

## 随伴する発話の韻律情報に基づく動作意図の理解

### Understanding the Intentions of Actions Based on the Prosodic Information of Accompanying Utterances

櫻井 晴章†      岡 夏樹†  
Haruaki Sakurai      Natsuki Oka

京都工芸繊維大学大学院工芸科学研究科  
Graduate School of Science and Technology, Kyoto Institute of Technology

#### 1. まえがき

多くの情報システムやロボットが、身近な存在として日常生活に浸透してきている。それに伴って、人間が安心して接することができ、かつなるべく負担が少なく済むインタフェースが必要になってきている。

例えば、人間が手を滑らせて皿を割ってしまったのを見て、それをロボットが真似をしてしまうのは避けたい。また、カーナビゲーションシステムにおいて、カーナビゲーションシステムが最適と思われる道を教示したとき、人間が意図して違う道を選択した場合なら、次からはそのようなルートを教示するように適応してほしいが、人間が違って違う道を選択した場合は適応させたくない。

このように、人間側からの指示を待つ従来型のコンピュータのようなインタフェースではなく、人間の動作が意図した通りであったのか、偶然そうなったのかを汲み取って、自然に適応できる知的インタフェースの実現が期待される。しかし、意図的かどうかを判別し、意図された動作を模倣する能力は幼児でも持っていると言われており [Tomasello 97][Meltzoff 95] にも関わらず、工学的にそれを実現する方法はまだ確立していない。

人間の意図的動作と偶発的動作を判別する情報としては、音声や顔の表情、身体動作、身振りなどさまざまな情報が考えられる。本研究では、人間が偶発的な動作を行った場合、特徴的なタイミングで特徴的な発話を行うであろうと考え、発話のタイミングとその韻律情報に着目して、意図的動作と偶発的動作の判別を試みた。

矢入ら [矢入 02] は、介護用スクータであるピークルの開発を行っており、例えば、障害物に近づく行為や、ハンドルを切る行為が行われている最中に、その行為の結果が意図通り（障害物の手前で止まる）になるか誤って衝突してしまうかを判別するという研究を行っている。超音波センサ・実速度・ハンドル角などの時系列データを用いて、停止・衝突するどれだけ以前に操縦者の行動結果が意図通りになるかどうかを判別できるか、またそのためにはどれだけ遡った幅のデータを用いる必要があるかを検討している。

これに対して本研究では、行動の結果を予測するのではなく、行動後に起こった事象が意図に沿ったものかどうかを事象の生起後に判別することを目的とするため、

行動完了後の行為者の発話や行動の情報も用いて判別を行う。

#### 2. 実験

仮想空間上で、人間のユーザ（被験者）とシステムがボール遊びをすることができる実験環境を構築した（図 1）。被験者は、キー入力により自分の側の手（図中に「ユーザの手」と表記された手）を移動したり、握ったり開いたりすることができる。ボールのある場所で手を握るとボールをつかむことができ、手を握った状態でボールに接触するとボールをはじくことができる。また被験者は、システムに対して音声入力により言葉をかけることもできる。システムの側の手（図中に「システムの手」と表記された手）は、システムにより操作（移動と開閉）される。

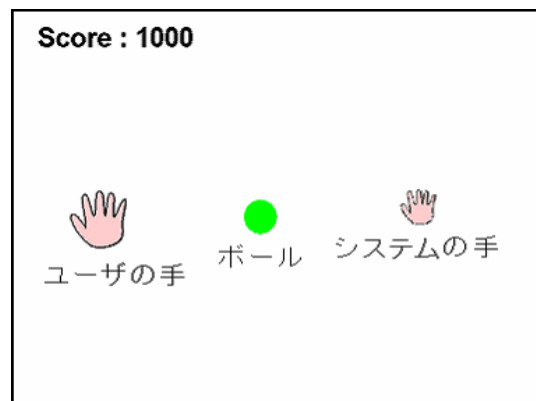


図 1：実験環境（仮想空間）の構成画面

この環境を用いて、ボールをつかむ動作やはじく動作を、お手本を見せながらシステムに教えることを被験者に指示し、その様子を 2 台のビデオカメラで記録した。被験者は、20 代の男性 9 名と女性 1 名の計 10 名である。

実験は、突発的な出来事への自然なリアクションを抽出するための「無意識的発話条件」での実験と、突発的な出来事への意識的なリアクションを抽出するための「意識的発話条件」での実験を行った。実験手順を表 1 に示す。4) において、突発的な出来事が起こった場合、それをシステムに警告する発話を行うことを指示した後、3) と同様に、ボールをつかんで投げる動作と、ボールをはじく動作をそれぞれ 3 分間ずつ教えさせた。指示の内容は次の通りである：「今度は意図しない動作が起こった場合には、システムにその動作をまねさせないために、警告する発話

† 連絡先: 岡夏樹, 京都工芸繊維大学工芸学部電子情報工学科, 〒606-8585 京都市左京区松ヶ崎御所海道町, 電子メール: oka@dj.kit.ac.jp

を行ってください。例えば“あっ”や“間違っ”という感じですよ。」

表 1: 実験手順

1) 実験環境の説明と操作練習 (5分)
2) 実験課題の説明
3) 「無意識的発話条件」での教示タスク
a) ボールをつかんで投げる行為をシステムに教えるタスク (3分)
b) ボールをはじく行為をシステムに教えるタスク (3分)
4) 「意識的発話条件」での教示タスク
a) ボールをつかんで投げる行為をシステムに教えるタスク (3分)
b) ボールをはじく行為をシステムに教えるタスク (3分)

### 3. 特徴量

反射的な発話や行動である場合も含めて、話者の感情や意図は韻律情報に反映されやすいと我々は考える。本研究では特に、動作完了前後の韻律情報(今回はパワーを使用)の変化や、動作完了後に発話を行うタイミングに注目し、意図的動作と偶発的動作の判別を試みた。

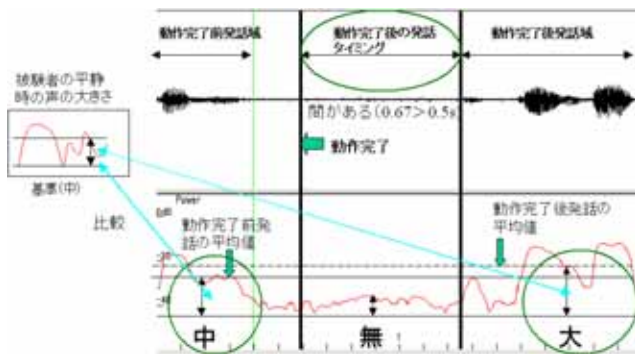


図 2: 特徴量のラベル付け

図 2 に、ある被験者の実験中に現れた発話“つかんで...つかんだら...”のタイミングとパワーの変化を示す。判別に用いた属性は、動作完了前発話のパワーの大きさ(f)と、動作完了後の発話タイミング(t)(動作完了時から動作完了後発話開始までの時間)と、動作完了後発話のパワーの大きさ(b)である。

発話の大きさは、被験者が普段平穏時に出す声の大きさを基準として、実際に聞いた声の大きさとパワーの波形から、“無”(発話なし)、“大”(基準よりも大きい)、“中”(基準とほぼ同じ)、“小”(基準よりも小さい)のいずれかに、発話タイミングは、発話がない場合を“無”、動作完了後 0.5 秒以内を“直後”、0.5 秒以上を“間がある”のいずれかにラベル付けを行った。

0.5 秒を基準として採用した理由は次の通りである。西本ら[西本 04]は、簡単な早押しゲームと記憶課題を用いた実験を行って被験者の反応時間を測定している。画面に矢印が表示されてから、その矢印の向きがその矢印の下にあ

る矢印の向きと同じであれば正解ボタン(キー)を、異なれば不正解ボタン(キー)を押すが、一定時間ごとに記憶課題を行わなければならない。2 単語記憶課題を行わせた被験者 5 人の反応時間の平均は約 0.5 秒であった。矢印の向きの判別は、正しいか正しくないかの判別であり、意図的か偶発的かの判別と同程度の反応時間を要するのではないかと考えた。また、2 記憶課題はほとんど負荷を感じることなく遂行できる課題であり、人間が反射的に行う行動の応答速度を測定したことになると考えた。そこで本実験では、動作の偶発的結果を知覚しての反射的な発話であるかどうかを、0.5 秒を基準として分けることができるのではないかと考えた。

### 4. 決定木による判別実験の結果と考察

動作完了前後の発話のパワーの大きさ(f) (b)と発話タイミング(t)の 3 属性を入力とした、意図的動作と偶発的動作の判別を、決定木アルゴリズム ID3[Quinlan 86]を用いたプログラムにより学習し、交差検定法を行い正解率を求め、表 2~4 の結果を得た。

表 2~4 では、被験者ごとの、ID3 による判別の正解数、意図的動作回数 P と偶発的動作回数 N、全体数(P+N)、ID3 による判別の正答率をまとめた。表中の「平均」の行は「10 人の被験者ごとに、個人ごとのデータで個々に学習し、個人ごとのデータで個々に評価した結果の、10 人分の平均」である。表中の「全体」の行は「10 人分のデータをひとまとめにして学習し、評価した結果」である。表 2 において、無意識的発話条件よりも意識的発話条件の方が正答率が高くなった。これは、意識的発話を行わせることで、ユーザが意図的動作時と偶発的動作時で区別しやすい発話を自然に行うようになっていたことが原因と考えられる。また表 3 において、つかむとはじくではほぼ等しい正解率となり、操作の難易度には差が見られなかったと考えられる。さらに、表 2~4 において「平均」の行の正答率がすべて 70%を超え、「全体」の行の正答率は「平均」の行の正答率よりも 10%ほど下回った。これは、意図的動作と偶発的動作それぞれが持つ特徴には個人差があり、両者を精度よく判別するためには、個人適応する必要があることを示す。

以上により、意図的動作と偶発的動作を判別するための特徴量として、動作完了前後の発話のパワーと発話タイミングを用いることがある程度有効であるが、精度の高い判別のためには個人適応する必要があることが明らかになった。

### 5. おわりに

本研究では、意図的動作と偶発的動作を判別するための特徴量として、動作完了前後の発話のパワーと、発話タイミングを用いることを提案し、仮想空間上で人とシステムがインタラクションする環境で採取したデータを用いた判別実験を行い、その判別性能を評価した。

今後の課題として、1) 音声以外の判別情報として、表情や視線、身振りなどの特徴量も抽出する、2) 教えることを意識していない状況で、意図的か偶発的かが判別可能かどうかを調べる実験、3) 意図的かどうかの判別とピッチ情報との関連性、4) 現実世界でボール遊びを行う場合と仮想空間上でボール遊びを行う場合との比較、5) 意図的

かどうかを自動で判別できる学習モジュールの構築, などを検討していきたい. 本研究の一部は国立情報学研究所との共同研究「人間とエージェントの相互適応」によるものである.

参考文献

[Tomasello 1997] Tomasello, M.: The Pragmatics of Word Learning, 認知科学 4(1):59-74, 1997.  
 [Meltzoff 1995] Meltzoff, A. N.: Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology 31: 838-850, 1995.  
 [矢入 2002] 矢入(江口)郁子, 矢入健久, 猪木誠二: 多次元時

系列センサデータからのピークル操縦者の意図認識の研究, 人工知能学会全国大会, 1D1-05, 2002.

[西本 2004] 西本 卓也, 高山 元希, 櫻井 晴章, 荒木 雅弘: 音声インタフェースのための対話負荷測定法, 電子情報通信学会論文誌 J87-D-II(2):513-520, 2004.

[Quinlan 1986] Quinlan, J. R.: Induction of Decision Trees. Machine Learning 1(1):81-106, 1986.

表 2: 無意識的発話条件と意識的発話条件の決定木学習結果の比較

被験者	無意識的発話条件				意識的発話条件			
	正解数	P:N	全体数	正解率 (%)	正解数	P:N	全体数	正解率 (%)
A	10	8:6	14	71.4	8	5:4	9	88.9
B	8	8:4	12	66.7	13	8:11	19	68.4
C	13	6:11	17	76.5	16	8:11	19	84.2
D	10	6:8	14	71.4	16	12:11	23	69.6
E	6	5:4	9	66.7	12	7:9	16	75.0
F	15	13:5	18	83.3	12	9:15	14	85.7
G	15	11:10	21	71.4	15	10:10	20	75.0
H	6	4:5	9	66.7	12	5:11	16	75.0
I	10	8:8	16	62.5	13	9:9	18	72.2
J	7	4:6	10	70.0	11	5:9	14	78.6
平均	10.0	7.3:6.7	14.0	70.7	12.8	7.8:9.0	16.8	77.4
全体	100	73:67	140	62.9	128	78:90	168	67.3

(被験者 A~J, 平均, 全体(被験者 A~J のデータを足し合わせた場合)の正解率を求めた. また意図的動作回数 P と偶発的動作回数 N を列挙した)

表 3: つかむ動作とはじく動作の決定木学習結果の比較

被験者	つかむ動作				はじく動作			
	正解数	P:N	全体数	正解率 (%)	正解数	P:N	全体数	正解率 (%)
A	12	9:6	15	80.0	6	4:4	8	75.0
B	14	8:11	19	73.7	9	8:4	12	75.0
C	18	7:15	22	81.8	10	7:7	14	71.4
D	17	11:13	24	70.8	10	7:6	13	76.9
E	11	8:7	15	73.3	8	4:6	10	80.0
F	9	9:3	12	75.0	15	13:7	20	75.0
G	14	10:9	19	73.7	15	11:11	22	68.2
H	13	5:12	17	76.5	6	3:5	8	75.0
I	11	6:8	14	78.6	16	11:9	20	80.0
J	13	5:11	16	81.2	6	4:4	8	75.0
平均	13.2	7.8:9.5	17.3	76.5	10.1	7.2:6.3	13.5	75.2
全体	132	78:95	173	65.3	101	72:63	135	66.7

表 4: 全体の決定木学習結果の比較

	正解数	P:N	全体数	正解率 (%)
平均	11.5	7.5:7.9	15.4	74.1
全体	200	151:157	308	64.8