

Title	音声明瞭度向上のための残響音声回復法に関する研究
Author(s)	戸井, 真智
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1918
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

音声明瞭度向上のための残響音声回復法に関する研究

戸井 真智 (310072)

北陸先端科学技術大学院大学 情報科学研究科

2005年2月14日

キーワード: 残響回復, 変調伝達関数 (MTF), パワーエンベロープ逆フィルタ処理, キャリア再生, 音声明瞭度.

残響環境下では, 壁や天井などによる反射音が直接音に歪みを与えるため, 音声の明瞭度を著しく低下するだけでなく, 音声認識における認識精度低下の原因にもなっている. そのため, これらの問題点を解決するためには, 残響の影響を抑圧しなければならない. また, 残響は, 環境の様々な変化と共に時々刻々と変動するため, この性質を考慮して残響回復を行う必要がある.

これまでの残響音声の回復法では, 室内伝達特性の逆フィルタ処理を用いた手法が数多く提案されている. しかし, これらの手法では, 時変的な室内伝達特性をその都度測定する必要があるため, 様々な音声アプリケーションに用いることは困難である. これに対して, 酒田らは, MTF 理論に基づいて, 室内インパルス応答を測定することなく, 観測した音声情報から残響音声を回復する手法を提案した. この手法は, 音声をパワーエンベロープとキャリアに分け, それぞれに対して回復処理を行うことにより, 回復した音声信号を波形で出力することができる. しかし, この手法には解決すべき点がいくつか残されている. まず, パワーエンベロープ回復処理部で, 改良の余地が考えられるということ. 次に, キャリア再生部において, 位相の制御を行っていないため, 回復音声の品質が良くないということ. また, 回復音声の明瞭度について何も考察されていないという検討事項もある.

本研究の目的は, 残響の影響を受けた音声から, 室内伝達特性の変動に依存せず, ブラインド的に残響音声を回復する手法を構築し, 残響によって低下した音声明瞭度を向上させることである. 本研究では, 室内インパルス応答を測定することなく回復処理を行うことができる酒田らの手法をベースに, 残響回復モデルの構築を行う. さらにこの手法は, パワーエンベロープの回復だけでなく, 残響の影響を受けたキャリアに対しても回復処理を行うので, 音声明瞭度を向上できる可能性があると考えられる.

本研究のモデルは, 主にパワーエンベロープの回復処理部とキャリア再生部で構成される. まず, パワーエンベロープの回復処理部では, MTF 理論に基づいたパワーエンベロープ回復法を利用する. この手法は, フィルタバンクによって帯域分割された音声

に対して、パワーエンベロープを抽出し、逆フィルタ処理を行うものである。従来法では、パワーエンベロープの共変調の性質と MTF の成立/不成立を考慮して、100 Hz 一定の帯域幅で帯域分割を行っていた。しかし、パワーエンベロープの共変調の性質は、音声帯域ごとに異なる。そのため、対象となる音声帯域ごとに適応的に帯域幅を定めてやることによって、さらに回復精度が向上できると考えられる。また、時間領域においても、従来法では音声信号の全区間一括で残響時間パラメータを決定していたが、適切な時間区間を定めて、それぞれの区間ごとにパラメータを決定することにより、さらに適切な値が得られると考えられる。本研究では、この考えに基づき、対象となる音声ごとに適応的に時間-周波数分割処理を行うパワーエンベロープ回復法の検討を行った。この提案法を相関と SNR、そして対数スペクトル歪み (LSD) を尺度として評価シミュレーションを行った。その結果、この時間-周波数分割処理によって、エンベロープの回復精度をさらに向上できることがわかった。

また、エンベロープを抽出する際に利用していたカットオフ周波数についても検討を行った。従来法ではカットオフ周波数を $f_c = 20$ Hz としてエンベロープ抽出を行っていた。しかし、この値はアドホックに決定した値であり、最適なエンベロープ抽出を行うカットオフ周波数であるという考察は無かった。そこで本研究では、MTF 理論が最も成立しやすいようにカットオフ周波数を定めて、その値でエンベロープ抽出を行うようにした。その結果、従来の一定のカットオフ周波数 $f_c = 20$ Hz を用いる場合に比べ、回復精度が向上した。特に、残響時間の長い音声に対しても、エンベロープの回復精度を維持できることがわかった。

次に、キャリア再生処理部では、推定された基本周波数 (F_0) を基に、有声/無声音ごとに音源信号を再生成し、フィルタバンクを通過させ、各チャンネルのキャリアを生成する。しかし、従来法では、再生成された音源信号の位相が制御されていなかったため、回復音声の音質が悪いという問題点があった。そこで本研究では、群遅延を操作することによって音源信号の位相を制御する手法を検討した。その結果、回復音声の音質を大きく改善することができた。

提案モデルがどれだけ残響回復できるかを総合的に評価するため、主観評価実験を行った。主観的評価では、音声品質、残響感、そして単語理解度について評価を行った。その結果、残響音声に比べて大きく残響感が解消され、また従来法よりも音声品質が向上する結果が得られた。それに伴い、提案モデルが、残響によって低下した音声明瞭度を回復できることも理解度試験からわかった。これより、主観的評価においても客観的評価においても、本研究の提案モデルの高い有効性が示された。