

# 対話における顔向けを伴う働きかけが 同意・不同意応答のタイミングに及ぼす影響

平山 高嗣<sup>\*1</sup> 大西 哲朗<sup>\*1</sup> 朴 惠宣<sup>\*1\*2</sup> 松山 隆司<sup>\*1</sup>

Effects of Approach with Face-turning Action on  
Timing of Agreement/Disagreement Response in Human-human Interaction

Takatsugu Hirayama,<sup>\*1</sup> Tetsuro Onishi,<sup>\*1</sup> Hye Sun Park<sup>\*1\*2</sup> and Takashi Matsuyama<sup>\*1</sup>

**Abstract** – When talking with someone, we convey intention to each other by verbal and non-verbal behaviors. In consensus building dialogue, the participants need to understand whether they agree or disagree. They reiterate approach to partner, such as statement of opinion, and response for it. In this study, we considered that the timing of listener's response for the approach by a speaker reflects listener's internal state (agreement/disagreement), therefore analyzed the multimodal timing structures between the approach and the response by utterance and body motion. Especially, we focused on an action that the speaker turns his face toward the listener when he approaches the listener and analyzed how it influences the timing structures. As the results, we confirmed that the timing structures relate to the internal state and the relations are controlled by *face-turning* action.

**Keywords** : *face-turning*, agreement/disagreement response, multimodal timing

## 1. はじめに

近年、ヒューマンマシンインタラクションに基づくユーザ支援システムの開発が進められている。我々は、現在、2者による旅行プランニングや商品選択といった合意形成対話を支援するシステムをデザインしている。ユーザにとって気の利いた情報提供や望ましい商品紹介を行うためには、ユーザの振る舞いから、意図や興味などの心的状態を推定する必要がある。人は合意形成対話において、基本的には言語を発話することによってやりとりを進める。言語は顕在化された情報であり、意識的に制御される。そのため、言語にはしばしば虚偽が含まれる。つまり、言語から心的状態を正確に推定することは難しい。そこで、本研究では、人が表出する非言語的な振る舞いに注目する。人は心的状態の表出を完全には制御できず、非言語的な振る舞いによって無意識的に表出してしまう<sup>[1]</sup>。

対話に表れる非言語的な振る舞いとしては、発話の韻律や表情、ジェスチャなどの身体動作がある。また、対話上の時間的特徴は非言語的な振る舞いが持つ特徴の一つとして重要視されている。従来研究において、対話の話者交替時の発話のタイミング構造が、発話の韻律などの非言語情報以上に話者の心的状態と関係が強い<sup>[2]~[4]</sup>ことが言及されている。しかし、これらの

研究は、発話のタイミングだけに注目し、対面対話で重要な役割を果たす身体動作などの非言語的な振る舞いがタイミング構造に与える影響を無視している。従来、話者ごとの発話と身体動作のタイミング構造については研究されている<sup>[5]</sup>が、複数人の間のマルチモーダルタイミング構造と、それと心的状態の関係についてはほとんど分析されていない。複数人対話ではないが、落語の役柄交替時のマルチモーダルタイミング構造を分析した研究では、肯定的回答と同意応答の身体動作の発生タイミングが先行発話にオーバーラップする頻度が、否定的回答と不同意応答より高いことを明らかにしている<sup>[6]</sup>。この現象は、観客への演出効果を狙った恣意的なものであると考えられるが、対話者の心的状態の表現手段としてマルチモーダルタイミング構造の重要性を示している。そこで、本研究では、合意形成対話における2者間の発話と身体動作に基づくマルチモーダルタイミング構造と心的状態の関係について分析する。

本研究では心的状態として、対話相手の意図に対する同意/不同意状態に注目する。これは合意形成対話において最も根源的な状態である。合意形成対話では自分と相手の間の同意/不同意状態を綿密に理解する必要がある。そのために、意見の主張や相手の意図の確認のための働きかけとそれに対する応答が頻繁に起こる。それらの行為の中で、我々は、働きかけを行う時に起こる相手への顔向けに注目する。顔向けは対話の中でどんな意味を持つのか？人は相手に視線を向け

\*1: 京都大学大学院 情報学研究科

\*2: 日本学術振興会

\*1: Graduate School of Informatics, Kyoto University

\*2: Japan Society for the Promotion of Science

ることで自分の行動に対するフィードバックを求める傾向がある [7],[8]。つまり、顔向けは反応を促す、時には反応を強要するため、働きかけの行為が含む意図確認の役割を明確化する信号と言える。また、視線を向ける行為は身体動作を確認する行為でもある。そのため、心的状態を反映した反応が身体動作に表れると考えられる。応答者は、顔向けを伴う働きかけに対して、発話と身体動作に基づく応答のタイミングに心的状態を強く反映させる可能性がある。顔向けは、対話相手の心的状態を顕在化する役割を持つと考えられる。

本論文では、「働きかけと応答の発話および身体動作のタイミング構造と、その応答の意図（同意／不同意という心的状態）の間には、強い関連性がある」という仮説を立て、合意形成対話の映像と音声の分析を通じて、その妥当性を検証する。特に、相手への顔向けが対話のタイミング構造に及ぼす影響に注目し、顔向けが対話の中で持つ意味について考える。また、発話と身体動作の発生タイミングを基本特徴とした同意／不同意状態の自動識別の可能性について検討する。

以下、次章では働きかけと応答の発話および身体動作について定義し、応答者の同意／不同意という心的状態と発話および身体動作のタイミング構造の関連性について仮説を述べる。3章では、仮説検証のための実験手法と、発話と身体動作の区間抽出手法について述べる。4章では、実験結果を分析し、仮説を検証する。そして、5章で顔向けの意味について考察し、6章では情報システムによる同意／不同意状態の自動識別の可能性について検討する。7章は結論である。

## 2. 働きかけに対する応答のタイミングと応答者の心的状態の関係

### 2.1 合意形成対話の構成要素

合意形成対話は、課題遂行型対話の一種であり、複数の言語的なやりとりから構成される。本研究では、合意形成対話の最も原初的な要素である働きかけとそれに対する応答で構成される1つのやりとりに注目する。働きかけとは、新しいやりとりを開始する機能を持ち、自己の心的状態を表出し、それに対する相手の同意／不同意を確認する行為と定義する。応答は、働きかけに対する反応の機能を持ち、聞き手が自己の内部状態に従って返答する行為と定義する。図1にやりとりの例を示す。

様々な身体動作が働きかけと応答の発話に伴って生じる。これは言語情報を強調したり、補完する非言語情報である。本研究では、頭部の回転による身体動作に注目する。働きかけの身体動作としては、相手への顔向けが挙げられる。これは心的状態を反映する振る舞いを観察する行為である。また、応答の身体動作と

しては、頷き、かしげ、首振りが挙げられる。これらは同意／不同意を表現することができる。

### 2.2 発話と身体動作による働きかけと応答のタイミング構造

話し手  $A$  が聞き手  $B$  に対して働きかけを行い、聞き手  $B$  がそれに対して応答を行う状況を考える。2者それぞれが発話と身体動作を行うため、2者間に以下の4つのタイミング構造  $I_{UU}$ ,  $I_{MU}$ ,  $I_{UM}$ ,  $I_{MM}$  が存在する。これらを図2に示す。発話と身体動作によるイベント  $E$  の発生時刻を  $T(E)$  と記述する。対話者  $X$  が発生したイベントを  $E^X$ 、イベントの開始、終了をそれぞれ  $E_s$ ,  $E_f$  と記述する。働きかけの発話を  $U(apr)$ 、応答の発話を  $U(res)$ 、働きかけの顔向けを  $M(apr)$ 、応答の身体動作を  $M(res)$  と記述する。

応答の発話に関するタイミング構造:  $I_{UU}$ ,  $I_{MU}$

話し手  $A$  による働きかけの発話に対する聞き手  $B$  による応答の発話のタイミング構造  $I_{UU}$  を次のように定義する。

$$I_{UU} = T(U(res)_s^B) - T(U(apr)_e^A) \quad (1)$$

発話意図は、特に日本語の場合、発話の終端付近で明らかになることが多い。そのため、働きかけの発話終了時刻が最も応答責任が生じる時刻であると考えられることから、それを  $I_{UU}$  の基準時刻とする。一方、応答発話のタイミングはその開始時刻に注目する。

話し手  $A$  による働きかけの顔向けに対する聞き手  $B$  による応答のタイミング構造  $I_{MU}$  を次のように定義する。

$$I_{MU} = T(U(res)_s^B) - T(M(apr)_s^A) \quad (2)$$

働きかけの顔向けは、その行為の開始時点で聞き手にそれが視認される。顔向けの開始時刻から応答責任が生じると考えられることから、顔向け開始時刻を  $I_{MU}$  の基準時刻とする。ただし、働きかけの顔向けは、働きかけの発話開始時刻より後に発生し、応答の発話開始時刻より前に発生するものと定義する。すなわち、話し手  $A$  が発生した顔向け  $M^*$  について、

$$T(M_s^*) \geq T(U(apr)_s) \quad (3)$$

$$T(M_s^*) \leq T(U(res)_s) \quad (4)$$

$$T(M_s^*) \leq T(M(res)_s) \quad (5)$$

を全て満たすものを、働きかけの顔向け  $M(apr)$  とみなす。ただし、これらの条件を満たす顔向けが複数ある場合は、最も遅く発生したものを働きかけの顔向けとする。また、式4の条件より、 $I_{MU}$  は常に正の値を取る。

話し手 A : この2つから選ぶなら, こっちがいいなあ. どう思う? 「働きかけ」  
 聞き手 B : うん, それがいいと思うよ 「応答」

図 1 働きかけと応答で構成されるやりとりの例

Fig. 1 An example of exchange composed of approach and response.

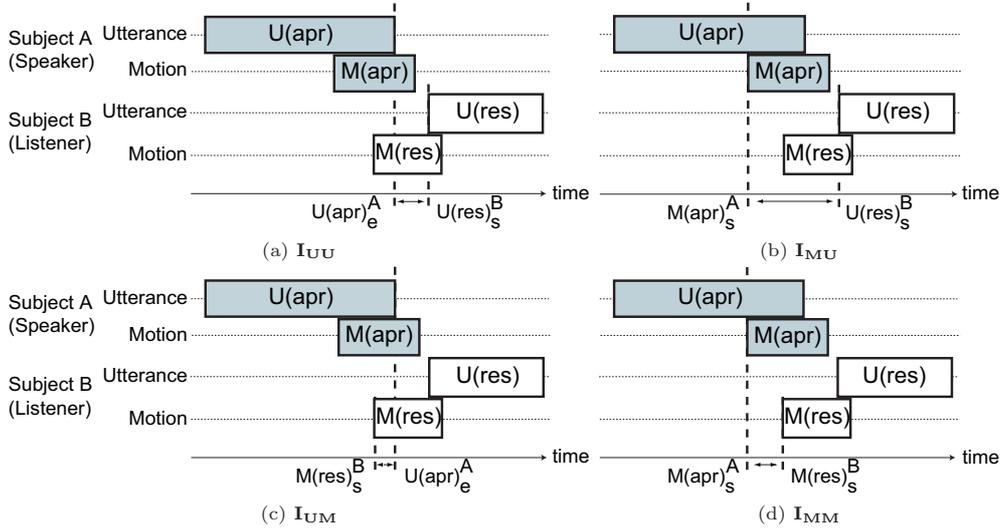


図 2 タイミング構造

Fig. 2 Timing structures.

応答の身体動作に関するタイミング構造:  $I_{UM}$ ,  $I_{MM}$   
 話し手 A による働きかけの発話に対する聞き手 B による応答の身体動作のタイミング構造  $I_{UM}$  を次のように定義する.

$$I_{UM} = T(M(res)_s^B) - T(U(apr)_e^A) \quad (6)$$

$I_{UM}$  は,  $I_{UU}$  と同様に, 働きかけの発話終了時刻を基準時刻とする. 応答の身体動作のタイミングは, その開始時刻に注目する. ただし, 応答の身体動作は, 働きかけの発話開始時刻から応答の発話終了時刻までに発生し, 働きかけの発話終了時刻より少し早い時刻以降に発生するもの, または応答の発話開始時刻より少し早い時刻以降に発生するものと定義する. すなわち, 応答側が発生した身体動作  $M^*$  について,

$$T(U(apr)_e) \leq T(M_s^*) \leq T(U(res)_e) \quad (7)$$

を満たし, かつ,

$$T(U(apr)_e) - T(M_s^*) \leq 500\text{msec} \quad (8)$$

$$T(U(res)_e) - T(M_s^*) \leq 500\text{msec} \quad (9)$$

のいずれかを満たすものを, 応答の身体動作  $M(res)$  とみなす. ただし, 以上の条件を満たす身体動作が複数ある場合は, 最も早く発生したものを応答の身体動作とする.

話し手 A による働きかけの顔向けに対する聞き手 B による応答の身体動作のタイミング構造  $I_{MM}$  を次

のように定義する.

$$I_{MM} = T(M(res)_s^B) - T(M(apr)_s^A) \quad (10)$$

$I_{MM}$  は,  $I_{MU}$  同様, 働きかけの顔向け開始時刻を基準時刻とする. また, 式 5 の条件より,  $I_{MM}$  は常に正の値を取る.

### 2.3 タイミング構造と応答者の心的状態の関係

課題遂行型対話において,  $I_{UU}$  が, 後続話者の発話が肯定的な場合より否定的な場合に長く調整されるという知見<sup>[2]</sup>がある. 合意形成対話において, 応答者は  $I_{UU}$  および  $I_{MU}$  を調整することで, 同意/不同意という心的状態を表現すると考えられる. そこで, 本研究の1つ目の仮説を挙げる.

#### 仮説 1: 心的状態に応じた $I_{UU}$ , $I_{MU}$ の調整

$I_{UU}$ ,  $I_{MU}$  は, 話し手の意図に対して応答者が同意する場合より, 不同意を表す場合に長く調整される.

対話者は基本的には言語によって情報伝達を行うが, 顔向けなどの身体動作が対話に一切影響を及ぼさないことを仮定できる状況は少ない<sup>1</sup>. そのため, 身体動作は, 対話者の心的状態の表出に何らかの影響を及ぼすと考えられる. 話し手による聞き手への顔向けは,

1: 身体動作が対話に影響を及ぼさない状況の一例として, 電話が挙げられる.

聞き手が表出する振る舞いを観察する行為であり、明確な応答を促し、時には応答を強要する行為である。そのため、聞き手は仮説1より  $I_{UU}$ ,  $I_{MU}$  を明確に調整することで心的状態を表現すると考えられる。そこで、本研究の2つ目の仮説を挙げる。

仮説2: 顔向けによる  $I_{UU}$ ,  $I_{MU}$  の変化

話し手が聞き手に顔向けを行うことで、応答者が同意を表す場合の応答タイミングと不同意のその差が、仮説1より大きくなる。

応答の発話と同様に、応答者の心的状態に従って、応答の身体動作のタイミング構造  $I_{UM}$ ,  $I_{MM}$  が調整されると考えられる。そこで、本研究の3つ目の仮説を挙げる。

仮説3: 心的状態に応じた  $I_{UM}$ ,  $I_{MM}$  の調整

$I_{UM}$ ,  $I_{MM}$  は、話し手の意図に対して応答者が同意する場合より、不同意を表す場合に長く調整される。また、同意と不同意のタイミング構造の差は、話し手が聞き手に顔向けを行うことでさらに大きくなる。

以下の章では、これらの仮説を検証するために、合意形成対話を観察し、話し手が働きかけのために

- (s1) 発話のみ行い、聞き手が同意
- (s2) 発話のみ行い、聞き手が不同意
- (s3) 発話と顔向けを行い、聞き手が同意
- (s4) 発話と顔向けを行い、聞き手が不同意

である場合について、応答者がタイミング構造  $I_{UU}$ ,  $I_{MU}$ ,  $I_{UM}$ ,  $I_{MM}$  をどのように調節するかを分析する。

3. 合意形成対話のコーパス構築

タイミング構造の分析のために、合意形成対話における働きかけと応答を記録したコーパスが必要となる。そこで、まず、観測する合意形成対話の設定について述べる。次に、観測データから発話と身体動作の区間を抽出する手法について述べる。

3.1 合意形成対話の観測

対話状況設定

合意形成課題は、2択クイズとした。選択肢が被験者にとって既知であり、解答が未知であると思われる問題を12問用意した。それらを表1に示す。被験者間の関係は立場が対等で友人関係にある同性とした。

対話の空間的な設定を図3に示す。本研究では、相手への顔向けの表出を明確にするために、横並びでの対話を分析した。被験者の対面側には、21インチディ

表1 合意形成課題  
Table 1 Consensus building tasks.

No.	課題内容
#1	売上上位の即席麺はどちらか
#2	原価が高い百円均一店の商品はどちらか
#3	売上上位の駄菓子品はどちらか
#4	売上上位の缶ビールはどちらか
#5	納税額が多いのはどちらの人物か
#6	新生児の名前として多いのはどちらか
#7	売上上位のおでんの具はどちらか
#8	おみくじの上位はどちらか
#9	外国人旅行者が多いのはどちらの県か
#10	人気が高いアニメはどちらか
#11	売上上位の雑誌はどちらか
#12	原価が高い百円均一寿司のネタはどちらか

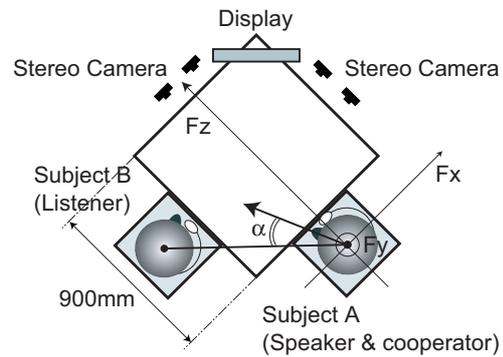


図3 対話空間の設定  
Fig.3 Experimental environment.

スプレイを配置し、2択クイズを表示した。これは、接客カウンタのような状況を想定している。また、身体動作として特に頭部動作に注目しているため、他の身体動作の発生を抑えるために、座位での対話とした。

実験協力者の導入

2.3節で述べた(s1)~(s4)の4つの実験条件を均一に分析するために、2名の被験者のうち1名に実験協力者を導入し(もう一方の被験者を真の被験者と呼ぶことにする)、実験条件を統制した。まず、対話が開始される直前に、それぞれの被験者に2択クイズを提示し、事前調査として個別に回答を得た。そして、真の被験者の回答に応じて、実験協力者が支持する選択肢を実験者が指示した。さらに、実験協力者に働きかけの顔向けの表出の有無についても指示した。これらの指示は真の被験者に察知されないように行われた。

また、実験協力者の振る舞いも統制した。まず、働きかけの発話について、常に「この2つなら、     が     」という内容になるように教示した。     は選択肢のどちらかを指す単語である。この統制は、働きかけの発話内容がタイミング構造に影響を及ぼすことを回避するためである。また、発話時の話速や韻律が、各実験において一定となるように教示した。さらに、働きかけの発話

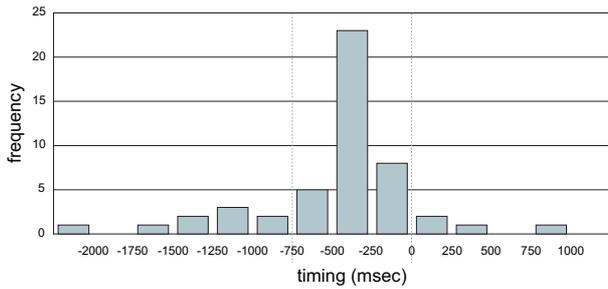


図4 動きかけの発話と顔向けとのタイミング  $I_{apr}$  の分布

Fig. 4 Distribution of  $I_{apr}$ .

終了時刻  $T(U(apr)_e)$  と顔向け開始時刻  $T(M(apr)_s)$  の間のタイミングが一定となるように統制した。そのタイミング  $I_{apr}(= T(M(apr)_s) - T(U(apr)_e))$  の分布を予備実験によって調べたところ、ほとんどの顔向けが発話にオーバーラップして発生することが分かった。  $I_{apr}$  の分布を図4に示す。そこで、発話にオーバーラップさせて顔向けを行うように実験協力者に教示し、観測データの中で  $-750\text{msec} \leq I_{apr} \leq 0\text{msec}$  の条件を満たした動きかけのみを分析対象とした。

これらの統制に従った動きかけが、同一対話内で複数回に渡って行われるのは不自然である。そこで、各課題の対話の先行発話者をルーレットで選出するというルールを設け、先行発話者に実験協力者が選出された時に、上記の統制に従った動きかけを最初の発話時のみに行わせた。ただし、ルーレットは実験者が意図的に操作できるように仕組んだ。表1の #2, 3, 5, 6, 7, 9, 10, 12 の課題で実験協力者を先行発話者として選んだ。そして、各課題の最初のやりとり（動きかけとそれに対する応答）だけを分析対象とした。このやりとりにおける真の被験者の心的状態は、課題遂行直前の事前調査によって確認できている。先行発話者に実験協力者が指名されなかった場合の対話は分析対象外とした。

#### 対話観測

実験協力者は4名、真の被験者は13名で、合計13組について実験を行った。各組について、(s1) ~ (s4) の4つの実験条件それぞれにつき2つずつ表1の課題を割り当てた。すなわち、104対話を収録した。

図3に示すように、発話の計測機器として、被験者にヘッドセットマイクロフォンを装着させた<sup>2</sup>。身体動作の計測機器としては、各被験者の対面側に頭部を撮影するステレオカメラを設置した。撮影された画像例を図5に示す<sup>3</sup>。



図5 撮影画像例

Fig. 5 Examples of captured images.

### 3.2 発話と身体動作の区間抽出

#### 発話区間の抽出

計測した音声に対し、音声分析ツール“Wavesurfer”<sup>4</sup>を使用し、笑い声や咳払いなどの非言語的な音声区間を削除し、発話区間を抽出する。ただし、400msec以上の無音区間は、発話間のポーズとみなす。そして、区間抽出された発話イベントに対して、動きかけと応答のラベルを付与する。

#### 身体動作区間の抽出

身体動作として観測する相手への顔向け、頷き、かしげ、首振りの頭部動作の区間を頭部回転量と並進量に基づいて抽出する。そのために、図3のように被験者座標系を定義する。被験者座標系は、被験者ごとに平均的な頭部位置を原点とし、原点からステレオカメラの基線中心への水平な直線を  $F_z$  軸、鉛直方向への直線を  $F_y$  軸、 $F_z$  軸と  $F_y$  軸に直交する直線を  $F_x$  軸と設定したものである。被験者  $A$  から  $B$  への顔向けは、 $F_y$  軸まわりの頭部回転方向へのベクトルと、被験者  $A$  と  $B$  の頭部位置を結ぶベクトルのなす角  $\alpha$  から抽出する。 $\alpha$  の時間変化の一次微分の極小値の絶対値が閾値以上となるときを顔向け開始時刻、極大値の絶対値が閾値以上となるときを顔向け終了時刻と定義する。また、被験者  $B$  の頷き、かしげ、首振り、 $F_x$  軸、 $F_z$  軸、 $F_y$  軸まわりの頭部回転量の時間変化の一次微分の絶対値が閾値以上となるときをそれぞれの身体動作の開始時刻または終了時刻と定義する。

これらの頭部動作量は、被験者の顔特徴点の3次元位置を計測し、あらかじめ計測しておいた顔特徴点の3次元モデルからの回転量および並進量を推定することで得る。顔モデルは、被験者座標系の原点に  $F_z$  軸を正面方向として配置する。まず、各カメラで撮影された顔画像にAAM (Active Appearance Model)<sup>[9]</sup>を適用し、顔特徴点を抽出する。図6に顔特徴点の抽出結果の一例を示す。次に、一組のステレオ画像の顔特徴点位置にステレオ法<sup>[10]</sup>を適用し、顔特徴点の3次元位置を計測する。そして、計測した顔特徴点の3次元位置が顔モデルをどの程度回転、並進したものを

2: 音声のサンプリングレートは 44.1kHz とした。  
3: 画像は 1024 × 768pixel, 30fps で撮影した。

4: <http://www.speech.kth.se/wavesurfer/>

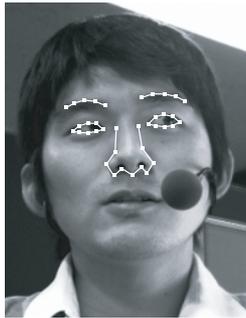


図 6 AAM によって抽出された顔特徴点  
Fig.6 Facial feature points extracted by AAM.

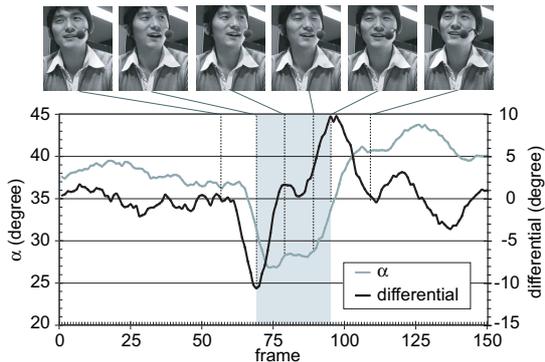


図 7 顔向け区間の抽出結果  
Fig.7 Extracted face-turning interval.

求める．計測した顔特徴点  $i$  の 3 次元位置からなるベクトルを  $Q'$ ，顔モデルの特徴点  $i$  の 3 次元位置からなるベクトルを  $Q$  とする．また， $3 \times 3$  の回転行列を  $R$ ， $3 \times 1$  の並進行列を  $T$  とする．回転行列および並進行列の推定は，

$$\Sigma^2 = \sum_{i=1}^n \| Q' - (RQ + T) \|^2 \quad (11)$$

を最小化する問題である．本研究では，式 11 の解法として，特異値分解を用いたアルゴリズムを採用する<sup>[11]</sup>．この解法では， $QQ'^T$  を特異値分解して

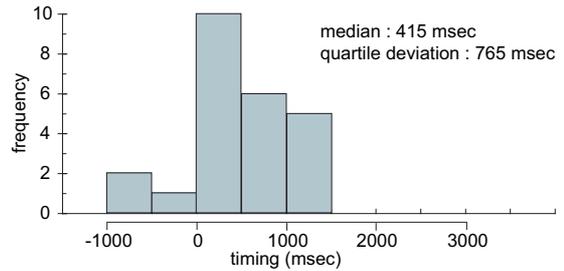
$$QQ'^T = USV^T \quad (12)$$

$$R = VU^T \quad (13)$$

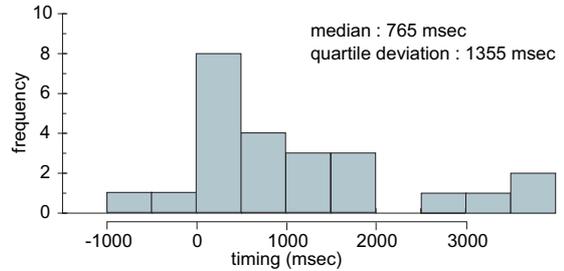
として回転行列  $R$  が求められる．また，計測した顔特徴点の 3 次元位置の重心と顔モデルの特徴点の重心の差から，並進行列  $T$  が求められる． $\alpha$  は  $R$  の  $F_y$  成分と  $T$  から求められる．図 7 に  $\alpha$  の時間変化とその微分値と，抽出された顔向け区間の例を示す．区間抽出された身体動作イベントに対して，働きかけと応答のラベルが付与される．

#### 計測結果の手動補正

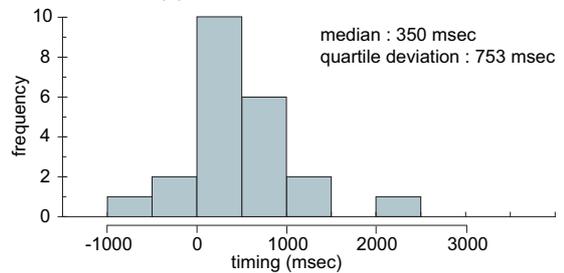
計測した発話と身体動作の開始時刻および終了時刻



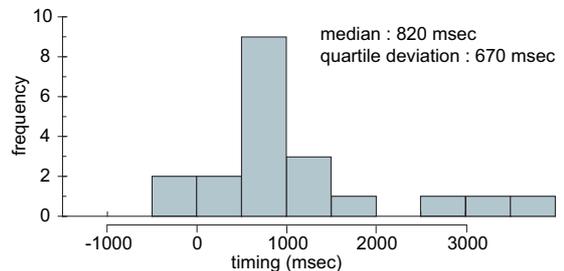
(a) 顔向け無し，同意の場合



(b) 顔向け無し，不同意の場合



(c) 顔向け有り，同意の場合



(d) 顔向け有り，不同意の場合

図 8  $I_{UU}$  の頻度分布  
Fig.8 Histograms of  $I_{UU}$ .

が誤差を含むため，アノテーションツール Anvil<sup>[12]</sup> を用いて，手動補正を行った．

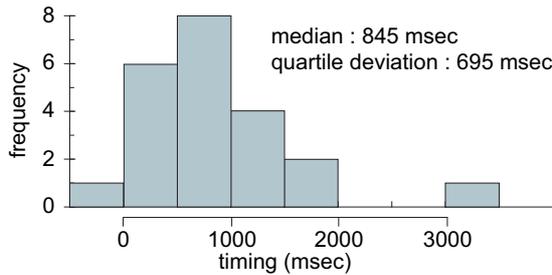
#### 4. タイミング構造と心的状態の関係の分析と仮説の検証

##### 4.1 応答の発話に関するタイミング構造分析

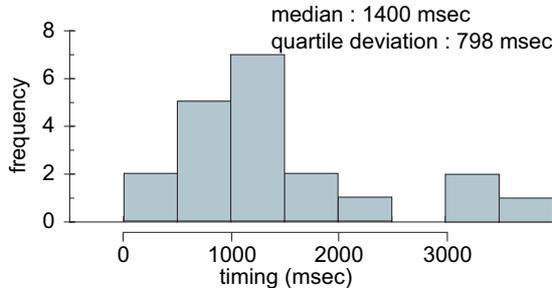
まず，応答の発話に関するタイミング構造  $I_{UU}$ ， $I_{MU}$  を分析する．応答の発話は，実験条件に関係なく，全ての働きかけに対して発生し，働きかけの顔向けが無い場合の同意応答が 24 回，不同意応答が 24 回，顔向けが有る場合の同意応答が 22 回，不同意応答が 20 回であった．

#### $I_{UU}$ の分布

コーパスから得られた  $I_{UU}$  の分布を図 8 に示す．



(a) 顔向け有り, 同意の場合



(b) 顔向け有り, 不同意の場合

図9 I<sub>MU</sub> の頻度分布  
Fig.9 Histograms of I<sub>MU</sub>.

これらの分布から抽出される統計量としては、外れ値の影響を抑えるために中央値と四分位偏差に注目する。働きかけの顔向けが無い場合は、同意の応答タイミングの中央値が 415msec、四分位偏差が 765msec、不同意の応答タイミングの中央値が 765msec、四分位偏差が 1355msec であった。また、いずれの場合も、0 ~ 500msec の範囲で最も頻度が高い。

一方、顔向けが有る場合は、同意の応答タイミングの中央値が 350msec、四分位偏差が 753msec、不同意の応答タイミングの中央値が 820msec、四分位偏差が 670msec であった。また、同意の場合に 0 ~ 500msec の範囲で最も頻度が高くなることに対し、不同意の場合は 500 ~ 1000msec の範囲で最も頻度が高くなった。

#### I<sub>MU</sub> の分布

I<sub>MU</sub> の分布を図9に示す。3.1節で述べたように、I<sub>apr</sub> が一定となるように分析データを統制したため、I<sub>MU</sub> は I<sub>UU</sub> と同様の分布と統計量を持つ。

#### 仮説1の検証

働きかけの顔向けが無い場合の I<sub>UU</sub> について、不同意の応答タイミングの中央値は同意より大きくなった。これは仮説1を支持する結果である。ただし、中央値検定の結果、これらの中央値に有意な差は認められなかった (有意確率  $p = 0.56$ )。

#### 仮説2の検証

働きかけの顔向けが有る場合は、無い場合と同様に不同意の応答タイミング I<sub>UU</sub> の中央値は同意より

表2 応答の身体動作 M(res) の発生頻度。  
Table 2 Occurrence frequency of M(res).

実験条件	M(res) 発生回数 / U(apr) 発生回数
(s1) 顔向け無し, 同意	14 / 24 (58.3%)
(s2) 顔向け無し, 不同意	20 / 24 (83.3%)
(s3) 顔向け有り, 同意	18 / 22 (81.8%)
(s4) 顔向け有り, 不同意	9 / 20 (45.0%)

大きくなった。中央値検定の結果、これらの中央値に有意な差が認められた ( $p = 0.013$ )。顔向けが無い場合には有意な差が認められなかったことから、顔向けによって同意と不同意の応答タイミングの差が顕著になったと言える。これは仮説2を支持する。

#### 4.2 応答の身体動作に関するタイミング構造分析

応答の身体動作に関するタイミング構造 I<sub>UM</sub>, I<sub>MM</sub> を分析する。応答の身体動作は、実験条件によって発生する頻度が変化した。その頻度を表2に示す。

#### I<sub>UM</sub> の分布

I<sub>UM</sub> の分布を図10に示す。同意の応答タイミングは 0 ~ 2000msec の範囲で分布の頻度が高くなり、不同意の応答タイミングは 0 ~ 1000msec の範囲と、2500msec 以上の範囲で分布の頻度が高くなった。

働きかけの顔向けが無い場合は、同意の応答タイミングの中央値が 820msec、四分位偏差が 720msec、不同意の応答タイミングの中央値が 770msec、四分位偏差が 2925msec であった。また、顔向けが有る場合は、同意の応答タイミングの中央値が 295msec、四分位偏差が 573msec、不同意の応答タイミングの中央値が 2690msec、四分位偏差が 2060msec であった。

#### I<sub>MM</sub> の分布

I<sub>MM</sub> の分布を図11に示す。3.1節で述べたように、I<sub>apr</sub> が一定となるように分析データを統制したため、I<sub>MM</sub> は I<sub>UM</sub> と同様の分布と統計量を持つ。

#### 仮説3の検証

I<sub>UM</sub> と I<sub>MM</sub> は I<sub>UU</sub> と I<sub>MU</sub> と同様の結果となった。しかし、応答者が必ずしも身体動作を発生させるとは限らないため、I<sub>UM</sub> と I<sub>MM</sub> のサンプル数が表2のように少なく、不均一となった。そのため、有意差の検定に基づく仮説の検証は困難である。

表2に示す応答の身体動作の発生頻度について、応答時の身体動作の発生率は、

- \* 働きかけの顔向けが無い場合は、応答者が不同意の場合に同意の場合より高い
- \* 働きかけの顔向けが有る場合は、応答者が同意の

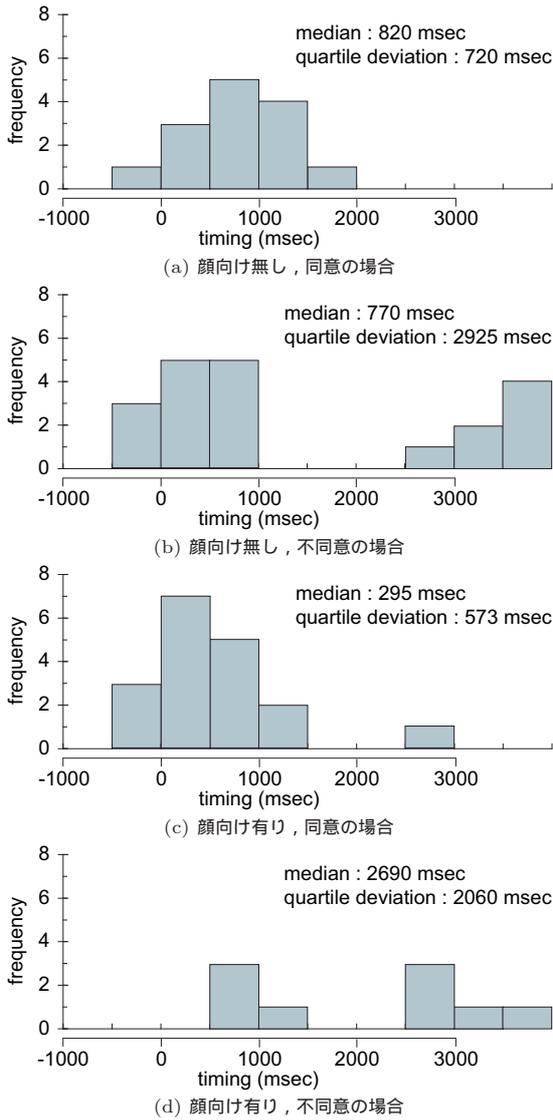


図 10  $I_{UM}$  の頻度分布  
Fig. 10 Histograms of  $I_{UM}$ .

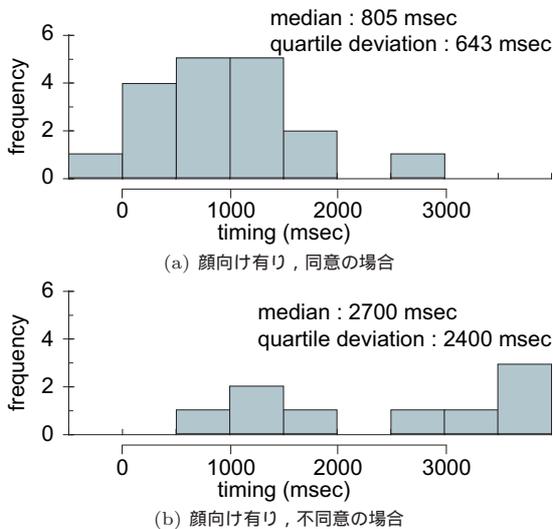


図 11  $I_{MM}$  の頻度分布  
Fig. 11 Histograms of  $I_{MM}$ .

場合に不同意の場合より高い

という傾向がある．この理由について，次章で考察する．

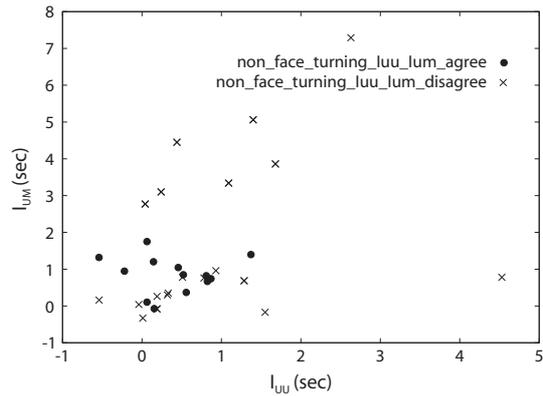
### 5. 考察-対話で発生する顔向けは何を意味するか?

まず，働きかけの顔向けが無い場合に，同意応答のタイミングが不同意応答より早くなった理由について考察する．藤原らの考察を引用すると [2]，応答者の応答タイミングが遅すぎると，解答に迷っていると働きかけ側に解釈されてしまう可能性がある．そのため，同意の応答タイミングは早くなると考えられる．一方，応答者が不同意を表す場合，合意形成対話では相手の意見を尊重して言葉を選んだり，熟慮する必要があるため，その応答タイミングは遅くなると考えられる．また，同意の場合に応答タイミングを早くし，不同意の場合にそれを遅くすることが，心的状態を表現する暗黙的なルールとなっている可能性もある．

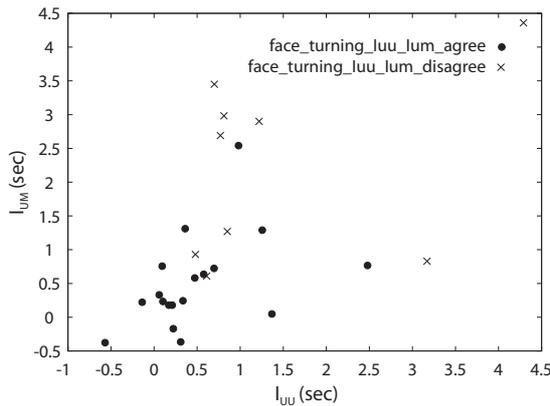
同意応答と不同意応答のタイミングの差は，働きかけのために顔向けを行うことによって大きくなった．顔向けは，聞き手の明確な応答を促し，時には応答を強要する行為である．これは，顔向けが有る場合の不同意応答のタイミングの四分位偏差（図 8 (d)）が，顔向けが無い場合（図 8 (b)）に比べて非常に小さくなっていることから確認できる．顔向けには，ある一定の時間区間内で発話による応答を促す効果がある．これらの結果より，働きかけの顔向けは応答のタイミングを制御すると考えられる．顔向けという明示的なイベントを発生させることで，対話者双方が対話のタイミング構造に対する感覚を無意識的に強めている可能性がある．

また，表 2 に示したように，顔向けが，応答者が同意を表現する場合に身体動作を促進し，不同意の場合に抑制するという結果を得た．働きかけの顔向けが無い場合は，応答者は身体動作を観察されるという圧力を受けないため，同意であることを伝える時には，身体動作を使って必要以上に表現する必要がなく，不同意であることを伝える時には，無意識的に身体動作を表出してしまうと考えられる．一方，顔向けが有る場合は，応答者が身体動作を表出しないことが一種の不同意表現となるため，同意であることを伝える時には，必要以上にうなずき，不同意であることを伝える時には，身体動作を表出しないことも多用すると考えられる．顔向けは身体動作という対話チャンネルを開き，対話に表れる応答の身体動作表出の意味（同意 / 不同意）を変化させる可能性がある．

話し手が働きかけの顔向けを行わない場合は，発話応答から同意 / 不同意を認識することになるが，同意応答と不同意応答のタイミングの差が小さいため，嘘



(a) 顔向け無しの場合の  $I_{UU}$  と  $I_{UM}$  の分布



(b) 顔向け有りの場合の  $I_{UU}$  と  $I_{UM}$  の分布

図 12  $I_{UU}$  と  $I_{UM}$  の 2 次元分布  
Fig. 12 2D distribution of  $I_{UU}$  and  $I_{UM}$ .

を含む発話応答，すなわち，偽りの同意 / 不同意を見破ることが難しい．顔向けを行うことによって，聞き手の身体動作による応答を観察することができ，応答のタイミングから同意 / 不同意を認識しやすくなる．聞き手の心的状態を理解できなかった場合や，より確実に理解したい場合に，相手に顔を向けて働きかけることは我々は普段のコミュニケーションで行っていることである．顔向けは，相手の心をプローブする行為の一つであると言える．

## 6. タイミング構造に基づく同意と不同意の識別

対話者間の同意 / 不同意状態を応答タイミングに基づいて情報システムが自動識別することが可能であるか検討する．まず，図 8 の  $I_{UU}$  の分布について，同意と不同意の識別境界の閾値を設定し，それぞれの識別成功率を求める．両識別率が均衡する閾値を導出し，その識別率を精度評価の指標とする．働きかけの顔向けが無い場合は，閾値が 500msec の場合に同意の識別率が 54.2%，不同意が 58.3% となった．顔向けが有る場合は，閾値が 700msec の場合に同意が 72.7%，不同意が 70.0% となった．働きかけの顔向けが行われる場合に識別率が大きく向上することが分かる．

識別は特徴ベクトルの次元を上げることによって精

度が向上する可能性がある．そこで，複数のタイミング構造を軸とする多次元の分布での識別を考える． $I_{MU}$  と  $I_{MM}$  は，3.1 節で述べた  $I_{apr}$  の統制によって，それぞれ  $I_{UU}$  と  $I_{UM}$  とほぼ同じ分布を持つため冗長な特徴であると考え， $I_{UU}$  と  $I_{UM}$  を軸とする 2 次元の分布で識別を行う．図 12 にその分布を示す．身体動作を伴った応答の  $I_{UU}$  と  $I_{UM}$  の組だけが識別の対象データである．識別境界の学習とテストデータの識別は Support Vector Machine<sup>[13]</sup> によって行った．学習データとテストデータのサンプリングは交差検定法を適用した．その結果，識別成功率は顔向け無しの場合の同意が 94.3%，不同意が 33.3%，顔向け有りの場合の同意が 85.6%，不同意が 60.4% となった．同意に比べて不同意が分散した分布を持ち，2 つの分布に重なりがあるため，妥当な結果であると言える．2 次元の分布による識別も，働きかけの顔向けが行われる場合に精度が向上しているが，表 2 に示すように，顔向けに対して不同意の身体動作が発生しにくく，識別サンプルが少なくなったため，さらに実験を行って検証する必要がある．

情報システムが対話者間の同意 / 不同意を識別するために，言語情報を用いることは言語に含まれる虚偽を認識することができない限り不可能である．また，従来知見<sup>[2]</sup>に基づいて，発話応答タイミングだけに注目すると，本実験の結果からは高い精度で識別することが困難であると言える．しかし，本研究で注目する顔向けを伴う働きかけが行われたやりとりを同定すれば，そのやりとりの  $I_{UU}$  によって識別精度を向上させることができる．また，身体動作による応答のタイミングを識別特徴に加えることで識別特性が変化し，条件によっては 80% 以上の精度で識別できる可能性がある．

## 7. 結論

本論文では，働きかけと応答の発話および身体動作のタイミング構造と応答者が相手に対して持つ同意 / 不同意の心的状態が関係するという仮説を立てた．この仮説を検証するために，実験協力者を用いた 2 者の合意形成対話を分析した．特に，話し手が意見を主張したり，聞き手の意図を確認する時に発話と共起する顔向けに注目し，それがタイミング構造に与える影響を分析した．

その結果，働きかけの発話に対する応答の発話のタイミングは，応答者が相手に不同意を表す場合に同意を表す場合より遅くなることを確認した．また，働きかけの発話に付随して顔向けを行うことで，同意と不同意の応答のタイミングの差が有意に大きくなることを確認した．顔向けが対話者に対話のタイミング構造

を強く意識させるトリガになっていると考えられる。さらに、その顔向けが、応答者が同意を表現する場合に身体動作を促進し、不同意の場合に抑制するという結果を得た。これは、顔向けを行わない場合と逆の結果であった。我々はこれが顔向けによって開かれた身体動作という対話チャンネルを使った対話の特徴であると考えられる。

今後は、まず、コーパスを充実させ、本研究で得られた結果の信憑性を確認する必要がある。そして、本研究で得られた知見に基づいてユーザの心的状態を推定するヒューマンマシンインタラクションのシナリオをデザインしたいと考えている。

#### 謝辞

本研究の一部は、科学研究費補助金 18049046 の補助を受けて行った。

#### 参考文献

- [1] T. J. Jacobs: On Unconscious Communications and Covert Enactments: Some Reflections on Their Role in the Analytic Situation; *Psychoanalytic Inquiry*, Vol. 21, pp. 4-23 (2001).
- [2] N. Fujiwara, T. Itoh, and K. Araki: Analysis of Changes in Dialogue Rhythm due to Dialogue Acts in Task-oriented Dialogues; *Springer-Verlag Lecture Notes in Artificial Intelligence(LNAI)*, Vol. 4629, pp. 564-573 (2007).
- [3] M. Yoshida and Y. Miyake: Relationship between Utterance Dynamics and Pragmatics in the Conversation of Consensus Building Process; *Proc. of the 15th IEEE Int. Symposium on Robot and Human Interactive Communication (ROMAN2006)*, pp. 641-645 (2006).
- [4] 長岡千賀, 小森政嗣, Maria Raluca Draguna, 河瀬諭, 結城牧子, 片岡智嗣, 中村敏枝: 協調的対話における音声行動の2者間の一致-意見固持型対話と聞き入れ型対話の比較-; *ヒューマンインタフェースシンポジウム 2003*, pp. 167-170 (2003).
- [5] 坊農真弓, 片桐恭弘: 対面コミュニケーションにおける相互行為的視点-ジェスチャー・視線・発話の協調-; *社会言語科学*, Vol. 7, No. 2, pp. 3-13 (2005).
- [6] 川嶋宏彰, 西川猛司, 松山隆司: 落語の役柄交替における視覚的「間合い」の解析; *情報処理学会論文誌*, Vol. 48, No. 12, pp. 3715-3728 (2007).
- [7] M. Argyle and J. Dean: Eye Contact, Distance and Affiliation; *Sociometry*, Vol. 28, pp. 289-304 (1965).
- [8] A. Kendon: Some Function of Gaze-direction in Social Interaction; *Acta Psychologica*, Vol. 26, pp. 22-63 (1967).
- [9] T. F. Cootes, G. J. Edwards, and C. J. Taylor: Active Appearance Model; *Proc. European Conf. on Computer Vision*, Vol. 2, pp. 484-498 (1998).
- [10] R. Hartley and A. Zisserman: Multiple View Geometry in Computer Vision; *Cambridge University Press* (2004).
- [11] K. S. Arun, T. S. Huang, and S. D. Blostein: Least-squares Fitting of Two 3-d Point Sets; *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, pp. 698-701 (1987).
- [12] M. Kipp: Gesture Generation by Imitation - From Human Behavior to Computer Character Animation; *Boca Raton, Florida: Dissertation.com* (2004).

- [13] T. Joachims: Making Large-scale SVM Learning Practical. *Advances in Kernel Methods - Support Vector Learning*; MIT-Press(1999).

(2002年1月1日受付, 1月1日再受付)

#### 著者紹介

平山 高嗣 (正会員)



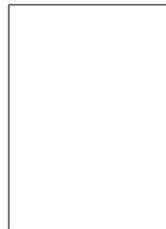
2000年金沢大学工学部電気情報工学科卒業。2002年大阪大学大学院基礎工学研究科修士課程修了。2005年同博士課程修了。2005年より京都大学大学院情報学研究科特任助手(現, 特任助教)。博士(工学)。顔画像認識, ヒューマンコンピュータインタラクションの研究に従事。電子情報通信学会会員。

大西 哲朗



2006年京都大学工学部電気電子工学科卒業。2008年同大学院情報学研究科修士課程修了。現在, 株式会社リクルート勤務。2006年情報処理学会 CVIM 卒論セッション最優秀賞。最近は, しぐさ等, 非言語的な情報が他者に与える印象に興味を持つ。

朴 恵宣



2001年嶺南大学校文科大学哲学科卒業。2004年慶北大学校電子電気コンピュータ大学院修士課程修了。2007年同博士課程修了。慶北大学校電子電気コンピュータ学部研究員を経て, 現在, 日本学術振興会外国人特別研究員, 京都大学情報学研究科外国人共同研究者。工博。ジェスチャ認識の研究に従事。

松山 隆司



1976年京大大学院修士課程修了。京大助手, 東北大助教授, 岡山大学教授を経て1995年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究科知能情報学専攻教授。2002年学術情報メディアセンター長, 京大評議員。2004年情報環境機構長。工博。画像理解, 分散協調視覚, 3次元ビデオの研究に従事。最近「人間と共生する情報システム」の実現に興味を持っている。1980年情報処理学会創立20周年記念論文賞, 1990年人工知能学会論文賞, 1993年情報処理学会論文賞, 1994年電子情報通信学会論文賞, 1995年第5回国際コンピュータビジョン会議 Marr Prize, 1999年電子情報通信学会論文賞, 2000年画像センシングシンポジウム優秀論文賞, 2004年, 2005年FIT優秀論文賞。IAPR, 情報処理学会, 電子情報通信学会フェロー。日本学術会議連携会員。