

音声生成を前提としたテキストレベルでのキャラクタ付与

沈 睿^{1,a)} 菊池 英明^{2,b)} 太田 克己^{3,c)} 三田村 健^{3,d)}

受付日 2011年6月25日, 採録日 2012年1月13日

概要: 近年, 擬人化インタフェースの研究開発がさかんに行われ, ユーザとそれらの人工物とのインタラクションが重要視されている. インタラクションを促進する手段の1つとして, ユーザにパートナー感や愛着を感じさせるため, エージェントのキャラクタ付与の必要性が高まっている. 筆者らは, 自動車内でロボティクス・エージェントが運転のサポートや情報コンテンツの提供などを行うインタフェースの開発を進めている. ここでは, 音声メディアの利用が前提となり, 音声表現によるエージェントのキャラクタ付与が必要である. 本稿では, ユーザにおけるキャラクタ像の形成を制御することを目的として, テキストレベルでのキャラクタ付与の手法を検討する. まず, (1) 自発音声に対する印象評定を行い, 高い評価が得られた音声の転記テキストを集めた. 続いて, (2) これらの転記テキストからキャラクタ像の形成に有効な言語パターンの抽出を行った. 最後に, (3) これらの言語パターンをニュートラル文章に適用してキャラクタ付与文章を生成し, それに対する印象評定実験を行って言語パターンの有効性を検証した. その結果, 抽出された言語パターンがキャラクタ像の形成に有効であることが分かり, 本研究のアプローチの有効性が示された.

キーワード: 自然言語処理, 印象評定, コーパス, 音声生成, 擬人化インタフェース

Towards the Text-level Characterization Based on Speech Generation

RAYMOND SHEN^{1,a)} HIDEAKI KIKUCHI^{2,b)} KATSUMI OHTA^{3,c)}
TAKESHI MITAMURA^{3,d)}

Received: June 25, 2011, Accepted: January 13, 2012

Abstract: Recently, the research in anthropomorphic interfaces is being conducted actively. Interactions between human users and those artifacts are considered of great importance. To make users have more partner feelings and favorable impressions, the characterization of agents, which is also one of the ways to enhance interactions, are becoming necessary. On the premise of using speech media, we attempt the characterization of Anthropomorphic vehicles interfaces that give routing and driving assistance by speech representation. In this paper, to control the formation of the types of impressions that perceived by users, the proposal is made to characterize anthropomorphic vehicles by text control. (1) Conducted an impression rating on spontaneous speeches and picked out the transcriptions of those speeches that were highly rated. (2) Extracted linguistic patterns that help to form impressions from the transcriptions. (3) Adapted those linguistic patterns into neutral texts to generate pattern-loaded texts, and using these texts, conducted an impression assessment to verify the effectiveness of those linguistic patterns. As a result, we found that those extracted linguistic patterns are effective in forming impressions and the approach in this study is considered feasible.

Keywords: natural language processing, impression assessment, corpus, speech generation, anthropomorphic interface

¹ 早稲田大学人間科学研究科
Graduate School of Human Sciences, Waseda University,
Tokorozawa, Saitama 359-1192, Japan
² 早稲田大学人間科学学術院
Faculty of Human Sciences, Waseda University, Tokorozawa,
Saitama 359-1192, Japan
³ 日産自動車総合研究所モビリティ・サービス研究所
Mobility Services Laboratory, Nissan Research Center,
Nissan Motor Co. Ltd., Atsugi, Kanagawa 243-0122, Japan

1. はじめに

近年, 擬人化インタフェースに関わる研究がさかんに

a) raymondshenrui@gmail.com
b) kikuchi@waseda.jp
c) ka-ohta@mail.nissan.co.jp
d) t-mitamura@mail.nissan.co.jp

われ、自動車メーカーにおいても、車載擬人化インタフェースの開発が進んでいる。ユーザと車載擬人化インタフェースとのインタラクションの促進の一環として、筆者らは、自動車内で安全で楽しい運転のサポート機能や情報コンテンツの提供機能などを小型のロボットなどで行う自動車用擬人化インタフェースの実現に向けた検討を行っている [1]。擬人化インタフェースを設計する際、ユーザにパートナー感や愛着を感じさせるキャラクター像 (2D エージェント、ロボット、車両など) を具現化することは重要な目標である。これまでエージェントの外見・動き [2], あるいはテキスト音声合成 (TTS, Text To Speech) [3], [4], 話し方 [5], 特にプロソディー [6], [7], [8], 音韻特徴 [7], [9], [10], [11] などによってキャラクター像を具現化しようとする研究が行われてきたが、発話内容 (TTS に与えるテキスト) によるキャラクター具現化の研究はまだ少ない。しかし、話者の外見、動きあるいは声の情報がいっさいなくても、たとえば日常発話 [12], 新聞記事 [13], 小説 [14] やブログ [15] を対象とした研究によって、テキスト情報のみから話者 (テキストの筆者) のキャラクタや属性を特定することが可能であることが分かっている。金水 [16] は、話者の人物像をすぐに思い浮かべることができる特定の言葉遣い (語彙・語法・言い回し・イントネーションなど) を「役割語」と定義した。たとえば、下記の2つの発話で、A は老人のキャラクタ、B は上品な女性のキャラクタを表現できる。

A: 「そうじゃ、わしが知っておる。」

B: 「そうですわよ、わたくしが存じておりますわ。」 [16]

しかしながら、このような表現が TTS の標準的な音声 (年齢、性別などの話者属性が反映されていない合成音声) で発せられれば聞き手は違和感をおぼえるであろう (将来的に合成音声の表現力が高まればこうした違和感がなくなる可能性はある)。したがって、文献 [12], [13], [14], [15], [16] のようにテキスト表現が人間に与える印象を扱う研究は数多くあるが、擬人化インタフェースの TTS システムからの出力において音声を用いることを前提とする場合、テキストレベルでの知見をそのまま適用することができない恐れがある。

そこで我々は、ユーザが擬人化インタフェースに対していざ印象を制御することを目的として、TTS による音声生成を前提としたテキストレベルでのキャラクタ付与の方法を提案する。提案する方法について図 1 を用いて説明する。本研究では、より自然なキャラクタを、擬人化インタフェースへ付与することを目指しているため、まず自発的な音声から着手した。第 1 に自発的な音声コーパスの多数の音声に対して印象評定を行い (Section 2: Rating), 特定のキャラクタ像を強く感じさせた音声の転記テキスト (Transcripts) を選定した。第 2 にそれらの転記テキストから特定のキャラクタ像の形成に寄与するテキストレベルの言語パターン (Patterns) を抽出した (Section 3:

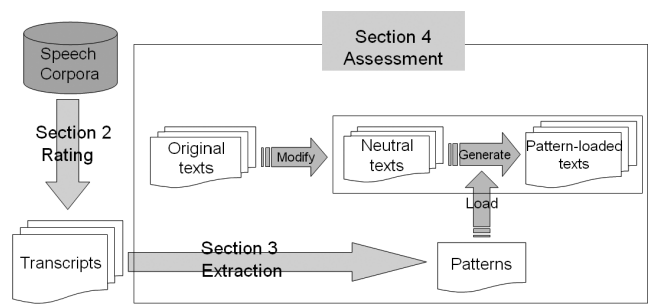


図 1 本研究の流れと応用

Fig. 1 Procedure and application.

Extraction)。なお、本研究ではキャラクター像の形成に有用な特定の言語表現や言葉遣いを「言語パターン」と呼ぶ。抽出された言語パターンを適用して (Load), 特定のキャラクター像を感じさせない文章 (Neutral texts. 以降、ニュートラル文章と呼ぶ) を特定のキャラクター像を感じさせる文章 (Pattern-loaded texts. 以降、キャラクタ付与文章と呼ぶ) に変換する (Generate) ことができる。この手法で作成したキャラクタ付与文章を TTS システムに入力する (TTS) と、特定のキャラクター像を感じさせる音声の合成が可能であるものとする。最後に、このキャラクタ付与方法の有効性を検証するため、ニュートラル文章とキャラクタ付与文章を用いて印象評定実験を行った (Section 4: Assessment)。すべての過程において、音声生成を前提としていることが本研究の特徴である。

以降、2章では、自発音声に対する印象評定の詳細について述べる。言語パターンの抽出と分類については3章で説明する。4章では、キャラクタ付与文章作成過程について説明し、言語パターンの、キャラクター像の形成に対する有効性を評価するための印象評定実験の詳細も述べる。最後の5章では本研究で行われた内容のまとめおよび今後の方向性と可能性の検討を行う。

2. 自発音声の印象評定

本章では、特定のキャラクター像を強く感じさせる音声の転記テキストの選定を目的とした、自発音声に対する印象評定について述べる。

2.1 評価者

本評定は、大学生男女 19 名の評価者による。

2.2 音声データ

本実験で使用した「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese; CSJ) [17] の模擬講演は談話のタイプがモノログであり、本研究が実現を目指す自動車用擬人化インタフェースにおける情報コンテンツ提供の発話と特性が類似している。さらに、「日本語話し言葉コーパス」のような大規模の音声コーパスであれば、話者の性別

や年齢のバランスも考慮可能である。CSJは、講演の自発音声からなる話し言葉研究用のデータベースであり、全体で約660時間の自発音声（語数にして約700万語）が格納されている。今回はその中の「模擬講演（Simulated Public Speaking; SPS）」を音声サンプル抽出の対象とした。SPSは、一般話者による日常的な話題についてのモノローグであり、聴衆は3～5名程度で、全体の傾向としてかなりリラックスした雰囲気では収録されている。収録する際、話者に原稿の読み上げは禁止されているが、講演内容のアウトラインを参照することは可能である[18]。我々はSPSから話者年齢と性別のバランスを考慮して30の音声ファイルを選び出した。また、これらの使用にあたり、10–12分の音声ファイルをそのまま使用するのとは長すぎると判断し、各音声ファイルの中盤約1分間の音声サンプルを切り出した。

2.3 キャラクタ像評価語

キャラクタ像の評価語として、「講演音声の印象評価尺度[19]」と、「成人の読み上げ音声における声質評価値と音響関連量の分析的検討[20]」を参考にしうえて、本研究の目的である自動車用擬人化インタフェース環境の需要を考慮して、「まじめさ」、「かわいさ/かっこ良さ」、「感情豊か」合計3評価語を扱うことにした。なお、本研究では、「かわいさ」は女性の、「かっこ良さ」は男性の好ましさを示す一般的な評価用語であると仮定して、音声サンプルの話者の性別によって切り替えている。

2.4 方法

順序効果を考慮して、19名の評価者を4つのグループに分け、グループごとに30サンプルをランダムに再生した。すべての音声サンプルの再生は1回に限定し、評価者は音声再生されている1分と、次の音声再生されるまでの15秒の間に評価を行う。こうすることで、できるだけ直感的な印象を答えさせることを狙った。なお、評定の前に、評価結果の安定性を保つために訓練として3回の評価を行った。設問は5段階評価（1：そう思わない、3：どちらでもない、5：そう思う）を用いた。疲労効果を軽減するため、評定は前半と後半に分け、間に5分の休憩を設けた。また、話題の影響を回避するため、それらは評価対象としないようにあらかじめ評価者に教示した。

2.5 結果

キャラクタ像の強さと次の処理に必要な転記テキスト数の確保の両方を考慮し、キャラクタ像評価語ごとの平均レーティングが4以上の音声サンプルの転記テキスト（全90件中の34件）を、次の処理における評価対象として選んだ。

3. 言語パターン抽出と分類

本章では、2章で選んだ音声サンプルの転記テキストと、それらから感じるキャラクタ像との関係を調べるため、1章で定義した「言語パターン」の抽出（特定のキャラクタ像の形成に寄与する表現や単語などの抽出）を行った。この作業を経て、特定のキャラクタ像を感じさせるテキストの生成（4.2節に詳述）が可能となる。

3.1 評価者

7名の評価者にキャラクタ像の形成に寄与すると思われる特定の単語や表現をマークしてもらった。評価者のうち3名は大学生で、4名は研究者であり、全員が2章の印象評定に参加している。

3.2 方法

評価者に2章で選定した音声の転記テキストを目視してもらい、「まじめさ」、「かわいさ/かっこ良さ」と「感情豊か」の3キャラクタ像の形成に役立った手がかりをあげてもらった。たとえば、転記テキストにある「絶対」という副詞が「感情豊か」のキャラクタ像の形成に貢献すると評価者が判断した場合、「絶対」をマークする。なお、単語レベルだけでなく、語句レベルや発話レベルでもマークすることを可能とした。時間制限は設けず、繰返しも可能とした。

3.3 結果

「3.2節方法」でマークされた結果から手作業で言語パターンを抽出した。より多くの言語パターンを抽出するほど、それから生成されるテキストのバリエーションが豊かになる。そのために、基本的には半数以上（4名以上）の評価者が同一箇所マークした箇所を重視したが、それだけに限定せず、半数未満（3名以下）の評価者のみからマークされた箇所も、キャラクタ像の形成に寄与すると判断した場合は参考にした。抽出できた63通りの言語パターンのうち、半数以上の評価者によるものは43通りで、全体の約7割である。続けて、抽出された言語パターンの全体の発話中の出現位置（セグメントの頭、文中、あるいは句末など）をもとに、形態論や語用論などの言語現象レベルの区分と、品詞や活用などの形態論情報を言語パターンに付与した。さらに、キャラクタ像の形成に有効だと思われる理由の推測などのコメントも情報として付与した。最後に、次の評価実験（4章に詳述）で用いるテキストを生成するため、抽出できた言語パターンの分類を行った。そのうち、統語論的分類（表の横軸：品詞など）とパターン適用時の操作による分類（表の横軸：置換あるいは挿入）を表1に示す。なお、半数未満の評価者がマークした言語パターンを「*」で示す。

表 1 抽出したパターンと分類
Table 1 Extracted patterns and grouping.

	助詞 Particle	副詞 Adverb	接続表現 Conjunction	接頭・接尾 Prefix/Suffix	オノマトペ Onomatopoeia	口語・音便 Colloquial/Euphonic changes	人称代名詞 Personal pronoun	その他 Others
かわいさ/かっこよさ Cuteness/Coolness	置換 In other words		かなり/たくさん/いっぱい*	それで/したがって>で*		てしまう>ちゃう/じゃう られる>れる*		このように>こう(挿発)
	挿入 Insertion	～よ ～よね	いっぱい* 凄(偏副詞, 挿発)	～や～とか*	お(食べ物) さん(擬人化対象)	ひゅって～		～んです
まじめさ Seriousness	置換 In other words			～た>～ました ～て>～まして			俺>僕*	その時>当時 た.です>である
	挿入 Insertion		取り敢えず*	～(れ)ど>～(れ)ども そこで として*				状況でした まじめに
感情豊か Expressiveness	置換 In other words		やはり>やっぱり あまり>あんまり* 少し>ちよつと>ちよつぱり			で(で)いて>て(て)で* なけれは>なさき* ではないか>じゃん(若者口調) てしまう>ちゃう/じゃう* ～の～が～ん～(れ)ど(も)		どうして>なぜ>何で...だろう 手出す>ちよつと出さ* このように>こう(挿発) そのような>そういう>それほどの>そんな
	挿入 Insertion	～よ ～よね ～かな(確信できない) ～(です)ね	やっぱり 全然(否定強調)* 凄く あんまり* 結構(ほど) 絶対(否定強調) かなり* 凄(偏副詞, 挿発) ずっと(状態継続)	～と～(た)～(とう)* ～や～とか そしたら(挿発)*		ぼつと* ひゅって びして	そっかじゃ ～ろっかな(未然形)	～たまらない(程度) ～じゃないですか(共感を求める) 何か* どのような(挿発) ～んです

4. キャラクタ付与方法検証のための印象評定実験

本章では、2章と3章の手続きを経て構築した「言語パターン」が、実際のキャラクタ像の形成にどの程度貢献するのか、実際にそれらを適用して文章を作成し、文章のテキスト表記のみに対する印象評定によって効果を評価した。

4.1 評価者

今回の実験では、11名の評価者が参加した。評価者全員が2章の印象評定実験と3章の作業には関わっていない。

4.2 実験用テキスト

以下に示す原文章、そしてそれを変換したニュートラル文章とキャラクタ付与文章の3種類のテキストを扱った。

(1) 原文章

自動車用擬人化インタフェースでの応用を考慮し、今回は実際のカーナビゲーションのTTSシステムで使われた横浜市の観光案内文章を原文章として使用した。1文章は150文字程度で、3-5の節からなる。内容は横浜市内1カ所の観光地の紹介である。このような文章を9個、原文章として選んだ。

(2) ニュートラル文章

1章でも述べたが、本稿では「ニュートラル文章」を、理想的で特定のキャラクタ像を感じさせないテキストとして定義する。我々はすべての原文章を単語、フレーズ、文の変換を通じてニュートラル文章に変換した(図2の左にニュートラル文章の例を示す)。変換する際に、単語自体が特定のキャラクタ像を感じさせないように注意した(動詞を原形に変換、形容詞の選択など)。

(3) キャラクタ付与文章

さらに、特定のキャラクタ像を感じさせることを目的として言語パターンを適用した「キャラクタ付与文章」を生成した。その際、なるべく多くの種類の言語パターンを用いるようにした。その理由として、本研究の第1目標は、

Neutral text

横浜ワールドポーターズ。輸入商品をメインとする店舗が多く入っている。雑貨店から、ファッションショップ、飲食店まで幅広くある店舗は、1つ1つを見て回るだけでも楽しく時間が過ぎせる。横浜に観光で訪れた時には土産を買うスポットとしてもお勧めできる。

Pattern-loaded text (cuteness)

横浜ワールドポーターズ。輸入商品をメインとするお店屋さんが凄く多く入っているんですよ。雑貨屋さんやファッションショップや飲食店とか凄く幅広いお店屋さんは、1つ1つを見て回るだけでも楽しく時間が過ぎせるんですよ。横浜に観光で訪れた時にはお土産を買うスポットとしてもお勧めできるんですよ。

図2 ニュートラル文章とキャラクタ付与文章のサンプル(「かわいさ/かっこ良さ」)

Fig. 2 A sample of neutral texts and pattern-loaded texts (Cuteness).

まず言語パターンを利用したキャラクタ像の形成の有効性を示すことなので、言語パターン間の関係は考慮していないからである。各言語パターンにおける特定のキャラクタ像の形成への貢献度を検証するには、キャラクタ付与文章の生成方法を精緻化する必要があると考えており、それは今後の課題である。

具体的には、「かわいさ/かっこ良さ」、「まじめさ」と「感情豊か」の3種類のキャラクタ像を検証するため、合計9(3×3短文章)キャラクタ付与文章を生成した。1文章に適用した言語パターン数の平均は8.22で、標準偏差は3.07である。9文章に対する言語パターンの適用状況を表2に示す。「3.3節結果」で述べたように、本研究で抽出した言語パターンは半数以上の評価者によるものと半数未満の評価者によるものがある。表2においては半数以上の評価者がマークした言語パターンの数を示し、各キャラクタ像の言語パターン全体に対する割合も示している。この表により、半数以上の評価者がマークした言語パターンの割合は「かわいさ/かっこ良さ」で0.87、「まじめさ」で1.00、「感情豊か」で0.86であり、結果的には半数以上の評価者により共通してマークされた言語パターンがどのキャラクタ像においても多く適用されていることを確認した。

表 2 キャラクタ付与文章に適用された言語パターン数

Table 2 The numbers of loaded patterns in the pattern-loaded texts.

“Major” denotes a number of the loaded patterns that marked by major of the evaluators (4 or more) in a single text
 “Total” denotes a number of all of the loaded patterns in a single text

Impressions	Text No.	Major	Total	Major/Total
Cuteness	1	6	7	0.87
	2	9	11	
	3	11	12	
Seriousness	1	6	6	1.00
	2	4	4	
	3	5	5	
Expressiveness	1	9	10	0.86
	2	9	11	
	3	6	7	

図 2 右の部分は 3 章の手順で抽出した「かわいさ/かっこ良さ」のキャラクタ像の形成に有効だと思われる言語パターン (図の斜体部分) を適用して生成したキャラクタ付与文章である。

4.3 実験方法

本研究では、音声生成への応用を前提としているため、刺激文章の提示方法も工夫した。音声を聴覚で聴くのに近い環境を、文字を見る際にも再現するため、PC のスライドアニメーションを用いて、刺激文章を一気に提示するのではなく、1 単語ずつ提示した。提示のスピードも事前に十分に検討され、人間の文章を見る速度にあわせて調節した。さらに、1 つの刺激文章の提示が終了すると、文章全体がすぐ消え、再度の目視を防いだ。刺激文章の提示順はランダムである。評価者に提示される文章を見てもらい、感じたキャラクタ像を答えてもらった。絶対評価と相対評価を両方確認するため、先に絶対評価、次に相対評価の順序で実施した。以下に、印象評定方法に関するパラメータと、それぞれの具体的な実験方法について述べる。

(1) 印象評定方法に関わるパラメータ

図 3 は印象評定方法に関わるパラメータの詳細である。評価者に提示する刺激 (input) に関しては、1) 刺激文 (sentence) は単一か、複数か、2) 対象キャラクタ像 (impression) は単一か、複数か、3) 言語パターン (pattern) が単一か、複数か、4) ニュートラル文章 (neutral) 提示の有無、のパラメータがある。また、形成されたキャラクタ像 (output) に関しては、a) 回答の対象キャラクタ像 (impression) は単一か、複数か、b) 設問を選択式 (choice) にする場合、単一選択か、複数選択か、c) 設問にスケール (scale) を用いるかどうかなどのパラメータが考えられる。

後述の印象評定実験では、キャラクタ付与文章に対する期待されたキャラクタ像が形成されたかどうかを調べるため図 3 下部にある I (絶対評価) を行った。以下に、本実験で設定した印象評定方法のパラメータを詳述する。

		impression			output	
		sentence	pattern	neutral	choice	scale
input	single	single	1	0	1	0
		multi	1		III	
	multi	single	1		IV	
		multi	0		II	

input	texts as stimuli
output	formed impressions
sentence	stimuli sentences
impression	target types of impressions
pattern	extracted patterns in Section 3
neutral	neutral texts
scale	assessment in a five-point scale

I	absolute assessment in Section 4.3 (2)
II	relative assessment in Section 4.3 (3)
III, IV	maybe the next steps of this study

図 3 印象評定方法に関わるパラメータ制御

Fig. 3 Parameters control in assessment.

刺激:

- 1) 1 文ではなく、まず全文章での効果を検証するため、刺激文を複数にした。
- 2) まず 1 種の対象キャラクタ像だけに着目するため、対象キャラクタ像を単一にした。
- 3) 4.2 節の (3) でも述べたように、本研究の第 1 目標は、まず言語パターンを利用したキャラクタ像の形成の有効性を示すことであるため、言語パターン間の関係を考慮せずに複数にした。
- 4) 絶対評価であるため、ニュートラル文章を提示しないことにした。

形成されたキャラクタ像:

- a) 評価者に悟られないよう、評価対象とされたキャラクタ像のほか、ダミーとされたキャラクタ像 (4.3 節の (2) で詳述) も含めるため、回答の対象キャラクタ像を複数にした。
- b) 回答の対象キャラクタ像を複数にしたため、設問を複数選択式にした。

絶対評価のほか、ニュートラル文章との比較でキャラクタ付与文章によるキャラクタ像の形成のされやすさが分かるため図 3 下部にある II (相対評価) を行った。以下に、設定した印象評定方法のパラメータを詳述する。

刺激:

- 1), 2), 3) において、上に述べた絶対評価 (I) と同じ理由でパラメータを設定した。
- 4) 相対評価であるため、ニュートラル文章を提示することにした。

形成されたキャラクタ像:

- a) まず 1 種の対象キャラクタ像だけに着目するため、回答の対象キャラクタ像を単一にした。
- b), c) 各評価対象とされたキャラクタ像の形成度を調べるため、設問に選択式ではなくスケールを用いた。

(2) 絶対評価

絶対評価では、キャラクタ付与文章のみを用いる。評価

表 3 絶対評価の混同行列結果

Table 3 Confusion Matrix of the results in the absolute assessment.

		Predicted		
		Expressiveness	Cuteness	Seriousness
Actual	Expressiveness	17	5	1
	Cuteness	3	16	0
	Seriousness	5	8	29
	others	16	9	7

表 4 絶対評価の結果

Table 4 Precision, Recall and F-value of the absolute assessment.

	Expressiveness	Cuteness	Seriousness
Precision	0.52	0.48	0.88
Recall	0.74	0.84	0.69
F-value	0.61	0.62	0.77

者に合計 9 (3 キャラクタ像 × 3 文章) のキャラクタ付与文章が提示され、感じたキャラクタ像を複数選択で選んでもらった。選択肢として、今回の実験で対象とされた「まじめさ」、「かわいさ/かっこ良さ」と「感情豊か」のほか、ダミーとして、モノローグ音声の印象評価尺度 [19] を参考にして選出した 5 評価語 (都会的, 活動性, 上品さ, 好悪, 癒し) も含めた。

(3) 相対評価

相対評価では、ニュートラル文章とキャラクタ付与文章を用いる。評価者には 9 個のニュートラル文章と 9 (3 印象 × 3 文章) 個のキャラクタ付与文章が一对一提示され、2 章で用いた 3 キャラクタ像評価語に対するどのくらい感じたかについて 5 段階評価 (1: ニュートラル文章のほうが形成されやすい, 3: 同じ, 5: キャラクタ付与文章のほうが形成されやすい) で答えてもらった。なお、提示順序はニュートラル文章を先とした。

4.4 結果と考察

絶対評価と相対評価の結果を以下に示す。はじめに、絶対評価の結果を表 3 に示す。

縦軸は評価者が感じたキャラクタ像、横軸は刺激に適用した言語パターンのキャラクタ像を示す。表の各数値は、全部で 33 回 (1 キャラクタ像につき 3 文章 × 11 名評価者) のうち、正解が選ばれた回数を示す。各項目は、‘Expressiveness’ が「感情豊か」、‘cuteness’ が「かわいさ/かっこ良さ」、‘seriousness’ が「まじめさ」、‘others’ がそのほかのダミー選択肢を意味する。

表 4 は絶対評価結果の適合率、再現率と F 値を示す。「かわいさ/かっこ良さ」と「感情豊か」に関して、適合率はやや低い (順に 0.48, 0.52) ことに対して、再現率 (順に 0.74, 0.84) と F 値 (順に 0.62, 0.61) は比較的高かったが、「まじめさ」に関して、いずれにおいても高い評価を得られた (適合率: 0.88, 再現率: 0.69, F 値: 0.77)。その理由として、絶対評価のアンケートに設けた選択肢がそれ

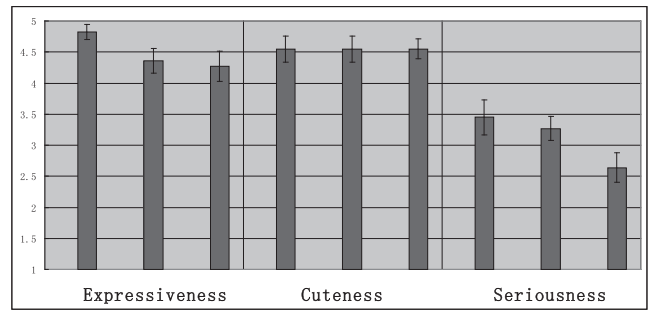


図 4 相対評価の結果—平均と誤差 (9 本の棒は刺激の 9 キャラクタ付与文章を意味する)

Fig. 4 Results of the relative assessment—average and error (each bar represents one pattern-loaded text).

ほど独立していないことが考えられる。具体的には、「かわいさ/かっこ良さ」と「感情豊か」に関しては、ダミーとしての 5 評価語 (「都会的」、「活動性」、「上品さ」、「好悪」、「癒し」) の意味空間と重なる (たとえば、「かわいさ/かっこ良さ」と「好悪」、「感情豊か」と「活動性」など) ことが考えられ、それにより「かわいさ/かっこ良さ」と「感情豊か」の適合率が低くなった可能性がある。評価実験におけるダミーとしてのキャラクタ像の選び方は今後の課題であるが、その中でも「かわいさ/かっこ良さ」と「感情豊か」の再現率と F 値を見る限り、比較的良好な結果が得られたと考える。

図 4 は相対評価の結果である。図 4 の縦軸の数字は 5 段階評価に対し、11 評価者の平均採点を意味し、横軸は対象のキャラクタ像評価語のグループであり、それぞれのグループの 3 つのバーは相対評価実験で用いられた刺激のキャラクタ付与文章を意味する。各バーのエラーバーは標準誤差を示す。図 4 より、「感情豊か」と「かわいさ/かっこ良さ」の両キャラクタ像において、5 段階評価の 3 (ニュートラル文章とキャラクタ付与文章のキャラクタ像の形成度が同じ) と比較して高い評価 (それぞれの平均得点は 4.48 と 4.55 であり、4 から 5 の間に集中する) が得られた。それに対し、「まじめさ」の得点は 3 よりもやや高い程度であった (平均得点 3.12 であり、2 から 4 の間に集中する)。これは、ニュートラル文章自体が「まじめさ」のキャラクタ像を想起させる文章である可能性が考えられる。そのため、ニュートラル文章と「まじめさ」のキャラクタ付与文章の印象空間における距離が、他のキャラクタ付与文章よりも近かったと解釈できる。したがって、「まじめさ」に関してニュートラル文章の変換の仕方を改善する余地があるものの、「感情豊か」と「かわいさ/かっこ良さ」のキャラクタ像の形成において、本提案手法の有効性が示された。

5. むすび

本稿では、ユーザと車載擬人化インタフェースとのインタラクションの促進の一環であり、自動車用擬人化インタ

フェースにおいてキャラクタ像を具現化することを目的とし、音声生成を前提としたテキストレベルでのキャラクタ付与方法を提案し、ユーザが擬人化インタフェースに対して印象の制御方法を検討した。

まず、自発性の高い音声コーパスから、印象評定によって特定のキャラクタ像を感じさせる音声の転記テキストを選出した。次に、得られた転記テキストを分析し、特定のキャラクタ像の形成に有効な「言語パターン」の抽出を行った。続いて、これらの言語パターンを用いて、特定のキャラクタ像を感じさせる「キャラクタ付与文章」を生成した。最後に、原文章から変換した特定のキャラクタ像を感じさせない「ニュートラル文章」と「キャラクタ付与文章」の両方を用いて印象評定実験を行うことによって、キャラクタ像の形成における言語パターンの有効性を確認した。

上記一連の手続きは、最終的に音声生成されることを前提としている。有効性が確認された言語パターンを用いれば、たとえば、「感情豊か」のキャラクタ像の形成に有効な言語パターンの適用したテキストをTTSシステムに入力し、インタフェースの外見、動きあるいは声などの要素も調整すれば、「感情豊か」というキャラクタが付与された擬人化インタフェースが実現可能である。またTTS技術においてより自発音声に近い音声合成が可能になれば、自発音声を前提とした提案方法が有効にはたらくと考えられる。

本稿での目標は、我々が提案した手法の有効性を示すことなので、4章の印象評定実験においては、言語パターン間の関係を明らかにすることは目的としなかった。しかし、実用化を考えると、言語パターンごとに、キャラクタ像の形成への貢献度は異なると考えられる。したがって、擬人化インタフェースへのキャラクタ付与の実用化に向けては、4.3節の(1)で述べたように、今後の課題として、各言語パターンのキャラクタ像の形成における貢献度を確認し、異なる言語パターン間の相乗効果の有無を解明する。たとえば、各キャラクタ像の形成における単一の言語パターンの貢献度を確認するため、図3のIIIの方法が適切なのではないかと考えている。さらに、図3のIVの方法により、言語パターン間の相乗効果が解明できるのではないかと考えている。言語パターンの有効性を定量化して示すことができれば、テキストへの適応および生成がより容易かつ効率的になると考えている。

参考文献

- [1] 長田純一, 坂井 晃, 尾藤真記子, 三田村健, 井上真人, 藤田善弘: PIVO2におけるロボティックエージェントのインタラクションデザイン開発, ヒューマンインタフェースシンポジウム論文集 CD-ROM, No.1225 (2008).
- [2] 河野 泉, 吉阪主句, 上窪真一: 擬人化キャラクタインタフェース構築支援: キャラクタコンポーザの開発, 日本情報処理学会全国大会講演論文集, Vol.1997, pp.87-88 (1997).
- [3] Rui, A., Rebordao, F., Shaikh, M.A.M., Hirose, K. and Minematsu, N.: How to Improve TTS Systems for Emotional Expressivity, *Proc. INTERSPEECH 2009*, Vol.2009, pp.524-527 (2009).
- [4] 山下洋一, 峯松信明, 徳田恵一, 小林隆夫, 広瀬啓吉: 擬人化音声対話エージェントにおける音声合成, 音講論集, Vol.1, pp.33-34 (2002).
- [5] Eskenazi, M.: Changing speech styles: Strategies in read speech and casual and careful spontaneous speech, *IC-SLP 92*, Vol.1992, pp.755-758 (1992).
- [6] Grimaldi, M. and Cummins, F.: Speech Style and Speaker Recognition: A Case Study, *Proc. INTERSPEECH 2009*, Vol.2009, pp.920-923 (2009).
- [7] 阿部匡伸: 異なる発話様式の特徴分析とその制御, 音響誌, Vol.51, pp.929-937 (1995).
- [8] 宮武正典, 匂坂芳典: 種々の発話様式にみられる韻律特徴とその制御, 信学論, Vol.73, pp.1929-1935 (1990).
- [9] 海木延佳, 匂坂芳典: 種々の発話様式における音韻継続時間長の分析, 音講論集, Vol.1991, pp.225-226 (1991).
- [10] 樋口宜男, 平井俊男, 匂坂芳典: 音声基本周波数モデルパラメータに対する発話様式の影響, 信学会講論集, Vol.1, pp.488-489 (1994).
- [11] 原 紀代: 種々の発生における時間長制御の分析, 音講論集, Vol.1991, pp.227-228 (1991).
- [12] 定延利之: 発見の「た」と発話キャラクタ, 言語, Vol.36, No.12 (2007).
- [13] 白井 諭, 阿部さつき, 矢部孝幸, 久保京子, 池原 悟, 横尾昭男: 新聞記事日本文における書き替え対象表現の分布, 言語処理学会第2回年次大会, Vol.1996, pp.37-40 (1996).
- [14] 望月朝香, 鈴木泰博: 小説における文体印象解析の試み, 情報処理学会研究報告 MPS, 数理モデル化と問題解決研究報告, Vol.2007, pp.179-182 (2007).
- [15] 岸本千秋: ウェブ日記に見られる話しことば的文体, 日本語学, 明治書院 (2008).
- [16] 金水 敏: ヴァーチャル日本語 役割語の謎, 岩波書店, 東京 (2003).
- [17] Maekawa, K., Koiso, H., Furui, S. and Isahara, H.: Spontaneous Speech Corpus of Japanese, *Proc. 2nd International Conference of Language Resources and Evaluation (LREC2000)*, Athens, Vol.2000, pp.947-952 (2000).
- [18] The Corpus of Spontaneous Japanese, available from (<http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/public/index.j.html>).
- [19] 山住賢司, 籠宮隆之, 榎 洋一, 前川喜久雄: 講演音声の印象評価尺度, 日本音響学会誌, Vol.61, pp.303-311 (2005).
- [20] 山下泰樹, 松本 弘: 成人の読上げ音声における声質評価値と音響関連量の分析的検討, 日本音響学会誌, Vol.62, pp.856-864 (2006).
- [21] Neubig, G., Mori, S. and Kawahara, T.: A WFST-Based Log-Linear Framework for Speaking-Style Transformation, *Proc. INTERSPEECH 2009*, Vol.2009, pp.1495-1498 (2009).



沈 睿

2005年中国華東師範大学中国語教育学科卒業，2009年早稲田大学大学院人間科学研究科言語情報科学修士課程修了．同年同大学院博士課程入学．音声言語コーパス，言語教育の研究に従事．



菊池 英明 (正会員)

1991年早稲田大学理工学部電気工学科卒業，1993年同大学大学院修士課程修了．同年(株)日立製作所中央研究所入社．早稲田大学理工学総合研究所助手，国立国語研究所非常勤研究員，早稲田大学人間科学部非常勤講師・専任講師を経て，2002年より早稲田大学人間科学学術院准教授．博士(情報科学)．音声言語，音声対話，ヒューマンインタフェースの研究に従事．人工知能学会，日本音響学会，ヒューマンインタフェース学会等各会員．



太田 克己

1997年福井大学工学部電気・電子工学科卒業．1998年日産自動車総合研究所電子情報研究所入社．2001年同車両交通研究所，2004年同第四技術研究所，2006年同モビリティ研究所，2010年同モビリティ・サービス研究所配属となり現在に至る．車両のIT化を中心に，近年は2007年東京モーターショーに出展したPIVO2に搭載されたロボティックインタフェースを車両適用するための研究開発に従事．社団法人自動車技術会員．



三田村 健

1990年慶應義塾大学大学院理工学研究科物理学専攻修了(修士)，同年日産自動車に入社，中央研究所電子研究所に配属．半導体MEMSセンサ，コックピットシステム，車両用Robotインタフェース研究に従事．(社)自動車技術会会員．