B2

実際の発話者と知覚される個人性を切り分けられる音声の探索的検討

林　大輔（東京大学大学院人文社会系研究科）・山上　精次（専修大学人間科学部）
daisuke.semmu@gmail.com, yamagami@psy.senshu-u.ac.jp

1. はじめに

ヒトは音声を聞いた時に，何を話しているのかだけではなく，発話者が「どのような人であるのか」を知ることができる（e.g. Krauss, Freyberg & Moresella, 2002）．そのような発話者の個人性に関する情報は，大きく 2 つに分けて考えることができる．1 つは，年齢や性別などの「どのような発話者であるのか」という個人性一般に関する情報であり，もう 1 つは，「発話者が誰であるのか」という個人の同定に関わる情報である．これらの情報の処理は，相互に関わり合っていると考えるのが自然である．個人性に依存して発話者の声道などは生理学的に変化し，またその結果，音声の音響特徴も変化する．これらの変化は，個人性一般に関する情報でも，個人の同定に関する情報でも同様であり，両者は必ずしも明確に区別できるものではない．

しかし，神経心理学研究において，個人性一般に関する情報はある程度以上知覚でき，知らない発話者の音声を弁別することはできるにも関わらず，よく知っているはずの発話者の音声を聞いても同定ができない音声失認という症例が報告されている（e.g. Hailstone, Crutch, Vestergaard, Patterson & Warren, 2010; Roswandowitz et al., 2014）．これらの研究は，個人性一般に関する情報の知覚と，発話者の同定が異なる処理メカニズムによるものである可能性を示唆している．

そのため，音声から発話者の個人性に関する情報を知覚する過程を明らかにするには，これらの処理の独立性や階層性を調べることが不可欠である．しかしこれまでの研究では，「実際の発話者が誰であるのか」ということと，その人が発した音声から「どのような個人性が知覚されるのか」という関係が，ある程度 1 対 1 に対応するものとして扱われてきた．これ自体は，日常を考えれば自然なことではある．しかしそのような条件では，「どのような個人性であるのか」ということと「誰であるのか」ということが密接に繋がり合っているため，個人性一般に関する情報の知覚処理と，発話者の同定に関わる知覚処理をそれぞれ区別した上で検討するには不十分である．

そこで本研究では，同一の発話者であっても異なった個人性を持って知覚されるよう意図して発話された演技音声を用いて，「実際の発話者」と「知覚される個人性」を切り分けられる音声を見出すことを目的として，聴取実験を行った．そのような音声が得られれば，今後の研究において，音声から発話者の個人性に関する情報を知覚するメカニズムを明らかにする上で有用であると考えられる．実験では，地声で朗読された音声であれば，読まれている文章が異なっても似た個人性を持っていると評定されることを確かめた上で，同一の発話者であっても異なった個人性を持って知覚される音声が存在することを示す．

2. 方法

本研究では，2 つの実験を行った．基本的な方法は 2 つの実験で同一であったが，実験 1 では女性話者の音声のみを用いたのに対し，実験 2 では男性話者の音声も合わせて用いた点が大きく異なっていた．

2.1. 参加者

実験 1 には，26 名の大学生が参加した（男性 10 名，女性 16 名）．平均年齢は 19.58 歳で，標準偏差は 0.90 歳であった．また実験 2 には，実験 1 とは異なる大学生 26 名が参加した（男性 8 名，女性 18 名）．平均年齢は 21.08 歳で，標準偏差は 6.26 歳であった．全ての参加者が実験目的を知らなかった．また，全ての実験は専修大学人間科学部心理学科の基礎実験 2 の授業内に行い，参加者は事前に倫理に関する説明を受け，実験参加の同意書に署名を行った．

2.2. 装置

音声の加工と再生，および反応の取得に MacBookAir Early 2014 (Apple) を用いた．ヘッドフォン (Sony) を用いて音声を呈示し，液晶ディスプレイ (HP 2311) を用いて画面を呈示した．音声の加工は，Audacity (R) 2.1.0 と GNU Octave 4.0.3 を用いて行い，実験プログラムは jsPsych を用いて作成した（de Leeuw, 2015）．

2.3. 音声

7 人の話者について，3 種類ずつ（A，B，C）の音声を用いた（表 1）．

話者 1, 2, および 7 の音声は，話速バリエーション型音声データベース（SRV-DB）内の，ATR25 文をプロが朗読した音声であった．話速は 8 モーラ/秒のものを用いた．話者 1 と 2 は女性であり，話者 7 は男性であった．A，B，C それぞれの音声は，別の文章を朗読したものであり，文章は話者間で共通であった．全て地声で朗読された音声であり，同じ発話者であれば似たような個人性を持って知覚されると考えられる音声であった．

話者 3，4，5，および 6 の音声は，同一の発話者であっても異なる個人性を持って知覚されるよう意図して発話された演技音声であった．具体的には，声優事務所である株式会社マウスプロモーションのボイスサンプルを，研究使用許諾を得て，実験者が適切と感じる音声を抽出して用いた．話者 3，4，5 は女性であり，実験者の聴取印象では，各話者の A の音声は幼い少女，B の音声は大人の女性，C の音声は少年の発話であるように知覚される音声であった．話者 6 は男性であり，実験者の聴取印象では，A および B の音声は青年，C の音声は少年の発話であるように知覚される音声であった．

このうち，実験 1 では話者 1〜5 の音声を，実験 2 では話者 4〜7 の音声を用いた．なおこれ以降，それぞれの音声は，たとえば話者 1 の音声を「音声 1A・音声 1B・音声 1C」といった形で表記する．

表1:各音声の情報の要約

話者	性別	データベース	実験者の聴取印象
1	女性	SRV-DB	全て地声による朗読
2			
3		マウスプロモーションのボイスサンプル	A(幼い少女) B(大人の女性) C(少年)
4			
5			
6	男性		A・B(青年)、C(少年)
7		SRV-DB	全て地声による朗読

表2:用いた表現語対

低い声	—	高い声
老けた感じの声	—	若い感じの声
女性的な声	—	男性的な声
張りのない声	—	張りのある声
弱々しい声	—	迫力のある声
細い声	—	太い声
澄んだ声	—	かすれた声
落ち着きのない声	—	落ち着きのある声

2.4. 音声の加工

全ての音声について，200ms 以上の無音部分は，200ms に短縮した．その上で，発話内容の影響を軽減するため，ランダムスプライシングを用いた（Scherer, 1971）．ランダムスプライシングは，音声の発話内容をマスクしながら，声質を保持する上で有用な音声の加工法だと考えられている（Teshigawara, 2004）．具体的には，音声を 250ms ずつのセグメントに分割し，各セグメントの立ち上がりと立ち下がりの 3ms ずつは，振幅を線形に変化させた．そのように加工した 250ms のセグメントを 20 個，もともと並んでいたセグメント同士が続くことがないように，ランダム順に並べて，5 秒間の音声を作成した．1 秒間の無音区間を挟んで，異なる順番で並べた 5 秒間の音声を繋げて，1 つの音声につき 11 秒間の加工音声を作成した．

2.5. 評定項目

木戸・粕谷（1999）に基づいて，声質を表現する日常表現語対を 8 つ用いた（表 2）．表現語対は「低い声–高い声」「老けた感じの声–若い感じの声」「女性的な声–男性的な声」「張りのない声–張りのある声」「弱々しい声–迫力のある声」「細い声–太い声」「澄んだ声–かすれた声」「落ち着きのない声–落ち着きのある声」であった．それぞれの表現語対について，「非常に・かなり・やや・普通・やや・かなり・非常に」の 7 件法を用いて，どちらの表現語対により近く知覚されるかの判断を求めた．評定値は，上記の内，左側の表現語が選択された時に小さく，右側の表現語が選択された時に大きくなるように 1〜7 で定量化した．

2.6. 手続き

実験は，1 人ずつ個室で行った．まず実験者が内容について教示を行い，参加者が課題を理解したら実験者は部屋の外に出て，参加者は 1 人で課題を行った．実験では，2.4 項のように加工した音声を 1 つずつ呈示し，それぞれの音声について，2.5 項で記述した 8 つの表現語対について評定を行った．1 つの画面に 8 つの表現語対を全て呈示して，参加者はそれぞれの表現語対について，マウスクリックを用いて 7 件法で評定を行った．8 つ全ての表現語対について評定を行った後，次の音声を呈示し，再び評定を行った．1 つの音声につき，1 人 1 回ずつ評定を行った．なお，音声の呈示順は参加者ごとにランダムであり，表現語対の呈示順や左右位置は音声ごとにランダムであった．

3. 結果

結果の解析の際に，ある音声に対するある参加者の評定値の一部に欠損値があった場合には，その参加者のその音声に対する評定値は全て削除した．欠損値を除くと，実験 1 は 1 つの音声当たり少なくとも 22 人分，実験 2 は 23 人分の評定値が得られた．各音声に対する各表現語対の評定値について，参加者間の平均をとり，得られた平均値を用いてまず，表現語を対象として平方ユークリッド距離についてウォード法を用いて階層的クラスター分析を行い，デンドログラムを作成した（図 1)．その結果，実験 1 と実験 2 で，類似した構造のデンドログラムが得られた．

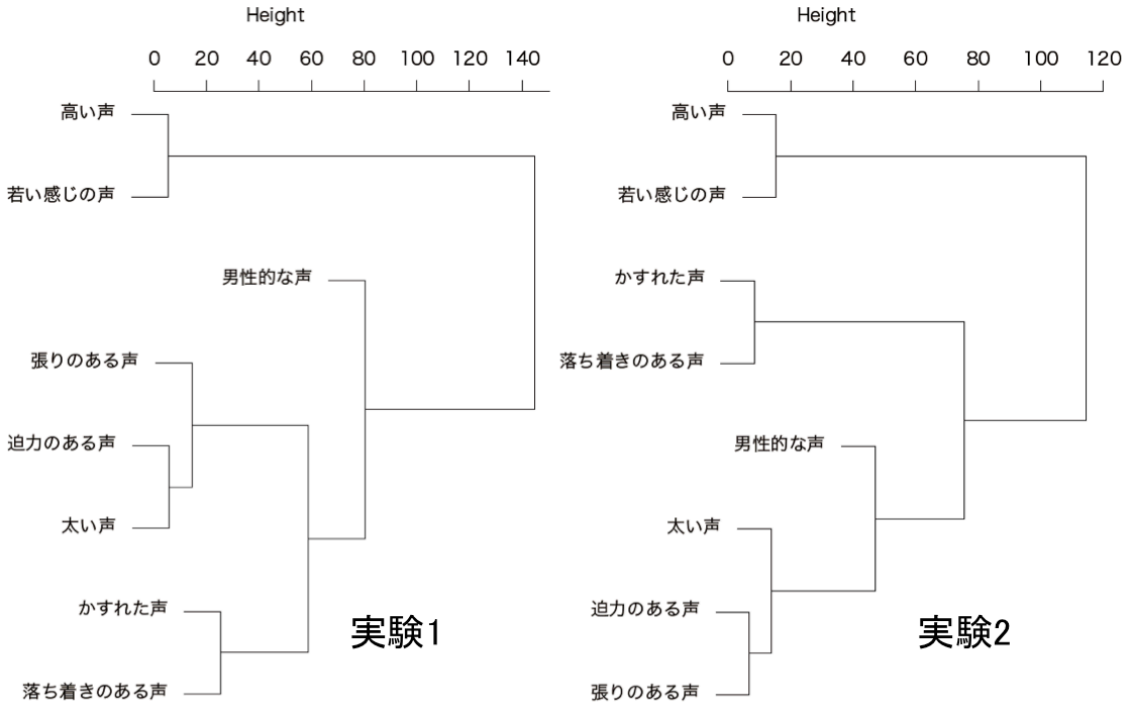

図 1: 表現語を対象としたクラスター分析の結果．（左）実験 1．（右）実験 2．

続けて，同様のデータを用いて，音声を対象としてクラスター分析を行い，デンドログラムを作成した（図 2 左，図 3 左）．得られたデンドログラムに基づいて，実験 1 では 3 つ，実験 2 では 4 つのクラスターに音声を分類した．実験 1 のクラスターはそれぞれ，実験者の聴取印象に基づくと，クラスター1 は少年，クラスター2 は幼い少女，クラスター3 は大人の女性の発話であるように知覚される音声が含まれていた．実験 2 のクラスターは，クラスター1 は幼い少女，クラスター2 は大人の女性，クラスター3 は少年，クラスター4 は青年の発話であるように知覚される音声が含まれていた．分類に基づいて，それぞれのクラスターに含まれる音声の，各表現語対に対する評定値を平均し，折れ線グラフで表した（図 2 右，図 3 右）．なお，折れ線グラフの背景は，図 1 に示したクラスター分析に基づいて，表現語対を 4 つのグループに分け，それぞれの区別がつくように灰色の濃さを変えた．

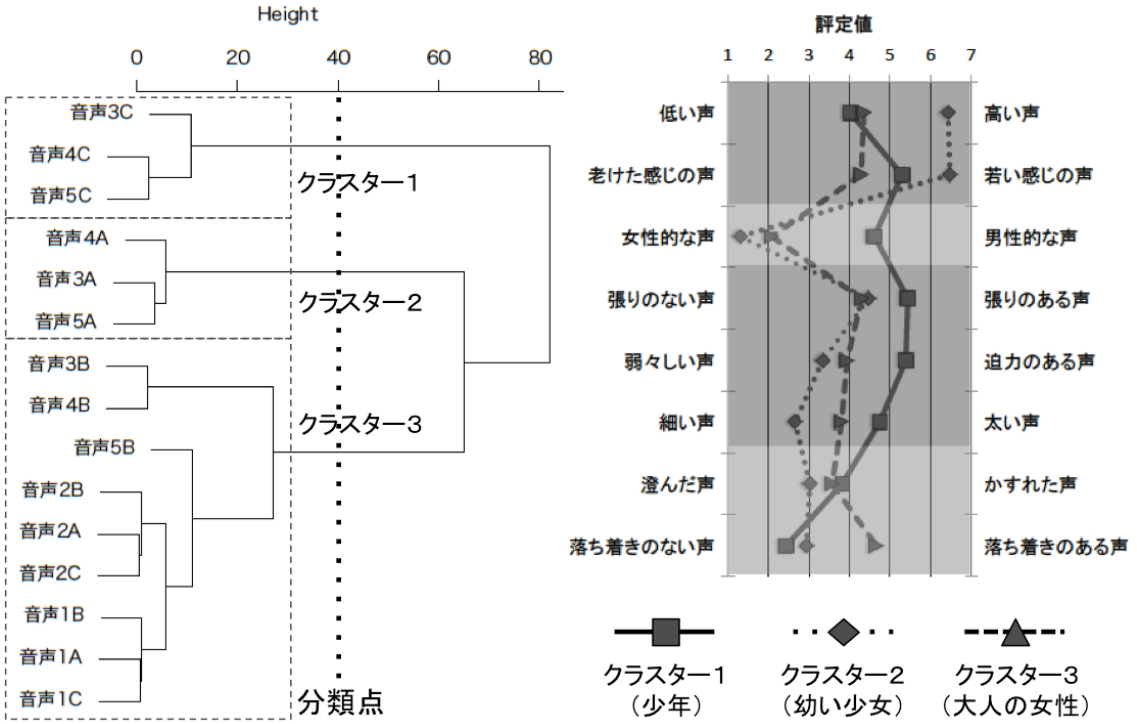

図 2: 実験１の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

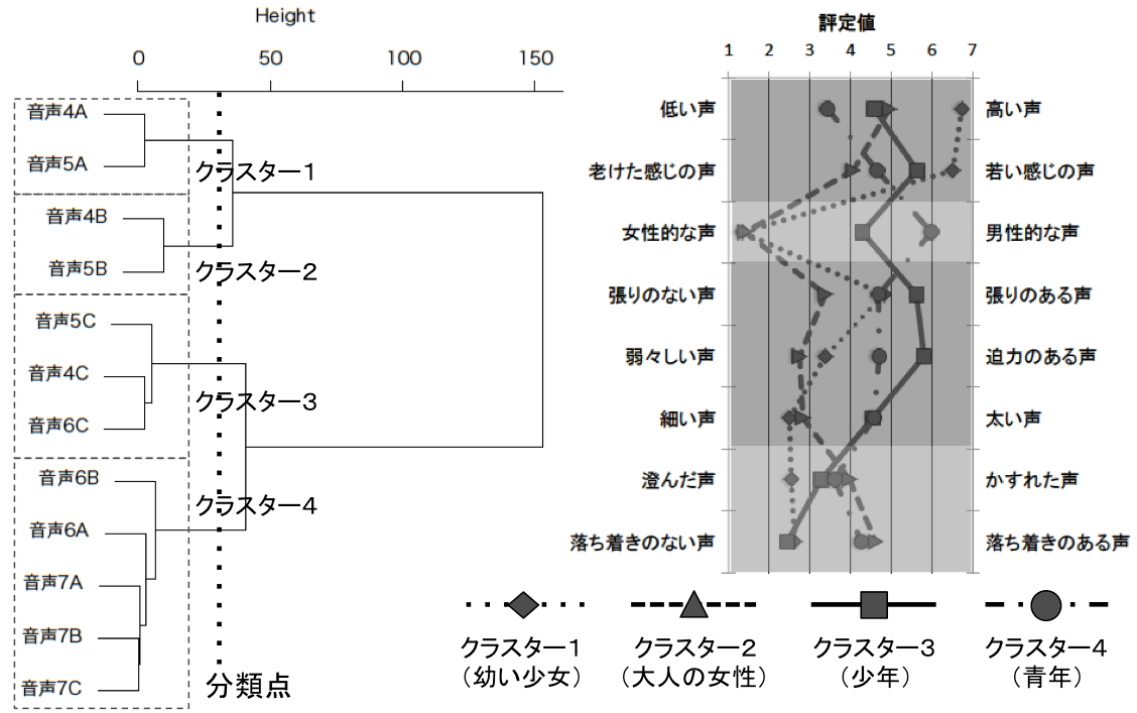

図 3: 実験２の音声の分類．（左）クラスター分析により得られたデンドログラム．（右）クラスターごとの評定値の平均．

4. おわりに

本研究の結果から，地声で朗読された音声は，同一の発話者であれば，文章が異なっても似た個人性を持って知覚されることが示された（図 2 左の音声 1A〜2C，図 3 左の音声 7A〜7C）．また，実験 1 では話者 3〜5 の 3 種類の音声がそれぞれ異なるクラスターに分類されたこと（図 2 左），実験 2 では話者 4〜6 の音声が異なるクラスターに分類されたことから（図 3 左），同一の発話者であっても異なった個人性を持って知覚される音声，すなわち「実際の発話者」と「知覚される個人性」を切り分けられる音声が見出されたと言える．

個人性の違いは声道などの生理学的特徴，ひいては音声の音響特徴に影響することから，音声の音響特徴と個人性の知覚との関係を調べることは，何が個人性の知覚に重要なのかを知る上で有用である（e.g. Kitamura & Akagi, 1995）．今回用いた音声の分類においては「性別」と「（声の高さと関連した）年齢」が強い影響を及ぼしていたことが見てとれるが（図 1 および図 2 右，図 3 右），音声の音響特徴について分析を行って知覚との関連を見ることで，何が個人性の知覚に重要なのかについて知見が得られると考えられる．

今回見出された音声を利用して，今後の研究において，たとえばある音声と発話者の結びつきを学習した時，知覚される個人性が異なっても同一の発話者であることが認識できるのか，などを調べていくことで，「実際の発話者」と「知覚される個人性」を切り分けた上で，音声から発話者に関する情報を知覚する過程を明らかにできるであろう．

参考文献

Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. Neuropsychologia, 48, 1104-1114.

木戸 博・粕谷 英樹．(1999)．通常発話の声質に関連した日常表現語の抽出．日本音響学会誌，55，405-411.

Kitamura, T., & Akagi, M. (1995). Speaker individualities in speech spectral envelopes. Journal of the Acoustical Society of Japan (E), 16(5), 283-289.

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. Journal of Experimental Social Psychology, 38, 618-625.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47(1), 1-12.

Roswandowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. Current Biology, 24, 2348-2353.

Scherer, K. R. (1971). Randomized splicing: A note on a simple technique for masking speech content. Journal of Experimental Research in Personality, 5(2), 155-159..

Teshigawara, M. (2004). Random splicing: A method of investigating the effects of voice quality on impression formation. Proceedings of Speech Prosody 2004, Nara, 209-212.