

アバター評価と“かわいい”音声： 音声の「かわいい」感情が及ぼす影響と加齢変化¹ Effects of CUTE voice on avatar-evaluation. and aging

原田 悦子[†], 安久絵里子[†], 鷹阪龍太[†], 目黒文乃[†], 葛岡英明[‡]
Etsuko T. Harada, Eriko Ankyu, Ryuta Takawaki, Ayano Meguro, & Hideaki Kuzuoka

[†]筑波大学 [‡]東京大学
[†]University of Tsukuba, [‡]Tokyo University
etharada@human.tsukuba.ac.jp

概要

アバター技術の新たな利用法として、仮想的な二者対話＝腹話術型会話支援システムを構築するに当たり、「かわいい」音声のアバター評価に与える影響を検討した。音声の物理的特性に対応した「かわいい」感情は若年成人には生起するが高齢者には見られないこと、音声のかわいい評価は高齢者でのみ外見のかわいい評価と相関を示すことが示された。音声の「かわいい」感情、かわいい感情と加齢の影響について検討が必要と考えられる。

キーワード：アバター評価(evaluation of avatar), 人工エージェント(agent), 音声のデザイン(voice design), 「かわいい」感情 (kawaii emotion), 加齢変化(aging), 腹話術形式の会話 (Ventriloquism-style conversations)

1. はじめに

近年、さまざまな人工エージェントが日常生活においても利用されているが、その一つにアバター利用システムがある。遠隔会議や遠隔接客システムなどで注目をされるアバター技術であるが、実際の話者がリアルタイムで話をしながら、複雑な操作なしで利用できるシステムとして大きな可能性があり、ヒューマンインタフェースデザインとしても興味深い存在である。

一般に、アバター利用システムは「現実の話者を隠す」あるいは「変形させる」形での対話形態が利用されるが、その可能性を考えると、もっと異なる形での様々な利用方法があると考えられる。その一つとして本研究では、一人の人が「一人二役をする」ような対話形式、いわゆる腹話術形式の会話をアバター技術の利用により構築するシステムを取り上げた。日本社会において

こうした腹話術形式の会話がしばしば取り上げられているのが、交通安全教室、詐欺予防講座といった、警察など公的機関による「講話」である。こうした講話では、児童生徒や一般市民、高齢者を対象として、多くの人に興味を持って話を聴いてもらい、その内容がよりよく理解されることを目的として、人形を用いた腹話術が用いられるが、その操作は複雑で、人形の体や口の動きに関する操作スキルの習得に長時間の訓練が必要とされること、加えて話者の演技分けと人形の操作を同時に行う必要もあって、認知的負荷の高い専門技術となっている²。その意味において、より容易な形で腹話術的な仮想的二者会話ができるシステムができるならば、その利用可能性は大きいと考えた。そこで、腹話術形式の交通安全講話を支援するアバター利用システムをとりあげ、その実現可能性の検討を行っている (鷹阪ら, 2022; 安久ら, 2022)。

ところで、アバター利用システムについては、その利用の有無自体が話題となりがちであるが、実際の利用を考える時、そのデザインがシステムの有効性や評価に大きく影響を与えることは想像に難くない。そのデザインは、アバターの視覚的デザインのみならず、その動きのデザインも大きな影響を持つであろう。また音声のデザインも考えられ、とりわけ腹話術型の会話を構築するシステムを考えたとき、「話者自身とは異なる声で話す」音声変換をかけること、その話し声のデザインをいかにするかという問題は大きな影響力を持つものと考えられる。

一般に、腹話術型対話において、仮想の対話相手とし

¹ 本研究は沖電気工業株式会社との共同研究の成果の一部である。また、本研究は筑波大学人間学群心理学類4年の目黒文乃の卒業研究「“かわいい”感性に対する音声の影響—アバター対話システムを用いた検討—」の一部をもとに再分析をしたものである。声質変換については、日立製作所の藤岡拓也氏ならびに孫慶華氏に音声ファイルの作成をしていただいた。また3本の講話台本の作成にあたり、茨城県警察本部交通部交通総務課の皆様にもさまざまにアドバイスをいただいた。いずれも記して感謝したい。

² 茨城県警・交通安全教室担当者によると、たまたまその担当に割り当てられた職員のうち、腹話術形式の講話に興味を持つ人がいる場合に、新たな講話実施可能者としての練習・指導が行われるため、実施者を常時確保することは難しい状況にあるという (2019年12月私信)。

て幼い子どもや幼児に見立てたキャラクターを想定することが多い。そのため、演者は仮想対話者としての「かわいい声」での発話を行うことが多く、そうした発話技術も求められていると考えられる。

「かわいい」は、対象に対する好意的な、多様な評価を表す言葉として日常的に用いられるが、入戸野 (2019) は、「かわいい」を外的な刺激特性ではなく、人の認知過程に生じる感情語として定義している。先行研究でのかわいい感情研究はその大部分が視覚刺激を対象とするものであるが、一方でアニメーションやVTuber等では、声の「かわいい」も魅力としてしばしば取り上げられている。しかし、聴覚刺激が対象の印象に与える影響として、声のフォルマントの高さや発話速度等が話者の性格印象に影響を及ぼすこと (内田・中畝, 2004; 内田, 2011)、楽器の音色ではピッチの高い音が「かわいい」という評価を得る傾向にあること (大倉・菅野, 2014) などが報告されているものの、実際の人の声を使用した「かわいい」音声の実験的検討は行われていない。これは聴覚音声刺激での「かわいい」が物理的な刺激特性上の定義・デザインが難しいことも反映しているものと思われる。

一方、かわいい感情の生起やその反応については、年齢による違いも指摘されている。豊島・入戸野 (2020) は、高齢者はかわいさを知覚させる属性に対して感情的反応が起こる一方で、キャラクター人形など現在の若者文化において「かわいい」とされるものには興味関心が低いという側面があることを報告している。上述のように本研究でとりあげる講話では、その主対象者層の一つに高齢者があることも含め、かわいい感情についてより一般的な特性を検討する必要があり、そのためには年齢による相違も検討対象とする必要があると考えた。

そこで本研究ではアバター利用場面として、腹話術形式の仮想的二者対話による交通安全講話支援システムをとりあげ、エージェント利用において「かわいい」音声を使う効果として、「かわいい」感情発生があるか否か、それは効果が音声のみならず外見評価にも影響を与えるか否かを検討した。またそこに加齢による変化があるか否かについても検討を行った。

2. 方法

実験の概要 高齢者向け交通安全講話の腹話術型ビデオ講話 3 題をビデオ作成し、アバターに「かわいい

音声」を付与した効果を、主観評価 (課題 1) ならびに接近動機の測定 (課題 2, アバター選好ならびに単調な課題での課題遂行数) により検討し、年齢群間比較を行った。かわいい音声として物理的な特性を操作した音声変換である VT-4 条件、声優の声をういた変換を行った声質変換条件をオリジナルの地声条件と比較した。

実験計画 音声条件 (地声条件/VT-4 条件/声質変換条件) × 年齢群 (高齢者群/若年者群)。音声条件は、課題 1 では参加者内要因、課題 2 では参加者間要因。

参加者 高齢者群 24 名、若年者群 (大学生) 24 名。課題 2 では年齢群ごとに 8 名 (男女同数) ずつ、音声 3 条件にランダムに割り当てた。

実験材料 オリジナル刺激として若年男性による地声条件刺激を録音し、そのデータを用いて変換をかけた音声刺激を作成した。VT-4 条件では、Roland 社製の音声変換機器 (VT-4 Voice Transformer) を用いて、地声条件の音声との変化が明らかで「かわいい声」と感じられるよう、地声条件の音声のピッチとフォルマントを +1.5 として録音した。声質変換条件では、藤岡・孫 (2018) により、子役も務める女性声優の声をモデル声として音声変換を行った。

アバターとして、外見から声質を想像しにくいロボット型を選定、ロボタンと命名した。ハンドトラッキング装置 (Ultraleap 社) により、手の動きでアバターを操作した (システム詳細については、鷹阪ら (2022) 参照)。ビデオ作製では、まず成人男性の地声条件での音声ファイルを作成、これに合わせてアバターの自然な動きを構成した動画を作成した。また地声条件音声をベースに 2 種の「かわいい声」音声を作成し、変換後の音声ファイルを合成して、刺激ビデオとした。作成したビデオ刺激の事例を図 1 に示す。



図 1 課題 1 の講話動画の課題画面例

手続き 音声なしのロボタンの自己紹介動画を見て、外見に関する評価を行った後、課題 1 では、警察官 (シルエット) とロボタンの会話ビデオ 1 本が提示され、直後に主観評価を求める試行を 3 試行を行った。主観評価

は、音声／外見のSD法（「かわいい」を含む形容詞7対）、および講話内容に関する4項目をVASで回答した。音声条件の呈示順は参加者間でカウンターバランスされた。3本のビデオに登場するアバターは胸に付けたシンボルで弁別され「別のロボット」とされ、全ビデオ視聴後に「この後も話を聞きたいのはどのロボタンか」という選好判断を求めた。ただしこの選好判断と実際に課題2で提示されたアバター音声とは独立であった。課題2「自由実施課題」では、「人の判断を求める」単調課題として発話の速さ評価課題を実施した。短文24文を3速で作成、3種の音声ファイルを作成した。課題では69試行が疑似ランダム順に自動再生されるが、参加者は音声話速を6件法で判断した後、課題を続行するか否かを選択するように求められ、実際に実施された課題数をカウントした。その後、参加者属性として「かわいいに対する態度」7項目、声をあてた動画等の聴取経験ほか評価6項目で尋ね、最後に事後インタビューが実施された。

3. 結果・考察

音声ならびに外見に対する「かわいい」評価 音声ならびに外見についてのSD評価の内、「かわいい-かわいくない」について、音声条件（3・参加者内：地声、VT-4、声質変換）×年齢（2：若年者、高齢者）×参加者性別（2：男性、女性）の3要因混合分散分析を行った。性別の要因については、先行研究での「かわいい」評価において性差による主効果、交互作用がさまざまに報告されていたために組み込まれた。

音声に対する「かわいい」評価（図2）では、年齢の主効果が有意であり（ $F(1, 44)=10.79, p=.002$, 偏 $\eta^2=.20$ ）、若年者より高齢者のほうが全体的にかわいい音声だと評価した。また、音声条件の主効果が有意であった（ $F(2, 88)=13.63, p<.001$, 偏 $\eta^2=.08$ ）が、年齢×音声条件の交互作用が有意であり（ $F(2, 88)=4.43, p=.01$, 偏 $\eta^2=.09$ ）、単純主効果検定の結果、高齢者群では音声条件間に有意な差は見られなかったが、若年者群では地声条件ならびにVT-4条件はそれぞれ、声質変換条件よりも有意に低く評価された（地声条件 vs. 声質変換条件, $t(44)=-3.83, p=.001$ ；VT-4条件 vs. 声質変換条件, $t(44)=-5.96, p<.001$ ）。性別ならびに性別との交互作用はいずれも有意ではなかった。これに対し、外見に対する「かわいい」評価は、いずれの主効果、交互作用とも有意ではなかった（図3）。

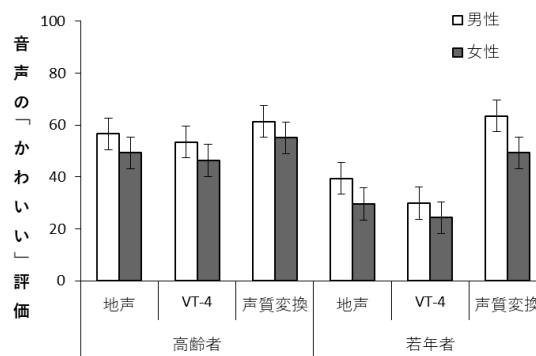


図2 音声に対する「かわいい」評価（エラーバーは標準誤差）。

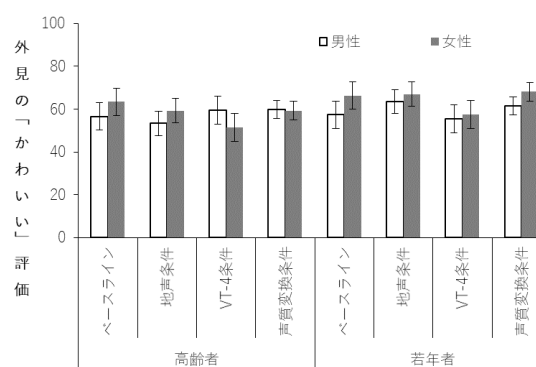


図3 外見に対する「かわいい」評価（エラーバーは標準誤差）。

接近動機 課題2の前に動画視聴のためのアバターの選択を求めた際の、音声条件ごとの選好数を年齢群ならびに性別にカウントした（表1）。アバターの選好数を従属変数として、年齢群あるいは性別ごとに2×3のカイ二乗検定を行った結果、アバターの選好数に有意な差はないことが示された（年齢群別, $\chi^2(2)=3.85, p=.15$ ；性別, $\chi^2(2)=1.85, p=.40$ ）。

表1 年齢群/性別ごとの各音声条件のアバター選好数

音声条件	年齢群			性別		
	若年者	高齢者	合計	女性	男性	合計
地声	5	11	16	9	7	16
VT-4	7	4	11	7	4	11
声質変換	12	8	20	8	12	20
合計	24	23	47	24	23	47

さらに課題2の課題実施試行数（図4）については、高齢者群24名中23名が全69試行を実施し天井効果が見られたため、若年者群のみを分析対象とした。実施試行数を従属変数として、性別（2）×音声条件（3）を要因とする2要因参加者間分散分析を行ったところ、い

ずれの主効果と交互作用も有意ではなかった。

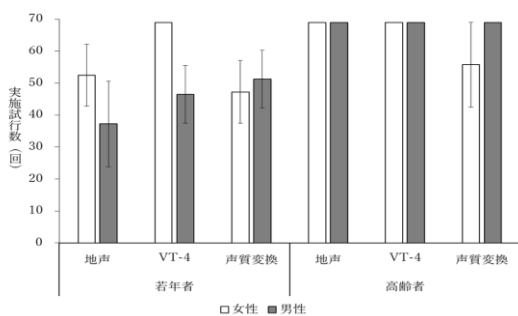


図4 課題2における実施試行数
(エラーバーは標準誤差を示す)。

音声／外見の「かわいい」評価とその関係性 以上のように、実験的に操作を行った音声条件は、外見に対する「かわいい」評価や接近動機としての指標への影響がほとんど見られなかった。一方、前述のように音声条件が音声に対するかわいい感情に影響を与えたのは若年成人に限られており、特に高齢者においては音声のかわいい評価が必ずしも実験的に操作された音声条件の影響を反映していなかったことから、各参加者による音声のかわいい評価値と外見のかわいい評価値との関係性について検討した(図5)。その結果、両者のかわいい評価値について、若年者群では有意な相関が認められなかった($r=.15$, $p=n.s.$)のに対し、高齢者群ではやや強い正の相関がみられた($r=.64$, $p<.01$)。

4. 結論

以上、本研究では若年男性が話し手となった腹話術形式の交通安全講話において、会話相手として画像で表示されるアバター(名:ロボタン)の評価において、1) 話者の音声を変換して音声刺激を「画像に充てる」ことにより、音声に対するかわいい感情がもたらされることがあること、2) 若年成人では音声のかわいい感情と、物理的音声特性との間に一定の対応関係があり、視覚的かわいい感情(外見へのかわいい評価)とは独立であるが、3) 高齢者では音声の「かわいい」評価が、音声の物理的的特性とは必ずしも関係しない一方で、外見のかわいい評価とは関連性が強く、可愛い感情が分化しておらず、包括的な評価となる可能性が示された。こうした実験結果を聞いたある高齢者は、「かわいい、というのは、そのモノ(ここではロボットアバター)全体がかわいいのでだから、声だけ、見た目だけかわいいというのはないと思う」と感想を述べており、豊島・入

戸野(2020)の加齢に伴うかわいい判断の変化と共にさらに検討を行っていく必要が示された。

なお本研究で利用した接近動機の指標については、いずれも十分な検出力を持ちえなかった可能性が示唆される。単調課題の実施試行数では、特に高齢者においてほとんどの参加者が「ロボットのために」全試行を実施する結果となり、それぞれの年齢群における利他的行動・態度等の影響とは独立に「かわいい」感情の影響を測定しうるか、検討が必要である。また、アバター選好判断において、本実験でのVT-4条件と声質変換条件を音声変換条件として合わせてカウントすると、若年では元音声5、変換後音声19である選好が、高齢者では元音声11、変換後音声12であり、その選択比率にやや偏りが見られた($\chi^2(1)=3.81$, $p=.051$)。この原因として、一部の高齢参加者で地声条件の方が聞きやすいという感想を述べており、音声変換という操作によって多少とも聴覚刺激としての情報量が減っている可能性、ならびに一般に「かわいい」音声は「高い声」となるために加齢による聴力変化から聞き取りにくくなることによっている可能性が否定できない。音声情報の処理は、揮発性を有すること、ならびに加齢に伴う感覚・知覚的变化の影響を大きく受けることから、音声デザインにおいて、感情的な受容性のみならず、その基本となる感覚知覚的な処理可能性についても考慮するべきであることが示唆された。

近年、生活の中の各種機器についてもさまざまな人工エージェント要素が組み込まれてきている。そうしたエージェントの音声デザインについて、その満たすべき要件は何か、音声のデザインとして人の認知過程から多層的な検討が必要とされているといえよう。

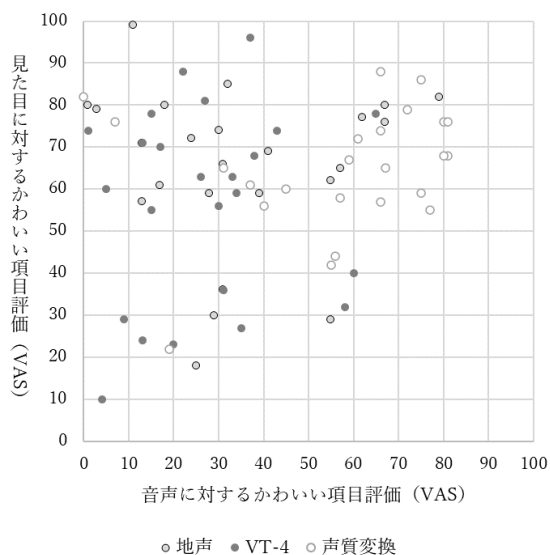
文献

- [1] 安久絵里子・鷹阪龍太・目黒文乃・葛岡英明・原田悦子(2022) アバターと音声変換技術が講話能力の向上に与える影響(2): 聞き手によるパフォーマンス評価. 第39回大会発表論文集.
- [2] 藤岡 拓也・孫 慶華(2018). 複数のネットワークの出力を考慮した DNN ベース声質変換手法. 信学技報, 117(393), 11-15.
- [3] 入戸野 宏(2019). 「かわいい」のちから: 実験で探るその心理. 化学同人.
- [4] 大倉 典子・菅野 涼(2014). かわいい人工物の系統的研究 第17報—「かわいい音」に対する基礎的検討— 信学技報, 114(52), 389-392.
- [5] 鷹阪龍太・安久絵里子・目黒文乃・葛岡英明・原田悦子(2022) アバターと音声変換技術が講話能力の向上に与え

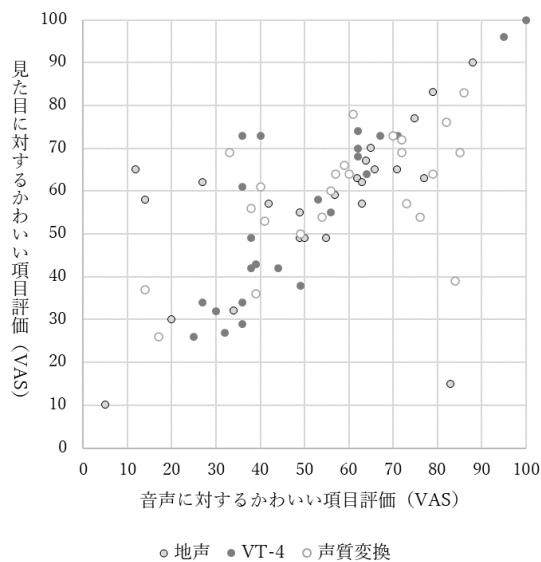
る影響(1)：話し手自身による主観評価による検討

- [6] 豊島 彩・入戸野 宏 (2020). 高齢者にとっての“かわいい”の概念と構成要素 老年社会科学, 41 (4), 409-419. https://doi.org/10.34393/rousha.41.4_409
- [7] 内田 照久 (2011). 音声の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. <https://doi.org/10.4992/jjpsy.82.433>
- [8] 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. <https://doi.org/10.4992/jjpsy.75.397>



a) 若年成人群



b) 高齢者群

図1 音声のかわいい得点と外見のかわいい得点の関連

a) 若年者群 ($r = .15, p = n.s.$), b) 高齢者群 ($r = .64, p < .01$)