実環境下における音声認識率向上のための残響除去技術の検討

大田健紘　柳田益造
同志社大学工学部　〒610-0394 京都府京田辺市多々羅都谷 1-3
E-mail: dtd0736@mail4.doshisha.ac.jp myanagid@mail.doshisha.ac.jp

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や残響，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，雑音や残響を除去する手法や残響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，残響がある環境に適応させる HMM 合成分解法[3]が挙げられる．

本稿では，残響除去法の一手法を開発したので，実環境下における評価実験により，その有効性を検討する．

２．提案手法の概要

２．１　残響除去の原理

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．

$$r(t) = s(t) \otimes h(t) + \sum_{i}^{N} n_i(t) \otimes h_i(t) \qquad (1)$$

ここで，$s(t)$ は元の信号，i はノイズ源の番号，$n_i(t)$ はノイズ，$h_i(t)$ 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．本稿では残響のみを扱うので，式（１）は式（２）のように簡略化することができる．

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

残響は元の信号の定数倍されたものがある時間遅れをもって加算されたものとして定義することができる．このことから，残響を除去するために式（３）を仮定する．

$$\hat{s}_k \cong r_k - \sum_{i=1}^{P} \alpha_i s_{k-l_i} \qquad (3)$$

ここで，α_i は第 i 経路の減算率，l_i は第 i 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．右辺第２項にある元の信号は未知なので，実際は観測信号を用いて近似する．

$$\hat{s}_k \cong r_k - \sum_{i=1}^{P} \alpha_i r_{k-l_i} \qquad (4)$$

２．２　処理の流れ

提案手法では，２本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，認識率を向上させることが目的である．

まず，２本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，２本のマイクのそれぞれ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．自己その差を，第 i 番目の経路の時間遅れ（l_i）とみなす．一方，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 i 番目の経路の減算率（α_i）を最適推定する．これらを，取り扱い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．

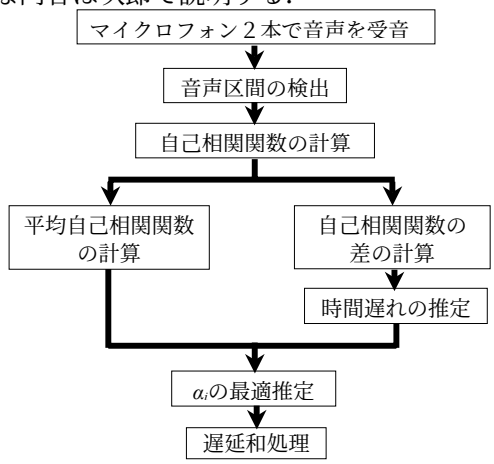


図１　残響除去の処理過程

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　時間遅れの推定

壁からの反射があると定数倍（１より小さい）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．

そこで，提案手法では 2 本マイクを用いて，各マイクの自己相関関数の最大値ではなくマイク間の自己相関関数の差の最大値を用い，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．各マイクで元の信号の特性によって自己相関関数が大きくな

Reverberation Removal for Improving Speech Recognition Rate in a Real Environment
Kenko Ota,Masuzo Yanagida
Faculty of Engineering,Doshisha University

る点は同じであるが，マイクの配置によって反射の影響を受ける点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からである．

まず各マイクについて検出した音声区間開始点から 512 点を用いて自己相関関数を計算する．このとき見かけ上の周波数分解能を上げるために 4096 点の FFT を用い０付加処理を施す．

そして，各マイクの受音波形の自己相関関数の差を各周波数について求め，それが最大になっている時間遅れを，反射の時間遅れとみなす．

３．３　α_i の最適推定

推定した時間遅れと，各マイクの平均自己相関関数を用いて，以下の方法により最適な減算率（α_i）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平均自己相関関数を使うことにより，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．減算率（α_i）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 0.001 以下になったときとする．

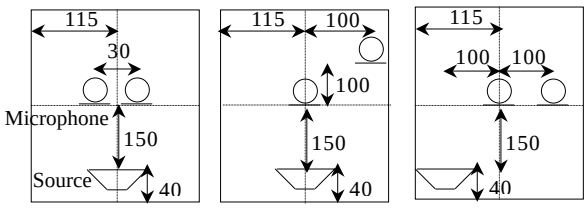
３．４　遅延和処理

推定した時間遅れに基づいて，２本のマイクの信号を遅延加算する．

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２本（高さ 1.25m）を用いた．それらの配置の仕方により３種類の状況を設定し，実験を行った．図２に３種類の状況を示す．



状況１　　状況２　　状況３

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による２５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	148
文法数	41

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．認識の際のパラメータは，分析フレームのサンプル点数 510 点，フレームシフト幅 255 点，10 個の文仮説が得られるまで探索を続けてそのうち１つを出力し，音響尤度計算時に行うガウシアン・プルーニングは精度が最も高いものを用い，7.2kHz 以上の周波数は遮断し，ビーム幅は 1000 とした．

クリーン音声の認識率，状況１，２，３それぞれについて処理前と処理後の認識率を表３に示す．

表３　音声認識率（%）() は標準偏差

	処理前	処理後
クリーン音声	92	
状況１	84	92
状況２	64	68
状況３	64	80
平均	70.7(11.5)	80(12)

５．検討

表３から，処理前と処理後を比較すると音声認識率は向上していることが確認できる．状況１～３について，正しく認識できるようになったのは８データ，誤って認識されるようになったのは１データである．処理後の音声認識率と処理前の音声認識率に有意差があることを検定するために符号検定を行ったところ，有意水準５%で有意差が認められた．このことから，提案手法を音声認識に用いることは有効であるといえる．

６．今後の課題

・実環境に近づけるために，雑音（テレビの音声や会話音声など）を含む音声データを用いる
・　ICA の前処理としての検討
・　時間遅れの推定精度の向上

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.