

声優のキャラクター演技音声を用いた音声知覚に関する実験研究

Experimental Research on Voice Perception with Character Voices by Professional Voice Actors

林 大 輔*

Daisuke HAYASHI

要 旨

本研究では、声優が様々な年齢・性別のキャラクターを演じたキャラクター演技音声を用いて、2つの実験を行った。まず実験1では、声質を表現する表現語対を用いて印象評定を行った。その結果、声優の演技音声は、声優の実年齢や性別ではなく、少年の声などの声優が意図したキャラクターに基づいて特徴付けられることが示された。続けて実験2では、発話者を弁別する実験を行い、異なるキャラクターとして知覚されていても、ある程度は発話者の違いを弁別できることを示唆する結果が得られた。これらの結果は、声優のキャラクター演技音声は、ヒトの音声知覚に関する処理過程を研究する上である種の「錯聴」としての有用性を持つことを示している。

キーワード：音声、声優、演技、印象評定、話者弁別

1. はじめに

ヒトとヒトとのコミュニケーションにおいて、音声は重要な役割を果たしている。ヒトは音声を聞いた時に、様々な情報を得ることができる。音声には大きく分けて、3種類の情報が含まれていると考えられており、それぞれ言語情報、パラ言語情報、非言語情報と呼ばれる（Fujisaki, 1996；森・前川・粕谷, 2014）。言語情報とは、書き言葉で表せるような言語の情報である。パラ言語情報とは、態度や意図といった随意的に発話者が込める言語ではない情報である。非言語情報とは、発話者の個人性や心身の状態といった、非随意的に音声に含まれる情報である。本研究では、非言語情報に主に含まれるような、どのような誰が話しているのかという発話者自身に関する情報に焦点を当てて実験を行う。

音声に含まれる発話者自身に関する情報は、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に関する一般的な情報であり、ここでは話者情報と呼ぶ。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ような、より個々の発話者に特有の情報であり、ここでは個人性情報と呼ぶ。これら2つの情報には、類似点が存在している。個々人で特性が違えば、発話者の声道なども生理学的に異なり、それに伴って音声の音響特徴も異なる。そしてそのような違いは、話者情報であっても個人性情報であっても同様である。たとえば話者情報の1つである年齢を例に挙げると、幼児期から青年期にかけての成長に伴って、生理学的な特徴である発話者の声帯膜様部や声道は長くなっていき（e.g., 栗田,

* 愛知淑徳大学人間情報学部

1988; Vorperian et al., 2009), 音響特徴である音声の基本周波数や第1, 第2フォルマント周波数は低くなっていく (e.g., 粕谷・鈴木・城戸, 1968; 森他, 2014)。個人性情報により近いと考えられる特徴についても、たとえば下咽頭腔の形状に着目した研究では、個々人の形状をMRIを用いて調べた上で、モデル化してシミュレーションを行った結果、形状の個人差が音声スペクトルの比較的高い周波数帯域における違いとなって表れることが示されている (Kitamura, Honda, & Takemoto, 2005)。このように、発話者によって生理学的な特徴が異なり、それに伴って音声の音響特徴が異なることは、話者情報でも個人性情報でも似通っている。

しかし一方で、これらの情報に関する知覚的な処理過程が、ある程度独立している可能性を示唆する神経心理学研究の知見も存在する。脳損傷患者を対象とした研究で、知らない発話者の音声と弁別する課題と、知っている発話者の音声を同定する課題の間で、二重乖離が示されている (Van Lancker & Kreiman, 1987)。すなわち、話者情報に基づいた音声の弁別は可能であるが個人性情報に基づいた個人の同定ができない患者と、逆に話者情報に基づいた音声の弁別はできないが個人性情報に基づいた個人の同定は可能な患者がそれぞれ存在することが報告されている。このうち前者のような症例を音声失認と呼ぶ。他にも、よく知っているはずの音声徐徐に認識できなくなる進行性の音声失認や (Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010)、顕著な脳損傷は認めないものの、よく知っているはずの発話者のことを音声に基づいて同定する課題が行えない発達性音声失認も、数は少ないが報告されている (Garrido et al., 2009; Roswadowitz et al., 2014)。これらの症例では、顔写真と個人の結びつけや音声に基づいた感情の知覚、知らない発話者の音声の弁別などの課題はできるにも関わらず、知っている発話者の音声と個人を結びつけることができない。そのため、音声失認の存在は、音声の個人性情報に基づいて個人を同定することが、ある程度他の情報を処理する過程とは独立している可能性を示唆しているといえる。

これらの知見を考慮すると、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するためには、話者情報と個人性情報を明確に切り分けた上で、それぞれの情報を処理する過程の独立性や階層性を調べることが不可欠である。しかし一般的な発話音声においては、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話者情報が知覚されるのかということは、ある程度1対1に対応している。そのため、話者情報と個人性情報は共変するのが通常であり、これらの情報は明確に区別されることなく研究が行われてきた。それに対して本研究では、声優のキャラクター演技音声に着目し、話者情報と個人性情報の切り分けを試みる。

声優はその名の通り「声で演じる俳優」である。自らの身体などの視覚情報を提示することなく、声のみを用いてキャラクターを演じる。そのため声優は、年齢や性別、体格などに関して身体的な制約が少なく、たとえば女性声優であれば、幼児から中高生、大人、老年まで様々な年代の女性を演じるとともに、少年の声や、場合によっては青年の声を演じることもある。すなわち声優の演技音声は、同一の発話者であっても、話者情報に相当する年齢や性別などの特性が様々に異なって知覚される音声だといえる。一方で、個人性情報というのは定義上、同一の発話者であれば同一である。よって、声優のキャラクター演技音声は、個人性情報と話者情報を切り分けて実験を行う上で、非常に適した音声であると考えられる。そのような音声を用いることで、話者情報に関する知覚処理と個人性情報に関する知覚処理をそれぞれ区別した上で検討することが可能となり、音声から発話者自身に関する情報を知覚するメカニズムを新たなアプローチを用いて明らかにすることができる。

以上を踏まえて本研究では、異なる特性を持ったキャラクターの音声として知覚されるよう声優が意図して発話した演技音声を用いて実験を行う。まず実験1では、同一の発話者であっても話者情報が実際に違って知覚されているのかについて、印象評定実験を通じて知覚的分類を行い、明らかにする。次に実験2では、そのように話者情報が違っていても、個人性情報に基づいて話者弁別が行えるのかについて調べる。そして結果を踏まえて、声優のキャラクター演技音声は、音声の知覚研究にどのように貢献しうるのかについて議論を行う。

2. 実験 1：印象評定

実験 1 では 2 つの印象評定実験を行い、音声の知覚的分類を試みた。基本的な方法は 2 つの実験で同一であったが、実験 1.1 では女性話者の音声のみを用いたのに対し、実験 1.2 では男性話者の音声も合わせて用いた。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話者情報を持っていると評定されることを確かめた上で、声優のキャラクター演技音声が同一の発話者であっても異なった話者情報を持って知覚されるのかを調べることを目的とした。

2.1. 方法

2.1.1. 参加者

実験 1.1 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。実験 1.2 には、実験 1.1 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。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。また、全ての実験はヘルシンキ宣言に従って行われた。

2.1.2. 装置

音声の加工と再生、および反応の取得に Apple MacBookAir Early 2014 を用いた。Sony 社のヘッドフォン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 5.0.3 を用いて作成した (de Leeuw, 2015)。

2.1.3. 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話者情報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話者情報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、各発

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		株式会社 マウスプロモーションの ボイスサンプル	A (若い少女)
4			B (大人の女性)
5			C (少年)
6	男性	SRV-DB	A・B (青年), C (少年)
7			全て地声による朗読

話者の A の音声は若い少女、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 1.1 では話者 1～5 の音声を、実験 1.2 では話者 4～7 の音声をを用いた。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。

2.1.4. 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、発話内容が評定に与える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（van Bezooijen & Boves, 1986; Scherer, Feldstein, Bond, & Rosenthal, 1985; Teshigawara, 2004a）。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しち下がりの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。

2.1.5. 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、上記のうち、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。

2.1.6. 手続き

実験は、1人ずつ静かな個室で行った。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、2.1.4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、2.1.5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。参加者が8つ全ての表現語対について評定を行った後、画面上の次の画面に進むためのボタンを押すと、次の音声呈示されるという流れであった。1つの音声につき、1人1回ずつ評定を行った。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。なお教示の際には、判断は試行の間で首尾一貫している必要はないので、1つ1つの音声について、その時に感じた印象に基づいて回答するよう強調した。

2.2. 結果

結果の解析は統計ソフトR 3.3.3を用いて行った（R Core Team, 2017）。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1.1は1つの音声当たり少なくとも22人分、実験1.2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、音声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）。得られたデンドログラムに基づいて、実験1.1では3つ、実験1.2では4つのクラスターに音声を分類した。実験1.1のクラスターはそれぞれ、実験者の聴取印象に基づく、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1.2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。

続けて、表現語対の評定値について主成分分析を行った。それぞれの実験における第1主成分と第2主成分を軸とし、各音声の主成分得点を表したパイプロットを作成した（図2）。累積寄与率は、実験1.1では、第1

主成分で 54.0%，第 2 主成分までで 87.0%であり，実験 1.2 では，第 1 主成分で 65.4%，第 2 主成分までで 90.8%であった。

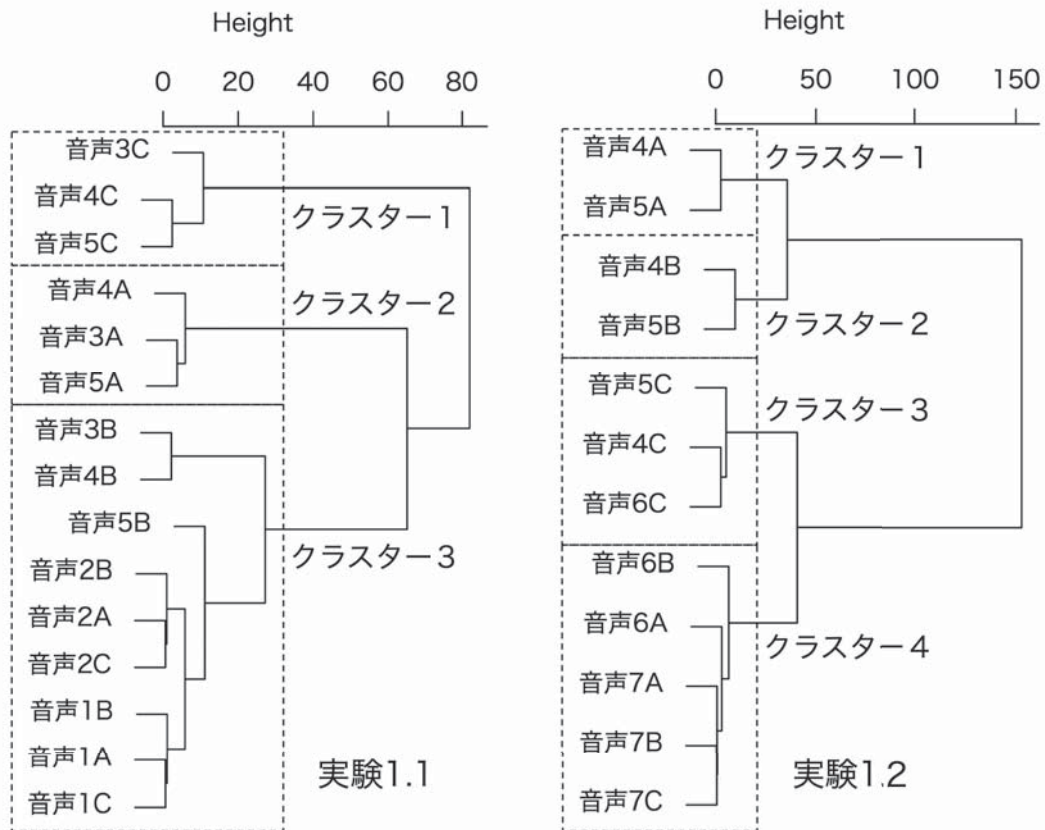


図1 クラスタ分析の結果のデンドログラム。左は実験 1.1，右は実験 1.2 の結果を示す。

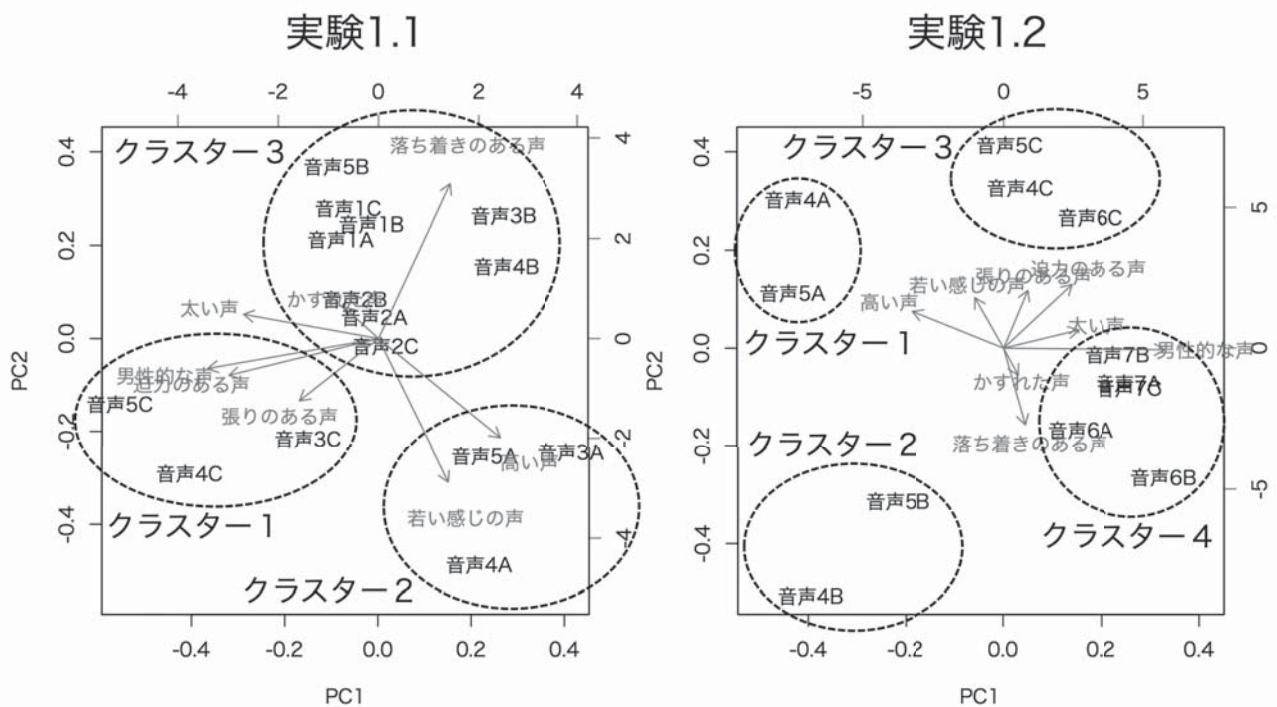


図2 主成分分析の結果のバイプロット。左は実験 1.1，右は実験 1.2 の結果を示す。

2.3. 考察

まず、クラスター分析の結果(図1)を見ると、実験1.1の話者1ならびに2の音声、実験1.2の話者7の音声がいずれも同じクラスターに入ったこと、また音声1A～1C、音声2A～2C、音声7A～7Cはデンドログラムにおいてそれぞれ近くに位置していることから、地声で朗読された音声については、男性話者でも女性話者でも、発話している文章が異なっても、同一の発話者であれば似た話者情報を持って知覚されることが示された。この結果は、話者情報と個人性情報が一般的にはあまり区別されないことと整合的な結果だと考えられる。

一方で、実験1.1では話者3～5の3種類の音声それぞれ異なるクラスターに分類されたこと、また実験1.2では話者4～6の音声異なるクラスターに分類されたことから、声優のキャラクター演技音声は、同一の発話者であっても異なった話者情報を持って知覚される音声、すなわち個人性情報と話者情報を切り分けられる音声だといえる。さらに、実験1.1および1.2の各クラスターは複数の発話者の音声を含んでおり、異なる発話者であっても似た話者情報を持って知覚される音声も見出された。これらの音声は、それぞれ似たキャラクター特性を持った音声だといえる。

続けて、主成分分析の結果(図2)について考える。まず、実験1.1の第1主軸については、右側に幼い少女や大人の女性、左側に少年として知覚される音声が含まれているため、第1主成分得点は性別を表すと解釈できる。実験1.1の第2主軸については、上側に大人の女性、下側に少年や幼い少女として知覚される音声が含まれているため、第2主成分得点は年齢を表すと解釈できる。同様に、実験1.2の第1主軸については、右側に青年や少年、左側に幼い少女や大人の女性として知覚される音声が含まれているため、正負は逆であるが、実験1.1と同様に第1主成分得点は性別を表すと解釈できる。実験1.2の第2主軸については、上側に幼い少女や少年、下側に大人の女性や青年として知覚される音声が含まれているため、やはり正負は逆であるが、実験1.1と同様に第2主成分得点は年齢を表すと解釈できる。そして、実験1.1のクラスター1(少年)は「男性的な声」「迫力のある声」「張りのある声」といった表現語、クラスター2(幼い少女)は「高い声」「若い感じの声」といった表現語、クラスター3(大人の女性)は「落ち着いたきのある声」という表現語と関連が高いように見受けられる。実験1.2については、実験1.1では用いなかった男性の音声に着目すると、少年の声(クラスター3)は「迫力のある声」「張りのある声」として評定される一方で、青年の声(クラスター4)は「落ち着いたきのある声」「男性的な声」として評定されている。これらを総合して考えると、年齢に関しては「落ち着いたきのある声」は大人の音声を、「高い声」や「若い感じの声」は子供の音声を表現する表現語であるように思われる。性別に関しては「男性的な声」と評定されるか否かである程度説明がつくと考えられるが、少年の声については「迫力のある声」「張りのある声」といった表現語で表される特徴も重要な役割を果たしていることが示唆される。

これらの結果は、声優がどのように音声の印象を操り、それぞれの年齢や性別を表現しているのか、そしてヒトがどのような声質の音声であれば意図した年齢や性別として知覚するのかを考える上で、示唆的である。1つの興味深い点は、少年の発話が、声の迫力や張りで表現されていると考えられることである。児童から思春期の子供の性別を判断する上では、基本周波数や第1～3フォルマント周波数が重要な手がかりになると考えられており(Perry, Ohde, & Ashmead, 2001; Skuk & Schweinberger, 2014)、「高い声」であるかどうかは1つの重要な手がかりになることが予想できる。しかし、10歳前後ではそれらの手がかりにまだ男女で大きな差はない(粕谷他, 1968)。そのため、男女の違いを声の迫力や張りで表現している可能性が考えられる。また、大人らしさを表す上で「落ち着いたきのある声」が利用されていることも示唆されている。年齢の違いを判断する上でも、基本周波数やフォルマント周波数は1つの重要な手がかりとなりうる(森他, 2014)。しかしそれ以外にも高齢化に伴って話速が低下するという報告があり(Smith, Wasowicz, & Preston, 1987)、話速は落ち着いた声かどうかの印象に影響を与えると考えられている(木戸・箕輪・粕谷, 2002)。また、発話の韻律的特徴も年齢の判断に影響を与えると考えられており(Maekawa, 2012)、基本周波数の変化が小さいと落ち着いた声という印象を与えることも報告されている(木戸他, 2002)。そのため、音声の落ち着いた程度を操作することで、より効果的に子供と大人の区別をつけている可能性は十分に考えられる。

ここで注意しなければならないのは、本研究で用いたボイスサンプルは、自身がどのような演技ができるのか、その幅を示すという性格のものであるため、より音声間の違いが際立つように、誇張した典型的な芝居が行われていると考えられることである。そのため、たとえば年齢が若いキャラクターの男女の違いを迫力や張りで表現している可能性があるとしても、必ずしも現実世界の少年が少女に比べて張りのある声を出していることを意味していない。それよりはむしろ、アニメーションの世界において求められる典型的な少年の芝居が、若い少女の芝居に比べて、張りがあり迫力がある声であることを示唆する結果ではないかと考えられる。ただ、そのような音声を聞いた時に実際に聴取者の中に「少年」という印象が生じることを鑑みると、私たちが実際に少年の音声に対してそのようなステレオタイプを抱いている可能性は大いに考えられる。そのため、演技音声の限界は考慮に入れる必要があるが、典型的な芝居においてどのように声の印象が操作されているのかを知ることは、ヒトが声に対して抱いているイメージを知ることに関わり、ヒトの音声知覚を調べる上で1つの有用な方法になりうるであろう。

3. 実験 2：話者弁別

実験 1 の結果、声優のキャラクター演技音声は、同一の発話者であるにも関わらず異なった話者情報を持って知覚される音声であることが示された。実験 2 ではそれらの音声を用いて 3 つの弁別実験を行い、知覚される話者情報が異なっても同一の発話者であることが分かるのかを調べた。実験 2.1 では、地声で朗読された音声であれば発話者が弁別できることを確かめつつ、話者情報が異なって知覚される音声を用いた時に発話者の弁別が行えるのかを調べた。実験 2.2 と 2.3 では、それぞれ異なる発話者の組み合わせを用いて、話者情報が異なって知覚される音声の話者弁別について、より詳細に調べた。

3.1. 方法

3.1.1. 参加者

実験 2.1 には大学生 16 名が参加した（男性 5 名、女性 11 名）。平均年齢は 19.44 歳で、標準偏差は 0.73 歳であった。実験 2.2 には大学生 18 名が参加した（男性 7 名、女性 11 名）。平均年齢は 19.50 歳で、標準偏差は 0.92 歳であった。実験 2.3 には大学生 8 名が参加した（男性 3 名、女性 5 名）。平均年齢は 20.00 歳で、標準偏差は 0.53 歳であった。各参加者は、3 つの実験のうち 1 つのみに参加した。それ以外は実験 1 と同様であった。

3.1.2. 装置

基本的な装置は実験 1 と同様であった。ただし、実験プログラムは GNU Octave 4.0.3 と PsychoPhysics Toolbox を用いて作成した (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997)。

3.1.3. 音声

実験 1 で用いた音声のうち、話者 1～5 の音声、すなわち女性話者の音声のみを用いて実験を行った。

3.1.4. 音声の加工

実験 1 と同様に、無音部分を短縮した上で、ランダムスプライシングを用いて音声を加工した。250ms のセグメントをランダム順に 4 個並べて、1 つの音声につき試行ごとに 1 秒間の音声を作成した。

3.1.5. 条件

実験では 3 つの条件を用いた。まず、話者 1 および 2 の音声、すなわち地声で文章を朗読した音声を用いた条件であり、地声朗読条件とした。この条件では、朗読された文章が異なる 3 つの音声を呈示し、そのうち 2

つは同じ発話者の音声で、1つだけが別の発話者の音声であった。次に、1試行に呈示される3つの音声全てが大きく異なる話者情報を持って知覚される話者情報相違条件を用いた。この条件は、話者3、4または5の音声を用いており、1試行に呈示される3つの音声それぞれが、1つは幼い少女、1つは大人の女性、もう1つは少年の発話として知覚されるよう設定した。そのうち2つは同一の発話者によるものであり、1つがそれらとは異なる発話者によるものであった。さらに、話者3、4または5の音声を使い、1試行に呈示される3つの音声全てが似た話者情報を持って知覚されるような話者情報類似条件を用いた。この条件では、3つの音声全てが幼い少女、大人の女性、あるいは少年の発話として知覚されるよう設定した。そのうち2つは同一の発話者によるものであり、1つがそれらとは異なる発話者によるものであった。

各実験では、これらのうち2つの条件を用いた。実験2.1では、話者1～4の音声を使い、地声朗読条件と話者情報相違条件を用いた。実験2.2では、話者3と話者4の音声を使い、話者情報相違条件と話者情報類似条件を用いた。実験2.3では、話者3と話者5の音声を使い、話者情報相違条件と話者情報類似条件を用いた。

3.1.6. 手続き

実験は、1人ずつ静かな個室で行った。実験者が内容について指示を行い、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。まず始めに、本実験と同じ流れを用いて、高さの異なる純音を弁別する練習セッションを行った。練習セッションでは回答後直ちに、画面上の文字を使って正誤がフィードバックされた。この練習セッションで十分に課題内容を理解し、正答できていることを確認した上で、本実験に移行した。

本実験では、1試行に3つの音声が続けて呈示された。試行が始まると、500msの無音区間の後、1つ目の音声呈示された。1000msの無音区間を挟んで2つ目の音声呈示され、さらに1000msの無音区間の後に3つ目の音声呈示された。参加者の課題は、3つの音声のうち発話者が異なると感じる音声、すなわち仲間外れをキー押して回答するものであった。3つの音声のうち、いずれが正解であるかは、試行ごとにランダムであった。なお、発話者が全員女性であることはあらかじめ指示した。本セッションでは正誤のフィードバックを行わず、参加者の回答が行われたら、直ちに次の試行を開始した。1条件当たり24回を繰り返し、1人の参加者当たり48試行を行った。条件の呈示順はランダムであった。

3.2. 結果

各実験において、条件ごとに各参加者の正答率を計算し、それらを集計してバイオリンプロットと参加者個々のデータをドットでプロットしたものを組み合わせたグラフを作成した（図3）。

まず、各条件における正答率とチャンスレベルとの差について、1標本の t 検定を行った。実験2.1においては、話者情報相違条件における正答率はチャンスレベルと有意な差がなかった($t(15) = 1.00, p = .33, r = .25$)。一方、地声朗読条件における正答率は、チャンスレベルよりも有意に高かった($t(15) = 12.18, p < .0001, r = .95$)。実験2.2の話者情報相違条件における正答率も、チャンスレベルと有意な差がなかった($t(17) = 0.21, p = .83, r = .05$)。それに対して、実験2.2の話者情報類似条件における正答率は、チャンスレベルよりも有意に高かった($t(17) = 5.15, p < .0001, r = .78$)。また、実験2.3における正答率は、いずれの条件でもチャンスレベルよりも有意に高かった（話者情報相違条件： $t(7) = 3.16, p = .016, r = .77$ ；話者情報類似条件： $t(7) = 7.72, p = .00011, r = .95$ ）。

続けて、条件間の正答率の差について対応のある t 検定を行った結果、実験2.1においては、地声朗読条件の方が話者情報相違条件よりも正答率が有意に高かった($t(15) = 12.35, p < .0001, r = .95$)。また、実験2.2および2.3では、いずれの実験においても話者情報類似条件の方が話者情報相違条件よりも正答率が有意に高かった（実験2.2： $t(17) = 3.18, p = .0054, r = .61$ ；実験2.3： $t(7) = 3.72, p = .0075, r = .81$ ）。

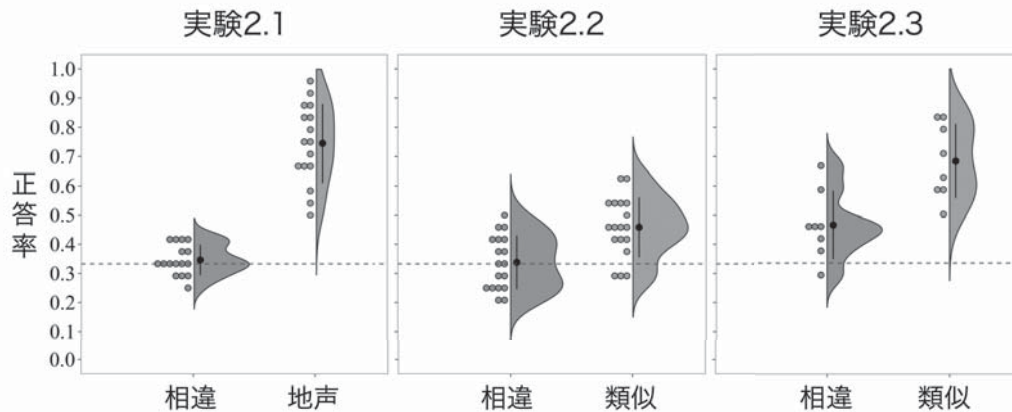


図3[†] 実験2の結果。左からそれぞれ実験2.1, 2.2, 2.3の結果を示す。各ドットはそれぞれの条件における参加者個々の正答率を表し、バイオリンプロットが併記されている。点線はチャンスレベルを表す。

[†] GitHubに公開されている、dgrtwoのgeom_flat_violin.Rとjbburantのhalf_violin_with_raw_data.Rを用いて作成した。

3.3. 考察

まず、地声朗読条件において、チャンスレベルよりも高い正答率が得られたが、正答率の平均は0.75程度であった。実験1.1のクラスター解析の結果（図1左）を見ると、話者1と話者2の音声を持つ話者情報は比較的似通っていること、また朗読音声は自然発話とはやや異なるため個々人の声の特徴が出にくかったこと、音声の呈示時間が1秒とあまり長くないことなどが、正答率が75%程度にとどまった要因だと考えられる。ただし、チャンスレベルよりは十分に高い正答率であり、効果量 r も0.95と大きいことから、地声での朗読であれば十分に弁別が行える課題であったといっていよう。

続けて、話者情報類似条件については、実験2.2および2.3のいずれでもチャンスレベルより高い正答率が得られた。この条件は、1試行に呈示される3つの音声全てが似た年齢・性別のキャラクターとして知覚される音声であり、話者情報が弁別の手がかりとして有用となりにくい条件である。この条件で弁別が行えていたことは、話者情報とは異なる手がかり、すなわち個人性情報に基づいた話者弁別が行えていたことを示唆する結果である。ただし、いずれの実験においても話者情報相違条件より高い正答率であったことも踏まえて考えると、似たような話者情報を持つ音声であっても、その中で話者情報の細かな発話者間の違いを利用していた可能性が十分に考えられるため、個人性情報のみを手がかりに弁別が行われたとまではいえない結果である。

話者情報相違条件については、話者3と4の音声を用いた実験2.1ならびに2.2ではチャンスレベルとの間に有意な差がなかった一方で、話者3と5の音声を弁別した実験2.3では、チャンスレベルよりも有意に高い正答率が示された。話者情報相違条件は、3つの音声全てが大きく異なる話者情報を持って知覚される条件であることから、実験2.3において弁別がある程度行えていたことは、話者情報とは異なる個々人に特有の成分、すなわち個人性情報に基づいて話者弁別が行えていたことを示唆する結果である。ただし、実験2.3は実験参加者が他の実験に比べて少ないため、チャンスレベルより高い正答率ではあるものの、さらなる検討が必要な結果であることも事実である。また、話者情報相違条件の正答率が、話者の組み合わせによって異なることも興味深い点である。実験1のクラスター分析や主成分分析の結果を踏まえて考えると、少なくとも実験1で用いたような表現語対で表されるような声質については、話者3と4が特に似ており、話者5だけが大きく異なっているという傾向は見取れない。実験間で参加者が異なることや人数が異なることもあり単純に比較はできないが、個人性情報の類似の程度が結果に影響していた可能性が考えられる。発話者によって個人の同定に関わる手がかりが異なる可能性については先行研究でも指摘されており（e.g., Van Lancker, Kreiman, & Emmorey, 1985）、個人性情報は必ずしも個人間で共通の特徴量で表されるとは限らない。個人性情報がどのような特徴量で表されるのか、個人の同定に関わる特徴量が個人間でどの程度共通しており、どの程度異なるのか、それらの特徴量を処理する過程がどのようなものであるのかといったことは、今後の重要な検討課題である。

4. 総合考察

本研究では、声優が異なるキャラクターを演じた音声を用いて、2つの実験を行った。まず実験1では、声質表現語を用いて印象評定を行い、声優の演技音声がキャラクターによって様々な年齢・性別を持って知覚されていることを示す結果が得られた。続けて実験2では、実験1で見出された異なるキャラクターとして知覚される音声を用いて発話者を弁別する実験を行い、たとえ異なる年齢や性別を持って知覚される音声であっても、条件によってはある程度発話者を弁別できることが示された。これらの結果を踏まえて本節では、音声知覚研究における声優のキャラクター演技音声の有用性について議論を行う。

声優のキャラクター演技に着目した研究としては、作中での役割に特有の声質として、善玉と悪玉のステレオタイプ的な声質に着目した研究が存在する (Teshigawara, 2004b; 勅使河原・伊藤・武田, 2005)。また、声優の芝居とはやや異なるものの、プロの物真似タレントによる物真似音声に着目し、本人の音声と音響特徴を比較した研究や (北村, 2007)、歌舞伎の女形が様々な年齢の女性を演じた音声の音響特徴を分析した研究が存在している (森他, 2014)。このように、演技音声に着目した研究は、これまでも行われてきている。しかし、個々の声優がどのようなキャラクターを演じるのか、どう演じ分けるのかに着目した研究はこれまでに行われていない。この「音声でキャラクターを演じる」という行為は、必ずしもアニメーションにおける声優の演技のような日常から離れた場所にだけ存在しているわけではない。たとえば電話をとる時に少し声が変わったり、初対面の人と話す場合と親しい相手と話す場合で話し方や声のトーンが違ったりすることは、多くの人にとって経験があることではないだろうか。そのような発話の制御は、聴取者が抱く印象に声質や話し方が与える影響を、自覚的あるいは無意識的に考慮していることが1つの要因だと考えられる。実際、発話者の声質や性別が音声に基づいたパーソナリティの判断に影響を与えることや (Aronovitch, 1976; Scherer, 1972, 1978)、声質が感情の判断に影響を与えること (池本・鈴木, 2009)、発話速度、抑揚の大きさなどの韻律的特徴や母音の明瞭性といった話し方が性格印象に影響を与えることがこれまでに報告されている (内田, 2002, 2005, 2009, 2011)。このようなある種のステレオタイプを利用して、状況に合わせて相手に与えたい印象を与えることが可能となるからこそ、音声を操作することが日常的に行われているのではないかと考えられる。そしてそのような音声操作の1つの極端な例が、声優がキャラクターを演じた音声であると捉えることができる。特に日本人は、たとえば感情を判断する時に、視覚情報である顔よりも聴覚情報である声を重視することが報告されており (Tanaka et al., 2010)、発話者について何らかの判断を行う際に、声に重きを置いていることが考えられる。このような、音声から受ける印象や話者内での声質の変化、それに基づいた知覚を調べる上で、様々な年齢・性別・性格・体格などのキャラクターを演じることのできる声優の音声は、非常に有用だと考えられる。

これまでに、ヒトが音声に基づいて発話者自身に関する情報を知覚する過程について、様々なアプローチを用いて研究が行われてきた (Belin, Bestelmeyer, Latinus, & Watson, 2011; Podesva & Callier, 2015; Schweinberger & Burton, 2011; Schweinberger, Kawahara, Simpson, Skuk, & Zaske, 2014; Young & Bruce, 2011 for recent reviews)。たとえば、音声を呈示して、発話者の年齢や体格などの情報をどの程度正確に知覚できるかを調べた研究や (Krauss, Freyberg, & Morsella, 2002)、逆再生やランダムスプライシング、ローパスフィルタをかけるなどの操作を行った際に、声質の知覚がどのように変化するかを調べるような研究が行われてきた (van Bezooijen & Boves, 1986; Scherer et al., 1985; Teshigawara, 2004a; Van Lancker et al., 1985)。あるいは、病的な声質をどのように評価するのかということも1つの重要な観点であり、様々な研究が行われている (Kreiman & Gerratt, 2010; Kreiman, Gerratt, Kempster, Erman, & Berke, 1993 for reviews)。また、個々人の特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、生理学的特徴と音声の音響特徴を結びつけたり、音響特徴とヒトの知覚的な評定との関係を調べたりする研究も行われ、どのような特徴量が年齢、性別や個人性の知覚に関わるのかについても調べられてきた (e.g.,

Kitamura & Akagi, 1995; Kitamura et al., 2005) (for review, see Schweinberger et al., 2014)。さらに近年では、工学的な技術の進歩もあり、音声の特定の音響特徴量を、他の特徴量を保持しながら高精度に操作することが可能となっている。たとえば、Tandem-STRAIGHT (Kawahara et al., 2008) や WORLD (Morise, Yokomori, & Ozawa, 2016) などは、高精度の音声分析合成システムの例である。これらのシステムを用いて音響特徴を精密に操作し、音響特徴量の変化に伴う知覚的評価の変化を調べたり (内田, 2005, 2009, 2011, 2017), 音声のモーフィングを行って順応実験が行われたりしている (Latinus & Belin, 2011, 2012; Schweinberger et al., 2008; Zaske, Schweinberger, & Kawahara, 2010)。また、顔も発話者に関する情報を伝えるものであることから、顔と声を結びつけて個人を同定する過程についても研究が行われている (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Robertson & Schweinberger, 2010; Smith, Dunn, Baguley, & Stacey, 2016; Zweig, Suzuki, & Grabowecy, 2015) (for review, see Yovel & Belin, 2013)。本研究はここに、新たな研究アプローチとして『錯聴』としてのキャラクター演技音声を用いた研究を提示したい。声優のキャラクター演技音声は、同一の発話者であるにも関わらず異なった話者情報を持って知覚される音声、すなわち個人性情報と話者情報を明確に切り分けて検討することが可能な音声である。演技音声は、厳密な実験研究を行う上では刺激の統制などにおいて不安が残ること、また生態学的妥当性についての懸念が存在するのは事実である。しかし、視覚や聴覚を始めとした知覚研究は、知覚システムをうまく「だます」ことのできる錯覚の発見や発明をきっかけに進展してきた面がある (柏野, 2010; 村上, 2010)。そして声優の演技は、視聴者を世界に引き込むために、実際には存在しない世界がリアルなものとして感じられるように、存在しないキャラクターが生きているかのように、いい意味で「だます」ことを1つの目的としている。したがって、錯聴としてキャラクター演技音声を用いることは、音声知覚に関する研究を進める上で、大きな可能性を秘めていると考えられる。そしてキャラクター演技音声を、これまでに用いられてきたような音響特徴量の分析や順応実験などと組み合わせることで、話者情報と個人性情報を明確に切り分けた上で、それぞれに関する選択的な知覚処理メカニズムの存在や、処理の階層性や独立性について検討することが可能となる。また、声の芝居の芸術的な側面についても考えることで、たとえば声の魅力のような、定義づけの難しい部分についてもアプローチできる可能性がある。声の魅力に関しては、「平均声」が平均的に魅力的であることを示した研究が存在するが (Bruckert et al., 2010), 女性の音声に対する好感度の評価が男女で異なっている可能性も示唆されている (横森・二宮・森勢・田中・小澤, 2016)。好みに関する個人差も考慮に入れると、まだ研究すべき部分が大いに残されている問題だといえる。本研究は、声優の声の芝居を知るという観点でも、ヒトの音声知覚を知るという観点でも、その一端を対象としたに過ぎない。本研究を端緒として今後、声優の演技音声を用いた研究がさらに活発になることで、ヒトの音声知覚に関する理解が進むとともに、声の芝居に対する理解も深まることを願っている。

5. 謝辞

本研究はJSPS 科研費 (18H05821) の助成を受けたものです。本研究を専修大学人間科学部心理学科の基礎実験2の授業内で実施するにあたり、専修大学の山上精次先生には、実験の実施を快諾いただき、また様々な面でご協力いただきました。この場を借りて感謝の意を表します。また、千葉大学の清水求さんにも、実験データを収集する際にご協力いただきました。ありがとうございました。所属声優の方々のボイスサンプルを研究に使用することに許諾をくださいました株式会社マウスプロモーション様にも感謝申し上げます。最後に、実験の実施を可能にした声優の皆様の声の演技に感謝と敬意を表します。

注：本論文のデータは、著者が専修大学で非常勤講師を務めていた時期に行った実験の結果である。本研究の一部は、第31回日本音声学会全国大会および日本基礎心理学会第37回大会で発表された。

参考文献

- Aronovitch, C. D. (1976). The voice of personality: stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology*, 99, 207–220.
- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102, 711–725.
- van Bezooijen, R., & Boves, L. (1986). The effects of low-pass filtering and random splicing on the perception of speech. *Journal of Psycholinguistic Research*, 15(5), 403–417.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 443–446.
- Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20, 116–120.
- Fujisaki, H. (1996). Prosody, models, and spontaneous speech. In Sagisaka, Y., Campbell, N., & Higuchi, N. (Eds.) *Computing Prosody*, Springer, 27–42.
- Garrido, L., Eisner, J., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., …Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia*, 47, 123–131.
- 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.
- 池本 真知子・鈴木 直人 (2009). 感情判別における声質の影響—単音節, 短文を用いて— 感情心理学研究, 16(3), 209–219.
- Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). 'Putting the face to the voice': Matching identity across modality. *Current Biology*, 13, 1709–1714.
- 柏野 牧夫 (2010). 音のイリュージョン 岩波書店
- 粕谷 英樹・鈴木 久喜・城戸 健一 (1968). 年令, 性別による日本語5母音のピッチ周波数とホルマント周波数の変化 日本音響学会誌, 24(6), 355–364.
- Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: a temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *Proceedings of ICASSP*, 3933–3936.
- 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.
- 木戸 博・箕輪 有希子・粕谷 英樹 (2002). 声質表現語の音響関連量に関する非線形分析—決定木による方法— 日本音響学会誌, 58(9), 586–588.
- 北村 達也 (2007). 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, 107(282), 1–6.
- Kitamura, T., & Akagi, M. (1995). Speaker individualities in speech spectral envelopes. *Journal of Acoustical Society of Japan*, 16(5), 283–289.
- Kitamura, T., Honda, K., & Takemoto, H. (2005). Individual variation of the hypopharyngeal cavities and its acoustic effects. *Acoustical Science and Technology*, 26(1), 16–26.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). "What's new in Psychtoolbox-3?". *Perception*, 36, ECVF Abstract Supplement.
- Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, 38, 618–625.
- Kreiman, J., & Gerratt, B. R. (2010). Perceptual assessment of voice quality: Past, present, and future. *Perspectives on Voice and Voice Disorders*, 20, 62–67.
- Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., & Berke, G. S. (1993). Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, 36, 21–40.
- 栗田 茂二郎 (1988). 声帯の成長, 発達と老化 音声言語医学, 29, 185–193.
- Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2, 175.
- Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief vowel stimuli. *PLoS ONE*, 7(7), e41384.

- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1–12.
- Maekawa, K. (2012). Prediction of non-linguistic information of spontaneous speech from the prosodic annotation: Evaluation of the X-JToBI system. *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 991–996.
- 森 大毅・前川 喜久雄・粕谷 英樹 (2014). 音声は何を伝えているか コロナ社
- Morise, M., Yokomori, F., & Ozawa, K. (2016). WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE transactions on information and systems*, E99-D(7), 1877–1884.
- 村上 郁也 (2010). 視覚研究ツールとしての錯視 光学, 39(2), 66–74.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, 109(6), 2988–2998.
- Podesva, R. J., & Callier, P. (2015). Voice quality and identity. *Annual Review of Applied Linguistics*, 35, 173–194.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Robertson, D. M. C., & Schweinberger, S. R. (2010). The role of audiovisual asynchrony in person recognition. *The Quarterly Journal of Experimental Psychology*, 63(1), 23–30.
- 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.
- Scherer, K. R. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, 40(2), 191–210.
- Scherer, K. R. (1978). Personality inference from voice quality: the loud voice of extroversion. *European Journal of Social Psychology*, 8, 467–487.
- Scherer, K. R., Feldstein, S., Bond, R. N., & Rosenthal, R. (1985). Vocal cues to deception: A comparative channel approach. *Journal of Psycholinguistic Research*, 14(4), 409–425.
- Schweinberger, S. R., & Burton, A. M. (2011). Person perception 25 years after Bruce and Young (1986): An introduction. *British Journal of Psychology*, 102, 695–703.
- Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., Robertson, D. M. C., Simpson, A. P., & Zaske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, 18, 684–688.
- Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zaske, R. (2014). Speaker perception. *WIREs Cognitive Science*, 5, 15–25.
- Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research*, 57, 285–296.
- Smith, H. M. J., Dunn, A. K., Baguley, T., & Stacey, P. C. (2016). Matching novel face and voice identity using static and dynamic facial images. *Attention, Perception, & Psychophysics*, 78, 868–879.
- Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research*, 30, 522–529.
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, 21(9), 1259–1262.
- Teshigawara, M. (2004a). Random splicing: A method of investigating the effects of voice quality on impression formation. *Proceedings of Speech Prosody 2004, Nara*, 209–212.
- Teshigawara, M. (2004b). Vocally expressed emotions and stereotypes in Japanese animation: Voice qualities of the bad guys compared to those of the good guys. *Journal of the Phonetic Society of Japan*, 8(1), 60–76.

- 勅使河原 三保子・伊藤 克亘・武田 一哉 (2005). 日本のアニメの音声に表された感情と性格—声のステレオタイプの音声学的研究— 電子情報通信学会技術研究報告, *105(291)*, 39–44.
- 内田 照久 (2002). 音声の発話速度が話者の性格印象に与える影響 心理学研究, *73(2)*, 131–139.
- 内田 照久 (2005). 音声中の抑揚の大きさと変化パターンが話者の性格印象に与える影響 心理学研究, *76(4)*, 382–390.
- 内田 照久 (2009). 音声の韻律的特徴と話者のパーソナリティ印象の関係性 音声研究, *13(1)*, 17–28.
- 内田 照久 (2011). 音声中の母音の明瞭性が話者の性格印象と話し方の評価に与える影響 心理学研究, *82(5)*, 433–441.
- 内田 照久 (2017). 話者の匿名性の確保を目的とした声道長の制御を模した声質変換の評価 日本音響学会誌, *73(3)*, 151–162.
- Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, *25(5)*, 829–834.
- Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: patterns and parameters Part I: Recognition of backward voices. *Journal of Phonetics*, *13*, 19–38.
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., & Lindell, R. G. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Acoustical Society of America*, *125(3)*, 1666–1678.
- 横森 文哉・二宮 大和・森勢 将雅・田中 章浩・小澤 賢司 (2016). 好感度評価の性差に着目した女性発話の音響特徴量分析 日本感性工学会論文誌, *15(7)*, 721–729.
- Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, *102*, 959–974.
- Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17(6)*, 263–271.
- Zaske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, *268*, 38–45.
- Zweig, L. J., Suzuki, S., & Grabowecy, M. (2015). Learned face-voice pairings facilitate visual search. *Psychonomic Bulletin & Review*, *22(2)*, 429–436.