



一般的・魅力的な男性の声質を用いた Voice Ownership Illusion による 顕在的・潜在的自尊心の変化の検証

國見友亮^{1,2)}, 木村健太^{2,3)}, 鳴海拓志¹⁾, 高道慎之介^{1,4)}, 持丸正明^{1,2)}

1) 東京大学 2) 産業技術総合研究所 3) 法政大学 4) 慶應義塾大学

概要: 自尊心とは自身に対する自己評価である。自尊心は否定的体験から自身を守るのに有効であり、短期的にも自尊心を向上させるのは重要である。自尊心に関係のある個人特性として、声の魅力がある。従来研究では、自身の声質を魅力的と感じている人は感じていないよりも自尊心が高いことが示されていた。本研究では魅力的な声質を利用した Voice Ownership Illusion によって、ユーザの自尊心の向上を引き起こすことを目的とする。本研究では、一般的・魅力的な男性の声質を用いた Voice Ownership Illusion によって、顕在的・潜在的自尊心の変化を検証した。その結果、声質の魅力を要因とした検証では、顕在的・潜在的自尊心の変化は確認できなかった。

キーワード: Voice Ownership Illusion, 発話主体感、声所有感、ボイスチェンジャー

1. はじめに

自尊心とは自身に対する自己評価である。自尊心が高いことは、否定的体験から自身を守るのに有効である [1]。短期的にストレスがかかる状況下で、自尊心を向上させることが重要ではないかと考えた。

ここで、ユーザの自尊心を向上を引き起こすのに、声質の魅力に着目した。従来研究では、自身の声を魅力的と感じる人は魅力的と感じない人よりも有意に自尊心が高いことが示された [2]。本研究では、体性感覚と共に聴覚提示を行う Voice Ownership Illusion を利用し、魅力的な声質を利用した Voice Ownership Illusion によって短期的に自尊心の向上を引き起こせるか検証する。

声の平均基本周波数に着目した従来研究では、男性の声は平均基本周波数が低いほど魅力的であり、女性の声は平均基本周波数が高いほど魅力的だと示されてきた [3]。本研究では、男性の声質の平均基本周波数に着目し、男性ユーザの声質を魅力的な男性の声質に変換し聴覚提示を行うことで、ユーザの自尊心が変化するか検証することとした。

本研究は、一般的・魅力的な男性の声質を用いた Voice Ownership Illusion によって、潜在的・顕在的自尊心が変化するか検証することを目的とする。

- RQ1-1: 平均基本周波数が低い男性の声を利用して、顕在的自尊心が変化するか?
- RQ1-2: 平均基本周波数が低い男性の声を利用して、潜在的自尊心が変化するか?
- RQ2-1: 声質の魅力に応じて主観的な声の所有感や発話主体感が変化するか?



図 1: 声質変換システム利用のイメージ

2. 声質変換システムについて

Voice Ownership Illusion とは、ユーザの声質を変調し聴覚提示を行うことで、聴覚提示された声質に対して発話主体感や声所有感を生じさせる錯覚である。本研究では、一般的・魅力的な男性話者の声質を用いた Voice Ownership Illusion によって、潜在的・顕在的自尊心の変化が引き起こされるか検証する。Voice Ownership Illusion に利用する声質変換システムのイメージを図 1 に示す。

2.1 声質変換システム

声質変換とは、ユーザの声質を任意の話者の声質に変化する技術である。本研究では、声質変換モデルをユーザに最適化する際の音声データの収録の手間が少なく、精度良く声質を変換可能な声質変換ソフトウェアである Paravo¹を利用することとした。Paravo は、声質変換先の話者として JVS コーパスに含まれる 100 名の話者のいずれかを指定できる。

本実験では、Paravo を用いた声質変換システムを構築し、一般的・魅力的な男性の声質を利用した Voice Ownership

¹<https://parakeet-inc.com/paravo>

Illusion を行うこととした。Fig. 1 に示すように、ユーザは骨伝導イヤホンを装着し、マイクに向かって発声を行うことで、ユーザの声質がターゲット音声の声質に変調された音声で聴覚提示される。

2.2 変換対象の声質

本実験では、Parakeet 株式会社²に問い合わせ、Paravo 内の変換先話者 ID と JVS コーパス内の話者 ID を照合し、所望の声質をもつ変換先話者 ID を決めることとした。本実験では、JVS コーパス [4] に存在する 49 人の男性話者の平均基本周波数を、声質の平均基本周波数を高精度に計算可能な WORLD[5] を用いて解析を行った。

従来研究では、一般的な男性の平均基本周波数は 121Hz とされていた [6]。そのため本研究では、Paravo の中で 121Hz に近い平均基本周波数を持つ男性話者、最も低い平均基本周波数 (約 101Hz) を持つ男性話者を変換先話者として利用することとした。以降では、前者と後者の声質をそれぞれ、一般的な男性の声質、魅力的な男性の声質と称する。

2.3 遅延

従来研究から、被験者の音声を遅延を伴い聴覚提示することで発話障がい引き起こされることが知られている [7]。遅延は、身体所有感錯覚における行為主体感や身体所有感にとって重要であると考えられている [8]。Voice Ownership Illusion においても遅延は重要なパラメータであると考え、音声入力から変換音声の出力聴覚までの遅延の計測を行った。

本計測では、実験で使用する読み上げ文章を読みあげた際の遅延の計測を行った。入力音声と出力音声はそれぞれ防音室内外で計測を行った。発話開始時点基準に遅延を確認したところ、約 90ms の遅延であることが確認できた。従来研究では 190ms 以内の遅延は行為主体間や身体所有感において影響はないと考えられており [8]、本実験で用いる声質変換システムの遅延も Voice Ownership Illusion に影響はないと考えられる。

3. 実験手法

本研究では一般的な・魅力的な男性の声質を用いた Voice Ownership Illusion によって、顕在的・潜在的自尊心の変化が引き起こされるか検証した。聴覚提示された声が自身の発声によって生じた声と感じる感覚を発話主体感、聴覚提示された声が自身の声と感じる感覚を声所有感と呼ぶ。主観的な声所有感や発話主体感のアンケートを用いて計測した。

本実験の手順について図 2 に示す。本実験では、顕在的・潜在的自尊心の変化を引き起こせたか確認するため、Voice Ownership Illusion 前にベースラインとなる自尊心の計測を行った。また、Voice Ownership Illusion 後に、顕在的・潜在的自尊心の計測を行い、聴覚提示による自尊心の変化の検証を行った。

3.1 Voice Ownership Illusion

本実験では、Voice Ownership Illusion に利用する声質に応じて自尊心が変化したか検証を行った。本実験では声

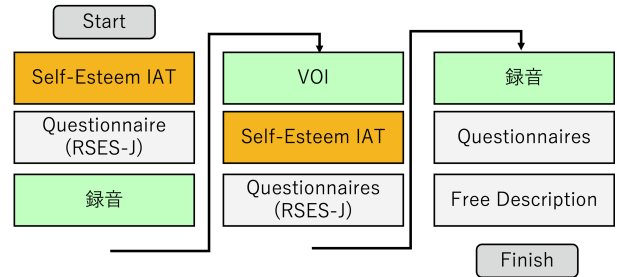


図 2: 魅力的な声質を利用した Voice Ownership Illusion による自尊心の変化の検証手順

質の魅力を変因とした被験者間比較を行うため、声質変換システムに利用する平均基本周波数を基準とし、魅力的な男性の声質について選定を行い、実験に利用する声質変換モデルの選定を行った。

Voice Ownership Illusion では、まず最初に、声質変換モデルの変換精度を高めるため、予め決められた文章の読み上げを行ってもらった。その後、ユーザに発声を行ってもらい、声質が変化されているか確認する。次に、ユーザの骨伝導イヤホンを介して聴覚提示される音量の調整を行うため、聴覚提示を受けながら ATR503 文 [9] の 10 文を読み上げてもらい、平静感情の 50 文 [10] を聴覚提示を受けながら読み上げてもらい。本実験は、実験参加者の声以外が声質変換システムにできるだけ入力されないよう、実験参加者は防音室の中で実験に参加した。

3.2 実験デザイン

実験参加者について説明する。本実験では、参加者を男性ユーザに限定し検証を行った。また除外基準として、聴覚障害と閉所恐怖症、20-40 歳ではない方とした。また、本実験ではユーザの声質を一般的な・魅力的な同性の声質に変換し聴覚提示を行い実験を行うため、被験者間比較によって検証を行う。

3.3 自尊心

自尊心の計測方法については、顕在的自尊心の質問紙である RSES-J[11] を用いた。潜在的自尊心の計測については、自己と自尊心の潜在的連合を計測する自尊心 IAT[12] を用いた。顕在的・潜在的自尊心の双方から、魅力的な声質を用いた Voice Ownership Illusion による自尊心の変化が引き起こされるか検証を行った。

3.4 発話主体感と声所有感の計測

本実験ではアンケート調査を用いて、発話主体感と声所有感の計測を行った。

3.5 その他の項目

3.5.1 Voice Similarity

Voice Similarity とは、聴覚提示された音声に変換対象の声質として聴取されているか確認する質問項目である。本実験では、男性の声質を用いた検証を行うため、声質変換の対象となる男性の声質として聴取されているか 7 段階のリッカート尺度で質問する。本実験では、声質変換の精度を担保するため、男性の声質に対しての Voice Similarity が

²<https://parakeet-inc.com/>

「どちらでもない」よりも低いスコアの場合、検証対象から除外することとした。

3.5.2 声質に対する印象

これにより、ユーザがアバターに対してどのようなイメージを持つかによって生起される。本実験では声質に対する印象を計測し、声質から想起される魅力従い自尊心の変化が引き起こされたか検証した。

4. 結果

本章では男性の一般的・魅力的な声質を用いた Voice Ownership Illusion による自尊心の変化についての検証結果を述べる。

4.1 主観的な発話主体感と声所有感

一般的・魅力的な声質に対する発話主体感と声所有感について、従来研究で用いられた発話主体感と声所有感のアンケートの日本語訳版 [13] を用いた。発話主体感については、聴覚提示された声は自身の発話によって生成された声であるという VoiceAgency を計測した。声所有感については、聴覚提示された声が自身の声だと評価する Own Voice、聴覚提示された声の特徴は自身の声と似ていると評価する VoiceFeatures を計測した。

VoiceAgency について、等分散性 (F test, $p > .05$) が棄却されなかったため、ウィルコクソンの順位和検定を実施した結果、VoiceAgency は魅力的な声質群と一般的な声質群の間に有意差は確認できなかった ($p = .786$, Cohen's $r = 0.07$)。

OwnVoice について、正規性 (Shapiro-Wilk test, $p > .05$) および等分散性 (F test, $p > .05$) が棄却されなかったため、スチューデントの t 検定を実施した結果、OwnVoice は魅力的な声質群と一般的な声質群の間に有意差は確認できなかった ($t(13) = -0.463$, $p = 0.651$, Cohen's $d = 0.239$)。

VoiceFeatures について、正規性 (Shapiro-Wilk test, $p > .05$) および等分散性 (F test, $p > .05$) が棄却されなかったため、スチューデントの t 検定を実施した結果、VoiceFeatures は魅力的な声質群と一般的な声質群の間に有意差は確認できなかった ($t(13) = -0.410$, $p = 0.689$, Cohen's $d = 0.212$)。

4.2 自尊心の変化

4.2.1 顕在的自尊心の変化

一般的・魅力的な声質を用いた聴覚提示によって、顕在的自尊心が変化したか質問紙を用いて計測した。結果を図3、図4に示す。顕在的自尊心の変化について、正規性 (Shapiro-Wilk test, $p > .05$) および等分散性 (F test, $p > .05$) が棄却されなかったため、スチューデントの t 検定を実施した結果、顕在的自尊心の変化は魅力的な声質群と一般的な声質群の間に有意差は確認できなかった ($t(13) = 0.514$, $p = 0.616$, Cohen's $d = 0.266$)。

4.2.2 潜在的自尊心の変化

一般的・魅力的な声質を用いた聴覚提示によって、潜在的自尊心が変化したか Self-Esteem IAT を用いて計測した。潜在的自尊心の変化について、正規性 (Shapiro-Wilk test,

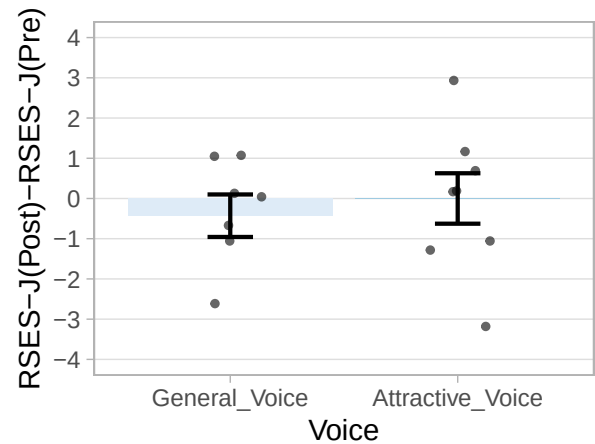


図 3: 顕在的自尊心の変化の結果。

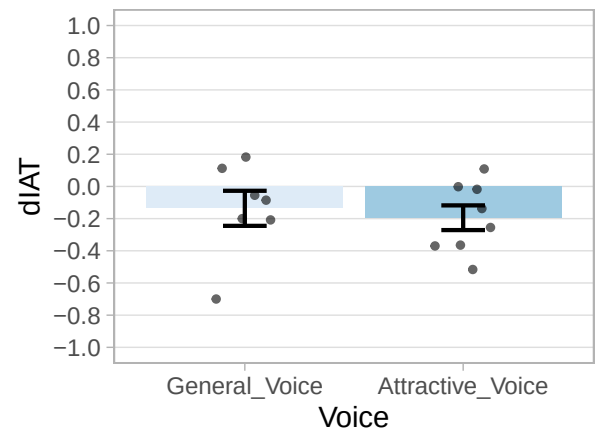


図 4: 潜在的自尊心の変化の結果。dIAT=SelfEsteem-IAT(Post)-SelfEsteem-IAT(Pre)。

$p > .05$) および等分散性 (F test, $p > .05$) が棄却されなかったため、スチューデントの t 検定を実施した結果、潜在的自尊心の変化について魅力的な声質群と一般的な声質群の間に有意差は確認できなかった ($t(13) = 0.448$, $p = 0.662$, Cohen's $d = 0.232$)。

5. 考察

5.1 RQ1-1, 1-2: 魅力的な声質の聴覚提示による自尊心の変化の検証

本実験では、従来研究を参考に、魅力的な声質を用いた Voice Ownership Illusion によって、自尊心に変化が引き起こされるか検証を行った。検証の結果、男性の一般的・魅力的な声質を用いた Voice Ownership Illusion によって、潜在的・顕在的自尊心の変化を引き起こすことができなかった。

本実験では、聴覚提示された声質に対して魅力的に聞こえたか7段階のリッカート尺度を用いて計測した。聴覚提示された声の魅力に関して、5 Points 以上のユーザは魅力的と感じており魅力有、4 Points 以下のユーザは魅力的と感じていなかった魅力無とする。声の魅力を確認したところ、一般的な男性の声質を提示した群は全ての参加者が提

示された声質を魅力無と回答しており、魅力的な男性の声質を利用した群は1人だけ魅力有と回答していた。

従来研究では、平均基本周波数を用いた個人性情報の変化について検証されている [14]。この研究では、ベースとなる声質の平均基本周波数から±2半音シフトの変化ではベースの声質から個人性情報が変化したと知覚できず、±4半音シフトの変化から声質の個人性情報の知覚に影響をしたと知覚できることを示していた。±2半音シフトは平均基本周波数を120Hzとした時の、106.9 Hz - 134.7 Hzに相当している。

一般的な男性の声質の平均基本周波数は約121Hzであるが、本実験で用いた魅力的な男性の声質は平均基本周波数が約101Hzであった。そのため、一般的な男性の声質と魅力的な男性の声質の間に個人性情報が変化するほどの平均基本周波数の変化が存在せず、聴覚提示に使用した魅力的な声質に対して魅力を感じられなかったのだと考えた。今後の実験では、聴覚提示される声質を要員とした実験を行う場合、要因に用いた声質の個人性情報に変化が存在するような実験設計を行い、声質を要因としたパラキート効果の検証を行う。

5.2 RQ2-1: 発話主体感と声所有感について

主観的な発話主体感と声所有感について、アンケート調査によって得た発話主体感と声所有感を用いた検証では、聴覚提示に用いられ一般的・魅力的な男性の声質の群間に有意差は確認できなかった。これらの結果から、本実験で利用した Voice Ownership Illusion では、聴覚提示に利用される声質の平均基本周波数の変化に応じ発話主体感や声所有感は変化しなかったと考えられる。

6. 結論

本章では、一般的・魅力的な男性の声質にを利用した Voice Ownership Illusion によって、声質の魅力に応じて男性ユーザの顕在的・潜在的自尊心の変化が引き起こされるか検証した。その結果、聴覚提示に利用した声質に応じた自尊心の変化は確認できなかった。今後は魅力的な声質の選定方法を再検討し、魅力的な声質を利用した Voice Ownership Illusion によって自尊心の変化が引き起こすことが可能か検証を行う。

謝辞 本研究は、JST、ACT-X、JPMJAX24CA の支援を受けたものです。

参考文献

- [1] Virgil Zeigler-Hill: Self-Esteem, 1st ed., Psychology Press, 2013.
- [2] 功刀友子: 自分の声を好きと評価する人はセルフ・エスティームが高いのか?: 大学生と専門学校生を対象として, 東亜大学紀要, Vol. 28, pp. 1-6, 東亜大学, 2019.
- [3] 須田仁志, 渡邊亞椰, 高道慎之介: 「キミは私の声, 好きかな?」大規模主観評価による声質好感度コーパスの構

築とその分析, 情報処理学会研究報告, No. 7, 2024 年 6 月.

- [4] Shinnosuke Takamichi, Kentaro Mitsui, Yuki Saito, Tomoki Koriyama, Naoko Tanji and Hiroshi Saruwatari: JVS corpus: free Japanese multi-speaker voice corpus, arXiv preprint arXiv:1908.06248, 2019.
- [5] Masanori MORISE, Fumiya YOKOMORI, and Kenji OZAWA: WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications, IEICE Transactions on Information and Systems, Vol. E99.D, No. 7, pp. 1877-1884, 2016.
- [6] Laetitia Bruckert, Patrick Bestelmeyer, Marianne Latinus, Julien Rouger, Ian Charest, Guillaume A. Rousselet, Hideki Kawahara and Pascal Belin: Vocal attractiveness increases by averaging, Current Biology, Vol. 20, No. 2, pp. 116-120, 2010.
- [7] 廣谷定男, 寛一彦, 辰巳格, 皆川泰代, 持田岳美, 渡辺眞澄: 聞くと話すの脳科学, 音響サイエンスシリーズ / 日本音響学会編, No. 17, コロナ社, 2017.
- [8] Ismail, MAF and Shimada, S: 'Robot' Hand Illusion under Delayed Visual Feedback: Relationship between the Senses of Ownership and Agency, PLOS ONE, Vol. 11, No. 7, e0159619, 2016.
- [9] Kurematsu, Akira, Takeda, Kazuya, Sagisaka, Yoshinori, Katagiri, Shigeru, Kuwabara, Hisao, and Shikano, Kiyohiro: ATR Japanese speech database as a tool of speech recognition and synthesis, Speech Communication, Vol. 9, No. 4, pp. 357-363, 1990.
- [10] 武石笑歌, 能勢隆, 加瀬嵩人: エントロピーに基づく音韻・韻律バランス感情依存文の設計と評価, 電子情報通信学会技術研究報告 = IEICE technical report: 信学技報, Vol. 115, No. 253, pp. 33-38, 電子情報通信学会, 2015 年 10 月.
- [11] Chizu Mimura and Peter Griffiths: A Japanese version of the Rosenberg Self-Esteem Scale: Translation and equivalence assessment, Journal of Psychosomatic Research, Vol. 62, No. 5, pp. 589-594, 2007.
- [12] 村上史朗: 自尊心の測定における潜在的認知指標の有効性: 潜在的連合テストを用いた検討, 東京大学大学院人文社会系研究科 博士論文, 2008.
- [13] Nami Ogawa, Jun Baba, and Junya Nakanishi: 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, pp. 18, 2024.
- [14] 北村達也, 川元広樹: 基本周波数のシフトが個人性知覚に及ぼす影響, 信学技報, Vol. 113, No. 134, pp. 33-37, 電子情報通信学会, 2013.