

遠隔操作型対話ロボットの操作者に適切な発声調整を促す 音量ゲージの開発と予備的評価

Voice Volume Gauge to Encourage Appropriate Vocal Adjustment of an Operator of a Teleoperated Social Robot: Development and Preliminary Evaluation

松島 圭吾^{1*} 窪田 智徳¹ 小川 浩平¹ 佐藤 理史¹
Keigo Matsushima¹ Tomonori Kubota¹ Kohei Ogawa¹ Satoshi Sato¹

¹ 名古屋大学大学院工学研究科

¹ Graduate School of Engineering, Nagoya University

Abstract: 遠隔操作型対話ロボットの操作インタフェースでは、操作者はロボットと対話相手との距離を把握しづらい。そのため、操作者は距離に応じて対話相手に聞こえやすく話すための発声の調整が難しく、また実際に聞こえやすく話しているか不安を感じてしまう。本研究ではこの問題を解決するため、操作者の現在の音量と最適な音量との差分をインタフェース上に表示することで、操作者に発声の調整を促す音量ゲージを開発した。本稿では、音量ゲージを実装し、予備実験でその使いやすさや操作者の不安感の軽減について検証した結果を報告する。

1 はじめに

遠隔操作型対話ロボットとは、操作者が遠隔地のロボットを操作することで、ロボットを介して遠隔地の人と対話できるものである。操作方法の例としては、ロボットの操作者は、遠隔地の映像・音声を画面上の操作インタフェースで見ながら、マイクを通じて自身の声を遠隔地のロボットから出力することで、遠隔地にいる相手とロボットを介して対話できる。このようなロボットは、人間が移動せずとも互いにコミュニケーションを取ることができるという利点から近年注目を集めている。例えば遠隔での高齢者のヘルスケア [1] からスーパーマーケットでの接客 [2]、博物館 [3] での案内で応用がなされている。

ロボットの操作インタフェースでは、操作者が取得できる情報が少ないため、対面対話と比べて対話の状況を把握しづらく、操作者と対話相手は自然に対話することが難しい。特に、操作者はロボットと対話相手との距離感を把握するのが難しいため、どのように発声すべきか判断できず、また相手に聞こえやすく話しているかについて不安を感じてしまうといった問題が未だあり、操作者にとっての対話しづらさが存在する。一方で、対面対話における人の発声では、相手との距離感に応じて音量や声質（声の高さ、抑揚、リズム等）

が変化することが知られている [4]。例えば、人は遠くの人には大きな声で、近くの人には小さな声で話すなど適切に音量を変化させ、またそれに伴って声質も変化する。しかし上記の問題により、ロボットを介した対話では、例えば操作者が必要以上に大きな声で話して対話相手がうるさく感じたり、あるいは必要以上に小さな声で話して対話相手が聞き取りづらくなるなど、対話の自然さが低減してしまう。ロボットを介した自然な対話を実現するためには、対話相手との距離や周囲の騒音レベルに応じて、操作者の適切な発声調整を可能にする手法の導入が不可欠である。

先行研究としては、ロボットの操作者の音量のみの調整に着目した、次の2つの研究が存在する。Hayamizuらは、ロボットと対話相手との距離および周囲の騒音レベルから、ロボットから出る操作者の音量を対話相手に聞こえやすいように自動調整するシステムを実現した [5]。Ruebenらは、操作画面上の対話相手を選択すると、その対話相手にとって聞こえやすいようにロボットから出る操作者の音量が自動で調整される操作インタフェースを実現した [6]。

上記の先行研究では音量のみに着目していたが、ロボットを介した自然な対話には、声質の変化も考慮する必要があると考える。人は音量変化に伴って声質も変化することから、ロボットから出る操作者の音量のみを変えると、対話相手にとっては声に違和感を感じる可能性がある。例えば、操作者が小さな声で囁くよ

*連絡先：名古屋大学大学院工学研究科
(愛知県名古屋市千種区不老町)
E-mail: matsushima.keigo.y5@nagoya-u.ac.jp

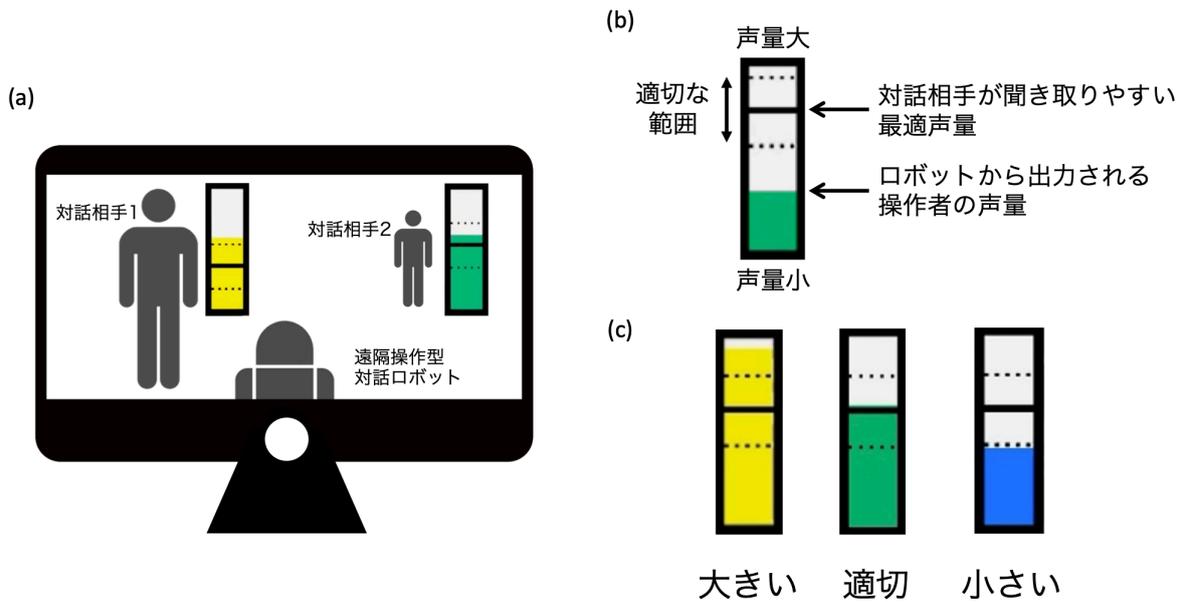


図 1: 音量ゲージの設計イメージ。(a) 操作インタフェース画面上での表示イメージ。ロボットと対話相手との距離および騒音によって最適音量が変動する。(b) 音量ゲージの各部の説明。音量ゲージには最適音量とその上下に適切な範囲、また操作者の音量がリアルタイムに表示される。(c) 音量ゲージの色の变化。操作者の音量が適切であると緑色、大きいと黄色、小さいと青色に変化する。

うに話している声の音量のみを上げた場合、人間の発声として現実的とは言い難い声になってしまい、ロボットの声として不自然になってしまうと考える。

我々は、操作者自身による発声すなわち音量と声質の調整が可能な操作インタフェースの実現を目指す。ここで、対話相手の状況に合わせて、声質も含めて発声を自動で調整することは現段階では難しく、また実現できたとしても自動調整時の遅延の発生は避けられない可能性が考えられる。そのため、本研究では操作者が自身で発声を調整するアプローチをとる。

本研究では、ロボットを介した自然な対話の実現を目指し、操作者自身での適切な発声調整を可能にする操作インタフェースを開発する。具体的には、操作者の現在の音量と、対話相手が聞き取りやすい最適音量を推定して、操作インタフェース上でリアルタイムに表示する音量ゲージ（次章で詳述）を開発する。音量ゲージを用いることで、操作者は自身の音量を対話相手にとって最適な音量になるよう調整できるようになり、その音量調整に従って声質も適切に変化すると考えられる。本稿では、音量ゲージの開発を行い、予備実験を通じて操作者にとっての音量ゲージの使いやすさと、対話相手にとって聞こえやすく発声できているかについての不安感を調査した結果を報告する。

2 提案手法

本研究では、操作者の適切な発声調整を可能にする手法として、操作インタフェース上に提示できる音量ゲージ（図 1）を提案する。本章では、音量ゲージのコンセプトおよび実装方法について述べる。

2.1 音量ゲージのコンセプト

音量ゲージでは、操作者の現在の音量と、対話相手が聞き取りやすい最適音量を推定して、操作インタフェース上でリアルタイムに表示する。操作者は、音量ゲージで示された自身の音量と最適音量との差分を見て、自身の音量を最適音量になるよう調整することができ、またこれに伴って声質も適切に変化すると考える。図 1 の音量ゲージの例では、操作者はより大きな音量で話すことで、自分自身の音量が対話相手にとって聞き取りやすい最適音量になることを読み取れる。尚、画面上に対話相手が複数いる場合を考慮し、各相手に対する音量ゲージをその人の顔の横に表示する。音量ゲージを用いることで、操作者は対話相手が聞き取りやすい発声で話せるようになり、またどのような発声をすべきかの不安感を低減し、ロボットを介した自然な対話ができるようになると思われる。

音量ゲージにおいて、最適音量は、ロボットと対話相手との距離および対話相手の周囲の騒音レベルから

推定して算出する。すなわち、最適音量を推定するモデルが必要となる。また、音量ゲージ上の最適音量の上下に一定の範囲を設定し、この範囲に合うような音量で操作者は発声できれば良いものとした。このような範囲を設定する理由は、単一の最適音量に合わせるよう発声するのは操作者にとって難しく感じられる可能性があり、一定の範囲内で発声できれば良いようにすることで、音量ゲージを使用しやすくなると思ったためである。尚、その範囲は最適音量を推定するモデルから、対話相手にとっての聞き取りやすさに問題のない範囲に設定する。また、操作者の現在の音量と最適音量との差が視覚的にわかりやすくなるよう、現在の音量がその範囲内にある場合はバーを緑色に、範囲を超える（最適音量より大きすぎる）場合は黄色に、範囲未満にある（最適音量より小さすぎる）場合は青色になるようにした。

2.2 音量ゲージの実装方法

本研究では、図1に示すような音量ゲージを開発した。音量ゲージのシステム構成を図2に示す。

開発においては、次の3つの要素を実現した。

1. 対話相手側の映像取得と映像上の対話相手の顔付近への音量ゲージの描画
2. 操作者の音量の測定
3. 対話相手が聞き取りやすい最適音量の推定

1つめについて、まず Kinect v2 を用いて、対話相手側の RGB 映像を取得した。さらに映像上の対話相手の顔検出を行い、顔付近に音量ゲージを重畳表示できるようにした。

2つめについて、マイクから操作者の声を取得し、Python の PyAudio ライブラリ¹を用いて、音のエネルギーの二乗平均平方根を計算することで、操作者の音量を測定した。尚、操作者側のマイクではノイズが含まれることがあったため、本稿ではノイズフィルタ²を用いており、操作者の音量はノイズフィルタを通した後の音量を測定して取得した。また、本来はロボット側スピーカから出力される操作者の音量を測定し、それを操作者の音量として音量ゲージ上に示すべきであるが、今回は操作者の音量と同じ音量の声がスピーカから出力されるよう設定することで、操作者の音量をそのままゲージに示しても問題ないようにした。

3つめについては、最適音量推定モデルを構築することで実現した。構築方法は次章で詳述するが、まず実験を行い、ロボットと対話相手との距離ごとに、対話

相手が聞き取りやすい最適音量を測定したデータを収集し、このデータを用いて、ロボットと対話相手との距離から最適音量を推定できるモデルを作成した。音量ゲージの実使用時には、ロボットの付近に Kinect を設置して、ロボットと対話相手との距離を取得することで、作成したモデルから最適音量を推定する。尚、最終的には対話相手の周囲の騒音情報も含めた最適音量の推定を目指す。本稿ではまず、騒音は一定の状況のみを扱うこととした。

3 最適音量推定モデル作成実験

本章では、先述の最適音量推定モデルを作成するために行った実験について述べる。実験は、Hayamizu ら [5] で同様のモデル作成のために行われた実験を参考にして行った。

3.1 実験方法

モデル作成のため、ロボットと対話相手との距離ごとに、対話相手が聞き取りやすい最適音量を測定したデータを収集する実験をまず行った。実験では、図3のようにスピーカと騒音計を設置し、実験参加者とスピーカとの距離を 50cm, 100cm, 150cm, 200cm, 250cm, 300cm, 350cm, 400cm, 450cm の 9 パターンに変えながら、各距離における最適音量を調べた。騒音計はスピーカから流す音声の音量を測定するためのもので、スピーカから 5cm 離れた位置に設置した。尚、騒音計は C 特性³のものを用いた⁴また実験はエアコンや換気扇を切った部屋で実施し、このときの騒音レベルは騒音計（C 特性）を用いて約 50dB であった。

最適音量の測定は、次の手順で行った。

1. 参加者に先述の 9 パターンのうちある地点に立ってもらおう
2. 後述の測定用音声を、音量を 70dB から約 0.5dB ずつ大きくしながら再生していく
3. 参加者が音声について「音量が大きすぎず小さすぎず、自分にとってだけではなく他の人にも聞き取りやすい」と感じた時点の音量を最適音量とする

最適音量の測定を頑健に行うため、各参加者は各距離で 4 回ずつ測定を行った。手順としては、参加者はまずスピーカとの距離 50cm 地点に立った状態から始め、

³<https://www.rex-rental.jp/faq/product/1138>

⁴騒音計には、人間が聴取した音の大きさの感覚に近くなるように、音圧レベルの測定値に補正をかける特性をもつものが存在するが、今回は音圧レベルそのものを測定するため、人の声の音域で補正をかけない C 特性を用いた。

¹<https://pypi.org/project/PyAudio/>

²<https://www.nvidia.com/ja-jp/geforce/broadcasting/broadcast-app/>

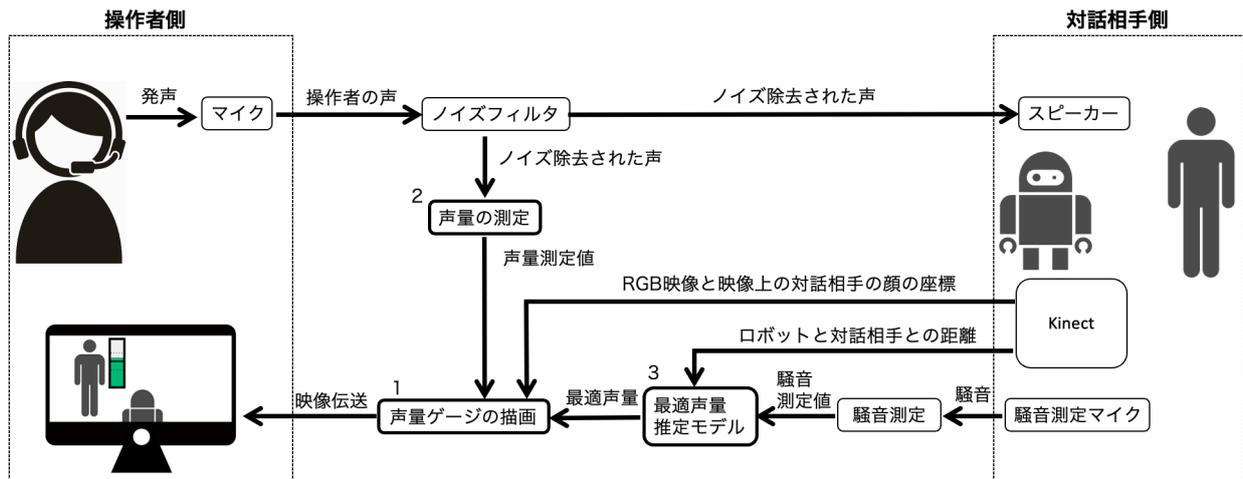


図 2: システム構成図

スピーカから離れていく順に各距離で測定し、450cm地点での測定後はスピーカに近づいていく順に各距離で測定した。また各地点では、上記の2と3を連続して2回行った。

測定用音声としては、Hayamizuら[5]の実験に倣って、参加者にとって十分親しみがあると考えられる47都道府県名を1つ発話する47種類の音声を用いた。上記2においては、音量を大きくするごとに異なる都道府県名の音声を再生した。また各音声は、すべての音素が同じ音量になるよう加工した。

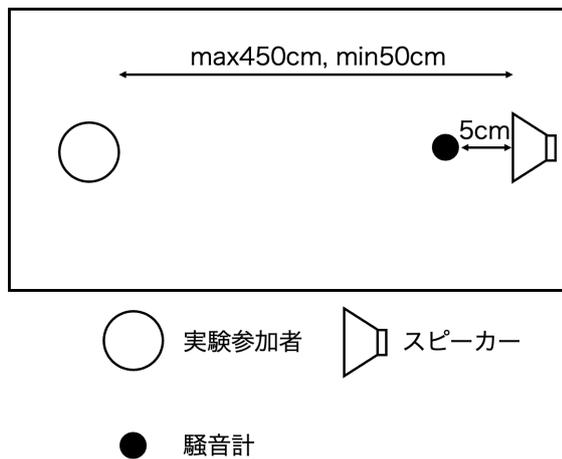


図 3: 最適音量推定モデル作成実験を行った環境。実験参加者とスピーカとの距離を50cm, 100cm, 150cm, 200cm, 250cm, 300cm, 350cm, 400cm, 450cmの9パターンに変えながら、各距離における最適音量を調べた。

3.2 結果と考察

実験参加者は9人であり、測定した最適音量の平均値と標準偏差をプロットした結果を図4に示す。この結果から、最適音量は距離の増加に伴い緩やかに増加していくことがわかった。よって、距離と最適音量の関係は二次曲線近似できると仮定し、最小二乗法を用いて最適音量を V 、距離を d として、最適音量推定モデル(以下の式)を算出した。

$$V = 0.0000345d^2 + 0.0450d + 78.340 \quad (1)$$

また結果から、各距離における最適音量の標準偏差は約5dBの範囲に収まると言える。そこで、2.1節で述べた音量ゲージの最適音量の上下の範囲は、上限5dB、下限5dBとした。

ここで、対話相手にとっての最適音量には個人差があると考えられる。そこで、作成した最適音量推定モデルで、一定の聞こえやすさを担保できているかを追加検討した。本検討では、先の実験でもっともスピーカから離れた450cm地点において、人が最低限聞き取れる最低可聴音量を調査し、その音量がモデルの下限とした音量を下回っていれば、本モデルで一定の聞こえやすさを担保できているものとした。最低可聴音量の調査は参加者3人で行い、参加者はスピーカとの距離450cm地点に立った状態にて、まず参加者が聞き取ることができない音量50dBから、約0.5dBずつ大きくしながら前述の音声測定用音声を再生していった。参加者には都道府県名を聞き取って復唱してもらい、それが再生された都道府県名と一致していた音量を最低可聴音量とした。その結果、参加者全員が約55dBで都道府県名を正しく聞き取れることがわかった。このことから、作成したモデルで対話相手にとっての一定

の聞こえやすさを担保でき、算出した最適声量で操作者が発声したときに、対話相手が聞き取れないといった問題は起こらないと考える。

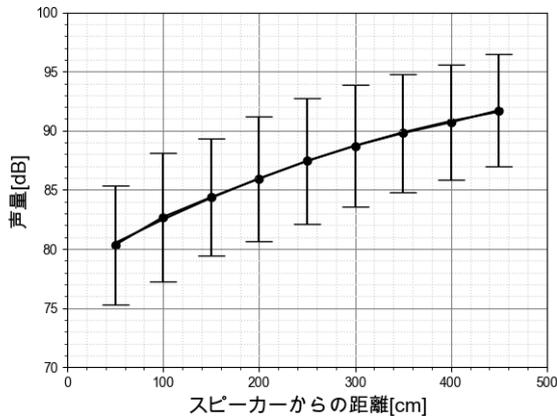


図 4: 測定した最適声量の平均値と標準偏差。

4 声量ゲージの予備的評価実験

本章では、開発した声量ゲージについて、操作者にとっての声量ゲージの使いやすさと、対話相手に聞こえやすく発声できているかについての不安感を調査した予備実験について記す。

4.1 実験方法

本実験では、実験参加者には遠隔操作型対話ロボットの操作者となってもらい、別室にいる実験者とロボットを介して簡単な対話を行う状況で声量ゲージを使用してもらい、評価してもらった。条件は、声量ゲージを使わずに対話してもらう条件（声量ゲージなし条件、図 5a）と声量ゲージを使って対話してもらう条件（声量ゲージあり条件、図 5b）の 2 条件を用意した。声量ゲージの使いやすさについては、あり条件でのみ評価し、声量ゲージにより対話相手に聞こえやすく発声できているかの不安感の評価は、2 条件で比較評価した。実験は被験者内計画とし、2 条件を行う順序はカウンターバランスした。

実験における対話内容は、次のようなものとした。参加者は、2 枚の絵のうちの好きな方を選び、対話相手に聞こえやすいように「私の好きなのは〇〇です」（〇〇は選んだものの名称）と発話する。

1. 実験者が対話相手となり、（ロボットとの距離が 75cm, 200cm, 400cm となる各地点において、）2 枚の絵を提示する。

2. 参加者は、2 枚の絵のうちの好きな方を選び、対話相手に聞こえやすいように「私の好きなのは〇〇です」（〇〇は選んだものの名称）と発話する。
3. この対話を、実験者とロボットとの距離が 75cm, 200cm, 400cm となる 3 地点にて、2 地点間の全移動 6 パターンをすべて含むように 7 回繰り返した⁵。

尚、上記の 3 地点は、Hole の対人距離の 3 分類 [7] に当たるよう決定し、75cm が個人距離、200cm が社会距離、400cm が公共距離に当たる。

実験手順として、まず参加者に対する全体の教示として、操作者の声量と同じ声量がロボットから出力されることと、対話相手にとって聞こえやすい声量で話すようにすることを伝えた。その後、各条件に沿って上記の対話を行い、後述の質問紙に回答してもらった。尚、各条件の開始前には操作画面の見方を説明し、特に声量ゲージあり条件では、ゲージの使い方を説明した。また参加者が慣れるまで、各条件でロボットを介して話す操作練習を行ってもらった。

測定方法としては、各条件終了後に質問紙（7 段階リッカート尺度）で測定した。声量ゲージの使いやすさを調べるためには、Schrepp らが作成した日本語版 User Experience Questionnaire (UEQ) の魅力 (Attractiveness)、効率 (Efficiency)、わかりやすさ (Perspicuity)、信頼性 (Dependability) の 4 項目を用いた [8]。UEQ の各項目を構成する質問を表 1 に示す。声量ゲージの有無による不安感の違いを調べるため、下記の 4 つの質問を用意して、これらの平均値を不安感の評定値とした。下記の質問のうち、1 と 2 は今回の実験のために著者らが作成したもので、3 と 4 は Heerink ら [9] によるエージェントの Anxiety 項目をもとに作成した。

1. 相手にとって、自分の声がうるさすぎたり小さすぎたりしないか不安に感じた
2. 相手にとって、自分の声が聞こえやすいような大きさと話せていたか不安に感じた
3. この操作画面を使わなければならないとしたら、相手に聞こえやすく話すことが出来なさそうで心配だ
4. この操作画面を使わなければならないとしたら、相手を不快にしまいそうで心配だ

4.2 結果と考察

声量ゲージあり条件における使いやすさの結果を図 6 に示す（魅力: Mdn = 4.58, IQR = [4.38, 5.75]; 効

⁵例えばある参加者は、次の順番で各地点で対話した: 75cm → 200cm → 75cm → 400cm → 200cm → 400cm → 75cm。



(a) 声量ゲージなし条件の操作インターフェース



(b) 声量ゲージあり条件の操作インターフェース

図 5: 声量ゲージなし条件の操作インターフェースと声量ゲージあり条件の操作インターフェース。実験者は操作者に 2 つの絵を提示する。

表 1: UEQ における、魅力、効率、わかりやすさ、信頼性の質問。

| 項目 | 質問数 | 質問 |
|--------|-----|-----------------------|
| 魅力 | 6 | 嬉しくない - 嬉しい |
| | | 悪い - 良い |
| | | 嫌いだ - 好きだ |
| | | 楽しくない - 楽しい |
| | | 魅力がない - 魅力がある |
| 効率 | 4 | 遅い - 早い |
| | | 効率が悪い - 効率がいい |
| | | 実用的でない - 実用的だ |
| | | 整理されていない - 整理されている |
| わかりやすさ | 4 | わかりにくい - わかりやすい |
| | | 覚えやすい - 覚えにくい |
| | | 複雑 - 簡単 |
| 信頼性 | 4 | ごちゃごちゃしている - スッキリしている |
| | | 予想がつかない - 予想が付きやすい |
| | | 妨げになる - 助けられる |
| | | 安全ではない - 安全だ |
| | | 期待に合う - 期待に合わない |

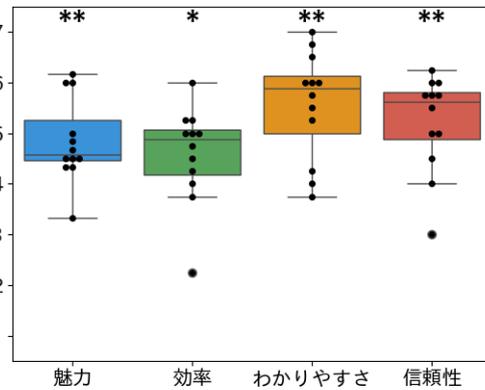


図 6: 声量ゲージあり条件における使いやすさの結果 (* : $p < .05$, ** : $p < .01$)。

率: $Mdn = 4.88$, $IQR = [4.06, 5.19]$; わかりやすさ: $Mdn = 5.88$, $IQR = [4.50, 6.38]$; 信頼性: $Mdn = 5.63$, $IQR = [4.63, 5.94]$ 。7 段階リッカート尺度の 4 (どちらでもない) をチャンスレベルとしたウィルコクソン順位和検定の結果、魅力、効率、わかりやすさ、信頼性の 4 項目全てがチャンスレベルよりも有意に大きかった (魅力: $W = 72$, $p = .0053$; 効率: $W = 54.5$, $p = .0305$; わかりやすさ: $W = 64.5$, $p = .0028$; 信頼性: $W = 63$, $p = .0042$)。

次に操作者の不安感の結果を図 7 に示す (声量ゲージなし条件: $Mdn = 5.38$, $IQR = [5.06, 6.19]$; 声量ゲージあり条件: $Mdn = 4.13$, $IQR = [3.00, 5.31]$)。ウィルコクソン順位和検定の結果、不安感の声量ゲージあり条件の方が声量ゲージなし条件よりも有意に小さかった ($W = 117.5$, $p = .0070$)。

図 6 の結果から、声量ゲージの使いやすさに問題はなかったと考える。これは、声量ゲージをインターフェース上の対話相手の顔の横に表示したことで、参加者は

対話相手の顔を見て対話をしながら声量ゲージを見て声量を調整するというのが無理なくできたことが一因に挙げられると考える。また、最適声量の上下に一定の範囲を設定し、参加者には最適声量の許容幅を与えたことから、参加者は最適声量に厳密に合わせなければならない難しさを感じることなく自身の声量を調整することができたことも影響していると考えられる。

図 7 の結果から、声量ゲージがあることによって、不安感は有意に軽減したことがわかった。声量ゲージのない操作インターフェースでは対話相手との距離感を掴むのが難しいため、参加者は声量ゲージのないインターフェースでは対話相手にとって聞こえやすく発声できているかを判断できず、一定の不安を感じていたと考える。声量ゲージがあることによって、参加者は自分がどれくらいの声量で発声しているのかを判断することができたことができたため、不安感を軽減できたと考える。

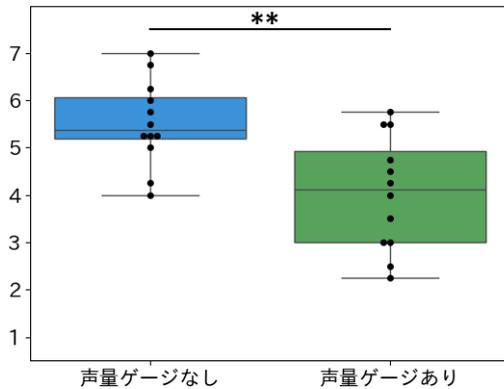


図 7: 不安感の結果。値が大きいほど不安感が高いことを意味する (** : $p < .01$)。

5 まとめ

現状の遠隔操作型対話ロボットの操作インタフェースでは、操作者は対話相手に聞こえやすく話すための発声の調整が難しく、また実際に聞こえやすく話せているか不安を感じてしまう。この問題を解決するため、本稿では操作者に発声調整を促す音量ゲージを提案し開発した。開発した音量ゲージについて、操作者にとっての使いやすさと、対話相手に聞こえやすく話せているかの不安感に着目した予備実験を行った結果、ゲージの使いやすさに問題はなく、また操作者の不安感を有意に軽減できることを確認した。以上のことから、従来の操作インタフェースにおける、操作者が対話相手に聞こえやすく話せているか不安に感じてしまうという問題は、音量ゲージを用いることで改善できると考える。

今後は、音量ゲージを用いることで自然な対話を実現できるかを検証するために、音量ゲージを使ったときに操作者の声量と声質がどのように変化したか、また対話相手にとって操作者の声が聞こえやすく自然だったかどうかについて調査していく。

謝辞

本研究は、JST ムーンショット型研究開発事業 JP-MJMS2011 (基礎研究の役割)、牧誠記念研究助成 (システム開発の役割) の支援を受けたものです。

参考文献

[1] Koceska, N. and Koceski, S.: Supporting elderly's independent living with a mobile robot platform,

Vol. 28, No. 5 (2022).

[2] Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., and Hagita, N.: A communication robot in a shopping mall, *IEEE Transactions on Robotics*, Vol. 26, No. 5, pp. 897–913 (2010).

[3] Claudio, G., Luca, G., and Luce, L. M.: Interaction design for cultural heritage. A robotic cultural game for visiting the museum's inaccessible areas., *The Design Journal*, Vol. 20, No. sup1, pp. S3925–S3934 (2017).

[4] Pelegrín-García, D., Smits, B., Brunskog, J., and Jeong, C.-H.: Vocal effort with changing talker-to-listener distance in different acoustic environments, *The Journal of the Acoustical Society of America*, Vol. 129, No. 4, pp. 1981–1990 (2011).

[5] Hayamizu, A., Imai, M., Nakamura, K., and Nakadai, K.: Volume adaptation and visualization by modeling the volume level in noisy environments for telepresence system, in *Proceedings of the second international conference on Human-agent interaction*, (2014).

[6] Rueben, M., Syed, M., London, E., Camarena, M., Shin, E., Zhang, Y., Wang, T. S., Groechel, T. R., Lee, R., and Matarić, M. J.: Long-term, in-the-wild study of feedback about speech intelligibility for k-12 students attending class via a telepresence robot, in *Proceedings of the 2021 International Conference on Multimodal Interaction*, (2021).

[7] Hall, E. T.: *The hidden dimension*, Vol. 609, Anchor (1966).

[8] Schrepp, M., Hinderks, A., and Thomaschewski, J.: Applying the user experience questionnaire (UEQ) in different evaluation scenarios, in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3*, Springer (2014).

[9] Heerink, M., Kröse, B., Evers, V., and Wielinga, B.: Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model, *International Journal of Social Robotics*, Vol. 2, No. 4, pp. 361–375 (2010).