5S-7

# 話者照合への影響を考慮した模倣音声の音響分析

坂本　香菜子†　　岩野　公司†
東京都市大学†

## １．はじめに

近年，音声による個人認証（話者照合）に対する期待が高まっており，国際的にも様々な研究や開発が進められている[1]．話者照合性能の評価を行うためには，本人（申告者）を正しく受理する事象と，他人（詐称者）を正しく棄却する事象の双方を考慮する必要がある．しかし，これまでの研究では，後者の事象を評価する際に，詐称者音声として，その話者が申告者（本人）として発声している，いわば詐称の意図を持っていない発声が用いられており，その影響が不明確であるという問題がある．

最も基本的な詐称手段として「模倣（物真似）」が挙げられる．文献[2]では，プロの物真似タレント 1 名による音声の特徴分析が行われており，基本周波数（$F_0$）パターンやフォルマント周波数が目標話者に近づく傾向があることを報告している．しかし，一般の素人が行う模倣音声の特徴や，それが話者照合の性能に与える影響については明らかになっていない．

そこで，複数名の一般の素人を対象に「本人として受理されようとする発声」と「他人を模倣してシステムを攻撃しようとする発声」を収録し，話者照合用音声データベースの構築を行う．本論文では，構築データベースを用い，模倣音声の音響特徴（基本周波数とケプストラム特徴量）の分析結果について報告する．

## ２．模倣音声を含む話者照合用データベースの構築

### 2.1　音声データの収録

物真似の素人として，本学学生の男性 6 名，女性 6 名を被験者とした音声収録を行った．模倣は異性間では行わないものとし，男女 2 グループに分けて収録を行う．収録は 2 日ごと，3 週間にわたり継続的に行った．

各話者は 1 日の収録で，

① 特に意図を持たず，本人の声として自然に行う発声（1 セッション）
② グループ内の他者（5 名）を詐称しようと努力して行った発声（5 セッション）
③ 過去に行った本人自身の発声を聴取した上で，本人として受理されようと努力して行った発声（1 セッション）

の計 7 セッションの発声を行う．各話者はそれぞれのセッションでランダムの 4 桁連続数字を 10 回発声する．初回は①のセッションのみ，2 回目以降は初回に収録された他者または本人の音声を聴いた上で②③の収録を行った．

### 2.2　模倣音声の主観評価

収録データの模倣音声に対し，「模倣の上手さ」を示すスコアを主観評価によって付与した．このスコアリングは，収録した 3 週間分の音声のうちの 4 日間分に対して実施し，収録被験者と異なる 5 名の被験者の主観評価によって行った．各被験者は，模倣対象者の音声と模倣音声の両者の音声を聴いた後，7 段階 (1: 全く似ていない～7: 非常によく似ている) で評価を行い，その 5 名の平均値を各回のスコアとした．

その結果，男子の平均スコアは 1.81（標準偏差 0.35），女子の平均スコアは 3.41（標準偏差 0.39）となった．スコアが低いことから全体的には模倣がうまくいっていないことがわかる．したがって，一般の素人による模倣の難しさが読み取れる．

## ３．模倣音声の音響特徴分析

本研究では，以下の 4 つの発声について特徴量を求め，その「発声間の距離」を用いて模倣音声の分析を行う．

- $\boldsymbol{U}_i$: ①による，発話者 $i$ の自然な発声
- $\boldsymbol{U}_j$: ①による，模倣対象者 $j$ の自然な発声
- $\boldsymbol{U}_{i,j}$: ②による，発話者 $i$ が対象者 $j$ を詐称しようと努力して行った模倣発声
- $\boldsymbol{U}_{i,i}$: ③による，発話者 $i$ が本人として受理されようと努力して行った発声

図 1 に発声空間の模式図を示し，今回分析

Acoustic analysis of imitated voices considering effects on speaker verification
†Kanako Sakamoto, Koji Iwano, Tokyo City University

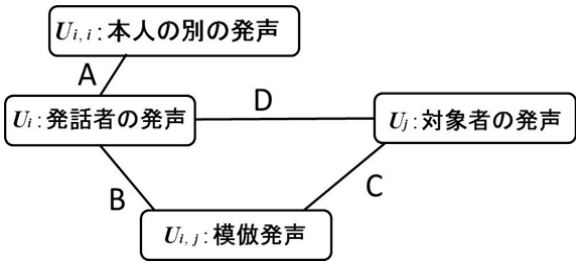


図 1　分析対象とする発声間距離

対象とする発声間距離（A～D）を表す．なお，分析対象データにはスコアリングを行った 4 日間分の音声を用いた．

### 3.1 基本周波数に基づく分析

発声 $U_a$, $U_b$ 間の距離 $D_f(U_a, U_b)$を，それぞれの発声の平均 $F_0$ 値 $F0_a$, $F0_b$ により式(1)で定義する．

$$D_f(U_a, U_b)=\left|\log(F0_a)-\log(F0_b)\right| \quad (1)$$

全話者を対象とした距離の平均値を表 1 に示す．模倣による移動距離（B）は，話者間距離（D）と同程度で，本人内の変動距離（A）よりも有意に大きい．一方，模倣発声と対象者発声の距離（C）も大きいことから，「模倣によって声の高さは変化しているものの，対象者には必ずしも近づいていない」傾向が読み取れる．

### 3.2　ケプストラムに基づく分析

話者認識で一般的な 12 次元 MFCC，その 1 次微分成分，対数パワーの 1 次微分成分の計 25 次元の特徴ベクトルを分析に使用する．各発声を 3 状態の HMM（混合数 1）でモデル化し，2 状態目のモデルパラメータ（25 次元の平均・分散ベクトル）をその発声の音響特徴とする．発声 $U_a$, $U_b$ 間の距離 $D_c(U_a, U_b)$ を式(2)のようにマハラノビス距離を用いて定義する[2]．$K$ はベクトルの次元数（25）であり，$\mu_{ak}$ と $\sigma_{ak}^2$ は，発声 $U_a$ の平均，分散ベクトルの $k$ 次元目の要素である．

$$D_c(U_a,U_b)=\sqrt{\frac{K\sum_{k=1}^{K}(\mu_{ak}-\mu_{bk})^2}{\sum_{k=1}^{K}\sigma_{ak}^2+\sum_{k=1}^{K}\sigma_{bk}^2}} \quad (2)$$

各距離の平均値を表 2 に示す．模倣による移動距離（B）は，本人内の変動距離（A）よりも大きいが，話者間の距離（D）に比べるとその値は小さく，また，模倣発声と対象者発声の間の距離（C）も大きいことから，「模倣の努力によって本人とは異なる声色を出そうとしているが，対象者には近づかない」傾向が読み取れる．

表 3 に，話者ごとの「模倣による移動離（B）」を「本人内の変動距離（A）」と「話者間距離（D）」の比で示す．話者 M01, M04, F01, F02 は，本人内の変動に比べ模倣による特徴変動が大きく，その距離は話者間距離の 7 割程度に達していることから，このような話者が照合性能に影響を及ぼす可能性が考えられる．

表 1　発声間の平均対数 $F_0$ 値の距離
（括弧内は標準偏差）

| A | B | C | D |
|---|---|---|---|
| 0.70E-02 (0.50E-02) | 2.10E-02 (1.54E-02) | 1.97E-02 (1.50E-02) | 2.97E-02 (2.33E-02) |

表 2　発声間のケプストラム特徴量の距離
（括弧内は標準偏差）

| A | B | C | D |
|---|---|---|---|
| 0.305 (0.374) | 0.385 (0.487) | 0.631 (0.598) | 0.527 (0.271) |

表 3　話者ごとの模倣による移動距離の様子

| 男性 | | | 女性 | | |
|---|---|---|---|---|---|
| ID | B/A | B/D | ID | B/A | B/D |
| M01 | 4.93 | 0.60 | F01 | 5.12 | 0.74 |
| M02 | 1.01 | 0.22 | F02 | 2.46 | 0.73 |
| M03 | 1.65 | 0.35 | F03 | 1.15 | 0.35 |
| M04 | 2.51 | 0.78 | F04 | 1.47 | 0.58 |
| M05 | 1.63 | 2.35 | F05 | 0.77 | 0.42 |
| M06 | 0.20 | 0.39 | F06 | 0.98 | 0.32 |

## 4．まとめ

本研究では，模倣発声を含む音声データベースの構築を行い，模倣音声の基本周波数，ケプストラム特徴量の分析を行った．その結果，発話者は模倣の努力は行っているが，対象者には近づかない傾向が読み取れた．ただし，模倣による特徴変動が大きい話者も存在し，話者照合性能に影響を与える可能性も示唆された．今後は，発話を単位としたより詳細な分析や，実際の話者照合性能との関連を調査する必要がある．

**謝辞**　本研究は JSPS 科研費 基盤研究（C）25330206 の助成を受けたものです．

## 参考文献

[1] 越仲他，“話者認識の国際動向，” 日本音響学会誌，vol. 69, no. 7, pp. 342-348, 2013.
[2] 北村，“物真似タレントによる物真似音声の分析，” 電子情報通信学会技術研究報告, vol. 107, no. 282, pp. 49-54, 2007.
[3] 中村他，“話し言葉音声の音響的・言語的特徴の分析，” 電子情報通信学会技術研究報告, vol. 106, no. 78, pp. 19-24, 2006.