

自己組織化マップを用いた音声認証
-男女ごとの適正周波数帯-
Voice authentication using Self-Organizing Maps
- proper frequency band for gender-

平澤 翼[†]
Tsubasa Hirasawa

納富 一宏[†]
Kazuhiro Notomi

斎藤 恵一[‡]
Keiichi Saito

1. はじめに

PC の普及により、所持できる情報量が増加し、それに伴い強固なセキュリティが求められてきている。しかし、現在の認証方法はパスワード認証や、IC カード認証が一般的である。これらは、紛失や、他人でも承認されてしまう問題がある。

そこで最近、終生不変、万人不同の性質を持つ行動的または身体的な特徴を用いて個人特定するバイオメトリクス認証が注目されている。

本研究では、バイオメトリクス認証の音声認証を用い、男女の音声に含まれる周波数の違いに着目する。ニューラルネットワークの一種である自己組織化マップによる学習を行い認証精度の算出を行う。男女ごと抽出する周波数帯を変化させて分析を行い、最も認証精度の高くなるものを適正周波数帯とする。

本稿では、男女ごと検討した適正周波数帯について述べる。

2. 音声認証と自己組織化マップ

2.1 音声認証

音声認証とは、行動的特徴を利用したバイオメトリクス認証の一種であり、人が発声した声に含まれる音声特徴量を用い、本人確認を行う認証方法である。

本研究では、あらかじめ決められたキーワードを発声するテキスト従属方式を採用した。

2.2 自己組織化マップ

自己組織化マップ (SOM : Self-Organizing Maps) とは、Kohonen によって提案されたニューラルネットワークの一種であり、入力層と出力競合層の 2 階層で階層型ネットワークを構成している。SOM の特徴は、高次元のベクトル集団を学習することにより 2 次元のマップ上にそれらのベクトル関係を写像することができることができ、類似するベクトルは近い位置に、類似しないベクトルは遠い位置に配置される^[1]

本研究では、音声特徴の分類を行うために SOM を使用した。SOM の出力例を図 1 に示す。

3. 実験

3.1 実験環境

機材と使用ソフトウェアを表 1 に、実験環境の写真を図 2 に示す。録音は雑音の入らない静かな室内で行う。

表 1 機材と使用ソフトウェア

マイク	SANWA MM-MCUSB13
録音ソフト	Sound Engine Free (Version 4.52)
特徴抽出	Audacity (Version 1.2.6)
基本周波数特定	Praat(version 5.2.28)

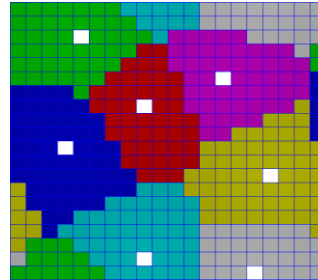


図 1 SOM 出力例



図 2 実験環境

3.2 実験方法

20 代前半の男性 10 名、女性 10 名の計 20 名の被験者に協力を仰いだ。キーワードは先行研究^[2]で採取した音声を再度使用するため「ありがとう」とし、発声した音声を採取した。音声を録音する際、サンプリング周波数 44.1kHz, 16bit, 2ch とした。録音は、被験者 1 名に対し、キーワード 1 回分の音声を 10 個採取した。

上記で得た音声サンプルの音声特徴を抽出するため、発声した波形すべてを選択範囲とし、高速フーリエ変換 (FFT: Fast Fourier Transform) を行う。本研究では、始点を固定し、終点を変化させた 5 種類の周波数帯を決定し、種類ごとの認証精度を算出する。特徴抽出を行う周波数帯の始点は、被験者から採取した音声の基本周波数を調べ、その中で最も低い値を基準として決定する。始点から終点までの特徴点の数は、先行研究^[2]で高い認証精度を示した 23 個とする。周波数帯の終点は、始点から一定間隔で 23 個目とし間隔は 5 種類用意する。

3.3 分析方法

FFT により出力された 23 個の特徴点を正規化し、録音された 1 名当たり 10 個の音声サンプルのうち、最初の 5 個を登録用とする。続いて、残りの 5 個を認証用として属性ベクトルを投入し、SOM を用いてマップを作成する。SOM のマップサイズは 70×70 とし、学習回数は 5 万回として認証精度を算出する。作成する SOM は、男女ごと周波数帯を変化させた 5 種類のものとする。

マップ上に表示された登録用のベクトルと、認証用のベクトルとのユークリッド距離の平均を求め、その値が設定した閾値よりも小さければ認証成功とする^[2]。SOM は毎回異なるマップが生成されるため、5 回作成した平均値を本研究の認証精度とする。

評価には、他人受容率 (FAR : False Accept Rate) と本人拒否率 (FRR : False Reject Rate) を用いた。FAR, FRR の定義式を以下に示す。

$$FAR = \frac{\text{他人受容回数}}{\text{試行回数}}, \quad FRR = \frac{\text{本人拒否回数}}{\text{試行回数}}$$

4. 結果

4.1 実験結果

基本周波数を調べた結果を図 3 に示す。図の下部に表示されている線が基本周波数を示すピッチ曲線である。

[†]神奈川工科大学 Kanagawa Institute of Technology

[‡]国際医療福祉大学情報教育センター Education Center of Medical Informatics, International University of Health and Welfare

この右側に表示されている周波数が、最も低い被験者の基本周波数であり、87.1Hz という値となった。FFT の分解能上この値に最も近い抽出可能な周波数は 86Hz であるため、抽出する周波数の始点は 86Hz とする。

始点 86Hz から FFT で分解可能な約 22,43,86,172,345 の間隔で 5 種類抽出する。したがって、この時の終点は 559,1033,1981,3875,7666 の 5 種類となる。

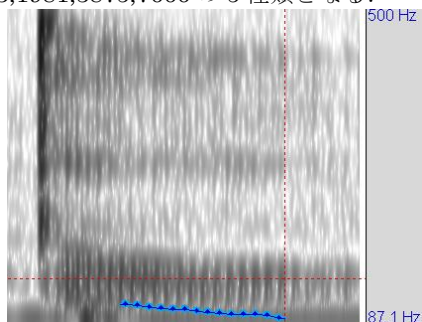


図3 基本周波数解析

4.2 分析結果

実験結果より得られた基本周波数を元に算出した認証精度を表2, 表3に示す。

表2 男性認証精度表

被験者	周波数帯	閾値	FAR[%]	FRR[%]	認証精度[%]	平均精度[%]
男性 10名	86Hz～ 559Hz	20.5	0.17	0.12	85.56	88.16
		19.5	0.14	0.12	87.11	
		18.5	0.11	0.12	88.34	
		19.5	0.16	0.12	86.11	
		16.0	0.07	0.06	93.67	
	86Hz～ 1033Hz	18.5	0.11	0.10	89.56	90.98
		18.0	0.09	0.10	90.34	
		18.5	0.11	0.12	88.67	
		17.5	0.07	0.08	92.34	
		17.0	0.06	0.06	94.00	
	86Hz～ 1981Hz	16.0	0.04	0.06	95.00	95.56
		15.0	0.02	0.02	97.78	
		17.0	0.07	0.08	92.45	
		15.5	0.05	0.06	94.56	
		15.0	0.02	0.02	98.00	
	86Hz～ 3875Hz	14.5	0.01	0.00	99.34	99.53
		14.0	0.01	0.00	99.56	
		14.5	0.01	0.00	99.45	
		13.5	0.00	0.00	99.89	
		14.0	0.01	0.00	99.45	
86Hz～ 7666Hz	16.0	0.01	0.02	98.34	98.33	
	15.0	0.01	0.02	98.34		
	16.0	0.01	0.02	98.45		
	16.5	0.03	0.04	96.56		
	14.0	0.00	0.00	100.00		

表3 女性認証精度表

被験者	Hz	閾値	FAR[%]	FRR[%]	認証精度[%]	平均精度[%]
女性 10名	86Hz～ 559Hz	18.5	0.11	0.10	89.45	89.96
		19.0	0.12	0.10	89.11	
		17.5	0.09	0.10	90.45	
		17.5	0.09	0.10	90.67	
		17.5	0.10	0.10	90.11	
	86Hz～ 1033Hz	18.0	0.11	0.12	88.56	87.11
		18.0	0.11	0.08	90.67	
		19.0	0.14	0.16	85.22	
		18.5	0.10	0.12	88.89	
		20.5	0.18	0.18	82.22	
	86Hz～ 1981Hz	19.5	0.12	0.12	88.22	87.31
		19.0	0.13	0.16	85.34	
		18.5	0.11	0.12	88.56	
		18.0	0.11	0.12	88.45	
		19.0	0.14	0.14	86.00	
	86Hz～ 3875Hz	19.0	0.12	0.08	90.00	88.71
		18.0	0.09	0.12	89.34	
		20.5	0.15	0.16	84.34	
		18.5	0.10	0.12	89.00	
		18.5	0.08	0.10	90.89	
86Hz～ 7666Hz	17.0	0.08	0.08	92.22	88.53	
	19.5	0.11	0.12	88.45		
	19.0	0.12	0.12	88.11		
	18.5	0.11	0.14	87.56		
	19.5	0.13	0.14	86.34		

5.考察

男性の認証精度表(表2)より、男性は抽出する周波数帯を変化させることで認証精度に差が表れることが確認できた。また、3875Hz までと 7666Hz までの 2 種類の周波数帯で平均 98% 以上の高い認証精度を得た。3875Hz までにおいては平均 99.5% 以上の非常に高い認証精度を示した。しかし周波数帯の終点が 3875Hz から低下するにつれて認証精度が大きく低下している。この結果は、4000Hz 以上の周波数で数値が不安定になることや、人間の音声特徴はおよそ 6000Hz までの周波数帯に表れることを示した先行研究^[3]と一致する。また、3875Hz よりも低い周波数帯での認証精度が低い理由として、抽出する周波数の間隔が狭いことから、特徴を十分に抽出することができなかったからだと考えられる。したがって、男性の音声認証を行う場合、被験者中最も低い基本周波数(86Hz)～4000Hz までの範囲が適切であると考えられる。

女性の認証精度表(表3)においては、どの周波数帯で特徴抽出を行ってもあまり変化がみられなかった。認証精度が最も高い値を示したものが 89.9% と男性と比較して非常に低く、最高値を示した周波数帯が 559Hz までと男性とは正反対の結果となった。したがって、本分析方法は女性被験者には適していなかったと考えられる。精度低下の原因は、同条件で分析を行うため、周波数帯の始点を男性と同一にしたが、今回の女性被験者の基本周波数は 200Hz 前後のため、86Hz～200Hz の間の特徴点が認証精度低下の原因になったと考えられる。そのため、現段階では女性に適した周波数帯は判断できない。したがって、更なる分析を行い適した周波数帯の特定が必要である。

6.おわりに

今回、SOM を用いて男女における適正周波数帯の検討を行った。分析結果より、男性で最も高い認証精度を示したのは 5 種類の周波数帯のうち、被験者中最も低い基本周波数(86Hz)～3875Hz までの周波数帯で、認証精度は 99.53% と非常に高い精度を示した。したがって、男性はこの周波数帯が適正であると考えられる。

一方、女性の適正周波数帯は、5 種類の周波数帯ともほぼ同じ認証精度を示した。また、その精度も 88% 前後と男性と比べて大きく低いことが確認された。どの周波数帯でも、実用化を考慮した場合認証精度が十分でないため、更なる分析を行い、認証精度を向上させる必要がある。

今後は、女性被験者の認証精度を向上させるため、今回とは異なる周波数帯での分析を行う必要がある。また、被験者の周波数特性が近いことで認証精度が低下した可能性があるため、今回とは異なる女性被験者の音声サンプルを採取し、分析する必要がある。

参考文献

- [1]大北正昭, 徳高平蔵, 藤村喜久朗, 権田英功: 自己組織化マップとそのツール, シュプリンガー・ジャパン社 (2008).
- [2]平澤翼, 納富一宏, 斎藤恵一: 音声認証における「なりすまし」の検証と対策の検討, 情報処理学会第 73 回全国大会, 3Y-2 (2011).
- [3]吉川昂, 石田秀春, 納富一宏, 斎藤恵一: 自己組織化マップを用いた音声認証における学習条件による比較, 電子通信情報学会 2009 年度 HCG シンポジウム, B8-1 (2009).