平成 29 年度 修士論文

2 体の同調エージェントを用いた
音によるコミュニケーションの研究

指導教員 北原 鉄朗 准教授

日本大学大学院 総合基礎科学研究科

地球情報数理科学専攻 情報科学部門

6116M10　棚橋 徹

2018 年 2 月 提出

概　　要

近年, 特定のタスク達成を目的としたタスク指向型コミュニケーションシステムだけではなく, コミュニケーション自体を目的とした非タスク指向型コミュニケーションシステムの開発が盛んに行われている. 非タスク指向型コミュニケーションシステムは言語情報を用いたコミュニケーションと非言語情報（身振り手振り, 音など）を用いたコミュニケーションに分けることができる. 言語情報を用いたコミュニケーションシステムとして代表的なのは雑談システムである. 雑談システムの問題点として, 音声認識や言語理解に誤りがある場合, 会話が成り立たなくなるという点が挙げられる. 一方, 非言語情報を用いたコミュニケーションシステムでは, 言語情報を用いた際に生じる問題が回避できる利点がある. しかし, 多くの研究では人間 1 人に対してシステム内の対話相手（エージェントと呼ぶ）が 1 体の 1 対 1 コミュニケーションを想定しており, コミュニケーションを維持するために人間もエージェントも（音を用いる場合）話し続けないとコミュニケーションを維持できない. このことから, 人間側の心理的負担が大きいことが予想される.

本研究では, 2 体のエージェントと人間が音の韻律的特徴 (本研究では音の振幅, 高さ, 長さ) のみを用いてコミュニケーションを行うシステムを提案する. 本システムでは, 2 体のエージェントが人間とコミュニケーションを取りつつ, 人間からの反応がない場合にはエージェント同士でコミュニケーションを継続する. これにより前述の問題を解決する. エージェントと円滑にコミュニケーションを行うためには, 時間的随伴性（行動を行った人に対して反応をすること）, エージェンシ（人間がコミュニケーションを取るべき適切な対象として捉えることができる特徴

を持つこと）, 人間に学習・適応する能力（人間の行動に無関係にシステムの挙動を決めるのではなく, 人間の行動に影響を受けて行動すること）が必要だと言われている. 「時間的随伴性」については, 各エージェントは, 人間または他のエージェントが発話を終えると, 正規分布に従う乱数により決められた時間待った後に発話を始めるものとする. 「エージェンシ」については, お化け型の外見が発話時に光ることで対話相手らしさを演出する. 「人間への学習・適応」は, エージェントが発する音を決める内部パラメータを対話相手（人間または他のエージェント）に合わせて変化させる（これを同調と呼ぶ）ことで実現する. また, 人間が発話を数秒間行わないとエージェント同士がコミュニケーションを始めるので, その様子を見ることでコミュニケーションはより円滑化されると期待される.

本システムの有効性を確認するため, 2 種類の評価実験を行った. 実験 1 ではエージェントが 2 体いることの有用性を検証するため, エージェントが 1 体のみのシステムとの比較を行った. 実験の結果, エージェントが 1 体のシステムが良いと答えた被験者と, エージェントが 2 体のシステムが良いと答えた被験者に分かれた. この評価の差は, どちらのシステムを先に使ったのかに依存するため, 今後, 両システムを長期的に試行し検証していく必要がある. また, 実験 2 は同調制御の効果を検証するため同調制御をしないシステムとの比較を行った. その結果, 6 人中 5 人は同調制御を行うエージェントの方が, 同調を行わないエージェントに比べて, 協調的だった, 会話を行うことができた, と回答した. このことから, 同調制御を行うことで, 協調的な会話を行うことができる可能性が示唆された.

実験の結果, 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示唆された. しかし, 2 体にすることによってコミュニケーションの維持ができるかについては, 検証が必要な結果となった. 加えて, 非言語コミュニケーションにおいても発話を行うタイミングや頻度も考慮に入れるべきということがわかった. 今後は被験者数を増やし本研究の結果が一般性のある結果であることを示していくとともに, 長期的な検証や本システムの有効性を確かめ, 発話頻度やタ

イミングも考慮に入れた持続的な非言語コミュニケーションの実現を目指したい.

目　　次

図 目 次

表 目 次

第1章　序　　論

本章では, 研究の背景, 目的を述べた後本論文の構成を述べる.

1.1　本研究の背景

近年, チケット予約や道案内など, 特定のタスクの達成を目的としたタスク指向型コミュニケーションシステムだけでなく, コミュニケーションそのものを目的とした非タスク指向型コミュニケーションシステムの開発が盛んに行われている. 非タスク指向型コミュニケーションシステムの中でも, 雑談システムと呼ばれる会話コミュニケーションシステムの開発, 研究が盛んに行われている [1, 2, 3, 4, 5, 6, 7, 8]. 雑談システムでは, 他愛のない会話や最近おきた出来事について話すことによって, ユーザを楽しませることや, 親近感や安心感を与える事が期待されている. こうした雑談システムは, 話題の制約が小さいことから, 音声認識や言語理解に失敗する可能性があり, その場合, 適切な応答文を生成出来ない場合がある. この問題を解決するために, 様々な研究が行われている [9, 10, 11, 12, 13, 14, 15, 16]

一方で非タスク指向型コミュニケーションは, 非言語情報を用いたものも存在する. 非言語情報を用いた場合, 音声認識や言語理解, 応答文生成などの問題を回避できる利点がある. 非言語情報とは, 身振り手振りをはじめ, 発話の韻律的特徴 (本研究では音の振幅, 高さ, 長さ), 表情などが含まれる . 例えば, 文献 [17] では, 単純かつ僅かな非言語情報のやり取りを繰り返し行うことで, 何らかの意図が伝わることができるとの考えから, 光るボタン型デバイスや振動デバイスを用いて遠隔地にいるユーザらのコミュニケーションを実現してきた. また, 単純な情報呈示によって,

ロボットの内部状態の表出を試みた研究 [18, 19, 20] も存在しており, 人とロボットにおける非言語情報を用いたコミュニケーションについても検討がなされている. しかし, これらの非言語コミュニケーションの研究は, ほとんどが人間1人に対してエージェント1体, もしくはロボット1体である. 1対1コミュニケーションでは人間が行動を起こし続けなければコミュニケーションの維持が出来ないことがあり, ユーザ側が疲れてしまうことや, システムに飽きてしまうことが予想される.

1.2 本研究の目的

本研究では, 非言語コミュニケーションにおいて, コミュニケーションの維持が難しいという問題を解決するために, 人と2体のエージェントが音の韻律的特徴のみを用いてコミュニケーションを行うシステムを提案する. 一般的に, ロボットとの円滑なコミュニケーションを行うためには,

(1) 時間的随伴性
(2) エージェンシ
(3) 人間への学習・適応していく能力

が重要とされている [21]. (1) の時間的随伴性とは, 自分が行動をとった際に, 相手が自分に対して行動を返すというものである. (2) エージェンシとは, ロボットはただの人工物ではなく, コミュニケーションを行うことができる対象としてユーザが捉えることができる特徴を持つ, ということである. また, (3) の人間への学習・適応していく能力とは, 人間の行動に無関係にロボットの行動を決めるのではなく, 人間の行動に影響を受けてロボットの行動を決定するということである. これらは, ロボットだけでなくエージェントでも同様であると考えられる. 以上から本システムでは, これらの仕様を満たす2体のエージェントが, 人間とコミュニケーションを行い, 人間からの反応がない場合にはエージェント同士でコミュニケーション

を継続する. これにより, コミュニケーションを維持する際の人間側の心理的負担が軽減されると期待する.

1.3 本論文の構成

本論文は次の構成からなる. 第 2 章では, エージェントとのコミュニケーションについて述べる. 第 3 章では, 本研究で作成したシステムの構成とその処理方法を述べる. 第 4 章では, 本システムの有用性を検証するために行った実験について述べる. 第 5 章では, 本研究の結論と今後の課題について述べる.

第2章　エージェントとのコミュニケーション

本章では, 2.1 節で人間とエージェントのコミュニケーションのあり方と, 2.2 節で非言語コミュニケーションに関する研究について述べる.

2.1　人間とエージェントのコミュニケーションのあり方

小松ら [21] では, 人間とロボットとの間の円滑なコミュニケーションに必要なものは,

(1) 時間的随伴性
(2) エージェンシ
(3) 人間に対して学習・適応していく能力

と述べている. (1) の時間的随伴性とは, 自分が行動をとった際に相手が自分に行動を示す (呼びかけたら返事をするなど) というものである. これは, 母子間の相互作用の研究 [22] で述べられているが, 生後二ヶ月未満の乳児でさえ, 自分のとった行動に対して相手が自分に行動を返すことに対して敏感なことが知られている. このことから, 時間定期随伴性がロボットとのコミュニケーションに重要な要素だということが予想されると述べられている. (2) エージェンシとは, ロボットをただの人工物ではなくコミュニケーションにおける適切な対象としてユーザが捉えることができる特徴を持つ, ということである. 人と人が対話を行うのは当たり前だ

が, ほうきやちりとりなどに対して話そうと思う人はいない. 一方でロボットに関しては, 人工物にも関わらず, 条件によっては乳児にもコミュニケーションにおける適切な対象として捉えられていることが報告されている [23]. また (3) の人間に対して学習・適応していく能力とは, 人間の行動に無関係にロボットの行動を決めるのではなく, 人間の行動に応じてロボットの行動を決定することである. 学習・適応した結果, 円滑なコミュニケーションが実現され, そのコミュニケーションが次の学習・適応するためのデータを提供するといった循環が存在する. またこの時に, 人間もロボットに適応し, 人間とロボットとが相互適応を行うことで, より円滑なコミュニケーションが生まれる. 実際に雑談システムでは, 人への学習・適応を行う研究が存在している [14, 15, 16, 24]. これらはロボットとのコミュニケーションだけでなく, エージェントとのコミュニケーションにおいても同様であり, 非言語情報におけるコミュニケーションにおいても同様であると考えられる.

2.2 非言語コミュニケーション

非言語コミュニケーションにおいては, ボタンを押したら遠隔地にあるボタンの色が変わるデバイス [17](図 2.1) や, 一方のデバイスを動かしたら他方も動くデバイス [17] (図 2.2) が存在する. これらの研究の目的は, 単純かつわずかな情報のやり取りを繰り返し行うことで, ユーザ同士が「つながっている」という感覚を得ることで安心を創出することである. 実際に, ボタンデバイスを用いた実験では, 終了後に被験者から「ボタンを使用しなくなると少しさびしく感じた」など, 「つながっている」という感覚を持っていたことが示唆される被験者も存在した. 以上からこの研究では, 2.1 節で述べた (1) をリアルタイムに光や振動を伝えることによって実現した. (2) は人と人のコミュニケーションのため, 満たしている. また, (3) に関しては考慮していない.

人とロボットにおけるコミュニケーションにおいては, 文献 [25] などが行われて

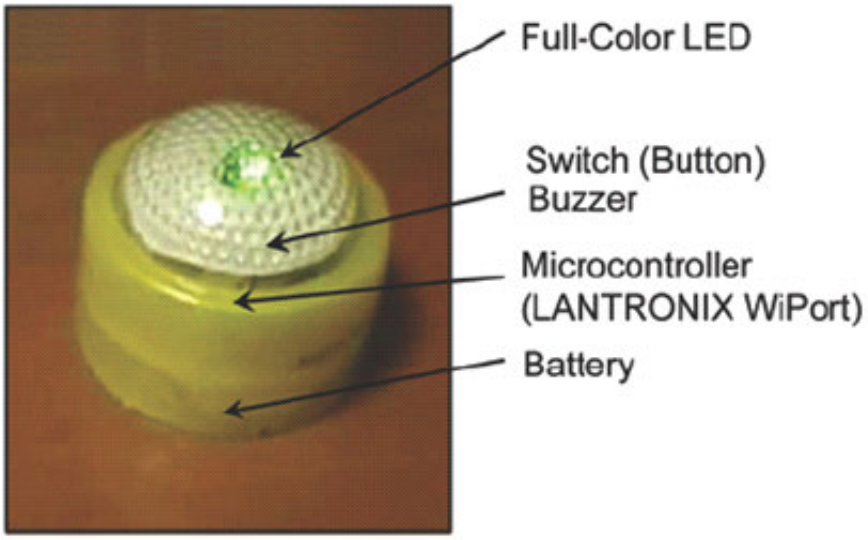

図 2.1: ボタン型デバイス [17]

図 2.2: 振動型デバイス COLOLO[17]

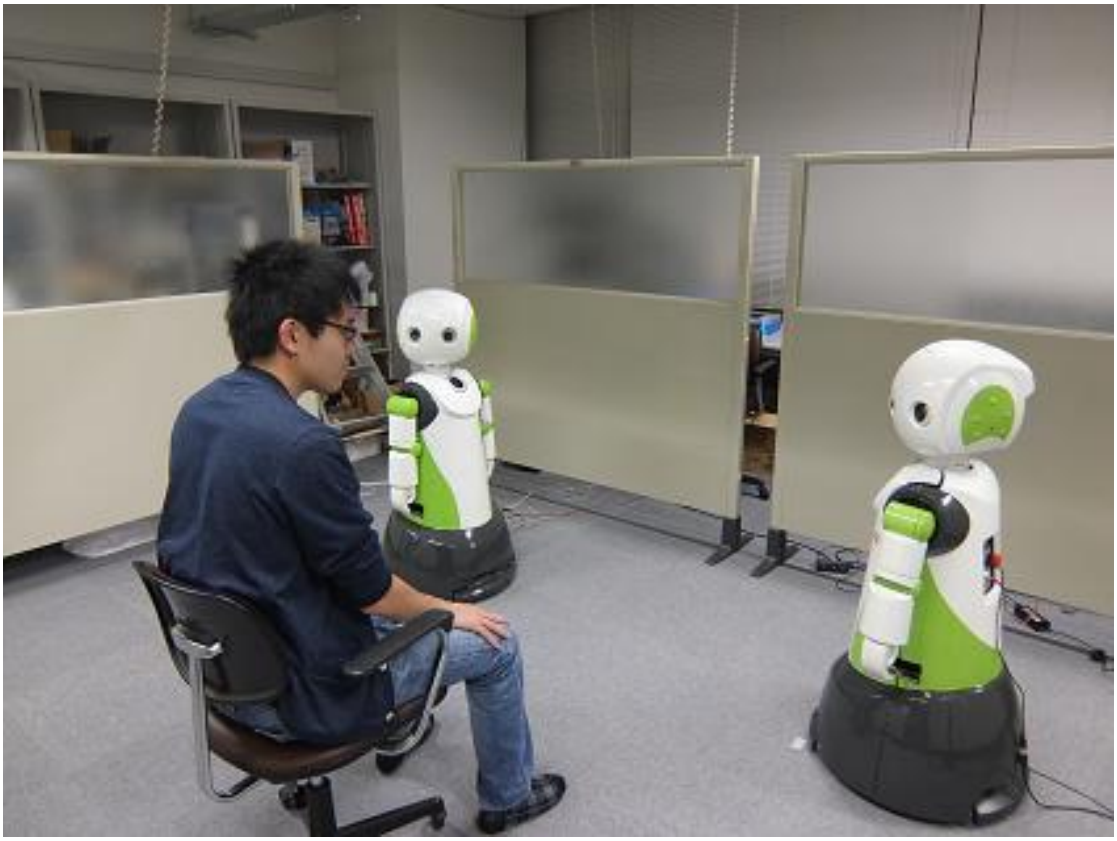

図 2.3: ロボット 2 体が参加者の後方に視線を向ける様子 [25]

いる. この研究では人型ロボット 2 体が, 人工音でコミュニケーションを行っている最中に, 順番に実験参加者の後ろに視線を向ける場合と 2 体が同時に後ろに視線を向ける場合の 2 種類の状況において, 参加者がどのぐらいロボットの視線を追従するのかを比較した研究である. 実験の様子を図 2.3 に示す. 結果は, 2 体のロボットが同時に視線を向けるときのみ, 参加者の視線が後方に誘導された. このことから参加者は, ロボットからの何らかの意思を感じたと考えられる. この研究では, 2.1 節で述べた (2) はロボット同士が人工音でコミュニケーションを行うなどの点で達成していると考えられる. しかし, (1), (3) は考えられていない.

他にも, 音のみを用いてコンピュータの態度をユーザが推定できるか, ということを確かめた研究 [26] が存在する. この研究では, 図 2.4 に示すようなビープ音を

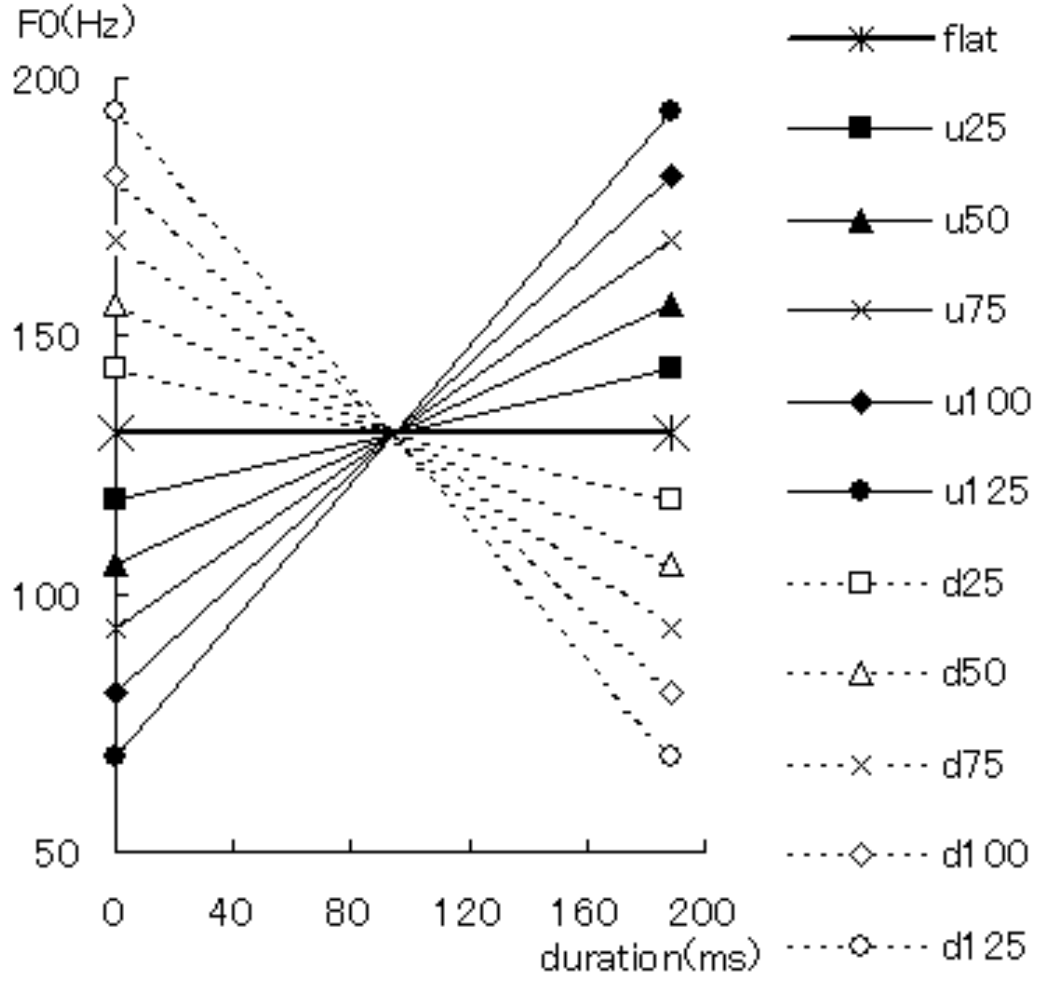

図 2.4: 11 種類の呈示したビープ音 [26]

用いて, ユーザにどのような感情に聞こえたかを答えてもらう実験を行った. その結果, 上昇調のビープ音では疑問・驚き, 変化の幅が小さく, 長い呈示時間の音は時間稼ぎ・躊躇, 下降調で呈示時間が短い音は同意・肯定といった態度をユーザが推定できる事が示された. このような知見を用いて, ロボットの内部状態を伝える研究 [18, 27] もなされている. これらの研究では, 2.1 節で述べた (1) に関しては, 適切な場面で音を流すことなどにより, 達成している. (2) に関しては, ロボットの見

た目や, 色などを用いることにより達成を図っている. (3) に関しては, 人間の知覚に合わせて beep 音を作成しているため, 適応していると考えられる.

しかし, いずれの研究においても 1 人の人間に対して 1 体のロボット, もしくはエージェントを想定している. そのため, コミュニケーションを続けるためには, 人間もエージェントも話し続ける, もしくは行動を起こし続けなければコミュニケーションの維持が出来ず, 人間側の心理的負担が大きいことが予想される.

第3章　システム構成

本章では作成したシステムの概要と構成, 各処理について述べる.

3.1　システム概要

本研究では, 非言語コミュニケーションにおいて, コミュニケーション維持が難しいという問題を解決するために, 人と 2 体のエージェントが音の韻律的特徴のみを用いてコミュニケーションを行うシステムの作成を行う. そのために,

(a) エージェントが発する音は言語情報を一切持たない.
(b) 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.
(c) エージェントは, ユーザだけでなく他方のエージェントともコミュニケーションを行う.
(d) 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.

という仕様を設ける. これにより, ユーザが音を発さない場合においてもエージェント同士がコミュニケーションを行うことにより, コミュニケーションが維持できると考えられる. また, 一般的に人とロボットの円滑なコミュニケーションを行うためには,

(1) 時間的随伴性
(2) エージェンシ
(3) 人間に対して学習・適応していく能力

が必要だとされている [21]. これは, ロボットだけでなくエージェントでも同様であると考えられる. 以上のことから,

- 発話者からの入力に対して, 返答する機能
- 会話者として捉えることができる外見
- 発話者の入力を反映させたエージェントの返答を生成する機能

が必要だと考えられる. したがって,

(e) エージェントは, ユーザからの入力が合った場合, 音を発して反応をする.
(f) エージェントはお化け型の外見を用意し, 音を発する際に光の明滅を行う.
(g) エージェントは韻律的特徴を発話者に同調する.

という仕様を設ける. (e) は, ユーザが行った発話に対して, エージェントが発話を行うということにより, 時間的随伴性を満たすと考えられる. (f) は, 会話相手らしさを演出することにより, エージェンシを満たすと考えられる. (g) における同調とは, エージェントの韻律的特徴が, 発話者の韻律的特徴の変化を追従することである. これによって, 人間への学習・適応を行う. これは, 既存研究 [28, 29, 30] などで述べられているとおり, 人間同士の協調的な対話では, 発話相手に韻律的特徴が同調することが知られているためである.

このような (a)∼(g) の仕様で想定される会話フローの例を図 3.1 に示す. 本研究で提案するシステムは, ユーザの発話を検出すると, その発話から振幅, セグメント長, 音の高さ (以後 F0) の 3 つの特徴量を抽出する. その抽出した特徴量に同調, すなわち真似するようにエージェントの特徴量を決定する. その後, その特徴量を用いてエージェントは音を生成し, 発話を行う. この際, ユーザもエージェントも言語情報を持たない音を発し, コミュニケーションを行う. このコミュニケーションを繰り返している最中, ユーザが数秒間発話しなかった場合, 他方のエージェントが発話を行う仕組みとなっている.

すなわち, システムは以下の流れで処理を行う.

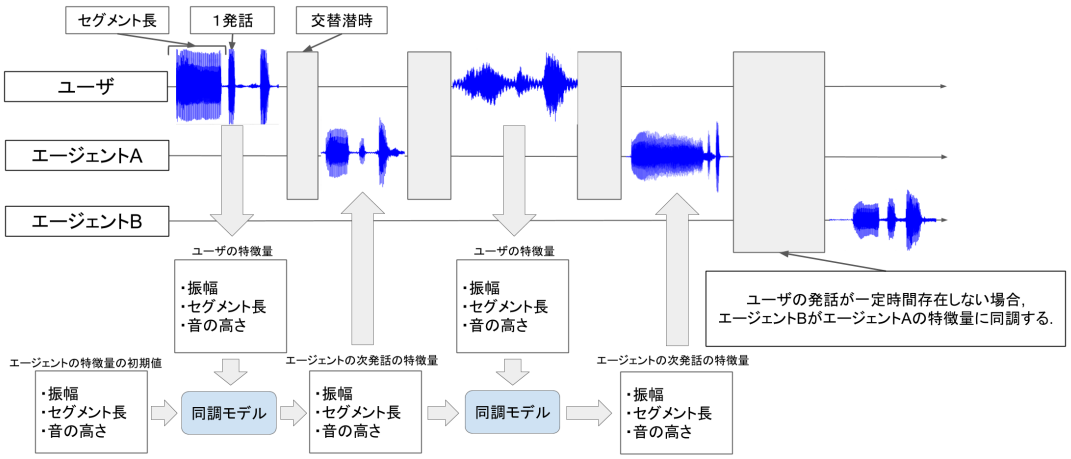

図 3.1: 会話フロー例

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. 同調制御で求めた値を平均とする正規分布からサンプリングした値を用いて, 正弦波を生成し, 発話する.
4. 数秒間ユーザからの入力が無い場合, 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を 3.3 節で, 1. を 3.4 節で, 2., 3. を 3.5 節で, 4. を 3.6 節で詳しく述べる.

3.2 外見の設計

エージェントは, Arduino[31] を用いて, 発話の際に音に合わせて LED の明滅を行う. また, [32] で購入したお化け型アクセサリー (図 3.2) を LED に被せる. これにより, ユーザにエージェントとして認識してもらいやすくし, コミュニケーションを行う適切な対象として捉えてもらえるようにした.

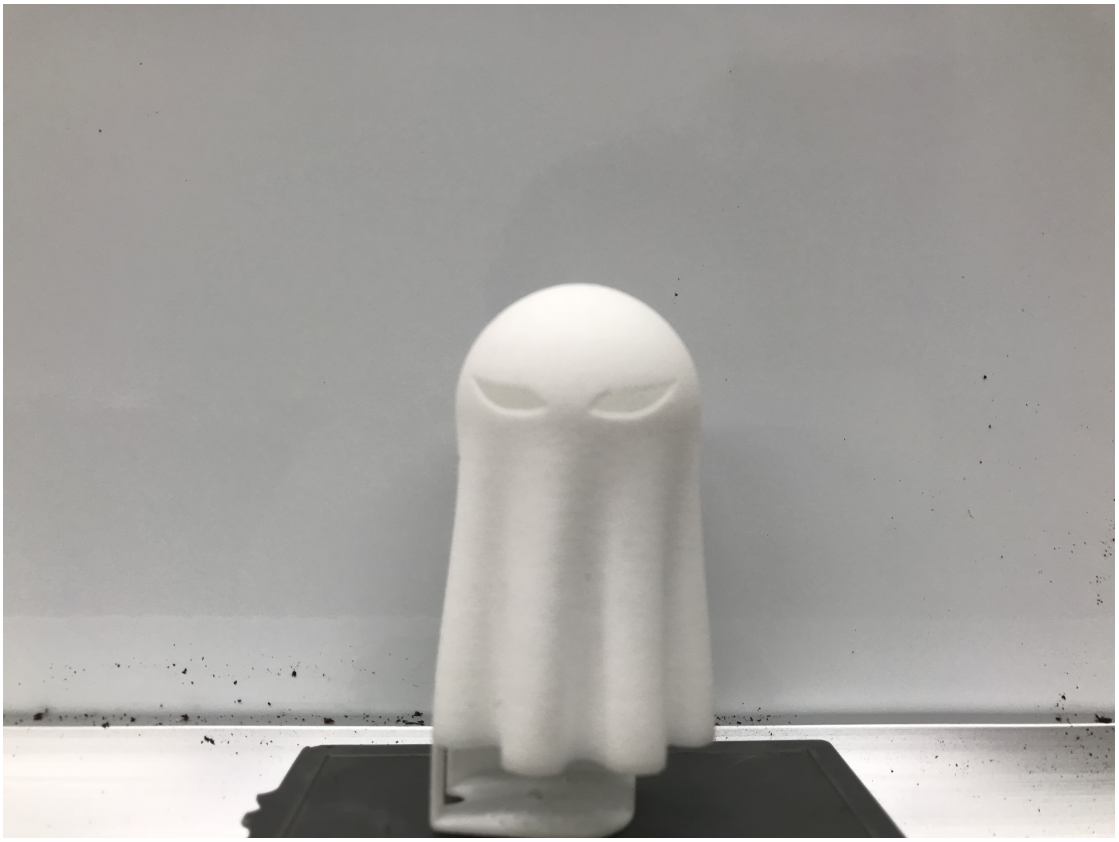

図 3.2: 本研究で使用したエージェントの外見

3.3 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.

3.4 特徴抽出

ユーザの発話から, 3.3 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, 基本周波数推定法である DIO[33] を用いて 5ms 毎に F0 を求める. この時, 無声区間では F0 欠損値となるので, 50ms 以上連続で欠損した箇所でセグメントを分割する. ただし, 200ms 未満のセグメントは削除する. これは, ユーザの発話が 3 セグメントあり, 各セグメントが短い場合もあることを考慮し実験的に定めた. 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: 各セグメントごとに二乗平均平方根 (以後 RMS) を求め, それを dB に変換して特徴量として用いる. 本研究で用いた RMS の式は

$$\mathrm{RMS} = \sqrt{2 \times \frac{1}{N} \sum_{i=1}^{N} {amp_i}^2} \tag{3.1}$$

である. ここで, N は各セグメントのサンプル数, amp_i はマイクからの入力信号のサンプリングされた値である. また, dB に変換するための式は,

$$y_{\mathrm{dB}} \quad = \quad 20 \times \log_{10} \mathrm{RMS} \tag{3.2}$$

である.

- F0: DIO で 5ms 毎に推定した値 $F0_{\mathrm{Hz}}$ を式 (3.3) で cent 単位に変換し, 各セグメントごとに平均を求め特徴量として扱う.

$$F0_{\mathrm{cent}} = 1200 \times \log_2 \frac{F0_{\mathrm{Hz}}}{440 \times 2^{\frac{3}{12}-5}} \tag{3.3}$$

- 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発話の特徴抽出は, 内部的に発話パラメータベクトルを参照し取得する.

3.5 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. システムを起動して, エージェントの t 回目の発話パラメータベクトル $\boldsymbol{x}_t$ は, 正規分布 $\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \Sigma)$ からランダムサンプリングされる. この時, $t+1$ 回目のエージェントの正規分布の平均値 $\boldsymbol{\mu}_{t+1}$ は

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \dot{\boldsymbol{\mu}}_{t+1} \tag{3.4}$$

$$\dot{\boldsymbol{\mu}}_{t+1} = \dot{\boldsymbol{\mu}}_t + \lambda(\dot{\boldsymbol{y}}_t - \dot{\boldsymbol{\mu}}_t) \tag{3.5}$$

によって決定する. ここで, λ は結合係数とし, 対話相手にどれだけ同調するかの重みを表す. 本稿では予備実験的に, $\lambda = 1.0$ とした. また, $\dot{\boldsymbol{y}}_t$ は発話相手の発話パラメータベクトルの一回微分で, 一発話前の発話パラメータベクトルとの差分を表す. すなわち,

$$\dot{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1} \tag{3.6}$$

この $\boldsymbol{\mu}_{t+1}$ を用いて, エージェントの次の発話特徴量 $\boldsymbol{x}_{t+1}$ を正規分布 $\mathcal{N}(\boldsymbol{x}_{t+1}; \boldsymbol{\mu}_{t+1}, \Sigma)$ からランダムサンプリングする．ただし, $\boldsymbol{\Sigma}$ は対角行列とし, 実験的に定める．また，各特徴量は互いに独立に同調する．

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．

3.6 沈黙時のエージェントの動き

ユーザが数秒間話さない場合, エージェントが発話を行うように設計した．エージェントが話した後, ユーザが話し始めるまでの時間を交替潜時とする．最初は, 交替潜時が閾値を超えた場合, エージェントは音を発する．閾値は, 予備実験的に 5 秒とした．この時発する音は, 他方のエージェントに同調するように発話パラメータベクトルを決定し発話を行う．

第4章　評価実験

評価実験では,

- エージェントが 2 体いる効果の検証 (実験 1)
- 同調制御の効果の検証 (実験 2)

を行った. 両実験とも, 6 名 (男性：5 名, 女性：1 名, 平均 20.8 歳) に対して行った. また, 3 セグメントを確実に分割して発声してもらうために, 被験者の発話は「ぱ」を用いることとした.

4.1　実験前練習

実験 1, 2 を行う前に被験者には発話の練習として, 「ぱ」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) するということを行った. 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, 「ぱ」を使ってどんな発話ができるかを被験者にわかってもらうためである. これを, 5 分間行ってもらった.

4.2　エージェントが 2 体いることの有用性の検証 (実験 1)

4.2.1　実験条件

エージェントが 2 体いる効果を検証するため, 提案システム (以下「2 体条件」) 及びエージェントを 1 体にしたもの (以下「1 体条件」) を用いて実験を行った. 両条件においても, エージェントは同調制御を行う. 各々の条件において, 被験者にはエージェントと「ぱ」を 3 回用いて好きなように会話をしてみてくださいと指示し, 5 分間使用してもった. 実験を行う順番は, 1 体条件から行う場合と 2 体条件から行う場合で人数が等しくなるようにした. 実験終了後, 文献 [12] を参考に作成した質問項目に 1 体条件の方だと感じた場合 1 に, 2 体条件の方だと感じた場合 5 に, 5 段階で評価してもらった.

4.2.2　実験結果

結果を表 4.1 に示す. 表の A∼F は被験者を表す. この結果から, 1 体条件が良いと答えた被験者と 2 体条件が良いと答えた被験者に分かれた. この評価の差は, 後に使ったシステムの方が良いと答える被験者が多かったことからシステムへの慣れによる順番に依存した結果となった. 2 体条件を先に行った被験者からは, 2 体のエージェントがいるとエージェントが音を発した際に, 自分に話しかけているのか, エージェント同士で話しているのかがわからない, という意見や, どのタイミングで会話に入ればよいのかわからないといった意見が出た. その結果, Q1 ∼ Q3, Q6 の評価では 1 体条件のほうが評価が高い被験者がいたと考えられる. このことから, 2 体条件は発話のタイミングが分からないため, エージェントと会話を行う事が難しいということがわかる. また, 1 体条件の際には, 無言の時に返事がないこ

表 4.1: 実験 1 結果 (1: 1 体条件, 5: 2 体条件)

	質問項目	A	B	C	D	E	F	平均
Q1	どちらのシステムが話しやすかったですか	4	1	4	1	4	1	2.5
Q2	どちらのシステムが会話できたと感じますか	4	1	3	1	4	1	2.3
Q3	どちらのシステムが話続けやすいですか	5	1	4	4	2	1	2.8
Q4	どちらのシステムが楽しかったですか	4	1	5	2	5	1	3.5
Q5	どちらのシステムが再度使いたいですか	5	1	4	2	5	1	3.0
Q6	どちらのシステムが話す時に迷いましたか	1	5	3	5	2	5	3.5
Q7	どちらのシステムが話さなければならないと感じましたか	1	1	1	1	1	1	1.0

とで不安を感じた, なんとなく焦ったなどの意見があった. 一方 2 体条件の時には, 自分が何も話さなくても会話が継続されるので良いという意見があがった.

4.2.3 考察

Q7 の結果や, 被験者の感想からも示される通り, 自ら話さなければならないということが緩和されたことから, 会話に参加するタイミングがつかめれば, 2 体条件のほうが会話の継続が行いやすい可能性が考えられる. 会話に参加することができていれば, 被験者の発話回数が条件によって違いがあると考えられることから, 2 体条件が好評価な被験者 A, C, E と被験者 B, D, F の発話回数の平均を比較した. すると, 2 対条件の際の発話回数の平均は, 被験者 A, C, E の平均は 12.7 回, 被験者 B, D, F の平均は 10.0 回であった. 1 体条件における両グループにおける発話回数の平均はそれぞれ 25.3 回と同じことから, 被験者 A, C, E は会話に参加するタイミングを見つけられたため, 2 体条件に高評価をつけた可能性がある. したがって, 順番だけではなく, 会話へ参加できたか否かがエージェントとの話やすさや, 話を続けやすかったかの評価を分けた可能性もある.

4.3　同調制御の効果の検証 (実験 2)

4.3.1　実験条件

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で実験を行った. 被験者には, エージェント同士がコミュニケーションを行っている中に, 好きなタイミングで入ってくださいと指示した. これは, 被験者にエージェント同士の会話を観察してもらう時間を設け, エージェントらがどのような発話を行っているかを見てもらうためである. 使用したエージェントは, 各特徴量の平均値が定数の正規分布から生成される同調なし条件のエージェントと, 提案システムである同調あり条件のエージェントである. 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 回ずつ行い, 1 回終わる度に 5 段階でエージェントを評価してもらった. 実験を行う順番に関しては, 同調あり条件から始める場合と, 同調なし条件から始める場合それぞれが等しい人数となるようにした.

4.3.2　実験結果

被験者の各エージェントに対する 1 ～ 4 回の平均評価値を表 4.2 に示す. 表 4.2 の A∼F は被験者を, 同調とは同調あり条件, 非同調とは同調なし条件の際の結果の平均を表す. Q10, 11 において, 6 人中 5 人が同調あり条件の方が同調なし条件に比べて評価平均値が上回っている. このことから, 同調あり条件の方が協調的で会話ができたと感じることが示唆された. 一方 Q8, 9 に関しては, 6 人中 2 人のみ同調あり条件の方が同調なし条件に比べて評価平均値が上回るに留まった. これは被験者から, エージェント A（図 3.1 に示したエージェント A と同義）が, 自分の声に反応する頻度が高く, エージェント B（図 3.1 に示したエージェント B と同義）と話したいときにも反応するために, 話の邪魔をした, 話す頻度が多すぎるなどの意見が聞かれた. また被験者の中には, 同調なし条件のエージェントを使用した際

にも, 自分の真似をしようとしてきていると感じたとの意見もあった. エージェントが真似してきていると感じる場合では, 自分の発話に反応していると感じたなどの意見もあった. また, 被験者からは会話がこのような非言語情報で行えたら楽だが, もう少し意思疎通ができると嬉しいとの意見も多くあがった.

表 4.2: エージェントの評価平均値 (1: そう思わない, 5: そう思う)

		A		B		C	
質問番号	質問内容	同調	非同調	同調	非同調	同調	非同調
Q8	エージェント A は協調的でしたか	3.5	4.0	2.5	2.5	4.8	4.5
Q9	エージェント A と会話できましたか	3.8	4.3	2.5	2.8	5.0	5.0
Q10	エージェント B は協調的でしたか	3.5	3.0	4.0	3.8	3.8	2.8
Q11	エージェント B と会話できましたか	3.0	2.3	4.0	3.0	4.5	3.0
Q12	会話にスムーズに入れましたか	3.5	3.5	3.5	2.5	4.5	5.0
Q13	楽しかったですか	4.8	5.0	3.3	3.0	3.5	3.5
Q14	真似してきいると感じましたか	3.3	4.0	2.3	2.8	3.5	3.8
Q15	様々な「ぱ」は言えましたか	3.8	3.8	3.3	3.0	4.3	5.0
		D		E		F	
質問番号	質問内容	同調	非同調	同調	非同調	同調	非同調
Q8	エージェント A は協調的でしたか	3.5	3.3	3.0	4.0	4.3	4.5
Q9	エージェント A と会話できましたか	4.8	4.8	3.3	4.0	4.3	5.0
Q10	エージェント B は協調的でしたか	3.0	2.8	2.5	2.0	2.3	1.8
Q11	エージェント B と会話できましたか	2.5	1.3	2.8	3.0	2.5	1.8
Q12	会話にスムーズに入れましたか	3.3	2.8	2.5	2.0	3.0	3.5
Q13	楽しかったですか	3.3	3.0	2.8	3.0	4.0	4.3
Q14	真似してきいると感じましたか	1.3	1.8	2.0	2.0	1.8	2.0
Q15	様々な「ぱ」は言えましたか	3.0	3.0	3.5	4.0	3.0	3.0

4.3.3　考察

エージェント B が 6 人中 5 人, 同調あり条件のほうが評価が高かったことから, 同調制御を行う事により, 円滑なコミュニケーションが行える可能性が示唆された. しかし, エージェント A に関する評価は同調あり条件の方が同調なし条件より良いという結果にはならなかった. 被験者から, エージェント A は話す頻度が多いという意見から, 実験中の動画を確認すると, 被験者の発話とエージェント A の発話が被ってしまうことが多々確認された. このことから, 被験者は発話を邪魔された, 会話に入れてもらえないなどの感想を持った可能性が高いことがわかった.

第5章　結　　論

本研究では, 非言語コミュニケーションにおいて, コミュニケーションの維持が難しい問題を, 韻律的特徴を同調する 2 体のエージェントを用いて解決を目指した. 韻律的特徴の同調は, ユーザの発話の振幅, 基本周波数, セグメント長の変化に合わせるよう設定した. またユーザが沈黙を行った場合, エージェント同士がコミュニケーションを行うようにし, ユーザが何を話していいかわからない場合においてもコミュニケーションが継続されることを目指した.

本研究で作成したシステムを用いて, 評価実験を行った. 評価実験は 2 つ行い, 実験 1 ではエージェントが 1 体のみのシステムとの比較を行った. 実験の結果から, エージェントが 1 体のシステムが良いと答えた被験者と, エージェントが 2 体のシステムが良いと答えた被験者に分かれた. これは, どちらのシステムを先に使ったのかに依存するため, 今後, 長期的な試行実験が必要であると考えられる. また, 実験 2 では韻律的特徴の同調制御の効果を検証するため同調制御を行わないシステムとの比較を行った. 実験 2 の結果から, 6 人中 5 人は同調制御を行うエージェントの方が, 同調を行わないエージェントに比べて, 協調的だった, 会話を行うことができた, と回答した. このことから, 同調制御を行うことで, 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, 被験者の発話に重なることが多々あることや, 被験者が会話に参加するタイミングが無い場合など, エージェントの発話の頻度やタイミングに関しては課題が残った.

本研究では, 2 体のエージェントと人間が音の韻律的特徴のみを用いてコミュニケーションを行うシステムを提案した. 実験の結果から, 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示唆されたが, 2体にすることによってコミュニケーションを維持することができるという結論は得られなかった. しかし, エージェントの発話頻度や発話タイミングを考慮することによって, 円滑で継続しやすいコミュニケーションが行える可能性が見いだせたのではないかと考えられる. 今後は, 被験者の増加や, 発話頻度やタイミングにも着目し持続的非言語コミュニケーションの実現を目指したい.

参考文献

[1] 尾崎健太郎, 小暮悟, 甲斐充彦, 小西達裕, 伊藤幸宏, 複数の社内機器操作と雑談を扱えるマルチタスク音声対話システムのユーザビリティの向上, 情処学研報, Vol. 2010–SLP–80, No. 6, pp. 1–6, 2010.

[2] A. Ritter, C. Cherry and W.B. Dolan, Data-driven response generation in social media, *Proc. EMNLP*, pp. 583–593, 2011.

[3] 大林祐司, 川端豪, 雑談可能な目的達成型音声対話システム, 情処学研報, Vol. 2012–SLP–94, No. 9, pp.1–5, 2012.

[4] D. Anastasiou, K. Jokinen, and G. Wilcock, Evaluation of WikiTalk - User Studies of Human-Robot Interaction, *HCI international*, pp. 32–42, 2013.

[5] 大西可奈子, 吉村健, コンピュータとの自然な会話を実現する雑談対話技術, NTT DoCoMo テクニカル・ジャーナル, Vol. 21, No. 4, pp. 17–21, 2014.

[6] S. Gandhe and D. Traum, SAWDUST: A SemiAutomated Wizard Dialogue Utterance Selection Tool for domain-independent large-domain dialogue, *Proc. SIGDIAL 2014 Conference*, pp. 251–253, 2014.

[7] R.E. Banchs and S. Kim, An Empirical Evaluation of an IR-Based Strategy for Chat-Oriented Dialogue System, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–4, 2014.

[8] 狩野芳伸, コンピュータに話が通じるか：対話システムの現在, 情報管理, 59, 10, pp. 658–665, 2017.

[9] 山崎史生, 芋野美紗子, 土屋誠司, 渡部広一, 会話システムにおける質問文の回答予測による音声認識補正システム (人工知能と知識処理). 電子情報通信学会技術研究報告, 114, 481, pp. 29–34, 2015.

[10] 藤原 裕樹, 山下晃弘, 松林 勝志, Twitter から獲得した会話データに基づく雑談対話システムの開発, 情報処理学会第 78 回全国大会, 3M-01, pp. 115–116, 2016.

[11] 木村幸士, 湯浅将英, 武川直樹, 多人数エージェントによる会話の雰囲気生成, HAI シンポジウム, 1E–3, 2007.

[12] 藤堂祐樹, 西村良太, 山本一公, 中川聖一, 単一対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, 研究報告音楽情報科学, 19, pp. 1–8, 2012.

[13] 藤堂祐樹, 西村良太, 山本一公, 中川聖一, 複数の対話エージェントを用いた雑談指向の音声対話システム, 電子情報通信学会論文誌 D 99.2, pp. 188–200, 2016.

[14] 大川良希, 井上聡, 対話者の感情を推定した応答文生成のための言語モデルの提案, 人工知能学会, 1K3-5, 2015.

[15] 福田拓也, 若林啓, 雑談システムにおけるトピックの関係性を考慮した発話選択手法, 人工知能学会, 1L3-2in1, 2016.

[16] 西尾友佑, 堀部美那, 宮本善成, 村井翼, 韓東力, ユーザの発話履歴を考慮した会話文の自動生成, 言語処理学会, D3-5, pp. 646–649, 2012.

[17] 打田恭平, 大垣史迅, 鈴木健嗣, 伝達意図交信のみに基づくシグナル通信ネットワーク, 情報処理学会 インタラクション, pp. 257–262, 2012

[18] 小松孝徳, 山田誠二, 小林一樹, 船越孝太郎, 中野幹生, Artificial Subtle Expressions: エージェントの内部状態を直感的に伝達する手法の提案, 人工知能学会論文誌 25.6, pp. 733–741, 2010.

[19] 小林一樹, 船越孝太郎, 山田誠二, 中野幹生, 小松孝徳, 音声対話におけるロボットの外見と明滅光源 ASE がユーザの印象に与える影響, HAI シンポジウム, 1A–5, 2011.

[20] 船越孝太郎, 小林一樹, 中野幹生, 小松孝徳, 山田誠二, 対話の低速化と Artificial Subtle Expression による発話衝突の抑制, 人工知能学会論文誌, 26.2, pp. 353–365, 2011.

[21] 小松孝徳, 開一夫, 岡夏樹, 人間とロボットとの円滑なコミュニケーションを目指して, 人工知能学会誌, 17, 6, pp. 679–686, 2002.

[22] C. Trevarthen, U. Neisser, The self born in intersubjectivity: The psychology of an infant communicating, 1993.

[23] 有田亜希子, 開一夫, インタラクティブなロボットに対する乳児の認識, 東京大学大学院 総合文化研究科 広領域システム科学系 修士論文, 2002.

[24] 速水達也, 佐野睦夫, 向井謙太郎, 神田智子, 宮脇健三郎, 笹間亮平, 山口智治, 山田敬嗣, 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, 情報処理学会論文誌, Vol.54, No.8, pp. 2109–2118, 2013.

[25] T. Ono, T. Ichijo and N. Munekata, Emergence of Joint Attention between Two Robots and Human using Communication Activity Caused by Syn-

chronous Behaviors, *Proc. of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1187–1190, 2016.

[26] 小松孝徳, 長岡康子, ビープ音からコンピュータの態度が推定できるのか?: 韻律情報の変動が情報発信者の態度推定に与える影響, ヒューマンインタフェース学会誌, 7, 1, pp. 19–25, 2002.

[27] S. Song, S. Yamada, Expressing Emotions through Color, Sound and Vibration with an Appearance-Constrained Social Robot, *Human-Robot Interaction(HRI)*, pp. 2–11, 2017.

[28] 長岡千賀, M. Draguna, 小森政嗣, 中村敏枝, 音声対話における交替潜時が対人認知に及ぼす影響, ヒューマンインターフェースシンポジウム, 22, pp. 171–174, 2002.

[29] 長岡千賀, 小森政嗣, M. Draguna, 河瀬諭, 結城牧子, 片岡智嗣, 中村敏枝, 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, ヒューマンインターフェースシンポジウム, pp. 167–170, 2003.

[30] 西村良太 北岡教英, 中川聖一, 音声対話における韻律変化をもたらす要因分析, 日本音声学会 音声研究, 13, 3, pp.66–84, 2009.

[31] Arduino, `https://www.arduino.cc/`, 2018 年アクセス.

[32] Lightclip: ninja Ghost, `https://www.shapeways.com/product/ZPFG39RLC/lightclip-ninja-ghost-iphone-4-4s`, 2018 年アクセス.

[33] 森勢将雅, 河原英紀, 西浦敬信, 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.

業績一覧

- 2018 年 3 月 第 80 回情報処理学会全国大会, 学生セッション, 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, 小林一樹 (信州大), 北原鉄朗.
- 2018 年 3 月 日本音響学会, 春季研究発表会, イビキ音のマスキングに関する一検討, 伊藤春菜, 棚橋徹, 北原鉄朗.
- 2017 年 3 月 第 79 回情報処理学会全国大会, 学生セッション, 音によるヒューマン・エージェント・インタラクションのためのプロトタイプシステムの試作, 棚橋　徹, 小林一樹 (信州大), 北原鉄朗.
- 2017 年 3 月 第 79 回情報処理学会全国大会, 学生セッション, パターン認識を用いた特定のベーシストの特徴の分析, 松浦佳輝, 棚橋 徹, 北原 鉄朗.
- 2016 年 12 月 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Relations on Prosody of Japanese Back-channeling Word "Hai" and Listener's Impression: An Investigation with Synthesized Voices, Tetsu Tanahashi, Tetsuro Kitahara.
- 2016 年 9 月 日本音響学会, 2016 年秋季研究発表会, 相槌「はい」において聴取者が受ける印象について, 棚橋　徹, 北原　鉄朗.
- 2016 年 7 月 HCI International 2016, Support System for Improving Speaking Skills in Job Interviews, Tetsu Tanahashi, Yumie Takayashiki, Tetsuro

Kitahara.

- 2016年 5月 音学シンポジウム, 音声の韻律分析及び表情の特徴抽出による面接支援システム, 棚橋 徹，高屋敷 弓恵，北原 鉄朗.

謝　　辞

本研究に際して, 研究の進め方や, 論文の作成, 発表方法など様々な点でご指導を頂きました北原鉄朗准教授に深く感謝いたします. また本研究を進めるにあたり, 根幹となるアイデアや鋭いご指摘, 議論をしてくださった信州大学の小林一樹准教授にも厚く御礼申し上げます.

本論文の内容について, 議論, ご指摘, ご精読いただきました本学森山園子教授, 宮田章裕准教授に感謝いたします.

また, 本論文を作成するにあたり, 被験者集めや励ましをくださった斎藤明教授, 数々のご指摘, ご意見をくださった同級生や, 本研究室の後輩にも感謝いたします. 考えれば, 研究室の皆様にはたくさんの励ましや激励をいただきました. そのような支えがあったからこそ, このような研究が行え, また一生懸命努力ができたのだと身にしみて感じています. 急遽募集したのにも関わらず長い実験に参加してくださった皆様にも感謝致します.

最後に, この論文に関わった皆様に再度感謝申し上げます. 誠にありがとうございました.