

音声認識のためのマイクロホンアレー技術

Microphone Array Technique for Speech Recognition

天田 皇 金澤 博史

■ AMADA Tadashi

■ KANAZAWA Hiroshi

実環境下で音声認識を行う場合、周囲の雑音により認識率が低下する問題がある。東芝では、音声認識の前処理として2本のマイクロホンを用いたマイクロホンアレーによる雑音抑圧技術の研究開発を行っている。この技術は目的音の追尾機能により、あらかじめ設定した範囲から到来する目的音のみを強調し、それ以外の方向から到来する信号を抑圧することができる。また、独自の2チャンネルスペクトルサブトラクション法により、暗騒音のような拡散性雑音も抑圧が可能である。

Speech recognition techniques in the real world suffer from the problem of environmental noise degrading speech recognition rates. Toshiba is developing a noise reduction technique that involves using a microphone array with two microphones for the speech recognition front end. This is able to enhance target speech coming from a predefined direction range and reduce signals coming from other directions. We have also developed a new 2-channel spectral subtraction method in order to reduce diffuse noise such as background noise.

1 まえがき

オフィスや家庭内などホームロボットの活躍が期待される実環境では様々な雑音が存在する。このような環境下で、ホームロボットにユーザーの発話を正しく認識させるためには、音声認識の前段で雑音を除去する技術が重要であり、近年、複数のマイクロホンを用いて雑音抑圧を行うマイクロホンアレー技術が注目されている。

ここでは、東芝が開発した2チャンネル話者追尾型マイクロホンアレー技術について述べる。

2 雑音抑圧の方法

指向性マイクロホン(単一指向性マイクロホン、ガンマイクロホンなど)を用いて特定の方向の音声を強調する方法はもっとも簡単な雑音抑圧方法の一つである。このような物理現象を利用したアナログ方式に対し、受信した信号をデジタル的に処理することで音声を強調する技術の研究も行われている。

スペクトルサブトラクション⁽¹⁾はその一つであり、信号の時間情報に基づく手法であるため、1マイクロホンで効果的な抑圧が可能である。また、実現が容易であるため現在広く用いられている。

これに対し、複数のマイクロホンを用いた手法(マイクロホンアレー)の研究も行われている。マイクロホンアレーは、音源から発せられた信号が各マイクロホンに到達するときに

生じる位相差などの空間情報を用いて雑音を抑圧する技術であり、1マイクロホンの手法に比べて雑音抑圧能力は高い。マイクロホンアレーは、計算量が増加するという問題があるが、マイクロプロセッサの処理能力の向上により、複雑な方式でも実時間動作が可能になってきた。

マイクロホンアレーは、大きく分けて固定型と適応型に分類される。固定型は遅延和アレー⁽²⁾に代表される方式で、事前にフィルタの特性を決定する。そのため、特性が安定している利点があるが、十分な抑圧性能を得るには多くのマイクロホンが必要となり、システム全体の規模が大きくなる問題がある。

これに対し適応型アレーは、例えばGriffith-Jim型アレー^{(2),(3)}に代表されるように、少ないマイクロホンを用いて、大きな抑圧性能を実現できる利点があるが、フィルタの特性を周囲の音声や雑音に基づき適応的に更新するため、想定された環境からのずれにより性能の劣化が起きる問題がある。

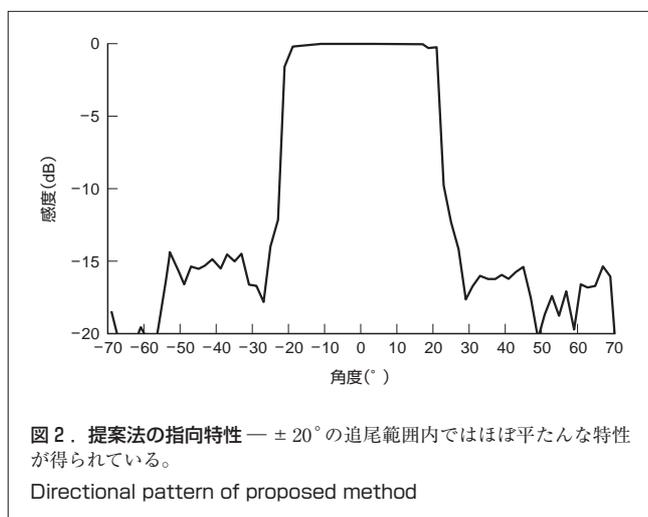
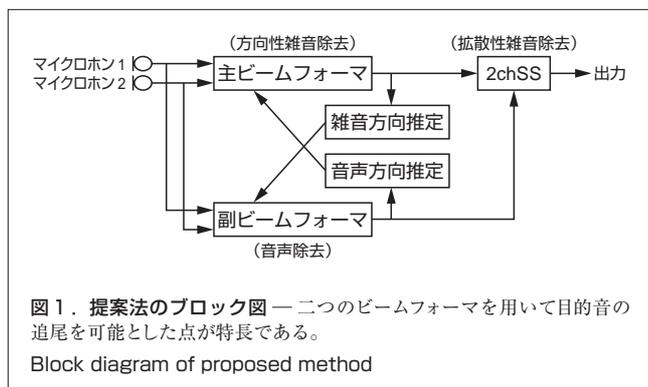
実環境で人の声を処理する場合、①話者の動きにより音源が一定でない、②音源が理想的な点音源ではない、③周囲の反射の影響を受ける、などの問題を考える必要がある。前述のGriffith-Jim型のマイクロホンアレーではこれらの問題は最終的には目的信号の除去という形で現れる。つまり、本来強調されるべき目的信号の一部が抑圧され、ひずみが生じる。

目的信号除去の問題に対するロバスト化手法として様々な方法^{(4),(5)}が提案され検討されている⁽⁶⁾。これらのロバスト化手法は、適応可能な範囲が狭いことや少数マイクロホンに

は不向きであるなどの問題があり、小規模でありかつ実用に耐えうるロバストな性能を実現することができなかった。

3 話者追尾型マイクロホンアレー

前述の目的音方向の変動に対するロバスト化を少数のマイクロホンを用いて実現する方法として、当社は2チャンネル話者追尾型マイクロホンアレーを提案している^{(6),(7)}。提案法のブロック図を図1に示す。提案法は従来の適応型のビームフォーマ(図1の主ビームフォーマ)に加え、副ビームフォーマを用いている点が特長である。二つのビームフォーマはいずれもGriffith-Jim型の適応型アレーを基本にしている。主ビームフォーマは目的音声を強調するように動作する。一方、副ビームフォーマは目的音声以外の方向の信号を強調するように動作し、その結果として目的音声方向に死角が形成される。この死角の位置を副ビームフォーマの伝達特性から求めることで、目的音声の到来方向を推定することができる。この推定結果を用いて主ビームフォーマの目的音声方向を逐次更新することで、目的信号の除去の問題を回避することができ、目的音方向の変動に対してロバストな特性を得る



ことができる。

提案法の指向特性を図2に示す。図2の横軸は目的信号の到来方向(角度)、縦軸が感度(出力パワー)を示す。追尾範囲は事前に設定する必要があり、今回は正面方向を0°とし±20°の範囲を追尾範囲とした。図2から、追尾範囲内では目的信号の到来方向が常に補正されているため、目的信号の除去は起きず、ほぼ平坦な特性が得られているが、追尾範囲を外れると急激に抑圧が始まるようすが確認できる。

4 拡散性雑音の抑圧

実環境に存在する雑音は、その分類方法の一つとして、方向性雑音と拡散性雑音に分類することができる。前者は特定の方向から到来する信号で、複数のマイクロホンで受信した場合、それぞれの信号に位相差を伴って観測される。つまり、類似した波形が到来方向に応じた時間だけずれた形になる。これに対し拡散性の雑音は暗騒音のように、音源が1方向に特定できない雑音であり、マイクロホン間の信号の相関が小さい。マイクロホンの本数が少ない場合Griffith-Jimなどの適応型アレーで大きく抑圧できるのは方向性雑音であり、拡散性雑音の抑圧能力は低い。

そこで、提案法ではビームフォーマの後処理として2チャンネルスペクトルサブトラクション(2chSS)を用い、拡散性雑音に対する抑圧性能を向上させている⁽⁸⁾。2chSSは従来の1チャンネルのSS⁽¹⁾を基本にしているが入力信号が二つある点が異なる。

通常のSSは、音声と雑音が混在した1チャンネルの入力信号から、なんらかの方法で雑音区間を判定し、その区間のスペクトルの平均化処理により推定雑音スペクトルを生成し、これをスペクトル領域で減算することで雑音を抑圧している。このため、雑音が定常な場合は、高い抑圧性能を発揮する。

SSの問題点の一つとして音声成分が推定雑音スペクトルに混入する問題がある。この問題は音声区間を誤って雑音区間と判定した場合に生じ、音声成分が混入した推定雑音を入力信号から減算するため、出力音声から音声成分が削られ、ひずみが生じる。

また、非定常雑音(ドアを閉める音など)の抑圧が困難であるという問題もある。非定常雑音は、パワー変動などの簡単な手法では目的音声と区別が難しく、音声信号と同様に扱われ、抑圧されずに出力されてしまう。

これらの問題を解決するため、提案法では、副ビームフォーマ出力から推定雑音を生成し、主ビームフォーマの信号から差し引く方法を用いている。

副ビームフォーマの出力は目的音声は除かれ、雑音信号のみで構成されている。したがって、全区間で雑音信号の推定

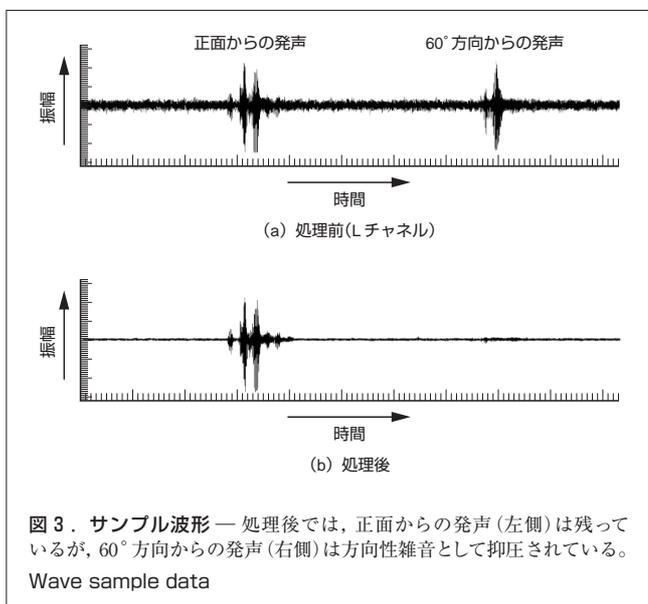
が可能であり、推定雑音に音声スペクトルが混入する問題を避けることができる。

また、非定常な信号が含まれていても、これが目的音ではないので非定常雑音として用いることができる。その結果、通常のSSでは抑圧が困難であった非定常雑音の抑圧を行うことも可能である。

厳密には、主ビームフォーマと副ビームフォーマは独立に適応処理を行っているため周波数特性が異なる。そのため、雑音推定には主ビームフォーマの信号も一部用いている。

5 実験

実験に用いた音声の波形例を図3に示す。図3(a)はステレオ収録した信号のLチャンネルの波形、図3(b)はこのステレオ信号をマイクロホンアレイで処理した波形である。収録は計算機室で行い空調機や計算機のノイズが含まれている。図3(a)の二つの発声のうち、前半は正面方向(0°)から、後半は右側約60°方向から発声した。追尾範囲は±20°とした。図3(b)から正面の発声(前半の発声)は残っているが、60°方向からの発声(後半の発声)は方向性雑音として抑圧されていることがわかる。また、2chSSの働きにより拡散性雑音も抑圧され、全体として正面方向の音声のみが強調されていることがわかる。このデータの場合、拡散性雑音の抑圧量は14.5 dBであった。



6 あとがき

二つのマイクロホンを用いた追尾型マイクロホンアレイについて提案した。提案法は二つのビームフォーマを備え、目的音方向に死角を形成する副ビームフォーマから目的音方向を検出し、主ビームフォーマの目的音方向を逐次更新することで、目的音方向の変動に対するロバスト化を実現している。また、二つのビームフォーマの出力を用いた2chSSを後処理として用いることで、突発性雑音も含めて拡散性の雑音を抑圧可能である。実験により、方向性雑音と拡散性雑音を効果的に抑圧できることを確認した。今後は反射や残響に対する検討を進めていく。

文献

- (1) S.F.Boll. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Trans. ASSP, 27, 2, 1979, p.113 - 120.
- (2) 金田 豊. 音響システムとデジタル処理. 東京, 電子情報通信学会, 1995, 265p.
- (3) L.J. Griffiths, et al. An Alternative Approach to Linearly Constrained Adaptive Beamforming. IEEE Trans. Antennas Propagation, 30, 1, 1982, p.27 - 34.
- (4) 宝珠山治, ほか. ブロッキング行列にリーク適応フィルタを用いたロバスト一般化サイドローブキャンセラ. 電子情報通信学会論文誌A, J79-A, 9, 1996, p.1516 - 1524.
- (5) J.E. Greenberg, et al. Evaluation of an Adaptive Beamforming Method for Hearing Aids. J.A.S.A., 91, 3, 1992, p.1662 - 1676.
- (6) 永田仁史, ほか. 話者追尾2チャンネルマイクロホンアレイに関する検討. 電子情報通信学会論文誌-A, J82-A, 6, 1999, p.860 - 866.
- (7) 永田仁史, ほか. 話者追従型2チャンネルマイクロホンアレイ. 東芝レビュー, 52, 10, 1997, p.47 - 50.
- (8) 井阪岳彦, ほか. マイクロホンアレイを用いた音声検出に関する検討. 日本音響学会講演論文集A, 1997-9, p.165 - 166.



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, 音声認識, 音声符号化分野の研究・開発
に従事。電子情報通信学会, 日本音響学会, IEEE 会員。
Multimedia Lab.



金澤 博史 KANAZAWA Hiroshi

研究開発センター マルチメディアラボラトリー 主任研究員。
音声認識, 音声対話, 音響信号処理分野の研究・開発に従事。
電子情報通信学会, 日本音響学会 会員。
Multimedia Lab.