アレー

3.1　アルゴリズム

任意の残響下で事前情報を用いずにアレー重みを解析的に求めることは困難と考え，提案法では，残響環境をある程度制限し，その条件下で動作するアレーの実現を試みる。具体

的には，残響下で事前学習した重み辞書からアレー重みを選択して用いる構成とし，重みの選択は受音信号から生成されるチャネル間特徴とアレー重みの対応付けにより行う。アレー重みの学習やチャネル間特徴との対応付けを限定された残響下で行うことで，残響の影響を直接扱わずに目的音除去を回避することができる[3]。

提案法のブロック図を**図1**に示す。二つのマイクロホンの入力信号$x_1(t)$，$x_2(t)$は，離散フーリエ変換（DFT）により周波数成分$X_1(k)$，$X_2(k)$にそれぞれ変換される。ここで，tは時間インデックス，kは周波数成分を表すインデックスである。$X_1(k)$，$X_2(k)$を周波数成分kごとにフィルタリングし，離散逆フーリエ変換（IDFT）することで出力信号$y(t)$を得る。フィルタリングにおいて，出力$Y(k)$は，次式で表される。

$$Y(k) = W_1^*(k)A_1(k)X_1(k) + W_2^*(k)A_2(k)X_2(k) \quad (1)$$

$A_1(k)$，$A_2(k)$は，走行雑音などの定常雑音を抑圧するプレフィルタの係数[3]，$W_1(k)$，$W_2(k)$は，主にアレーの指向性を制御する複素数の係数である。＊は共役複素数を表す。

アレー重みを決定する重み算出部（FW）の構成を図1の下段に示す。まず，X_1，X_2からチャネル間の信号の関係を表す特徴量fが算出される。次に，事前学習で求めておいた複数の代表的な特徴量の中から，算出された特徴量との距離を最小とする候補を選ぶ量子化が行われ，そのインデックスiが出力される。特徴量として，今回はコヒーレンス（信号間の関連の度合い）と一般化相互相関関数を用いた[3]。選択部では，インデックスに対応する$W_1(k)$，$W_2(k)$が重み辞書から選択され出力される。重み辞書に格納されている係数は，次節で述べる事前学習によりインデックスごとに最適化されている。

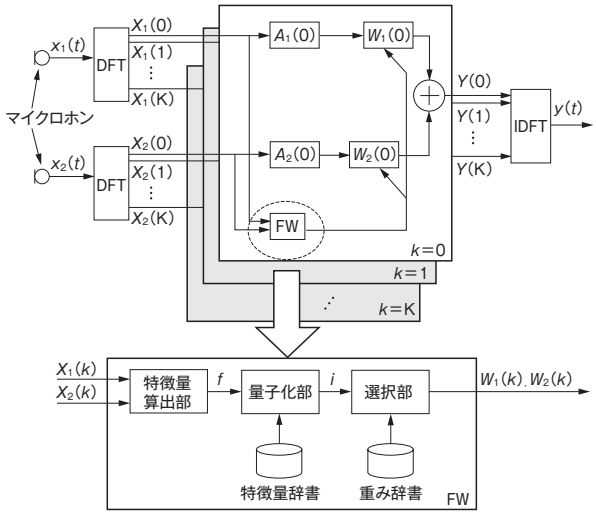

図1. 提案法のブロック図 — 入力信号$x_1(t)$，$x_2(t)$に事前学習されたアレー重み（$W_1(k)$，$W_2(k)$）で重みづけて出力信号$y(t)$を得る。

Block diagram of newly developed method

これらの処理は周波数ごとに行われ，全周波数（K+1成分）の出力$Y(0)$～$Y(K)$を得た後，これらをフーリエ逆変換することで出力信号$y(t)$を得る。

3.2 学習方法

アレー重みの事前学習の方法について述べる。事前に対象とする残響環境下で，音源位置を変えながら生成したチャネル間特徴をLBG（Linde-Buzo-Gray）アルゴリズムなどでクラスタリングし，各クラスタにアレー重みを対応付ける。

アレー重みの学習はクラスタごとに行う。ここでは理想的な出力信号$S(l)$（学習時には既知）とアレー出力の2乗誤差J_iの最小化により求める。

$$J_i = \sum_{l \in C_i} \left(\boldsymbol{W}_i^H \boldsymbol{X}(l) - S(l) \right)^2 \quad (2)$$

ただし，$\boldsymbol{W}_i = (W_{i1}(k), W_{i2}(k))$はアレー重み，$\boldsymbol{X} = (X_1(k), X_2(k))$は入力信号の周波数成分，$l$はフレーム番号，$C_i$はクラスタ$i$に所属するフレーム番号の集合，$H$は共役転置である。以上の処理は周波数ごとに独立に行う。なお，この学習は実際に使用する環境で行うのが最良であるが，今回は計算機内で生成した仮想残響空間を用いて行った。

3.3 指向特性

提案法の指向性を測定した。測定は学習時と同一の仮想残響空間で行い，マイク間隔は15 cm，音源とマイク間の距離は0.8 mである。評価に用いた残響空間のサイズと音源位置は**図2**のとおりであり，図は平面図で，Hは高さを示す。

空間内で音源位置を変えながらアレーの入力と出力の比を測定した。**図3**は残響なしの場合，**図4**は残響時間（RT）が50 msの場合における，アレーの入出力比と角度との関係をプロットしたものである。提案法の学習はいずれの場合もRT=50 msで行った。比較対象として固定型アレーの代表である遅延和アレー（DS法）[4]と適応型アレーとしてよく用いられるGriffiths-Jim型アレー（GJ法）[4]の値も示す。プレフィルタは，比較対象と条件をそろえるため，この実験では用いてい

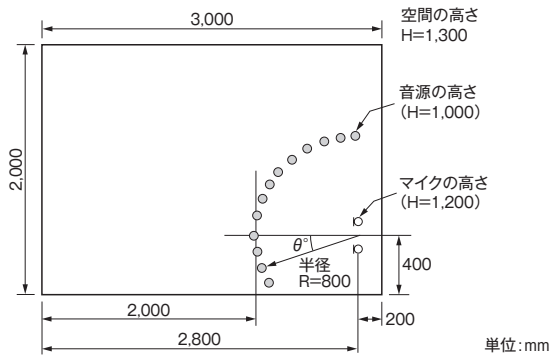

図2. 仮想残響空間 — 音源位置（○）を変えながらマイクロホンアレーの入出力比を測定した。

Virtual reverberant room

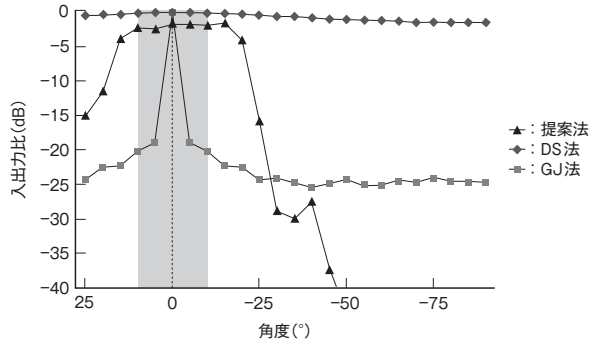

図3. 指向特性（残響なし）— 音源位置とアレーの入出力比の関係を示す。Griffiths-Jim型は，急しゅんな指向特性を示している。

Directional characteristics (anechoic)

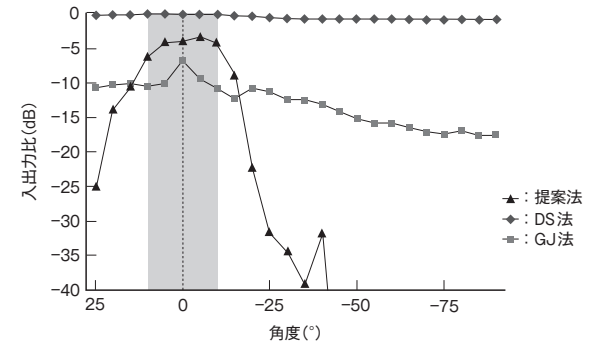

図4. 指向特性（RT=50 ms）— 音源位置とアレーの入出力比の関係を示す。提案法は，残響環境でも良好な指向特性を示している。

Directional characteristics (reverberation time (RT): 50 ms)

ない。図中の帯状の区間が入力信号を強調する区間，それ以外が抑圧する区間である。

DS法は，残響の有無にかかわらず，強調と抑圧の差が小さい。これはマイク数が2と少ないためである。一般に，遅延和で鋭い指向性を得るには多くのマイクが必要である。GJ法は，残響なしの場合には強調区間の中央で鋭い指向性を形成しているが，RT=50 msの場合はピークが下がり，抑圧区間でも抑圧量が低下している。ピークが下がる現象は目的音が削られていることを意味し，適応型のアレーの弱点である目的音除去が発生していることがわかる。提案法は，残響の有無によらず強調と抑圧の差が20 dB以上あり，良好な指向特性が得られている。残響なしでビーム幅が広がっているが，学習時の条件と一致するRT=50 msでは帯状区間を外れると減衰が始まり，設計どおりの動作となっている。

4　実験

4.1　実験条件

提案法を自動車内で収録したデータを用いて評価した。運転席又は助手席から発声された音声データに，実走行で収録した雑音を重畳し評価データとした。強調すべき音声の品質を評価するため運転席音声の認識率を，また，妨害音の抑圧能力を評価するため，助手席音声の抑圧量を測定した。音声認識実験では従来型アレーとの比較実験と，プレフィルタの効果を確認する実験の2種類を行った。アレーは車内中央のルームミラー付近に設置し，マイク数は2とした。

学習では，計算機シミュレーションで生成した車内とほぼ同じサイズの仮想残響空間内で，音源の位置を変えながら生成した音声データに，評価時とは異なる車種の雑音データを重畳した学習データを用いて運転席付近からの信号を強調し，それ以外からの信号を抑圧するように学習した。音声認識にはMFCC（Mel-Frequency Cepstrum Coefficient）とその動的特徴（Δ（1次の変動成分）とΔΔ（2次の変動成分））を音声特徴とするTriphone音響モデル(注1)を用い，認識語いは100都市名とした。

4.2　運転席の音声認識率

アレー方式の比較実験結果を**図5**に示す。横軸は走行条件の違いであり，高速走行（highway），市街地走行（city），信号待ち（idle），及びエンジン停止（clean）の4条件である。縦軸は，100都市名を発声したときの認識率である。評価に用いた方式は，提案法のほかに，1チャネルの生データ（無処理），これにMMSE（Minimum Mean Square Error）[5]による雑音抑圧を施した方式（MMSE法），及び代表的なアレーであるDS型とGriffiths-Jim型アレー（GJ1法，GJ2法）である。無処理とMMSE法は，左右のチャネルのうち性能が良いほうを選んだ。GJ1法は常に適応フィルタの更新を行い，GJ2法は事前に指定した音声区間で適応を止める設定とした。アレーの重み辞書などの設定は，走行条件によらず同一とした。提案法のプレフィルタは，ほかのアレーとの比較のためここでは用いていない。音声データには，車内で測定したインパルス応答を無残響の音声データに畳み込む“畳込みデータ”を用いた。音声認識の音響モデルは，雑音のない環境で学習したクリーンモデルを用いた。

DS法，GJ1法，GJ2法は無処理と同等以下の認識率であり，アレーとしての効果が得られていない。特にGJ1法は劣化が著しく，目的音除去により音声にひずみが生じていることが推察される。GJ2法は音声区間で適応を停止することにより，目的音除去の問題は回避できていると考えられる。提案法はこれらより高い認識率を示し，残響下での効果が確認できる。しかし，MMSE法と比較すると同等程度の性能である。これは，走行雑音のように音源方向が特定できない雑音（拡散性雑音）に対して，音源方向を重要な手がかりとするアレー処理は，その効果を十分発揮できないためと考えられる。

（注1）　前後の音素を考慮した三つ組音素の特徴を表したモデル。

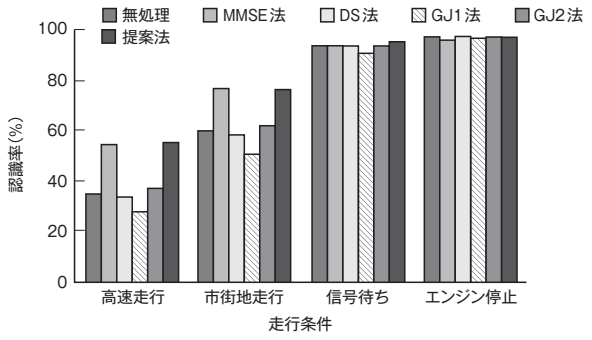

図5. 運転席の音声認識率（アレー方式）— 提案法以外のマイクロホンアレーは無処理と同等以下の性能であり，マイクロホンアレーの効果が発揮できていない。

Results of experiments on speech recognition rates of array method

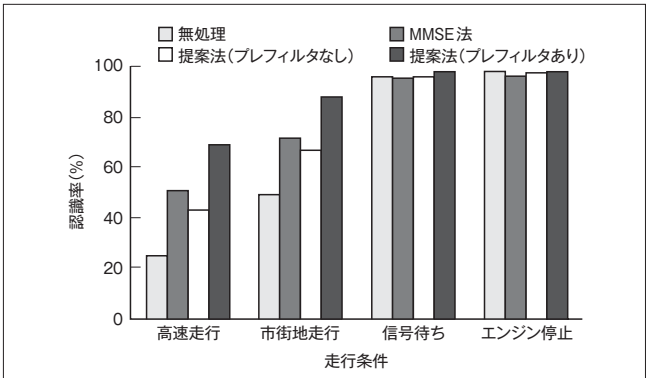

図6. 運転席の音声認識率（プレフィルタの効果）— 提案法は，1チャネルのMMSE法よりもすべての条件で認識率が高い。

Results of experiments on speech recognition rates with prefilter effect

プレフィルタは，このような拡散性雑音を前処理で除く目的で導入されたものである。プレフィルタを用いた場合の認識率を**図6**に示し，比較対象として，プレフィルタなしとMMSE法の結果も示す。この実験では車内で実収録した音声データを用いた。また，音声認識の音響モデルは方式ごとに再学習を行った。実験結果から，提案法はすべての走行条件でMMSE法を上回り，走行雑音に対しても効果が確認できた。

4.3　助手席音声の抑圧量

アレーの特徴は，信号の到来方向を利用した雑音抑圧である。この性能を確認するため，助手席の音声（畳込みデータ）の抑圧能力を測定した。その結果を**図7**に示す。ここで，抑圧量はアレーの出力パワーを入力パワー（重畳雑音なし）で割った値と定義し，値が小さいほど大きな抑圧を意味する。提案法は助手席の音声を抑圧できており，運転席方向以外の音を抑圧する事前学習が正しく機能していることが確認できる。GJ法は，クリーン条件で10 dB程度の抑圧であり，提案法より8 dB以上抑圧能力が低い。DS法は，マイク数が2ではほとんど抑圧効果がない。なお，無処理とMMSE法はアレーではないため，音源方向に基づく処理はできない。

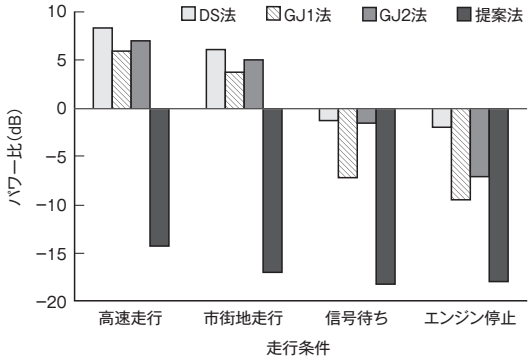

図7. 助手席音声の抑圧量 — 提案法は，ほかの方式に比べて高い抑圧性能を示す。

Results of experiments on passenger speech suppression performance

5　あとがき

残響下における適応型アレーの問題点である目的音除去を解決する一手法を提案した。提案法は，事前学習により残響の影響を考慮したアレーを構成できる点が特徴である。計算機シミュレーションにより指向特性の評価を行い，残響下でも目的音方向に対して感度を維持し，妨害音方向に対しては高い抑圧性能を発揮できることを確認した。

また，実車データによる評価を行い，提案法は運転席音声に対してMMSE法を上回る認識率を示し，助手席音声に対しては，遅延和アレーより15 dB以上高い抑圧性能を示すことを確認した。今後は，車載音声認識の前処理やハンズフリー通話用ノイズキャンセラとして，製品化を行う予定である。

文　献

(1) Flanagan, J.L., et al. Spatially Selective Sound capture for speech and audio processing. Speech Communication. **13**, 1-2, 1993, p.207－222.
(2) Harrison, W. A., et al. A New Application of Adaptive Noise Cancellation. IEEE Trans. ASSP. **34**, 1, 1986, p.21－27.
(3) 天田　皇．“重み選択型マイクロホンアレーの自動車走行雑音に対する音声認識率の改善”．日本音響学会講演論文集（秋）．山梨，2007-09，日本音響学会．2007，p.159－160.
(4) 大賀寿郎，ほか．音響システムとディジタル処理．電子情報通信学会．1995，265p.
(5) Ephraim, Y., et al. Speech Enhancement Using a- Minimum Mean Square Error Short-Time Spectral Amplitude Estimator. IEEE Trans. ASSP. **33**, 2, 1984, p.443－445.

天田　　皇　AMADA Tadashi
研究開発センター　マルチメディアラボラトリー研究主務。
音響信号処理，音声符号化の研究・開発に従事。電子情報通信学会，日本音響学会，IEEE会員。
Multimedia Lab.