

リアルタイム音源分離技術

稗方孝之・森田孝司

技術開発本部 生産システム研究所

当社は、複数の音声や各種環境音が混ざり合った環境から、目的とする音源信号のみをリアルタイムかつ高精度に抽出できる技術を開発し^{注)}、同時に、手のひらサイズの小型装置に実装することに成功した(図1)。工場の生産ラインにおける暗騒音下での異常音の聞き分け、様々な周囲騒音や他人の会話などのなかでの音声指示により動作するカーナビや人間型ロボットの音声認識システムへの音声入力、屋外の喧騒下で使用する携帯電話や工場内の騒音下で使用するインカムなどに対するクリアな音声入力など、目的とする音だけを確実に抽出することが必要とされる多様な用途への展開が期待されている。

複数の音源が混在する環境から目的とする音源を抽出する音源分離の手法としては、従来からアレイ状に配列されたマイクロフォンに対する入力音を利用する方式があるが、あらかじめ音源方向の特定が必要であるなど自由度が少なく、より柔軟に対応できる手法が望まれていた。当社で開発した音源分離技術は、異なる位置に配置された複数のマイクロフォンに対して、各マイクロフォンで観測された信号のみを利用して元の音源信号を推定する手法であり「ブラインド音源分離」と呼ばれている。この手法では種々の環境に対して柔軟に対応できる一方で、各マイクロフォンでの観測信号を解析し、複数の音源信号に分離するために膨大な演算量の信号処理が必要となり、十分な分離性能をリアルタイムで実現することは非常に難しかった。

そこで当社では、分離信号が互いに独立となるようにフィルタ行列を最適化する「独立成分分析法」と、2つの信号成分を周波数ごとにパワー比較する「バイナリマスク処理」という2つの手法を合わせた独自の2段階音源分離アルゴリズムを開発した(図2)。独立成分分析法は分離性能が高い反面その演算量が非常に大きくリアルタイム処理には不向きであるが、一方でバイナリマスク処理は比較的演算量が少なくリアルタイム処理に適しているが分離性能は十分ではない。これらのそれぞれの長所と短所を補完しつつ、リアルタイムかつ高性能な音源分離技術を実現した。

また同時に、今回開発した2段階音源分離アルゴリズムを手のひらサイズの小型装置に実装し、その有効性を確認した。装置には2つのマイクロフォンを搭載しており、2つの音源を分離

できる。上記アルゴリズムは浮動小数点 DSP (Digital Signal Processor) に搭載されており、リアルタイムに2つの音源信号に分離して出力する。分離性能として目的音源信号に対する非目的音源信号の残留量は、最大 28dB である(表1)。

当社では、社内外の生産設備における音響診断システムへの応用のほか、クリアな音を必要とする様々な製品やシステムに対する技術供与も同時に進めていく予定である。

応用例

- ・機械装置の音響診断に
工場内の暗騒音下における特定の設備・機器・配管などの動作音を対象とした異常音の抽出
- ・携帯電話のノイズ除去マイクとして
街なかの雑踏のなかでの通話や工場内の騒音環境下での音声通信
- ・自動車搭載の音声インターフェースとして
カーナビの音声による操作やハンズフリーでの音声通話に対するクリアな音声入力
- ・人間型ロボットの耳の役割として
まわりにいる人の声を認識して自然な音声対話を実現



図1 リアルタイム音源分離装置

表1 音源分離装置の仕様

| | |
|-------------|--|
| 使用プロセッサ | Texas Instruments 社製浮動小数点 DSP (Digital Signal Processor) TMS320VC6713 (動作周波数 200MHz) |
| 入出力インターフェース | 2ch マイク入力, 2ch スピーカ/ライン出力 (4ch マイク入力まで拡張可能) |
| サンプリング周波数 | 8kHz (音声帯域に対応) (16kHz/32kHz 対応可能) |
| 音源分離性能 | 最大 28dB (残信号の音圧は約 1/25) |
| 処理遅延時間 | 約 200msec |
| 電源 | 単 3 乾電池 2 本 |

注) 本技術は、奈良先端科学技術大学院大学 音情報処理学講座 猿渡洋助教授と共同で開発した。

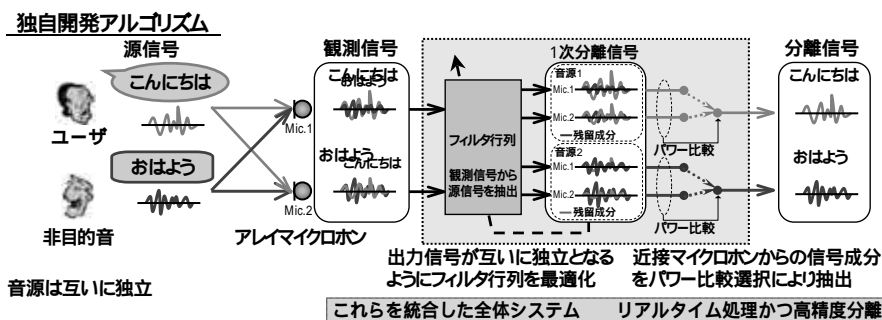


図2 2段階音源分離アルゴリズム