

言語情報を持たない音を用いた ヒューマンエージェントインタラクションシステムの開発

棚橋 徹^{a)} 小林 一樹^{b)} 北原 鉄朗^{c)}

^{a)} 日本大学大学院 総合基礎科学研究科

^{b)} 信州大学 学術研究院

^{c)} 日本大学 文理学部

1. はじめに

これまで様々な雑談システムが開発されてきたが、そもそも何を話せばいいのかわからない、1対1の対話だと対話の継続に疲れてしまう、などの問題が発生すると考えられる。そうした考えから本研究では、言語情報に基づかずに、音の韻律的特徴のみを用いて人と2体のエージェントによる持続的非言語雑談コミュニケーションを行える研究に取り組んでいる。

韻律的特徴によるコミュニケーションで課題となるのは、人間の発話の何をどうエージェントの発話に反映させるかである。人間の発話とは無関係にエージェントの発話が決まれば、いずれ人間側がそれに気づいて興味を失う可能性がある。

本稿では、エージェントが人間あるいは別のエージェントに発話パラメータ(韻律)を同調させることで、この課題の解決を図る。人間同士の協調的な対話では、韻律が互いに同調することが知られており^{1)~3)}、ヒューマンエージェントインタラクションにおいても、同調を導入することで協調性のあるコミュニケーションを実現できると期待する。

2. システム構成

2.1 システム概要

本研究の目標はユーザとの持続的なコミュニケーションを、言語情報を持たない音で行うことである。この目標を達成するために、以下の仕様を満たすエージェントを作成する。

- 2体のエージェントが存在し、各エージェントが音を発する。
- エージェントらが発する音は言語情報を一切持たない。
- エージェントらは韻律的特徴をユーザに同調する。
- ユーザが数秒間沈黙すると、片方のエージェントが発話する。

本稿では、このような仕様を満たすシステムを作成した。システムは、以下の流れで処理を行う。

- (1) ユーザからのマイク入力があったら、韻律的特徴(振幅, F0, セグメント長)を抽出する。
- (2) ユーザの韻律的特徴を基に同調制御を行う。
- (3) 同調制御によって求められた値を正規分布の平均とし、ランダムサンプリングする。
- (4) 正規分布からランダムサンプリングした値を用いて、正弦波を生成し、発話する。
- (5) 数秒間ユーザからの入力がない場合、エージェントは他方のエージェントに同調するように韻律的特徴を決定し、発話する。

以降では、エージェントの音の設計について2.3節で、(1)を

2.4節で、(2)、(3)、(4)を2.5節で詳しく述べる。

2.2 外見の設計

エージェントは、Arduinoを用いて、発話の際に音に合わせてLEDの明滅を行う。加えて、ユーザに対話相手として認識させるために、シンプルな人形型の外見⁴⁾を採用した。

2.3 エージェントの音の設計

エージェントが発する音は3セグメントからなり、各セグメントは振幅, F0, セグメント長の3つの特徴量からなる。ただし、セグメント内で振幅, F0は一定で、各セグメントの間は100msとする。エージェントは1発話に対して、9次元(3セグメント×3特徴量)の特徴ベクトル(以後発話パラメータベクトル)を持ち、この特徴ベクトルを基に、正弦波を生成しエージェントは音を発する。ユーザも、エージェントとの発話の自由度を揃えるため同様の制約を設ける。

2.4 特徴抽出

ユーザの発話から、2.3節と同様の9次元の発話パラメータベクトルを抽出する。まず、DIO⁵⁾を用いて5ms毎にF0を求める。この時、非有声区間では欠損値となるので、50ms以上連続で欠損した箇所ではセグメントは分割する。ただし、200ms未満のセグメントは削除する。その後、以下の方法によりセグメント毎にパラメータを求める。

- 振幅: 二乗平均平方根(RMS)を求め、dBに変換して特徴量として用いる。
- F0: DIOを用いて5ms毎に推定した値を440Hz=5700centになるようにcent単位に変換し、時間平均を求める。
- セグメント長: F0推定で検出された有声フレームの区間をセグメント長とする。

ただし他方のエージェントの特徴量は、内部的に特徴量を参照し取得したものをを用いる。

2.5 同調制御

エージェントの音発話パラメータベクトルを決定した後、これに基づいて正弦波を生成する。これは、あらかじめ設定した多次元正規分布に基づいたランダムサンプリングによって決定する。ここでの同調制御は、この正規分布の平均ベクトルを更新することで実現する。

同調制御を行うために、VanderPol方程式を文献⁶⁾を参考に相互作用を起こすよう拡張する。システムを起動して、エージェントの t 回目の正規分布の平均を μ_t とする。この時、 $t+1$ 回目のエージェントの正規分布の平均値 μ_{t+1} は

$$\mu_{t+1} = \mu_t + \dot{\mu}_{t+1}$$

$$\dot{\mu}_{t+1} = \mu_t + \ddot{\mu}_{t+1}$$

$$\ddot{\mu}_{t+1} = \epsilon(\mu_t^2 - 1)\dot{\mu}_t - \mu_t + \lambda(y_t - \mu_t) \quad (\epsilon > 0, \lambda > 0)$$

によって決定する。ただし、 μ_t^2 は要素ごとの演算である。ここで、 ϵ は粘性係数、 λ は結合係数で、対話相手にどれだけ同調するかの重みを表す。 y_t は対話相手の直近5発話の発話パラメータベクトルの平均である。この μ_{t+1} を用いて、エージェ

Human-Agent Interaction System using Non-Linguistic Sound by Tetsu Tanahashi (Nihon University), Kazuki Kobayashi (Shinshu University) and Tetsuro Kitahara (Nihon University)

表 1 実験 1 結果 (1: 1 体条件が良い, 5: 2 体条件が良い)

	質問項目	評価
Q1	どちらのシステムが話しやすかったですか	2
Q2	どちらのシステムが会話できたと感じますか	3
Q3	どちらのシステムが話し続けやすいですか	2
Q4	どちらのシステムが楽しかったですか	4
Q5	どちらのシステムを再度使いたいですか	5

ントの次の発話特徴量 x_{t+1} を正規分布 $\mathcal{N}(x_{t+1}; \mu_{t+1}, \Sigma)$ からランダムサンプリングする。ただし, Σ は対角行列とし, 実験的に定める。

このようにして求められた発話パラメータベクトルを基に正弦波を生成し, エージェントは音を発する。

3. 評価実験

評価実験では, エージェントの個体数効果の検証 (実験 1) と同調制御の効果の検証 (実験 2) を行った。実験には, 1 名 (女性, 22 歳) が参加した。実験参加者が, 3 セグメントを確実に分割して発声しやすいうように, 発話は「ば」のみを用いることとした。

3.1 実験前練習

実験 1, 2 を行う前に参加者には発話の練習として, 「ば」を 3 つ使って韻律的特徴量を変更しながら発話してもらい, それに対してエージェントが返答 (440Hz, 1 セグメント 1 秒を 3 回繰り返す正弦波) するやりとりを何度か行ってもらった。これは, エージェントに対して非言語情報で話すことを恥ずかしいと感じたり, 「ば」を使ってどんな発話ができるかを参加者にわかってもらうためである。

3.2 実験 1

エージェントの個体数効果を検証するため, 2 体条件と 1 体条件とを設定して実験を行った。両条件において, 被験者にはエージェントと「ば」を用いて好きなように会話を試みるように指示し, 5 分間使用してもらった。終了後, 文献⁷⁾を参考に作成した質問項目に 5 段階で回答してもらった。

結果を表 1 に示す。Q1, Q3 では 1 体条件の方が評価が高かった。これは被験者から, 好きなタイミングで話すことができるため, Q1, Q3 に関しては 1 体条件の方が良いとの意見をもらった。しかし, 自分が発話しないとエージェントは発話しないため, 対話の継続には疲れてしまうという意見もあった。また Q4, Q5 では, 2 体条件の方が評価が高かった。これは, 自分が話さなくてもエージェントが発話する時や, エージェント同士で会話を行う時があるため, 会話を継続を行うには疲れにくいとの意見をもらった。この結果から, エージェント同士の会話が行われる事で, ユーザが会話の継続に疲れてしまうということが軽減される可能性が示唆された。

3.3 実験 2

同調制御の効果を検証するため, 同調あり条件と同調なし条件とで実験を行った。参加者には, エージェント同士のコミュニケーション中に, 好きなタイミングで入るように指示した。これは, 参加者にエージェント同士の会話を観察させる時間を設けるためである。本システムへの不慣れの影響を避けるため, 同調あり条件となし条件を交互に 4 回実験を繰

り返し, 各条件が終わる度に 5 段階でエージェントを評価してもらった後, 2~4 回の平均評価値を用いた。

参加者の各エージェントに対する 2~4 の平均評価値を表 2 に示す。結果, すべての質問項目に対して同調あり条件の評価の平均値が高かった。このことから, 今回の参加者にとって同調なし条件よりは, 同調あり条件の方が協調的なエージェントで, コミュニケーションを楽しむ事が出来たと感じたと考えられる。また, 参加者が同調条件の場合にエージェントが同調していることに 4 回中 1 回気づき, その際には評価が高くなった。これは, 参加者から「エージェントが自分の真似をしようとしてきて親しみを感じた」とのことであった。この結果から, 言語情報を用いず, 韻律的特徴を同調することにより協調的な会話を行うことができる可能性が見出された。

表 2 エージェントの評価平均値 (1: そう思わない, 5: そう思う)

	質問項目	非同調	同調
Q6	エージェント A は協調的でしたか	3	3.33
Q7	エージェント A と会話できましたか	3	3.33
Q8	エージェント B は協調的でしたか	3	3.33
Q9	エージェント B と会話できましたか	3	3.33
Q10	会話にスムーズに入れましたか	2.67	4
Q11	楽しかったですか	3	4
Q12	様々な「ば」は言えましたか	2.67	3.67

4. おわりに

本研究では持続的非言語雑談コミュニケーションの実現のために, 韻律的特徴を同調する 2 体のエージェントを作成し, 実験を 2 つ行った。実験 1 の結果から, 1 被験者対 1 エージェントよりも 1 被験者対 2 エージェントの方が, 会話の継続に疲れてしまうことが軽減される可能性が見出された。また実験 2 の結果から, 同調するエージェントを用いた場合, 協調的な会話を行うことができる可能性が示唆された。しかし, 現状では 1 名しか実験できておらず, 一般性のある結果になるとは言い難い。今後は, 実験参加者を増やし, 本システムの有効性を更に検証する予定である。

謝辞 本研究は, JSPS 科研費 16K16180 号, 新学術領域研究「認知的インタラクションデザイン学」(No.JP26118005) から支援を受けた。

参考文献

- 1) 長岡千賀 他, 音声対話における交替潜時が対人認知に及ぼす影響, HIS 論文, 22, pp. 171-174, 2002.
- 2) 長岡千賀 他, 協調的対話における音声行動の 2 者間の一致一意見固持型対話と聞き入れ型対話の比較, HIS 論文, pp. 167-170, 2003.
- 3) 西村良太 他, 音声対話における韻律変化をもたらす要因分析, 音声研究, 13, 3, pp.66-84, 2009.
- 4) Lightclip: ninja Ghost, <https://www.shapeways.com/product/ZPFG39RLC/lightclip-ninja-ghost-iphone-4-4s>, 2018 年アクセス.
- 5) 森勢将雅 他, 基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法, 信学論, J93-D, 2, pp. 109-117, 2010.
- 6) 速水達也 他, 交替潜時と韻律情報に基づく会話同調制御方式と情報収集を目的とした会話エージェントへの実装, 情処学論, 54, 8, pp. 2109-2118, 2013.
- 7) 藤堂祐樹 他, 単一对話エージェントと複数対話エージェントを用いた音声対話システムの分析と評価, MUS, pp. 1-8, 2012.