

ヒューマンコミュニケーションの「場」が読めるロボットの研究開発 (041307003)

The Robot's Ears Project

研究代表者

ニック・キャンベル

国際電気通信基礎技術研究所 音声言語コミュニケーション研究所 音声音響処理室学研究室

Nick CAMPBELL

ATR Spoken Language Communication Research Laboratories

研究分担者

定延 利之[†] 千原 國宏^{††} 井村 誠孝^{††}

Sadanobu Toshiyuki, Chihara Kunihiro, Imura Masataka,

ダミアン ドウシャン^{††} 岩橋 直人^{†††} 中川 明子^{†††}

Damien Douchamps, Iwahashi Naoto, Nakagawa Akiko

[†]神戸大学 国際文化学部 国際文化学科 情報コミュニケーション論講座

^{††}奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 像情報処理学講座

^{†††}国際電気通信基礎技術研究所 音声言語コミュニケーション研究所 音声音響処理室学研究室

研究期間 平成 16 年度～平成 18 年度

本研究開発の概要

本研究プロジェクトは、人間同士のインタラクションをセンシングし、会話の「場」の流れを読む基礎技術の研究開発である。本プロジェクトで提案する技術は、二人以上の参加者から構成される会話を対象として、参加者のふるまいの画像・音声情報から、非言語的動作プリミティブを抽出し、創発的手法に基づいて場の流れを認識する技術である。会話の場の流れは、主に体の向きや視線方向・身振り手振り・声のトーン・あいづち・うなずきの挿入といった非言語的手段によって表現されている。この技術が実現されることにより、人間同士のインタラクションを言葉に表されるレベルより深い関係性ニュアンスのレベルにおいて理解し、人間と機械との自然なインタラクションを実現するための対話処理技術が可能となる。

Abstract

This research project investigates tools and methods for the monitoring of discourse flow in a meetings environment. The technology we propose collects and extracts non-verbal information from video and audio data captured by observation of participants in a round-table discussion of between two and a dozen participants. This information is then used to estimate the discourse participation status of each of the members present for recording the flow of activity throughout the meeting and for discriminating among the different types of interaction between the members.

1. まえがき

非言語情報プリミティブを抽出・分析処理することによって、人間の活動をモデル化する試みは、既にいくつかなされている。顔表情による感情の表出については、ATRをはじめとするいくつかの研究機関において感情表情の認識および合成の研究が進められている。また、研究代表者は科学技術振興機構 CREST 発話様式プロジェクト代表者として音声合成の観点から感性音声の研究を実施してきた。一方、米国 MIT の Affective Computing グループでは、生体信号を検出するセンサーやデバイスを用いて、人間の感情的状態を認識・モデル化する研究を進めている。

本研究プロジェクトでは、ヒューマンコミュニケーションの意図を理解する基盤技術によって、次世代ヒューマンインターフェース技術を提案する。ここでのコミュニケーションの場は、発話内容よりも発話情報が表現する参加者の意図のセンシングに着目する観点に立つ。

2. 研究内容及び成果

会話が行われる場面では、人々は、参加者は誰か、誰がいつ発話をするか、提示された情報を参加者がどのように受け止めたか、会話中の話題がどのように展開してい



図1 キャプチャーシステムの1例、全方位カメラにマイクアレイによるデータ収集装置

くかなど、言語的には明示されないさまざまな情報をお互いに交換しながら会話を進行させている。「ヒューマンコミュニケーションの場を読む」ことは、このようなコミュ



図2 キャプチャーシステム出力の1例、全方位カメラによって人を認識し、トラッキングする

ニケーションの流れを制御するために非言語情報を活用する状況を対象とする点でセンシング技術の第一歩である。

言語内容レベルより深い関係性ニュアンスのレベルにおいて理解するための非言語情報認識システムは、以下の5段階で研究を行った。

- (1) キャプチャーシステムの設計・開発
- (2) 自然対話のデータベース設計・収集
- (3) 対話音声データの分析・モデル化
- (4) ノンバーバル情報の認識アルゴリズム
- (5) システム統合化・評価

図1は、キャプチャーシステム、図2はキャプチャーシステムの出力画面を示す。ここでは9人の認識、顔の抽出によって、動作情報を計算する。図3は参加程度を示し、それぞれの対話アクティビティと発話状況によって、高次の情報を元にして対話コミュニケーションの流れのトラッキングを表示する。

(1) キャプチャーシステムの設計・開発

創発的な人物認識システムの一部である、ロバストかつ自動的に自然対話データ収集を可能とする。自然性は保たれ、参加者はカメラを意識せず、対話流れのすべてをモニターする技術である。

(2) 自然対話のデータベース設計・収集

「場」は、日常的人間の気持ちのやり取りを含む、友人間の談話から、ビジネスミーティングや、多人数ソーシャルイベントまでの3種の層を設定し、それぞれの対話コーパスを作成した。これは、基礎研究のための重要な知的財産となる。

(3) 対話音声データの分析・統計モデル化

異なる対話環境の対話構造や表現形式などの音声と動作情報の処理アルゴリズムを検討し、統計モデルの学習を行った。ノンバーバルな音声情報、つまり「短発話の発話様式による情報」などを処理することである。その短発話の辞書を作成(約250種類)し、それぞれの短発話の種類ごとの韻律分析情報から、言い方(発話様式)による話者意図や情動情報を判断する学習の改良も行った。

(4) ノンバーバル情報の認識アルゴリズム

非言語情報の特徴量の変化と参加者の意図判断のため、自動データから人手のデータへの統計的対応づけ・音声情報と映像情報の統合・相互行為や参与のタイプの分類・非言語行動から意図への対応づけのモデル化・現存する対話行動モデルへの対応づけ等の研究から対話インタラクション処理アルゴリズムを開発した。

(5) システム統合化・評価

高次の情報を数多く取得することが出来る統合システムは、会話参加者の意図・態度あるいは感情的状態の情報を人間の約6割程度認識可能とした。

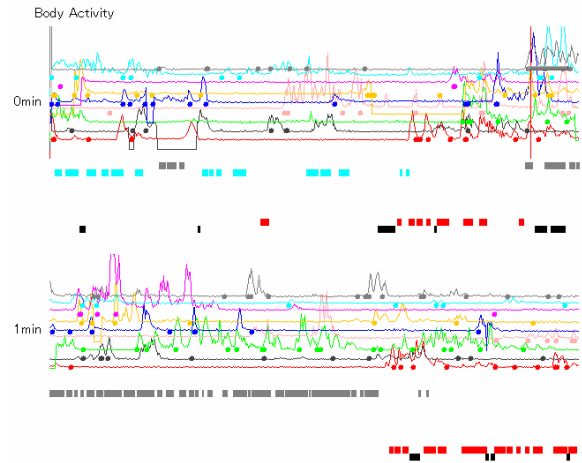


図3 認識された9人毎の動作・音声トラッキング結果

3. むすび

ユビキタスコンピューティング世代の基本技術として、今後、ヒューマンセンシングが最も重要な技術となる。本研究は、ノンバーバル情報処理の研究を基礎とするヨーロッパEU大規模共同研究プロジェクト「Cost 2102」の立ち上げの先導となった。日本側(本研究代表者)はプロジェクトのアドバイザーである。「場」が読める研究プロジェクトで提案した技術である、動作センサー技術と音響センサー技術の複合システムによって、会話への参加人数の判断、発話者の特定、聞き手の情動状態の認識レベルに始まり、家電や車内技術の利用も充分期待できる。テレビが部屋の人数とそれぞれの動作関係が判ったら、自分(テレビ)が差動すべきかどうかを判断できたり、車が運転者や同乗者の状態をセンシングすることによって反応を変えて安全走行できたりする。

【誌上発表リスト】

- [1] Campbell, N., "Working with very sparse data to detect speaker and listener participation in a meetings corpus", in Proc LREC w/s Multimodal Corpora: from Multimodal Behaviour Theory to Usable Models, Genova, Italy, May 2006.
- [2] 定延利之, "話し手は言語で感情・評価・態度を表して目的を達するか?—日常の音声コミュニケーションから見えてくること—", 自然言語処理 Vol.14 No.3 (2007年4月)
- [3] Mayumi Bono and Yasuhiro Katagiri, "An analysis of non-linguistic and paralinguistic cues for turn management in presentation talk", Peter Slezak (ed) Perspectives on Cognitive Science. Elsevier series.

【申請特許リスト】

- [1] Nick Campbell, "A method for detecting listener reactions in conversational speech", PCT

【登録特許リスト】

- [1] Nick Campbell, "行動指針決定装置及びコンピュータプログラム", 日本, (2005/8/25)
- [2] Nick Campbell, "行動指針決定装置", PCT, (2005/10/18)

【報道発表リスト】

"日認知メディア情報学研究室本語の「音」に挑む", 日経新聞(朝刊), 2005/9/4

【本研究開発課題を掲載したホームページ】

<http://feast.atr.jp/non-verbal/>