

身体化された音声インタラクション

Embodied Vocal Interaction

小野 哲雄^{*1*2} 鎌島 正幸^{*2*3} 本田 喜久^{*2*3}

Tetsuo Ono Masayuki Kamashima Yoshihisa Honda

神田 崇行^{*2} 今井 倫太^{*2*3} 石黒 浩^{*2*4}

Takayuki Kanda Michita Imai Hiroshi Ishiguro

^{*1} 公立はこだて未来大学 ^{*2} ATR 知能ロボティクス研究所 ^{*3} 慶應大学 ^{*4} 大阪大学
Future University-Hakodate ATR Intelligent Robotics and Communication Lab. Keio University Osaka University

In this paper we propose a model of embodied vocal interaction. This model aims at clarifying the mechanism of interaction emerged by body and cognitive entrainments. To explore the validity of our model, we carried out experiments consisted of four conditions: interactions through a headset, a microphone-speaker, face-to-face, and a robot. We show the results and discuss the difference between mobile and ubiquitous computing, and the development of entertainment systems based on our proposed model.

1. はじめに

携帯電話で話をしていて「お辞儀」をする人を街中でよく見かける。しかし、インタホンで話をしていて「お辞儀」をする人はあまり見かけない。このような身体動作はどういう意味があり、コミュニケーションにどのような影響を与えているのだろうか。また、使うメディアによって本当に違いはあるのだろうか。

本稿では、上記のような身体動作をコミュニケーションにおいて本質的なものにとらえ、「身体化された音声インタラクション」のモデルを提案する。さらに、いくつかのメディアを用いた人-人のインタラクション実験、および、人-ロボットの実験をとおして、本モデルの妥当性の検証を行う。

2. 身体化された音声インタラクションのモデル

我々のこれまでの研究[小野 01, Ono 01]から明らかになったことは、空間情報を円滑に伝達するためには、対話者間で「システムを創る」ことが重要であるということである。この「システムを創る」際には、対話者同士があたかもダンスを踊るかのよう「身体」を同調的に動作させることが重要な役割を果たすことも確認された。つまり、図 1 に示すように、身体的な引き込み (body entrainment) が対話者間に関係 (relation) を構築し、この関係に基づき共有視点 (joint viewpoint) を形成する。この共有視点が円滑な情報伝達を可能にしていると考えられる。このようなメカニズムを我々は「認知的引き込み」(cognitive entrainment) とよぶ。この認知的な引き込みは、発話権の交代や談話における文脈の形成など、一般的なコミュニケーションにも影響を与えていると考えられる。

上記の「身体同調的動作」の「身体」は物理的な身体である必要はないだろう。つまり、自分の身体と「一体感」が持て、「境界を創らない」ものであってもよいだろう。本稿ではそれらを総称して「身体化されるメディア」(embodied media) とよぶ。このようなメディアをとおして音声インタラクションを行うとき、人はあたかも対面しているかのような身体動作を行う可能性がある (図

1 参照)。逆にいえば、このような身体動作を引き出すメディアを構築することが、円滑なコミュニケーションを実現する手段になる。具体的にいうと、携帯電話やヘッドセットは対話者に一体感を与え、境界を創らずにインタラクションをすることができるのではないだろうか。また、人の身体動作と同調的な動作をするロボットも同様であろう。しかし、「身体化されないメディア」(unembodied media) と考えられるインタホン (外部に固定されたマイクとスピーカ) は、上記のような身体動作を引き出さないことが予想される。

本稿で提案するモデルは、認知的なメカニズムを説明するだけでなく、モバイルコンピューティングとユビキタスコンピューティングの本質的な違いを明らかにし、さらにロボットシステムの設計方法やエンタテインメントコンピューティングの研究にも有益な示唆を与える。つまり、人とインタラクションする物全般に適用可能な考え方である。

3. インタラクション実験

本稿で紹介する実験の概要について述べる。本実験では、順路案内の場面における「空間情報の伝達」と「身体動作」の関係について調べ、前節で提案したモデルの検証を試みる。具体的には、あるビル内の目的地までの順路を被験者 A のみに教え、その情報を別の被験者 B に伝達するというタスクを与える (道案内タスク)。A、B ともにこのビルを訪れたことがないことを前提とする。この道案内のインタラクションにおける身体動作をモーションキャプチャシステム Vicon で計測し、情報の伝達と身体動作の関係について調べる。本実験では、以下の 4 条件を設定し、各条件 5 組ずつ実験を行った。

- ・**ヘッドセット条件**: A、B にヘッドセットを着けてもらい、衝立越しに A が B に目的地までの順路を教える
- ・**マイク・スピーカ条件**: A、B がマイクとスピーカの前に立ち、衝立越しに A が B に目的地までの順路を教える
- ・**ロボット条件**: A がロボットに目的地までの順路を教える。ロボットは、Wizard of oz (WOZ) 法により、人と同様のあいづち、身体表現ができるように実験者が操作を行った
- ・**対面条件**: A と B が対面し、目的地までの順路を教える

連絡先: 小野哲雄, 〒041-8655 函館市亀田中野町 116-2 公立はこだて未来大学, Tel: 0138-34-6132, tono@fun.ac.jp

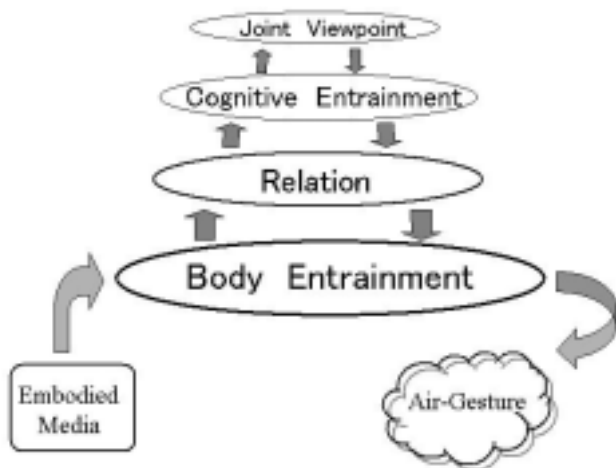


図1 身体化された音声インタラクションのモデル



(a) ヘッドセット条件 (b) マイク・スピーカ条件
(c) ロボット条件 (d) 対面条件
図2 インタラクション実験における被験者の身体動作

4. 実験結果と考察

まず、「身体動作」に関する実験結果について簡潔に述べる。図2に、インタラクション実験における被験者の様子を示す。まず、ヘッドセット条件では、2人の被験者は衝立により仕切られ、互いはまったく視界に入らないにもかかわらず、引き込むような身体動作が現れた(図2(a))。しかし、マイク・スピーカ条件では、2人の被験者の間に、まったく身体動作は現れなかった(図2(b))。一方、ロボット条件では、ロボットがリアルタイムでは人と同程度の身体表現が出来なかったにもかかわらず、被験者は身体動作を行いながらロボットに説明を行っていた(図2(c))。最後に、被験者同士の対面条件では、体の向き、腕、首を同調的に動作させながら、円滑なインタラクションを行っていた。モーションキャプチャのデータから「身体動作」の積算量の多い順に並べると以下のとおりであり、マイク・スピーカ条件と他の3条件の間に有意な差があった。

対面 > ヘッドセット > ロボット >> マイク・スピーカ

次に、「空間情報の伝達」に関する実験結果について述べる。空間情報の伝達の指標としては、順路を教えられた被験者が実際に目的地にたどりつけたかどうか、およびその所要時間を調べた。まず、対面条件では、全員の被験者が目的地へたどりつくことができた。しかし、ヘッドセット条件とマイク・スピーカ条件

では、5人中3人のみが目的地へたどりつくことができた。また、所要時間については、3条件に有意な差はなかった。

最後に、質問紙による被験者の主観的な評価の結果について述べる。この結果では、マイク・スピーカ条件において、順路を教える被験者の「教えやすさ」の評価点が有意に低くなっていた。つまり、対面条件はもとより、ヘッドセット条件、ロボット条件よりもマイク・スピーカを使って教える場合に不自然さを感じていたと考えられる。

以上の実験結果から、本稿で提案した「身体化された音声インタラクションのモデル」(図1参照)について考察を行う。人と人が対面して情報伝達を行う条件では、被験者間の身体的動作に引き込みがおり、情報伝達も正確に行われていた。このことは、同調的な身体動作に基づく「認知的な引き込み」により両者が「共有視点」を獲得していたことを示唆する。

また、対面条件と同様に、「身体化されるメディア」としてのヘッドセットを用いた条件でも、被験者間に引き込まれるような同調的な身体動作が見られた。2人の被験者は、衝立をはさんでいるため、互いがまったく見えないことを考えると、これらの同調的な身体動作は大変興味深い。しかし、「身体化されないメディア」としてのマイクとスピーカを用いた条件では、同調的な身体動作はまったく現れなかった。つまり、図1に示したように、「身体化されるメディア」は身体的な引き込みを起こさせ、スムーズなインタラクションを実現するが、「身体化されないメディア」では違和感のあるインタラクションとなっている。このことは、被験者の質問紙による評価からもうかがえる。

「身体化されるメディア」として、当然ロボットを考えることもできる。ロボット条件では、被験者はマイクとスピーカを用いてインタラクションするのは異なり、対面条件で見られたような身体動作を行いながら、順路を教えている。今回の実験では、WOZ法を用いているが、現在、本稿で提案したモデルに基づくロボットの動作機構を開発中である。この機構は、人と同調的な身体動作を行うだけでなく、人に「認知的な引き込み」を起こさせる機構も併せ持つ。このため、このロボットシステムは、インタフェースとしてのメディアだけではなく、コミュニケーションや言語の研究にも示唆を与えることが期待される。

5. おわりに

本稿では、身体化された音声インタラクションのモデルを提案し、実験によりモデルの検証を試みた。提案したモデルおよび実験結果は、音声インタラクションだけではなく、ウェアラブルとユビキタスコンピューティングの本質的な相違や、ロボットシステムの開発、およびエンタテインメントコンピューティングの研究[松原03]への有益な示唆を含むものである。

参考文献

- [小野 01] 小野哲雄他: 身体表現を用いた人とロボットの共創対話、情報処理学会論文誌、Vol. 42, No. 6, pp. 1348-1358.
- [Ono 01] Tetsuo Ono, et al.: A Model of Embodied Communications with Gestures between Humans and Robots. Proceedings of Twenty-third Annual Meeting of the Cognitive Science Society (CogSci2001), pp. 732-737.
- [松原 03] 松原仁、小野哲雄: 身体的エンタテインメントとロボット、電子情報通信学会総合大会チュートリアル講演、2003.