

人工物による個人の人格を再現するための音声対話手法

竹内 玄 角 薫

概要

本論文では、人工物との会話を通して、想定する人の人格の再現に必要なデータを自然に引き出す技術の研究を紹介する。昨今、音声対話が身近なものになってきている。個人の特徴を理解している音声対話システムは個人に特化した内容の対話が可能であるが、個人の特徴を表した情報をたくさん入力するには手間がかかるため、手間をかけずに個人を再現する方法が必要である。本研究ではソーシャルスタイル理論を用いてユーザのコミュニケーションのスタイルを分類し、Pepper による音声対話機能、各種センサー、モーションを利用して対話システムを開発したものである。本システムの検証として個人の情報を正確に自然に引き出せていたかを調査するために被験者との対話による評価実験を行った。

キーワード：AI, 人格, 性格分析

Research on spoken dialogue technology to reproduce personal character

GEN TAKEUCHI KAORU SUMI

In this paper, we introduce research on technology to naturally derive data necessary for reproducing the personality of supposed person through conversation with artificial object. In recent years, spoken dialogue is getting familiar. If a spoken dialogue system that understands the characteristics of individuals, it is capable of dialogue of contents specialized to individuals. However it takes time and trouble to input a lot of information expressing the characteristics of individuals, so we need a method to reproduce an individual. In this research, we classified user's communication style by using social style theory and develop the system using speech dialog function, various sensors, and motion of Pepper. In order to investigate whether personal information could be drawn accurately and naturally as a verification of this system, we performed an evaluation experiment by dialogue with the subjects.

1 背景と目的

昨今、音声対話技術の発展に伴い音声対話が身近なものになってきている。身近な例ではスマートフォンの iPhone に搭載されているアプリケーションソフトウェアの Siri[1] がある。Siri は自然言語処理を用いて、質問に答える、推薦、Web サービスの利用などを行うアプリケーションであり、女性秘書のような擬似的な人格を持った AI と会話をしてリマインダー、天気予報、株式情報、テキストメッ

セージ、Eメール、カレンダー等のアプリケーションの操作を音声対話により行うことができる。また、IBM 社の Watson[2] は質問応答システム・意思決定支援システムで、ユーザーからの質問への自動応答、プロセスやアプリケーションについてのユーザーへの説明、難しい作業のガイドなどを音声対応により行うことができる。

個人の特徴を表した情報をたくさん入力すると個人に特化した内容の会話ができるがそれには手間がかかる。手間をかけずに個人を再現する方法が必要で

あり、そういった個人の情報を引き出すためには人工物と人間との間に信頼感が重要であり、それには身体性が重要になるのではないかと考える.[3]
そこで本研究では対話する相手の人格を再現するために必要な情報を対話の中から自然に引き出す技術を開発する。また、身体性に関して人型ロボットである Pepper を利用してデータを自然に引き出す技術を開発する。

2 関連研究

関連研究としては服部峻らの Web 知識を用いた時空間依存な対話システムの試作 [4] では対話システム内に擬似人格をあらかじめ用意しており、対話の情報をインターネット上から取得するという方式を取っているが、新たな情報を対話相手から引き出すということではできていない。また、対話の中から情報を引き出す手法として金子稜らの連想と常識判断を用いた会話からの話者情報取得法 [5] では対話内から対話相手に関する情報を引き出していくことができているが自然な対話内で情報を取得できてはならず、直接的な質問からしか情報を引き出せていない。Slater らの研究 [6] や Valster の研究 [7] では対話エージェントのロボットアバターに感情や感情的な行動を加える研究を行っており、現実の人間のように感情によって行動することでユーザーからより親しめるといった評価を得ていたが、まだまだ人間の反応と比べユーザーが違和感がある部分もあり用意された擬似人格も 1 パターンしかなかった。本研究では性格判断にソーシャルスタイル理論を用い、被験者のコミュニケーションの仕方を分析し、分析結果を話し方やモーションに当てはめた。

3 対話システムについて

3.1 Pepper について

本研究では pepper を利用する。Pepper (ペッパー) とは感情を認識できるという大きな特徴を持ったヒューマノイドロボットである [8]。pepper には音声認識機能や QiChat 等の機能を利用すれば対話システムを比較的簡単に構築することが出来る。また、pepper は入力した分をしゃべるときに勝

手に身振りをつけてくれる機能もある。この機能により人間に近い自然な対話が可能になる。

また、使用する機能として画像認識機能がある。画像認識を応用してユーザーの顔を認識して、その顔の画像からおおよそのユーザーの年齢を予測することが出来る。また同様にユーザーの顔の画像から表情を読み取りユーザーの感情を予測する機能も利用していく。

pepper を利用する利点として pepper が人型のロボットであるという点も挙げられる。人型であるためにモニターなどの無機質な非人型なハードよりもユーザーが抵抗なく使用できるといった点も期待できると考える。ユーザーが抵抗なく利用できると会話が進み、ユーザーから入力される情報も増えより多くの情報を入手できると考えられる。

4 個人の人格するための音声対話技術

本研究はユーザーがイメージする人物を想定者とし、想定者の人格を性格診断し、診断結果に最も近い擬似人格を提供する。そのあと、その対話システムとユーザーに簡単な会話してもらいユーザーの会話内の情報からさらに疑似的な人格を細分化させたり調整していくものを現在想定している。現在使用しようと考えているシステムは softbank 社提供の Pepper の開発者ツールの中にある QiChat Script の DiaLog を利用して音声対話システムを実現しようとしている。QiChat Script を利用すると Dialog ボックスを配置したときにある程度融通の利く対話システムを組み立てることができるからである。また、QiChat Script を利用したときに Pepper 内のシステムを利用すると DiaLog 内の音声認識機能も利用することができ、これを利用することで人間と対話しているときのようなより自然な対話が行え、またチャットボットなどのきーぼど入力に慣れていない高齢層などにも抵抗なく利用してもらえると考えた。

4.1 性格分析について

性格分析についてはソーシャルスタイル理論 [9] を利用する。ソーシャルスタイル理論とは 1970 年

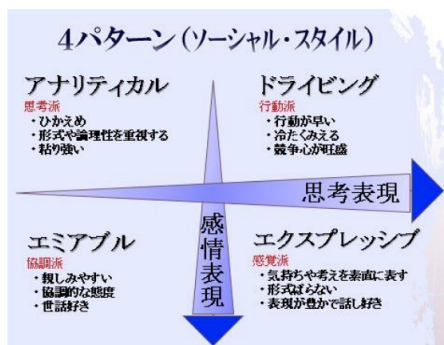


図1 ソーシャルスタイル理論

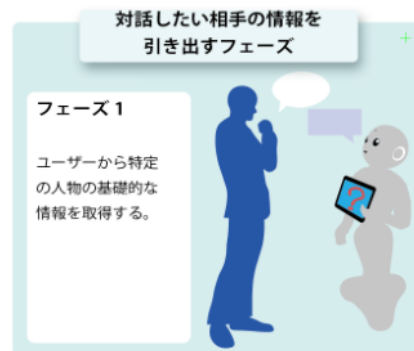


図2 フェーズ1

代に社会学者 David Merrill らが提唱した理論であり、人間がコミュニケーションをする時の4種類の傾向があり、それぞれアナリティカル（思考派）、ドライビング（行動派）、エミアブル（協調派）、エクスプレッシブ（感覚派）の四種であり、これらは人間の感情表現度と思考表現度で測る.[図2] このうち思考表現度はグラフの横軸で表され、まわりの方からみて、その方の考えや主張がどのくらい力強いのか、またまわりの方がその方にどのくらい影響されるか、ということの度合いを意味する。この値はユーザーへの質問をしていき、その答えで測る。また、ユーザーの質問に答える速度も参考にする。感情表現度はどのくらい自分の感情を表しているか、または他の人の感情を理解しているかわかる、ということの度合いを意味し、グラフの縦軸で表される。この値はユーザーとユーザーの想定者は似ていると仮定し、ユーザーの思考表現度の値を参考にする。ユーザーの感情表現度は単純にユーザーが感情を表情に出すかどうかを基準にする。ここでは pepper に搭載されている画像認識の表情検出機能を使い、表情の割合や、質問の回答により明るく気さくな人間かを確かめる。

4.2 Pepper との会話について

フェーズ1 [図2] は基礎的な情報を引き出す質問をユーザーにしていくフェーズである。このフェーズではまずユーザーについて Pepper が質問していく。想定者はユーザーの友人だと限定することによってユーザーと想定者の情報は酷似していると考えて、ユーザーの情報から想定者の情報を予測す

るためである。ここでの質問の例はユーザーの名前、ユーザーの出身地、想定者の名前、想定者とユーザーがいつ出会ったかなどである。例えば、

会話例1

Pepper 「あなたの名前は何ですか？」
 ユーザー 「私の名前は” □□” です」
 Pepper 「よろしくね。” □□” さん。あなたの話したい人の名前は何ですか？」
 ユーザー 「彼の名前は” ○○” さんです」
 Pepper 「” ○○” さんですね？少し” □□” さんと” ○○” さんについて質問をしていきますね」

このように質問していきユーザーと想定者の情報を引き出していく。

フェーズ2[図3] ではフェーズ1で取得した情報をもとに想定者の情報を予想し、その予想した情報を Pepper がユーザーと対話しながら確かめていく。また対話する際にはコールドリーディング [10] の技術を使いユーザーが対話したい相手の情報を引き出していく。コールドリーディングとは話術の一種であり、相手の事前の情報がなくても相手の外観や会話の中から相手の情報を引き出し、相手に自分が情報を持っているように思わせる話術である。この一つに性格を読むために”対極的な傾向”を述べるというテクニックがあります。例えば、

会話例2



図3 フェーズ2

Pepper 「〇〇さん（想定者）は優しく誰にでも手を差し伸べますことありますが、時に自分の中のルールに反する事には一切の妥協を許さない部分がありませんでしたか？」
ユーザー 「確かに彼は...」

このように客観的に見ると誰にでも当てはまる対極的な傾向とも言えることをいうことで相手に当てはまっていると思わせることができる。具体的な質問はあたかも Pepper が想定者の情報をすでに持っているかのように対話を続けていきさらに情報を引き出していく。例えば、画像認識にユーザーの年齢の情報を取得した後に想定者とユーザーの年齢は近いと予想して

会話例3

Pepper 「〇〇さん（想定者）の年齢は大体 20 歳くらいでしたね」
ユーザー 「いいえ」
Pepper 「あれ？違いましたか？すみません。ド忘れてしまったようです。何歳でしたっけ？」
ユーザー 「彼は 30 歳です」
Pepper 「そうでした。30 歳でしたね」

といったように情報の正誤性を高めていく。フェーズ3[図5]ではフェーズ1とフェーズ2で引き出した情報から割り当てた擬似的な人格と対話してもらい、そこで出た対話内の情報からさらに人格の情報

	思考派	協調派	感覚派	行動派
しぐさ	なし	あいづち	手を広げる	手を顎に
口調	丁寧 客観的 早い	穏やか 客観的 遅い	上から 主観的 早い	下から 主観的 遅い
セリフ例	あります	あるんだ	あるんですよ	あるんだよね

図4 ([11]より) ソーシャルスタイルとしぐさの関連

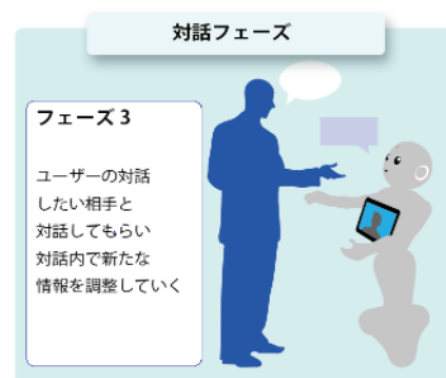


図5 フェーズ3

を追加したり調整していく。また、福井ら [11] が論文内で4種のソーシャルスタイルに合わせたキャラクターエージェントのモーションを作り、そのしぐさで各々のソーシャルスタイルの印象を与えていた。本研究でも pepper に福井らが提案したモーションを搭載してユーザーに印象を与える。[図4]

これらのフェーズ内の対話によって個人の人格に関わる情報を引き出していく。ユーザーのに対する質問を形成するときには相手の発言したうちの単語の種類から質問の種類を形成する。例えば、相手の会話で土地に関する単語が出てきた場合には土地に関する質問をする。特に質問の形式としてはできるだけ解放の質問の方式をとる。開放の質問とは質問に対する回答が YES/NO の二択で終わらないような質問のことであり、解放の質問をすることにより通常の質問よりも引き出せる情報が増え対話が続きやすくなる。また、これらの質問や対話を通して引き出した情報は擬似人格を形成する情報として [図

取得した情報	
名前	タケウチゲン
性別	男
血液型	A
あだ名	ゲンサン
出身地	北海道 - 札幌
好きな食べ物	らーめん
趣味	映画

図6 人格情報

	口調	標準モーション	文脈に合わせたモーション
思考派1	丁寧・早い	なし	なし
思考派2	丁寧・早い	なし	あり
行動派3	断定的・早い	手を顎に	なし
行動派4	断定的・速い	手を顎に	あり
協調派5	穏やか・遅い	相槌	なし
協調派6	穏やか・遅い	相槌	あり
感覚派7	気さく・速い	手を広げる	なし
感覚派8	気さく・速い	手を広げる	あり

図7 作成したモーションセット

6のように保存する。保存した単語は連想辞書を利用して新たな話題や質問を生成する。生成した新しい質問や話題による対話によって新たな情報を引き出し、今まで取得した情報の正確性を高める。

5 Pepperのモーションに関する予備実験

5.1 はじめに

本研究では予備実験としてPepperに用意した性格モーションが意図した機能を果たしているかことを確認するために行った。この評価を受けてシステムモーションの改良をした。

5.2 目的

本実験はPepperのモーションとセリフが意図した印象を与えることができているかのテストである。意図した印象とは、ソーシャル・スタイルの思考派、行動派、協調派、感覚派の4種類を意図したものである。口調、しゃべるスピード、モーションの組み合わせが異なる8種類のPepper用のモーションセットの好感度の順位とそれぞれのモーションがどのソーシャル・スタイルの印象を受けたかを調査し、本実験のために改良するために利用する。

5.3 方法

8種類のソーシャル・スタイルに合わせた口調、しゃべるスピード、モーションの組み合わせを持つPepperのモーションセットを作成した。[図7 8種類のソーシャル・スタイルに合わせたモーションセットにはそれぞれ1, 2, 3, 4, 5, 6, 7, 8を割り当てている。このうち1, 2が思考派, 3, 4が行動派, 5, 6が協調派, 7, 8が感覚派のモーションセットである。思考派のモーションセットは丁寧でゆっくりした口

調で話して、しぐさは行わない。行動派キャラクターエージェントは断定的で早い口調で話して、腕を組むしぐさを行う。協調派のキャラクターエージェントは穏やかにゆっくりと話して、あいづちを打つしぐさを行う。感覚派のキャラクターエージェントは気さくに早い口調で話して、手を広げるしぐさを行う。また、2, 4, 6, 8のモーションセットには、指定したモーションとモーションの間に文脈に合わせた自然なモーションを挟んだモーションセットである。これらの組合せを持たせたキャラクターエージェントが話しているシーンを見せたときに、各ソーシャル・スタイルのどのタイプだと感じたかを選択肢で選んでもらった。被験者は21~23歳の男性10人である。被験者には以下の4つの作業を行ってもらった。

- ・被験者のソーシャル・スタイル診断を行った。
 - ・4種類のソーシャル・スタイルの説明をして、4種類のソーシャル・スタイルを理解してもらった。
 - ・口調、しゃべるスピード、モーションの組合せが異なる8種類のPepperが話すシーンを見て、どのソーシャル・スタイルだと感じたかを選択肢で選んでもらった。
 - ・口調、しゃべるスピード、モーションの組合せが異なる8種類のPepperが話すシーンを見て、被験者が好感を持てる順番に並び替えてもらい、自由記述のアンケートを行った。
- また、Pepperがしゃべった内容の例としては以下のとおりである。

・思考派のキャラクタエージェント

「こんにちは、わたしはプレゼンターです。
今日はあなたにおすすめの観光都市を紹介したいとおもいます。
あなたはオーストラリアのケアンズという都市をご存知ですか？
それではオーストラリアの都市ケアンズのオススメスポットを 3 つ紹介させていただきます」

・行動派のキャラクタエージェント

「こんにちはわたしはプレゼンターです
オーストラリアの都市ケアンズのオススメスポットを 3 つ紹介します」

・協調派のキャラクタエージェント

「こんにちは、わたしはプレゼンターです。今日はおすすめの観光都市を紹介しますね。
あなたはオーストラリアのケアンズという都市を知っていますか。
それでは、オーストラリアの都市ケアンズのオススメスポットを 3 つ紹介するね。」

・感覚派のキャラクタエージェント

「こんにちは、わたしはプレゼンターです。今日はおすすめの観光都市を紹介しますよ。
あなたはオーストラリアのケアンズという都市を知っているかな。
それでは、オーストラリアの都市ケアンズのオススメスポットを 3 つ紹介するよ」

5.4 結果

結果としては [図 8] のようになった。思考派 1 については、10 人中思考派であると感じた被験者は 6 人。行動派であると感じた被験者は 3 人。協調派であると感じた被験者は 1 人。感覚派であると感じた被験者は 0 人で正答率は 60 % だった。思考派 2 については、10 人中思考派であると感じた被験者は 1 人。行動派であると感じた被験者は 5 人。協調派であると感じた被験者は 2 人。感覚派であると感じた被験者は 2 人で正答率は 10 % だった。行動派 3 に

	正答率	思考派	行動派	協調派	感覚派
思考派1	60%	6人	3人	1人	0人
思考派2	10%	1人	5人	2人	2人
行動派3	60%	4人	6人	0人	0人
行動派4	40%	4人	4人	2人	0人
協調派5	0%	6人	1人	0人	3人
協調派6	40%	3人	2人	4人	1人
感覚派7	30%	1人	1人	5人	3人
感覚派8	80%	0人	1人	1人	8人

図 8 モーションセットの印象結果

については、10 人中思考派であると感じた被験者は 4 人。行動派であると感じた被験者は 6 人。協調派であると感じた被験者は 0 人。感覚派であると感じた被験者は 0 人で正答率は 60 % だった。行動派 4 については、10 人中思考派であると感じた被験者は 4 人。行動派であると感じた被験者は 4 人。協調派であると感じた被験者は 2 人。感覚派であると感じた被験者は 0 人で正答率は 40 % だった。協調派 5 については、10 人中思考派であると感じた被験者は 6 人。行動派であると感じた被験者は 1 人。協調派であると感じた被験者は 0 人。感覚派であると感じた被験者は 3 人で正答率は 0 % だった。協調派 6 については、10 人中思考派であると感じた被験者は 3 人。行動派であると感じた被験者は 2 人。協調派であると感じた被験者は 4 人。感覚派であると感じた被験者は 1 人で正答率は 40 % だった。感覚派 7 については、10 人中思考派であると感じた被験者は 1 人。行動派であると感じた被験者は 1 人。協調派であると感じた被験者は 5 人。感覚派であると感じた被験者は 3 人で正答率は 30 % だった。感覚派 8 については、10 人中思考派であると感じた被験者は 0 人。行動派であると感じた被験者は 1 人。協調派であると感じた被験者は 1 人。感覚派であると感じた被験者は 8 人で正答率は 80 % だった。正答派が多いモーションは思考派 1、行動派 3、協調派 6、感覚派 8 であった。また、その中で正答数が 50% を超えたモーションは思考派 1、行動派 3、感覚派 8 であった。

また、印象が良かったのは感覚派 8 のモーションであった。

5.5 考察

思考派モーションについて2よりも1の方が評価が高かった。この理由としてはソーシャルスタイルの説明としてあまり動作が少なく機械的というものがあ、モーションがより少ない思考派1のモーションが思考派の印象を与えたと考えられる。なので思考派のモーションは思考派1のモーションをもとに制作する。特に動作は極力行わず機械的にするようにする。行動派モーションについては3よりも4の方が評価が高かった。この理由としては話が早く淡々としていると評価があったためだと考えられる。そこで、行動派のモーションは口調として早く淡々とした行動派3のもの元にして作成する。協調派モーションについては協調派6の方が5よりも高かった。この理由としては身振り手振りが多いのが協調派6だったのでより感じられたのではと考えられる。しかしどちらのモーションも過半数を超えることはなかったため、改善が必要と考える。具体的には問かけの様に、相談を持ちかける様に話し口調もよりゆっくりで穏やかにする。感覚派モーションについては感覚派7よりも感覚派8の方が高かった。身振りや、語尾、相手に質問したりが親しみやすさがあり、話している言葉がくだけてる感じがした。そのため、感覚派8のモーションを元に感覚派のモーションを作成する予定である。

6 今後の展望

pepperの実機を利用し対話システムやモーションのテストを行った。性格分析システムの作成も行った。用意した4種類のPepperのモーション[図9の印象がソーシャルスタイル各性格の印象であること確認するために行うPepperのモーションの予備実験を行った。現在は直接的な質問によりユーザーから情報を引き代しているが、より自然な会話から情報を引き出すことができると考えられる。具体的にはコールドリーディングなどの会話技術に対話システムに組み込み、相手の無意識の中に情報を引き出させるようになると考えられる。このような技術が発達することにより容易に現実の人間に近いAIを生み出せるようになると考えられる。そして、

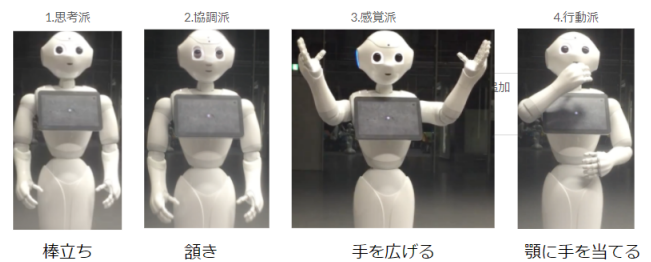


図9 Pepperのモーション

このシステムユーザーが年代や性別環境にかかわらず話したい相手と擬似的に会話できるようになると考えられる。

また、今後の実験予定としては、Pepperと対話してもらい、どれだけ情報を正確に取得できていたかを評価してもらうPepperと対話の評価実験を行う予定である。

参考文献

- [1] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, Russ Webb: "Learning from Simulated and Unsupervised Images through Adversarial Training", 2017
- [2] IBM solution: 「Watson 事例 日本航空 (JAL) バーチャルアシスタント「マカナちゃん」」, <<https://www.ibm.com/blogs/solutions/jp-ja/watson-jal-makana/>>,2017.
- [3] 神田 崇行, 宮下 敬宏, 長田 拓, 配川 有, 石黒 浩: 「人口ロボット相互作用における人型ロボットの外見の影響」, 『日本ロボット学会誌 Vol.24』 (2006) No.4 P497-505, 2006, 2010
- [4] 服部峻: 「Web 知識を用いた時空間依存な対話システムの試作」, 『電子情報通信学会技術研究報告. AI, 人工知能と知識処理』, 2010
- [5] 金子稜: 連想と常識判断を用いた会話からの話者情報取得法, 情報科学技術フォーラム講演論文集, 2014
- [6] Stuart Slater, David Burden: "Emotionally Responsive Robotic Avatars as Characters in Virtual Worlds", Games and Virtual Worlds

- for Serious Applications, 2009
- [7] Michel Valstar, Tobias Baur, Angelo Cafaro: "Ask Alice: An Artificial Retrieval of Information Agent", Proceedings of the 18th ACM International Conference on Multimodal Interaction Pages 419-420, 2016
 - [8] SoftBank Group All rights reserved: 「ロボット—ソフトバンク」, <<https://www.softbank.jp/robot/>>, 2017.
 - [9] David W Merrill, Roger H Reid.(1981) *Personal Styles & Effective Performance*. CRC Press.
 - [10] ジャック・シェーファー, マーヴィン・カーリンズ, 栗木 さつき (2015). : 『元 FBI 捜査官が教える「心を支配する」方法』. 大和書房.
 - [11] 福井 亮介, 角 薫: 「非言語情報を用いた説得するキャラクターエージェントシステム」, 『日本デジタルゲーム学会 2015 年度年次大会 (DiGRA2015)』, pp.133-136, 日本デジタルゲーム学会 (2016.2)