

多値確信度伝達のためのエージェントによる集団音声表出

Expression of Agent's Collective Voices to Convey Multi-Level Confidence

本戸丈裕^{1*} 小林一樹²
Takehiro Hondo¹ Kazuki Kobayashi²

¹ 信州大学大学院

¹ Graduate School of Science and Technology, Shinshu University

² 信州大学 学術研究院

² Academic Assembly, Shinshu University

Abstract: 本研究では、エージェントによるシンプルかつ直観的な多値確信度の伝達手法を提案する。これまで、シンプルかつ直観的な確信度の表出方法として Artificial Subtle Expression が提案されているが、2つの状態の表出を前提としたものであった。そこで提案手法では、エージェントの音声を重ね合わせた集団音声によって多値の確信度伝達を目指す。エージェントが3つの選択肢のうち1つを予想する経路選択ゲームを用いて参加者実験を行い、提案手法により多値の確信度が伝達可能であるかを調査した。参加者実験の結果、人数が多い音声の方がエージェントの提案を拒否されにくいことが示唆された。

1 はじめに

一般生活のIoT化により、Google社のGoogle HomeやiRobot社のRoombaといった機器が生活に溶け込むようになった。こういった機器の使用時には人間の命令を入力とし、機器からは動作や音声、光などの出力を行うといったやり取りが発生し、機器の普及に伴い必然的に人間と人工物のコミュニケーションの機会は増加していくこととなる。人間と人工物が円滑なコミュニケーションをとるためには、人工物側から人間への伝達を、人間が理解可能な形式とすることが重要である [1]。

人間同士のコミュニケーションは言語情報と非言語情報の大きく2つに分けられ、ジェスチャーや表情などもコミュニケーションに盛り込まれる [2]。非言語情報の中でも表情や小さな動作などの微細なものは Subtle Expression [3] と呼ばれ人間同士のコミュニケーションにおいて重要な役割を持つとされる。人間と人工物のコミュニケーションにおいても非言語情報は重要であるとして研究がされており [4]、具体的な手法として、例えば小林 [5] は、人間に酷似した顔ロボットの開発を目的とし、20箇所の制御点を実装することで世界共通の6つの基本表情を表出している。また、Ishiharaら [6] は子供型ロボットに呼吸時の胸のふくらみや肩をす

くめる動作といった非常に細かい動作を実装している。こういった手法は本来の人間のコミュニケーションプロトコルに基づくため、人間は人工物の出力の意図を学習する必要がなく負担を抑えられるという利点がある一方、実装にかかる費用や時間が問題となる。

リアルな人間の表出を再現する手法に対し、人工的で非常に単純な表出方法である Artificial Subtle Expression (ASE) が提案されており、ブープ音 [7] や振り返り速度 [8]、振動 [9] など様々な人間の知覚に働きかけたアプローチが提案されている。ASEは低コストでの実現が可能であり、既存のシステムに大きな変更を加えることなく組み込めることが利点であるが、主に2つの内部状態の伝達を前提としており、情報量が少ない問題がある。

そこで本研究では、シンプルで直感的なエージェントの多値確信度伝達を目的とする。提案手法では、複数の音声を重ね合わせることで集団圧力を表現し、重ね合わせの数を変化させることで、ユーザへの異なる確信度の伝達を狙う。

2 集団音声による多値確信度伝達

人間には少数派意見を持った者が斉一性の圧力により多数派意見に合わせるように振る舞う同調行動を示す傾向がある。Asch [10] の実験によると、非常に容易な課題にもかかわらず、意図的に誤答するサクラに同

*連絡先： 信州大学大学院総合理工学研究科工学専攻
〒380-0928 長野県長野市若里 4-17-1
E-mail: 19w2109b@shinshu-u.ac.jp

調し、誤った回答をする実験参加者が認められ、集団内の多数派の人数が3~4人のときに最も高い同調率となったことが報告されている。また小泉 [11] は、カナダでカリブーを狩っているカリブー・エスキモーに2人で歌を歌ってもらったところ、互いに合わせるができなかったが、一方で、クジラを狩っているクジラ・エスキモーは、リズムに合わせて大勢がピッタリと歌うことができた。小泉はこの理由を、生活のために集団で働くことのない文化圏では、リズムを合わせるが存在しないのではないかと考えている。すなわち、発声がそろっていることは集団的であることを意味する。

本研究では、同意見を持った集団の構成員の増加により同調圧力が高まること、タイミングをそろえた発声は集団性に結びつくと考えられることから、異なる声質の合成音声を用意し、これらを重ね合わせ同時に発話する人数を調整することで多値の確信度伝達を目指す。具体的には、重ね合わせ人数が多いほど人間はエージェントの確信度を高く受け取ると予想し、調査を行う。合成音声を作成するにあたり、Weber-Fechnerの法則により人間の感覚量は刺激量の対数に比例するとされることを踏まえ、著者らの検討により、1つの台詞につき、1人のみの音声(1人音声)、2人を重ね合わせた音声(2人音声)、4人を重ね合わせた音声(4人音声)の計3パターンを作成した。

3 実験

参加者実験により、異なる重ね合わせ人数のエージェントの集団音声が、異なる確信度としてユーザに伝達できるかを調査する。実験には信州大学、大学院の男子学生8名(平均年齢22.6歳)が参加した。

3.1 経路選択ゲーム

実験では図1に示す経路選択ゲームを使用する。このゲームでは道中出现する分かれ道を選択しつつ進んでいき、最短経路でのゴールを目指すことが目的である。分かれ道は最小27個、最大177個出現し、後述する終了条件のいずれかを満たしたときにゴールにどり着く。

ゲームは画面奥方向に道路を自動的に進んでいく。道中では、3方向への分かれ道が出現し、分かれ道の手前で停止する。分かれ道では、ゲーム画面左下に設置されたエージェントとしての予想システムが道の番号を示す「いちばん」「にばん」「さんばん」のいずれかを発話する。実験参加者はエージェントの予想を聞いた上でいずれかの方向への道を選択する。選択されると再び道路を自動的に進んでいき、終了条件のい



図 1: 経路選択ゲーム

れかを満たすまで分かれ道での道の選択と自動的な道路の進行が繰り返される。終了条件は次のとおりであり、この条件は実験参加者には説明していない。

終了条件

- 実験参加者がエージェントの予想とは違う方向を選択した、すなわち、実験参加者がエージェントの予想を拒否した回数が10回以上である、かつ27個以上の分かれ道通過。
- 177個の分かれ道通過。

3.2 エージェント音声

エージェントの音声は「いちばん」「にばん」「さんばん」の3方向と、1人音声、2人音声、4人音声の3パターンをかけ合わせた全9種を用意した。図2に「いちばん」を例として音声のスペクトログラムを示す。経路選択ゲームでは各分かれ道において、まず9種の音声を各3回ずつ、計27回がランダムな順序で表出され、28回以降では9種の内からランダムで表出される。エージェントの音声は音声作成ソフト CeVIO CS6 を用いて十分に区別できる4人分を作成し、音声編集ソフト Audacity を用いて合成した。また、合成時の各音声の組み合わせは、すべて同一のものとした。

3.3 実験手順

実験は、実験参加者が机の上に設置されたモニターに向かって着席した状態で行われた。まず、実験参加者は実験と経路選択ゲームについての説明を受け、音量調整を行い、ゲームを行う。このとき、音声はモニター付属のスピーカから出力される。その後、表1に示すアンケートとインタビューを受ける。実験参加者にはゲームについて以下の内容が教示された。なお、ゲー

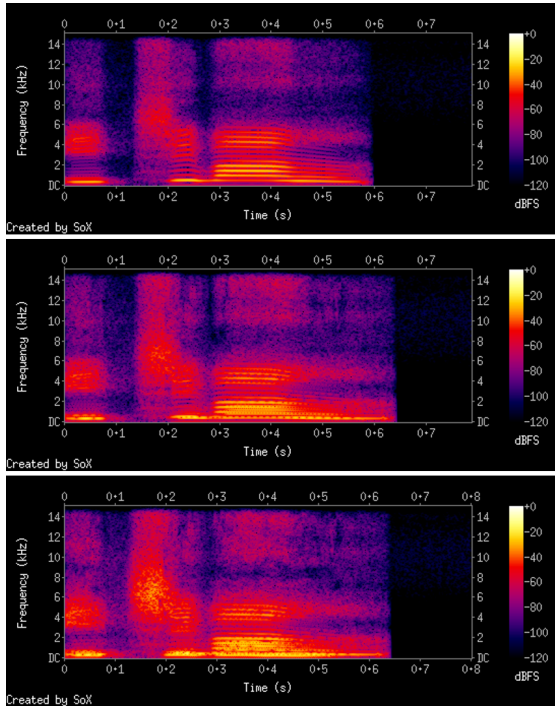


図 2: 「いちばん」のスペクトログラム (上から順に 1 人音声, 2 人音声, 4 人音声)

ムプレイ時とアンケート回答時は実験者は退室し, 実験参加者 1 名のみでの状況で行われた。

教示内容

- 分かれ道ではエージェントにより予想が提示される。
- 例えば左折を繰り返しても, 同じ場所には移動することはない。
- 各分かれ道での予測精度はおおよそ 70%である。
- エージェントの予想を聞いてから道を選択する。
- エージェントの予想に従うかは自由である。
- 最短経路では 27 個の分かれ道でゴールにたどり着く。
- できるだけ短い経路でのゴールを目的とする。

4 結果

エージェントの集団音声の実験参加者の経路の選択に与えた影響を調査するために, 経路選択ゲームにおいて, 実験参加者のエージェントの予想に対する拒否率を計測した。図 3 に計測と分析の結果を示す。音声の重ね合わせ人数ごとに集計を行ったところ, 1 人音声提示の平均拒否率は 0.376 (S.D.=0.346), 2 人音声では 0.188 (S.D.=0.0182), 4 人音声では 0.0722 (S.D.=0.0746) であった。ただし, 各音声の提示数は実験参加者ごとの

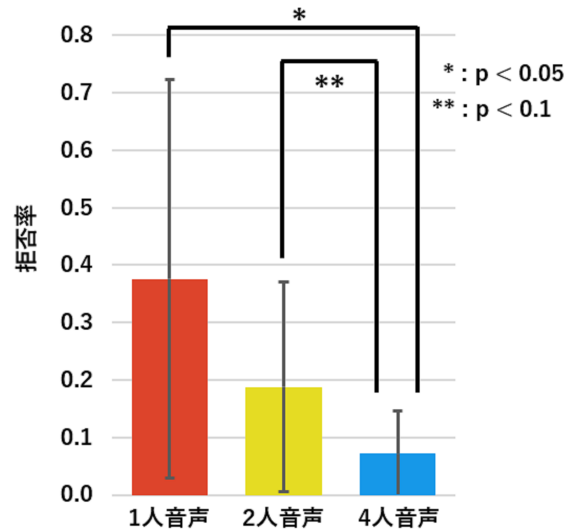


図 3: 各エージェント音声に対する拒否率

経路選択ゲームの進行状況により異なるため一定ではなく, 実際の平均提示数は表 2 のとおりであった。

3 水準の Friedman の検定 (水準: 重ね合わせ人数 1 人, 2 人, 4 人) の結果, 1 人音声, 2 人音声, 4 人音声での拒否率に統計的有意傾向が確認された ($\chi^2(2) = 5.07$, $p = 0.079$)。下位検定として Durbin-Conover 法による多重比較の結果, 1 人音声-4 人音声間では有意差が確認された ($p = 0.035$)。また, 2 人音声-4 人音声間では有意傾向が確認された ($p = 0.060$)。

アンケートにおいて, Q1 では 7 名が少なくともいずれかの番号に複数のパターンが存在することを認識しており, 「いちばん」の平均認識数は 2.38, 「にばん」では 2.75, 「さんばん」では 2.38 であった。各番号での認識パターン数について Friedman の検定を行ったところ, 統計的有意差は確認されず ($\chi^2(2) = 2.24$, $p = 0.327$)。発話音声による違いは認められなかった。

5 考察

実験の結果, 1 人音声-4 人音声間と 2 人音声-4 人音声間では有意差が認められたが, 1 人音声-2 人音声間では差が認められなかった。アンケート項目 Q4 への回答や実験後のインタビュー結果では, 人数が多いほうが信頼できた, という意見が多数見受けられた。このことから, 重ね合わせ人数が最大の 4 人での音声提示を信頼し, その他の音声は 4 人音声と比較して信頼するに足りないと判断されたことが示唆される。また, 今回作成した 3 パターンの重ね合わせ人数では 1 人音声と 2 人音声の差が明確でなく, 多値の確信度表出には適切でなかったことが考えられ, 今後の検討が必要である。

表 1: アンケート項目

No.	質問内容	回答方法
Q1	選択肢で提示された各番号について、その音声は何パターンあると思ったか.	1 人音声, 2 人音声, 4 人音声それぞれで 1,2,...,5,6 以上からの選択式
Q2	道案内の仕方に何か意味を感じたか.	自由記述
Q3	選択の際、どのくらい素早く考えて選ぶことを意識したか.	1~7 からの選択式 高い値ほど「意識した」
Q4	進む道を選択する際に決め手とした情報や戦略などはなにか.	自由記述
Q5	実験に関する感想など.	自由記述

表 2: 各音声の平均提示数 (標準偏差)

	1 人音声	2 人音声	4 人音声
いちばん	6.88 (2.53)	6.13 (2.64)	7.38 (4.14)
にばん	6.00 (2.83)	7.00 (3.42)	6.50 (3.12)
さんばん	6.25 (3.81)	6.50 (3.89)	6.25 (2.82)

実験参加者全体の傾向として人数の多い音声拒否されにくい傾向がある。一方で個人に着目すると、1 人音声のみを拒否した参加者や、1 人音声では内容が明確に聞こえ、人数が多いと説得力があるように感じるとして 2 人音声を多く拒否した参加者がいた。エージェントの表出に対する確信度の捉え方には個人差があり、いくつかの分類が存在する可能性があることが考えられ興味深い。今後は実験参加者数を増やし、一般的な傾向を調査するとともに、捉え方の分類が必要と考える。

また、本研究で作成した音声は、全 4 人分の音声を作成し、単に合成したものであった。そのため、音量については人数ごとの調整は行っていない。著者らの検討により、重ね合わせ人数が増加する際、音量の上昇は自然であるとして本稿の設定で実験を行ったが、実験参加者は人数ではなく音量の差によって拒否率に違いを出した可能性がある。そのため実際の音量を計測し、音量、人数の 2 要因についての調査が必要である。

6 まとめ

本研究では、音声の重ね合わせ人数の違いにより、多値の確信度表出が可能であるかを調査した。Weber-Fechner の法則に基づき、1 人音声、2 人音声、4 人音声の 3 パターンを作成し、経路選択ゲームにおいて実験参加者のエージェントの提案に対する拒否率を計測したところ、1 人音声-2 人音声、1 人音声-4 人音声間に統計的な差が認められた。実験の結果から、人数が多い音声の方がエージェントの提案を拒否されにくいことが示唆された。今後は多値確信度伝達のための適切な重ね合わせ人数の検討とともに、集団音声の捉え方の分類や音量要因を含めた調査をする予定である。

参考文献

- [1] Kobayashi, K., Yamada, S.: Motion Overlap for a Mobile Robot to Express its Mind, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.11, No.8, pp.964-971 (2007)
- [2] Kendon, A.: Do Gestures Communicate? A Review, *Research on Language and Social Interaction*, Vol.27, No.3, pp.175-200 (1994)
- [3] Karen K. Liu, Rosalind W. Picard: Subtle Expressivity in a Robotic Computer, in *Proceedings of CHI2003 Workshop on Subtle Expressivity for Characters and Robots*, pp.1-5 (2003)
- [4] Bartneck, C., Reichenbach, J.: Subtle Emotional Expressions of Synthetic Characters, *International Journal of Human-Computer Studies*, Vol.62, No.2, pp.179-192 (2005)
- [5] 小林 宏: 表情豊かな顔ロボットの開発と受付システムの実現, *日本ロボット学会誌*, Vol.24, No.6, pp.708-711 (2006)
- [6] Ishihara, H., Asada, M.: Design of 22-DOF pneumatically actuated upper body for child android 'Affetto', *Advanced Robotics*, Vol.29, No.18, pp.1151-1163 (2015)
- [7] Komatsu, T., Yamada, S., Kobayashi, K., Funakoshi, K., Nakano, M.: Artificial Subtle Expression: Intuitive Notification Methodology of Artifacts, *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.1941-1944 (2010)
- [8] 寺田 和憲, 山田 誠二, 小松 孝徳, 小林 一樹, 船越 孝太郎, 中野 幹生, 伊藤 昭: 移動ロボットによる Artificial Subtle Expressions を用いた確信度表出, *人工知能学会論文誌*, Vol.28, No.3, pp.311-319 (2013)
- [9] Komatsu, T., Kobayashi, K., Yamada, S., Funakoshi, K., Nakano, M.: Vibrational Artificial Subtle Expressions: Conveying System's Confidence Level to Users by Means of Smartphone Vibration, *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp.1-9 (2018)
- [10] Solomon E. Asch: Opinions and Social Pressure, *Scientific American*, Vol.193, No.5, pp.31-35 (1955)
- [11] 小泉文夫, 音楽の根源にあるもの, 青土社 (1977)