

Melodic Emotional Expression: 音楽重畳による AI アシスタントの社会的存在感の強化

高橋ともみ^{†1} 田中一晶^{†1} 小林賢一郎^{†2} 岡夏樹^{†1}

概要: AI スピーカ等のエージェントの音声操作に抵抗を感じる日本人は多く、普及の歯止めとなっている。この問題は、エージェントの感情表現が希薄であるために、感情が意図通りに伝わりにくく社会的存在感が低下してしまうことに起因すると考えた。まずこの前提を検証した結果、エージェントの音声として用いられることの多い平坦な合成音声では、適切に感情表現を行わないと感情や意図が多義的に捉えられ、社会的存在感が低下することが分かった。そこで本研究では、エージェントが伝達すべき感情に合致した BGM (Background Music)、効果音、擬態語を合成音声に付与して感情表現を聴覚的に強調する手法を提案し、その有効性を検証した。その結果、合成音声に BGM を付与することで感情を意図通りに伝達できるだけでなく、エージェントの人間らしさや話しかけやすさを向上させる効果が見られ、提案手法によって社会的存在感を強化できる可能性が示された。

キーワード: 感情表現, Background Music, 効果音, 擬態語, 対話エージェント

Melodic Emotional Expression: Melody Superposition Enhances Social Presence of an AI Assistant

TOMOMI TAKAHASHI^{†1} KAZUAKI TANAKA^{†1}
KENICHIRO KOBAYASHI^{†2} NATSUKI OKA^{†1}

1. はじめに

日常生活において AI スピーカや AI アシスタント等の対話型 AI エージェントがみられるようになってきた。しかしながら、KDDI 株式会社の「日本人の音声操作に対する意識調査 2017」[1]によると、7 割を超える日本人が音声操作に抵抗感を感じていることが分かっており、その抵抗感がエージェントの普及を妨げる要因となっている可能性が考えられる。そこで本研究では、この音声操作への抵抗感を低減することを目指す。

人がエージェントへ話しかける際に抵抗感が生じるのは、社会的存在感が低く、エージェントが人から社会的に接してもらいにくい[2][3]ことが起因していると考えた。社会的存在感とは人がエージェントから人間のような性質を感じ取る度合いであり、ヒューマンエージェントインタラクションの分野において古くから社会的存在感を強化する手法が研究されてきた[4][5][6][7][8][9]。その基本的な手法はエージェントの外見や振る舞いを人間らしくすることである。しかしながら、対話型 AI エージェントはユーザが運転中や料理中等の作業中であっても支援できることが利点であり、そのようなエージェントに視覚的な工夫を適用することは必ずしも適切とは言えない。

そこで、我々は聴覚的に人間らしさを向上させる可能性

として、エージェントの感情表現に着目した。既存の AI スピーカ等のエージェントには平坦な合成音声を用いられていることが多く、感情表現が乏しいことがエージェントの社会的存在感を低下させる一因となっていることが考えられる。合成音声の韻律を調整して感情を表現する技術(以降、感情的な合成音声)も存在するが、メラビアンの実験結果[10]から感情的な合成音声のみではエージェントの感情表現としては不十分であると考えた。これを改善するため、本研究ではエージェントの聴覚的な感情表現として音情報 (BGM, 効果音, 擬態語) を合成音声に重畳して再生する方法を提案し、その効果を検証する。

2. 音情報重畳手法の効果検証

提案手法の効果を検証するために実験を行った。実験では、AI スピーカが実験協力者に対し、メールを受信したことを通知する場面を想定した。なお実験協力者には、AI がメールの内容に応じて通知を行っていることを教示をした。実験は図 1 に示す環境で行った。本来であれば AI スピーカから直接音声流れることが望ましいが、イベント会場の一角で行ったため雑音への対策として、実験協力者にはヘッドフォンをつけて通知を聴いてもらった。

実験手順は次の通りである。

1. 「メールが一通届きました」という合成音声のみを再生
2. 「メールが一通届きました」という合成音声に音情報を付与して再生
3. 2 種類の通知音声に対する印象評価アンケート

^{†1} 京都工芸繊維大学
Kyoto Institute of Technology
^{†2} TIS 株式会社
TIS Inc.

上記の実験を、合成音声の種類（平坦／感情的）と音情報の種類（BGM／効果音／擬態語）を組み合わせで行った。さらに、表現する感情をポジティブ／ネガティブの2通りで行った。したがって、 $2 \times 3 \times 2$ の合計 12 条件で実験を行った。合成音声には VOICEROID2 結月 ゆかり [a] を使用した。



図 1 実験環境

合成音声への 3 種類の音情報の付与方法についてそれぞれ述べる。BGM の場合は、合成音声開始の直前からフェードインし、合成音声の終了の直後にフェードアウトさせた。効果音の場合は、合成音声の開始直前に再生し、合成音声の開始時には終了するようにした。これらの 2 種類の音情報に使用した音源は、いずれもインターネット上にアップロードされたフリー素材を使用した。なお、使用した音源のポジティブ／ネガティブの評価についてはあらかじめ簡単な印象調査を行った。最後に、擬態語の場合は、ポジティブな感情を表現する際には「にこにこ」、ネガティブな感情を表現する際には「しょぼん」という擬態語を、合成音声の再生直後に再生した。

アンケート項目は下記の通りである。2 種類の通知のそれぞれについて評価してもらった。

1. 通知されたメールの内容について (7 段階 SD)
 [ネガティブなメール - ポジティブなメール]
2. 通知をした AI について (7 段階 SD)
 [機械的 - 人間らしい]

3. 結果

印象評価アンケートは、2 節に記載した形容詞対のうち左側を 1、右側を 7 として点数化した。このデータに対し、音情報ごとに 2 要因分散分析を行った。

3.1 ポジティブな感情の表現

まず、図 2 は音情報として BGM を用いた場合の結果である。メールの内容 ($F(1, 18) = 100.660, p < .001$)、人間ら

しさ ($F(1, 18) = 39.710, p < .001$)、話しかけやすさ ($F(1, 18) = 25.462, p < .001$) の全項目において、BGM の主効果が見られた。メールの内容については、合成音声要因の主効果も有意傾向であった ($F(1, 18) = 3.176, p = .092$)。したがって、音情報として BGM を用いれば、合成音声の種類を問わず、感情の伝わりやすさ、人間らしさ、話しかけやすさを向上させられることがわかった。

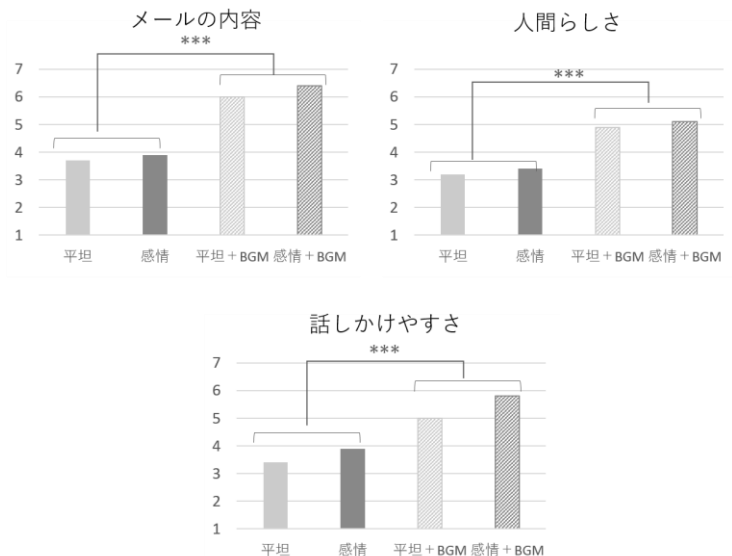


図 2 ポジティブな BGM を重畳した場合の結果

次に、音情報として効果音を用いた場合の結果を述べる。メールの内容 ($F(1, 18) = 59.648, p < .001$)、話しかけやすさ ($F(1, 18) = 5.345, p < .01$) の項目において、効果音の主効果が見られた。人間らしさについては交互作用が有意傾向であったため ($F(1, 18) = 3.927, p = .063$)、TukeyHSD による下位検定を行ったところ、合成音声が平坦である場合のみ効果音により有意に人間らしさが向上していた ($p < .05$)。したがって、音情報として効果音を用いれば、合成音声の種類を問わず、感情の伝わりやすさ、話しかけやすさを向上させられることがわかった。また、合成音声が平坦な場合には効果音を重畳することで人間らしさを向上させることが可能であった。

さらに、音情報として擬態語を用いた場合の結果を述べる。メールの内容において擬態語の主効果が見られた ($F(1, 18) = 10.055, p < .01$) が、そのほかには有意な結果は見られなかった。したがって、音情報として擬態語を用いると、合成音声の種類を問わず、感情の伝わりやすさを向上させられることがわかった。

3.2 ネガティブな感情の表現

まず、音情報として BGM を用いた場合の結果を述べる。メールの内容においてのみ、BGM の主効果と ($F(1, 18) = 13.926, p < .01$)、合成音声の主効果が見られた ($F(1, 18) = 8.901, p < .01$)。したがって、音情報として BGM を用いれば、合成音声の種類を問わず感情の伝わりやすさを向上さ

a) 株式会社エーアイが開発し、株式会社 AHS が個人向け製品にパッケージングした入力文字読み上げソフトウェア。

せられることがわかった。また、BGMの有無を問わず合成音声を感情的にした場合にも感情の伝わりやすさを向上させられることがわかった。

次に、音情報として効果音を用いた場合について述べる。メールの内容においてのみ交互作用が見られたため ($F(1, 18) = 8.727, p < .01$), TukeyHSDによる下位検定を行ったところ図3に示す結果が得られた。したがって、音情報として効果音を用いれば、合成音声の種類を問わず感情の伝わりやすさを向上させられることがわかった。また、効果音を付加しない場合には合成音声を感情的にすることも感情の伝わりやすさが向上することがわかった。

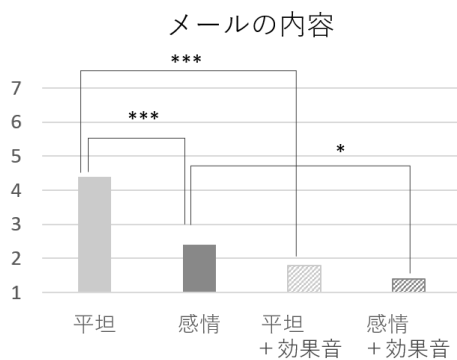


図3 ネガティブな効果音を重畳した場合の結果

さらに、音情報として擬態語を用いた場合の結果を述べる。メールの内容において擬態語の主効果が有意であり ($F(1, 18) = 81, p < .001$), 合成音声の主効果が有意傾向であった ($F(1, 18) = 3.491, p = .078$)。また、話しかけやすさにおいても擬態語の主効果が見られた ($F(1, 18) = 4.455, p < .05$)。したがって、音情報として擬態語を用いれば、合成音声の種類を問わず、感情の伝わりやすさ、話しかけやすさを向上させられることがわかった。また、擬態語の有無を問わず、合成音声を感情的にした場合にも内容の伝わりやすさの向上には効果があることがわかった。

4. 考察

前章で得られた結果より、ポジティブな感情を表現する場合には、音情報を重畳した感情表現手法が有効であると考えられる。特に、感情の伝わりやすさ、人間らしさ、話しかけやすさのすべての項目を向上させられたBGMを用いることが、感情表現の観点のみならず社会的存在感の観点からも適切であるといえる。

一方、ネガティブな感情を表現した場合は、感情の伝わりやすさについてはすべての音情報において重畳手法に効果が見られたが、人間らしさや話しかけやすさを向上させる効果はほとんど見られなかった。アンケート後に行った簡単なインタビューでは、ネガティブなメールの場合は通知までネガティブにしてほしくないという意見が多く聞かれたことから、ネガティブな感情表現を伴う通知を行うと

いうタスク自体が話しかけやすさに悪影響を及ぼしたことが考えられる。唯一話しかけやすさを向上させることが出来た擬態語は、茶化しているような感じがしたという意見が得られており、ネガティブな内容であることを伝えつつも深刻さを感じさせないことで話しかけやすい印象を与えることが出来たと考えられる。したがって、ネガティブな感情を表現する際には、用途に応じて適切な音情報を使い分けることが望ましいといえる。

今回は適切と思われる音情報を実験者があらかじめ用意したが、今後は文章や状況から適切な音情報を生成する手法を検討したい。

謝辞 貴重なデータ収集の機会や AI スピーカのご提供等、本研究にご協力下さいました TIS 株式会社 織田村明雄様に深く感謝致します。

参考文献

- [1] “日本人の音声操作に対する意識調査 2017”, <http://news.kddi.com/kddi/corporate/newsrelease/2017/10/05/2726.html>, (最終検索日: 2019年5月28日)
- [2] M.K. Lee, U. States, S. Kiesler, J. Forlizzi, and P. Rybski. Ripple Effects of an Embedded Social Agent: A Field Study of a Social Robot in the Workplace. Proc.CHI2012, pp.695–704, 2012.
- [3] K. Tanaka, N. Yamada, H. Nakanishi, and H. Ishiguro. Teleoperated or Autonomous?: How to Produce a Robot Operator’s Pseudo Presence in HRI. Proc. HRI2016, pp.133-140, 2016.
- [4] N. Yee, J.N. Bailenson, and K. Rickertsen. A Meta-analysis of the Impact of the Inclusion and Realism of Human-like Faces on User Experiences in Interfaces. Proc.CHI2007, pp.1–10, 2007.
- [5] A. Zarak, D. Mazzei, M. Giuliani, and D.D. Rossi. Designing and Evaluating a Social Gaze-control System for a Humanoid Robot. IEEE Transactions on Human-Machine Systems, vol.44, no.2, pp.157–168, 2014.
- [6] N. Koyama, K. Tanaka, K. Ogawa, and H. Ishiguro. Emotional or Social?: How to Enhance Human-robot Social Bonding. Proc. HAI2017, pp.203–211, 2017.
- [7] C. Breazeal. Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies, vol.59, no.1-2, pp.119–155, 2003.
- [8] D. Cameron, S. Fernando, E. Collins, A. Millings, R. Moore, A. Sharkey, V. Evers, and T. Prescott. Presence of Life-like Robot Expressions Influences Children’s Enjoyment of Human-robot Interactions in the Field. Proc. AISB2015, pp.36-41, 2015.
- [9] A. Pereira, R. Prada, and A. Paiva. Improving Social Presence in Human-agent Interaction. Proc. CHI2014, pp.1449–1458, 2014.
- [10] A. Mehrabian. Nonverbal Communication. Aldine Transaction, 1972.