

混ざった音をリアルタイムに聞き分ける

小野 順貴 (NII)、春原 政浩、春田 智穂 (リオン)

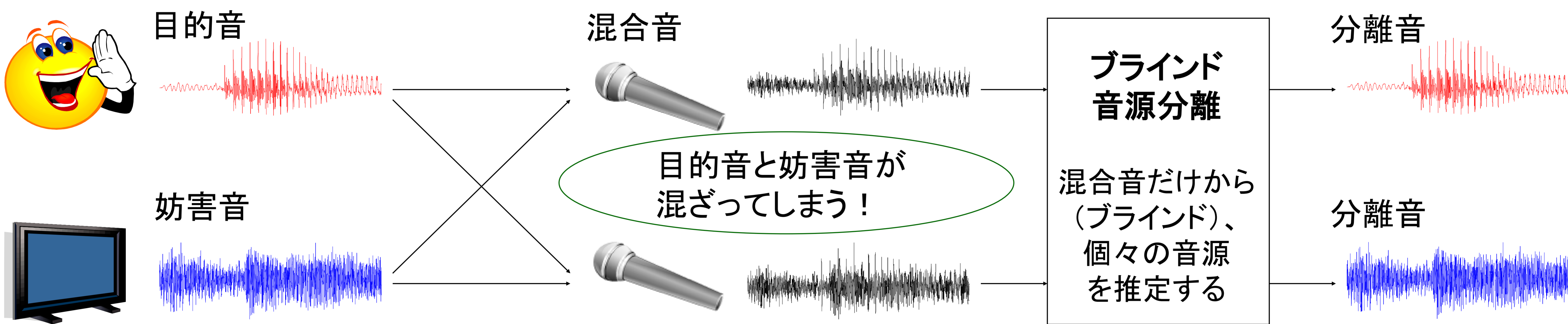
どんな研究？

実環境には様々な音が存在し、それらは混ざり合って聞こえてきます。混ざった音の中から、特定の音を認識したり聞き分けたりするために、音源分離という信号処理技術を研究しています。

何のために？

音源分離は音声認識をはじめ、様々な音響システムに役立ちます。本研究では補聴器への応用を目指し、リアルタイム性の高い音源分離システムを実現しました。

状況設定



研究内容

● どうして混ざった音源が分離できるの？

2つの音が混ざった場合を考えてみましょう。2つのマイクロフォンで音を録音すると、音が伝わってくる方向によって、録音した信号には強度差や時間差が生じます。2つの音が混ざっていても、それらの音が異なる方向から伝わってくる場合には、強度差や時間差が異なります。(図ではわかりやすさのため、混ぜずにかいています。)(A)

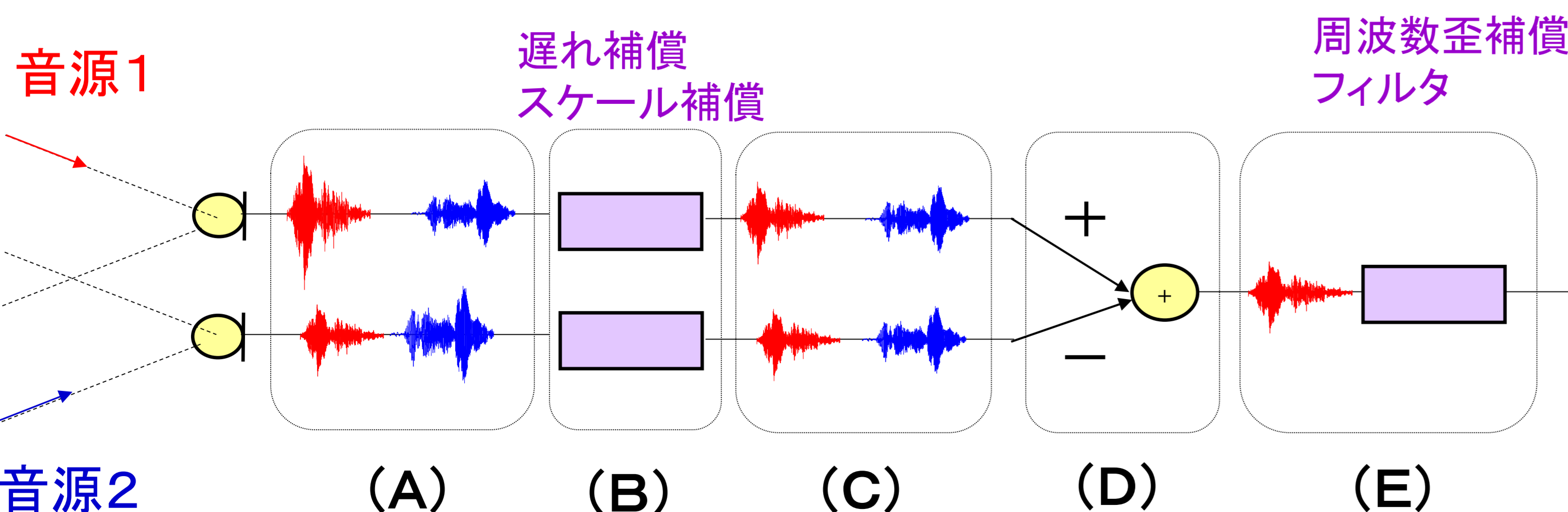
よって、例えば音源2に対して、2つの録音の間で、強度差や時間差がどのように違ってくるかがわかっていたとすると、録音2の大きさを変え、時間をずらすことによって(B)、録音1と録音2の間で、音源2の強度と時間が同じになるようにあわせることができます。(C)

音は波なので、混ざっていても、波形としては実は足し算されているだけです。そのため、(C)のように強度と時間をあわせてから引き算すれば、音源2だけ消すことができます。(D)

ただし、この引き算のせいで、音源1の音が歪んでしまいますが、これは簡単なフィルタ操作で元に戻すことができます。(E)

実環境では、残響などの影響があるので、2つの録音の間で生じる強度差や時間差は異なるため、一度信号をたくさんの周波数成分に分解し、それぞれの周波数帯域ごとにこの処理を行います。

また、2つの録音の間での強度差や時間差は本当はわかりませんが、分離した2つの信号ができるだけ独立になるように、強度差や時間差を推定しながら音を分離するのが、独立成分/ベクトル分析という技術です。

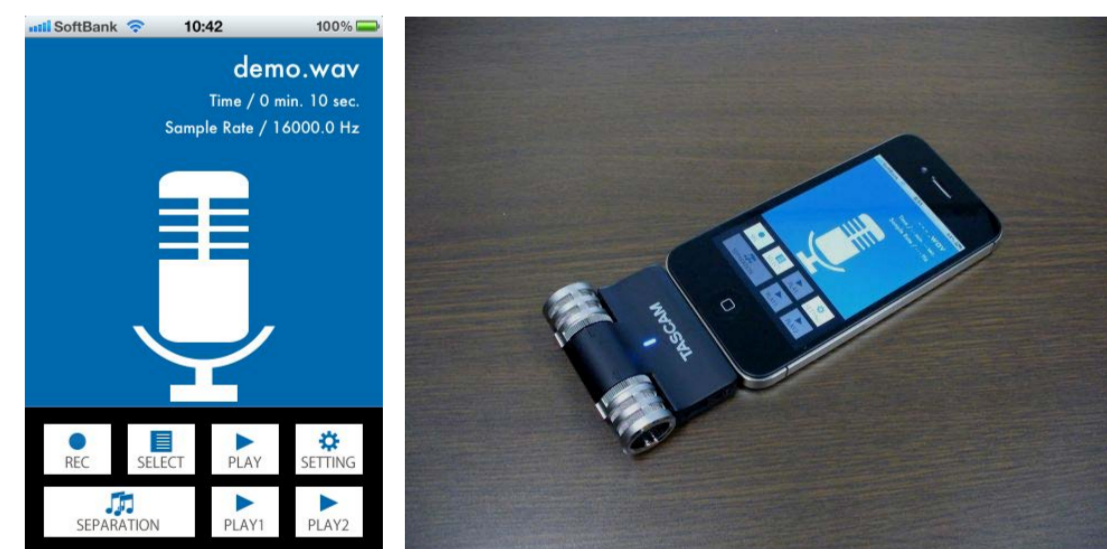


● どうやってリアルタイムに計算しているの？

リアルタイムに音源分離を行うためには、
1) 音信号に対する計算をその音より短い時間で行う (高速性)
2) 過去の情報だけを使って計算する (因果性)
の2つが必要となります。

高速化：補助関数法に基づく高速独立ベクトル分析 (2011年～)

NII小野研究室ではまず、補助関数法という最適化の枠組みを用いて、少ない計算量で音源分離フィルタを推定できる高速アルゴリズムを導出しました。またこれを用いて、一度混合音を録音し、録音全体を使って音を分離する、オフラインの音源分離システムを開発しました。上図はiPhone上で動作する音源分離システムです。



オンライン化

次に、録音全体を使って音の分離フィルタを推定する代わりに、次々と入力されてくる混合音を使って音の分離フィルタを逐次的に更新していく、オンライン型のアルゴリズムを開発しました。これは、株式会社東芝との共同研究により行われました。

低遅延化

これまでの音源分離アルゴリズムは、音信号を、フレームと呼ばれる適当な長さ (例えば、256ms) ごとに分割して処理する必要がありました。この場合、たとえ計算機が無限に速くても、入力 (録音) から出力 (再生) までの間に1フレーム分の時間遅れが生じてしまいます。これをアルゴリズム遅延と呼びます。NII小野研究室とリオン株式会社は共同研究により、音の分離をほぼ因果的なFIRフィルタにより行い、この遅延を10ms以下に抑える新しいアルゴリズムを開発しました。

