

帯域パワーのみによるワードスポッティングの検討

5H-1

青木昌紀 小池敏和 福島学 荒井秀一 城戸健一

(千葉工業大学)

1.はじめに

音声のマッチングにおいて帯域パワーを用いる場合、1/3ないし1/6オクターブバンドの数十帯域が用いられる[1][2]。しかし、帯域数を少なく、かつ、マッチングの精度の劣化を少なくできれば、計算量を減らすことができ、有用である。

一方、音声の特徴がよく現れる帯域は、1k~2kと3k~4k付近であるとの報告がある[3][4][5]。

本研究では、ワードスポッティングに用いる特徴量の検討として、帯域を少なくした場合についての検討を行った。これらの帯域を用いた2帯域、および、それ以外の2帯域、500~4kHzの1/3OctBandの帯域を用いた場合の結果とを比較した。

2.実験

実験では、標準単語と正検出単語との距離と、ニュースの他の音声の距離との比較を行った。約20秒間のニュースと、それから切り出した音声を標準単語とし、距離を調べた。標準単語、及び、ニュースをフレーム分割し、標準単語をニュース上で1フレームづつずらしながらDPマッチングで距離を調べる。予め、標準単語として使用した単語「九州」とニュース上の他の時間に出現する「九州」との距離を調べる。実験を行うとき、DPマッチングの経路の3/4以内で、予め調べた距離以下になれば検出されたとした。

実験条件は次の表の通りである。

表1.データ

ニュース	15秒間 (NHK、男性話者)
標準単語	「九州」 (ニュース中)
サンプリング周波数	12kHz
フレーム長	21.3[ms]
フレーム周期	2.6[ms]

表2.特徴量

特徴量	帯域パワー
帯域	500~4kHz 1/3octBand
全帯域数	10帯域

標準単語は、子音と母音の区別が明瞭な単語として「九州」を用いた。子音と母音の区別は、帯域パワーの近似直線より求めた傾き[1]を使って調べた。実験で使用した標準単語のスペクトルの傾きを図1に示す。図1のように、実験で使用した標準単語は、子音と母音とが明瞭に分かれている。

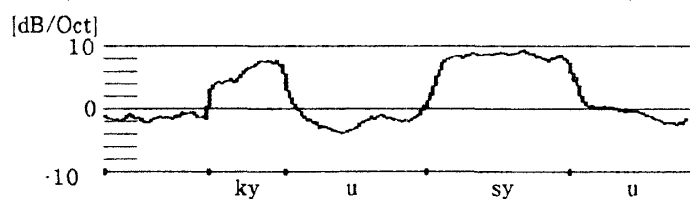


図1. 「九州」のスペクトルの傾き

A study of wordspotting using Band power

Masaki Aoki, Toshikazu koike, Manabu Fukushima, Shuiti Arai, Ken'iti Kido (Chiba Institute of Technology)

3.結果

帯域による正検出単語と湧き出し単語数を表3~5に示す。検出された単語が目的とする単語なら正検出し、他の単語なら湧き出し単語とした。また、低域と高域との境は2.5kHzとした。

表3. 1k~2k[Hz]と3k~4k[Hz]の組み合わせ

組み合わせ帯域 kHz	検出数 (正検出/湧出)
1.5kHzと4.0kHz	2/2
1.2kHzと3.1kHz	2/0

表4. 低域のみの組み合わせ

組み合わせ帯域 kHz	検出数 (正検出/湧出)
0.5kHzと0.8kHz	2/5
0.5kHzと1.0kHz	2/4
0.5kHzと2.0kHz	2/0
1.0kHzと2.5kHz	2/1

表5. 高域のみの組み合わせ

組み合わせ帯域 kHz	検出数 (正検出/湧出)
3.1kHzと4.0kHz	2/2

表5. 全帯域

組み合わせ帯域 kHz	検出数 (正検出/湧出)
全帯域 (10チャンネル)	2/0

4.検討

以上の結果より、2帯域のみでワードスポッティングをおこなっても、離れた帯域を使えば1/3octBandの10帯域全部を使用した場合と違いは見られない。しかし、近接した帯域を使用すると湧き出し単語が多くなる。そのため、近接していなければ、2帯域でも検出できると言える。今後、別の単語、別の話者についても検討する必要がある。

[参考文献]

- [1] 三輪、新津、牧野、城戸：“音声スペクトルの概略形とその動特性を利用した音声認識システム”、日本音響学会誌、Vol.34、No.3、pp186-193、1978/3
- [2] 坂井、中川：“不特定話者、連続音声向き単語音声の識別”、情報処理学会誌、Vol17、No7、pp650-657、1976/6
- [3] 北村、赤木：“音声のスペクトル包絡に含まれる個人性について”、信学技報、SP93-146、pp17-23、1994/3
- [4] 村上、板倉：“音声認識に有効な周波数帯域について”、信学技報、SP94-61、pp1-5、1994/09
- [5] 青木、小池、福島、荒井、城戸：“ワードスポッティングにおける合成単語法に対する補助的方法の検討”、日本音響学会平成8年度春季研究発表会講演論文集、Vol1、pp119-120、1996/3