

Title	残響音声からの音声特徴量抽出法と 再合成に関する研究	音源波形
Author(s)	酒田, 恵吾	
Citation		
Issue Date	2003-03	
Type	Thesis or Dissertation	
Text version	author	
URL	http://hdl.handle.net/10119/1672	
Rights		
Description	Supervisor:赤木 正人, 情報科学研究科, 修士	

残響音声からの音声特徴量抽出法と 音源波形再合成に関する研究

酒田 恵吾 (110049)

北陸先端科学技術大学院大学 情報科学研究科

2002年2月14日

キーワード: 変調伝達関数 (MTF)、パワーエンベロープ逆フィルタ処理、残響時間、フィルタバンク.

1 はじめに

残響音声から元の音源信号を回復することは、遠隔会議システムや音声認識において大きな課題である。これまでの残響抑圧法では、室内伝達特性の逆フィルタを用いた手法が多く提案されている。これらの手法は、時間変動する室内伝達特性をその都度正確に測定する必要があり、実用化は困難である。

古川らは MTF の理論に基づき [1]、室内伝達特性を測定せずに、観測信号の情報のみから残響音声のパワーエンベロープを回復する手法を提案している [3]。しかしその手法には解決すべき問題が残されている。また時間波形としての回復を考えると、キャリアに関する処理も考える必要がある。

本研究では室内伝達特性を測定せずに、残響音声の回復処理を行う手法を提案する。音声信号をエンベロープとキャリアでモデル化し、エンベロープ、キャリアそれぞれで回復処理を行う。エンベロープの処理に関しては、古川らのパワーエンベロープ回復法の問題点を解決し、改善した手法を提案する。キャリアの処理に関しては、残響音声中から F_0 が推定されたと仮定して、その F_0 情報を基にキャリアを再合成する処理を提案する。そしてそれぞれの処理から出力されたエンベロープとキャリアから音声を合成することで、時間波形としての残響音声の回復を行う。

2 パワーエンベロープ回復法の問題点

古川らが提案しているパワーエンベロープ回復法には以下の問題点が挙げられる。

1. 帯域分割処理において適切な帯域分割幅の決定をする際に、MTF 成立/不成立の検討がされていない。

2. 音声の重要な特徴を有する低帯域での回復効果が小さい。

この二つの問題点に対して検討を行った。

適切な帯域分割幅を設定するとき、パワーエンベロープの共変調、MTF 理論成立/不成立の二点を考慮する必要がある。そこで、帯域分割処理における適切な帯域分割幅を検討する際に、この二点に着目してそれぞれの調査を行った。最初にパワーエンベロープの共変調に関する調査を行った。そして帯域分割幅を狭くするにつれパワーエンベロープが共変調とみなすことができる結果が得られた。次に MTF 理論成立/不成立に関する調査では、帯域分割幅を狭くするにつれ MTF 理論が適用できない結果が得られた。以上二つの結果から、この二点はトレードオフの関係にあり、また、適切な帯域分割幅は 300 から 400 Hz の範囲であることがわかった。

2.1 低帯域の回復効果についての検討

2.1.1 低帯域の回復効果が小さい原因

低帯域で回復効果が小さい原因について説明する。音声の低帯域では、音声間の無音区間が長い状況が多く見られた。古川が提案した従来の残響時間推定法は、音源信号のパワーエンベロープは変調度 1 と仮定しているため、残響音声パワーエンベロープの変調度 1 となる状況下では適用できてないことがわかった。故に長い無音区間が存在する場合にも適用できる残響時間推定法を提案した。

2.1.2 長い無音区間に適用できる残響時間推定法の提案

パワーエンベロープ逆フィルタ処理 [2] によるパワーエンベロープの移動変化量から残響時間 T_R を推定する手法を提案した。パワーエンベロープ逆フィルタ処理は、回復処理後のパワーエンベロープを処理前よりも時間方向とは逆方向へ移動させる働きがある。逆フィルタ処理のパラメータ T_R の変化に対し、逆フィルタ処理後のパワーエンベロープ $\hat{e}_x(t)^2$ の移動変化量が急激に減少する点を適切な回復処理を行う境界条件と推測して、その地点の残響時間の値 \hat{T}_R を回復に最も適切なパラメータ値として決定する。この方法は音声の無音区間が長い場合でも適用できる。

従来法と提案法の比較のため同条件で残響音声を対象にパワーエンベロープ回復シミュレーションを行ったところ、低帯域で、従来法では音声の無音区間が長い低帯域で回復効果が得られなかったのに対し、提案法では回復効果が得られた。この結果から提案した推定法の有効性が示せ、低域での回復効果を上げることができた。

3 キャリア再合成法

キャリア再合成法は、F0が存在する有声音の区間と、F0が存在しない無声音区間に分けてそれぞれ処理を行う。有声音区間では、PIFM音源モデルを参考に、残響中から推定したF0から調波複合音のキャリアを作成する。無声音区間では、白色ガウス過程から生じた雑音による、不規則な波形構造のキャリアを作成する。

再合成されたキャリアのスペクトログラムを評価したところ、有声音区間ではF0に対応した縞模様、無声音区間では不規則な模様と、実音声の場合と同様の模様が見れた。音声としての特徴を持ったキャリアを作成することができた。以上から提案した手法の有効性が示せた。

4 提案法の評価のためのシミュレーション

提案したモデルの評価のためシミュレーションを行う。オリジナル信号音声ATR音声データベース (mau /sinbun/) に、 $T_R=0.5$ 秒の室内インパルス応答を畳み込んで得られた残響音声から、提案法による回復処理を行う。フィルタバンクの帯域分割幅は400 Hzとした。評価尺度としては、音声としての回復の評価をするため対数スペクトル距離 (LSD) を用いる。オリジナル、残響音声、回復音声の3つの音声に対して平等に測定するため、各エンベロープを提案したキャリア再合成したキャリアにそれぞれ合成する。その3つの音声に対して評価を行う。(Original, Reverberant, Dereverberant) OriginalとReverberantのLSD、OriginalとDereverberantのLSDをそれぞれ測定し、LSDの改善度で評価を行う。その結果、音声が存在する区間で大体1 dBの改善度が得られた。

オリジナル、残響音声、パワーエンベロープ回復処理後のそれぞれのパワーエンベロープと、キャリア再合成法で作成したキャリアで3つの音声を合成する。(Original, Reverberant, Dereverberant)。OriginalとReverberant、OriginalとDereverberantの対数スペクトル距離 (LSD) をそれぞれ計算し、評価を行った。その結果、音声の存在する区間で1 dB以上の回復効果が得られた。

5 まとめ

本研究では、観測した残響音声の情報のみから残響音声の回復処理を行う残響回復処理モデルを提案した。エンベロープ回復部では、適切な帯域分割幅の検討と低帯域での回復効果の二つの問題点に取り組んだ。適切な帯域分割幅の検討では、パワーエンベロープの共変調、MTF理論成立/不成立の二点に着目し、適切な帯域分割幅は300から400 Hzの範囲と決定した。低帯域での回復効果については、低域では音声間の無音区間が長い多く存在し、その場合では従来の回復法は適用できないことがわかった。そして無音区間が長い場合にも適用した回復処理法を提案し、低域の回復精度を上げることができた。

キャリア再合成部では、F0からキャリアを作成する手法を提案した。そして音声の特

徴をもつキャリアを再合成することができた。

提案モデルの評価のためシミュレーションを行った。その回復効果がみられ、提案モデルの有効性が示された。

参考文献

- [1] Schroeder, M.R., “Modulation Transfer Functions:Definition and Measurement”, Acoustics, Vol. 49, pp.179-182, 1981.
- [2] 広林, 山淵, “帯域分割を用いたパワーエンベロープ逆フィルタ処理の残響抑圧効果,” 信学論 A, Vol. J83-A, No. 8, pp. 1029-1033, 2000.
- [3] 古川, 鷓木, 赤木 “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.