

会話エージェントによる情報提供 非言語情報による重要概念の強調

PROVIDING INFORMATION THROUGH CONVERSATIONAL AGENTS:
EMPHASIZING IMPORTANT CONCEPTS USING NONVERBAL INFORMATION

中野 有紀子¹・村山 敏泰²・西田 豊明³

¹MS. (Media Arts and Sciences) 社会技術研究システム ミッションプログラム 統括研究グループ
会話型知識プロセス研究サブグループ 研究員 (E-mail: yukiko@ristex.jst.go.jp)

² 社会技術研究システム ミッションプログラム 統括研究グループ
会話型知識プロセス研究サブグループ 研究員 (E-mail: tmuraya@ristex.jst.go.jp)

³工学博士 京都大学大学院情報学研究科 (E-mail: nishida@i.kyoto-u.ac.jp)
社会技術研究システム ミッションプログラム 統括研究グループ
会話型知識プロセス研究サブグループ リーダー

共通認識を確立するための情報提供は安心・安全な社会作りには不可欠である。本論文では、社会への情報提供をより円滑に行うことを目指し、自然な動作、特に適切にジェスチャーを行う会話エージェントの提案とその情報提供システムへの統合について述べる。まず、実際のプレゼンテーションのビデオデータを分析し、分析結果に基づき人間のジェスチャー使用についてのモデルを確立する。さらに、このモデルをエージェントによるプレゼンテーション自動生成システムに統合し、その有効性について情報工学、社会技術の両側面から議論する。

キーワード：会話エージェント，ジェスチャー自動生成，マルチメディアコミュニケーション環境，コミュニケーション科学，Empirical Study。

1. はじめに

安心・安全な社会を実現するためには、共通認識の確立、知識共有は必要不可欠であり、そのための教育や社会コミュニケーションは、重要な活動である。我々は情報工学的な知見から、このような活動を支援するコミュニケーション基盤の構築を進めている。本論文では、特に社会技術としてのエージェント技術に焦点を当て、顔や体を使ってジェスチャー等の非言語情報を利用しながら、適切に知識や情報を伝達するアニメーションキャラクター(以降、エージェントと呼ぶ)について提案する。

中野¹⁾は知識流通に寄与するエージェントの機能を分類しており、その中で、コミュニケーションを円滑化させるために、エージェントによる非言語情報の表出が重要であることを述べている。一般的には難しいと感じられる専門的な話をエージェントを使って説明する場合には、重要な概念や話のポイントで適切にジェスチャー等を使って強調することにより、わかりやすさを向上させ、より自然性の高いコミュニケーションが実現すると期待される。

そこで、本論文では、実際のプレゼンテーションデータを収集し、どのような場合にジェスチャーが用いられるのかを詳しく分析する。特に、ここでは発話中の言語的特徴とジェスチャー使用との相関関係に着目し、ジェスチャー決定モデルを構築する。さらに、このモデルをエージェントによるプレゼンテーション自動生成システムに統合する。

本論文の構成は以下の通りである。次章では、ジェスチャーに関する心理学、コミュニケーション科学、そして言語学の研究を概観する。3章では、プレゼンテーションのデータを収集し、ジェスチャーの出現とそこでの言語的特徴との関係について分析する。さらに4章では、この分析結果に基づき、テキストからエージェントによるプレゼンテーションを自動生成するシステムを構築する。さらに、本システムの社会技術としての有効性を議論する。最後に将来課題について議論する。

2. 関連研究

2.1. ジェスチャーに関する心理学的研究

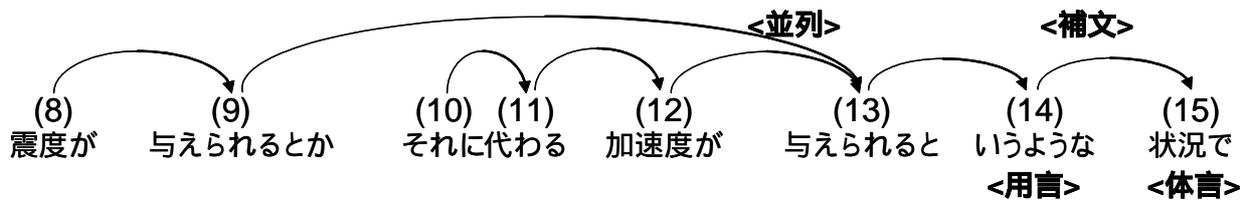


Fig. 1 : 文節間の係り受け関係

人間のコミュニケーション行動に関する従来研究では、様々な非言語情報の中でも特にジェスチャーは音声言語の理解を高める効果を持つと報告されている。音声でジェスチャーを伴って与えられた場合には、理解の正確性は、音声のみの場合と比べて約2倍に向上し²⁾、さらに、口の動きや顔の表情よりもジェスチャーを伴った刺激のほうがよりよく理解されることが実験により実証されている³⁾。

2.2. 言語学とジェスチャー研究

言語表現とジェスチャーの出現との関連について、従来研究において、以下のような知見が得られている。

参照表現の言語的分量: McNeill⁴⁾ は、Firbas⁵⁾ により提案された「伝達のダイナミズム」(Communicative Dynamism (CD)) - 所与のメッセージがコミュニケーションを前進させる度合い - という概念を変数として用い、CD が大きくなるほどジェスチャーが出現しやすいと述べている。さらに McNeill は、CD の尺度として「参照表現を構成する言語的情報の量」を取り上げている。例えば、代名詞は通常の名詞よりも CD が低く、名詞は修飾句/節を伴う名詞よりも CD が低い。この McNeill の議論は、文中の修飾関係を調べることにより、CD を、さらにはジェスチャーの出現位置を推定できることを示唆している。

主題・題述: さらに、McNeill は、文の主題は通常最も CD が低く、一般的にはジェスチャーを伴わない、ジェスチャーは通常題述(主題について述べている部分)に付随すると論じている。ここで注目すべき点は、英語の場合は、主題・題述の認定が語順だけからでは困難であるが、日本語の場合には、提題助詞(eg. 「は」等)と呼ばれる主題を有標化する言語的手段が備わっているということである。従って、日本語に特化すると、提題助詞で有標化されていれば、統語解析によって文の主題部分を同定し、ジェスチャーが出現しにくい場所として識別することができるはずである。

さらに、WH 疑問文においては、文頭の疑問詞は文内容の方向を示す標識であり、疑問詞で尋ねているものが焦点であると考えられるが、疑問詞を主題とするのか題述

とするのかは議論が分かれている^{6, 21)}。従って、文の情報構造において特殊な位置づけにある疑問詞というものがジェスチャーを誘発する要因になるのかどうか、データ分析により明らかにする必要がある。

新・旧情報: 主題・題述による文の情報構造と深くかかわる概念として新・旧情報がある。通常、旧情報は題述性が低く、新情報は高い。このことを利用すると、初めて言及された表現であるか否かを調べることによって、表層表現から題述性が推定でき、さらには、題述性の低い旧情報よりも、高い新情報の方がジェスチャーを伴いやすいと予想される。

対比的関係: 談話中の重要な概念間の対比関係を明確化するためにアクセントが用いられるが⁷⁾、Cassell & Prevost⁸⁾ は、この研究⁷⁾ と Kendon の韻律とジェスチャーについての理論⁹⁾ に基づき、意味表現からの対比的ジェスチャーの生成を行っている。このことを統語的に捉えてみると、同様の関係を統語的な並列構造を見つけることにより同定できる可能性がある。

統語分析として係り受け分析を用い、実際の発話を分析した例を Fig. 1 に示す。文節(8)~(9)と(10)~(13)は並列関係にある。また、文節(14)は補文であり、文節(15)の体言に係り、修飾している。つまり、文節(15)は節によって修飾された言語的分量の大きい体言である。

2.3. 計算機科学におけるジェスチャー生成の問題

2.1 節では、人間同士のコミュニケーションにおいて、ジェスチャーが重要な役割を担っていることを示したが、ここで、人対コンピュータのコミュニケーションについて考えてみると、Reeves and Nass¹⁰⁾ が主張するように、人がエージェントに対して、人に対してと同様の反応をすると仮定すれば、音声言語と同期して適切なジェスチャーを自動的に生成することは、より効果的なエージェントを実現するための重要な研究課題となる。

ジェスチャーの自動生成に取り組んだマルチモーダル生成に関する研究では、ジェスチャーは説明システムにおける教示内容^{11, 12)}、仮想学習環境におけるタスク状

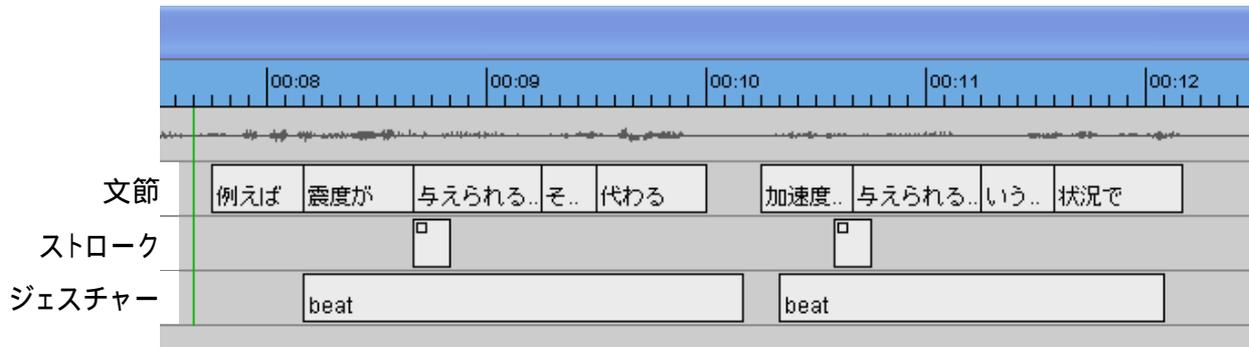


Fig. 2 : Anvil によるジェスチャーアノテーション例

況¹³⁾ ,あるいは会話エージェントにおける発話意図¹⁴⁾ に応じて決定される。しかし、これらのシステムでは、コンテンツを追加するために、コンテンツ作成者(例えば、学習教材を用意する教師など)が論理的な意味表現を記述しなければならず、大きな障害となる。一方、Cassellら¹⁵⁾は、テキストを入力すると、それに対するエージェントの動作を自動的に決定し、合成音声と同期したエージェントの動作スケジュールを出力するツールを提案している。しかし、従来研究では、入力されたテキスト中のどのような情報を使ってジェスチャーを決定すべきか、表層的な言語情報からジェスチャーをどの程度適切に決定することができるのかについての議論は少なく、また関連した言語学理論は計算モデルとして利用できるように形式になっていない。

以上の議論に基づき、本研究では以下の課題に取り組む。

- ・言語表現を解析することによって得られる語彙的/統語的情報はテキストにジェスチャーを付与する上で、有効であるのか。
- ・もしそうであるならば、テキストからどのような情報を抽出し、それをエージェントの動作決定機構の中でどのように利用するのか。

まず、次章では、実際のプレゼンテーションデータを収集し、その分析結果について報告する。

3. 分析

どのような言語的特性がジェスチャーの出現に関係しているのかを調べるために、実際のプレゼンテーションデータを収集し、その分析を行う。

3.1. データ収集

社会技術に関連した研究発表会約7時間分のプレゼンテーションと質疑応答を録画した。録画にはデジタルビデオカメラを用い、明瞭な音声を録音するために、マイク入力を直接ビデオの音声チャンネルに入力した。また、

ビデオから書き起こしを作成した。

収集したデータの中から7人の発表者につき、約15分の発表時間のうち中間部分約3分(計約21分)のデータを分析した。分析対象となったデータは合計2124文節、その中に343のジェスチャーが含まれていた。

ジェスチャーのアノテーションには行動分析用ツールAnvil[16]を用いた。Fig.2に分析画面の一部を示す。最上段に音声波形、次に各文節の発生時間、3段目にジェスチャーストローク発生時間、そして、最下段にジェスチャー開始から終了までの時間が示されている。このようなツールを用いることにより、複数のコミュニケーションモダリティー(例えば、言語とジェスチャー)が同期している様子を視覚的に捉えることができる。この例では、「震度が与えられる」と「加速度が与えられる」という並列関係を構成する部分において、ジェスチャーが発生していることがわかる。

3.2. データ分析

(1) ジェスチャーの分析

ジェスチャーは準備期、ストローク、消失期とから構成され、ストロークはもっとも目立った音節と共起することが多い⁹⁾。そこで、ジェスチャーの準備期の開始から消失期の終了までを1ジェスチャーと認定し、その継続時間に加え、ストロークを打った時間もアノテーションに加えた。約半数のデータに関しては、3人の分析者が議論してジェスチャーの出現位置を認定した。ジェスチャーの認定方法について一致した基準が得られた後、一人の分析者が残りのデータを分析した。

(2) 言語情報の分析

日本語係り受け解析器¹⁷⁾を用いて、以下の項目に関して各文節の特徴を計算した。

- (a) 所与の体言が節や補文により修飾されているか (参照表現の言語的分量)

Table 1 : 分析結果

条件		ジェスチャー出現率	
[C1]	参照表現の 言語的分量	(a) 節により修飾された体言	0.382
[C2]		代名詞, その他の体言	(b) フ格 & (d) 新情報
[C3]	(c) WH疑問詞		0.414
[C4]	(e) 並列句		0.477
[C5]	強調の副詞	(f) 強調の副詞	0.244
[C6]		(f') 強調の副詞に後続	0.350
[C7]	(g) Cue word		0.415
[C8]	(h) 数詞		0.393
[C9]	その他(ベースライン)		0.101

- (b) 体言に後続している助詞の種類(「は」、「が」、「を」、その他)
 - (c) 所与の文節が疑問詞を含むか
 - (d) 所与の文節中の全ての内容語が1つ前の文中で既に述べられているか(新/旧情報)
 - (e) 所与の文節が並列句を構成しているか
- 以上の統語的な要因に加えて、以下の語彙的な特徴に関しても分析を行った。
- (f) 所与の文節が強調の副詞 (eg. 大変, 非常に) を含むか、あるいは強調の副詞を含む文節の直後の文節であるか
 - (g) 所与の文節が cue word (eg., それでは, 従って) を含むか、あるいは cue word を含む文節の直後の文節であるか
 - (h) 所与の文節が数詞 (eg., 数千人, 99 回) を含むか

Caselli⁽²⁾ は、ジェスチャーは話の区切りを示すと述べており、これは、談話の区切りを示す cue word がジェスチャーと共に起しやすいうことを示唆している。

以上の分析方法を用い、ジェスチャーの出現位置と各文節の言語的特性との相関関係を調べた。

3.3. 分析結果

分析結果を Table 1 に示す。ジェスチャー出現のベースラインは 10.1% である(つまり、約 10 文節に 1 回の頻度でジェスチャーが出現する)。体言が節によって修飾されている場合、ジェスチャーの出現確率は 38.2% である。代名詞やその他の体言が格助詞の「を」を伴って文節を構成し、かつその文節に新情報(前の文で言及されていない内容語)が含まれている場合には、28.1% の場合においてジェスチャーが共起している。

さらに、並列構造の一部となる文節であれば、ジェス

チャーが起こる確率は 47.7% であり、疑問詞を含む文節である場合には 41.4%、cue word を含む文節である場合には 41.5% となる。また、強調の副詞に続く文節である場合や数詞を含む文節においてもジェスチャーの出現確率が比較的高い(出現確率はそれぞれ、35%、39.3%)。

以上に示すように、Table 1 に列挙された言語的特徴を持つ文節では、それ以外の文節に比べて、ジェスチャーの出現確率が約 3~5 倍となる。また、これらの言語的特徴によって、観測されたジェスチャーの約 75% を説明することができる。これらの分析結果から、ジェスチャーを使うべき場所と使わないべき場所とを識別する上で、語彙的・統語的情報が有用であることが明らかになった。

4. システムの実装

4.1. CAST

本節では、会話エージェントシステム CAST(The Conversational Agent System for neTwork applications)について述べる。CAST のシステム構成を Fig. 3 に示す。

CAST はテキストを入力とし、エージェントのアニメーションスケジュールとエージェントの発話となる合成音声を手動的に計算し、出力する。Fig. 3 に示すように、CAST は (1) エージェント動作決定機構 (Agent Behavior Selection Module (ABS)), (2) 言語タグ付与機構 (Language Tagging Module (LTM)), (3) 音声合成装置 (TTS), (4) エージェントアニメーションシステム (RISA) の 4 つの主要モジュールからなる。CAST に入力されたテキストは、まず ABS に送られる。ABS がそのテキストを LTM に送ると、言語情報がタグ付けされ、ABS に返される。ABS は、この言語情報を用いて、どの文節でどんな動作を用いるかを決定する。そして、表情やジェスチャー等のエージェントの動作タグをテキストに付与する。最後に、これがタイムスケジュールに変換されてエージェントアニメーションシステムで実行される。

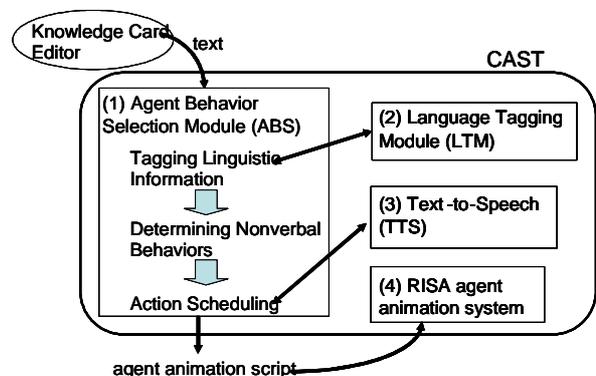


Fig. 3 : CAST システム構成図

(1) エージェント動作の決定

(1-1) 言語情報タグの付与

まず ABS では、LTM が入力テキストを解析し、2.2 節で議論された言語情報を自動的にテキストに付与する。例えば、Fig. 1 の文節(9)に以下のような言語情報が付与されたとする。

```
{テキスト ID:1, 文 ID:1, 文節 ID:9, 係り受け_from:8,
  係り受け_to:13, 文節タイプ:用言, 言語的分量: NA,
  格: NA, WH疑問: false, 新/旧情報: 新, 並列関係: 13,
  強調副詞: false, Cue-Word: false, 数詞: false}
```

例えば上の例では、この文節が含まれるテキストの ID が 1、文の ID が 1 であり、この文節の ID は 9 である。文節 8 がこの文節に係っており、この文節は文節 13 に係る。この文節は新情報を伝達し、文節 13 と並列関係にある。

(1-2) ジェスチャーの決定

次に ABS は、各文節に対し、ジェスチャーを付与すべきかどうかを Table 1 に示すデータの分析結果に基づいて決定する。例えば、先に示した例では、文節(9)は Table 1 の[C4] (文節が並列構造の構成要素である)の場合に当てはまるが、この場合には、システムは 47.7%の確率でジェスチャーを該当文節に付与する¹。

現在のシステムでは、ジェスチャーの形態のデフォルトとしてビートジェスチャーを採用している。ビートジェスチャーとは、手を上下に振るような身振りであり、発話の意味内容とは直接的に関連せず、発話の中で強調される部分に出現しやすい。一方、強調される文節中の概念に対して、特定のジェスチャーがエージェントアニメーションシステムのライブラリに定義されている場合

```
<Gaze type="towards">
  震度が
  <Gesture_right type="contrast" handshape_right="stroke1 @2">
    与えられるとか
  </Gesture_right>
  それに
  代わる
  加速度が
  <Gesture_right type="contrast" handshape_right="stroke2 @2">
    与えられると
  </Gesture_right>
  いろいろな
  <Gesture_right type="best" handshape_right="stroke1">
    状況で
  </Gesture_right>
  ...
```

Fig. 4 : エージェント動作スクリプト例

¹ 複数のルールに適合する場合も発生するが、現段階では、もっともジェスチャー出現確率の高いルールを適用している。

(例えば、「大きい」という概念を表現するジェスチャーがライブラリに既に登録されている場合には、それが優先して用いられる。

ジェスチャーが決定されると、エージェント動作タグが XML 形式で付与される。ABS ではジェスチャーに加えて、表情や視線などの動作も決定される。¹⁸⁾ Fig. 4 にエージェント動作の XML の例を示す。この XML では、2 番目と 6 番目の文節で対比のジェスチャーを行い、8 番目の文節でビートジェスチャーを行うことを指示している。

(2) スケジューリング

最後に、動作スクリプトの XML はタイムスケジュールに変換される。XML のテキスト部分を音声合成装置に入力することにより、ABS は音素や文節区切りの時間情報を合成エンジンから取得し、これをもとにエージェント動作のタイムスケジュールを計算すると同時に、読み上げの合成音声をファイルに保存する。ABS から出力されたタイムスケジュールは、アニメーションのコマンド列としてエージェントアニメーションシステムで解釈され、スケジュール通りに実行される

4.2. 社会技術への適用

(1) 適用例

前節で述べた CAST は、IPOC¹⁹⁾ における、エージェントコンポーネントとして統合してゆく予定である。IPOC は、社会技術への会話エージェント技術の適用を目的として開発されたストーリー型コミュニケーション環境であり、利用者は会話エージェントとのインタラクションを通して、体験的に安心・安全に関する知識を得ることができる。例えば、災害時等の状況の画像、映像、エージェントアニメーションを交えたコンテンツを IPOC 上で表現することにより、利用者は災害時への備えや教訓を知ることができる。

Fig. 5 に IPOC の概要を示す。まず、コンテンツ作成フェーズでは、コンテンツ作成者は、背景中の対象物に関連する説明や体験談を知識カード化し¹⁹⁾、それを対象物に付箋紙を貼り付けるような感覚で、背景に埋め込んでいく。最後に、知識カードが保存される際に、ストーリーテキストが、CAST に送られる。CAST では、エージェントのジェスチャー等の非言語情報を決定され、音声合成エンジンにより読み上げの音声とエージェントの動作スケジュールが作成、出力される。

コンテンツ体験フェーズでは、ユーザがエージェントに「この壊れた家について話して。」、「これ(オブジェクトのクリックにより指示)について話して。」等の働きかけを行うと、カードに記述されたストーリーがエージェントにより読み上げられる。このとき、コンテンツ作成フェーズで計算された動作スケジュールが音声と同期し

て実行される。

(2) 有効性と検討課題について

エージェントアニメーションを用いた教育・啓蒙コンテンツはテレビやWeb ページで一般的になりつつあるが、IPOC を利用することにより、表現力豊かなマルチメディアコンテンツを一般の人々が容易に視聴・作成することが可能になる。また、コンテンツ作成のコストも大きく削減することができる。このような特徴を持つ IPOC を用いると、専門的な知識を持たない人や、関心が高くない人にもアニメーションを見る感覚で、低い負荷で学習してもらえらる。また、そのようなコンテンツを専門家が手軽に作成・配布する環境が実現する。このようなシステムの重要な構成要素となる CAST は、社会における知識共有基盤構築のための要素技術として利用できると考える。

一方、会話 エージェントを社会技術に適用する際に、十分検討が必要な点もある。エージェントの社会的な役割や、エージェントが発信する情報に対して誰が責任を持つのかといった点を明確にしておかないと、情報の信頼性にかかわる問題となる。また、専門家による適切なフォローアップも専門家と一般市民とのコミュニケーションを継続・発展させていくためには不可欠である。

5. まとめと今後の課題

本稿では、テキストにジェスチャーを自動的に付与し、ナレーション音声と同期したエージェントアニメーションを出力するメディア変換技術を提案した。そのために、まず、実際のプレゼンテーションのデータを分析し、テキストにジェスチャーを付与するために有効な語彙的/統語的情報を統計的に明らかにした。特に、文節が並列構造を構成素する場合には、約半数の場合においてジェスチャーが付随し、また、節によって修飾されている場合にもジェスチャーの出現頻度が高くなることがわかった。これらの結果は、助詞のタイプや新/旧情報といった局所的な情報を用いて推定された主題・題述性よりも、係り受け関係から得られる文全体の構造的情報のほうがジェスチャー決定にはより有用であることを示唆している。次に、エージェントシステム CAST を実装し、さらに CAST を統合したコミュニケーション環境 IPOC について述べた。

人間のジェスチャー使用をモデル化した本論文での提案手法は、エージェントに人間のコミュニケーションプロトコルを模倣させることを目的としたものであったが、それが実際の人対エージェントのコミュニケーションに効果的であるのかを評価実験により調べる必要がある。例えば、提案モデルに基づいてジェスチャーを生成するエージェントとランダムにジェスチャーを行うエージェ



Fig. 5 : IPOC の概要

ントとを比較し、どちらとコミュニケーションした場合により情報の伝達が促進されたかを比較することは可能であろう。

また、本稿で提案したモデルにより、データの全てを説明できるわけではない。残りはさらにモデルを精緻化していくことにより、説明されなければならない。そのためには、意味や語用論の情報を統合したモデルへと改良していくことが今後の課題である。Hovy²⁰⁾ は、文章の構造や使用する単語を変えることにより、同じ事象についても、読み手は、肯定的、否定的、あるいは中立的等の異なる印象を持つことに着目し、意味内容は同じでも語用論的な効果が異なる文章を自動生成するシステムを構築している。この研究では、語用論的な情報を言語表現の選択に利用しているが、本論文で提案した手法を拡張することにより、語用論的な情報を非言語表現の決定や選択に利用することができる。これにより、同じテキストが読み上げられた場合でも、エージェントによる非言語行為によって強調される点が変化し、それを見る人々はそのストーリーに対して持つ印象が変化すると考えられる。さらに、社会的役割や立場といった要因を非言語情報選択において考慮することができれば、異なる立場のエージェントを実現することができるであろう。

最後に、IPOC は背景とエージェントとを統合することにより、よりリアリティの高いコミュニケーション環境を提供しているが、このようなシステムにおいては、視聴者が観察できる世界とストーリーの内容がうまく適合していることが重要である。例えば、今話題になっている対象物にズームアップする、あるいはアングルを変えることにより、ストーリー中に明示的には示されていない視点を明確化することに役立つであろう。

参考文献

- 1) 中野有紀子, 知識流通のためのメディア技術 インタフェースエージェントの利用 . 社会技術論文集, 2003.
- 2) Berger, K.W. and G.R. Popelka, Extra-facial Gestures in Relation to Speech-reading. *Journal of Communication Disorders*, 1971. 3: pp. 302-308.
- 3) Rogers, W., The Contribution of Kinesic Illustrators towards the Comprehension of Verbal Behavior within Utterances. *Human Communication Research*, 1978. 5: pp. 54-62.
- 4) McNeill, D., *Hand and Mind: What Gestures Reveal about Thought*. 1992, Chicago, IL/London, UK: The University of Chicago Press.
- 5) Firbas, J., On the Concept of Communicative Dynamism in the Theory of Functional Sentence Perspective. *Philologica Pragensia*, 1971. 8: pp. 135-144.
- 6) Halliday, M.A.K., The linguistic study of literary texts, in *Essays on the Language of Literature*, S. Chatman and S.R. Levin, Editors. 1967, Houghton Mifflin: Boston.
- 7) Prevost, S.A. An Informational Structural Approach to Spoken Language Generation. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics*, 1996, (Santa Cruz, CA), pp.
- 8) Cassell, J. and S. Prevost. Distribution of Semantic Features Across Speech and Gesture by Humans and Computers. In *Proceedings of Workshop on the Integration of Gesture in Language and Speech*, 1996, (Newark, DE), WIGLS, pp. 253-270.
- 9) Kendon, A., Some Relationships between Body Motion and Speech, in *Studies in Dyadic Communication*, A.W. Siegman and B. Pope, Editors. 1972, Pergamon Press: Elmsford, NY. pp. 177-210.
- 10) Reeves, B. and C. Nass, *The Media Equation: how people treat computers, televisions and new media like real people and places*. 1996, Cambridge: Cambridge University Press.
- 11) Andre, E., T. Rist, and J. Muller, Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence*, 1999. 13: pp. 415-448.
- 12) Rickel, J. and W.L. Johnson, Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition and Motor Control. *Applied Artificial Intelligence*, 1999. 13(4-5): pp. 343-382.
- 13) Lester, J.C., B. Stone, and G. Stelling, Lifelike Pedagogical agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. *User Modeling and User-Adapted Interaction*, 1999. 9(1-2): pp. 1-44.
- 14) Cassell, J., M. Stone, and H. Yan. Coordination and Context-Dependence in the Generation of Embodied Conversation. In *Proceedings of INLG 2000*, 2000, (Mitzpe Ramon, Israel), Association of Computational Linguistics, pp. 171-178.
- 15) Cassell, J., H. Vilhjalmsson, and T. Bickmore. BEAT: The Behavior Expression Animation Toolkit. In *Proceedings of SIGGRAPH 01*, 2001, (Los Angeles, CA), ACM Computer Graphics Press, pp. 477-486.
- 16) Kipp, M. Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology*, 2001, pp. 1367--1370.
- 17) Kurohashi, S. and M. Nagao, A Syntactic Analysis Method of Long Japanese Sentences Based on the Detection of Conjunctive Structures. *Computational Linguistics*, 1994. 20(4): pp. 507-534.
- 18) Nakano, Y.I., T. Murayama, and T. Nishida, Multimodal Story-based Communication: Integrating a Movie and a Conversational Agent. *IEICE Transactions on Information and Systems*, 2004. E87-D, No.6: pp. 1338-1346.
- 19) 村山敏泰, 中野有紀子, 西田豊明, 没入型会話環境を活用したストーリー交換による情報共有システムの提案. 社会技術論文集, 投稿中.
- 20) Hovy, E.H., *Pragmatics and Natural Language Generation*. *Artificial Intelligence*, 1990. 43(2): pp. 153-198.
- 21) 野田尚史. 『「は」と「が」』, 新日本語文法選書 1, 東京: くるしお出版
- 22) Cassell, J., *The Development of the Expression of Time and Event in Narrative*, in *Department of Psychology/Linguistics*. 1991, The University of Chicago: Chicago, IL. p. 233.

謝辞

本研究は、社会技術研究システム ミッション・プログラム 「安全性に係わる社会問題解決のための知識体系の構築 (平成13~14年度は日本原子力研究所の事業, 平成15年度からは科学技術振興機構の事業) の研究として行われた。

PROVIDING INFORMATION THROUGH CONVERSATIONAL AGENTS: EMPHASIZING IMPORTANT CONCEPTS USING NONVERBAL INFORMATION

Yukiko NAKANO¹, Toshihiro MURAYAMA², and Toyooki NISHIDA³

¹MS. (Media Arts and Sciences) Research Institute of Science and Technology for Society (RISTEX)
(E-mail: yukiko @ristex.ac.jp)

² Research Institute of Science and Technology for Society (RISTEX) (E-mail: tmuraya @s-tech.co.jp)

³Ph.D. (Engineering), Professor, Graduate School of Informatics, Kyoto University (E-mail: nishida@i.kyoto-u.ac.jp)

Establishing shared knowledge is indispensable to construct a safe and secure society. With the goal of providing information technologies helping knowledge sharing in a society, this paper proposes a conversational agent that can produce a presentation with natural nonverbal expressions, such as gestures and facial expressions. We also describe an information providing system into which the conversational agent is integrated. First, analyzing human presentation video-data, we establish a model of human gesture usage. Then, we implement the model into a conversational agent who works as a presenter. Finally, we will discuss the effectiveness of such systems in terms of information sharing in a society.

Key Words: *Conversational agent, Automatic gesture generation, Multimedia communication environment, Communication science, Empirical Study*