L-007

画像と連想語を用いた音声認証システムの開発

A development on voice authentication system with associative words and pictures

河合 博之† 納富 一宏† 斎藤 恵一‡
Hiroyuki Kawai Kazuhiro Notomi Keiichi Saito

1．まえがき

近年，パソコンやスマートフォン等の情報通信端末の普及が著しい．特に最近では，スマートフォンのシェアが爆発的に伸びている．総務省の調査[1]によると，平成 25 年末のスマートフォン保有率は，前年比 13.1 ポイント増と急速に普及が進んでいることがわかる．またこれと同時に，セキュリティの重要度も増してきている．現在主流の本人認証手段として，予め設定したパスワードを入力することで，本人のみが端末を操作可能にする「パスワード認証」があるが，この方法はパスワードの忘却や盗み見による盗難のリスクが存在する．このような問題を解決する手段として，人間の持つ身体的，行動的特徴を認証に用いる「バイオメトリクス認証」が近年注目されている．本研究では中でも，人が話す声の特徴を利用した，音声認証に着目した．音声であればマイクさえあれば導入できるため，コストを抑えることができる．また，音声認証は心理的抵抗が低いといったメリットもある[2]．本研究では話者が発話内容を登録する際，画像を何枚か提示し選択させ，そこから連想されるワードを発話内容とし，認証時にも何枚かの画像の中から登録に使用した画像を選択させるという手法を提案する．これにより認証時において発話内容・話者の声質・画像選択の三重チェックが可能になるため，よりセキュリティが強固なると考えられる．また，登録時に他人と同じ画像を選択した場合でも，そこから連想されるワードは人によって差が生じるため，安全性を損なわず音声認証を使うことができると考えられる．本研究では画像を 10 枚用意し，実験をおこなった．分析には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(SOM:Self-Organizing Maps)を用いて分析し，認証精度が十分であるかについて検証をおこなう．

2．システム概要

提案システムの構成図を図 1 に示す．提案手法の最大の特徴は，登録時および認証時に画像を用いる点にある．登録する音声を決めるときに画像から連想したワードを鍵とすれば，複数の利用者が同じ画像を選択したとしても，連想されるワードは人によって差がある為，この差を特徴点とみなすことができ，安全性の向上に貢献できると考えられる．また，認証時にも登録に用いた画像を選ばせることで，利用者の声質を模倣したなりすましに抑止をかけることが可能となる．これら一連の流れは，利用者にあまり負担をかけずにおこなうことができる為，手軽にセキュリティ度を上げることが期待される．

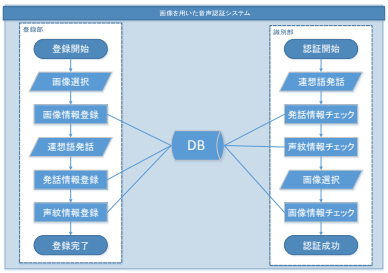

図 1　システム構成

3．実験

3.1　実験方法

20 代の男性 10 名の被験者に協力を仰いだ．被験者はまず予め提示された 10 枚の画像を見てもらい，そこから 1 枚を選択してもらう．実験に使用した画像を図 2 に示す．選択した画像から連想語を一語思い浮かべてもらい，発話内容をマイクで採取する．このとき，発話の長さや品詞は限定せず，自由に発話してもらった．なお，発話は 1 人あたり 10 回とした．

上記で得た音声サンプルから特徴点を抽出する為，MFCC(Mel-Frequency Cepstrum Coefficients)を行う．本研究では 1 つの音声サンプルに対し，分析窓長 25ms，シフト長 10ms，次元数 12 の条件で MFCC をおこない，各次元別に MFCC 値の相加平均を求め，これを特徴点とした．

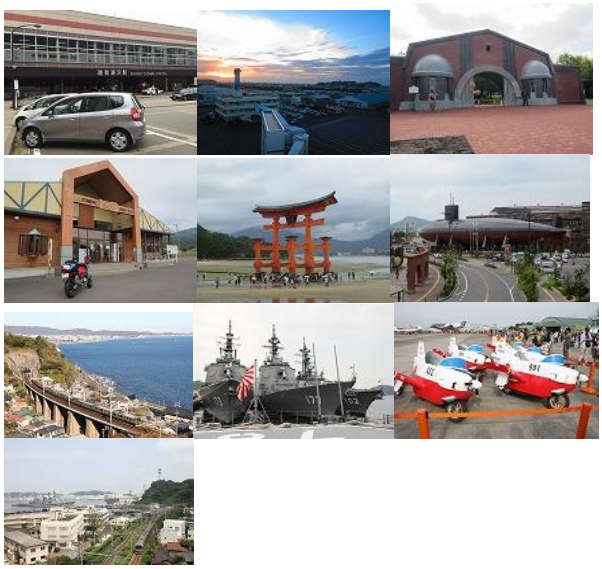

図 2　実験に用いた画像

† 神奈川工科大学 Kanagawa Institute of Technology
‡ 国際医療福祉大学 International University of Health and Welfare

Copyright © 2015 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

3.2 分析方法

MFCC をおこなった音声サンプルを 12 次元の属性ベクトルとし，SOM に投入する．10 回録音した音声サンプルのうち前半 5 個を登録用としてマップを作成し，後半 5 個を識別用として精度計算に用いた．SOM のマップサイズは 30×30～70×70 までの 5 段階とし，マップを作成した．マップ上に表示された登録用ベクトルと認証用ベクトルとの平均ユークリッド距離を求め，設定した閾値より低ければ認証成功とする．閾値の決定には，他人受容率(FAR:False Accept Rate)(1)と本人拒否率(FRR:False Reject Rate)(2)を用いた．なお本研究では，認証時に画像を選択することを考慮し，FAR は使用した画像枚数で割った値とした．また，FAR と FRR の交点を EER(Equal Error Rate)(3)と呼ぶ．FAR と FRR はトレードオフの関係にある為，本研究では EER の点を閾値とし，100%から EER を引いたものを認証精度とした．マップは毎回ランダムに生成される為，それぞれのマップサイズで 5 回実験を繰り返しその平均を認証精度とした．

$$FAR(\%) = \frac{\text{他人受容回数}}{\text{試行回数}} \times \frac{1}{\text{画像枚数}} \times 100 \quad \cdots\cdots (1)$$

$$FRR(\%) = \frac{\text{本人拒否回数}}{\text{試行回数}} \times 100 \quad \cdots\cdots (2)$$

$$EER(\%) = \frac{FAR + FRR}{2} \quad \cdots\cdots (3)$$

4. 結果

4.1 実験結果

それぞれの被験者が，選択した画像番号とそこから発話した連想語を，表 1 に示す．また作成した SOM の一例を図 3 に示す．マップ内に表示されている点が投入した属性ベクトルである．属性ベクトルにはどの被験者かを示すアルファベットがラベリングされている．マップを見ると，被験者別にデータがうまくクラスタリングされていることがわかる．

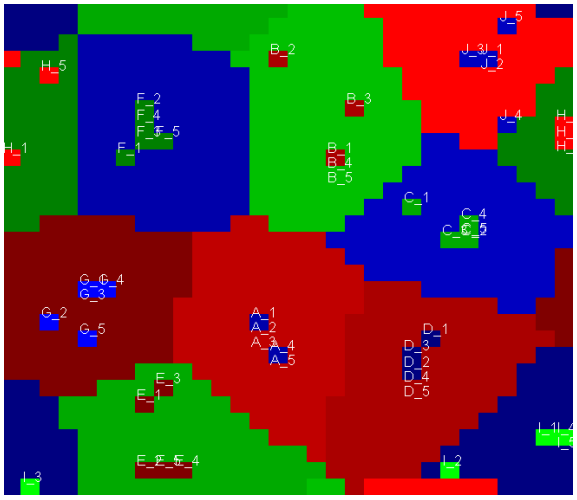

図 3　SOM(30×30)

4.2 分析結果

実験結果より得られたデータをもとに分析した認証精度を表 2 に示す．

表 1　選択した画像番号と連想語

被験者	画像番号	連想語
A	8	船場
B	8	船
C	5	雲
D	1	車
E	3	北海道
F	5	鳥居
G	7	海
H	5	鳥居
I	1	駅
J	1	駅

表 2　認証精度

サイズ	閾値[%]	FAR[%]	FRR[%]	EER[%]	認証精度[%]
30×30	5.60	0.13	0.40	0.27	99.73
40×40	7.60	0.18	0.80	0.49	99.51
50×50	8.70	0.05	1.60	0.82	99.18
60×60	10.50	0.04	0.00	0.02	99.98
70×70	13.30	0.09	0.40	0.24	99.76

5. 考察

表 2 を見ると，画像を用いることでさまざまな連想語を発話していることがわかる．今回の実験では画像の選択にやや偏りがあるが，これは画像の枚数に対して，被験者の人数が少なかったためだと思われる．また，同じ画像を選択しても，そこから異なる連想語を発話する場面が確認できた．よって画像から連想されるワードを鍵とする本研究の手法は有効であると考えられる．

認証精度を見ると，どのマップサイズでも安定して 99%以上の成功率を確認した．よって今回の実験の規模ではマップサイズが認証精度に与える影響は極めて少ないことが考察される．本研究では画像を 10 枚使用して実験を行ったが，認証時に提示される画像の枚数が増えれば増えるほど FAR はその分低下する．しかし枚数が増えすぎると，利用者が画像を探し出すのに時間がかかることが予想されるため，今後は提示する画像をどこまで増やすかの検証も必要になる．

6. おわりに

本研究では画像を用いた音声認証手法についての検討をおこなった．分析の結果，どのマップサイズにおいても 99%以上の認証精度を確認した．今後は本手法の実用化を目指すべく，デモシステムの構築と検証を行う予定である．

参考文献

[1]総務省：平成 26 年度版情報通信白書　インターネット利用状況：
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc253120.html(参照：2015/6/6)
[2] 瀬戸洋一：ユビキタス時代のバイオメトリクスセキュリティ，p.23，日本工業出版社(2003)

Copyright © 2015 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.