

# 情動刺激が発話と身振りのタイミングに与える影響

東京工業大学大学院 総合理工学研究科 知能システム科学専攻 ○沖津 健吾, 三宅 美博

## Effects of emotional stimuli to the timing of utterance and gesture in dialogue

○Kengo Okitsu and Yoshihiro Miyake Tokyoh Institute of Technology, Interdisciplinary Graduate School of Science and Engineering

**Abstract:** In this study, to develop the speech dialogue interface for the smooth dialogue with human, we discussed the effects of the emotional states to the user speech timing - utterance and gesture. Concretely developing the timing model of utterance and gesture, we compared the human response time between the emotional states - anger state and happy state. As a result, we found the possibility that emotional states affect the user speech timing.

### 1 はじめに

#### 1.1 発話と身振りのタイミングモデル

人間は対話コミュニケーションを通じて、他者との意思疎通を図ることができる。その際、言語情報だけでなく、音声の韻律やジェスチャーといった非言語情報も重要な役割を果たしている<sup>[1]</sup>。このような非言語情報の中でも発話タイミングが円滑なコミュニケーションにおいて重要な要素として注目されてきた<sup>[2-6]</sup>。例えば、渡辺らは発話とうなずきのリズムに引き込み現象が観察されることを示し、様々なインタフェースに應用している<sup>[6]</sup>。

我々の研究グループでは、対話における発話と身振りのタイミングの関係を同時に解析している研究がなされていないことに着目し、双方を包括的に解析する研究を行ってきた。具体的には、2人の話者が指示と応答を繰り返す対話において、指示者が意図的に発話速度を変化させ、被指示者がそれを認知している場合と、そうでない場合では、被指示者の発話と身振りのタイミングモデルが異なることを示した<sup>[7]</sup>。また、意図的に発話速度を変化させた場合において時間的特徴量を計測し解析を行った結果、特徴的な3つの相関関係を発見した<sup>[8]</sup>。さらに、この相関関係をもとにした発話と身振りのタイミングモデルをヒューマノイドロボットに実装し、そのようなタイミング制御を行っている場合としない場合とを比較して印象評価実験を行った所、タイミング制御を行っている方が被験者にとって好ましいという結果が得られた<sup>[9]</sup>。これらのことから、発話と身振りのタイミングモデルは人との円滑なコミュニケーションを実現する対話インタフェースにおいて非常に重要な要素であると考えられている。

#### 1.2 本研究の目的

情動の認知・表出や感情の伝達は、コミュニケーションにおいて非常に重要な要素である<sup>[10]</sup>。人間はコミュニケーションにおいて非言語情報を用いて情動状態を伝達

していると考えられており、その認知と表出のメカニズムの解明が求められている。これまで人の表情データや韻律情報データを基にした情動表出モデルを持つインタフェースの構築といった研究がなされてきた<sup>[11]</sup>。しかし、対話コミュニケーションという状況において、情動状態が発話と身振りのタイミングといった時間的側面にどのように影響を与えるかについての研究はまだない。そこで本研究では対話コミュニケーション状況において、情動刺激により被験者に喚起される情動状態が発話と身振りのタイミングに与える影響を調べることを目的とする。具体的にはPC上に表情や身振りなどの非言語チャンネルを持たないアバタを構築して、アバタの指示に対して人間側が応答する指示-応答対話を用意し、その際の被験者の反応の時間的特徴量の計測・解析を行う。

### 2 実験手法

#### 2.1 被験者

被験者は、22歳から24歳まで(平均23.6歳)の10人(うち女性2人)であった。

#### 2.2 対話課題と時間的特徴量の定義

本研究ではアバタと被験者の間にターゲットとしてマウスを用意して、アバタが「おしてください」と指示したのに対し、被験者が「はい」と応答してマウスのボタンを押すという課題を用いた。実験で使用・計測した時間的特徴量は図に示すように、アバタの指示発話長(IU:Duration of instruction utterance), 交替潜時長(SP:Duration of switching pause), ボタンを押す動作の開始時間と応答発話開始時間までの時間差(PT-RT:Duration of Push and Response Timing), ボタン押し動作の長さ(Push:Duration of Push)の4つであった。

#### 2.3 実験装置と刺激

アバタと音声はApple社のMacBookPro(late 2008, 15.4inch)を用いて提示された。アバタは武藤ら<sup>[12]</sup>が

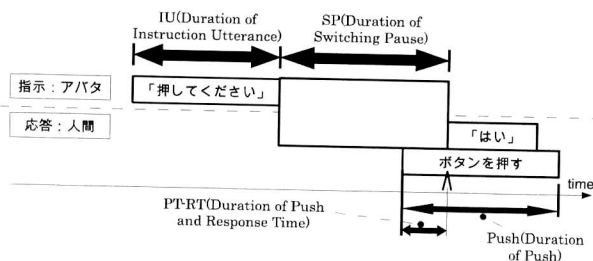


Fig. 1 Indices of dialogue.



Fig. 2 Avatar and facial stimulus.

用いた画像を本実験用に加工して使用した。アバタは表情や身振りを一切持たないものであり、被験者はディスプレイから1.1mの位置にアバタと対面して座るよう指示された。被験者の音声は被験者の前に固定して置かれたマイクで取得し、発話開始時刻は音圧が一定値を超えた瞬間の時刻とした。マウスを押しに行く動作の終了時刻は、被験者がターゲットのマウスを押しした瞬間とし、動作の開始時刻は被験者の手元にもう一つあるマウスを被験者が離れた瞬間とした。

アバタが発する音声は音声合成ソフト EasySpeech Version 1.0.0.7<sup>[13]</sup> を用いて作成した。本実験では音の高さ [Hz] を変えず早さ [語/分] のみを操作した4種類のwavファイルを音声刺激として使用しており(ファイルの長さ: IU=608,872,1029,1477[msec])、早い順に“早め”, “やや早め”, “やや遅め”, “遅め”の4段階で被験者に教示した。情動喚起刺激として、情動の神経科学や感情の心理学で広く用いられている Pictures Of Facial Affect (POFA)<sup>[14]</sup> を使用した。今回は普通、怒り、喜びの3種類の表情刺激を使用し、枚数はそれぞれ普通14, 怒り17, 喜び18であった。実験はマイクにノイズが入らないように防音室環境で行った。アバタを含めた視覚刺激は全て視角が水平約8.4°, 鉛直約12.4°になるように被験者に提示した。また、アバタの顔部分、顔表情写真の中心部分がスクリーンの中心に来るように提示し、被験者の目の高さスクリーンの中心が来るように調整した。被験者からスクリーンの中心までの距離は1.1[m]、被験者の手元のマウスとターゲットのマウスとのボタン間の距離は約30[cm]とした。

## 2.4 実験課題

### 2.4.1 練習

まずアバタとの対話に慣れてもらうために練習を行ってもらった。はじめに行った教示は以下の3つである。

- アバタに話しかけられたら発話「はい」と頷きとボタンを押すという3つ応答を、自分にとって自然なタイミングで行ってください(応答の統制)
- 両手を机の上に置いてください(姿勢の統制)
- スクリーンの中心を見てください(注視点の統制)

練習手順は、まずスクリーンに4つの発話長の条件のうちアバタがどの早さで話すかを提示される。その後アバタが表示され被験者がクリックをすると試行が開始され、クリックから1~3秒後に流れるアバタからの「おしてください」という指示発話に対して応答をしてもらい、その際の反応時間を計測した。これを4つの発話長条件に対してそれぞれ連続して4回ずつの計16試行を行ってもらった。

### 2.4.2 表情カテゴリー

被験者が正しく情動を認知できるかを確認するため、表情弁別課題を行ってもらった。提示される表情画像を普通(Neutral)、怒り(Anger)、喜び(Happy)のいずれかにカテゴリー化してもらおうよう教示し、ミスをした画像については正解を確認してもらい正しくカテゴリー化が行えるようにした。

### 2.4.3 本実験

本実験は2つのセクションに分けて行った。2.4.1練習に加えて、試行開始のクリックから1~3秒後に1秒間表情画像を提示し、その1秒後に指示発話が始まるようにした。はじめのセクションでは、1つの発話長の条件に対して15回試行を行ってもらい、始めの5回は普通表情、次の10回は怒り表情をランダムに提示した。開始前被験者には正しく情動が惹起されるように練習試行時の教示に加えて以下の教示を行った。

- 提示された表情と同じ感情を感じるようにしてください
- 普通から感情表情に変わったら感情を感じるようにしてください
- 表情写真は対面している相手として見ないでください。話しかけているのはアバタで、表情写真はあなたの感情を喚起するためのものです。

• 1条件15回の試行が終わったら、気持ちが落ちつくまで次の条件の試行を始めないでください。

4つの発話条件それぞれに対して行ってもらい、計60回の試行を行ってもらった。3~5分の休憩の後、次のセクションでは怒り表情を喜び表情に変えて同様の実験を行い計60回試行を行ってもらい、2セクション合計120回の試行を行ってもらった。つまり、普通(neutral)、怒り

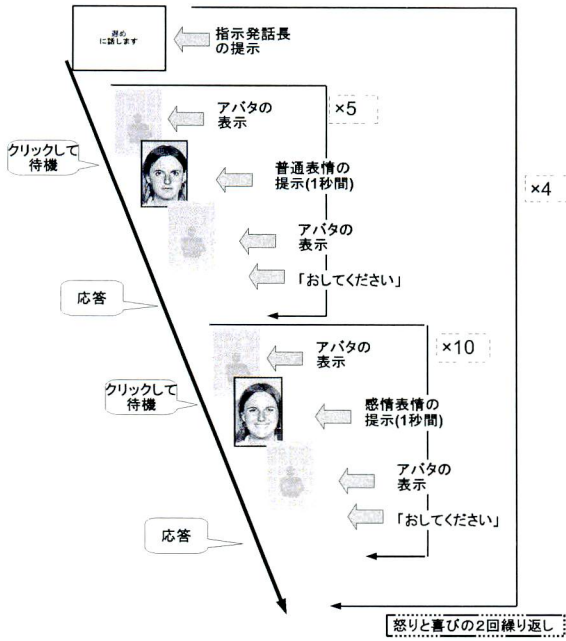


Fig. 3 Experiment.

(anger)、喜び (happy) 表情それぞれについて 40 試行分の反応時間データを得たことになる。

#### 2.4.4 アンケート

本実験後、表情と同じ感情をどれくらい強く感じられたかを 5 段階 (1:全く感じなかった、5:とても感じた) で評定してもらった。

#### 2.5 実験データの解析

3つの時間的特徴量のデータを、各被験者毎に以下の12の条件の組合せそれぞれ10試行分のデータを得た。これ

[発話長の条件 × 4]		[情動状態の条件 × 3]	
早め、やや早め	やや遅め、遅め	普通 (neutral)	怒り (anger)
		喜び (happy)	

らのデータから、音声の認識ミスや被験者のボタン押し損ねによる明らかな失敗試行のデータを除いた後、Excelの統計ツールボックスなどを用いて解析を行った。

### 3 結果

#### 3.1 アンケートの結果

アンケートの評定は怒り:  $mean = 3.8 (SD = 0.7888)$ , 喜び:  $mean = 4.3 (SD = 1.0593)$  であり、被験者毎の結果は以下の表のようになった。subject Dを除いて、概ねう

まく感情を感じられていたことが分かる。以降のデータ解析では subject D のデータは除くものとする。

Table 2 Score of questionnaire

subject	Anger	Happy
A	4	4
B	4	3
C	4	5
D	2	2
E	4	4
F	5	5
G	5	5
H	3	5
I	4	5
J	4	5

#### 3.2 情動刺激の違いによる時間的特徴量の変化

被験者全体で共通した傾向として、Pushは全ての被験者で neutral の時より happy の時の方が短いという傾向が見られた (Fig.4)。次に neutral、anger、happy を要因とした平均値の差の検定を行った。手順はまず等分散性の検定 (Bartlett's test) を行い、等分散性の仮定が棄却された場合は Steel-Dwass の多重比較、棄却されなかった場合、正規性の検定 (Kolmogorov-Smirnov test) の後 Tukey の多重比較を行った。その後、多重比較によって見られた有意差・傾向の組合せにより、どの情動状態に影響を受けたかを考察した (Table.3)。結果は Table.4 のようになった。被験者 E 以外で情動刺激の違いにより平均値が変化していることが分かる。

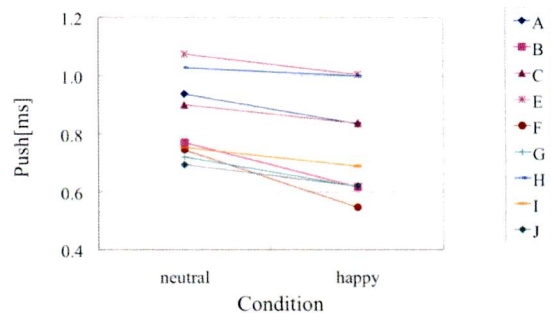


Fig. 4 Push graph

#### 3.3 時間的特徴量間の相関係数の変化

情動刺激毎にタイミングモデルが変化するかを調べるため相関係数の変化の検定を行った。具体的には、情動条件の違いによる標本相関係数の同等性の検定、またさらに母相関係数が少なくとも一つは 0 ではないことを検定し、前者の帰無仮説が棄却かつ後者の仮説が採択されたら情動状態で異なるタイミング相関が見られたとした。

Table 3 Correspondence table.

	1	2	3	4	5	6	7	8
neutral-anger		○			○		○	○
neutral-happy			○			○		○
anger-happy				○		○	○	○

- … 有意差 ( $p < 0.05$ ) または傾向 ( $p < 0.1$ )
- 1~5 … 情動状態の違いによる変化はなし
- 6 … 喜び情動状態に影響を受けた
- 7 … 怒り情動状態に影響を受けた
- 8 … 情動状態に影響を受けた

Table 4 Changes of average.

subject	SP	Push	PT-RT
A	6	3	4
B	7	8	4
E	2	4	1
F	2	5	6
G	1	8	7
H	5	1	6
I	8	8	7
J	1	6	6

結果は Table.5 のようになり、被験者 B,C 以外で相関係数の変化が見られた。

#### 4 考察

本研究では、アバタとの対話コミュニケーション状況において、情動刺激により喚起される情動状態が発話と身振りのタイミングに与える影響について調べた。その結果として、被験者全員に、事前に提示される情動刺激の違いによる発話と身振りのタイミングへの効果が認められたことから、人の表情刺激により喚起される情動状態が、対話における発話と身振りのタイミングに影響を及ぼす可能性があることが分かった。今後は被験者間に共通する変化傾向の更なる調査を目指して研究を進める予定である。

Table 5 Changes of coefficient of correlation.

subject	IU-SP	IU-PT_RT	IU-Push	SP-PT_RT	SP-Push	PT_RT-Push
A						○
B						
C						
E	○	○	○	○		○
F			○			
G				○		
H	○					
I	○					
J	○					

- … 情動状態の違いで異なるタイミング相関係数

#### 5 参考文献

- [1] 渡辺: コミュニケーションにおける身体性; ヒューマンインタフェース学会誌, Vol.1, No.2, pp.14-18 (1999).
- [2] Condon,W., Sander,L.: Synchrony demonstrated between movements of the neonate and adult speech; Child Development, Vol.45, No.2, pp.456-462 (1974).
- [3] Matarazzo, J.D., Weitman, M., Saslow, G., Wiens,A.N.: Interviewer influence on durations of interviewee speech; Journal of Verbal Learning and Verbal Behavior, vol.1, pp.451-458, (1963).
- [4] 長岡: 対人コミュニケーションにおける非言語行動の2者間相互影響; 対人社会心理学研究, Vol.6,pp.101-112, (2006).
- [5] Webb, J.T.: Interview synchrony: An investigation of two speech rate measures in an automated standardized interview; In B. Pope and A.W. Siegman (Eds.), Studies in dyadic communication New York: Pergamon, pp.115-133 (1972).
- [6] 渡辺, 大久保, 中茂, 檀原: InterActor を用いた発話音声に基づく身体的インタラクションシステム; ヒューマンインタフェース学会論文誌, Vol.2, No.2, pp.21-29, (2000).
- [7] 山本, 平野, 小林, 高野, 武藤, 三宅: 対話コミュニケーションにおける2種類の発話タイミング相関; ヒューマンインタフェースシンポジウム 2007 講演会予稿集, pp.631-634 (2007).
- [8] 阿部, 山本, 武藤, 三宅: 対話における発話と身体動作のインタラクションの解析; ヒューマンインタフェースシンポジウム 2008 講演会予稿集, pp.287-290 (2008).
- [9] 高杉, 山本, 武藤, 阿部, 三宅: コミュニケーションロボットとの対話を用いた発話と身振りのタイミング機構の分析; 計測自動制御学会論文集, Vol.45, No.4, pp.215-223 (2009).
- [10] 村井: 社会化した脳; エクスナレッジ, (2007).
- [11] 後藤, 加納, 加藤, 国立, 伊藤: 感性ロボットのための感情領域を用いた表情生成; 人工知能学会論文誌 Vol. 21, No. 1 pp.55-62 (2006).
- [12] 武藤, 高野, 大良, 小林, 山本, 三宅: 音声対話インタフェースにおける発話タイミング制御とその評価; ヒューマンインタフェースシンポジウム 2007 講演会予稿集, pp.639-642 (2007).
- [13] <http://easyspeech.jp/>
- [14] Ekman, P., Friesen, W.V.: Measuring facial movement. ; Environ. Psychol. Nonverbal Behav. 1, 56?75. (1976).