論　　文

女性声優の演技音声における年齢・性別の表現と関連する音響特徴量*

林大輔*1　　森勢将雅*2

[要旨]　本研究では，女性声優が異なる年齢・性別を意図した演技音声を用いて，聴取実験ならびに音響特徴量の解析を行った。その結果，意図した 5 キャラクター（幼い少女・中高生程度の少女，大人の女性，老婆，少年）の聴取印象がそれぞれ異なっていることが示された。また，キャラクター表現と関連する音響特徴量について，女性 4 キャラクターについては実際の加齢変化との関連が見て取れた一方で，少年声については特有の表現が用いられている可能性が示唆された。考察では，より幅広いキャラクター表現を対象とすることで研究を発展させられる可能性を議論しつつ，メディア芸術におけるキャラクター表現の理解深耕や工学応用に向けた展望を述べる。

キーワード　　音声，声優，演技，話者情報，個人性情報，音響特徴量
Voice, Voice Actor, Acting, Speaker Information, Identity Information, Acoustic Features

1.　はじめに

ヒトが発する音声には, 言語をはじめとした様々な情報が含まれており, その中には発話者自身に関する年齢や性別といった情報も含まれている[1,2]。これらの発話者自身に関する情報は, 発話者の身体と密接に関わっている。たとえば，大人に比べて子供の方が，あるいは男性に比べて女性の方が平均的に体が小さいため, 子供の方が声帯も声道も短く, 結果として基本周波数とスペクトル重心は高くなる[2,3,4,5,6,7]。このように，それぞれの人の身体的な特性が声には反映されることもあり, 年齢や性別などの一般的な情報だけでなく「声の持ち主が誰であるか」という個人性に関する情報も，声から知ることができる。すなわち，通常，発達や病気による変化は別として, ある時点におけるある人の声に含まれる情報は，言語や感情，意図などの情報が変わることはっても，年齢や性別，個人性に関する情報は基本的に変化しないと考えられる。

しかし，同一の人であっても，異なる年齢や性別に聞こえる音声を発することは可能であり, その顕著な例が声優の演技音声である。声優は様々な年齢のキャラクターを演じることがあり，また特に日本においては，女性声優が少年，すなわち男性を演じることも珍しくない[8]。つまり声優の演技音声は, 年齢や性別といった情報と「誰であるのか」という個人性に関する情報が乖離しうる音声であり, ヒトが音声から発話者に関する情報を処理し, 知覚するメカニズムを理解する上で有用な刺激となりうる[9]。

では，声優は同一の人物でありながら，どのように年齢や性別を演じ分けているのであろうか。この点について，これまでにいくつかの研究が行われてきている。たとえば，声優のボイスサンプルを用いて，異なる年齢・性別のキャラクターを演じた音声についての聴取実験を行った研究や[9]，同一の声優が男性と女性を演じた際の音響特徴量の違いについて調べた研究が存在する[10,11]。また，年齢や性別ではなく，より広い意味での“キャラクター”に関する演技音声について，その音響特徴量に関する研究が行われている[12,13,14,15,16,17,18]。しかしながらこれらの研究では，実際のアニメやゲーム，ドラマ CD，あるいは声優事務所のボイスサンプルなど, 既存の音声データを抽出して聴取実験や音響特徴量の解析を行っており, 音声間で読み上げている文章が異なっていたり, 研究対象となる声優が 1 人のみなど限られた条件となっている。そこで本研究では，同一の文章を，複数の声優が異なる年齢・性別を意図して読み上げた音声を用いて, 聴取実験ならびに音響特徴量の解析を行う。その際，演じ分けとの関連を検討するため, ある程度聴取印象との関係が示されている, 解釈性の高い音響特徴量を対象とする。その上で，年齢・性別の演じ分けと関わる音響特徴量について，聴取印象と音響特徴量との関係について重回帰分析を用いて明らかにする。

---

* Acoustical features relating to the representation of age and gender in acting voices by female voice actors.

*1 日本たばこ産業株式会社
*2 明治大学
（問合先：林大輔　〒105-6927　東京都港区虎ノ門 4-1-1 神谷町トラストタワー E-mail: daisuke.hayashi@jt.com）

2. 方法

2.1 音声の収録

株式会社マウスプロモーションへの業務委託を通じて，音声の収録を行った。10 名の女性声優に，共通の 5 キャラクター（幼い少女・中高生程度の少女，大人の女性，老婆，少年）を演じながら，指定した文章を読み上げてもらった。その際，幼い少女と少年は同年代を想定している旨を教示した。文章は，日本声優統計学会が，Wikipedia から文章を抜き出す形で作成した音素バランス文をもとにして[19]，JSUT コーパス[20]で句読点が追加され（JSUT コーパス版の改変済み声優統計コーパス voiceactress100），つくよみちゃんコーパス[21]でイントネーションなどの補足情報が追加されたものから，10 文章を選定して用いた。音声はスタジオの防音室で収録され，ノイズ除去などの後処理は行わずに，48 kHz，24 bit の wav ファイルで納品された。

2.2 音声の主観評価実験

クラウドワークスで募集した 24 歳～67 歳の一般の方 100 名（男性 65 名，女性 35 名，平均年齢 40.5 歳，標準偏差 7.91 歳）を対象に，オンラインで実験を実施した。参加者には同意説明書を画面上で提示し，ボタン押しにより同意を取得した。なお，実験は事前に日本たばこ産業株式会社の倫理委員会の承認を得て（RE-SE-2022-01），株式会社イデアラボ協力のもと行われた。

評価対象は，女性声優 10 名×5 キャラクターの計 50 音声であった。参加者の負担低減のため，全音声を声優 2 名ずつ（10 音声ずつ）の 5 セットに分けて，参加者はいずれか 1 セットについて評定を行った。そのため，1 音声当たりの評価者は 20 名であった。音声は「森永のおいしい牛乳は，濃い青色に，牛乳瓶をあしらったデザインの，パック牛乳である。」という文章を読み上げているものを用いた。音声の提示時間は，1 音声につき 10 秒程度であった。

評価項目は，声質に関する日常表現語であり，先行研究[22]が収集・抽出した 25 語を用いた。それぞれ 7 件法（全く当てはまらない～とてもよく当てはまる）で評定した。評定課題は jsPsych 6.3.0 を用いて作成した[23]。

参加者は自身の PC で実験に参加した。実験はヘッドフォンまたはイヤフォンを装着した状態で実施し，音量は始めに参加者自身が適切に調整した。課題内容を画面上で説明したのち，JVS コーパス[24]に含まれる，本試行と同一の文章を読み上げた音声を用いて，練習試行を 3 試行実施した。その後の本試行では，20 音声を参加者ごとにランダム順で 1 つずつ提示して評定を行った。実験は 1 人当たり 15～30 分程度であった。

2.3 音声の音響特徴量解析

まず，各音声ファイルについて，音素ラベリングを行った。48 kHz・24 bit で収録された音声を 16 kHz・16 bit にダウンサンプリングし，連続音声認識ソフトウェアである Julius 4.3.1 を用いて自動音素ラベリングを行った[25]。音響モデルは，monophone モデルである hmmdefs_monof_mix16_gid を用いた。

次に，音声分析合成システムである WORLD [26] (Harvest [27], D4C edition [28])の Python ラッパーである pyworld 0.2.8 を Python 3.6.9 上で用いて，48 kHz・24 bit である元の音声ファイルを対象に，基本周波数・スペクトル包絡・非周期性指標の時系列情報を音声ごとに抽出した。自動音素ラベリングの結果に基づいて，それぞれの時系列情報から母音部分のデータのみを取り出して，後段の分析を行った。その際，基本周波数の推定結果が 0 だったセグメントや，音声ごとの中央値より 2 倍以上値が大きなセグメントは解析から除外した。これらの基本周波数の時系列データから除外したセグメントは，スペクトル包絡ならびに非周期指標の時系列データからも除外した。

続けて，得られた基本周波数・スペクトル包絡・非周期性指標それぞれの時系列データに基づいて，様々な音声の音響特徴量を算出した。本実験では，全ての音声で読み上げた文章（韻質）が共通していたため，広義の声質（音声波から知覚される韻質以外の聴覚上の特質[2,29]）に着目して，ヒトの聴覚印象との関連がこれまでに示されている以下の音響特徴量を算出した。

- 基本周波数（mel）（中央値，標準偏差）
- スペクトル重心（Hz）（中央値）
- スペクトル傾斜（dB/Oct）（中央値）
- 倍音振幅差分（dB）（中央値）
- 平均パワー（dB）（中央値，標準偏差）
- 非周期性指標（単位なし）（中央値）

以下，それぞれの特徴量について，基本周波数・スペクトル包絡・非周期性指標いずれの時系列データに基づいて算出したかによって，項を分けて詳細を記載する。なお，上記に加えて，音素ラベリングの結果を用いて得られた発話時間の長さおよび原稿の読点における無音時間の長さに基づいて

- 話速（モーラ/秒）
- ポーズ長（秒）

を算出した。

2.3.1 基本周波数に基づく特徴量

基本周波数は，知覚的な声の高さと強く関連している音響特徴量であり[30]，標準偏差で表すことのできる基本周波数の変化の大きさは抑揚の大きさととらえることができる。本研究では，ヒトの聴取印象との関連に関

心があるため, WORLD で抽出した各時間点における基本周波数 f（Hz）を，ヒトの高さの知覚的尺度であるメル尺度に以下の式に基づいて変換した上で[31]，時系列データの中央値ならびに標準偏差を算出した（単位は mel）。

$$\mathrm{mel}(f) = 1127.01048\,\mathrm{Log}\left(\frac{f}{700}+1\right) \ldots (1)$$

2.3.2　スペクトル包絡に基づく特徴量

WORLD で抽出したスペクトル包絡に基づいて, スペクトル重心・スペクトル傾斜・倍音振幅差分・平均パワーを算出した。

スペクトル重心は, 知覚的な明るさと相関することが報告されている[32]。以下の式に基づいて[33], 各時間点におけるスペクトル重心を求め, 時系列データの中央値を算出した（単位は Hz）。

$$S_c = \frac{\sum_{k=0}^{\frac{N}{2}} f(k)A(k)}{\sum_{k=0}^{\frac{N}{2}} A(k)} \ldots (2)$$

$$f(k) = \frac{f_s}{N}k \ldots (3)$$

なお, fs は標本化周波数（48 kHz）に, N は FFT 長（2048）に，$A(k)$ は離散周波数番号 k における振幅（WORLD の出力したパワースペクトルの平方根）に対応する。

スペクトル傾斜はスペクトル包絡の傾きを表し, 基本的に負の値となるが，値が小さい（つまり傾きが急峻）なほど高周波成分のパワーが小さく，値が大きい（つまり傾きが緩やか）なほど高周波成分のパワーが大きいことを表す。そして値が大きいほど（つまり高周波のパワーが大きいほど），知覚的な印象としての気息性（息漏れがあり柔らかい感じ）が高く知覚されると言われている[34,35]。縦軸を対数パワー，横軸を周波数の対数として，最小二乗法で線形単回帰（直線近似）を行った際の回帰係数をスペクトル傾斜として各時間点において算出し，その時系列データの中央値を用いた（単位は dB/Oct）。その際，基本周波数や低域の倍音構造の影響を除外するため，1000 Hz 以上の周波数成分のみを用いて分析を行った[36]。

倍音振幅差分は低域の調波構造に関する音響特徴量であり, 第 1 倍音の振幅から第 3 倍音の振幅を引いた差分を算出した。倍音振幅差分について，自発音声と演技音声を比べた際に, 自発音声よりも演技音声において値が大きくなると報告されている[37]。第 1 倍音の対数パワーから第 3 倍音の対数パワーを引いた差分を各時間点において算出し, その時系列データの中央値を用いた（単位は dB）。

平均パワーは声の大きさと関連する音響特徴量であり，その標準偏差は声の大きさの変化（声の強弱のつけ方）としてとらえることができる。各周波数における対数パワーの平均値を各時間点で算出し, その時系列データの中央値と標準偏差を用いた（単位は dB）。

2.3.3　非周期性指標に基づく特徴量

非周期性指標は声のかすれ具合と関連するような雑音成分であり[38]，波形全体のパワーに対する非周期的な成分のパワーの割合として定義される[33]。WORLD で抽出した非周期性指標について, その時系列データの中央値を算出して用いた（単位なし）。

3.　結果

3.1 音声の主観評価結果

データの解析には，R 3.6.1 を用いた[39]。各音声に対する各評価項目の評定結果について, 参加者間の平均値を算出し，得られたデータに基づいて 25 の評価項目の情報を縮約するために主成分分析を行った。第 3 主成分までの主成分負荷量ならびに寄与率を表-1 に示した。

続けて, 各音声について第 3 主成分までの主成分得点を算出し, キャラクター（幼い少女・中高生程度の少女, 大人の女性，老婆，少年）ごとに分けて箱ひげ図をプロットした（図-1）。キャラクター間の主成分得点の差について，JASP 0.18.3[40]を用いて一要因分散分析を行った。まず第 1 主成分において，キャラクターの主効果が有意であった（F (4, 45) = 127.9, p < .001，η^2 = 0.92）。Tukey の補正に基づいて多重比較を行った結果，老婆＞大人の女性＞中高生程度の少女≒少年＞幼い少女という関係が示された（＞は ts (9) > 4.0, adjusted ps < .01 であり，≒は t (9) = 0.42, adjusted p = .993）。次に第 2 主成分について，キャラクターの主効果が有意であった（F (4, 45) = 68.4, p < .001，η^2 = 0.86）。Tukey の補正に基づいて多重比較を行った結果, 大人の女性＞中高生程度の少女＞老婆≒少年≒幼い少女という関係が示された（＞は ts (9) > 4.2, adjusted ps < .001 であり，≒は ts (9) < 0.52, adjusted ps > .98）。最後に第 3 主成分について，キャラクターの主効果が有意であった（F (4, 45) = 8.6, p < .001，η^2 = 0.43）。Tukey の補正に基づいて多重比較を行った結果，老婆・中高生程度の少女・幼い少女の間の差は有意でなかった（ts (9) < 2.0, adjusted ps > .31）。少年は老婆・中高生程度の少女・幼い少女の 3 キャラクターよりも値が高く（ts (9) > 3.6, adjusted ps < .01），大人の女性との差は有意でなかった（t (9) = 2.4, adjusted p = .14）。大人の女性は幼い少女よりも値が高く（t (9) = 3.2, adjusted p = .02），老婆・中高生程度の少女の 2 キャラクターとの差は有意でなかった（ts (9) < 1.5, adjusted ps > .60）。

表-1　音声の主観評価の主成分分析結果

	主成分負荷量		
	第1主成分	第2主成分	第3主成分
若い感じの声	**-.99**	.05	.00
かわいい声	**-.97**	.04	-.17
渋い声	**.95**	-.17	.05
明るい声	**-.94**	-.02	.13
高い声	**-.93**	.02	-.14
生き生きとした声	**-.92**	-.03	.32
子供っぽい声	**-.91**	-.37	-.10
かすれた声	**.91**	-.30	-.09
通りの良い声	**-.89**	.28	.27
つぶれた声	**.88**	-.37	-.01
澄んだ声	**-.87**	.44	.03
がらがら声	**.87**	-.39	.00
だみ声	**.85**	-.42	.00
ドスの効いた声	**.84**	-.30	.27
響きのある声	**-.84**	.24	.23
張りのある声	**-.81**	.09	.54
太い声	**.79**	-.17	.42
落ち着きのある声	**.74**	.63	-.01
金きり声	**-.72**	-.34	-.17
色っぽい声	.02	**.96**	-.02
品のある声	.40	**.89**	.00
女性的な声	-.14	**.68**	-.43
鼻の詰まった声	.02	**-.64**	-.27
迫力のある声	.43	-.27	**.75**
弱々しい声	.64	-.03	**-.69**
寄与率	.74	.14	.05
累積寄与率	.74	.88	.93

3.2 音声の特徴量解析結果

音声の各音響特徴量について，各キャラクター内で話者間の平均値と標準誤差を算出して，表-2 に示した。

3.3 主観評価と音響特徴量との関係

質問紙項目の第 1 主成分から第 3 主成分までの主成分得点を目的変数，標準化した各音響特徴量を説明変数として，50 音声について，AIC（赤池情報量基準）に基づいてステップワイズ法で重回帰分析を行った（表 3）。VIF（Variance Inflation Factor）は全ての変数において 5 よりも小さかったため，多重線形性の懸念はないと判断した。

4.　考察

4.1 演技音声の主観的印象

評定項目の主成分負荷量（表-1）ならびに各キャラクターにおける主成分得点（図-1）に基づくと，第 1 主成分は年齢（主成分得点が高いほど年齢が高い），第 2 主成分は「大人の女性らしさ」（色っぽい声，品のある声，落ち着きのある声，澄んだ声，女性的な声などの評定項目の負荷量が高い），第 3 主成分は「少年らしさ」（張りのある声，太い声，迫力のある声などの評定項目の負荷量が高い）を表現していると考えられる。第 3 主成分までで累積寄与率が 93%であり，これら 3 つの要素で今回の演技音声において意図した 5 キャラクター（幼い少女・中高生程度の少女，大人の女性，老婆，少年）の主観的印象の違いが表現できると考えられる。

女性 4 キャラクター（幼い少女・中高生程度の少女，大人の女性，老婆）については，基本的に年齢の違いが大きな要素だと考えられるが，それに加えて「大人の女性らしさ」らしきものが出ているのは興味深い点である。4.2 項でも記載する通り，たとえば話し方の 1 つである話速は，6〜7 歳児とお年寄りは同じくらいの速度である一方で，若い大人はそれらよりも速いことが報告されており[41]，年齢を重ねるにつれて線形に変化する要素だけでなく，大人の女性においてピークを迎える要素も存在することが，今回の結果に繋がっている可能性が考えられる。

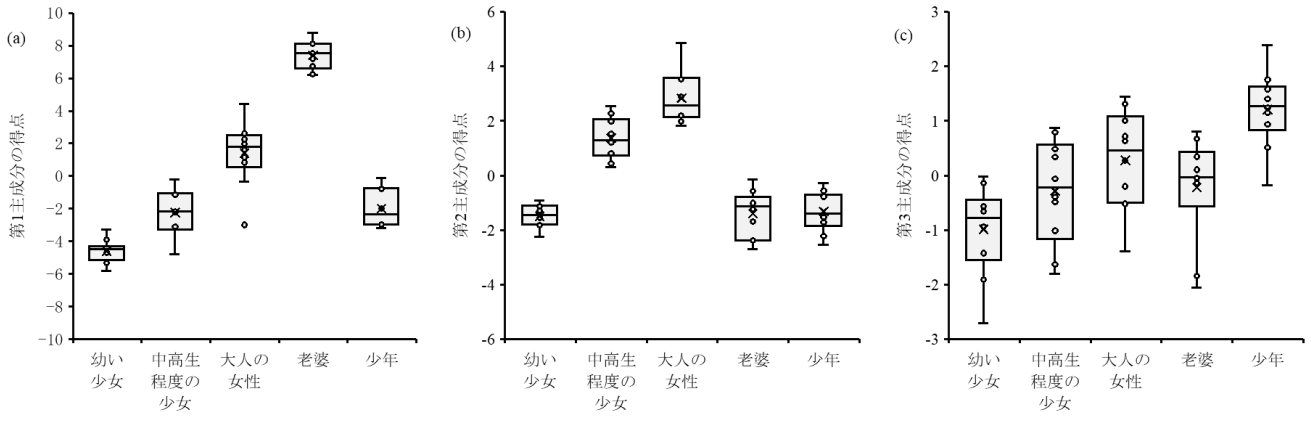

図-1　キャラクターごとの主成分得点 (a)第 1 主成分 (b)第 2 主成分 (c)第 3 主成分

表-2　キャラクターごとの音響特徴量（それぞれ上段は話者間平均，下段は標準誤差を表す）

	時系列データの中央値						時系列データの標準偏差		話速（モーラ/秒）	ポーズ長（秒）
	基本周波数（mel）	スペクトル重心（Hz）	スペクトル傾斜（dB/Oct）	倍音振幅差分（dB）	平均パワー（dB）	非周期性指標（単位なし）	基本周波数（mel）	平均パワー（dB）		
幼い少女	336.53	1958.47	-16.67	7.99	75.20	0.59	79.67	6.64	6.56	1.81
	15.88	80.36	0.65	0.77	1.34	0.02	4.16	0.27	0.22	0.18
中高生程度の少女	271.63	1694.08	-14.48	9.20	73.53	0.66	73.47	6.75	7.32	1.83
	13.97	61.51	0.51	0.82	0.89	0.02	5.24	0.34	0.15	0.18
大人の女性	225.83	1470.16	-14.10	12.88	70.02	0.67	62.22	6.64	7.28	1.91
	10.43	62.31	0.52	0.65	1.19	0.01	4.12	0.51	0.11	0.17
老婆	197.36	1525.15	-15.58	11.43	68.70	0.65	55.35	6.00	6.05	2.20
	8.29	90.95	0.76	1.36	1.10	0.02	3.14	0.35	0.18	0.35
少年	259.60	1800.57	-16.39	8.13	72.75	0.62	72.57	7.10	7.03	1.79
	12.08	69.95	0.69	1.11	1.16	0.02	3.40	0.45	0.17	0.19

表-3　主成分得点と音響特徴量との重回帰分析結果

		標準化偏回帰係数（β）		
		第1主成分	第2主成分	第3主成分
時系列データの中央値	基本周波数	**-0.67**	0.35	**-1.27**
	スペクトル重心			
	スペクトル傾斜	-0.14	**0.76**	**-0.86**
	倍音振幅差分	**0.20**		0.17
	平均パワー	-0.14	-0.40	**0.83**
	非周期性指標			
時系列データの標準偏差	基本周波数	-0.14		**0.61**
	平均パワー	-0.07		
話速		**-0.42**	**0.38**	0.19
ポーズ長				-0.17
（切片）		6.05E-16	2.94E-16	-1.18E-15
調整済みR^2		**0.9022**	**0.5322**	**0.4800**

太字下線：p < .001、太字：p < .01、下線：p < .05
文字のみ：p > .05、空欄：モデル選択で削除

また「少年らしさ」については，「張りのある声」「太い声」「迫力のある声」などが主観的印象として関連が深く，これは声優のボイスサンプルを用いた先行研究と同様の傾向である[9]。先行研究では 3 名の女性声優がそれぞれ異なる文章を読み上げたボイスサンプルを用いていたが，今回，同一の文章であっても同じ傾向であったこと，また 10 人の女性声優の音声を用いてある程度共通した傾向が示されたことは，女性声優におけるある種の典型的な「少年らしさ」の表現が存在していることを示唆している。一方で，4.2 項にも記載する通り，本来，少女の声と少年の声には大きな差はないため，これは日本のアニメにおいて特徴的に発達してきた表現である可能性が考えられる[8]。

4.2 意図したキャラクターと音響特徴量との関係

意図した 5 キャラクターごとの音響特徴量（表-2）の違いをみると，まず女性 4 キャラクター（幼い少女・中高生程度の少女，大人の女性，老婆）については，実際の加齢変化との関連が見て取れる。たとえば，基本周波数は声帯の長さと，スペクトル重心は声道の長さと身体的には関連しており，年齢を重ねて身体が大きくなるにつれて，いずれも値が低くなっていくことが知られているが[2,3,4,5,6,7]，老婆のスペクトル重心を除いて，整合的な音響特徴量の変化が示されている。また話速については，幼い子供や老人に比べて，大人の方が速いことが報告されているが[41]，これも整合的な音響特徴量が得られている。

少年声については今回，「幼い少女と同年代」を想定した演技を依頼した。すなわち，声変わり前の少年声を演じることを求めた。身体的特徴と関連する基本周波数およびスペクトル重心を考えると，成人の場合にはいずれも男性の方が低い一方で，声変わり前の男女では大きな差がないことが報告されている[2,3,4,5,6,7]。しかしながら，幼い少女，中高生程度の少女，少年で基本周波数とスペクトル重心を比較すると，基本周波数は少年が最も低い一方で，スペクトル重心については幼い少女と中高生程度の少女の間に少年が位置している。すなわち，基本周波数とスペクトル重心に着目すると，本来は特徴量に大きな差がない声変わり前の男女を演じ分ける際に，通常は身体と関連して相関した変化をするこれらの特徴量が，基本周波数は低くなる一方でスペクトル重心は高くなるという通常と異なる変化をしていることが見て取れる。少年らしさを表現する際に，単にどちらの特徴量も低くするわけではなく，基本周波数は下げつつスペクトル重心を上げることで「幼い少女ではないが単に年齢を重ねた女性ではない」声質を実現している可能性が考えられる。

スペクトル重心が高いと，同じ基本周波数でも知覚的な声の高さが高く感じられることが報告されており[5]，そのような錯覚を活用することで，実際には差がない声変わり前の男女の声の区別をつけている可能性も考えられる。女性声優の少年声は日本のアニメで独特に発展してきたものであり[8]，「なぜこの声が少年に聞こえるのか」を深掘りすることで，ヒトの音声知覚やアニメを始めとしたメディア芸術における音声表現に関する理解を深めることができるかもしれない。

4.3 演技音声の主観的印象と音響特徴量との関係

本項は主観評価結果の主成分得点と音響特徴量との関係についての考察であるため，4.2 項が「意図したと考えられるキャラクターごとの音響特徴の違い」であっ

たのに対して，「実際の聴取時の主観的な印象を決める特徴量」に関する考察となる。すなわち声優が，どのような特徴量を操作することで主観的な聴取印象を適切にコントロールしているかに関する考察である。

まず，第 1 主成分の年齢については，音響特徴量として基本周波数の低さや基本周波数の変化の小ささ（すなわち抑揚の小ささ），話速の遅さと関連していた。基本周波数や話速は 4.2 項にも記載した通り，実際の加齢変化とある程度整合的だと考えられる[2,3,4,41]。また，話速が遅く抑揚が小さいと「落ち着きのある声」として知覚されることが報告されており[42]，落ち着きによって加齢を表現している可能性が考えられる。

第 2 主成分と第 3 主成分は「大人の女性らしさ」と「少年らしさ」をそれぞれ表現していると考えられ，これらの間で標準化偏回帰係数が逆になっている特徴量として基本周波数，スペクトル傾斜，平均パワーが挙げられる。基本周波数は女性の方が男性より高く[2,3,4]，スペクトル傾斜も女性の方が男性より大きいことが報告されており[43,44]，少年における「男性らしさ」の表現として整合的だと考えられ，1 人の女性声優の女性役と男性役の基本周波数の差異について検討した先行研究とも整合的である[10]。また，パワーが大きいと「迫力のある声」として知覚されることが報告されており[42]，「迫力のある声」は「男性的な声」として評価される傾向も示されている[9]。スペクトル傾斜と関連する気息性は「息漏れがあり柔らかい感じ」であり，スペクトル傾斜が低い声はその逆として「迫力/張りのある声」として知覚される可能性も考えられる。このように複数の特徴量を適切にコントロールすることで，主観的な印象としての声の「女性らしさ」と「男性らしさ」を表現し分けている可能性が示された。

4.4 本研究の限界と今後の展望

本研究では女性声優の演技音声における年齢・性別の表現のみを対象としたが，実際にキャラクターを演じる際には，これらの要素だけでは不十分である。たとえば，キャラクターの性格も適切に表現する必要がある。先行研究において，声質は発話者のパーソナリティの判断に影響を与えることが知られている[45,46,47]。音響特徴を操作した実験では，発話速度[48,49,50,51]，声の高さ[52]，抑揚やイントネーション[51,53,54]，母音の明瞭さ[55]といった声質と発話者の性格印象との関係が報告されている。関連して社会的印象についても，発話の韻律的特徴によって変化することが知られている[56,57]。音響特徴量の操作ではなく人間による演じ分けの場合でも，同一声優による異なる性格を持つキャラクターの演技音声を分析した研究や[16]，メイド喫茶における「ツン」と「萌え」という異なるタイプ（役割）を演じ分けた音声間で韻律的特徴が異なることを示した研究[14]，アニメーションにおける善玉と悪玉ではステレオタイプ的な声質が異なっていることを報告した研究[17,18]，アニメ映画である『ズートピア』を対象に，異なるパーソナリティのキャラクター間で聴取印象や音響特徴量を分析した研究[12]などが存在している。キャラクターの声のステレオタイプ性と音響特徴量との関係について調べた研究もあり[13]，声のステレオタイプの 3 次元モデルについて提案した論文などを踏まえると[58]，「与えたいキャラクター印象を与える声」の表現についてより理解を深めるには，さらなる検討が必要である。

また，顔と声との認知的処理には様々な共通点があることや[59]，顔だけを見て「その人の声がどれか」を一定程度は当てることができること[60]，顔と声を関連づけるには言語内容よりも韻律のような話し方が重要であることを踏まえると[61]，キャラクターの見た目から受ける印象と声に対して抱く印象を適切に一致させることも重要となってくる。逆に，そのようなキャラクターに対してヒトが抱く印象や，与えたい印象と声質との関係が明らかになれば，たとえばゲームキャラクタの声を自動推薦するシステムを作るような工学システムへの応用が考えられる[15]。

性格印象や社会的印象のような，ある程度キャラクターの中で固定されているものに加えて，特定のキャラクター内でも変化する要素，たとえば感情をはじめとした心身の状態に関する表現も重要である。音声に含まれる感情については多様な研究が行われてきているが[2,62,63]，中でも演技音声に着目したものとして，たとえば感情を表出した音声を対象とした研究において，自発的な発話と演技発話の間で音響特徴量が異なることが報告されており[37,64,65]，演技音声の方が自発音声よりも感情の表出が明快で大げさなものとなりやすいことが知られている[66]。他にも，嫌悪感情を意図した演技音声に関する研究や[67]，感情とはやや異なる疲労状態を表現する演技発話の特徴量に関する研究もおこなわれており[38]，“リアル”な心身の状態と演じられた心身の状態でどのように音声が異なってくるのかも含めて，適切なキャラクター表現を考える上では検討していく必要がある。

また，自発音声と演技音声の違いだけでなく，実際のアニメにおける“リアル”な演技音声と，今回の実験にあたって収録を依頼した演技音声では，表現が異なっている可能性にも留意が必要である。今回は音声収録において，1 人で，同じ文章の読み上げにおいて，異なる年齢・性別のキャラクターを演じることを求めたため，実際のアニメ収録のような掛け合いもなく，またセリフがキャラクターを表現しえない条件となる。そのため，よ

りステレオタイプな, 極端で分かりやすい典型的な演技をしている可能性が高い。実際のアニメにおける声質を調べた研究も様々に存在しているが[68,69,70,71], これらの研究が対象とした演技音声と, 今回のような演技音声では, その表現や主観的印象, 音響特徴が異なっている可能性もあり, この点についても今後比較検討を行っていく必要がある。

今後の展望としては, キャラクターや表現したい主観的印象と音響特徴量との関係が明らかになることで, 上述したような, 推薦システムなどの工学応用が考えられる[15]。あるいは, 同一の発話者が異なるキャラクターを演じた際の音響特徴量の変化が明らかになることで, 「誰かの声になる」とは別の形で, 発話者の個人性を保ったままで「表現したいキャラクター性の声になる」ような声質変換の実現に繋がるかもしれない[9,72]。また, 声質変換ではなく, たとえば「適切なキャラクターが表現できているかをフィードバックするシステム」を用いたトレーニングであったり, あるいは適切に表現できている声質をフィードバックすることで目指す表現が明確になった状態でトレーニングが可能となったりなど, 声による演技のトレーニングに活用できる可能性も考えられる。声によるキャラクター表現は, アニメに限らず, たとえば昨今隆盛を見せる VTuber（バーチャル Youtuber） でも重要であると考えられ[73], 声とキャラクターとの関係についての理解が深まることで, メディア芸術のさらなる発展に繋がることも期待できるであろう。

5. おわりに

本研究では, 同一の文章を複数の声優が異なる年齢・性別を意図して読み上げた音声を用いて, 聴取実験ならびに音響特徴量の解析を行った。その結果, 意図した 5 キャラクター（幼い少女・中高生程度の少女, 大人の女性, 老婆, 少年）が演じ分けられていたことが聴取実験から示された。また, 女性 4 キャラクターについては実際の加齢変化との関連が見て取れる音響特徴量の違いが示された。一方で少年声については「少年らしく聞こえる」ように特有の音声表現が用いられている可能性が示唆された。本研究では女性声優の演技音声における年齢・性別の表現のみを対象としており, キャラクター表現の全てを検討できているわけではないが, 今後さらなる検討が行われることで, メディア芸術におけるキャラクター表現の理解深耕や工学応用が期待できるであろう。

文　献

[1]H. Fujisaki, “Prosody, models, and spontaneous speech,” in *Computing Prosody*, Y. Sagisaka, N. Campbell and N. Higuchi, Eds. (Springer, New York, 1996), pp. 27-42.

[2]森大毅, 前川喜久雄, 粕谷英樹, 音声は何を伝えているか : 感情・パラ言語情報・個人性の音声科学（コロナ社, 東京都, 2014）

[3]栗田茂二朗,”声帯の成長, 発達と老化 : とくに層構造の加齢的変化,” 音声言語医学, **29**(2), 185-193 (1988).

[4]I. R. Titze, “Physiologic and acoustic differences between male and female voices,” *Journal of the Acoustical Society of America*, **85**(4), 1699-1707 (1989).

[5]内田照久, 森勢将雅, ”声のピッチ感の錯覚と疑似歌声・疑似ささやき声による検討,” 情報処理学会論文誌, **61**(4), 807-816 (2020).

[6]H. K. Vorperian, S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert and L. R. Gentry, “Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study,” *Journal of the Acoustical Society of America*, **125**(3), 1666-1678 (2009).

[7]H. K. Vorperian, S. Wang, E. M. Schimek, R. B. Durtschi, R. D. Kent, L. R. Gentry and M. K. Chung, “Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study,” *Journal of Speech, Language, and Hearing Research*, **54**(4), 995-1010 (2011).

[8]石田美紀, アニメと声優のメディア史 : なぜ女性が少年を演じるのか（青弓社, 東京都, 2020）

[9]林大輔, “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,” 愛知淑徳大学論集―人間情報学部篇, **9**, 49-62 (2019).

[10] 丸島歩, “女性声優による役柄の性別の異なる音声の音響的特徴：基本周波数に着目して,” 大阪経済法科大学論集, **115**, 23-33 (2020a).

[11] 丸島歩, “女性声優の演技音声にあらわれるジェンダーの表現：母音フォルマントに着目して,” 年報新人文学, **17**, 165-139 (2020b).

[12] A. Crochiquia, A. Eriksson, M. A. S. Fontes and S. Madureire, “A phonetic study of Zootopia characters' voices in Brazilian Portuguese: the role of stereotypes,” *DELTA: Documentação de estudos em lingüística teórica e aplicada*, **36**(3), 1-46 (2020).

[13] 石井沙季, 伊藤克亘, “キャラクター音声のステレオタイプ識別のための音響分析,” 情報処理学会第 81 回全国大会講演論文集, **4**, 695-696 (2019).

[14] S. Kawahara, “The prosodic features of the "moe" and "tsun" voices,” *Journal of the Phonetic Society of Japan*, **20**(2), 102-110 (2016).

[15] 酒井えりか, 伊藤彰教, 伊藤貴之, “ゲームキャラクタと声質の傾向分析,” 映像情報メディア学会技術報告, **40**(11), 123-124 (2016).

[16] 佐藤茉奈花, “同一声優による異なる性格を持つキャラクターの演技音声の分析,” 社会言語科学会第 47 回大会発表論文集, **2-3**, 29-32 (2023).

[17] M. Teshigawara, “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 (2004).

[18] 勅使河原三保子, 伊藤克亘, 武田一哉, “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 (2005).

[19] 日本声優統計学会, “声優統計コーパス,” https://voice-statistics.github.io/ (参照 2025-05-14).

[20] R. Sonobe, S. Takamichi and H. Saruwatari, “JSUT corpus: free large-scale Japanese speech corpus for end-to-end speech synthesis,” *arXiv preprint*, **1711.00354**, 1-4 (2017).

[21] 夢前黎, “つくよみちゃんコーパス｜声優統計コーパス（JVS コーパス準拠）,” https://tyc.rei-yumesaki.net/material/corpus/ (参照 2025-05-14).

[22] 木戸博, 粕谷英樹, “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**(6), 405-411 (1999).

[23] J. R. de Leeuw, R. A. Gilbert and B. Luchterhandt, “jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments,” *Journal of Open Source Software*, **8**(85), 5351 (2023).

[24] S. Takamichi, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji and H. Saruwatari, “JVS corpus: free Japanese multi-speaker voice corpus,” *arXiv preprint*, **1908.06248**, 1-4 (2019).

[25] 河原達也, 李晃伸, “連続音声認識ソフトウェア Julius,” 人工知能学会誌, **20**(1), 41-49 (2005).

[26] M. Morise, F. Yokomori and K. Ozawa, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” *IEICE transactions on information and systems*, **E99-D**(7), 1877-1884 (2016).

[27] M. Morise, “Harvest: A high-performance fundamental frequency estimator from speech signals,” Interspeech, pp. 2321-2325 (2017).

[28] M. Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” *Speech Communication*, **84**, 57-65 (2016).

[29] 日本音響学会(編), 新版 音響用語辞典（コロナ社, 東京都, 2003）, p. 194.

[30] 古川茂人, “聴覚の心理物理学,” 内川惠二(編), 聴覚・触覚・前庭感覚（朝倉書店, 東京都, 2008）, p. 75.

[31] S. S. Stevens and J. Volkmann, “The relation of pitch to frequency: A revised scale,” *The American Journal of Psychology*, **53**(3), 329-353 (1940).

[32] E. Schubert and J. Wolfe, “Does timbral brightness scale with frequency and spectral centroid?,” *Acta Acustica united with Acustica*, **92**(5), 820-825 (2006).

[33] 森勢将雅, 音声分析合成（コロナ社, 東京都, 2018）

[34] J. Hillenbrand, R. A. Cleveland and R. L. Erickson, “Acoustic correlates of breathy vocal quality,” *Journal of Speech, Language, and Hearing Research*, **37**(4), 769-778 (1994).

[35] J. Hillenbrand and R. A. Houde, “Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech,” *Journal of Speech, Language, and Hearing Research*, **39**, 311-321 (1996).

[36] 横森文哉, 二宮大和, 森勢将雅, 田中章浩, 小澤賢司, “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 (2016).

[37] R. Jurgens, K. Hammerschmidt and J. Fischer, “Authentic and play-acted vocal emotion expressions reveal acoustic differences,” *Frontiers in Psychology*, **2**(180), 1-11 (2011).

[38] 生野琢郎, 森勢将雅, “演技発話による疲労の表現によって生じる音色変化の分析,” 電子情報通信学会技術研究報告, **117**(393), 39-42 (2018).

[39] R Core Team, “R: A language and environment for statistical computing” [Computer software] (2017).

[40] JASP Team, “JASP (Version 0.18.3)” [Computer software] (2024).

[41] B. L. Smith, J. Wasowicz and J. Preston, “Temporal characteristics of the speech of normal elderly adults,” *Journal of Speech, Language, and Hearing Research*, **30**(4), 522-529 (1987).

[42] 木戸博, 箕輪有希子, 粕谷英樹, “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 (2002).

[43] H. M. Hanson, “Glottal characteristics of female speakers: Acoustic correlates,” *Journal of the Acoustical Society of America*, **101**(1), 466-481 (1997).

[44] D. H. Klatt and L. C. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” *Journal of the Acoustical Society of America*, **87**(2), 820-857 (1990).

[45] C. D. Aronovitch, “The voice of personality: stereotyped

judgments and their relation to voice quality and sex of speaker," *Journal of Social Psychology*, **99**(2), 207-220 (1976).

[46] K. R. Scherer, "Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception," *Journal of Personality*, **40**(2), 191-210 (1972).

[47] K. R. Scherer, "Personality inference from voice quality: the loud voice of extroversion," *European Journal of Social Psychology*, **8**(4), 467-487 (1978).

[48] 内田照久, "音声の発話速度の制御がピッチ感及び話者の性格印象に与える影響," 日本音響学会誌, **56**(6), 396-405 (2000).

[49] 内田照久, "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 (2002).

[50] 内田照久, "音声の発話速度と休止時間が話者の性格印象と自然なわかりやすさに与える影響," 教育心理学研究, **53**(1), 1-13 (2005a).

[51] 内田照久, "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 (2009).

[52] 内田照久, 中畝菜穂子, "声の高さと発話速度が話者の性格印象に与える影響," 心理学研究, **75**(5), 397-406 (2004).

[53] 内田照久, "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 (2005b).

[54] 内田照久, "未知のイントネーションから想起される話者の性格印象と方言地域の特徴," 音声研究, **10**(3), 29-42 (2006).

[55] 内田照久, "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 (2011).

[56] P. Belin, B. Boehme and P. McAleer, "The sound of trustworthiness: Acoustic-based modulation of perceived voice personality," *PLoS ONE*, **12**(10), e0185651 (2017).

[57] P. McAleer, A. Todorov and P. Belin, "How do you say 'Hello'? Personality impressions from brief novel voices," *PLoS ONE*, **9**(3), e90779 (2014).

[58] 勅使河原三保子, "声に関するステレオタイプの解明に向けて：音声に基づく人物像の知覚の 3 次元モデル," 駒澤大学外国語論集, **27**, 1-19 (2019).

[59] G. Yovel and P. Belin, "A unified coding strategy for processing faces and voices," *Trends in Cognitive Sciences*, **17**(6), 263-271 (2013).

[60] M. Kamachi, H. Hill, K. Lander and E. Vatikiotis-Bateson, "'Putting the face to the voice': Matching identity across modality," *Current Biology*, **13**, 1709-1714 (2003).

[61] K. Lander, H. Hill, M. Kamachi and E. Vatikiotis-Bateson, "It's not what you say but the way you say it: matching faces and voices," *Journal of Experimental Psychology: Human Perception and Performance*, **33**(4), 905-914 (2007).

[62] 重野純, 本心は顔より声に出る：感情表出と日本人（新曜社, 東京都, 2020）

[63] 田中章浩, 顔を聞き，声を見る：私たちの多感覚コミュニケーション（共立出版, 東京都, 2022）

[64] P. Laukka, D. Neiberg, M. Forsell, I. Karlsson and K. Elenius, "Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation," *Computer Speech and Language*, **25**(1), 84-104 (2011).

[65] C. E. Williams and K. N. Stevens, "Emotions and speech: Some acoustical correlates," *Journal of the Acoustical Society of America*, **52**(4B), 1238-1250 (1972).

[66] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, **40**(1-2), 227-256 (2003).

[67] 俣野文義, 小口純矢, 森勢将雅, "嫌悪感情を意図して発話された日本語演技音声の音響特徴量分析と話者間比較," 日本音響学会誌, **81**(1), 64-72 (2025).

[68] 原雄太郎, 伊藤克亘, "声優の発話の音響特徴量分析及び確率モデルの作成," 情報科学技術フォーラム講演論文集, **8**(2), 369-372 (2009).

[69] C.T. Ishi, A. Utsugi and I. Ota, "Voice types and voice quality in Japanese anime," *Proceedings of the 20th International Congress of Phonetic Sciences*, pp. 3632-3636 (2023).

[70] R. L. Starr, "Sweet voice: The role of voice quality in a Japanese feminine style," *Language in Society*, **44**(1), 1-34 (2015).

[71] A. Utsugi, H. Wang and I. Ota, "A voice quality analysis of Japanese anime," *Proceedings of the 19th International Congress of the Phonetic Sciences*, pp. 1853-1857 (2019).

[72] 高道慎之介, "音声アバターを選ぶ時代：ボイスチェンジャー技術の動向," 電気学会誌, **141**(2), 93-96 (2021).

[73] 松本大輝, "フィクショナル・キャラクターとしてのVTuber," 岡本健, 山野弘樹, 吉川彗(編著), VTuber 学（岩波書店, 東京都, 2024）

英文アブストラクト

Acoustical features relating to the representation of age and gender in acting voices by female voice actors.

In this study, we conducted listening experiments and analyzed acoustic features using acting voice by female voice actors who intended to represent different ages and genders. The results demonstrated that the listening impressions of the five intended characters (a young girl, a teenage girl, an adult woman, an elderly woman, and a young boy) were subjectively distinct from each other. Furthermore, regarding the acoustic features related to character expression, for the four female characters, it was observed that there was a correlation with actual age-related changes, whereas for the young boy's voice, it was indicated that there was an unique expression to represent “boy”. In the discussion section, we explore the possibility of advancing the research by targeting a broader range of character expressions, while also presenting prospects for deepening the understanding of character expression in media arts such as anime and its applications for voice engineering.