複数自己相関関数の多数決に基づいた遅延時間推定による反射音除去法

G-010

Removing Reflected Waves Using Delay Time Detected by Majority Decision on Auto-correlation Functions

大田健紘† 柳田益造†
Kenko Ohta Masuzo Yanagida

1．はじめに

近年の音声認識技術の進歩に伴い，音声認識率は向上し，接話マイクを用いるならカーナビゲーションシステムなどにまで使えるようになってきている．しかし，実環境下における音声認識率はまだまだ実用レベルに達していない．実環境下における音声認識では，周囲の壁からの反射音や，その他の雑音の影響により音声認識率が極端に低下する．実環境下での音声認識率を向上させる手法として，雑音や反射音を除去する手法や反射音や雑音に適応させる手法がある．前者としては，雑音を除去するスペクトラムサブトラクション法（SS 法）[1]や，ケプストラム平均除去法（CMS 法）[2]が代表的である．後者としては，クリーン音声から作成された HMM を雑音や反射音がある環境に適応させる HMM 合成分解法[3]が挙げられる．

反射音抑圧処理としては一般的に室内伝達関数の逆フィルタが用いられるが[4]，伝達関数が既知でかつ不変である場合でしか用いることができない．

一方で伝達関数の測定を必要としない回復方法として，変調伝達関数(MTF：Modulation Transfer Function)に基づいた方法[5]を古川らは提案している．これは，MTF に基づいて音源信号と伝達特性をモデル化しており，パワーエンベロープの回復を目的としている．この方法では伝達特性のパラメータである残響時間と振幅項の決定方法を与えているが，まだ適切なものが提案されていない．

本稿では，自己相関関数を用いた，伝達関数の測定を必要としない反射音除去法を提案し，実環境下における評価実験により，その有効性を検討している．

2．提案手法の概要

2．1　反射音除去の原理

実環境における音声は，周囲の雑音，空間伝達中に受ける歪みや壁などからの反射の影響を受けてマイクに受音される．このときマイクで受音した音声は式（1）で表せる．

$$r(t) = s(t) \otimes h(t) + \sum_{u}^{N} n_u(t) \otimes h_u(t) \qquad (1)$$

ここで，$s(t)$は元の信号，uはノイズ源の番号，$n_u(t)$はノイズ，$h_u(t)$はノイズ源からマイクまでのインパルス応答，⊗は畳み込み演算である．本稿では反射音のみを扱うので，式（1）は式（2）のように簡略化することができる．

$$r(t) = s(t) \otimes h(t) \qquad (2)$$

反射音は直接音の定数倍されたものがある時間遅れをもって直接音に加算されたものとして仮定することができる．以上の仮定から，クリーン信号を推定するための式を式（3）のように書くことができる．

$$\hat{s}_k \cong r_k - \sum_{j=1}^{P} \alpha_j s_{k-l_j} \qquad (3)$$

ここで，α_jは第j経路の減衰率，l_jは第j経路の時間遅れ，Pは減算する反射の数である．但し，反射面では全周波数を均等に反射するものと仮定する．そして式（3）を逐次計算に書き直し，右辺第２項にある元の信号は未知なので，初期条件としては観測信号を用いて近似する．

$$\hat{s}_k^{(j)} \cong \hat{s}_k^{(j-1)} - \alpha_j \hat{s}_{k-l_j}^{(j-1)} \quad j = 1, \cdots, P$$
$$j = 1 \quad \hat{s}_k^{(0)} = r_k \qquad (4)$$

2．2　自己相関関数を用いた時間遅れの推定法

提案手法では，複数のマイクロフォンを用いることにより，時間遅れの推定精度を向上させ，式（4）にしたがってクリーンな音声を得て，認識率を向上させることが目的である．

壁からの反射があると定数（１より小さい）倍された時間遅れ信号が元の信号に加算されるので，反射の影響を受けている時間遅れで自己相関関数が大きくなるはずである．しかし，元の信号自身の自己相関関数に凹凸があるため，ある時間遅れで自己相関関数が大きくても，反射の影響なのか信号自身のものなのかの判断が困難になる．

提案手法では３本以上のマイクを用い，ある１本のマイクの自己相関関数とそれ以外のマイクの自己相関関数の平均との差の極大値を用いることで，反射の影響を受けている時間遅れの推定精度を向上させる．マイクと壁の相対位置によって反射の影響を受ける遅れ時間は異なっているはずなので，マイク間で自己相関関数の差が大きくなっている時間遅れは，反射による時間遅れを表すと解釈できる．

2．3　処理の流れ

まず，複数のマイクロフォンで受音した音声それぞれについて音声区間の開始点を検出する．検出した開始点を始点として各マイクの自己相関関数を計算する．次に，各マイクの自己相関関数の極大点を与える時間遅れの差を求める．その差を，第j番目の経路の時間遅れ（l_j）とみなす．一方，当該マイクロフォン以外の自己相関関数の平均を平均自己相関関数と呼ぶことにし，元の信号の自己相関関数と仮定する．これを最適化の基準として用い，第j番目の経路の減衰率（α_j）を最適推定する．これらを，取り除きたい反射の数だけ推定し，式（4）に従って減算を行っていく．図１に処理過程を示す．各処理の具体的な内容は次節で説明する．

3．各処理の説明

3．1　音声区間の検出

２重閾値法を順方向に用いることにより音声区間の検出を行っている．

3．2　α_jの最適推定

2.2 の方法で推定した時間遅れと，各マイクの自己相関関数を用いて，以下の方法により最適な減衰率（α_j）を推定する．

空間的に異なる位置に配置されたマイクによる受音波形の自己相関関数の平均を用いることにより，伝達特性が平均化され，元の信号の自己相関関数に近づくため，それ

†同志社大学工学部

を元の信号の自己相関関数と仮定し，最適化の基準として用いる．減衰率（α_j）は適当な初期値から出発し，最急降下法により求める．収束条件は，推定した時間遅れでの自己相関関数と，同一の時間遅れでの平均自己相関関数との差が 10^{-5}以下になったときとしている．

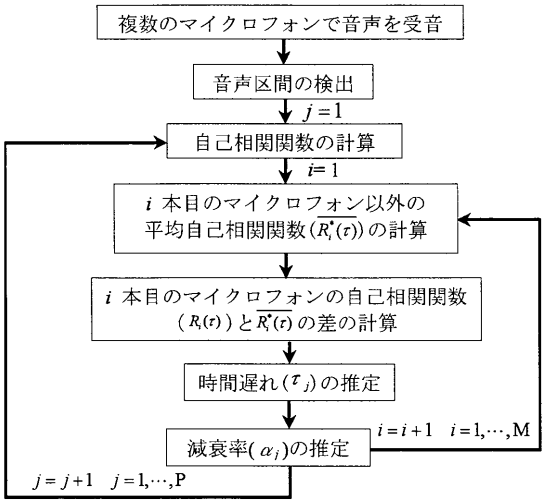

図１　反射音除去の処理過程

４．評価実験

４．１　実験環境

提案手法の有効性を確認するために，壁からの反射のあるリビングルームシミュレータ（白色ノイズによる残響時間＝440ms）で実験を行った．実験にはクリーン音声を流すためのスピーカ（高さ 1.25m）１個と，それを受けるマイク３本（高さ 1.25m）を用いた．図２にマイクとスピーカの配置図を示す．

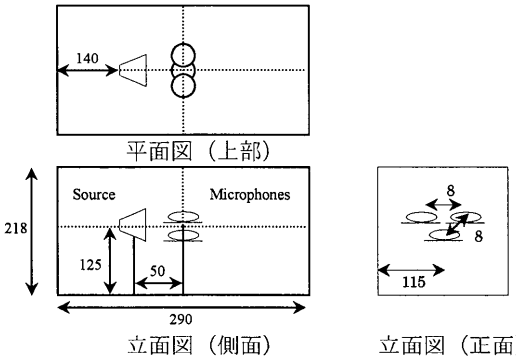

図２　評価における実験スピーカとマイクの配置

４．２　評価実験の条件

実験に用いた音声データは，接話マイクを用いて録音した男性２名，女性１名による合計１５３発話である．音声データの収録条件を表１に示す．

表１　音声データの収録条件

標本化速度	16ksamples/sec
量子化精度	16bits

認識で用いた単語数などの条件を表２に示す．

表２　認識辞書の語彙数および文法数

語彙数	99
文法数	13

４．３　実験結果

提案手法で処理した音声データを音声認識システム JULIAN を用いて処理前後の認識率を比較した．

処理前と処理後の各マイク認識率を表３に示す．

表３　各マイクによる音声認識率（%）

	マイク１		マイク２		マイク３	
	前	後	前	後	前	後
M1	88	88	88	84	82	88
M2	61	65	67	69	61	61
F1	82	78	82	82	63	75

実験ではマイクを３本用いているため，多数決を出力とした場合，即ち２本以上のマイクで正しく認識できたものを成功とみなし，認識率を再計算した結果を表４に示す．

表４　音声認識率の再計算（%）(**:有意水準１%)

	処理前	処理後
M1	84	88
M2	67	67
F1	76	84**

５．検討

反射音除去処理前の認識率と処理後の認識率を比較したところ，若干しか向上していないことがわかる．そこで，認識に失敗しているデータについて調査を行ったところ，以下のことがわかった．

- 音声区間の検出に失敗している
- 反射音除去処理後に摩擦音や撥音が消えている
- α の推定処理で適切に収束していない

これらの問題点を解決することで，提案法はまだまだ改善できるということが示された．

６．今後の課題

前節で示した問題点を解決するために

- 音声区間検出の高精度化
- 摩擦音とその他の音の区別

を行う．そして，提案法を用いることによって，スペクトルのレベルでクリーン音声に近づいていることを確認するために

- スペクトル距離による評価

を行う．さらには，反射音の影響により分離性能が落ちるとされている ICA の性能を向上させるために

- ICA の前処理としての検討

を行う．

参考文献

[1]北岡教英，赤堀一郎，中川聖一，“スペクトラムサブトラクションと時間方向スムージングを用いた雑音環境下音声認識，”電子情報通信学会論文誌，vol.J83-D-II，No.2，pp.500-508，2000.

[2]鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄，“音声認識システム，”オーム社，第１章，pp.14-15，2001.

[3]三木一浩，西浦敬信，中村　哲，鹿野清宏，“マイクロフォンアレーと HMM 分解・合成による雑音・残響下音声認識，”電子情報通信学会論文誌，vol.J83-D-II，No.11，pp.2206-2214，2000.

[4] Miyoshi, M. and Kaneda, Y., "Inverse filtering of room acoustics, "IEEE Trans. ASSP, Vol. 36, No. 2, pp. 145-152, Feb. 1988.

[5]古川正和，鵜木祐史，赤木正人，“MTF に基づいた残響音声パワーエンベロープの回復方法，”信学技報，SP2002-15，pp. 49-54，2002.