

Wavelet を用いた特徴量抽出法とその高精度化手法の評価

松井 清彰^{1,a)} 能勢 隆^{1,b)} 伊藤 彰則^{1,c)}

概要: 音声認識の普及のために、より安価な音声認識システムの実現が必要である。音声認識の低演算量化に関しては様々な先行研究が行われているが、特徴量抽出処理に関しては研究が不十分である。そのため我々は、Wavelet 変換を用いた新しい低演算量特徴量抽出法およびその高精度化手法について提案してきた。本論文では、Haar Wavelet 及び Daubechies Wavelet の 2 種類の Wavelet を用いて特徴量抽出を行い、その性能を MFCC と比較した。その結果、高精度化手法を用いることで、若干の認識率の向上が見られた。また、フレーム間の動的特徴量である Δ 特徴量及び MFCC と同様に、DCT 出力の高次削減によって、さらに認識率を向上させることができた。一方、計算時間に関しては、最もシンプルな Wavelet を用いることで、MFCC の 5 倍以上の計算速度を確保できることが分かった。

1. はじめに

近年、我々の身の回りには、音声認識機能を備えた機器が多く表れるようになってきた。Apple 社から発売された「Siri」や、NTTDocomo の「しゃべってコンシェル」等、スマートフォンに音声認識機能が搭載され、音声を用いた様々な検索、メール作成、スケジュール閲覧等の機能の操作、さらには簡単な会話まで行うことができるようになってきた。しかし、例えば 100 円ショップに売られているような道具や玩具に音声認識を用いることは、未だ難しい。

身の回りのものに対して、より気軽に音声認識を取り入れるためには、安価な音声認識システムの開発が必要であるが、その場合、必然的に CPU も小規模なものになってしまう、従来の音声認識システムが満足に動作しない、あるいは消費電力が大きいという問題点がある。そのため、小規模な CPU 上でも低消費電力で動作する低演算量音声認識システムの開発が重要になってくる。

頑健で高性能な組込音声認識実現のためには大きな課題がいくつもあり [1]、低演算量音声認識の実現のため、今まで様々な研究が行われてきた。例えば、Han らは、フレーム処理及びフィルタバンク処理に工夫を加えることで、MFCC の計算量を軽減させる手法を提案した [2]。また、Novuk らは、2 パスデコーディング処理を用いたデコーディングの改善策を提案した [3]。一方、メモリ削減に関しては、Entropy Pruning [4,5] を用いた言語モデル圧

縮や、GMM 音響モデル圧縮 [6] の他、近年ではタブレットを対象にした比較的規模の大きい組込音声認識への取り組み [7] も行われている。しかし、特徴量抽出においては、先行研究が十分ではない。そこで我々は、新たな低演算音響特徴量抽出法として、Haar Wavelet 変換を用いた手法 [8]、及びその高精度化 [9] について提案してきた。本論文では、Haar Wavelet 及び Daubechies Wavelet の 2 種類の Wavelet を用い、提案手法を用いた特徴量抽出を行った時の性能について、従来手法との比較を行った。

2. Wavelet 変換

2.1 離散 Wavelet 変換の概要

Wavelet 変換 [11] は、周波数解析手法の一つであり、主に画像処理や信号のノイズ処理等に用いられる。FFT では、入力音声を正弦波の線形結合とみて各成分に分解していくが、Wavelet 分析の場合、基底関数には Wavelet 関数を用いる。また、デジタル信号のような離散的な信号に対しては、離散 Wavelet 変換が用いられる。

離散 Wavelet 変換による信号の分解は、1 組のハイパスフィルタ (HPF) とローパスフィルタ (LPF) を繰り返し適用させていく処理とみなすことができる (図 1)。Wavelet 変換の分析の深さは、レベルで表される。HPF 出力はそのレベルでの分析結果となり、LPF 出力は次のレベルでの入力となる。次レベルでは、前レベルの LPF 出力を入力とし、新たな LPF 及び HPF をかけて信号を分解する。これを繰り返していくことで、信号の分解を行う。これにより、周波数領域で見ると、信号は、高域側から半分ずつの帯域に分けられ、分析が行われることになる。図 2 は、

¹ 東北大学
Tohoku University, Sendai, 980-8579, Japan
a) k_matsui@spcom.ecei.tohoku.ac.jp
b) tnose@m.tohoku.ac.jp
c) aito@spcom.ecei.tohoku.ac.jp

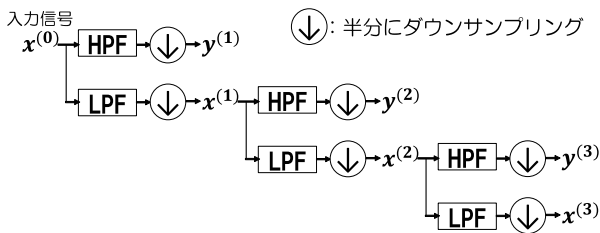


図 1 離散 Wavelet 変換のフィルタ分析 (3 レベルの処理の場合)

周波数分析の様子を表したものである。フィルタの中心周波数はナイキスト周波数 f_q で決まり、レベル 1 では f_q から $f_q/2$ までの HPF が、レベル 2 では $f_q/2$ から $f_q/4$ までの BPF がかけられている。レベル 2 以降の BPF は、現在のレベルでの HPF と、前のレベルでの LPF から作られている。本研究では、Wavelet 関数として、Haar [10] 及び

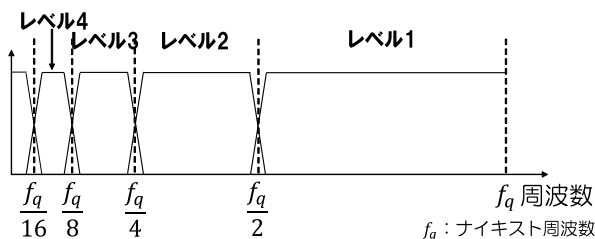


図 2 離散 Wavelet 変換による周波数分析の様子

Daubechies [11] Wavelet を取り上げた。

2.2 Haar Wavelet と Daubechies Wavelet

Haar Wavelet は、最も単純な Wavelet である。

k レベル目の入力を $\mathbf{x}^k = x_0^{(k)}, x_1^{(k)}, \dots, x_{N_k-1}^{(k)}$ とした時、Haar Wavelet 変換により求められる次のレベルの入力 $\mathbf{x}^{k+1} = x_0^{(k+1)}, x_1^{(k+1)}, \dots, x_{(N_k-1)/2}^{(k+1)}$ 及び k レベル目の出力 $\mathbf{y}^{k+1} = y_0^{(k+1)}, y_1^{(k+1)}, \dots, y_{(N_k-1)/2}^{(k+1)}$ は、

$$x_n^{(k+1)} = \frac{x_{2n}^{(k)} + x_{2n+1}^{(k)}}{2} \quad (1)$$

$$y_n^{(k+1)} = \frac{x_{2n}^{(k)} - x_{2n+1}^{(k)}}{2} \quad (2)$$

のようになる。ただし、 $0 \leq n \leq (N_k - 1)/2$ である。このようにして、 k レベル目のフィルタ出力及び次レベルの入力を求め、レベルを上げて繰り返し計算をしていく。その後、各レベルで得られたフィルタ出力 $\mathbf{y}^{(k+1)}$ について、二乗和をとり、サンプル数で正規化することで、帯域パワー p_k を得る。

$$p_k = \frac{1}{N_k} \sum_{n=0}^{N_k-1} (y_n^{(k)})^2 \quad (3)$$

Haar Wavelet は、2 点から新たな 1 点を出力する計算の特性上、分析レベルが上がる毎に入力点数が半分になって

いく。従って、最初のサンプル点数が 2 の累乗であることが理想である。MFCC と同程度のフレーム幅で、かつこの条件を満たすようにフレーム幅を決めていくと、16kHz でサンプリングされた音声については、16ms の窓幅で分析を行った場合、1 フレームあたりのサンプル数が 256 点 ($=2^8$) となり、8 次元の帯域パワーを得る。

帯域パワーを得た後は、MFCC と同様に対数を取って離散コサイン変換 (DCT) を行う。このようにして、MFCC と同様の特徴量を得ることができる。

Daubechies Wavelet は、より複雑な Wavelet である。そのため、予め係数の数列が与えられている [12]。数列の要素数は、ある自然数 N を選んだ時、その 2 倍 ($2N$ 個) となる。例えば、 $N = 2$ の Daubechies の場合、初期に与えられる点は 4 点であり、

$$p_m = [0.4830, 0.8365, 0.2241, -0.1294] \quad (4)$$

となる。これを用いた Daubechies Wavelet の計算は、次のようになる。

$$x_n^{(k+1)} = \sum_{m=0}^{2N} p_m x_{2n+m}^{(k)} \quad (5)$$

$$y_n^{(k+1)} = \sum_{m=0}^{2N} q_m x_{2n+m}^{(k)} \quad (6)$$

ただし、 $q_m = (-1)^m p_{1-m}$ である。

各レベルの出力を計算した後は、Haar Wavelet と同様、各レベル毎の帯域パワーを求め、対数を取って DCT を施す。

Haar 及び Daubechies のフィルタとしての特性を比較する。図 3 は、Haar Wavelet 及び $N = 2$ の時の Daubechies (Daubechies2) Wavelet を用いて信号の分析を行った時のフィルタ群の特性を示したものである。Daubechies2 Wavelet は、Haar に比べ、急峻な特性が得られていることが分かる。

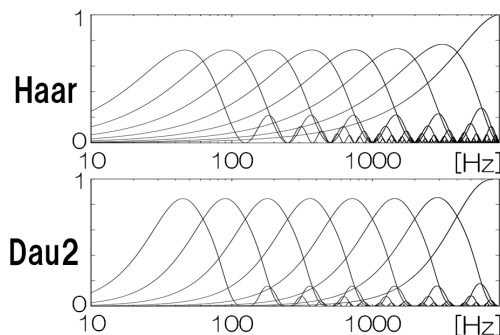


図 3 Haar と Daubechies2 Wavelet のフィルタ特性の比較

2.3 Wavelet の利点と欠点

Wavelet 特徴量の最大の利点は、その計算時間の少なさである。Wavelet 変換に要する計算時間は $O(n)$ であり、これは FFT の計算時間 $O(n \log n)$ よりも高速である。また、FFT では複素数演算が必要であったが、Haar Wavelet 変換を用いた場合、整数のみでの計算が可能である。Daubechies Wavelet の計算には実数が必要であるが、前節で述べた係数 p_m を整数を用いた分数で表して有理数化することで、整数のみでの演算を行うことができる。整数のみで計算が行えることの利点としては、メモリ削減の他、その後の対数変換 [13] 及び DCT [14] を高速に行えるということが挙げられる。

一方で、Wavelet 変換を用いた分析では、周波数分解能が低いことが欠点となる。サンプリング周波数を 16kHz とし、窓幅を 16ms とした時、1 フレームあたりのサンプル点数は $256 (= 2^8)$ 点になる。すなわち、この条件下では 8 帯域の分析が可能になるが、MFCC では、25ms 程度の窓幅で 20 帯域以上の分析を行っている。また、Wavelet 変換を用いた特徴量抽出法は、分析帯域を直接変化させることは難しい。サンプル点数を増加させることで帯域数を増加させることは可能だが、1 帯域増やすのに 2 倍の個数のサンプル点数が必要であり、帯域数を M から $2M$ に倍増させた場合、サンプル点の個数は 2^M 倍と膨大になり、計算時間でのメリットが失われてしまう。この問題を解決するために、我々は、ダウンサンプリングを用いた高精度化手法を提案した。

3. ダウンサンプリングを用いた Wavelet 特徴量の高精度化

Wavelet 分析によって得られるフィルタの中心周波数は、ナイキスト周波数で一意に決まる。図 4 は、ナイキスト周波数が 8kHz の場合の中心周波数の分布を示したものである。これを見ると、中心周波数が、ナイキスト周波数 f_q に対し、 $f_q/2, f_q/4, f_q/8, \dots, f_q/256$ と分布している様子が分かる。従って、ダウンサンプリングを行い、音声信号のナイキスト周波数をずらすことで、フィルタ全ての位置をずらすことが可能にある。

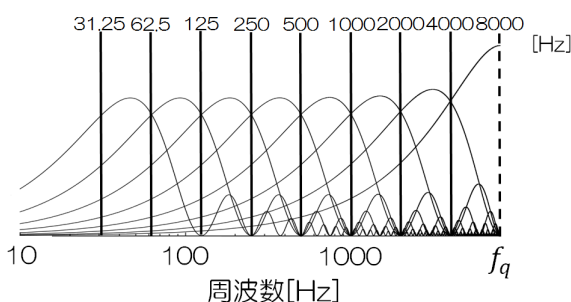


図 4 Haar Wavelet のフィルタ群における中心周波数 (16kHz, 8 レベルの場合)

フィルタをずらす位置としては、ずらす前のフィルタの中心周波数の丁度真ん中に、新しいフィルタの中心周波数をずらすのが理想的である。すなわち、 $1/\sqrt{2}$ 倍にダウンサンプリングを行えばよい。しかし、ダウンサンプリング処理を簡潔に行うため、倍率を簡単な整数比である $2/3$ 倍に近似してダウンサンプリングを行った。

簡略的なダウンサンプリング処理について説明を行う。通常、 $2/3$ 倍へダウンサンプリング処理を行う場合、2 倍にアップサンプリングしてから $1/3$ 倍へのダウンサンプリングを行うが、計算時間の短縮のため、簡潔な計算により直接 $2/3$ 倍へのダウンサンプリングを行った。具体的には、まず音声信号のサンプル点を連続する 3 点で一つのまとまりとする。次に、 k 番目のまとまり内の点 $x_0^{(k)}, x_1^{(k)}, x_2^{(k)}$ に対し、

$$y_0^{(k)} = \frac{x_{2n}^{(k)} + x_{2n+1}^{(k)}}{2} \quad (7)$$

$$y_1^{(k)} = \frac{x_{2n}^{(k)} + x_{2n+1}^{(k)}}{2} \quad (8)$$

という計算を行い、ダウンサンプリング後の出力 $y_0^{(k)}, y_1^{(k)}$ を得る。この処理は、Haar Wavelet の 1 レベル目における LPF 処理と同様である。これにより、 $2/3$ 倍へのダウンサンプリング及び折り返し歪抑制のための LPF 処理を行う。

ダウンサンプリングを用いた高精度化特徴量抽出の手順を図 5 に示す。フレーム処理を行った音声に対し、まずそのまま Wavelet 変換を行い、 M 帯域の特徴量を得る。次に、音声信号にダウンサンプリングを行い、その信号に対してもう一度 Wavelet 変換を行う。そうすると、それぞれで得られる特徴量の分析帯域は均等にずれているため、これらの特徴量を組み合わせて、 $2M$ 帯域の特徴量を得ることができる。

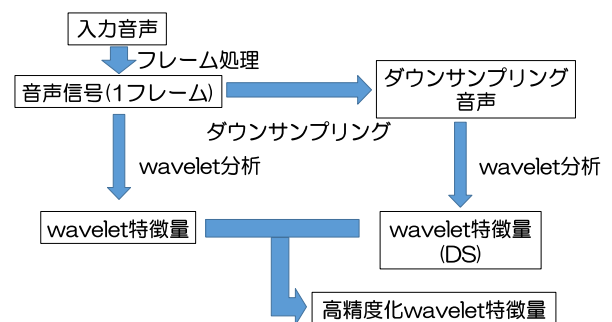


図 5 ダウンサンプリングを用いた Wavelet 特徴量の高精度化

3.1 3 倍高精度化処理

上記と同様の原理により、周波数分解能を 3 倍にすることが可能である。この場合、中心周波数はそれぞれ $1/2^{1/3}, 1/2^{2/3}$ 倍になればよいが、2 倍高精度化の時と同様に、近い整数比である $4/5, 2/3$ 倍へと近似を行った。このようにして、元

音声及び2種類のダウンサンプリング音声からそれぞれ特徴量を得ることで、帯域数を3倍にすることが可能になる。

4. 評価実験

提案手法の性能を図るため、認識率及び計算速度について、実験により比較を行った。

4.1 認識率評価

4.1.1 実験条件

東北大・松下単語音声データベース [15] より、話者 60 名 (男性 30 名, 女性 30 名) を使用した。このデータベースには、各話者ごとに 212 種類の単語発話が収録されており、この単語は、全ての音素が出現するようにバランスされている。男女各 10 名ずつを学習データ、残りの 40 名を認識データとして使用した。

また、使用特徴量として、MFCC 及び Haar Wavelet, Daubechies Wavelet を使用した。MFCC は、窓幅を 25ms, フレームシフトを 10ms とし、 Δ 特徴量を用いた 26 次元の特徴量を使用した。Wavelet 特徴量に関しては、高精度化 3 パターン (無し, 2 倍, 3 倍) と Δ 特徴量の有無 (有, 無) について条件を変化させ、各特徴量につき 6 パターンを用いて比較を行った。また、MFCC と同様に、DCT 後に得られる特徴量の高次成分を削減することで認識率が向上することが考えられるため、合わせて検討を行った。窓幅は、高精度化手法を用いない場合は 16ms, 高精度化手法を用いる場合は 24ms とし、フレームシフトはいずれの場合も 10ms とした。音響モデルには、5 状態の triphone HMM を用いた。混合数は 1 とした。

4.1.2 実験結果

実験結果は、図 6 のようになった。また、Haar Wavelet 及び Daubechies4 Wavelet における、次元削減による認識率の変化はそれぞれ図 7, 図 8 のようになった。

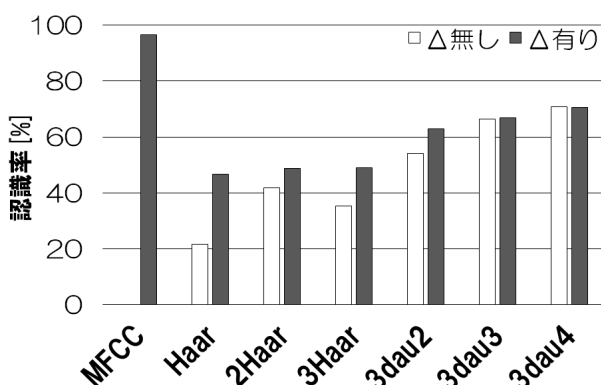


図 6 東北大・松下 212 単語を学習データに用いた認識実験の結果

なお、Daubechies Wavelet に関しては、認識率に見られる傾向が Haar Wavelet と同様であったため、3 倍高精度化条件のみを記載している。3 組のグラフのうち一番右

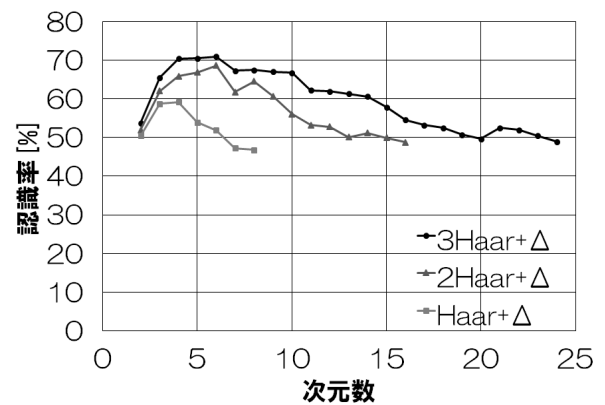


図 7 Haar Wavelet における次元削減による認識率の変化

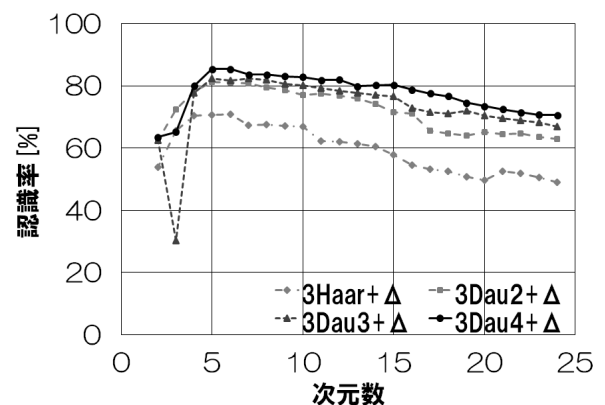


図 8 各 Wavelet 毎の次元削減による認識率の変化

の「 Δ +次元削減」の項目は、次元削減を行った時の認識率の最大値を示している。

各種 Wavelet 特徴量は、 Δ 特徴量を用いることで、認識率の向上が実現できた。そのため、 Δ 特徴量は Wavelet 特徴量にも有効である。また、MFCC の認識率は 96.51% と、非常に高い値である。

次に、高精度化と次元削減の結果について見ていく。図 7 は、Haar Wavelet について、高精度化手法と次元削減の関係を調査したものである。結果として、高精度化の有無に関わらず、次元削減を行うことにより認識率は向上し、さらに次元削減を行うと、急激に認識率が落ちるといった傾向が明らかになった。また、高精度化手法を用いた場合、次元削減によって到達する認識率の最大値がより高くなるのが分かった。

また、図 8 では、Wavelet の種類と次元削減について調べたグラフである。すべての条件で、3 倍高精度化を行い、また Δ 特徴量も使用した。その結果、複雑な Wavelet を用いるほど、認識率が向上することが分かった。特に、Haar と各種 Daubechies では、認識率に一定の開きがあった。

4.2 計算速度評価

4.2.1 実験概要

東北大・松下単語音声データベースの話者 60 名を用い

て、特徴量抽出に要する時間を計測した。各ファイルの処理に要した時間の平均と、音声ファイルの長さの平均を算出し、リアルタイムファクタを求めた。リアルタイムファクタ xRT は、次式で求められる。

$$xRT = \frac{\text{音声ファイルの処理時間}}{\text{音声ファイルの長さ}} \quad (9)$$

xRT が小さいほど、実時間処理に適しているといえる。

4.2.2 実験条件

認識率評価実験と同様、東北大・松下 212 単語より、話者 60 名を使用した。総発話数は 12767 発話となり、平均発話時間は 1.346 秒であった。

特徴量については、MFCC は Δ 特徴量を用いた 26 次元を使用した。Haar Wavelet 及び Daubechies Wavelet については、高精度化の有無のみを変化させ、 Δ 特徴量を用いなかった。

測定にあたっては、ファイル入出力の時間は除き、特徴量抽出処理に要する時間のみを計測した。また、各音声ファイルごとに処理時間を算出し、それを全データに渡って平均し、平均処理時間を算出した。これと前述の平均発話時間を用いて、リアルタイムファクタ xRT を算出した。

また、実験は、Raspberry Pi [16] 上で行った。Raspberry Pi は、ARM プロセッサを搭載したシングルボードコンピュータで、クロック周波数は 700MHz、語調は 32 ビットである。また、計算においては浮動小数点を扱え、乗算器も搭載されている。

4.2.3 実験結果

実験結果は、図 9 のようになった。MFCC の xRT は、

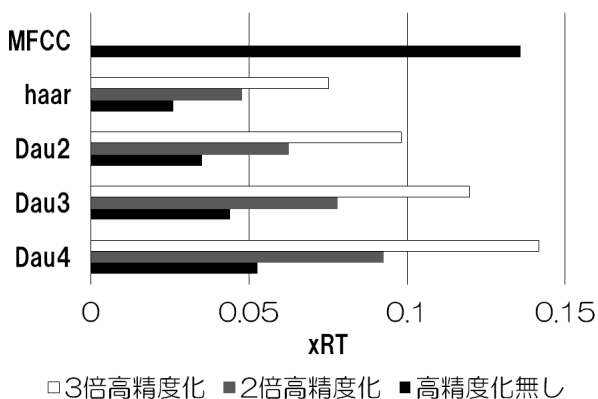


図 9 計算速度比較実験の結果

0.136 であった。ただし、実際の音声認識処理には、この後にモデルと辞書を用いたデコーディングが行われるので、この値単体でリアルタイム性を議論することはできない。一方、Haar Wavelet 特徴量を用いた場合、 xRT は 0.0260 で、MFCC の 5 倍以上の速度で計算が行えていることが分かる。Daubechies4 Wavelet は、認識率では他の Wavelet よりも高かったが、 xRT では 0.142 と MFCC よりも計算

速度が遅くなった。

Haar Wavelet に高精度化手法を用いない場合、計算速度は MFCC の 5 倍以上になるが認識精度は 20% と非常に低かった一方で、Daubechies4 Wavelet に 3 倍高精度化を用いた場合、計算時間は最低値であったが 70% 程度の認識率を出すことができた。このことより、認識率と計算速度はトレードオフの関係になっていることが分かった。さらに、 Δ 特徴量及び次元削減を行うことで、どの Wavelet 特徴量に対しても、大幅に認識率を向上させることができた。特に、次元削減は効果が大きいことに加え、特徴量抽出処理の計算時間を圧迫しないため、積極的に用いるべきである。また、要求されるタスクの難度に応じて可能な限りシンプルな Wavelet を用いることで、高速な特徴量抽出が可能になる。

タスクの難度については、中川が 1990 年に行ったシミュレーションが一つの指針となる [17]。このシミュレーション結果として、誤認識率は、語彙数の対数あるいは平方根にほぼ比例すると結論付けられている。すなわち、タスクの難度を測る重要な指標の一つとして語彙数が挙げられる。従って、特徴量を選択する場合は、目的とするタスクの語彙数に着目することで、想定される認識率の目安を知ることができる。

そこで、語彙数と認識率の変動について、検証を行った。

4.3 計算速度評価

4.3.1 実験概要

4.1 節で行った実験を、語彙数を変化させながら行い、認識率の変化を検証した。

4.3.2 実験条件

使用データ、音響モデルの条件は 4.1 節と同様である。特徴量については、Haar Wavelet 及び Daubechies Wavelet を用いた。また、すべての条件で 3 倍高精度化、及び Δ 特徴量を用いた。更に、用いる次元数については、図 8 において、最大の認識率を得られた時の次元数を選択した。すなわち、Haar では 5 次元、Daubechies2, 3, 4 では、それぞれ 4, 6, 4 次元の Wavelet 特徴量を使用した。実験は、認識に用いる単語をランダムに変化させながら各条件毎に 50 回行い、認識率の平均値を記録した。なお、50 回の試行の間に、212 単語全てが、少なくとも 1 回は出現した。

4.3.3 実験結果

実験結果は、図 10 のようになった。グラフより、語彙が少なくなるほど、認識率が高くなる傾向が示された。また、語彙数が少ない場合は用いる Wavelet による認識率の差は小さいが、語彙数が増えるにつれて、Haar Wavelet 等のシンプルな Wavelet は認識率が急激に悪化した。これは、語彙数の増加によって音響的に似た単語が増えることで、シンプルな Wavelet では区別が難しくなったためと考えられる。システムとして実用に耐えうる境界として、認識率

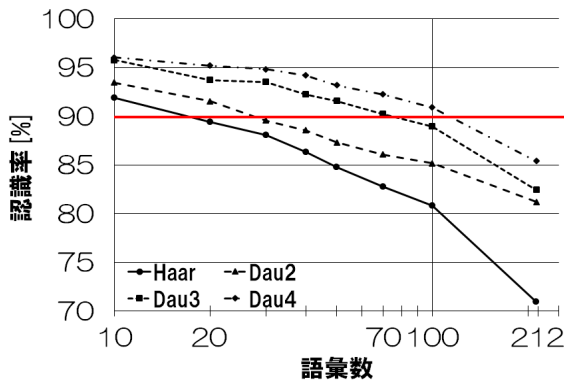


図 10 語彙数による認識率の推移

90%以上を設定した場合, Haar, Daubechies2, 3, 4 では, それぞれ 10, 20, 70, 100 単語までを使用できるという結果になり, 10 単語程度の語彙を持つ単語認識システムであれば, Haar Wavelet を用いても十分な認識精度を得られることが分かった. しかし, Daubechies2, 3 Wavelet を用いることで, MFCC より高速に, かつ Haar より大語彙での認識が行える.

5. まとめ

Wavelet 特徴量を用いた特徴量抽出は, MFCC より高速な計算が可能であるが, 認識精度が劣るという欠点を抱えていた. 本論では, 音声信号に対してダウンサンプリングを行うことで, ナイキスト周波数をずらし, フィルタの分布をずらすことによる高精度化手法を提案した. また, 認識率及び計算速度を実験によって比較したところ, 認識率と計算速度は, トレードオフの関係になっていることが分かった. また, 具体的な語彙数と用いる Wavelet による認識率の関係性について実験を行い, 10 語彙程度の極めて小語彙であれば, Haar Wavelet のような非常にシンプルな Wavelet でも十分認識に用いることが可能であることが分かった. これらの結果を踏まえ, タスクの難度や語彙数に応じて適切な精度の特徴量を使用するべきである. 今後は, 他種の Wavelet を用いた分析や, 特徴量抽出における他の処理 (対数変換, DCT 等) の計算速度向上について取り組んでいきたい.

参考文献

- [1] J. Cohen, "Embedded speech recognition applications in mobile phones: status, trends, and challenges," *ICASSP 2008*, pp. 5352–5355, 2008.
- [2] W. Han, C. Chan, C. Choy and K. Pun, "An efficient MFCC extraction method in speech," *ISCAS 2006*, pp. 145–148, 2006.
- [3] M. Novuk, R. Humpl, T. Krbec, K. Bergl and J. Sedivy, "Two-pass search strategy for large list recognition on embedded speech recognition platforms," *ICASSP 2003*, pp. 200–203, 2003.
- [4] A. Stolcke, "Entropy-based pruning of backoff language models," *In Proc. DARPA broadcast news transcription*

- and understanding workshop*, pp. 270–274, 1998.
- [5] C. Chelba, T. Brants, W. Neveitt and P. Xu, "Study on interaction between entropy pruning and Kneser-Ney smoothing," *Interspeech 2010*, 2010.
- [6] C. Levy, G. Linares and J.P. Bonastre, "GMM-based acoustic modeling for embedded speech recognition," *Interspeech 2006*, 2006.
- [7] X. Lei, A. Senior, A. Gruenstein and J. Sorensen, "Accurate and compact large vocabulary speech recognition on mobile devices," *Interspeech 2013*, 2013.
- [8] 松井清彰, 伊藤彰則, "組込型音声認識システムのための低演算特徴量抽出法", 情報処理学会第 75 回全国大会, Vol. 2, pp. 527–528, 2013.
- [9] 松井清彰, 千葉祐弥, 能勢隆, 伊藤彰則, "組込型音声認識システムのための Haar-Wavelet 変換を用いた音声認識特徴量抽出法の高精度化", 日本音響学会春季講演論文集, pp. 71–72, 2014.
- [10] A. Haar, "Theorie der orthogonalen funktionen-systeme," *mathematische Annalen*, Vol. 69, pp. 331–371, 1910. (G. Zimmermann(Translation), "On the Theory of Orthogonal Function Systems," (<https://www.uni-hohenheim.de/gzim/Publications/haar.pdf>) (accessed 2015-02-03))
- [11] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE transactions on information theory*, Vol. 36, No. 5, pp. 961–1005, 1990.
- [12] I. Daubechies, "Ten Lectures on Wavelet," SIAM, 1992.
- [13] article:Integer Log functions (from EmbeddedGurus), (<http://embeddedgurus.com/stackoverflow/2008/05/integer-log-functions/>) (accessed 2015-02-03)
- [14] N. Brahimi and S. Bouguezel "An efficient fast integer DCT transform for images compression with 16 additions only", *2011 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, pp. 71–74, 2011.
- [15] 牧野正三, 二矢田勝行, 真船裕雄, 城戸健一, 「東北大・松下単語音声データベース」音響誌, Vol. 48, No. 12, pp. 899–905, 1992.
- [16] Raspberry Pi Foundation, (<http://www.raspberrypi.org/>) (accessed 2015-02-03).
- [17] 中川聖一, "音声認識・理解システムの評価とデータベース", 電子情報通信学会誌, Vol. 73, No. 12, pp. 1304–1310, 1990.