

Title	騒音環境下における車室内発話音声の分析とその合成に関する研究
Author(s)	竹山, 佳成
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1994
Rights	
Description	Supervisor: 赤木 正人【鷗木祐史】, 情報科学研究科, 修士

騒音環境下における 車室内発話音声の分析とその合成に関する研究

竹山 佳成 (410075)

北陸先端科学技術大学院大学 情報科学研究科

2006 年 2 月 9 日

キーワード: Speech Synthesis, Speech Recognition, Lombard Effect.

1 はじめに

自動車内ではロンバード効果等の発話歪みが発生し、車室内音声認識システムの精度を低下させる一因となっている。HMM を用いた音声認識では発話歪みに対応した音響モデルを学習するために車室内で収録された巨大な音声データベースが必要となるが、現在の音声データベースのほとんどは雑音下で収録されたものではなく、車室内発話音声サンプルの数は多くない。原因として車室内での様々な状態における音声収録が困難なことが挙げられる。

そこで、本稿では車室内音声認識システムの精度を改善するため、音響モデル学習用にクリーンな音声から車室内発話音声を合成する手法について述べる。自動車内を特徴付ける要因として自動車の速度に着目し、自動車の速度を関数とした音響的特徴量の変換関数を構築する。調査した特徴量は母音部における基本周波数、パワー、フォルマント周波数(第1, 第2, 第3)、スペクトル傾斜、音韻長である。

2 提案手法

クリーンな音声の音響的特徴量に対し、自動車の速度に応じた変換を行い、その後合成することにより車室内発話音声を得る。本研究ではこれらのための音声分析変換合成系として STRAIGHT を採用する。

自動車の速度と特徴量との関連を調べるため、まず実際に車室内発話音声サンプルの収録を行った。データの詳細を表に示す。速度ごとに音声を STRAIGHT に読み込ませた後、音韻区間のしゃべり始めとしゃべり終わりを除いた定常状態であると思われる部分について特徴量の測定を行い、平均値をその雑音レベルにおける代表値とした。その後、各速度の代表値に対しクリーン環境に対する変化量を調べた。得られた値から近似曲線を求めることにより、速度に応じた特徴量の変換関数を構築した。

3 分析結果および考察

各特徴量についてクリーンを基準とした変化量を調べた。

基本周波数およびパワーについては各音韻に目立った差はなく、それぞれ増加傾向にあることがわかった。これは走行雑音による発話歪みが顕著に現れた結果と考えられる。このことからこれらについては各音韻毎にではなく、全体を1つの変換関数でまとめてもよいと考えられる。

フォルマント周波数については全体的に増加傾向であるものの音韻間で差があった。また、いくつかの音韻において値にばらつきがあった。これらについては話者や音韻によって値の変動が激しいものがあり、代表値を正確に求めることができなかった可能性がある。また、一般的にロンバート効果ではフォルマント周波数のシフトが起こるとされるが低い周波数帯と高い周波数帯ではシフト方向が異なることが報告されているため、各音韻毎や周波数帯でわけて考えて変換関数を構築する必要があると思われる。

スペクトル傾斜については値にややばらつきがあるものの、速度が上がるにつれて値が増加し、傾斜がゆるやかになっていく傾向が見てとれた。発話歪みによる高域成分の増加が起こったためと思われる。

音韻長に関しては各音韻毎の違いではなく、音韻の順番と強い関連が見られた。第一モーラと最終モーラについて増加傾向が見てとれた。特に最終モーラについては強い増加傾向にあった。そのため、音韻の順番によってルールを作る必要があると思われる。

4 車室内発話音声の合成

今回得られた結果を用いて車室内発話音声の合成を行った。前節を元に各特徴量についての変換関数を構築した。基本周波数およびパワーについては1つの関数、フォルマント周波数およびスペクトル傾斜については音韻毎に5個の関数、音韻長についてはモーラ毎に3個の関数を求めた。

主観的な聴覚印象について述べると、速度が上がるにつれ話者によって多少差はあるが実際の収録音声、合成音声ともに、声の高さについては段々と高く、声の大きさについては段々と大きくなっていくのが聞いてとれた。そのため基本周波数およびパワーについては今回のルール化で問題ないと感じた。フォルマント周波数およびスペクトル傾斜については図2にあるように多少ばらつきがでたためか、声質に違和感を感じるものがあった。今回フォルマント周波数において第2、第3と高域部分にいくほど値の変動が激しかったため、実際の音声と合成音声ではこれらの値が離れており、声質に違いが出た可能性がある。

5 合成音声の評価

本手法を用いて合成した音声は車室内音声認識のための音響モデルの学習において有用であるかの評価を行う。音声認識においては音声の特徴量ベクトルとして MFCC などが用いられることが多い。そのため、合成した音声と収録した車室内発話音声とでこれらの値が近いほど学習にとって有用であると考えられる。そこで合成音声と収録音声に対し特徴量ベクトルの分布を調べ、その分布間をマハラノビス距離を用いて測定することで評価を行った。その結果クリーン音声と比べて合成音声は車室内発話音声に近づいているとの結果が得られた。これは音響モデルの学習において合成音声はクリーン音声よりも有用である、ということが期待される。

6 おわりに

本稿ではクリーンな音声と車室内発話音声の音響的特徴量を自動車の速度毎に比較し、速度に応じた特徴量の変換関数を構築し、車室内発話音声の合成を行った。その結果合成音声はクリーン音声よりも音響モデルの学習において有用であるという結果が得られた。本システムを用いて大量の車室内発話音声を作成し、それを音響モデル学習に用いることで車室内音声認識システムの性能向上が期待できる。