情動刺激が発話と身振りのタイミングモデルに与える影響

沖津 健吾*1　　吉田 祥平 *1　　三宅 美博 *1

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

Kengo Okitsu*1,　Shohei Yoshida*1　and　Yoshihiro Miyake*1

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.**

Keywords : **emotion , dialogue , switching pause, timing, facial stimuli**

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 [1]。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。例えば、Condon らは母子コミュニケーションにおいて、音声リズムと身体リズム間の相互作用が重要な役割を果たしていることを示した [2]。 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応用している [3]。また、Matarazo らや Webb, 長岡らは、発話長、発話速度、反応潜時などが、話者間で同調する事を報告している [4−6]。

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 (「積み木を取ってください」) と応答（「はい」と言って頷き、机の上の積み木を取る）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 [7]。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、３つの相関関係が検出された [8]。

この時解析した時間的特徴量は、発話に関する特徴量 3 種類 (指示者発話長、被指示者発話長、交替潜時長) 、身体動作に関する特徴量 2 種類 (被指示者の頷き動作、把持動作の時間長)、被指示者の発話と身体動作に関する特徴量 2 種類 (頷き開始から発話開始までの時間長、把持動作開始から発話開始までの時間長) の計 7 種類でその内以下の組合せで相関関係が見られた。

1. 指示者の発話長 —交替潜時
2. 交替潜時 —頷き開始から発話開始までの時間長
3. 交替潜時 —把持開始から発話開始までの時間長

さらに、これら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い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 [9]。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　情動、感情と対話コミュニケーション

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。ここで、情動 (emotion) とは「自分の周りの事物・事象が、自分が生きていくうえで有益か有害かを速やかに評価することと、そのような評価に基づく身体の反応」[10] であるとする。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 [11]。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

*1: 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
*1: Tokyou Institute of Technology, Interdisciplinary Graduate School of Science and Engineering

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。

1.3 本研究の目的

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-応答対話系を用意し、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。条件として、対話の前に情動刺激を提示された場合と特に情動を喚起させない別の刺激を提示された場合を設定し、条件間で時間的特徴量に差が見られるかを調査する。

2. 実験手法

2.1 被験者

被験者は、22歳から24歳まで (平均22.6歳) の5人 (男性4人, 女性1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。

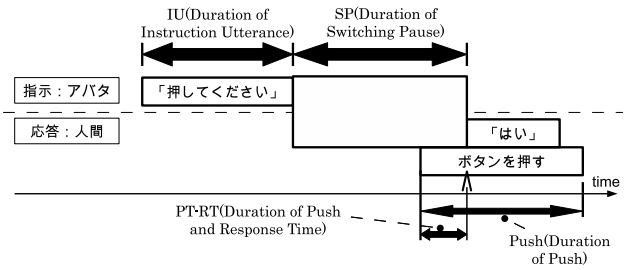


図1 時間的特徴量
Fig. 1 Indices of dialogue.

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[12]が発話タイミング制御とその評価を調査する際に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した瞬間とした。

図2 アバター
Fig. 2 avatar.

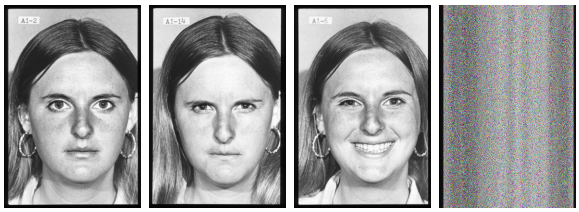


図3 表情刺激 (一番右はランダム画像))
Fig. 3 facial stimulus.

アバタが発する音声は音声合成ソフトEasySpeech Version 1.0.0.7[13]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)[14]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、被験者の目の高さに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

本研究では以下の3つの条件でアバタとの指示-応答対話を被験者に行ってもらった。

1. 事前に何も画像を提示しない (blank 条件)
2. アバタとの対話が始まる直前にランダム画像を提示 (random 条件)
3. アバタとの対話が始まる直前に表情画像を提示 (emotion 条件)
ここでランダム画像とは、表情画像のピクセルを縦方向にランダムに並び替えた画像のことである。

blank 条件は被験者のアバタとの対話の練習のために、random 条件はタスクに画像提示が入る事の効果を調べるために、emotion 条件は本研究の目的である、情動状態によって時間的特徴量に変化が現れるかを調べるために行った。全ての条件のタスクにおいて、アバタの発する指示発話長 (IU) ４種類をこちらで制御して提示した。

それぞれの条件で共通して行った教示は次の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)
以下にそれぞれの条件毎の詳細を説明する。

1.blank 条件

教示を行った後、実験に慣れるために練習を行ってから本実験に入った。

本実験は、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 10 回ずつの計 40 試行を行ってもらった。

2.random 条件

blank 条件に加えて、試行開始のクリックから 1〜3 秒後に 1 秒間ランダム画像を提示し、その 1 秒後に指示発話が開始するようにした。練習を行った後、4 つの発話長条件に対してそれぞれ連続して 10 回ずつの計 40 試行の本実験を行ってもらった。

3.emotion 条件

random 条件のランダム画像の代わりに表情画像を提示した。提示される表情画像は普通 (Neutral)、怒り (Anger)、喜び (Happy) の 3 種類を選び、この条件の実験開始前に被験者に表情弁別課題を行ってもらった。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ライズが行えるようにした。

実験は 2 つのセクションに分けて行った。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共通事項に加えて以下の教示を行った。

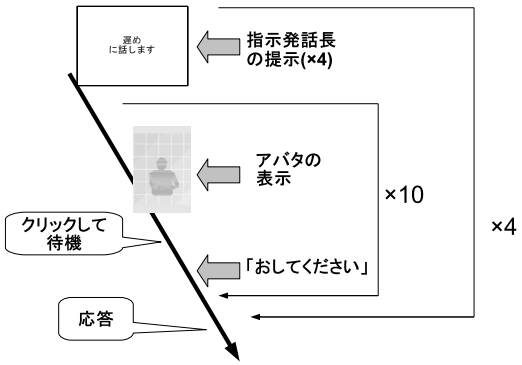


図 4　blank 条件
Fig. 4　blank condition.

・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
・普通から怒り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 (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。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

実験により、被験者の応答の交替潜時長 (SP)、動作の開始と応答発話開始の時間差 (PT-RT)、動作の長さ (Push) の 3 つの時間的特徴量のデータを取得した。被験者毎に以下の 20 の条件の組合せそれぞれ 10 試行分のデータを得た事になる。 これらのデータか

表 1　実験条件の組合せ
Table 1　Condition.

[発話長の条件× 4]	×	[タスクの条件× 5]
早め、やや早め		普通 (neutral), 怒り (anger)
やや遅め、遅め		喜び (happy),blank,random

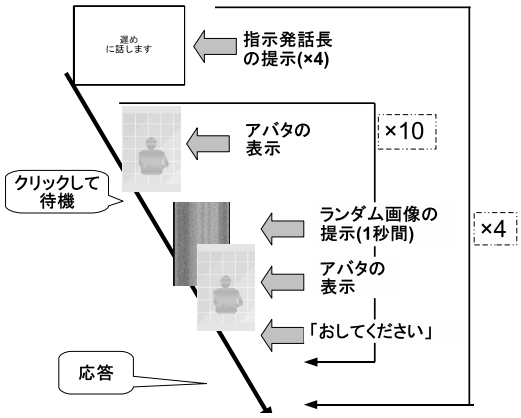


図 5　random 条件
Fig. 5　random condition.

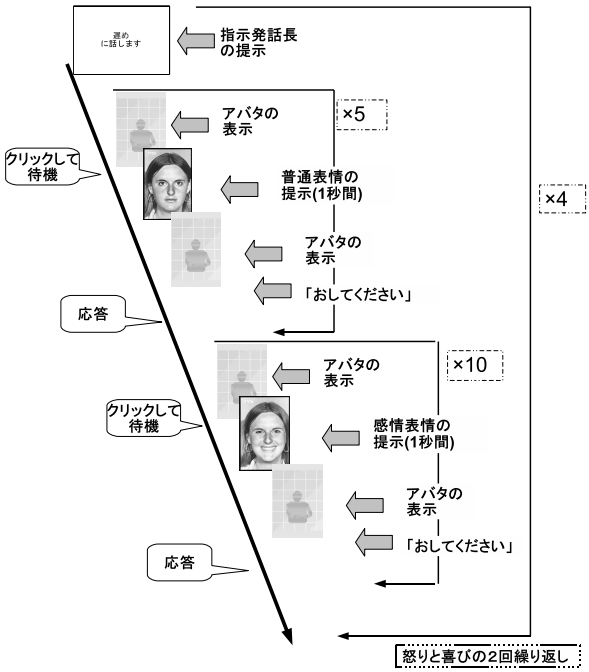


図 6　emotion 条件
Fig. 6　emotion condition.

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、Excell の統計ツールボックスを用いて解析を行った。

3.　結果

3.1　アンケートの結果

最初に、実験中に行った表情弁別課題の結果と、感情を感じた強度に関するアンケートの結果を示す。表情弁別課題のミスの数は平均 2.6、標準偏差 0.5477 であり、ほぼ問題なく表情を弁別することができていた。どれくらい感情を感じられたかのアンケートの評定は怒り:$mean = 3.6(SD = 0.8944)$, 喜び:$mean = 3.6(SD = 1.1402)$ であった。被験者毎のアンケート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。

表 2　感情評定アンケート
Table 2　How do you feel emotion.

subject	Anger	Happy
A	4	4
B	4	3
C	4	5
D	2	2
E	4	4

3.2　情動刺激の違いによる時間的特徴量の変化

次に、それぞれの条件の組合せにおける各変数の平均及び標準偏差を示す。

表 3　交替潜時長 (SP) の平均と標準偏差
Table 3　Mean and SD of SP.

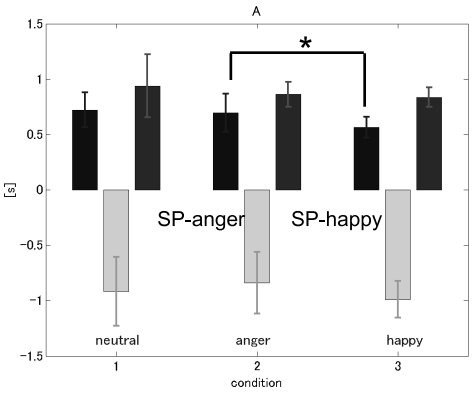
subject	blank	random	neutral	anger	happy
A :mean	0.643	0.645	0.722	0.696	0.564
SD	0.228	0.228	0.156	0.173	0.093
B :mean	0.562	0.634	0.671	1.039	0.592
SD	0.153	0.169	0.216	0.509	0.167
C :mean	0.960	0.772	0.797	0.656	0.631
SD	0.182	0.155	0.806	0.208	0.167
D :mean	0.229	0.314	0.509	0.527	0.553
SD	0.144	0.076	0.191	0.107	0.134
E :mean	0.918	1.018	1.144	1.269	1.319
SD	0.224	0.264	0.242	0.541	0.507

表 4　PT-RT の平均と標準偏差
Table 4　Mean and SD of PT-RT.

subject	blank	random	neutral	anger	happy
A :mean	-0.340	-0.553	-0.917	-0.840	-0.989
SD	0.314	0.351	0.311	0.279	0.166
B :mean	-0.625	-0.728	-1.021	-1.136	-0.931
SD	0.368	0.340	0.354	0.484	0.417
C :mean	0.681	0.313	-0.056	0.031	-0.490
SD	0.130	0.341	0.925	0.422	0.411
D :mean	-0.173	-0.240	-0.261	-0.312	-0.286
SD	0.184	0.079	0.172	0.101	0.141
E :mean	-0.582	-0.578	-0.746	-0.772	-0.731
SD	0.244	0.109	0.147	0.224	0.206

表 5　Push の平均と標準偏差
Table 5　Mean and SD of Push.

subject	blank	random	neutral	anger	happy
A :mean	1.289	1.238	0.939	0.862	0.836
SD	0.294	0.198	0.284	0.114	0.089
B :mean	1.217	1.103	0.772	1.081	0.618
SD	0.293	0.321	0.272	0.436	0.131
C :mean	1.840	1.369	0.901	0.860	0.838
SD	0.290	0.309	0.302	0.233	0.230
D :mean	0.484	0.495	0.518	0.486	0.589
SD	0.068	0.065	0.098	0.070	0.091
E :mean	1.212	1.004	1.075	1.175	1.005
SD	0.226	0.126	0.163	0.320	0.231



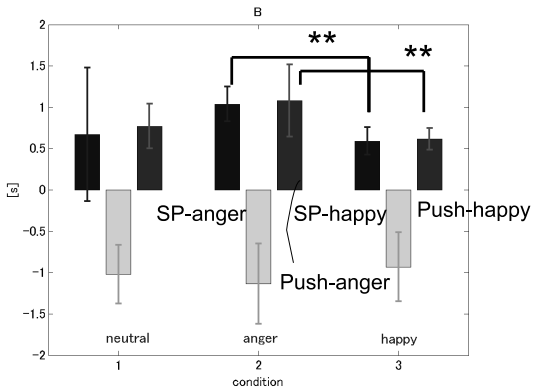


図 7 例：subject A,B のグラフ (*:p < 0.05),**:$p < 0.01$
Fig. 7 subject A,B.

バートレット検定を行った所、subject C の Push 以外で等分散性が保証されなかった。そこでその他についてクラスカル・ウォリス検定を行った所、全員の全ての特徴量に対して有意差が見られた ($p < 0.01$)。

情動状態による時間的特徴量の変化を見るため、被験者毎で Scheffe の多重比較を行い anger 条件と happy 条件の比較に有意差が出ている変数がないか解析した。すると、被験者 E 以外は anger-happy 間の比較で有意差のある変数が見られた (Table 6)。比較の様子の例を Fig 7 に示した。このことから、ほとんどの被験者は直前に喚起された情動が異なると、発話または動作のタイミングが変化していることが分かった。

表 6 怒りと喜びの多重比較で有意差が出た特徴量の組合わせ (*:$p < 0.05$,**:$p < 0.01$)
Table 6 Combination .

subject	parameter
A	SP*
B	SP**,Push**
C	PT-RT*
D	Push**

表 7 交替潜時と動作-発話の開始時間の差の相関
Table 7 SP PT-RT correlation.

subject	blank	random	neural	anger	happy
A	-0.0311	-0.319	-0.215	-0.472	-0.175
B	0.240	0.0265	-0.306	-0.0629	0.242
C	0.524	0.427	0.584	0.620	0.714
D	0.774	0.463	0.624	0.226	0.515
E	-0.453	-0.603	-0.646	-0.320	-0.815

表 8 発話長と交替潜時長の相関
Table 8 IU SP correlation.

subject	blank	random	neural	anger	happy
A	0.764	0.886	0.600	0.689	-0.205
B	0.0203	0.457	-0.0671	0.257	0.0203
C	0.253	-0.477	-0.162	-0.554	-0.340
D	-0.151	-0.571	-0.202	-0.594	-0.357
E	0.785	0.826	0.782	-0.817	0.879

3.3 時間的特徴量間の相関関係

最後に、タスクの条件毎の特徴量間の相関関係について示す。Table 7,8 は被験者毎の特徴量間の相関係数の例である。これらの表から、neutral 条件、anger 条件、happy 条件を比較してみると、提示した表情刺激の違いによってそれぞれの被験者で変化は認められるが、被験者間で共通するような一定した傾向は見られなかった。つまり、今回の実験では情動刺激の違いによって被験者に共通する変化の傾向は見られなかったものの、何かしらの影響を与えうることが分かった。

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、対話の直前に怒り表情を提示する条件と喜び表情を提示する条件とでは、ほとんどの被験者において交替潜時長やボタン押し動作時間の長さなどの時間的特徴量に差が見られる事を示した。また、怒り表情を提示した場合と喜び表情を提示した場合の違いによって、時間的特徴量間 (指示発話長-交替潜時、交替潜時-動作と発話開始の時間差 etc.) の相関係数が変化していることも示した。このことは、人の表情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を意味していると考えられる。従って本研究により、人間と自然な対話を行う対話インターフェースの構築のためには、情動状態の違いに応じて変化する発話と身振りタイミングモデルが必要になってくる可能性が示唆された。

しかし本研究には今後考慮しなければならないいくつかの問題点がある。

まず、タイミングの変化に被験者間で共通した傾向が見られていない点が挙げられる。この原因としては、データ数の不足、または個人差の大きさが考えられ、今後被験者数を増やしてデータを増やしていく事で、タイミングの変化に被験者間で傾向があるかを調査する必要があると考えている。

次に、情動状態の定量化ができないのでこのままだとモデル化ができない点が挙げられる。問題となる理由は、感情評定アンケートの結果で低い評点を与えているにも関わらず、特徴量には情動刺激条件間で差が出ている被験者 D のケースもあることから、主観的アンケートの結果からモデルを構築することは困難であると考えられる。また、実際に発話をしている瞬間の人の情動状態をリアルタイムで計測することは現在の技術ではほぼ不可能に近い。

そこで今後の方針としては、IAPS を情動喚起刺激に用いることを考えている。IAPS とは International Affective Picture System[15] の略称であり、700 枚以上ものスライドから構成される感情絵データベースのことである。これらは感情価 (pleasure), 覚醒 (arousal), 支配性 (dominance) の 3 次元から各スライドを評価されている。また、表情刺激では提示された表情と同じ感情になるように意識してもらう教示が必要であるが、IAPS の場合は被験者にただ見てもらうだけで情動が喚起されることなどから、情動状態に応じた発話と身振りのタイミングモデルの構築のための情動喚起刺激として適していると考えられる。

このような方針のもと、今後さらなる研究を進める予定である。

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] 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).

[4] Matarazzo, J.D., Weitman, M., Saslow, G., Wiens,A.N.: Interviewer in?uence on durations of interviewee speech; Journal of Verbal Learning and Verbal Behavior, vol.1, pp.451-458, (1963).

[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] 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).

[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] 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, 國立, 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).

[15] Lang, P.J., Bradley,M.M., Cluthbert, B.N.: The internationalAffective Picture System(IAPS); Photographic Slides The center for Research in Psychophysiology, University of Florida.(1995).