対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンシェルジェ

河原達也[†1] 川嶋宏彰[†1]
平山高嗣[†1] 松山隆司[†1]

1. 研究のねらいとコンセプト

膨大で多様な情報源にアクセス可能な現在においても，我々は漠然とした問題の解決や相談をする際に，専門家／詳しい人との直接対話に頼ることが多い．これは対話的コミュニケーションが，問題の所在や状況を把握したり，選好や制約条件を明確にした上で，適切な解を見いだすのに効果的であるためである．例えば，パソコン使用時のトラブルなどはマニュアルやヘルプ集を調べたりするより，詳しい人に聞く方がてっとり早い．また，旅行先で見所やレストランを探すときも，ガイドブックをいろいろ見るより,（高級ホテルに泊っていれば）コンシェルジェに聞いた方が気の利いた場所を教えてもらえる．このように，膨大な情報空間とユーザを効率的につなぐエージェントを実現するのが，我々のねらいである（図 1 参照）．

現在の情報システムは一般に，ユーザの明示的な指示に応じてシステムが反応し，情報を提示する**リアクティブ**なモデルに基づいている．現状の Web のサーチエンジンに代表される情報検索システムも，ユーザが指定したキーワード列にマッチした検索結果を提示し，ユーザがクリックしたり，キーワードをさらに追加するという流れになっている．これは，問題や意図（＝検索の目標）が明確で，キーワードの組合せで表現されることを前提とし，キーボードとポインティングデバイス (GUI) をインターフェースとしている．

これに対して我々が情報を欲する場合に，**いつもこのように検索の目標が明確で，キーワードで表現できるとは限らない**．例えば，京都の観光でも，数ある名所の中で具体的に行く場所や「○○が見たい」と決めている人は案外少ない．また，旅行先などで食事する場所を探す際も,「○○が食べたい」といった希望がなくても，ガイドブックを見たり，コンシェルジェに紹介されて，興味が喚起されて,「これにしようかな」と決定することが多い．

図 1 「情報爆発時代」を快適に暮らすためのコミュニケーション

我々は，**リアルタイムのインタラクションを介して，ユーザの興味を喚起し，意図や選好を顕在化しながら，情報を検索・提示する**システム，いわば「**情報コンシェルジェ**」の実現を目指している．そのために，複数のモダリティ・多様なインタフェース（ロボットを含む）を利用して，システム側からも積極的に気の利いた情報を気の利いたタイミングで提示する**プロアクティブ**なインタラクションの枠組みを研究している（図 2 参照）．

本稿ではまず，音声対話によって，大規模情報ベースにアクセスしながら情報検索・提示を行うシステム「京都版ダイアログナビ」を紹介する．このシステムは，観光地の紹介・案内のタスクにおいて，ユーザが

†1 京都大学 情報学研究科

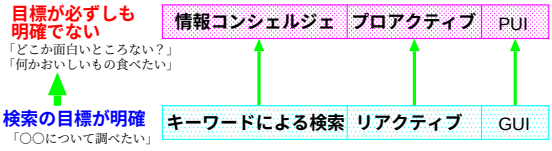


図 2 キーワード検索から情報コンシェルジェへ

興味を持ちそうな情報を質問形式で積極的に提示する枠組みを導入している．これは，言葉によるやりとりに基づく「情報コンシェルジェ」である．

ただし，真にプロアクティブなインタラクションを実現するには，ユーザの興味や心的状態を的確に察知（センシング）することが鍵となる．従来からも，ユーザの興味状態や非明示的な指示をマルチモーダルセンサ群を用いて推定しようとするパーセプチュアル・ユーザインターフェース (PUI) は研究されてきた．しかしながら，非明示的な指示や無意識レベルの心的状態をリアクティブに認識するのは（人間でも）容易でない．そこで我々は，システムも主導権を持ってインタラクションを行う**プロアクティブ**なモデルを導入し，システム側から提示された情報に対するユーザの反応を計測する “**Mind Probing**” という枠組みを提案しており，これについても紹介する．これは，画像による提示と視線滞留による興味推定を組み合わせた「情報コンシェルジェ」である．

プロアクティブな提示と興味推定は，人間どうしの対話では自然に行われている．例えば，販売員が商品を勧めたり，コンシェルジェがレストランを紹介する場合においても，いくつかの代替案を提示しながら，客の反応を探るのが通例である．その際に，客が明示的に希望や “Yes/No” を表明しなくても，非言語情報からそれらを推察できるのが優秀な店員／コンシェルジェといえよう．我々は，ユーザの興味や反応を探る上で，対話における非言語情報，具体的には対話の「間」やあいづちなどの特徴の解明が鍵と考えており，このような「気のつく情報コンシェルジェ」を実現するための基礎となる研究についても紹介する．

2. 京都版ダイアログナビ –音声対話による情報検索・質問応答・情報推薦–

本章では，大規模情報を対象とした音声対話システムの構成論について概観し，Wikipedia を用いた観光情報コンシェルジェ「京都版ダイアログナビ」について紹介する．

現状の情報検索システムでは，キーワード列にマッチする数多くの候補がディスプレイに表示されて，ユーザがそれらを 1 つずつチェックするというインタフェースとなっている．これに対して，音声対話によりインタラクティブにユーザに適した情報を検索し，効率的に提示する枠組みを考える．従来の音声対話システムが主に関係データベース (RDB) の検索を対象としていたのに対して，構造を持たない大規模な情報を扱えるようにするためには，図 3 に示すような方法論の大きな転換を必要とする[1)]．

ここでは，大規模情報・知識ベースとして Wikipedia を想定する．ただし，百科事典を引くというスタイルではなく,「専門家／ガイド」との対話を実現する．その際に，エージェントやロボットというインタフェースを想定すると，従来のようにディスプレイに複数の候補を提示し，ゆっくり見てもらうということができない．すなわち，第一候補で正しい検索結果を得る必要があるとともに,（聞き取りにくい合成音声で）長々と文章を読上げることも避ける必要がある．そこで，対象とするドメイン（＝専門分野）をある程度限定するとともに，情報を小出しにインタラクティブに提示することで対応する．また，質問応答機能も導入する．

我々は，京都の観光地の紹介を対象と（Wikipedia の関係文書をあらかじめ抽出）して，図 4 に示す枠組みを考えた．本システムでは，ユーザ・システム双方が主導権をとることができる．ユーザ主導の検索・質問応答 (pull) モードでは，ユーザが照会した名所・寺院などについて書かれた文章を要約して数文で提示したり，具体的な質問（「いつ建てられたか？」など）に対して端的に応答する．システム主導の情報推薦 (push) モードでは，現在の話題に関連する事項について，ユーザの興味を引き出す形で提示する．具体的には,「この庭園は何で有名か知っていますか？」のように，質問形式を採用した．これは，図 5 に示す例のように，tf-idf 値[⋆1]の高い固有名詞など (NE:Named Entity) に対して，質問応答技術を逆過程で適用することで実現される．質問形式による案内は，人間の観

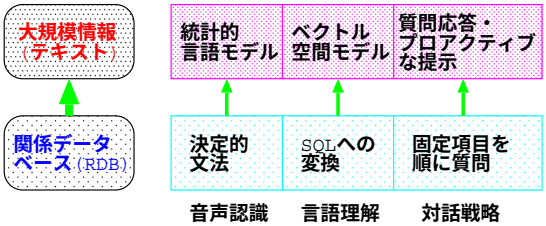


図 3 音声対話システムの大規模情報への展開

⋆1 その文書によく出現し（tf:単語頻度），他の多くの文書にあまり出現しない（idf:文書頻度の逆数）ことを示す指標

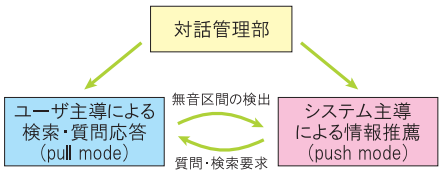


図 4 プロアクティブな音声対話の枠組み

原文：イギリスのエリザベス女王が**石庭を**絶賛したことが海外のマスコミでも報道されて，...
⇓（対象の名詞を対応する疑問詞に置き換え）
- イギリスのエリザベス女王が**何を**絶賛したことが海外のマスコミでも報道されて，...
⇓（対象の名詞が係る/係られる文節以外を削除）
- イギリスのエリザベス女王が何を絶賛した
⇓（疑問文に変換）
質問：イギリスのエリザベス女王が何を絶賛したか知っていますか？

図 5 システム主導の質問の生成例

光ガイドもしばしば用いる方策である．

このようなコンセプトに基づく情報案内エージェントを「京都版ダイアログナビ」として実装した[2]．現在の実装では，デフォルトは検索モードで，数秒間の沈黙を検出すると推薦モードに移行するようになっている．本システムによる対話例を図 6 に示す．この対話例におけるシステムの応答は，Wikipedia の龍安寺のエントリなどから自動的に生成されたものである．

本システムを，京都大学博物館の 2006 年夏季の企画展示において運用を行った．3ヶ月の運用期間中，のべ 2500 名のユーザの利用があり，25000 発話が収集された．期間の前半の 1497 対話のうち，581 対話（のべ 1635 回）で検索モードから推薦モードへの移行があり，このうち，さらに逆方向への遷移が行われたのが 248 対話（のべ 516 回）あった．多くの場合において，システム主導のプロアクティブな情報提示により，ユーザの興味が喚起されて対話が継続していることがわかる．

このような音声対話による情報案内は，ユーザが明確なゴールを持たなかったり，実際にコンテンツを見て興味や好みがわいてくるような場合に効果的であると考えている．なお現在のシステムでは，単に対話の「間」を検出すると情報推薦を行うようにしているが，より多くの非言語情報からユーザの興味や心的状態を察知する機構をとり入れるべく研究を行っている．

U１： 龍安寺について教えてください．
S１： 龍安寺は，京都市右京区にある臨済宗妙心寺派の寺院です．山号を大雲山と称します．本尊は釈迦如来，創立者は細川勝元，開山は義天玄承です．（沈黙を検出）
S２： では，イギリスのエリザベス女王がこの寺の何を絶賛したか知っていますか？
U２： 教えて．
S３： イギリスのエリザベス女王が１９７５年に公式訪問した際に，石庭を絶賛したことが海外のマスコミに取り上げられて，一躍有名になりました．
U３： どうやって行くの？
S４： 龍安寺は市バスの龍安寺前からすぐです．
U４： ありがとう．
S５： どういたしまして．

図 6 「京都版ダイアログナビ」との対話例

3. Mind Probing –能動的な働きかけと反応観察を行う情報コンシェルジェ–

本章では，大画面情報端末を用いてプロアクティブに情報を提示する枠組み・システムについて紹介する．大画面情報端末によって様々な選択肢（商品や観光情報）を提示する際に，画面を見ているユーザの興味を察知できれば，それに関係する情報を次々と提示・推薦する情報ナビゲーションが可能になる．

冒頭でも述べたように我々は，システムが主導権を持って働きかけを行うプロアクティブ・インタラクションモデルを提案している．これは，単に受動的に人の姿勢や動きを観察するのではなく，システム側が積極的に提示情報を変化させ，それに対するユーザの反応（視線などの非言語情報）を計測する．そして，これら提示と反応の組から人の心的状態を探る枠組みであり，商品販売などのプロービングになぞらえて “Mind Probing” とよぶ[3]．

予備的な観察実験として，被験者に 50 型ディスプレイに提示した選択肢（４分割した領域のうち３つに表示）から，好みのものを選択するタスクを行ってもらった．図 7 は，各時刻（横軸）においてどの選択肢（縦軸）を見ていたかを示した一例である．この視線の動きから，被験者は前半で画像や解説文を「読み込む」状態に，後半は「比較評価・選択」を行おうとしている状態にあったと推測され，この後半の視線の動きが興味を強く反映していると考えられる．

しかし，視線の滞留パターンは，提示情報の種類や複雑さなどの要因の影響も受け，単純に情報を提示するだけでは，「読み込む」状態と「比較・選択」状態を分離することは困難である．そこで，これら二つの状態の分離を容易にするために，情報提示の初期段階では各項目の情報を順次切り替えて排他的に表示（順

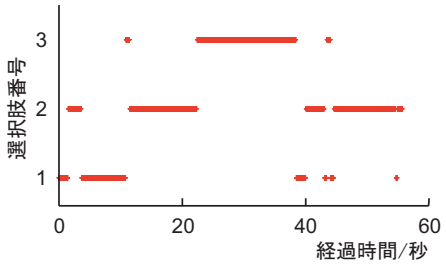


図 7 大画面端末において選択肢を閲覧する視線の滞留パターン例

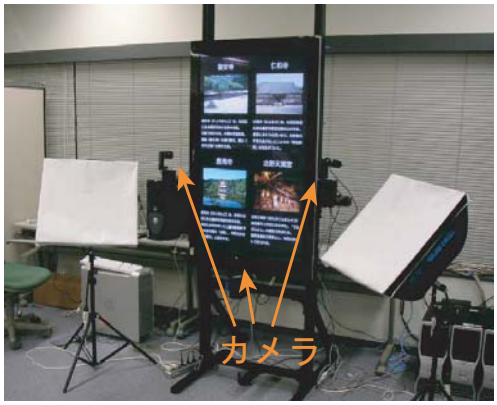


図 8 Mind Probing システムの外観

次提示モード）し，一通り表示し終えてからすべての項目を同時に表示（一覧提示モード）するプロービングの枠組みを設計した．順次提示モードは，画像やテキストの提示タイミングを，ユーザが情報を得るのに要する認知時間に合わせて設計しておくことで，ユーザを「読み込む」状態へと導く．これに続く一覧提示モードは，順次提示で得た情報に基づいて自由に「比較・選択」することを促し，このときの視線パターンにより興味を探る．現時点での被験者数はまだ少ないものの，約 78%の正解率で興味を推定できている．

この枠組みに基づいて，情報通信研究機構 (NICT) 知識創成コミュニケーション研究センターにおいて，図 8 に示すような大画面情報端末の構築を行った．本システムでは，3 台のカメラによる処理結果を統合することで，画面から 1m 程度離れた立ち位置において，画面上で平均 10cm 程度の誤差で視線を計測することが可能になっている[4]．この視線計測に基づいて，図 9 に示す提示フローによって，推定した興味に応じて情報をナビゲートしていくことが可能である．さらに，多様な非言語情報と言語情報を用いたプロービング手法を導入することで，商品案内などの情報提供をより柔軟に行う情報コンシェルジェの開発を進めている．

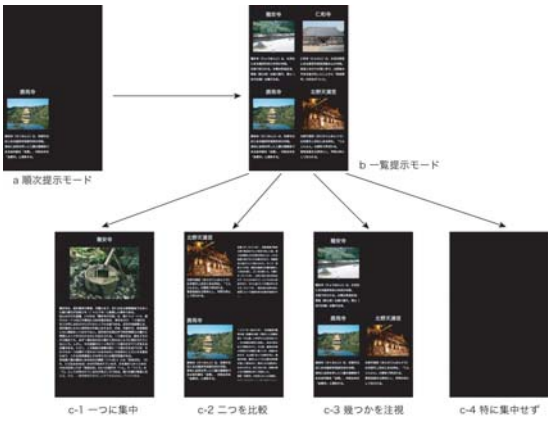


図 9 興味推定に基づく情報提示フロー

4. 対話における非言語情報を介したユーザの興味・反応の察知にむけて

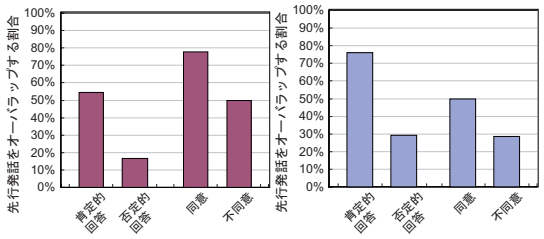
プロアクティブな情報コンシェルジェを実現するには，リアルタイムのインタラクションにおいて，ユーザの興味や反応を察知することが鍵となる．特に，対面の音声対話においては，いちいち言葉で確認しなくても，提示された情報に興味を示しているか，否定的であるかを察知できることが望ましい．前述の Mind Probing システムにおいては，数個の選択肢において視線の滞留を手がかりとしていたが，音声対話においては,「間」やあいづちなどの非言語情報にユーザの興味や反応が表出されると考えられる．本章では，そのような観点から我々が行っている研究を紹介する．

4.1 人間どうしの対話における「間」の分析

まず，対話の「間」，発話タイミングに着目した分析を行った．我々は，模範的な対話を演じていると考えられる漫才を分析対象とした．漫才には，意図や心的状態を表現するための間合いに関するこつが凝縮されていると考えたからである．

発話行為 (DA: Dialog Act) と発話タイミングとの関係について，ボケ役に対するツッコミ役の応答について調べた．質問に対する肯定的回答と否定的回答，陳述・意見に対する同意と不同意のそれぞれにおいて，オーバラップ（相手の先行発話が終了する前に開始した）発話が現れた割合を図 10(a) に示す．否定的な応答の方が，肯定的な応答に比べて，オーバラップの割合が小さい．肯定的な場合には，発話タイミングを早めることでその意味合いを強め，否定的な場合には，タイミングを遅らせることで十分考慮した返答であることを表現していると考えられる．

次に，落語についても分析を行った．落語では 1 人



(a) 漫才における発話行為と発話タイミングとの関係　(b) 落語における発話行為と頭部動作タイミングとの関係

図 10 漫才・落語における発話タイミングの分析

で複数の役柄を演じ分けるが，頭部を左右にふりむける動作（顔向きの切替え）が役柄交替を表現するために用いられている．そこで，先行役柄の発話終了時刻に対する頭部動作開始タイミング（「視覚的な間合い」）を調べたところ，図 10(b) に示すように，漫才と同様の傾向が見られた．

上記の分析に用いたサンプル数は十分多くはないが，いずれも模範的な対話を演じていると考えられることから，発話における「間」と発話者の意図や反応との関係を示唆するものといえる．なお，一般の人どうしの模擬対話においても，否定・拒否を示す応答の方が，肯定・受諾を示す応答に比べて，発話タイミングが遅くなることが報告されている[6]．

4.2 人間どうしの対話におけるあいづちの分析

次に，あいづちに着目した分析を行った．人間のプロのガイドが京都の観光地を案内し，ユーザが 1 日に回る場所を決めるというタスクの模擬対話を対象に分析を行った．ガイドはユーザの希望に沿うようにいくつかの名所を順番に紹介しながら，ユーザの反応を探り，詳細な説明を続けるか，別の名所に切り替えるか，といった判断をしている．その際に，ユーザの非言語的な情報を手がかりにしていると考えられる．ここでは，ある名所の説明中になされたユーザのあいづち（「はい」「うーん」など）に着目し，その発話タイミング（先行するガイド発話終了時との時間差）及び頻度（当該話題区間におけるガイドの発話数で正規化した回数）と，その名所が実際に訪問先として選ばれたかとの関係を調べた．選ばれた場合を「肯定的」，そうでない場合を「否定的」と分類している．

結果を図 11 に示す．早いタイミングで（100msec 以内に）なされたあいづちは大半が肯定的な反応であり，あいづちが多くなされた（頻度 0.3 以上の）場合もほぼ肯定的といえる．この 2 つの特徴を統合することで，適合率 84%の精度（再現率は 59%）で肯定的な場合を検出できる．これは，明示的な確認がなくても，ガイドが説明を続けたり，決定を促すことができることを示唆している．

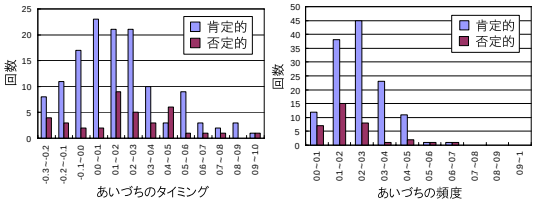


図 11 情報推薦に対する反応とあいづちのタイミング・頻度の関係

4.3 システムとの対話における「間」の分析

上記のような人間どうしの音声対話の分析から得られる知見が,（現状の）音声対話システムに適用できるかは自明でない．まず，システムに対する発話スタイルは，人間どうしの対話と大きく異なる．例えば，前節で述べたあいづちの現象は，システムとの対話においてはほとんど見受けられない．システム側からあいづちをうつことで，自然な対話システムを実現しようとする研究[5] もあるが,「見かけ」や合成音声の質，さらには会話能力全般を向上することが必要と思われる．

前述の「京都版ダイアログナビ」の運用で収集された対話コーパスにおいて，発話タイミングを分析したところ，人間どうしの対話の分析結果と比較して，(1) 全般に発話タイミングが 2 秒近く遅い（「間」があく），(2)「肯定」や「受諾」の場合が「否定」や「拒否」よりもタイミングが遅い，ことが観測された．特に後者は，前記の分析結果とも整合しないものである．

そこで，ユーザがエージェントを対話相手としてどのように認識しているかによって分類を行った．具体的には，エージェントに対する挨拶や名前の呼びかけ，エージェントに関する質問のいずれかを行ったユーザを「擬人的に接するユーザ」と分類し，そうでないユーザを「道具的に接するユーザ」と分類した．

図 12 に，このユーザ分類と発話行為 (DA) 毎の発話タイミングの分析結果を示す．システムからの情報推薦に対する応答に着目すると，擬人的に接するユーザでは，人間どうしの対話と同様に,「受諾」の方が「拒否」よりも応答が早い傾向が見られる．これは，インタフェースの効果を示すものである．

4.4 自然な対話における「間」と視線の分析

対話における働きかけは基本的には言語を発話することによって行われるが，対面の場合には，視覚に作用する身体動作にも効果があると考えられる．そこで，働きかけに頻繁に付随する相手への顔向け（視線の投げかけ）についても分析を行った．

働きかけと応答が繰り返される合意形成対話の音声と映像を解析した結果，顔向けを伴う働きかけに対す

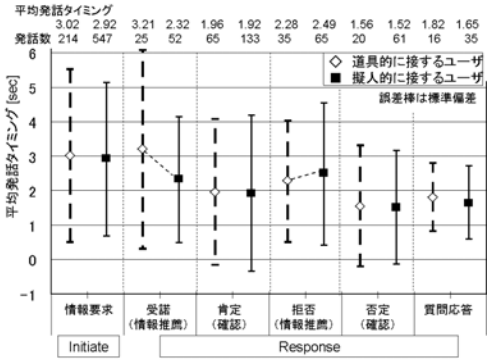


図 12 道具的／擬人的に接するユーザの発話タイミングの比較

る不同意の応答タイミングが同意の場合よりも遅くなり，それらのタイミングの差が顔向けを伴わない場合より大きくなる傾向が見受けられた．顔向けは，対話者に対話の時間構造を強く意識させるトリガーになっており，相手の心的状況をプローブする行為の一つであるといえる．これは，マルチモーダルなインタラクションの重要性を示唆するものであり，ロボットなどを用いたインタフェースを設計する際の指針になる．

5. 今後の展開

特定領域研究「情報爆発 IT 基盤」の研究項目 A03「情報爆発時代におけるヒューマンコミュニケーション基盤」では，上記で紹介した以外にも，プロアクティブなインタラクションモデルに関して，盛んに研究が進められている．例えば久野らは，美術館において絵画の解説を行うロボットを試作している．解説の終了時にロボットが人の方へ顔を向け，ロボットの発話終了時と顔向け動作との間合いを適切に同期させることで，人の振り向きやうなずきを引き出すことができることを示している[7]．また中野らは，大画面を用いた情報コンシェルジェとして，視線などから推定されたユーザの態度に応じて，適応的に商品説明を行う会話エージェントの開発を進めている[8]．

より基礎的な研究として，三宅らは，対話において合意が形成されるにしたがって，２人の話者の「間」が同調してくることを明らかにしている．今後，本特定領域研究の一環として構築されている「IMADE ルーム」において，マルチモーダルインタラクションコーパスの大規模な収集と分析が進められ，さらに多くの知見が得られることが期待される．

本稿で紹介したシステムはまだ初期段階のものであり，これらの知見を総合的に活用することで,「気の利く情報コンシェルジェ」の実現に近づいていくものと考えている．

謝辞 本稿で紹介した研究は，科研費特定領域研究「情報爆発 IT 基盤」の研究項目 A03 の一環として行われたものである．また，これらの研究開発の一部は，情報通信研究機構 (NICT) の翠輝久，水口充，佐竹純二，小林亮博，小嶋秀樹，柏岡秀紀の各氏の貢献におうものであり，深く感謝する．

参 考 文 献

1) 河原達也. 話し言葉による音声対話システム. 情報処理, Vol.45, No.10, pp. 1027–1031, 2004.
2) 翠輝久, 河原達也, 正司哲朗, 美濃導彦. 質問応答・情報推薦機能を備えた音声による情報案内システム. 情報処理学会論文誌, Vol.48, No.12, pp. 3602–3611, 2007.
3) 水口充, 浅野哲, 佐竹純二, 小林亮博, 平山高嗣, 川嶋宏彰, 小嶋秀樹, 松山隆司. Mind Probing: システムの積極的な働きかけによる視線パタンからの興味推定. 情報処理研究報告 HCI-125, pp.1-8, 2007.
4) 佐竹純二, 小林亮博, 平山高嗣, 川嶋宏彰, 松山隆司. 高解像度撮影における実時間視線推定の高精度化. 電子情報通信学会技術報告 PRMU-107-491, pp.137-142, 2008.
5) Norihide Kitaoka, Masashi Takeuchi, Ryota Nishimura, and Seiichi Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *Journal of Japanese Society for Artificial Intellignece*, Vol.20, No. 3 SP-E, pp. 220–228, 2005.
6) 藤原敬記, 伊藤敏彦, 荒木健治. タスク指向対話における相互の対話意図を考慮した対話リズムの分析. 言語・音声理解と対話処理研究会, SIG-SLUD-A701, pp. 45–50, 2007.
7) Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, Matthew Burdelski, Michie Kawashima, Hideaki Kuzuoka. Precision Timing in Human-Robot Interaction: Coordination of Head Movement and Utterance. CHI 2008, pp.131-139, 2008.
8) 石井亮, 中野有紀子. ユーザの注視行動に基づく会話参加態度の推定 —会話エージェントにおける適応的会話制御に向けて—. 情報処理学会第 70 回全国大会, No.5, pp.271–272, 2008.