

半自律テレプレゼンスロボットの不快刺激の学習と回避行動の表出

Learning of Unpleasant Stimuli and Expressing Avoiding Behavior of Semi-Autonomous Telepresence Robot

大澤 正彦 *1*2

Masahiko Osawa

今井 倫太 *1

Michita Imai

*1 慶應義塾大学大学院理工学研究科

Keio University, Graduate School of Science and Technology

*2 日本学術振興会 特別研究員

Research Fellow of Japan Society for the Promotion of Science

テレプレゼンスロボットの発展により没入感の高いシステムが出現しているが、没入感が高い操作環境であるがゆえに遠隔操作者に不快な感覚も提示してしまう状況も発生している。本論文では、バイノーラルマイクと Head Mount Display (HMD) を利用し遠隔操作者の社会的存在感を高めるように工夫した手持ち型テレプレゼンスロボットを作成し、テレプレゼンスロボットを用いたコミュニケーションにおける不快刺激回避の手法を提案する。手持ち型テレプレゼンスロボットのユーザビリティに関する調査を行ったところ、バイノーラルマイクを用いたことによる不快な音刺激が増加することが示唆された。そこで操作者が感じる不快感覚の回避並びに、不快感覚を回避していることを伝える回避行動を実現するために、手持ち型テレプレゼンスロボットに対して Echo State Network (ESN) を用いた音の不快感を検出する機能と、検出した不快感をもとに自律的な回避行動の表出を調停する Accumulator モデルを実装した。結果、手持ち型テレプレゼンスロボットは遠隔操作者が不快な思いをする前に不快刺激を避けることができるため、ロボットの周囲の人に不快刺激を回避していることを示しつつ、より不快感の少ない快適なコミュニケーション環境を遠隔操作者に提供できる。

1. はじめに

遠隔コミュニケーションにおいて相手と対面に対話している感覚の源である社会的存在感 [Nakanishi 09, De Greef 01] を高く保つことは重要である。例えば遠隔コミュニケーションに音声だけでなく映像も利用することによって社会的存在感を向上させることが明らかになりつつある [De Greef 01, Isaacs 94]。しかし対面した場合と比較すると十分な存在感を提示できていない [Jancke 01, Fish 90]。そこで遠隔コミュニケーションを目的としたテレプレゼンスロボットが注目されている。

テレプレゼンスロボットを用いた遠隔コミュニケーションによって社会的存在感を向上させるために、遠隔操作者が実際に会話相手と同じ環境にいるようリアルな感覚を提示することは重要である。Toshima らが提案しているテレヘッド [Toshima 14] は遠隔操作者と同じ形状に設計されたダミーヘッドを用いることで、よりリアルな音体験をすることができる。Tachi らの研究では TELESAR というロボットの開発が進んでおり、映像や音声だけでなく触覚や温冷感をも遠隔操作者に提示する [Fernando 12]。

既存研究 [Toshima 14, Fernando 12] で挙げられている特殊なデバイスを用いて遠隔操作者の没入感や臨場感を高める研究は、用いたデバイスによって生じるデメリットに関して考察されていない場合が多い。例えばバイノーラル録音と呼ばれる方法で集音し遠隔操作者に音を提示する場合には、ロボットと直接コミュニケーションする会話相手が遠隔操作者に提示される音を考慮し適切な音量や距離感で発話する必要がある。しかし、遠隔操作者の快適さを保つ手法には未着手である。

本論文では、バイノーラルマイクを用いて遠隔操作者の没入感や臨場感を高めた手持ち型テレプレゼンスロボットをテストベッドとして構築し、遠隔操作者が感じる不快刺激の回避方法について研究する。手持ち型テレプレゼンスロボット自体は自

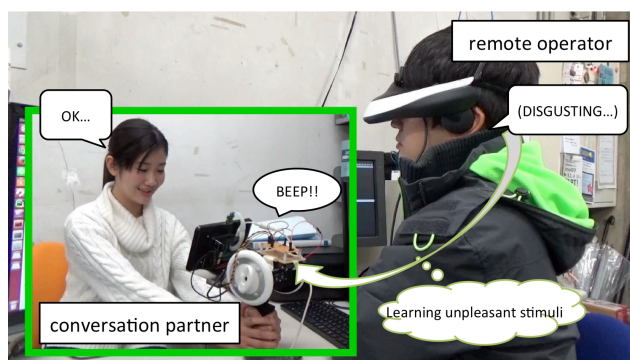


図 1: 不快刺激を検出して自律的に回避行動を表出する手持ち型テレプレゼンスロボット

ら移動する機能を持たず、会話相手が手で持ってコミュニケーションをとるデザインとなっており、コミュニケーション相手の意思でロボットとの会話距離を調整できる。またバイノーラルマイクにより、遠隔操作者はロボットの周囲の音をリアルティ高く聞くことができる。会話距離が会話相手の支配下にあることで、会話相手由来の高臨場感が与えるデメリットを引き起こしやすくなっている。

手持ち型テレプレゼンスロボットのユーザビリティの調査を行い、発見された遠隔操作者が感じる不快刺激に対して自律的に回避行動を表出する仕組みを組み込んだ。手持ち型テレプレゼンスロボットに対して、Echo State Network (ESN) を用いた音の不快感を検出する機能と、検出した不快感をもとに自律的な回避行動の表出を調停する Accumulator モデルを実装した。不快音の遮断と同時に、ロボットのディスプレイ部の首ふり動作およびビープ音によって会話相手に不快感を伝えることができる。遠隔操作者が、操作することなしに不快音を回避できるので、操作者が操作を開始するまでに不快音を聞くことをなくすることができる。

連絡先: 大澤 正彦, 慶應義塾大学大学院理工学研究科/日本学術振興会 特別研究員, 神奈川県横浜市港北区日吉 3-14-1 慶應義塾大学理工学部情報工学科 今井研究室, TEL:045-566-1749, Fax:045-566-1749, mosawa@ailab.ics.keio.ac.jp

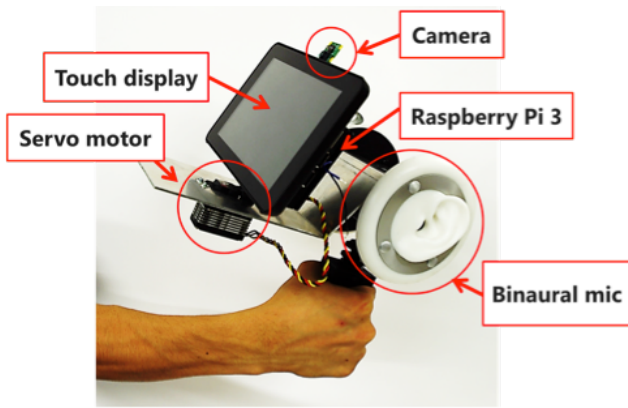


図 2: 手持ち型テレプレゼンスロボットの外観

2. 手持ち型テレプレゼンスロボットとユーザビリティの調査

2.1 手持ち型テレプレゼンスロボット

手持ち型テレプレゼンスロボットの外観を図 2 に示す。ロボットは Raspberry Pi 3 によって制御される。Raspberry Pi 3 には専用カメラモジュール、専用タッチディスプレイモジュール、USB Audio Interface、GPIO からサーボモータがそれぞれ接続されている。さらに USB Audio Interface には Binaural マイクである 3Dio Free Space Binaural Microphone が接続されており、遠隔操作者はロボットが取得した情報を知覚することで、没入感や臨場感が促進されるように工夫されている。

ロボットは移動のためのアクチュエータを持たない。ロボットとインタラクションする会話相手は、バイノーラルマイクに接続されているハンドグリップを握ってコミュニケーションをする。ロボットとの距離が会話相手によって支配されているので、会話相手由来の臨場感に関連するデメリットが生じやすい構造となっている。手持ちの構造にすることで、会話相手が引き起こしやすい不快刺激の詳細な検討がしやすいデバイスとなっている。ロボットや遠隔操作者の内部状態を表出するために利用できるサーボモータを 1 つ搭載している。サーボモータは Raspberry Pi 3、タッチディスプレイ、カメラを一体化したケースに接続されており、ケースごと左右に回転させることができる。

2.2 映像視聴によるユーザビリティと不快刺激の考察

2.2.1 刺激映像と実験条件

事前に手持ち型テレプレゼンスロボットを利用した自由会話による予備実験を行い、遠隔操作者が不快に感じる刺激を 15 項目選出した。さらに不快刺激を含む約 1 分間のシナリオを作成した。女性話者 2 名と男性話者 1 名に 2 種類の設置環境条件 (持ち型条件・置き型条件) と 2 種類の音環境条件 (バイノーラル条件・モノラル条件) の組み合わせの 4 つの条件下においてシナリオを行ってもらい、録画した。録画された映像の中から不快刺激が最も強いと思われる男性話者による映像を採用した。

2.2.2 実験参加者

20 代の男性 10 名と女性 1 名の実験参加者に録画した映像を視聴してもらい、実験終了後アンケートによる主観評価を行った。

2.2.3 予測される結果

結果として予想されるのは、以下の 2 点である。

- バイノーラル条件の場合のほうが臨場感や没入感が向上し、ユーザの満足度は向上する一方で、マイクを触られたり、大声を出されたりした場合に不快に感じる場合が増加すること
- 持ち型条件では一方的に映像を視聴する場合に関しては、手振れや会話相手の移動にともなう大きな揺れを遠隔操作者が不快に感じること。

2.2.4 実験結果

図 3(a) はバイノーラル条件とモノラル条件を比較した際の主観評価実験の結果である。楽しさと臨場感において有意差が認められたほか、またつかいたいという項目に関しても優位傾向が認められた。一方図 3 (b) は持ち型条件と置き型条件で比較した場合の結果であり、映像のみやすさという項目で有意差が認められた。一方、交互作用に関して検定を行ったところ、有意差が認められた項目はなかった。

次に図 4 に遠隔操作者が不快に感じた刺激の数に関する結果を示す。バイノーラル条件とモノラル条件を比較した際には、バイノーラル条件のほうが不快に感じる刺激が多い傾向が示唆されているが、持ち型条件と置き型条件を比較した際には有意差は認められなかった。また、交互作用についても有意差は認められなかった。したがって、映像の視聴者の観点の場合に関しては、バイノーラルマイクの影響が大きいが示唆されている。

最後に図 5 に遠隔操作者が刺激を不快であると報告した割合について示す。遠隔操作者は映像に関する刺激よりも、音に関する刺激を不快と報告する割合が高く、特に過半数の被験者が不快と報告した刺激は「マイクに指を入れられる」、「息を吹きかけられる」、「大声を出される」の 3 項目であった。

2.2.5 考察

バイノーラルマイクを用いることで、ユーザの満足度は向上する一方で、不快に感じる刺激が多いことが示唆された。一方で持ち型にすることででも不快に感じる刺激が増大すると予想されたが、今回の実験条件のように映像を視聴する条件では有意な差が認められなかった。そこで遠隔操作者が不快と感じやすい音を学習し、自動的に検出・回避するシステムは有効であると考え、実装することにした。

3. 不快刺激の学習と回避行動の表出

本章では、前章で明らかになったバイノーラルマイクを用いることに起因する手持ち型テレプレゼンスロボットのデメリットを解消するために作成した、不快な音刺激の検出器と不快刺激を検出した際に表出する回避行動について説明する。

3.1 不快刺激の学習

はじめに不快刺激を学習するために作成したデータセットの詳細について説明する。ケーススタディの結果から遠隔操作者の過半数が不快と報告した「マイクに指を入れられる」、「息を吹きかけられる」、「大声を出される」の 3 種類の刺激を不快刺激とした。2 人の話者によって 3 種類の不快刺激を左右のマイクに対してそれぞれ 1 分間ずつ提示し録音した。一方不快でないデータセットとして利用するために、走れメロスの冒頭部の約 2.5 分を 2 人の話者によって朗読し、録音した。

次に検出器の構成について述べる。検出器は 2 つのモジュールから構成される。1 つめは時系列データを扱うニューラルネットワークとして頻繁に用いられる Echo State Network (ESN) である。ESN はランダムな結合から構成される中間層を持ち、

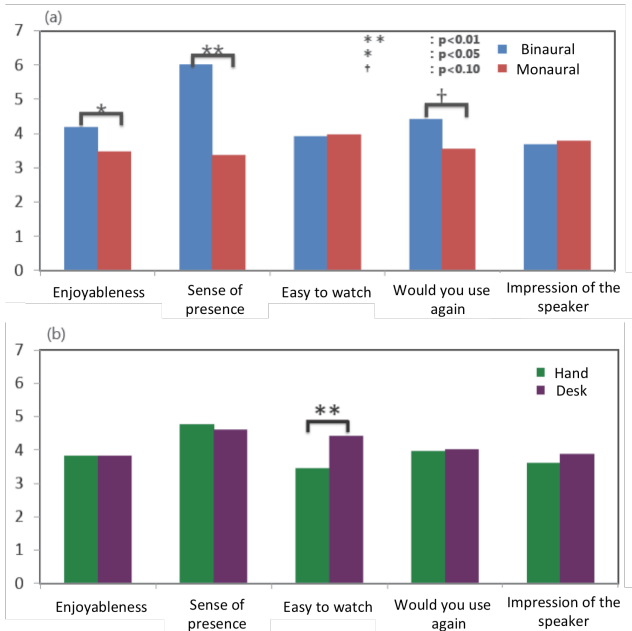


図 3: システムに関する評価. (a) バイノーラル VS モノラル (b) 持ち型 VS 置き型

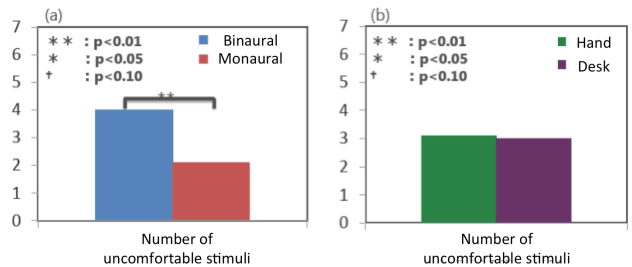


図 4: 不快に感じた刺激の数. (a) バイノーラル VS モノラル (b) 持ち型 VS 置き型

中間層と出力層の間の結合のみが線形回帰によって決定されるため比較的高速にパラメータが決定される。入力を 32 次元、中間層を 300 次元、出力を不快度に対応する 1 次元にそれぞれ設定した。入力層と中間層との結合確率を 3%, 中間層内の結合確率は 10%にそれぞれ設定し、各結合は-0.5 から 0.5 の間の一様乱数によって決定した。

2 つめのモジュールは ESN の出力を調停する Accumulator モデル [大澤 16] である。Accumulator モデルは、以下の式であらわされる累積証拠 A^t が閾値を超えた場合に対応する行動を表出させる。

$$A^t = rA^{t-1} + y^t, \quad (1)$$

ここで r は割引率、 y^t は ESN の出力をそれぞれ表す。また、 A^t が閾値を超えた場合には A^t を 0 にリセットすることで、同じ行動が連続して出力され続けることを防ぐことができる。本システムにおいては閾値を 1.5、割引率を 0.3 とし、 A^t が閾値を超えた場合に 3.2 で説明する回避行動を表出する。

3.2 回避行動の表出

回避行動は 2 種類の振る舞いを同時に提示する。1 つめの回避行動は小松らの研究 [Komatsu 07] で報告されている、人工物がネガティブな内部状態を表していると人間が感じやすいピーブ音である。[Komatsu 07] においてすべての実験参加者

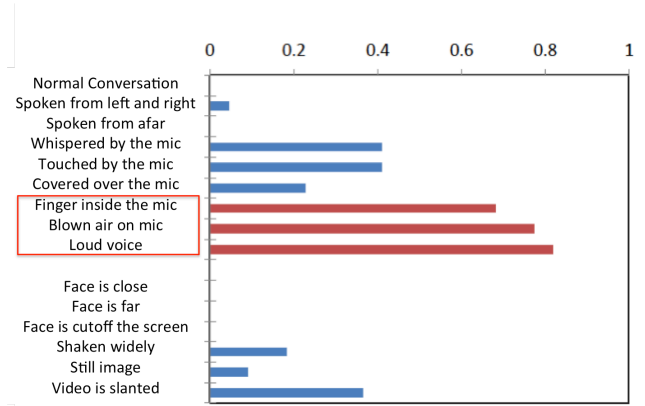


図 5: 刺激が不快に感じられた確率

がネガティブな内部状態を表していると感じたものの中から、今回は表出時間が 418ms, 周波数が 256Hz から 6Hz に一様効果するものを採用した。2 つめの回避行動はディスプレイの回転である。回転する角度は正面を向いている状態を基準におよそ 30 度程度であり、動作開始から停止までの時間は約 1.5 秒である。また、回避行動を表出すると同時に遠隔操作者に対する音声ストリーミング再生を停止する。

4. 評価実験

本章では、章で説明した自律的な機能が遠隔操作者の感じる不快感への影響を調べることを目的として行った評価実験について説明する。

4.1 実験条件

下記に示す 3 つのシステムを用いて、3 種類の不快刺激を提示した場合に遠隔操作者が感じる不快感の程度を調べた。

Automatic 自律的な回避行動を表出するシステム (提案)

Manual 操作者が自ら回避行動を表出するシステム

Immovable 回避行動を表出しないシステム

一方提示した不快刺激は、ケーススタディの結果から過半数の遠隔操作者が不快と報告した 3 つの刺激である。テレプレゼンスロボットは会話相手として参加する実験者とともに研究室の学生居室に設置した。一方実験参加者はテレプレゼンスロボットの遠隔操作者として別室でコミュニケーションをとる。会話相手となる実験者は 3 つのシステムそれぞれに対して、不快刺激を提示した。なお、システムと刺激はいずれも実験参加者ごとに順序はランダムとした。

4.2 実験参加者と評価方法

実験参加者は 20 代の男性 10 名と女性 2 名である。各システムに関して実験を行うたびにアンケート調査による主観評価を行い、すべての実験終了後にインタビューを行った。評価項目はシステムに関する 5 項目について 7 段階評価と、刺激の不快感に関する 3 項目について 7 段階評価である。

4.3 予測される結果

実験結果として予想されるのは下記の 2 つである。

- 回避行動が適切である場合、回避行動を表出する 2 つのシステムは、回避行動を表出しないシステムと比較して高い満足度と少ない不快感

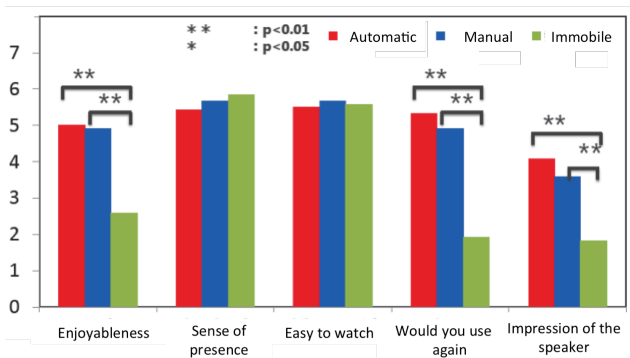


図 6: Results of subjective evaluation.

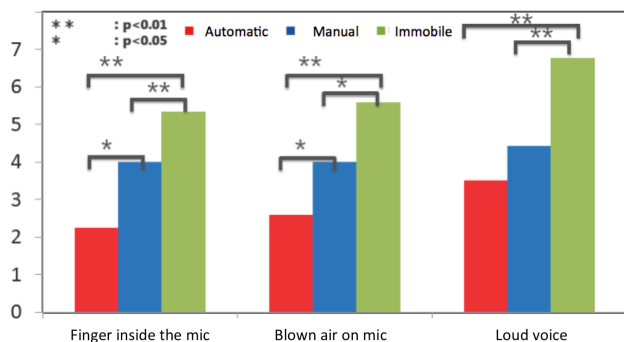


図 7: Comparison of number of stimuli the remote operators felt were uncomfortable.

- 不快刺激の検出精度が十分である場合、手動で回避行動を表出する場合と同等のシステムに関する評価や不快感

4.4 実験結果

図 6 にシステムの評価に関する結果を示す。各項目に関して分散分析を行ったところ、楽しさ、また使いたいか、話者に対する印象の 3 項目において、5%有意水準で有意差が認められたため、多重の t 検定を行った。t 検定の結果回避行動を表出するシステムとしないシステムとのすべての組において 1%有意水準で有意差が認められた。図 7 に不快感に関する結果を示す。分散分析の結果すべての群に関して 5%有意水準で有意差が認められた。多重 t 検定の結果、図中に示すように多くのシステム間で有意差が認められた。

4.5 考察

図 6 に示した結果から、回避行動をとるシステムと回避行動をとらないシステムとの間に有意な差が認められたことから、回避行動をとることの有意性が示唆されており、仮説ともよく合致する結果となった。次に、不快感に関する 3 項目のうち 2 項目において手動で回避行動をとるよりも自動的に回避行動をとる方が感じる不快感が優位に少ないという結果になり、当初の仮説には反している。自律的な回避行動が不快感を軽減させた要因として、手動で回避行動を表出する場合には遠隔操作者が不快な刺激を感じてから回避行動を表出するが、自動のシステムの場合には遠隔操作者が不快刺激を感じるよりも先に回避行動が表出される場合があり、複数の遠隔操作者が不快刺激が提示されたことを気づいていなかったことがインタビューの結果から明らかになった。

5. おわりに

本研究では、没入感を高めたテレプレゼンスロボットのデメリットに着目し、ケーススタディで明らかになった不快な音刺激を学習し自律的に回避行動を表出するシステムを作成した。今回作成したシステムは事前にはカテゴライズされた一般性のある刺激を扱ったが、今後は動的にシステムが遠隔操作者にパーソナライズし、適応していくシステムも重要であると考えられる。

参考文献

- [De Greef 01] De Greef, P. and Ijsselstein, W. A.: Social presence in a home tele-application, *CyberPsychology & Behavior*, Vol. 4, No. 2, pp. 307–315 (2001)
- [Fernando 12] Fernando, C. L., Furukawa, M., Kurogi, T., Kamuro, S., Minamizawa, K., Tachi, S., et al.: Design of TELESAR V for transferring bodily consciousness in teleexistence, in *Intelligent Robots and Systems, International Conference on*, pp. 5112–5118IEEE (2012)
- [Fish 90] Fish, R. S., Kraut, R. E., and Chalfonte, B. L.: The VideoWindow system in informal communication, in *Proceedings of the conference on Computer-supported cooperative work*, pp. 1–11ACM (1990)
- [Isaacs 94] Isaacs, E. A. and Tang, J. C.: What video can and cannot do for collaboration: a case study, *Multimedia Systems*, Vol. 2, No. 2, pp. 63–73 (1994)
- [Jancke 01] Jancke, G., Venolia, G. D., Grudin, J., Cadiz, J. J., and Gupta, A.: Linking public spaces: technical and social issues, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 530–537ACM (2001)
- [Komatsu 07] Komatsu, T. and Yamada, S.: How do robotic agents' appearances affect people's interpretations of the agents' attitudes?, in *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pp. 2519–2524ACM (2007)
- [Nakanishi 09] Nakanishi, H., Murakami, Y., and Kato, K.: Movable cameras enhance social telepresence in media spaces, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 433–442ACM (2009)
- [Toshima 14] Toshima, I. and Aoki, S.: Perception of delay time of head movement in using an acoustical telepresence robot: TeleHead, *Advanced Robotics*, Vol. 28, No. 15, pp. 997–1004 (2014)
- [大澤 16] 大澤 正彦, 芦原 佑太, 島田 大樹, 今井 倫太, 栗原 聡: 前頭前野 Accumulator を用いた複数の機械学習器の調停と認知アーキテクチャへの応用, 汎用人工知能研究会人工知能学会 (2016)