5S-08

音声信号からの身長推定手法の研究

宮　昊†　　浜田　宏一†　　荒井　正之†

帝京大学理工学部電子情報工学科†

1. はじめに

音響信号から，それに関連する物理量を推定する研究が行われている．例えば，声で人の身長が推定できれば，犯罪捜査に役立つ．また，身長はあまり変化しないことから，音声認証への活用も可能である．

声が作られるのは，気管の入口にある声帯である．息を吐きながら声帯の付いている声門を微妙に開け閉めすることで声が生まれる[1]．また，声そのものではないが，声道の長さである声道長は身長体重と関係あることが示されている[2]．声道長が変化すれば，それに応じて声の特徴も変化するはずであると考え，声で身長を推定する研究を行うこととした．

2. 実験用音声データの収集

本学科所属の学生と教員に被験者を依頼し，日本語 50 音の音声データを取得した．個人が特定できないように音声データは ID で管理し，身長・体重のデータと一緒に記録した．

録音時の設定を表 1 に示す．信号の劣化を抑えるため，低域カットフィルタの機能はオフにした．

表 1　録音設定

保存形式	Wav 16bit
サンプリング周波数	44.1kHz
チャンネル	ステレオ
低域カットフィルタ	Off
出力レベル	-12db

3. 使用する特徴量の検討

声道・声帯に関連する特徴量はいくつか存在しているが，今回は MFCC，ケプストラム係数，フォルマント周波数という三つの特徴量を抽出して実験を行った．MFCC とケプストラム係数は 13 次まで，フォルマント周波数は第 5 フォルマントまでを抽出した．同じ特徴量に対しても，全データの平均と最初の 1 フレームのデータに分けて実験を行った．

MFCC (Mel-Frequency Cepstrum Coefficients)

MFCC は，人間の音声知覚の特徴を考慮して，音響特性を表す特徴量である．人間の聴覚には，周波数の低い音に対して敏感で，周波数の高い音に対して鈍感であるという性質がある．メルフィルタバンクを使用することで，低周波成分ほど分解能を高く，高周波成分になるほど分解能を低くなるように，人間の音声知覚に沿う特徴量に変換することができる．

ケプストラム係数

ケプストラム係数は MFCC に似ているが，人間の音声知覚の特徴を考慮しない特徴量である．人間の声においては，音源特性と声道特性を分離することができる．ケプストラム係数の低い部分は声道特性（スペクトル包絡）に対応している．逆にケプストラム係数の高いところは音源特性（音の微細構造）を表すことができる．

フォルマント周波数

フォルマント周波数は声道形状を表す特徴量である．図 1 に示されるように，音声スペクトル包絡には，山と谷が存在している．山のピックのところに対応する周波数をフォルマント周波数という．周波数の低いものから順に第一フォルマント周波数（F1），第二フォルマント周波数（F2）…と呼ばれる．

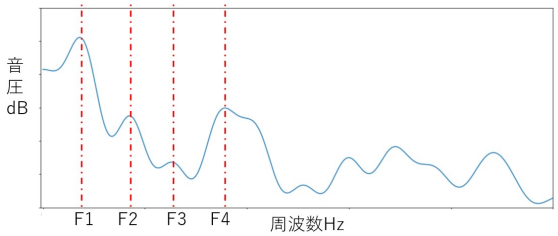

図 1　スペクトル包絡とフォルマント

4. 実験結果と考察

4.1 評価方法

各特徴量からの身長推定結果を R^2（決定係数）と RMSE（平均平方二乗誤差）を用いて評価する．R^2 は回帰分析の当てはまりの良さを表しており，1 に近づくほど精度が高い．

A Study of Height Estimation Methods from Audio Signals

†Hao Gong, Koichi Hamada, Masayuki Arai, Department of Information and Electronic Engineering, Faculty of Science and Engineering, Teikyo University

RMSE は真の値と予測値の差を二乗した値を平均し，平方根をとったものである．単位はデータと同じであり，今回の身長の推定実験では，単位は cm とする．

4.2 実験と考察

図 2 は MFCC を特徴量として用いた場合の実験結果である．各母音に対し，線形回帰と，ランダムフォレストの 2 種類の手法で予測を行った．音声データは，それぞれの手法に対し最初の 1 フレームを用いる場合と，平均を用いる場合を試行した．この実験で最も予測精度が高かったのは，母音として「お」を用い，予測手法がランダムフォレスト（平均）の場合で，R^2 は 0.6 程であった．

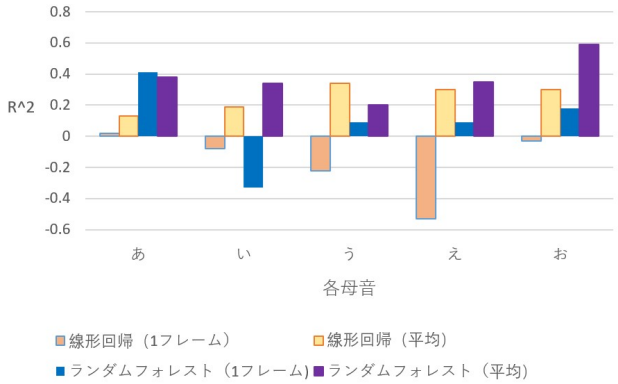

図 2　MFCC での推定結果

図 3 はケプストラム係数を用いた実験結果である．最も良い結果は，「あ」に対するランダムフォレスト（1 フレーム）の方法で，R^2 は 0.7 以上であることがわかる．

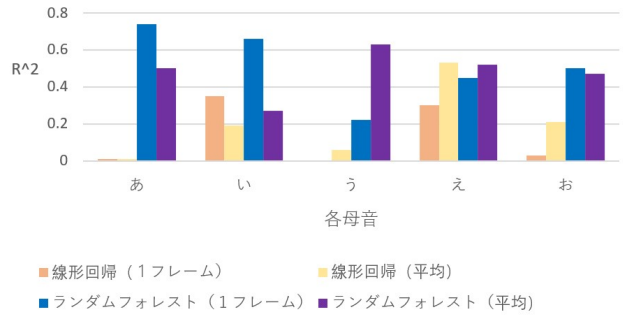

図 3　ケプストラム係数での推定

フォルマント周波数を用いる実験は，現在プログラムの実装中であり，結果はできていないが，身長と各特徴量の相関係数を調べた結果，表 2 に示す結果が得られた．「う」の第 3 フォルマント周波数が身長との相関が最も強いことが確認できた．

表 2　身長とフォルマント周波数の相関係数

	F1	F2	F3	F4	F5
あ	0.00	−0.06	−0.19	−0.06	0.12
い	−0.46	0.18	−0.25	−0.21	−0.02
う	0.20	0.12	**0.74**	0.30	0.16
え	0.07	0.04	−0.35	−0.13	0.07
お	0.14	0.36	0.29	0.25	−0.33

MFCC とケプストラム係数を用いる身長推定実験において，最も精度の高い結果を表 3 にまとめる．MFCC，ケプストラム係数どちらを使用しても，ある程度身長推定ができる見込みを得た．また，ノイズの影響が低減できることから，データの時間平均値で推定したほうが良いと考えたが，最初の 1 フレームのみを用いた方が，推定精度が高い場合もあることが分かった．どの母音に対する推定精度が最も高いのかも，今後の検討課題である．

表 3　精度の最も高い予測方式の詳細

特徴量	MFCC	ケプストラム係数
R^2	0.59	0.74
推定誤差（RSME）	3.78 cm	2.97 cm
推定方法	ランダムフォレスト（平均）	ランダムフォレスト（1 フレーム）
対象とする母音	お	あ

5. おわりに

音声信号から得た特徴量から，身長推定が可能な見込みを得た．今回，全て声道に関する特徴量を利用してきたが，声帯に関する基本周波数（F0）も利用して実験を行おうと考えている．さらに，複数の特徴量を同時に利用して，少量のデータで高精度の結果が得られないかの検討を行う．深層学習を用いた推定手法も試行予定である．

参考文献

[1] John Morton: Acoustic features mediating height estimation from human speech , Acoustical Society of America 134, 4072 (2013)
[2] 小林真優子：声から身体情報を求める，情報処理学会第 99 回音楽情報科学研究会，2013-MUS-99-50 (2013)