展望論文

声における重ね合わせに関する認知科学研究の可能性：量子論的世界観を踏まえた自己観と人間観の展望

林 大輔 [1,*]

[1] 日本たばこ産業株式会社

Cognitive science research about superposition in voice: Possible future views of self and humanity based on quantum perspective

Daisuke Hayashi [1,*]

[1] *Japanese Tobacco Inc.*

This paper explores the potential of cognitive science research into "superposition in voice," a concept inspired by quantum theory, to examine how voices can simultaneously embody multiple identities or interpretations. Focusing on three perspectives (self, individuality, and humanity), the paper refers to various research relating to each perspective in multiple disciplines. It shows that the advancements in human skills related to voice in pop culture such as voice actor and VTuber, as well as in engineering technologies such as speech synthesis and voice conversion, are blurring the boundaries between “self and others,” “speaker and character,” and “human and machine.” It suggests that these blurred boundaries can be viewed as "superposition in voice." Based on this point of view, the paper discusses the various questions relating to each of the three perspectives, and proposes the possible research approaches in cognitive science, such as quantum cognition and projection science. It also highlights the importance of a digital archive of voices in pop culture, as well as the ethical, legal, and social issues surrounding voice engineering. Overall, the paper advocates interdisciplinary research that bridges the gap between the humanities and engineering in order to redefine future views of the self and humanity.

Keywords: voice（声），superposition（重ね合わせ），pop culture（ポップカルチャー），voice engineering technology（音声工学技術），view of self（自己観），view of humanity（人間観）

Received xx June 2025; Accepted xx xxx 2025; Available online at J-STAGE xx xxx 2026

1. はじめに

声は日常にあふれており，私たちは当たり前のように声に触れ，声を用いて日々を過ごしている．私たちは，「私の声」というものを持っていると感じ，声を聞いた時にそれが「あの人の声である」と認識できていると思い，声でコミュニケーションしている相手が「誰であるのか」を理解しているように思われる．しかし，声は曖昧性を内包しており，自己と他者，話者とキャラクター，あるいは人間と機械といった，時には一見相反するもの同士の“間”をゆらぎうる．それらのゆらぎは，ポップカルチャーにおける人間の声の技術，あるいは音声合成や声質変換などの音声工学技術によって様々な形で生じている．そのようなゆらぎの状態を，量子論的世界観に基づいて，複数の解釈の「重ね合わせ」状態として考えることで，量子的な状態の「重ね合わせ」と関連付けて検討でき，認知科学の観点からそのような状態を表現し，説明するような研究を行える可能性がある（布山・西郷, 2022）．

本稿では，声における自己と他者の“間”と関わる「私らしさ」，話者とキャラクターの“間”と関わる「その人らしさ」，そして人間と機械の“間”と関わる「人間らしさ」という 3 つの観点について，声を題材にこれまでに行われてきた研究を概観しつつ，そ

* 責任著者．E-mail: daisuke.hayashi@jt.com

れぞれ声における重ね合わせがどのように生じているのか，その点についてどのような論考があるのかについて検討を行う．その上で，今後どのような研究の方向性が考えられるかについて展望を述べ，量子論的世界観に基づいた声に関する学際的な研究により，新たな「自己観」や「人間観」がどのように提起されうるのかについて議論を行う．

2. 声における「私らしさ」

私らしさ，すなわち自己に関する認知は大きな学問上の問いであり，これまでに様々な研究が行われてきている（嶋田, 2019; 田中, 2023）．私たちは声を発する時，自分の声を常に聞いているため，声は自己を規定する重要な 1 つの要因たりうる．本章では，声と自己との関係についての研究を概観した上で，声において自己と他者がいかに重なりうるのかについて検討を行う．

2.1 「私らしさ」の認知

声は呼気が声帯を振動させ，その振動が声道を通って口から外に出て，空気を振動させながら伝わり，聞き手の耳に届いて「声」として知覚されるものである．そのように，口から出て空気を伝わって耳に入ってくる音を気導音と呼ぶが，声を発した本人である話者が聞くのは気導音だけではない．発声時の声帯振動が頭蓋骨を通じて，直接内耳に伝わってきた音も話者は聞いており，そのような音は骨導音と呼ばれる（日本音響学会, 1996）．そして気導音と骨導音が合わさった「声」が，普段話者自身が聞いている声（自己聴取音声）であり，これが通常，私たちにとっての「自分の声」である．

骨導音があることで，自己聴取音声は話者以外の聞き手が聞いている声とは異なっており，録音した自身の声を聞くといつもと違って聞こえるため「自分の声じゃないように感じる」ことになる．そのようないつもと違う録音音声について，日常的な感覚として「好きではない」という話を聞くことは多く，実際，先行研究では自身の録音音声に対してネガティブな反応を示すことが報告されている（Daryadar & Raghibi, 2015; Holzman & Rousey, 1966）．一方で，自分の声は他者の声に比べて高く評価するという研究も存在している（Hughes & Harrison, 2013）．また，自分の声だけでなく，自分に似た声も評価が高いとも報告されており（Peng et al., 2019），自分の声に似たアバターから教わる方が，メタバース上での教育ゲームの効果が高いという研究も存在する（Kao et al., 2021）．このように話者自身の録音音声に対する知見は多岐にわたっており，個人特性と録音音声に対する評価との関係についても検討が行われている（Lundh et al., 2002; Yanagida et al., 2023）．

録音音声ではなく，骨導音込みの自己聴取音声に着目した研究も行われている．自己聴取音声は，気導音に比べて低周波数ほど強く，高周波数ほど弱く聞こえると言われているが（日本音響学会, 1996），様々なフィルタを用いて録音音声を変換し，自己聴取音声との類似度を評価した研究では，どのフィルタが最も自己聴取音声と類似した音声を生成できるかについては個人差が大きいことが報告されている（Kimura & Yotsumoto, 2018）．また関連して，録音音声を複数のフィルタで変換した音声を用いて，自己聴取音声と最も類似した音声を聞いている時と最も異なる音声を聞いている時の脳活動を fMRI を用いて比較し，「自分の声」の処理と関連する脳領域を調べた研究も行われている（Hosaka et al., 2021）．近年では，自己聴取音声を工学的に生成することで，「自分の声」に講義をしてもらった際の課題成績を調べた研究や（福田他, 2025），「自分の声」を用いて英語の発音学習を行った研究など（山中他, 2025），自己聴取音声の工学応用に関する研究も行われている．

では，そのような声における「私らしさ」は変容しうるのだろうか．この点について，Gallagher（2000）におけるミニマル・セルフとナラティブ・セルフそれぞれと関連すると思われる研究が存在している．ミニマル・セルフは，記憶などを失ってもなお残る「私」という感覚であり，行為主体感と身体所有感からなる．このうち，「この声は私が発したものである」という感覚である行為主体感と関連して，声の高さを工学的に操作してフィードバックすると，声に対する行為主体感が弱くなることが報告されている（Ohata et al., 2022）．また，本来は自分で発した声であるにも関わらず，声に対する行為主体感が失われることで，発話主が誤帰属されて他者の声として感じられ，それが幻聴として報告されるという幻聴の自己モニタリング仮説も提案されている（浅井・丹野, 2010）．さらに昨今では，「自身の声を他者の声に変換してリアルタイムでフィードバックする」声質変換技術が発展しており（Sisman et al., 2021），発話に対して他者の声をフィードバックすると，他者の声を自分の声のように感じる，すなわち他者の声に対す

る所有感が生じる可能性が示唆されている（中川・田中, 2021）．そのような“錯覚”を活用して，フィードバック音声をよく知られているキャラクターの声に変換した際の認知課題の成績を検討した研究や（Kunimi et al., 2024），遠隔操作でロボット接客をする際に，ロボットらしい声をフィードバックすることで，「ロボットになりきりやすくなる」ことを示した研究などが存在している（Ogawa et al., 2024）．

これらの研究では，変換した音声のリアルタイムフィードバックを行うことで，行為主体感を含めたミニマル・セルフに働きかけつつ，「私らしさ」という感覚をも変容させている，すなわちナラティブ・セルフにも影響している可能性が考えられる．ナラティブ・セルフは，アイデンティティとしての自己でもあり，自伝的記憶などを含めた「私らしさ」という感覚である．より長期的に声を変換して日常を過ごしている方々を対象とした研究として，國見他（2024）では，ソーシャル VR プラットフォームの 1 つである VRChat において，声質変換技術あるいは発声の工夫を用いて「身体的な性別とは異なる性別の声」を日常的に発しているユーザを対象にアンケート調査を行っている．その結果，声を変えて話した際に「別人になったような感覚」があること，一方でそのような場合でも「現実世界での自分と繋がりがある」とは感じることが示された．そのように，ソーシャル VR 上で理想のかわいさを表現するために，声もかわいいものとなるように努力している方々が存在しており（Bredikhina, 2022），長期的に声を変えて過ごすことで「私らしさ」という感覚も変容しうることが示唆されている．そのようなユーザの 1 人であるバーチャル美少女ねむは，「自分の声が変わった瞬間，ついさっきまで普通に仕事をしていたはずの自分の心がくるりと裏返って，美少女としての私が目覚めます…ボイチェンは，私の心の美少女スイッチなのです（バーチャル美少女ねむ, 2022, p.179）」と述べており，声を変えることは，ミニマル・セルフだけでなくナラティブ・セルフにも影響しうることが示唆されている．

2.2. 自己と他者の重ね合わせ

前節において概観した，声における「私らしさ」の認知に関する研究を踏まえつつ，本節では，声において「自己」と「他者」の重ね合わせが生じている可能性について検討を行う．

前節のおわりに述べたように，ソーシャル VR 上で声を変えて話しているユーザは，「別人になったような感覚」を抱きつつ，同時に現実の自己との連続性は感じている（國見他, 2024）．「現実と VRChat の自己は別々の自己であると認識しながらも自己連続性を感じている可能性がある（國見他, 2024, p. 4）」ことは，「自己でもあるが，他者でもある」状態，すなわち自己と他者の重ね合わせ状態が生じていることを示唆している．また，國見他（2024）では身体的な性別とは異なる性別の声を用いているユーザを対象としていたが，ソーシャル VR のユーザの中には，見た目（アバター）はかわいい美少女でありつつ，声は地声のままである男性ユーザも存在している（Bredikhina, 2022）．これは見た目と声が不一致な状態であり，違和感を抱きうるとも考えられるが，「現在のソーシャル VR では『女性アバターから男性の声が出る』ようなパターンが非常に多く，それもだんだんと当たり前のこととして受け入れられるようになってきています（バーチャル美少女ねむ, 2022, p.181）」との言及もあり，見た目の性別と声の性別が異なる状態は「女性であり，男性でもある」という重ね合わせ状態として，そのまま受け入れられている可能性が考えられる．

ソーシャル VR 上でのこれらの重ね合わせは，技術やサービスの発展に伴って昨今生じているものだと考えられるが，以前から，「声を用いて別人になる」ことは行われており，その顕著な例がアニメにおける声優だと考えられる．声優は声でキャラクターを演じ，表現するが，たとえば千葉繁は「役に自分が寄るシステムと，自分に役を引きつけるシステム，ふたつの演技システムがあるんです．僕はいつも後者で演技をやっています．こうなると見た目のキャラクターが違うだけで，どの役も本質的には自分なんです（藤津, 2019, p.59）」と述べており，キャラクターという別人になりつつも，それも自分である，すなわち「キャラクターであるが，自己でもある」ような重ね合わせ状態だと捉えることができる．また，そのように「声で演じる」ことは声優に限った話ではなく，「演劇や歌では，声はキャラクターやパーソナリティーを演じるための表現の一部として，意識的に用いられる…声は仮面や化粧と同じように，素顔を隠し，自分とは違った別の時代，別の身分，別の性格をもつ役柄を演じるための道具の一つである（増野, 2014, p. 196）」．そして演じる行為を行うことは，自己ではない何者かになるというよりは，「自己の二重化，つまり『私でありながら，私でないもの（たとえば役

柄）でもある』状態（増野, 2014, p.205）」であると述べられており，「自己であり，他者でもある」状態を生じさせることこそが，演じることだとも捉えることができる. 認知神経科学研究において, 物語の登場人物になり切っている時には自分自身について考えるのと同じ部位が活動することが示されており（Broom et al., 2021），一方で“fictional”な一人称視点でいる時, つまり何かの役になっている時には, 全体として脳活動が抑えられ，自己という感覚が弱まる可能性を示唆する知見も存在することから（Brown et al., 2019），演じている時の自己は，自己でありつつ自己ではない状態だと考えることができそうである.

そして声で演じることは，実は私たちにとって特別なものではなく，置かれている場面や自身の役割に合わせて意識的あるいは無意識的に声質を操作することで，望ましい印象を与えることは日常的に行われている（モクタリ, 2008）. たとえば，電話で話す時と対面で話す時で声質が違っていたり，初対面の相手と話す時と親しい相手と話す時で声質が変わったりすることは珍しくない. また, 百貨店の売り子，電車の車内アナウンスなど，職業に応じてそれに合う声質，発声が存在しているようにも見受けられる（Grawe, 2018）. このように私たちは，仕事などの社会的役割に応じて発せられる「自発型演技音声」を使っており（宮島他, 2013），「たとえば商売人の声も，その人自身の声でありながら，その人のふだんの生活の声とは別のものでもある. 限られた文脈の中で，自分の声を変形させることで，商売や職業に適したパーソナリティーを確立する（増野, 2014, p.205）」との言及にもある通り, 私たちは声を“変える”ことで，自己と「演じるべき役」を日常的に重ね合わせているのかもしれない.

3. 声における「その人らしさ」

私たちは声を聞いた時に, 話者の年齢や性別, 性格印象などの「どのような人なのか」に関する情報（話者情報）を知ることができる. また，個人の同定と関連するような, より個々の話者に特有の「誰なのか」に関する情報（個人性情報）についても知ることができる. それらの情報と，声道などの生理学的特徴，その特徴に起因する音声の音響特徴，そしてヒトの知覚的な特性との関係について，様々な研究が行われてきている（本多, 2018; 森他, 2014; Schweinberger et al., 2014）. そしてこれらの話者情報と個人性情報を統合して得られる認知が, 声における「その人らしさ」の認知だと考えられる. 本章では, 声とその人らしさとの関係についての研究を概観した上で，声において話者とキャラクターがいかに重なりうるのかについて検討を行う.

3.1 「その人らしさ」の認知

話者情報に含まれる年齢や性別については，加齢に伴う成長や性差に基づく体格差に依存して，声帯や声道などの調音器官に変化が生じ，そのことが音響特徴に，ひいては知覚に影響する. たとえば, 子供よりも大人の方が，あるいは女性よりも男性の方が，平均的に体が大きいため, 声帯も声道も長く, 結果として基本周波数とスペクトル重心が低くなる（栗田, 1988; 森他，2014; Titze, 1989; 内田・森勢，2020; Vorperian et al., 2009; Vorperian et al., 2011）. 生理学的特徴や音響特徴だけでなく, 音声を聞いた際, 話者の年齢や体格などの情報をある程度正確に知覚できることを示した研究や（Krauss et al., 2002），声の性別に対する順応と残効が生じることを示した研究が存在しており（Schweinberger et al., 2008），話者情報が認知的に処理されていることが示されている.

声を聞いた際に話者に対して抱く印象も話者情報の一部だと考えると, 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（森他, 2014; 日本音響学会, 2003, p. 194））は話者の性格印象に影響を与えることが，様々な研究において示されており（Aronovitch, 1976; Belin et al., 2017; McAleer et al., 2014; Scherer, 1972, 1978; 内田，2000, 2002, 2005a, 2005b, 2006, 2009, 2011; 内田・中畝, 2004），声に基づいた人物像の知覚の 3 次元モデルなども提案されている（勅使河原, 2019）. また, 声質の印象についても，日常表現語を用いた記述が行われている（木戸・粕谷, 1999）.

次に，名状しがたい「この人の声」という感覚と関わる個人性情報については，生理学的特徴の個人差に着目した検討が行われている. たとえば声道の形状の個人差と関連する音響特徴を検討し，高周波成分における違いが個人の同定の手がかりとなることが示されている（本多, 2018; Kitamura et al., 2005）. 一方で，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も指摘されている（Van Lancker, Kreiman, & Emmorey, 1985）. 認知的な処理としては，声の個人性に対する順応と残効が生じることを示した研究が存在している（Latinus & Belin, 2011）.

これらの話者情報と個人性情報は完全に独立な関

係ではない. たとえば, 記憶課題においてターゲット音声を同定する際，音声の特徴を自由記述で言語化する, すなわち話者情報を記述すると, ターゲット音声と似た異なる音声を「ターゲット音声である」と誤って回答することが増えてしまう（井上, 2022）. あるいは，年齢や性別などの話者情報が違って聞こえる複数の音声の中から同一の話者の音声を正しく弁別することは，話者情報の違いを弁別するよりも難しいことが示唆されている（林, 2019）. これらの研究は, 話者情報を無視して個人性情報のみに基づいて「誰の声か」を認識することの難しさを示しており, 話者情報と個人性情報が合わさって「その人らしさ」の認知が形作られていることの傍証だと考えられる.

そしてアニメを始めとしたポップカルチャーの中では，話者情報を人間の声の技術によって適切に操作することで, キャラクター表現がなされている. そのようなキャラクター演技音声を対象とした研究としては，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（佐藤, 2023），同一声優の異なる年齢や性別を演じた音声を比較した研究（林, 2019; 林・森勢, 2025; 丸島, 2020a, 2020b），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（石井・伊藤, 2019），善玉と悪玉の声質の違いを調べた研究（Teshigawara, 2004; 勅使河原他, 2005），実際のアニメ映画における異なる性格のキャラクターの声を比較した研究（Crochiquia et al., 2020）などが存在している. あるいは，メイド喫茶の「ツン」と「萌え」という異なるタイプのメイドそれぞれがどのような発話をしているかを調べる研究も行われている（Kawahara, 2016）. このような人間の声の技術による話者情報の操作は, ポップカルチャーにおけるキャラクター表現の重要な要素だと考えられる.

そして話者情報を操作してキャラクターを表現していても, 話者が同一の個人であれば, 個人性情報, すなわち「誰であるのか」は同一であり, その情報は声に埋め込まれていると考えられる. すなわち,「アニメーションにおいてキャラクターが発する音声は, 現実に存在する声優によって発せられた音声でもある（稲岡, 2016, p. 199）」のだが,「キャラクターの声は『役者の声』として視聴者に届いてはいない…役者の身体から発しつつも,『図像』の力によって，役者の身体から切り離されて視聴者に響いているのがキャラクターの声である（藤津, 2018, pp. 112-113）」. 声優は,「画面内のキャラクターの動きや演出に合わせて, 声の高低や濃淡などを変えたりする技術や, 感情を音声に乗せる技術など，細かな基礎的技術の集積（稲岡, 2016, p. 120）」により，キャラクターの音声が現実に存在する声優によって発せられた音声であることを「鑑賞者に意識させずにキャラクターによって発せられた音声として鑑賞者に認識させるという特質（稲岡, 2016, p. 199）」を持っている. このことは, 声優自身の語りとしても表れており, たとえば声優の久川綾は「このキャラを久川綾がやっているんだと意外に思われたい（藤津, 2017, p.238）」と述べ, また声優の森川智之も「僕の声優観というのは, 自分があまり出てはいけないというものです…僕のファンの方が僕目的で映画館に入ったとしても，それをすっかり忘れて作品を楽しめるように演じたいと思っています（藤津, 2019, pp. 177-178）」と述べている. このように, 本来は個人性情報が同一である中で, それを意識させない，すなわち身体を伴った人間の声ではなくキャラクターの声として聞こえるように話者情報を適切に操作することが，ポップカルチャーにおける人間の声の技術だと考えられる.

3.2 話者とキャラクターの重ね合わせ

前節において概観した, 声における「その人らしさ」の認知に関する研究を踏まえつつ, 本節では, 特にポップカルチャーにおける人間の声の技術に着目して, 声において「話者」と「キャラクター」の重ね合わせが聞き手の中で生じている可能性について，複数の状況に分けて検討を行う.

3.2.1 アニメにおける声優の声

まず始めに, 前節でも言及した, アニメにおける声優の声に着目する. 前節で記載した通り, 声優のキャラクター演技音声は, 声優自身の声でありつつ, それを意識させず，キャラクターの声として聞こえるように技術が用いられている. しかし,「アニメーションにおいてキャラクターによって発せられたとみなすことのできる音声は，現実に存在する声優と作品内に存在するキャラクターという二重の発生源を有する（稲岡, 2016, p. 122）」以上，キャラクターの声としてしか聞こえないとは言えず，そこに身体を持った人間としての声優を感じることは避けがたい.「現実に私たちがアニメーションを鑑賞する仕方を思い起こすと,『声優の声は声優の声であることを隠す』という点が常に重視されているとは必ずしも断言できない…作品を鑑賞する際にも，画面に連動して聴

こえてくる音声がどの声優によって発せられたものかを十分に承知した上で鑑賞するという場合も珍しくはない（稲岡, 2016, p. 121）」のが実際であり，「聴取者たちは，一方においてはキャラクターの同一性と物語世界を成立させる声として声優の声を聞いており，他方においては，声優が各作品を横断して複数のキャラクターと世界を成立させている単一の声であると了解している（黒嵜, 2018, p. 190）」．イコン・インデックスという話者情報・個人性情報とそれぞれ関連が深い概念を導入した論考（ジン, 2010）を踏まえて，「視聴者は特定の声を通じてキャラクターの『人格・のようなもの』を了解すると同時に，声の『インデックス』性に基づく声優の身体を思い浮かべる（鈴木, 2014, p.112）」とも述べられている．そしてそれは，「アニメーションにおいてキャラクターが発する音声を，キャラクターが発したものと受け取るか，声優が発したものと受け取るか，概念上は区別できても，それを聴覚的印象の水準で区別することはきわめて困難である（稲岡, 2016, p. 123）」ことに起因しているとも考えられ，その結果として「テレビアニメを観るとき…実際の発話者から声が発せられているとも，キャラクターから声が発せられているとも，どちらにも確信を得ていな（黒嵜, 2018, p. 194）」い状態が生まれる．すなわち，アニメにおける声優の音声は「話者の声でもあり，キャラクターの声でもある」声だと考えることができ，話者とキャラクターが重ね合わせ状態で存在しているととらえることが可能である．そしてだからこそ，「キャラクターと声優を同一視する――キャラクターを観るときに声優のイメージをそのキャラクターに重ね，逆に声優を観るときにキャラクターのイメージをその声優に重ねる――という認識の基盤が，少なくともアニメ声優に関心を寄せるファンの間ですでに共有されている（程, 2023）」という事態が生じると考えられる．

また，日本のアニメでは女性声優が少年キャラクターを演じることが多く（石田, 2020），そのような状況では，話者とキャラクターの重ね合わせは，男性と女性との重ね合わせでもある．すなわち，女性声優の少年声は「女性でもあり，男性でもある」声としてとらえることができると考えられる．

3.2.2 二・五次元文化における声

前項で言及したアニメにおける声優の声は，基本的に声優自身の身体は見えない状態での声であり，「声優の声でもありキャラクターの声でもある」ような重ね合わせ状態が，比較的起こりやすい状況だと考えられる．では，声優の身体が聞き手に見えている状態で，声優がキャラクターとして発した音声は，アニメの場合と同様に重ね合わせ状態を生じうるのであろうか．

声優について，「その技能が専業化していく中でキャラクター文化と結びつき，メディアミックスハブとしての役割を声優が担っていくことになる（永田, 2024）」と指摘されており，現代の声優の仕事は「キャラクターを様々なメディアで演じること」となっている．様々なメディアとは，ラジオやイベント登壇，メディア取材，ゲーム，派生作品なども存在するが，その 1 つとして，声優が「キャラクターとして」舞台に立ち，キャラクター名義の曲を歌ってライブを行うような，声優/キャラ・ライブコンサートが存在している．そのようなコンサートは，「マンガ，アニメ，ゲームなどの視覚情報が優先される二次元の虚構世界と，身体性を伴う経験を共有する三次元の“現実”世界のあいまいな境界でおこなわれる文化実践を，『二・五次元文化』として定義する（須川, 2021, p. 16）」ならば，二・五次元文化の 1 つとしてとらえることができる．ここでの二・五次元は「二次元の虚構キャラクターが，三次元の現実で具現化し，虚構性を限りなく内包した人間の生身の身体，いわば『虚構的身体（”virtual corporality”）』と呼びうるものによって実践されること（須川, 2021, p. 15）」であり，声優/キャラ・ライブコンサートはまさにその好例だと考えられる．そして声優/キャラ・ライブコンサートにおいて，「声優は『キャラ』の模倣ではなく，『キャラ』そのものの生成として，オーディエンスによって経験される（川村, 2024, p. 107）」．そしてその際に，「オーディエンスの情動を最も強く触発する記号性は，いうまでもなく声」であり，「『キャラ』の分配的領域を構成する声が声優によって発せられることで，『キャラ』と声優を不可分のものとみなすリアリティが横断個体的に経験される（川村, 2024, p. 107）」ことになる．これは「声を媒介にした声優の身体とキャラクターの一体化（須川, 2021, p. 118）」であり，「登場人物と同じ声色による歌声が聞こえてきた際には，声優が歌っているのではなく，登場人物が歌っているという想像が生じることがある（シノハラ, 2017, p.86）」．そしてそのような事態の一例として，藤田茜が演じた荒木レナというキャラクターのライブについて，「舞台上で発せられた『跳ぶよ』を，荒木レナの言葉として（虚構的に）聞くか，藤田茜の言葉として聞くかは

未確定でもあり，観客に委ねられている（シノハラ, 2017, p.93）」と述べられている．すなわち，声優自身の身体が見えない状態に限らず，声優の身体が聞き手に見えている状態であったとしても，声優がキャラクターとして発した音声は「話者の声でもあり，キャラクターの声でもある」声であり，話者とキャラクターが重ね合わせ状態で存在していると考えることができる．なお，視聴者は役者とキャラクターをごちゃ混ぜにする傾向があることも示されており（Goldstein & Filipe, 2018），このような重ね合わせ状態は声優に限らず，二・五次元一般，あるいは「キャラクターを演じながら声を発している人」を見る状況一般で生じうるものであるかもしれない．

3.2.3 VTuber の声

本項では，前項までの内容を踏まえつつ，昨今隆盛を見せるバーチャル YouTuber（VTuber）の声について検討を行う．VTuber も「二・五次元文化の一例（須川, 2021, p. 15）」として捉えることが可能である．しかし，前項で述べた声優/キャラ・ライブコンサートのような状況とは異なり，VTuber は動画サイト上での配信であってもライブであっても，基本的に「中の人（配信者・演者）」の物理的身体が聞き手の目にさらされることはなく，「パーソンの身体は後ろに隠れ，ペルソナの身体だけが鑑賞者の前に現れている（難波, 2018, p. 119）」（もちろん，例外はあるものの）．そのため，前項の二・五次元文化における声とは項を分けて考えていく．ここでの問いは，「鑑賞される VTuber とは誰だろうか．たとえば，VTuber が笑ったというとき，誰が笑っているのだろうか．VTuber が愛らしいというとき，誰が愛らしいのだろうか（難波, 2018, p.117）．」というもの，すなわち VTuber が声を出すとき，誰が声を出しているのか，あるいは誰の声として認知されているのであろうか，というものである．

VTuber のあり方は非常に多様であり，一意に定めて検討することは難しい．この点は，「いっぽうで，ある特定の VTuber は，その外見から一般に想像しうるような声や性格を逸脱し，かつ，そのパーソンの経験や考えをあからさまに語っているようにみえる．たほうで，動画内にとどまらず，SNS においておロールプレイに徹しており，あくまでもキャラクタとして鑑賞者に鑑賞される VTuber もおり（難波, 2018, p. 120）」などと指摘されている．ただし，ここまでに主として述べたアニメとの比較を考えると，アニメにおける声優の演技はあくまで「作品世界におけるキャラクターを演じる」ものであるのに対して，多くの場合で VTuber は，作品世界の中だけで生きているのではなく，視聴者の生きるこの物理世界と何らかの意味で地続きの世界で生きているように思われる．それはたとえば「月ノ美兎は水を飲む」と題した論考が存在することや（新, 2018），VTuber の存在論に関する論考の冒頭に「月ノ美兎は，カフェオレを飲んでいた（篠崎, 2024, p. 271）」「紫咲シオンが，とある深夜配信中にお腹が空いてカレーメシを食べ始めた出来事を考えてみよう（本間, 2024, p. 323）」のような文が存在することからも示唆されている．この点を踏まえて，「VTuber は生身の人間である中の人（配信者・演者）と同一の存在である」と考えることも可能である（篠崎, 2024）．その場合，VTuber の声については「もはやこれらのキャラクターが喋るとき，その向こうに本当の発声者がいることを，誰もが疑っていない」「彼ら彼女らの声は…あくまでもそれぞれ「個人」の身体に固定された声であるかのように振る舞っている（黒嵜, 2018, p. 195）」ととらえることができるだろう．「実質的には演者は顔を出していないのにも拘わらず，キャラクターの顔によって「声」を発している存在が視聴者に明確に見える状態だと思ってもらえる（皇牙, 2018, p. 65）」という記述も合わせて考えると，「VTuber の声は，中の人（配信者・演者）の声である」と考えることも可能だと考えられる．このような状況は，ある意味で VTuber というキャラクターの声と，中の人（配信者・演者）という話者の声が重ね合わせ状態にあるととらえることもできるかもしれない．

他方で，「VTuber は生身の人間である中の人（配信者・演者）と同一の存在である」とは必ずしも言い切れない．視聴者は中の人（配信者・演者）と VTuber を分けて見ており（Lu et al., 2021），少なくとも視聴者，すなわち声の聞き手にとっては「演者とキャラクターは別物だという考えが主流（皇牙, 2018, p. 65）」だとも考えられている．「パーソンそのものは，しかし，直接わたしたちとコミュニケーションをとっているわけではない（難波, 2018, p.118）」のである．この点について，アニメにおける声優とも関連付けて検討できる考え方として，「技術的所産」として VTuber をとらえた場合に，「現在の典型的な VTuber の場合，キャラクターの想像に対して配信者のなす貢献が極めて大きい…その話し声や歌声は基本的に配信者の貢献分である（松本, 2024, p.318）」と考える

ことができる．この場合の VTuber の声は，3.2.1 項で述べたアニメにおける声優の声と同じような形で，話者とキャラクターの重ね合わせ状態として理解できるかもしれない．ただし前述したとおり，アニメにおける声優の声と VTuber の声は「作品世界におけるキャラクターの声か否か」という違いが一般には存在するため，3.2.1 項の議論をそのまま当てはめることができるかについては留意が必要である．

いずれにしても，「『声』は，音声コミュニケーションがメインになるメタバースの世界においても VTuber として配信するにあたっても，アイデンティティの印象を左右する決定的な要素（関根, 2024, p.186）」であり，「私たちは，配信者の声だけが吹き込まれている実写動画に対しても，それを VTuber の声として鑑賞することが可能である（山野, 2024, p. 83）」ことを考えても，VTuber という存在において声は 1 つの重要な要素である．そのため，VTuber の声における重ね合わせを検討することは，声における重ね合わせ状態の理解にとって有用な 1 つの観点であると考えられる．

4. 声における「人間らしさ」

我々は声を聞いて，その背後に人間を想像する．「人間は声の向こうに誰かの身体を，そして名前をもつ一つの人格を想定せずにいられない（増野, 2014, p.217）」と語られる通り，発する「人間」と声は深く結びついたものであるが，音声合成技術の発展に伴い，人間ではなく機械が発する音声が日常的に使われるようになってきている．本章では，声と人間らしさとの関係についての研究を概観した上で，声において人間と機械がいかに重なりうるのかについて検討を行う．

4.1 「人間らしさ」の認知

声の人間らしさを定義するのは難しいものの，1 つの指標として「読み上げの自然さ」を挙げることができる（三井, 2010）．これは聞き手にとっての「言語内容の聞きやすさ」と関わると思われ，音声合成による機械の声を人間の声に近づける，というアプローチを考えた場合，まずは読み上げの自然さについて，人間と遜色ないものを目指すことが考えられる．以前は人間の肉声に比べて機械音声はこの点が劣っており，たとえば館内放送を想定した音声を用いた研究において，肉声の方が聞きやすいことが示されていた（三谷, 2014）．しかし，技術の進歩は著しく，十分なデータ量と計算量があれば，人間と遜色のない，すなわち主観評価が肉声と統計的に違わない機械音声を合成することは，すでに技術的には可能となっている（Tan et al., 2024）．

では，すでに十分に「人間らしい」機械音声が出来上がっているかといえば，依然として課題は残っている．たとえば，読み上げの自然さだけでなく，感情をうまく機械音声に乗せることも，人間らしい機械音声を実現する上では重要な課題だと考えられる（Qi et al., 2024）．また，対話場面を考えると，実際の人は親密さによって声や話し方が変わっていくため（Chiba & Ito, 2024），そのような変化を反映させることも，自然な対話音声を実現するには重要かもしれない．関連して，共感的な対話音声の生成や（Saito et al., 2022），笑い声の合成（Xin et al., 2023）など，より自然な人間の発話に近い機械音声の合成に向けた研究が進められている．朗読的な発話よりも自発的な発話をする機械音声で話すエージェントの方が，人間との会話に近いと評価されることも示されており（Iizuka & Mori, 2022），そのような自発音声合成についても研究が行われている（Matsunaga et al., 2023）．さらに，人間相手に話す状況だけでなく，AI 同士で話している状況で，より自然な人間のように感じる対話を目指した研究も行われている（Mitsui et al., 2023; Nguyen et al., 2023）．話者が人間の場合と機械の場合で比較した研究をみると，人間の声と機械の声いずれに対しても，感情的な発話の模倣が起こることを示した研究や（Cohn et al., 2024），人間の歌声と機械音声の歌声を比較した際に，定型発達者だと人間の歌声をよりポジティブに評価する一方で，自閉スペクトラム症の方では人間の歌声と機械音声の歌声で評価に差がないことを示した研究などが存在している（Kuriki et al., 2016）．このようにまだ課題は存在しているものの，「人間らしい」機械音声については様々な研究が進展しており，将来的には人間と区別のつかない対話が可能な機械音声が合成できるようになると考えられる．

全く区別のつかない状態となる前段階では，肉声と機械音声の使い分けも重要となる．たとえば，相手が人間の場合と機械の場合で，話しかけ方や話す内容が違うことが示唆されており（山田他, 2005），適切な使い分けが必要となるだろう．また，同じ機械音声であっても，自動電話応答場面とキャラクタ音声合成サービスではユーザに求められるものが変わって

くることも報告されており，前者では明瞭性や韻律の自然性，肉声感などが重要となる一方で，後者ではそれらのある種の「人間らしさ」はあまり重要ではなく，キャラクタ性や韻律制御の自由度が重要となることが示されている（三井, 2010）．このように，人間の声と機械の声を場面に合わせて適切に使い分けながら，徐々に機械の声は社会実装されていき，そして将来的には，聴覚印象として人間の声と機械の声は区別がつかない状態になっていくと考えられる．

4.2 人間と機械の重ね合わせ

前節において概観した，声における「人間らしさ」の認知に関する研究を踏まえつつ，本節では，声において「人間」と「機械」の重ね合わせが生じる可能性について検討を行う．

すでに社会の様々な画面で機械音声は活用されており，駅などでのアナウンス音声で機械音声を聞く機会も増えてきている．また，動画サイト等では合成音声キャラクターを用いた動画も多く投稿されている．現状，これらの音声は多くの場合，人間と区別がつかない状態までは達しておらず，その意味で「人間の声でもあり，機械の声でもある」という解釈が重なった状態は生じていないと考えられる．ただしここで，重ね合わせ状態とは直接関わらないものの，機械音声においては人間の声とは異なる形で「キャラクターと声」の関係が生じていることを指摘したい．機械音声とキャラクターとの関係について，最も多く語られているのは歌声合成ソフトであるボーカロイド，その中でも初音ミクである．初音ミクの声は，声優の藤田咲の声をサンプリングしたことが明示的に知られている．しかし，初音ミクの声は，藤田咲の声として知覚されるわけではなく，「ボーカロイドの素材は声優や歌手が提供した人間の声であるが，元の持ち主の身体からは，はっきりと切り離されている（増野, 2014, pp. 216-217）」．そのように人間の声としては知覚されない初音ミクの声は，他方で絵を含めたキャラクターとは重なる存在であり，アニメを含め多くの場合に先に絵があった上でそこに声が当てられるのに対して，「初音ミクが声と絵という二つの要素から成り立つもので，とりわけ前者の存在を前提としながら後者が要請されていた（さやわか, 2015, p.26）」ことは，キャラクターと声の関係を考える上で重要である．また，初音ミクの声は確かに藤田咲の声としては認知されないものの，何らかの繋がりを持っていることは考えられ，「藤田咲さんの声が断片化されて，ちょっと違う形になって初音ミクとして出てくるわけなんですね．でも，元の声にあった『私，可愛いでしょ』という雰囲気は，ある種ランダム化されて残っている（柴, 2014, p. 105）」のような指摘が存在している．さらに，初音ミクは「人間ではない」からこそ，人間が声を発する場合以上に，性別を超えられる可能性があり，「ミクは女性で女声だが，音源を低音に操作すれば男声にすることもできる…男であり，女．女であり，男（村瀬・須川, 2014, p.91）」との言説が存在している．

いずれにせよ，ここまでの検討では，「人間の声でもあり，機械の声でもある」という解釈が重なった状態は生じていない．しかし前節で述べたように，研究レベルでは人間と遜色のない機械音声を合成することは技術的に可能となっており（Tan et al., 2024），将来的にはそのような機械音声が社会実装されると考えられる．その際，物理的な身体も伴わせようと思うと，人間と区別のつかない物理身体も生み出す必要があり，技術的なハードルは高くなる．一方で，声のみであったり，バーチャル環境であったりすれば，この点は容易にクリアできる．たとえば CoeFont の「おしゃべりひろゆきメーカー」や Cotomo の「緑川光の声」のように，タレントや声優などの有名人の声をサンプリングし，別のキャラクターではなく「その人の名前」を用いた機械音声が存在しているが，それらの音声が人間と遜色がつかない状態となった場合，事前に知らされなければ，その有名人の肉声なのか機械音声なのか区別がつかない状態が生じると考えられる．そのような声は，少なくとも聞き手にとっては「機械の声」でありながら「人間の声」でもある状態，すなわち人間と機械の重ね合わせ状態となる．また，たとえばソーシャル VR プラットフォーム上で，自律的に会話できる AI エージェントが実装された際に（倉井・平木, 2023），その音声が人間と区別がつかなければ，話し手の情報を知らない聞き手にとって，コミュニケーション相手は「機械だが，人間でもある」状態となるだろう．さらに新しい形として，人間である中の人（配信者・演者）が存在する VTuber とは異なり，背後に人間がいない AITuber が動画サイト等で配信することが行われている．AITuber はモデルの見た目だけを取り出すならば，VTuber と大きな違いはないため，その存在が「人間らしく」なる上では，モデルの動きに加えて，声の人間らしさが重要な要素になると考えられる．AITuber のある配信について，

「それまでの『AITuber』の発話は，まだ『機械』の感じが強く残っていたのですが，今回アップデートされたアイリーンさんの発話は，自然な人間の発話にかなり近いものになっていた」ことに基づき「AITuber は実在の配信者と実質的に同じように存在しうる（すなわちバーチャルに存在しうる）という可能性を感じました（山野, 2025, p. 78）」と述べた言説も存在しており，技術の進歩に伴って，聞き手にとって「機械の声であるが，人間の声でもある」という解釈の重ね合わせ状態は，より一般的に生じる事例となっていくのかもしれない．

5 今後の展望

本稿ではここまで，声における「私らしさ」「その人らしさ」「人間らしさ」という 3 つの観点から先行研究を概観した上で，それぞれの観点で「自己と他者」「話者とキャラクター」「人間と機械」の重ね合わせ状態が生じている可能性について検討してきた．それを受けて本章では，声における「私らしさ」「その人らしさ」「人間らしさ」それぞれと紐づく問いを提起しつつ，主に認知科学の観点からどのような研究の方向性が考えられるかについて議論を行う．

5.1 声における「私らしさ」に紐づく問い

声における「私らしさ」と重ね合わせについては，日常場面も含めた「声で演じる」行為との関連もありつつ，声質変換技術の進展による「自己」と「他者」の重ね合わせ状態が生じている可能性が示された．これは，声質変換技術により，新たな自己を形成していける可能性を示唆している．音声の基本周波数を操作してフィードバックすることで感情状態を誘導できることを示した研究や（Aucouturier et al., 2016），自信のある状態の声をフィードバックすることによる緊張緩和の可能性を示した研究もあり（成瀬他, 2019），自身の発した声を操作することで，話者の状態はコントロールできる可能性がある．それは「自己」という感覚でも同様であり，キャラクタ性の獲得をはじめ（倉田他, 2021），声をデザインすることで新たな自己をデザインすることが可能となっていくかもしれない（Arakawa et al., 2021; 高道, 2021）．その際，理想の自己を表現する声を用いることも可能だろうし，逆に画一的な声を使ってアイデンティティを「隠す」こともできる（Hatada et al., 2023）．「『声』は…コミュニケーションを通して私たち自身の存在を示す，音響世界における『アバター』とも言うべき，重要なアイデンティティの一つ（バーチャル美少女ねむ, 2022, p.178）」であり，技術の発展により，「自己」という感覚はある意味でゆらぎ，そして新たに形成されていくと考えられる．そのような状況において，声が変わることでゆらぐ「私」は，いかに織り上げられてナラティブとしての私になっていくのであろうか，という問いは，これからの時代において重要になってくるであろう．

そしてそのように，望む自己をデザインしていけるかもしれない可能性がある一方で，「声はヒトの身体と深く一体化した圧倒的な存在感をもつ音（増野, 2014, p.12）」でもある．すなわち，私たちは物理的な身体を有しており，それは声を変えたとしても変わるわけではない．声を変えても変わらない，あるいは変えられない，身体と紐づいたままならぬ「私」はいるのか，という問いについても，合わせて検討していくことが，声における「私らしさ」を包括的に考えていく上で，ひいては「自己とは何か」を改めて考えていく上で重要だと考えられる．

5.2 声における「その人らしさ」に紐づく問い

声における「その人らしさ」と重ね合わせについては，アニメや二・五次元文化，VTuber などのポップカルチャーにおける人間の声の技術によって，「話者」と「キャラクター」の重ね合わせ状態が聞き手の中で生じている可能性が示された．このような重ね合わせ状態は，声に対する「好き」という感覚についての問いを喚起している．声のよさや魅力については様々な研究が行われており（Weiss et al., 2021），声の高さとの関係や（Jones et al., 2010），平均すると魅力が高まるという研究（Bruckert et al., 2010），他方で平均が必ずしも魅力を高めるわけではないという研究（Ostrega et al., 2024）などが存在している．よさや魅力に関する評価の個人差についても，評価の性差と音響特徴量との関係についての研究（横森他, 2016）や，性差や年代差も含めて大規模な音声・参加者を対象に行った研究が存在している（Suda et al., 2024）．これらの研究では基本的に，ある声は「特定の話者の声」であり，そこに重ね合わせは想定されていない．しかし 3.2 節で検討したように，特にポップカルチャーにおいては声において話者とキャラクターの重ね合わせが生じているとすると，声を聞いて心惹かれて好きだと感じた場合に，私たちは「誰の」声に心惹かれているのだろうか？ポップカルチャーにおける声について，「繰り返し聴取は，声の演戯性が持つ声

そのものの魅力へと没頭させ，まさに声への『偏愛』『中毒現象』を生じさせる…声優によって発せられる声は，一方で，その実在性により架空のキャラクターの想像をヴィヴィッドなものとし，他方で，その演戯性によりキャラクターへの愛着を持たせている（シノハラ, 2017, p.78）」と指摘されているが，その「偏愛」や「愛着」のようなこだわりの対象は，声優の声なのであろうか，あるいはキャラクターの声なのであろうか．

また，3.2 節では女性声優の少年声について「女性でもあり，男性でもある」声として言及している．「声優の受容には二つの傾向が存在する．一つは，年齢と性別，さらには外見上の声優とキャラクターの乖離を認識しながら，声優の演技を楽しむ傾向である．もう一つは，年齢と性別，さらには外見上のキャラクターと声優の一致を求めながら，声優の演技を楽しむ傾向である（石田, 2020, p. 197）」との指摘を踏まえた際に，女性声優の少年声は「前者の中核に位置する（石田, 2020, p. 197）」と考えられる．男性と女性は一見相反するものであるため，このような状態が生じることは，単純な話者とキャラクターの重ね合わせ状態以上に，矛盾や曖昧性を声が内包しうることを示している．そしてそのように解釈が定まらず，複数の解釈が並列している不定性があるからこそ生じる美的体験の存在が示唆されており（布山・西郷, 2022），声における「その人らしさ」と声に対する「好き」という感覚との関係を十分に理解する上では，声が内包しうる矛盾や曖昧性，不定性があるからこそ生じるよさや魅力とはどのようなものか，という問いについても検討していく必要があるだろう．

5.3 声における「人間らしさ」に紐づく問い

声における「人間らしさ」と重ね合わせについては，音声合成技術の発展に伴って「人間」と「機械」の重ね合わせ状態が生じうることが示された．すなわち，コミュニケーションの相手が人間であるか機械であるか，少なくとも声だけでは区別がつかない状態が訪れると考えられるが，そのような状況において，私たちは「誰と」コミュニケーションを通じてつながりたいと感じるのであろうか．たとえば絵画を用いた研究では，実際にはどちらも機械が描いた絵画であっても，「人間が描いた」とラベルづけされた絵画の方をポジティブに評価することや（Bellaiche et al., 2023），人間の描いた絵画と機械の描いた絵画を区別することはできなくとも，なぜか人間の描いた絵画の方がポジティブに評価されたことが報告されている（Samo & Highhouse, 2023）．このように「人間の方がいい」と評価されるケースもある一方で，たとえばロボットを用いた研究では，日本人の若年女性だとロボットでも人間でも自己開示の程度が変わらないことが示されていたり（Uchida et al., 2020），高齢者のライフレビューでは話し相手がロボットの方がポジティブとなりうる可能性が示されていたりする（Ueda et al., 2021）．また，ボーカロイドの歌声は機械の歌声であるが，「ボカロを抵抗なく聴くリスナーも，それを声と等価なものとしてではなく，全く別物として受容している（宮本, 2017, p.36）」との指摘があり，人間と比較するのではなく，機械の声は機械の声として楽しむことも考えられる．すなわち「いつでも人間がよい」わけではなく，逆に「いつでも機械がよい」わけでもないと考えられるため，つながる相手としての人間と機械は何が似ていて何が異なっているのか，という問いがこれから重要になってくると考えられる．

そして私たちは「声の向こうに誰かの身体を，そして名前をもつ一つの人格を想定せずにいられない（増野, 2014, p.217）」としたら，人間と機械の区別がつかなくなった世界において，私たちは声の“背後”にどのような存在を感じ，どのようにつながりたいと思うのであろうか．相手がいて初めて成立するコミュニケーションという場において，話し相手が人間か機械かも分からない状態で，私たちは声の“背後”にいると感じる存在とつながっている，あるいは理解し合えていると感じることができるのであろうか．このような声における「人間らしさ」に紐づく問いは，改めて「人間とは何か」について考える契機となるだろう．

5.4 認知科学研究のアプローチ

前節まで，声における「私らしさ」「その人らしさ」「人間らしさ」それぞれと紐づいて喚起されうる問いを検討してきた．それらを踏まえつつ本節では，主に認知科学の観点から，これらの問いに対してどのような研究の方向性が考えられるかについて検討を行う．

まず本稿では，声における様々な重ね合わせ状態の可能性を考えてきたが，いずれの場合においても，対象となる「声」は物理的には 1 つである．つまり生じている重ね合わせは，ヒトの認知における解釈の重ね合わせであり，認知科学の研究対象となる．量子

論的世界観を踏まえた認知科学研究として，たとえば文章理解における複数解釈の重ね合わせ状態について，量子確率論の公理に基づいて数理モデルで表現した研究があり（布山・西郷, 2022），そのようなアプローチを用いることで，主として人文社会学的な様々な論考において言及されてきた声における重ね合わせ状態について，異なる観点からの研究が可能となるかもしれない．

また，本稿で検討した声における重ね合わせ状態は，認知科学の概念であるプロジェクション・サイエンスと相性がよいように思われる．たとえば自己と他者の重ね合わせについては，アバターによる自己変容と関連付けて議論されているゴーストエンジニアリングと紐づけて検討ができるであろう（鳴海, 2019）．話者とキャラクターの重ね合わせについては，たとえば二・五次元舞台についてプロジェクション・サイエンスの観点から検討がなされており（久保, 2022），プロジェクション・サイエンスの領域で蓄積されてきた知見を活かした研究が考えられる．また，主に話者とキャラクターとの重ね合わせ状態に関する検討と関連する，哲学・美学領域でフィクション鑑賞の理論として用いられるメイクビリーブ（想像力）に関する議論は，プロジェクションという概念との親和性が高いと考えられる（松本, 2017; ウォルトン, 2016）．想像の引き金として働く事物がプロップ（小道具）であり，アニメにおける声優の声は，キャラクターの「声についての虚構的真理を生成するプロップ（松本, 2017, p.106）」である．そして，「虚構上のキャラクターである荒木レナに対して，実在する人間である藤田茜の声があてられることで，少なくとも音声面について，荒木レナの想像には実体が与えられる…藤田茜について，その声が荒木レナの声であると想像（シノハラ, 2017, p.70）」するとあるように，話者とキャラクターの重ね合わせ状態が生じる上で，この想像力に関する議論は重要である．そしてこのようなフィクション鑑賞における想像力に関するの理論は，プロジェクション・サイエンスにおける「表象を特定の人や事物に投射（プロジェクション）する」こととの相性がよいように思われ，これらの議論を接続することで，声における重ね合わせ状態についてより深い理解が得られるかもしれない．

先述したように，声における重ね合わせ状態はヒトの認知・解釈によって生じるものではあるが，重ね合わせ状態となりやすい「声」は存在している可能性がある．より具体的には，ポップカルチャーにおける人間の声の技術を考えた際に，たとえば「女性声優の少年声」には，通常の少年の声や他の演技音声とは異なる特有の「女性声優の少年声らしさ」に関する音響特徴が内包されている可能性がある（林・森勢, 2025）．また，アニメにおける声優の声を考えても，「アニメらしい声」についての調査を行った研究や（林・大杉, 2021），実際のアニメにおける声質を調べた研究が様々に存在している（原・伊藤, 2009; Ishi et al., 2023; Starr, 2015; Utsugi et al., 2019）．あるいは，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についても報告されており（Jurgens et al., 2011; Laukka et al., 2011; Scherer, 2003; Williams & Stevens, 1972），アニメやキャラクター，演技はいずれも「声」に通常の自発発話とは異なる特徴を持っていると考えられる．そうだとすると，たとえば「重ね合わせ状態にある声を情報科学的に認識する研究」や，「重ね合わせ状態にある声とそうでない声の聴取印象の違いに関する研究」なども，方向性としては考えられるかもしれない．

5.5 技術と社会との関係

本稿で検討した声における重ね合わせ状態は，主にポップカルチャーにおける人間の声の技術と，音声合成や声質変換などの音声工学技術と深く関わるものであった．今後の展望を考える上では，それぞれの技術と社会との関係についても検討することが重要である．

まず人間の声の技術については，「今あるものを，未来にどのように残していくのか」というアーカイブの観点が重要となってくるだろう．研究用の音声データベースやコーパスは多数存在しており，また歴史音声のデジタルアーカイブに関する検討も行われているが（Saeki et al., 2023），ポップカルチャーにおける人間の声に関するデジタルアーカイブについては，メディア芸術に関するデジタルアーカイブの取り組みの中でも十分に言及されていないのが現状である（三原, 2022）．声における重ね合わせ状態についての研究を行う上でも，また文化を守り伝えていく上でも，人間の声の技術のデジタルアーカイブについて検討していくことは重要である．

また，音声工学技術については，新たな科学技術の社会実装を考える際に必ず検討すべき点として，倫理的・社会的・法的課題（ELSI）についての議論が不可欠である（標葉・見上, 2024）．音声工学技術の ELSI については，人の容姿における肖像権のように，声の

人格権に関する法的な検討が行われていたり（荒岡他, 2023），論文のシステマティックレビューを通じてELSIに関する論点の抽出を試みた研究が存在したりするものの（Barnett, 2023），十分な議論がなされているとは言い難い状況であり，産官学民を巻き込んださらなる検討が重要となるだろう．

5.6 本論文の限界

本稿の限界として，声と深く関係するにもかかわらず，「言語」「音楽（歌声）」「多感覚（見た目などとの関係）」に十分に言及できなかった点を挙げたい．いずれも声について研究する上で非常に重要な要素であるにもかかわらず，一部言及されている場合はありつつ，本稿では十分な検討が行えていない．それはこれらの要素が重要でないことを意味するものでは全くなく，本稿が「声単体の」「歌声ではない発話音声における」「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（森他, 2014; 日本音響学会, 2003, p. 194））」について主に検討したことに起因する．声における重ね合わせ状態についての研究を進めていく中で，これらの要素については別の機会に整理を行っていきたい．

6 おわりに

本稿では声における「私らしさ」「その人らしさ」「人間らしさ」それぞれと関係する重ね合わせ状態について検討し，今後認知科学を始めとした諸領域において，どのような研究が行えるかについて展望を述べた．検討を通じて提起された「私らしさ」「その人らしさ」「人間らしさ」と紐づく問いは，いずれもこれからの社会における新たな「自己観」や「人間観」と繋がりうるものである．このことは，量子論的世界観に基づいて，声における「自己と他者」「話者とキャラクター」「人間と機械」の重ね合わせ状態について検討することで，未来の自己観や人間観がどのように形作られていくのかについて検討できることを示唆している．また本稿で述べた議論は，人間の技術（アート）と工学技術（エンジニアリング），あるいは人文社会系と理工系など，対比されやすいものを混ぜ合わせて検討することの価値を示唆している．声を題材とした研究の中で，様々な“間”を知り，繋いでいくことで，未来の日常における自己や人間についての理解が深まっていくことに本稿が寄与できればと願っている．

文　献

Arakawa, R., Kashino, Z., Takamichi, S., Verhulst, A., & Inami, M. (2021). Digital speech makeup: Voice conversion based altered auditory feedback for transforming self-representation. *Proceedings of the 2021 International Conference on Multimodal Interaction*, 159-167.

荒岡 草馬・篠田 詩織・藤村 明子・成原慧 (2023). 声の人格権に関する検討 情報ネットワーク・ローレビュー, *22*, 24-44.

Aronovitch, C. D. (1976). The voice of personality: stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, *99*(2), 207-220.

浅井 智久・丹野 義彦 (2010). 声の中の自己と他者：幻聴の自己モニタリング仮説 心理学研究, *81*(3), 247-261.

新 八角 (2018). 月ノ美兎は水を飲む ユリイカ, *50*(9), 92-99.

Aucouturier, J. J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, *113*(4), 948-953.

バーチャル美少女ねむ (2022). メタバース進化論 仮想現実の荒野に芽吹く「解放」と「創造」の新世界 技術評論社

Barnett, J. (2023). The ethical implications of generative audio models: A systematic literature review. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 146-161.

Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications*, *8*:42, 1-22.

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

Bredikhina, L. (2022). Babiniku: what lies behind the virtual performance. Contesting gender norms through technology and Japanese theatre. *Electronic Journal of Contemporary Japanese Studies*, *22*(2), 1-22.

Broom, T. W., Chavez, R. S., & Wagner, D. D. (2021). Becoming the King in the North: identification with fictional characters is associated with greater self-other neural overlap. *Social Cognitive and Affective Neuroscience*, *16*(6), 541-551.

Brown, S., Cockett, P., & Yuan, Y. (2019). The neuroscience of Romeo and Juliet: an fMRI study of acting. *Royal Society Open Science*, *6*(3), 181908.

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*(2), 116-120.

Chiba, Y., & Ito, A. (2024). Speaker intimacy estimation in chat-

talks based on verbal and non-verbal information. *IEEE Access*, *12*, 184592-184606.

Cohn, M., Bandodkar, G., Sangani, R. B., Predeck, K., & Zellou, G. (2024). Do people mirror emotion differently with a human or TTS voice? Comparing listener ratings and word embeddings. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1-10.

Crochiquia, A., Eriksson, A., Fontes, M. A., & Madureira, S. (2020). A phonetic study of Zootopia characters' voices in Brazilian Portuguese dubbing: the role of stereotypes. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *36*(3), 1-46.

Daryadar, M., & Raghibi, M. (2015). The effect of listening to recordings of one's voice on attentional bias and auditory verbal learning. *International Journal of Psychological Studies*, *7*(2), 155-163.

藤津 亮太 (2017). 声優語:アニメに命を吹き込むプロフェッショナル 一迅社

藤津 亮太 (2018). 声優論 : 通史的, 実証的一考察 小川 昌宏・須川 亜紀子 (編著) アニメ研究入門【応用編】: アニメを究める 11 つのコツ (pp. 93-117) 現代書館

藤津 亮太 (2019). プロフェッショナル 13 人が語るわたしの声優道 河出書房新社

福田 航希・阪井 瞭介・松下 嶺佑・國見 友亮・高道 慎之介 (2025). ELEVATE：学習者自身の自己聴取音声で聴く講義システム インタラクション 2025 論文集, *1B-21*, 314-319.

布山 美慕・西郷 甲矢人 (2022). 解釈の不定性の価値と量子認知による文章解釈研究の展望 認知科学, *29*(1), 100-119.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14-21.

Goldstein, T. R., & Filipe, A. (2018). The interpreted mind: Understanding acting. *Review of General Psychology*, *22*(2), 220-229.

Grawe, G. (2018). 日本文化における「声」 立命館言語文化研究, *29*, 155-173.

原 雄太郎・伊藤 克亘 (2009). 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.

Hatada, Y., Barbareschi, G., Takeuchi, K., Kato, H., Yoshifuji, K., Minamizawa, K., & Narumi, T. (2024). People with disabilities redefining identity through robotic and virtual avatars: A case study in avatar robot cafe. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, *61*, 1-13.

林 大輔（2019）. 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.

林 大輔・森勢 将雅 (2025). 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 Jxiv.

林 大輔・大杉 尚之 (2021). アニメにおける声質に対する印象の調査:日常場面およびドラマとの比較 PsyArXiv.

Holzman, P. S., & Rousey, C. (1966). The voice as a percept. *Journal of Personality and Social Psychology*, *4*(1), 79-86.

本多 清志 (2018). 実験音声科学:音声事象の成立過程を探る コロナ社

本間 裕之 (2024). 「身体」と「魂」としての VTuber 岡本 健・山野 弘樹・吉川 彗 (編著) VTuber 学 (pp. 323-337) 岩波書店

Hosaka, T., Kimura, M., & Yotsumoto, Y. (2021). Neural representations of own-voice in the human auditory cortex. *Scientific Reports*, *11*(1), 591.

Hughes, S. M., & Harrison, M. A. (2013). I like my voice better: Self-enhancement bias in perceptions of voice attractiveness. *Perception*, *42*(9), 941-949.

Iizuka, T., & Mori, H. (2022). How does a spontaneously speaking conversational agent affect user behavior? *IEEE Access*, *10*, 111042-111051.

稲岡 大志 (2016). 堀江由衣をめぐる試論 : 音声・キャラクター・同一性 フィルカル, *1*(2), 112-140.

井上 晴菜 (2022). 言語的符号化が標的音声の話者同定に与える影響 心理学研究, *93*(4), 320-329.

石田 美紀 (2020). 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

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

石井 沙季・伊藤 克亘 (2019). 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.

ジン リーファン (2010). 日本のアニメーションにおける音声の機能 :「GHOST IN THE SHELL/攻殻機動隊」「新世紀エヴァンゲリオン」「蒼穹のファフナー」を中心に 北海道大学大学院文学研究科研究論集, *10*, 235-251.

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57-62.

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, *2*(180), 1-11.

Kao, D., Ratan, R., Mousas, C., & Magana, A. J. (2021). The effects of a self-similar avatar voice in educational games. *Proceedings of the ACM on Human-Computer Interaction, 5(CHI PLAY)*, 1-28.

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

川村 覚文 (2024). 聖なるもの，情動，プラットフォーム：声優/キャラ・ライブコンサートにみるリアリティの複数性 須川 亜紀子（編）2.5 次元学入門 (pp. 81-113) 青土社

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*(6), 405-411.

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of “own voice.” *PLoS ONE*, *13*(6), e0199443.

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.

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

久保(川合) 南海子 (2022). 「推し」の科学 : プロジェクション・サイエンスとは何か 集英社新書

國見 友亮・畑田 裕二・木村 健太・鳴海 拓志・持丸 正明 (2024). 声質変化に伴う自己意識の変化についてのアンケート調査 第 29 回日本バーチャルリアリティ学会大会論文集, 2D1-08.

Kunimi, Y., Kimura, K., Matsumoto, K., Takamichi, S., Narumi, T., & Mochimaru, M. (2024). Character-voice embodiment impacts on the cognitive task performance with the voice ownership illusion. *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments 2024*, 1-10.

倉井 龍太郎・平木 剛史 (2023). ソーシャル VR プラットフォームにおけるエージェント API の提案 第 28 回日本バーチャルリアリティ学会大会論文集, 3D1-08.

倉田 将希・高道 慎之介・佐伯 高明・荒川 陸・齋藤 佑樹・樋口 啓太・猿渡 洋 (2021). リアルタイム DNN 音声変換フィードバックによるキャラクタ性の獲得手法 研究報告音声言語情報処理(SLP), *2021*(31), 1-6.

Kuriki, S., Tamura, Y., Igarashi, M., Kato, N., & Nakano, T. (2016). Similar impressions of humanness for human and artificial singing voices in autism spectrum disorders. *Cognition*, *153*, 1-5.

栗田 茂二朗 (1988). 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.

黒嵜 想 (2018). 縫い付けられた声 ユリイカ, *50*(9), 188-195.

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, *2*(175), 1-12.

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

Lu, Z., Shen, C., Li, J., Shen, H., & Wigdor, D. (2021). More kawaii than a real-person live streamer: Understanding how the otaku community engages with and perceives virtual YouTubers. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-14.

Lundh, L. G., Berg, B., Johansson, H., Nilsson, L. K., Sandberg, J., & Segerstedt, A. (2002). Social anxiety is associated with a negatively distorted perception of one's own voice. *Cognitive Behaviour Therapy*, *31*(1), 25-30.

丸島 歩 (2020a). 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.

丸島 歩 (2020b). 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.

増野 亜子 (2014). 声の世界を旅する 音楽之友社

松本 大輝 (2017). その歌は緑の髪をしている : ボーカロイドとメイクビリーブ フィルカル, *2*(2), 96-142.

松本 大輝 (2024). 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 (pp. 307-322) 岩波書店

Matsunaga, Y., Saeki, T., Takamichi, S., & Saruwatari, H. (2022). *Improving robustness of spontaneous speech synthesis with linguistic speech regularization and pseudo-filled-pause insertion*. arXiv.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say ‘Hello’? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), e90779.

三原 鉄也 (2022). メディア芸術に関する行政施策の展開とデジタルアーカイブ デジタルアーカイブ学会誌, *6*(1), 15-19.

三谷 雅純 (2014). 生涯学習施設の館内放送はどうあるべきか : 聴覚実験による肉声と人工合成音声の聞きやすさの比較 人と自然, *25*, 63-74.

三井 康行 (2010). 音声合成の利用シーンと要求される品質との関係 研究報告音声言語情報処理(SLP), *2010*(8), 1-4.

Mitsui, K., Hono, Y., & Sawada, K. (2023). *Towards human-like spoken dialogue generation between ai agents from written dialogue*. arXiv.

宮島 崇浩・菊池 英明・白井 克彦・大川 茂樹 (2013). 演技指示の工夫が与える音声表現への影響 : 表現豊かな演技音声表現の獲得を目指して 音声研究, *17*(3), 10-23.

宮本 直美 (2017). 嗜好対象としての歌声 : クラシック歌唱からポピュラー歌唱へ 嗜好品文化研究, *2017*(2), 26-37.

モクタリ 明子 (2008). 自然発話にみられる人物像に応じた個人内音声バリエーション 神戸大学総合人間科学研究科博士学位論文

森 大毅・前川 喜久雄・粕谷 英樹 (2014). 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

村瀬 ひろみ・須川 亜紀子 (2014). ジェンダー論 アニメから読み解くジェンダー :「力」と「美」を超えて 小川 昌宏・須川 亜紀子 (編著) 増補改訂版アニメ研究入門 : アニメを究める 9 つのツボ (pp. 76-95) 現代書館

永田 大輔 (2024). アイドル声優はなぜ問題となるのか : メディア史の中の「キャリア」 メディウム, *5*, 19-34.

中川 優奈・田中 章浩 (2021). 自分と他人の声の境界は変化するか 信学技報, *121*(177), 25-30.

難波 優輝 (2018). バーチャル YouTuber の三つの身体 : パーソン, ペルソナ, キャラクタ ユリイカ, *50*(9), 117-125.

鳴海 拓志 (2019). ゴーストエンジニアリング : 身体変容による認知拡張の活用に向けて 認知科学, *26*(1), 14-29.

成瀬 加菜・吉田 成朗・高道 慎之介・鳴海 拓志・谷川 智

洋・廣瀬 通孝 (2019). 自信声フィードバックによる緊張緩和手法の提案：クラウドソーシングを利用した自信声加工パラメータの推定 第 24 回日本バーチャルリアリティ学会大会論文集, 1B-05.

Nguyen, T. A., Kharitonov, E., Copet, J., Adi, Y., Hsu, W-N., Elkahky, A., Tomasello, P., Algayres, R., Sago, B., Mohamed, A., & Dupoux, E. (2023). Generative spoken dialogue language modeling. *Transactions of the Association for Computational Linguistics*, *11*, 250-266.

日本音響学会(編) (1996). 音のなんでも小事典 : 脳が音を聴くしくみから超音波顕微鏡まで 講談社

日本音響学会(編) (2003). 新版 音響用語辞典 コロナ社

Ogawa, N., Baba, J., & Nakanishi, J. (2024). Investigating effect of altered auditory feedback on self-representation, subjective operator experience, and task performance in teleoperation of a social robot. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1-18.

Ohata, R., Asai, T., Imaizumi, S., & Imamizu, H. (2022). I hear my voice; therefore I spoke: The sense of agency over speech is enhanced by hearing one's own voice. *Psychological science*, *33*(8), 1226-1239.

Ostrega, J., Shiramizu, V., Lee, A. J., Jones, B. C., & Feinberg, D. R. (2024). No evidence that averaging voices influences attractiveness. *Scientific Reports*, *14*(1), 10488.

皇牙 サキ (2018).「声」という商品のパッケージとしての VTuber ユリイカ, *50*(9), 64-65.

Peng, Z., Wang, Y., Meng, L., Liu, H., & Hu, Z. (2019). One's own and similar voices are more attractive than other voices. *Australian Journal of Psychology*, *71*(3), 212-222.

Qi, T., Zheng, W., Lu, C., Zong, Y., & Lian, H. (2024). Pavits: Exploring prosody-aware vits for end-to-end emotional voice conversion. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 12697-12701.

Saeki, T., Takamichi, S., Nakamura, T., Tanji, N., & Saruwatari, H. (2023). Selfremaster: Self-supervised speech restoration for historical audio resources. *IEEE Access*, *11*, 144831-144843.

Saito, Y., Nishimura, Y., Takamichi, S., Tachibana, K., & Saruwatari, H. (2022). *STUDIES: Corpus of Japanese empathetic dialogue speech towards friendly voice agent.* arXiv.

Samo, A., & Highhouse, S. (2023). Artificial intelligence and art: Identifying the aesthetic judgment factors that distinguish human-and machine-generated artwork. *Psychology of Aesthetics, Creativity, and the Arts*, Advance online publication.

佐藤 茉奈花 (2023). 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.

さやわか (2015). キャラの思考法：現代文化論のアップグレード 青土社

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*(4), 467-487.

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

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., Robertson, D. M. C., Simpson, A. P., & Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, *18*(9), 684-688.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zaske, R. (2014). Speaker perception. *WIREs Cognitive Science*, *5*, 15-25.

関根 麻里恵 (2024). メタ VTuber コンテンツの表象文化研究：「匿名性」「有名性」「声」「ジェンダー」から考える 岡本 健・山野 弘樹・吉川 彗 (編著) VTuber 学 (pp. 179-195) 岩波書店

柴 那典 (2014). 初音ミクはなぜ世界を変えたのか？ 太田出版

嶋田 総太郎 (2019). 脳のなかの自己と他者 : 身体性・社会性の認知脳科学と哲学 共立出版

標葉 隆馬・見上 公一 (編)(2024). 入門 科学技術と社会 ナカニシヤ出版

シノハラ ユウキ (2017). メディアを跨ぐヴィヴィッドな想像：『Tokyo 7th シスターズ』における「跳ぶよ」というセリフの事例から フィルカル, *2*(2), 60-95.

篠崎 大河 (2024). 実在する配信者としての VTuber 岡本 健・山野 弘樹・吉川 彗 (編著) VTuber 学 (pp. 271-288) 岩波書店

Sisman, B., Yamagishi, J., King, S., & Li, H. (2021). An overview of voice conversion and its challenges: From statistical modeling to deep learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 132-157.

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

Suda, H., Watanabe, A., & Takamichi, S. (2024). Who finds this voice attractive? A large-scale experiment using in-the-wild data. *Proceedings of Interspeech 2024*, 3165-3169.

須川 亜紀子 (2021). 2.5 次元文化論 : 舞台・キャラクター・ファンダム 青弓社

鈴木 真吾 (2014). サウンド/ヴォイス研究 アニメを奏でる 3 つの音：アニメにとって音とは何か 小川 昌宏・須川 亜紀子 (編著) 増補改訂版アニメ研究入門 : アニメを究める 9 つのツボ (pp. 96-119) 現代書館

高道 慎之介 (2021). 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.

Tan, X., Chen, J., Liu, H., Cong, J., Zhang, C., Liu, Y., Wang, X., Leng, Y., Yi, Y., He, L., Soong, F., Qin, T., Zhao, S., & Liu, T-Y. (2024). Naturalspeech: End-to-end text-to-speech synthesis with human-level quality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(6), 4234-4245.

田中 彰吾 (編著)(2023). 自己の科学は可能か：心身脳問題として考える 新曜社

程 斯 (2023). アニメ・キャラクターとアニメ声優を同一視する現象の原理的基盤をめぐって Phantastopia, *2*, 122-139.

Teshigawara, M. (2004). 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 (2019). 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.

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.

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

Uchida, T., Takahashi, H., Ban, M., Shimaya, J., Minato, T., Ogawa, K., Yoshikawa, Y., & Ishiguro, H. (2020). Japanese young women did not discriminate between robots and humans as listeners for their self-disclosure-pilot study. *Multimodal Technologies and Interaction*, *4*(35), 1-16.

内田 照久 (2000). 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.

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 (2005a). 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.

内田 照久 (2005b). 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 (2006). 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.

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・森勢 将雅 (2020). 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.

内田 照久・中畝 菜穂子 (2004). 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.

Ueda, A., Takahashi, H., Yoshikawa, Y., Ishiguro, H., & Nomura, H. (2021). Do robots facilitate life review narratives of older adults? A preliminary study. *Gerontechnology*, *20*(2), 1-12.

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

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., & Gentry, L. R. (2009). 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.

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (2011). 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.

Walton, K. L. (1990). Mimesis as make-believe: On the foundations of the representational arts. Harvard University Press.
（ウォルトン・ケンダル 田村 均 (訳) (2016). フィクションとは何か：ごっこ遊びと芸術 名古屋大学出版会）

Weiss, B., Trouvain, J., Barkat-Defradas, M., & Ohala, J. J. (Ed.) (2021). *Voice attractiveness: Studies on sexy, likable, and charismatic speakers*. Springer Singapore.

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

Xin, D., Takamichi, S., Morimatsu, A., & Saruwatari, H. (2023). Laughter synthesis using pseudo phonetic tokens with a large-scale in-the-wild laughter corpus. *Proceedings of Interspeech 2023*, 17-21.

山田 真也・伊藤 敏彦・荒木 健治 (2005). 対話相手の音声の品質を考慮した対話状況での言語的・音響的特徴の分析および様々な観点からの考察 情報処理学会研究報告音声言語情報処理(SLP), *2005*(127), 67-72.

山中 涼雅・大佐 健人・藤原 朱里・耿 浩彭・齋藤 大輔・峯松 信明・井上 雄介 (2025). 学習者本人の自己聴取音の声質で合成されたモデル音声を用いた発音学習とその効果 2025 年春季日本音響学会研究発表会講演論文集, 1165-1168.

山野 弘樹 (2024). VTuber の哲学 春秋社

山野 弘樹 (2025). 「VTuber 文化」を哲学する第 29 回 VTuber スタイル, *2025*(6), 78.

Yanagida, H., Ijima, Y., & Tawara, N. (2023). Influence of personal traits on impressions of one's own voice. *Proceedings of Interspeech 2023*, 5212-5216.

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.