ISS-P-122

逆フィルタ処理による残響除去のための深層学習を用いた残響除去

桃瀬 裕基† 堀内 靖雄† 黒岩 眞吾†
† 千葉大学大学院融合科学研究科

1. はじめに

マイクロホンで録音した音声には，壁などで反射し，様々な時間遅れでマイクに到達する乗算性雑音(残響)が含まれる．これは音声品質を下げる要因の一つである．近年，音声認識の分野で深層学習を用いた残響に頑健な手法が多く提案され，成果が出されている． そこで本研究では深層学習を用い，時間領域における残響除去フィルタの推定を行う手法を提案する．

2. 提案手法 1:周波数領域での残響除去

残響音声の対数パワースペクトルから残響のない対数パワースペクトルを推定する Denoising Autoencoder(DAE)を構成する[1]．このときスペクトルの微細構造が失われてしまうため，残響音声の微細構造を加えた上で，時間領域に戻し再合成する手法を試みる．

3. 提案手法 2:時間領域での残響除去

Deep Neural Network(DNN)を用いて残響音声の特徴量から残響時間のみを推定し，推定された残響時間により文献[2]の手法を用いて時間領域における残響除去フィルタを作成する．この時間領域残響除去フィルタを残響音声に適用することで残響除去音声を作成する．

4. 実験条件

本研究では日本語大語彙連続音声認識研究を目的とした新聞読み上げコーパス(JNAS)を用いた．DNN，DAE の学習には，120 名の各話者 10 発話，合計 1200 発話を学習セットとして用いた．また評価セットとして，学習に用いていない 8 名の各話者 10 発話，合計 80 発話を用いた．学習セットには，残響下連続数字音声認識コーパス(CENSREC-4)[3]に付属している全 8 種類の残響インパルス応答のうち 4 種類のいずれかをランダムに畳み込んだものを用いる．評価セットには，学習に用いた 4 種類のインパルス応答，また学習には用いていない 4 種類のインパルス応答を用いて，残響環境におけるクローズ，オープンの評価を行う． 評価は 5 段階の音声品質評価指標である PESQ[4]を用いる．PESQ がより高い程音声品質が良いと考えられる．

5. 実験結果・考察

表 1 に評価セット全体における PESQ の平均の値を示す．また，図 2 に各手法の残響除去音声のスペクトログラムを示す．残響音声と比較し，提案手法の双方で PESQ が向上した事が表 1 により確認できる．

表 1 PESQ の結果

残響環境	残響音声	提案手法 1	提案手法 2
クローズ	2.61	2.80	2.64
オープン	2.53	2.67	2.54

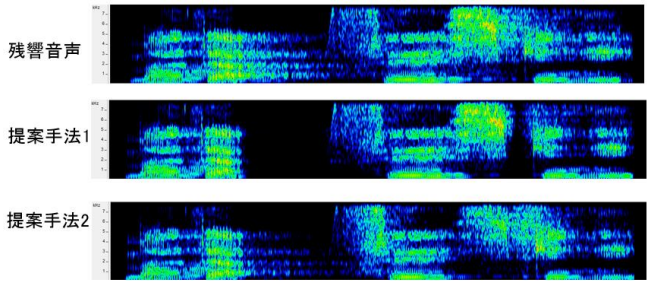

図 2 スペクトログラム

提案手法の比較では，提案手法 1 の方が性能は高い．また，スペクトログラムにおいても提案手法 2 より提案手法 1 の方が残響を抑制できていることが確認できる．しかし実際に音声を聞くと，提案手法 1 では残響は非常に抑制されているが，音声の自然性が低下していた．この要因として，周波数領域での処理のためのフレーム間不連続が生じてしまった事が考えられる．一方提案手法 2 では，残響は提案手法 1 ほどではないがある程度抑制され，発話音声の自然性の低下は感じられなかった．

5. まとめ

深層学習を用いた残響除去手法を提案し，従来手法と比較を行い，利点と欠点を明らかにした．

参考文献

[1]Takaaki Ishii et al, “ Reverberant speech recognition based on Denoising Autoencoder ”, Proc. Interspeech 2013, 2013/8.

[2]古川 正和ら, “MTF に基づいた残響音声パワーエンベロープの回復方法”, 電子情報通信学会技術研究報告. SP, 音声 102(35), 49-54, 2002/4/19

[3] T. Nishiura et al.,“ Evaluation framework for distant-talking speech recognition under reverberant environments —newest part of the censrec series—, “ Proc. LREC’08, 2008.

[4]ITU-T 勧告 P.862: Perceptual evaluation of speech quality (PESQ)