

プレゼンテーション映像における発話内容を用いた話者の動作理解*

7D-4

西谷 正志

中村 裕一

大田 友一†

筑波大学 電子・情報工学系†

1 はじめに

人間のコミュニケーションでは、音声言語によらないメッセージ伝達が重要な役割を果たしている。身体動作（表情、視線、身振り）はこのようなノンバーバル言語の1種であり、近年ではヒューマンインターフェースへの応用が検討されている。

プレゼンテーションにおいても話者の身振り動作が発話による言語的説明と同時に行われ、両者の協調によって話者の意図が伝達される場合が多い。

本研究はこの事実を着眼点とし、プレゼンテーション映像における動作情報と言語情報を統合し、話者の意図の推定と動作の解釈を可能にする手法を提案する。本稿では、身体動作と音声言語の時間的な同時性や連続性に基づく解析を行い、人間の身体動作の解釈と発話内容の理解のために、双方の情報を協調して用いることの有効性を示す。

2 ノンバーバル言語

身体動作の認識が最もよく研究されているものとして手話が挙げられる。手話では動作と、それが表す言語との間にはっきりとした規約があるため、その規約を基に翻訳を行うことができる。

それに対して、一般的な身体動作には無意識に行われる動作や、言語的な意味付けがなされていない動作が多数含まれる。Ekman らによると、それらは以下のように分類される [1]。

- (1) 標識 (emblem): 言語翻訳可能な動作
手話やサインなど言語的な同意がなされた動作
- (2) 例示子 (illustrator): 発話内容の補足強調
発話に付随する空間的な表現や指示動作など
- (3) 情感表示 (affect display): 表情による情緒表示
主に表情による怒り、喜びなどの情緒の表れ
- (4) 調整子 (regulator): 会話の流れの調整
主に視線の動きによる発話の促進や話者の交替
- (5) 適応子 (adaptor): 環境に適応する動作
腕組み等の発話内容と無関係に表れる動作

また、身体動作によって伝達されるメッセージの多くは、文脈や状況に依存することが多い [2]。

そのため、プレゼンテーション中の話者が意図的に行った動作でも、動作情報だけではその意味を理解するために不十分である。

3 身体動作と自然言語の対応

動作と言語の間には辞書的な意味関係だけではなく、実際には相互関係がいくつも存在している [4]。そこで、本研究ではプレゼンテーションに表出する動作と発話に含まれる言語の間に存在する関係に注目し、それを用いて動作と言語を相互に対応付けることを行う。

このような関係には図1で示すように、(1) 動作や姿勢（動作の静止状態）が持つ言語の意味を用いた対応関係、(2) 身体動作から特定した注視・指示対象の位置や知識情報を用いた対応関係がある。また、プレゼンテーションにおける身体動作は音声言語と同時に表出する場合が多く、(3) 動作と発話された言語の時間的な対応関係は重要な糸口となる。

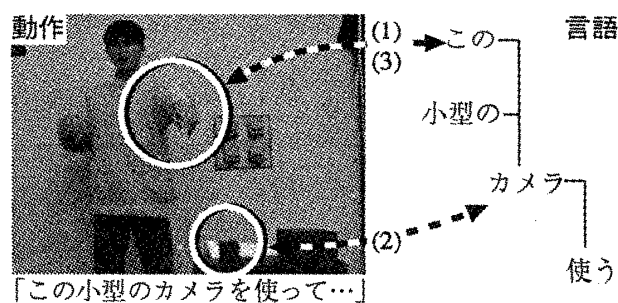


図1: 動作と言語のリンク

4 動作と言語の時間的対応関係の解析

ここで提案する枠組を実現するために、カメラや接触型センサから話者の動作を取得するシステムを構築し、発話内容の単語列との対応関係を推定する実験を行った。

4.1 動作情報と言語情報の取得

動作情報の取得にはカメラによって撮影される動画像を用いるが、プレゼンテーションを行う発話者に複数の磁気センサを装着し、装着部位での位置・方位角の情報を併用することで、画像での三次元情報の欠落を補い、動作抽出の高速化をはかる。

言語情報に関しては映像に付随する音声言語の自動認識が現在の段階では難しいため、発話内容を自然言語の文章として手動で与える。ただし、各文節が映像フレームのどこで発話されたのかを表す情報を同時に与え、言語と動作の時間的な同時性・連続性を残す。

4.2 対応関係の解析

動作に関しては身体動作において特に重要である手の動きに注目した。そこで、発話者の胴体位置を原点とした三次元座標系での右手位置の移動速度と移動方向の変化を用いて、動作情報のセグメンテーションを行う。

*Analysis of Nonverbal Expressions with Spoken Dialog

†Masashi NISHITANI, Yuichi NAKAMURA, Yuichi OHTA

‡University of Tsukuba

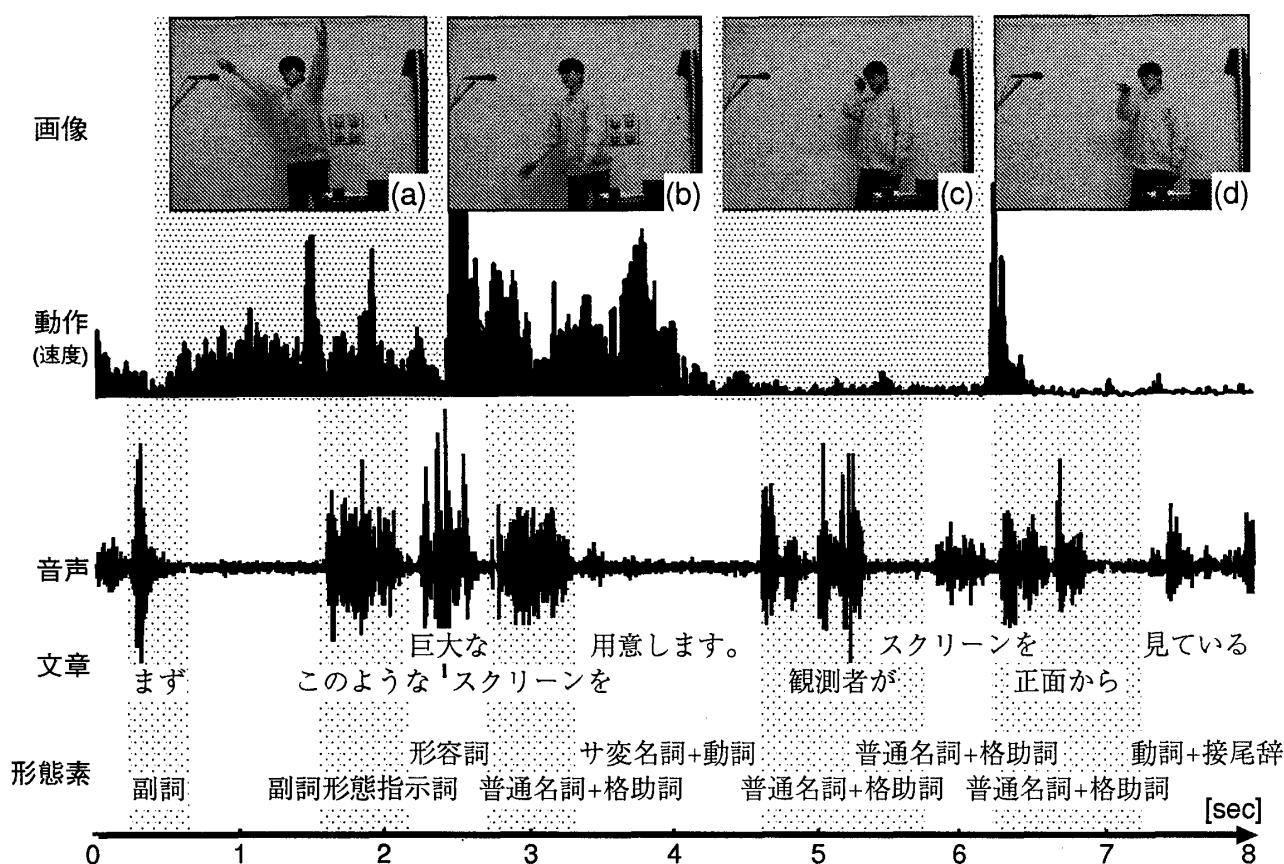


図2: 動作と言語の時間的対応

一方、言語に関しては発話文を形態素解析して形態素に分解した。発話文の構文解析は行っていないが、形態素間の係受け関係は比較的容易に求めることができる。

図2は動作と言語の時間的変化を表し、各セグメントが表出する時間から推定される両者の対応関係と Ekmanらの分類法に基づく動作種別を表1に示す。

図表で(a)の動作は時間的に対応する副詞形態指示詞「このような」の係受け関係から、動作が「スクリーン」を表現している例示子であると推定できる。ただし、図表で(d)の「人指し指で視線方向を示す」動作が「見る」という言語的意味を持つ標識であると推定するためには、身体動作が持つ言語的意味に関する知識が必要である。

実験結果が示すように、動作情報と言語情報の間には比較的わかりやすい対応関係が存在することが確認できた。このようなことから、人間の動作解釈や発話内容の理解のために本手法が有効であると考えられる。

5 おわりに

身体動作の解釈と発話内容の理解のために、プレゼンテーション映像における話者の動作情報と言語情報を統合する手法を提案した。さらに、両者の時間的な対応関係を推定する実験を行い、本手法の有効性を示した。

今後は両者の意味的な対応関係を含めた解析を行い、身体動作のより詳細な解釈を進めていく予定である。

表1: 対応関係の推定

動作	言語		動作種別
(a)	このような	副詞形態指示詞	例示子
	巨大な	形容詞	
	スクリーン	普通名詞	
(b)	用意	サ変名詞	
	する	動詞	
(c)	観測者	普通名詞	
	スクリーン	普通名詞	
(d)	正面	普通名詞 (空間語)	標識
	見る	動詞	

参考文献

- [1] Ekman, P. & Friesen, W.V.: "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding", *Semiotica* 1, pp49-98, (1969)
- [2] Herskovits, A.: "Language and Spatial Cognition", Cambridge University Press, (1986)
- [3] Kurokawa, T.: "Gesture Coding and a Gesture Dictionary for a Nonverbal Interface", *IEICE Trans. Fundamentals* E75-A, pp112-121, (1992)
- [4] McNeil, D.: "Psycholinguistics", Harper & Row (1987)