

マルチモーダル対話における参照行為

Referring Actions in Multi-Modal Dialogues

加藤 恒昭* 中野 有紀子*
Tsuneaki Kato Yukiko I. Nakano

* NTT 情報通信研究所

NTT Information and Communication Systems Labs., Yokosuka 239, Japan.

1996年9月26日 受理

Keywords: referring action, multi-modal dialogues, generation, natural language.

Summary

An empirical study on referent identification requests or referring actions in multi-modal dialogues is reported. Through the examination of a corpus obtained by experiments in which experts explained the installation of a telephone with answering machine features in two situations, spoken-mode dialogue and multi-modal dialogue, referring actions in multi-modal dialogues were well analyzed and compared with those in spoken-mode dialogues from two perspectives: information communicated and the style of goal achievement. One of the findings obtained through this study is that the availability of pointing actions in multi-modal dialogues reduces the amount of information verbally communicated. The kind of information reduced from the spoken-mode situation depends on contextual status. This fact suggests that pointing actions are different from actions that communicate a specific type of information such as object location and figures. Rather, they are actions to activate the object referred to in the interlocutor's mental space. As communicating information on an object plays the same role in referring actions, pointing actions and verbal communication can be compared in this standpoint. Another finding is that in spoken-mode dialogues, as compared to multi-modal dialogues, the speakers realize identification requests as a series of fine-grained steps, and try to achieve them step by step with frequent confirmation.

1. はじめに

名詞句等で言語的に表現したり、指差し等を用いて直示することによって、事物を相手に同定させる行為を、参照もしくは参照物同定要求と呼ぶ。参照は、すべての対話で現れ、特に物理的な対象の操作や組立に関する指示を与えるような対話においては、中心的な役割を果たしている。従って、すべてのマルチモーダル対話システムは、この参照行為を実現するためのなんらかの枠組みを持っていないといけない。

ところが、この参照行為がどうあるべきかについて、研究者の間で意見の一致がみられていないように思われる。言語情報と視覚的情報というようなモード間の関係について見てみると、例えば、COMET [Feiner

90] は、説明の意味内容を確定した後に、どの部分をどのモードで表現するかを決定するという順次的な枠組みをとっており、与えられた意味内容をどのように表現するかという点でのみモード間の相互作用が生じる。[Maybury 93] の枠組みでは、モードに依存しない修辞ゴールの達成のためのプランニングを通じて意味内容の決定と表現の実現が同時に行われ、その両者でモード間の相互作用が生じると主張されている。ところが、そこで示されている参照のためのスキーマでは、視覚的な手段が利用可能であるか否かにかかわらず言語的な参照表現の内容は同一であり、この例を見ると、指差し等の視覚的な参照行為は単に付録的に用いられ、言語による参照行為に影響を与えないと考えられているようである。これらの対極として、[André 94, Wahlster 91] では、cross-modal references と名

付けられた複数モード利用時に初めて可能となるモード間の参照関係が考慮されており、異なるモードで伝達される情報の間に密接な関係があることが前提とされている。また、具体的にどのような表現が選好されるかについても、システム毎に相違が見られる。[Neal 91]では、視覚的な手段を積極的に用いるべきであると主張され、言語的手段の利用は極力抑えられている。一方で、[Claassen 92]では、Saliency [Alshawi 87]と関連した談話的要因が重視され、対象物が現在可視であるかどうかの考慮は決定木の葉近くになって初めて行われる。

本稿はこれらの問題に経験論的に取り組むものである。つまり、実験的に収録された人間-人間対話の観察を通じて、マルチモーダル対話での参照行為において、どのような種類の情報がどの程度の量、伝達されているか、それらはどのような様式で伝えられているのかを明らかにし、それらに基づいて参照行為生成について考察する。その結果をマルチモーダル対話システム設計の指針へと発展させてゆくことが最終的な目的である。このようなアプローチはシステム設計を目的とした上述の様々な取り組みと補完しあうもので、参照行為生成の問題に新しい光を投げかけると期待できる。

以下、2章で、対話データ収集のために行った実験について説明する。3章では、参照行為において、どのような情報が伝達されているかの結果を報告する。4章では、参照行為の達成の仕方について、利用可能なモードとの関係を述べる。5章ではこれらの結果について考察し、マルチモーダル対話システム設計への含意を述べる。

2. 実験と対話データ

実験は、マルチモーダル対話システム設計のための様々なデータを提供できるような対話コーパスを収集するために行われた。タスクとしては、留守番電話機能付の電話機*1の初期設置を選んだ。そこで初心者は、専門家の教示に従って、箱詰めされた電話機セットを取り出して組み立て、その後、音量設定、時計合わせ、応答メッセージ録音等、8項目の初期設定を行う。最後に、幾つかの機能ボタンについての説明を受ける。

ヘッドセットを用いて音声のみで説明が行われる音声対話状況（以下、音声状況）と、音声と視覚情報の伝達が可能であるマルチモーダル対話状況（以下、MM状況）とを設定し、説明を収録した。共に、専門家と

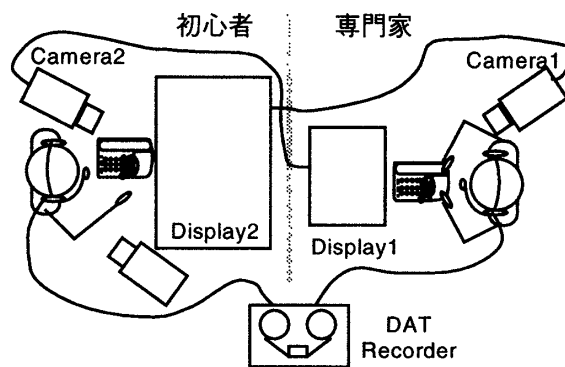


図1 マルチモーダル対話状況

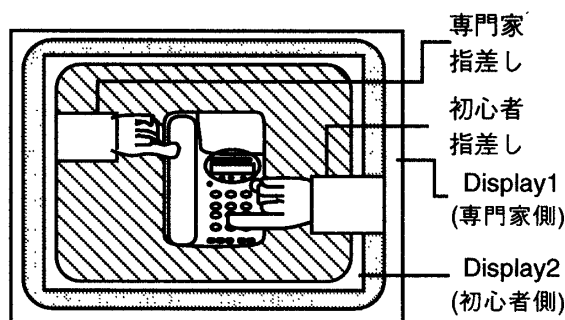


図2 マルチモーダル対話状況での視覚情報

初心者は自分の手前に電話機を置き、それを自由に操作することができる。初心者は教示を受けながら、実際に組立と初期設定を行っていく。なお、本稿では議論しないが、これらに加えて独白状況での説明の収録も行っている。本実験については [Kato 95] に詳しい。

音声状況で音声の双方向の伝達が可能であると同様、MM状況では、音声と視覚情報の双方向の伝達が可能である。この様子を図1に示す。Camera1によって撮影された専門家側の電話機とそこへの指差しがDisplay2に映される。初心者はこの画面へ指差しを行うことができ、その様子はCamera2によって撮影され、Display1に表示される。これにより、専門家は、図2のように、自分側の電話機とそこへの自分の指差し（の映像）とそこへの初心者の指差しを見ることができる。結果として、専門家側電話機とそこへの両者の指差しは完全に共有される。視覚的情報のこのような共有は対面状況と同じであるが、その一方で一般の対面状況と異なり、視線や傾き等、指差し以外の動作は共有されない。この状況は、ちょうどディスプレイへの画像出力と音声による言語出力を持ち、ユーザからのマウス等による指示動作と音声入力を受け付けることができるマルチモーダル対話システムの模倣となっていると考えられる。

この実験によって収録され、書き起こされた対話コー

*1 NTT製ハウディレボンス IM-3300を用いた。

パスのうち、同じ5人の専門家によって行われた、MM状況と音声状況での、初心者役の被験者を換えて行った各3回ずつの説明(従って、5専門家×2状況×3説明=30説明)を今回の調査の対象とした。初心者役の被験者は30説明すべてで異なっている。それらの説明から、音量設定、時計合わせ、応答メッセージ録音等の初期設定を行なっている部分に現れる参照表現を抽出した。この部分での参照行為は、対象物を同定させ、それに対してある行為を行わせるという文脈で行われる。登録ボタンを同定させ、それを押させる、音量スイッチを同定させ、それを大の側に設定させる等である。この文脈は、[Cohen 84, Oviatt 91]におけるポンプ組立を題材とした調査での文脈に近いものである。彼らはそこでの参照がどのような言語行為によって達成されるか等をキーボード対話、電話対話、独白とで比較しているが、本実験の結果はこの研究と比較可能であると考えている。

参照表現は、専門家がその対象についての操作の指示を開始した時点から専門家と初心者がその対象物について合意を確立したと見做せる時点までの専門家の発話の部分であるととらえ、初心者の明示的もしくは暗黙の問い返しによって付加されたとみられる専門家の発話の部分もそれに含めた。本作業は、書き起こしテキストとビデオ/オーディオテープを併用して行った。更にMM状況においては、その発話に伴って対象物への指差しやその他の動作があったかを注釈づけた。以下に、抽出された部分の例を示す。ここで、記号//はそこに初心者のあいづち等が入ったことを示し、[]で囲まれた部分は、その発話に合わせて{}内に示す指差し等の行為があったことを表している。

音声状況 ではまず、あの、カセットカバーを開けてください。// そうすると、ボタンが並んでると思いますけれども // それの左から2番目、登録ボタンを押してください。

MM状況 [まずカセットカバーを開けてください。{ 本体-指差し、カセットカバー-指差し、開ける }] 登録ボタンを押します。

3. 参照において伝達された情報の種類と量

どのような情報が伝達されることによって参照行為が行われるかと対話状況との関係について調査した。ここでは情報を広い意味で捉え、代名詞の利用や「先程使った」というような文脈的言及を含めて、以下のように分類した。

代名詞 「これ」「それ」等の代名詞の利用

文脈参照 「以前使った」「先程の」等、談話中で対象物が既に参照されたことがあるという記述

一般名称 「ボタン」「スイッチ」等、対象物の種類を示す名称で、電話機内でその対象を一意化するのには不十分なもの

固有名詞 「登録ボタン」「回線切替スイッチ」等、電話機内でその対象を一意化するのに十分な名称

形状 「丸い」「赤い」「大きい」等、対象物の形状、大きさ、色についての記述

位置 「本体右上の」等、対象物がどこにあるかという位置に関する記述

紋様 「留守と書かれた」等、対象物やその近傍に記入されたマークや文字列等についての記述

機能 「音量を調節する」等、対象物が果たす役割、機能についての記述

関連物 対象物が属す集団やその近傍の事物についての言及

その他 「2列に並んだ」等、対象物そのものではなくそれが属す集団についての記述等、上記のいずれにも分類できないもの

これらの数え上げにおいて、代名詞、一般名称、固有名詞については、単純にその延べ数を数え、それ以外のものについては、同じ内容を伝達しているものは字面上の反復があっても、ひとつと数えた。つまり、異なり数をとった。

また、MM状況については、参照表現に伴って行われた指差し行為の有無を集計した。ここでは、指差しが行われた場合はすべて1とし、複数回であるかや、継続時間は考慮しなかった。

談話的要因を考慮して、各状況について2種類の参照についての集計を行った。第一は、対象物が最初に説明対話中に導入される際の参照行為(以下、初回参照と呼ぶ)であり、12の対象物*²に関する(従って、12対象物×5専門家×3説明=180件の)参照行為を平均し、一回の参照でどのような情報がどの程度伝達されるかを見た。第二は、この初回参照の後の、異なる設定項目の中での参照行為(以下、再参照と呼ぶ)で、これは4つの対象物*³に関して同様の集計を行った。例えば、登録ボタンは時計の設定で参照された後、

*2 カセットカバー、スピーカボタン、登録ボタン、時刻パンチボタン、ワンタッチボタン1、留守ボタン、応答1ボタン、消去ボタン、DP/PBスイッチ、受話音量ボタン、スピーカ音量つまみ、呼出音量切替スイッチ

*3 カセットカバー、スピーカボタン、登録ボタン、応答1ボタン

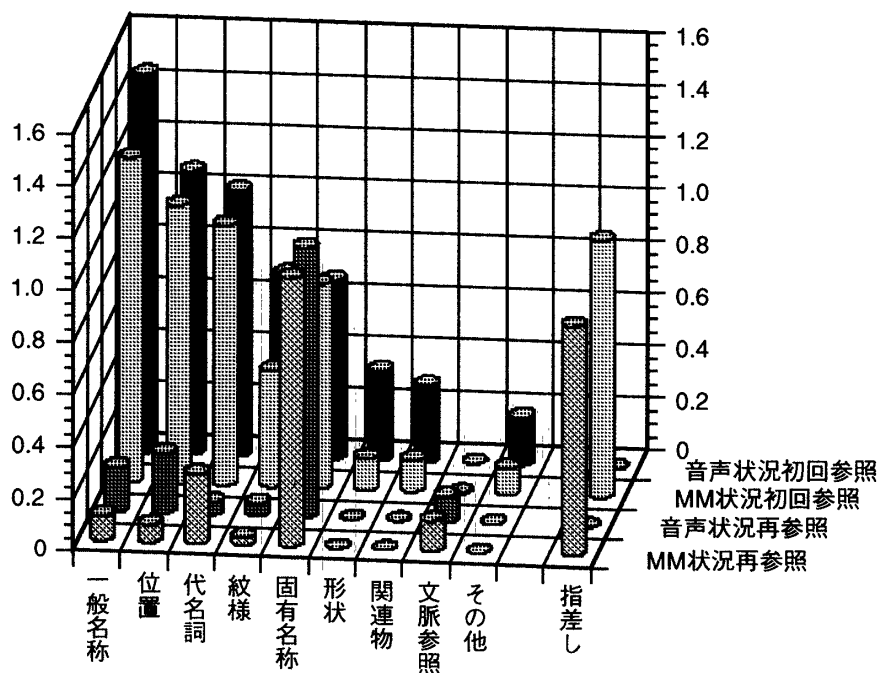


図3 参照行為において伝達される情報の種類と量

ワンタッチダイヤルの設定で参照されており、後者の参照が再参照となる。結果を図3に示す。情報の種類については音声状況初回参照での利用の順に並べてある。なお、初回参照については12対象物に関する結果であり、再参照は4対象物に関する結果であるので、これらは直接比較の対象となるものではないが、この4対象物だけを取り出して初回参照について集計しても、図に示したのとほぼ同様の傾向を示すことを確認している。

結果は次のようにまとめられる。以下において、有意差の検定は、各対話での着目した対象物についての反応の平均値を算出して、これを分析データとし、対話状況を被験者内要因とした分散分析に基づいて行った*4。

- MM状況においては、指差しが頻繁に用いられる。注釈付テキストによれば、これらの指差しはコ系の代名詞を伴っており、cross-modal referencesになっていると考えられる。
- 初回参照において、MM状況では、音声状況と比較して、言語による情報の伝達が少なくなっている ($F(1,14) = 12.83, p < 0.01$)。形状、紋

様、関連物においてその差が有意である (順に、 $F(1,14) = 20.39, p < 0.01$, $F(1,14) = 19.50, p < 0.01$, $F(1,14) = 19.96, p < 0.01$)。音声状況と比較してMM状況での利用が減少する情報は、音声状況での利用頻度が比較的低いものである。なお一般名称の伝達も音声状況で多い ($F(1,14) = 5.00, p = 0.04$) が、これについては延べ数であることと他の情報伝達の際に主辞となることによると考えられる。

- 両状況で、再参照に利用される言語的な情報の量は、初回参照に較べ大きく減少する (音声状況で、 $F(1,14) = 158.69, p < 0.01$, MM状況で、 $F(1,14) = 57.04, p < 0.01$)。再参照においても、MM状況では、音声状況と比較して、一部の言語的な情報の伝達が少なくなっている。位置の情報について、その差が有意である ($F(1,14) = 7.00, p = 0.02$)。一方、代名詞の利用はMM状況の方が多 ($F(1,14) = 7.53, p = 0.02$) が、この代名詞も殆どがコ系である。

4. 参照行為達成の様式

参照行為がどのように行われ、どのような過程でそのゴールが達成されるか対話状況との関係を明らかにするために、幾つかの機能的特徴について調査した。

調査した第一は、参照に用いられる表層言語行為の分布の違いである。これは [Cohen 84] においても調査

*4 分散分析による F 値、及び以降の z 値、 χ^2 値の計算方法と意味づけについては、[Hoel 76] や [Cohen 95] 等の参考文献を参照のこと。すべての検定において、 p 値は棄却率と呼ばれ、ここでの主張が誤っているという仮説が真である時に、それを棄却してしまっている危険性 (確率) を示す。一般には $p \leq 0.05$ の時に、統計的な有意性を主張しうる。

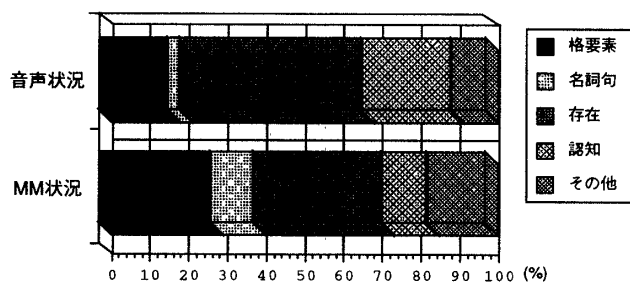


図4 参照行為で利用される発話の形式

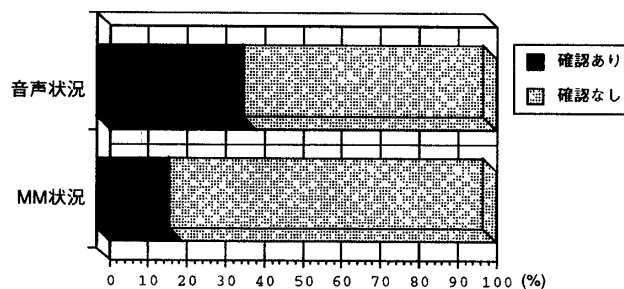


図5 参照表現中の確認の有無

されたものであり、そこでも述べられているように、参照表現が、行為を依頼する命令文等の中に格要素として吸収されてしまい、しかもイントネーション等によって陽にマークされることなく現れている場合は、参照行為の達成は独立したゴールとして実現されておらず、一方、行為の依頼とは異なる文で行われる場合、独立したゴールとして実現されていると考えることができる。表層言語行為は以下の様に分類した。

格要素 他の行為を依頼する命令文等の格要素として特にマークなく現れる。この場合のみ、参照行為が独立したゴールとして実現されていないと考える。
例 留守ボタンを押してください。

名詞句 名詞句片として対象物の記述が現れ、それによって参照行為が行われる。
例 留守ボタン、このボタンですが、これを押してください。

存在 対象物の存在を述べる文によって、参照行為が行われる。
例 ダイヤルボタンの右上に留守ボタンがあります。それを押してください。

認知 「見る」「開ける」等、その対象を認知するための行為が依頼されることによって、参照行為が行われる。
例 ダイヤルボタンの右上を見てください。そこにある留守ボタンを押します。

その他 上記のいずれにも分類されない発話によって参照行為が行われる。

図4に初回参照においてこれらが用いられた割合(%)を示す。状況間の差は有意であり($\chi^2(1) = 26.72$, $p < 0.01$)、音声状況では、MM状況に比べ、参照行為の達成が独立したゴールとして、つまりここでの分類では格要素以外の形式によって、実現されることが多い($z = 2.49$, $p < 0.01$)。

第二の特徴として、抽出した参照表現中に、疑問文もしくは付加疑問文による確認が含まれるかを調査した。

この特徴も参照行為の達成が独立したゴールであることと関連する。初回参照において確認行為が見られた割合を図5に示す。状況間の差は有意である($z = 3.98$, $p < 0.01$)。

最後に、ゴール達成の過程に関する特徴として、詳細化に関連する表現の存在について調査した。これは[Oviatt 91]や[Clark 90]でも調査されている。ここでは、発話境界やイントネーションについては配慮せず、詳細化に関連する表現を以下のように分類した。

同格表現 対象を説明する名詞句に続く同格的な名詞句によって、説明が詳細化されているもの。

例 留守と書かれたボタン、赤い大きなボタンを押してください。

補足表現 対象を説明する名詞句を含んだ文や節に続く補足的な名詞句や文によって、対象の説明が詳細化されているもの。

例 留守と書かれたボタンを押してください。赤い大きなボタンです。

修正表現 統語構造は上記のいずれかと同じであるが、意味内容が直前の説明と矛盾しており、それを訂正しているもの。

例 右端のボタンを押してください。左端です。

初回参照において、ひとつの参照行為中にこれらの表現が平均何回現れたかを図6に示す。同格表現、修正表現については有意な差はなく、補足表現については、音声状況で多いという弱い傾向が見られた($F(1, 14) = 3.53$, $p = 0.08$)。

5. 議 論

5・1 情報伝達と指差しの役割

音声状況で言語的に伝達された情報を、指差し行為が利用可能であるMM状況で伝達されたものと比較することにより、指差しの役割について考察することができる。MM状況において言語的に伝達される情報が

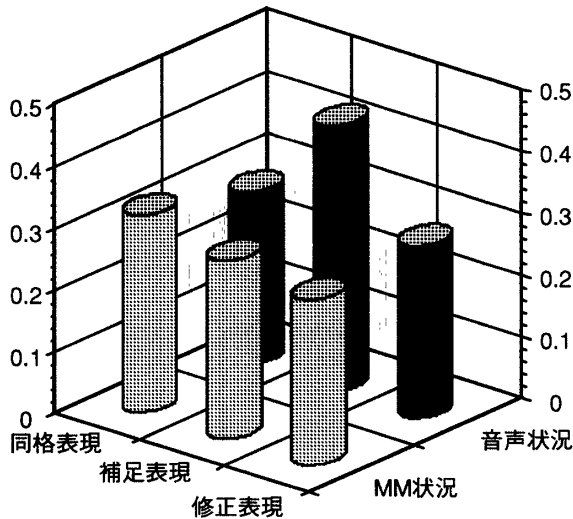


図6 参照における詳細化と関連する表現

少なくなることは、指差しの利用が単に付録的なものではなく、モード間に相互作用が生じていることを示している。この点で、それを否定する Maybury が示したスキーマは我々の結果と一致しない。また、例えば、「場所の情報を言葉で伝える代わりに指差しを用いて伝えた」というように指差しの役割のある種の情報伝達と捉える考え方がある。これは事前に発話すべき意味内容が決定していて、その後でそれをどのモードで伝達するかを決定する COMET の枠組みに近い。しかし、これも我々の実験の結果と一致しない。我々の結果では初回参照では形状などの情報の伝達が、再参照では位置の情報の伝達が減少しており、これらの情報が指差しによって伝達されていることになるが、初回参照とそれ以降の参照とで指差しに異なる役割を考えるのは不自然で煩雑である。

Appelt は、参照行為において、話者は対象を活性化するために、それを他と区別して一意化するために必要十分な情報を選択して利用するという枠組みを提案している [Appelt 85]。Dale らは、参照における適切性と効率性の原則を提示し、最も効果的なものから順に必要な充分となるまで情報を選択するという参照における内容決定方式を提案している [Dale 91]。これらの枠組みを一步進め、様々な情報の伝達や指差しなどの行為が対象を活性化する効果とそれを利用するコストによって特徴付けられると考え、必要な活性化の効果をj得るための最も安価な組み合わせによって参照を行っていると考え、我々の結果を説明することができる。この説明では、指差し行為は何らかの情報を伝達する行為ではなく、指差し行為そのものとして、様々な情報の伝達行為と効率の観点で比較される。

例えば、音声状況ではある対象を活性化するのに、効率的なものから3つ、例えば、一般名称、位置、形状、を選んで伝達することで充分であったとする。この同じ対象を MM 状況で参照する場合は、まず、極めて効率的な行為として指差しを用いて、不足分に言語による情報の伝達を用いると考える。指差し行為によってある程度活性化されるので、言語的な情報は、例えば、2つで充分で、一般名称と位置だけが述べられることになる。その結果として、形状の情報の利用が減少するように見える。音声状況での利用頻度が比較的低い情報が MM 状況で伝達されなくなっているが、これは、このように指差しで底上げして、その不足分を言語的な情報伝達で補うと考えると説明できる。

必要な活性度は、談話的要因*5によって変化し、既に談話中に導入された事物を改めて参照する際は、より少ない活性度で充分であると考えられる。更に初回にラベリングした固有名称も利用可能である。例えば、音声状況では固有名称と位置の2つで十分な活性化が行われ、MM 状況では、指差しを利用することで、言語的に伝達される情報は固有名称だけとなる。この時は、位置の情報が減少するように見える。

もちろん、コスト最小については聞き手の役割を考慮した協調的行為としての観点 [Clark 90] で考えなければならぬし、厳密には問題となるのは結果として得られたものの最適性でなく、それを生成する過程の最適性である。しかし一次近似として、上述したモデルは充分妥当なものであると考える。

5.2 ゴールの達成と対話のモード

[Oviatt 91] では音声独白と音声対話を比較して、後者の方が参照を独立したゴールとして実現することが多いと述べられている。[Cohen 84] では、キーボード対話と音声対話を比較して、同様の観点から、音声対話の方が、より細かいゴールの達成を行っているという結果を報告している。我々の結果では、音声状況では、MM 状況に較べて参照を独立したゴールとすることが多いことが明らかになった。つまり、音声対話においては、説明者が参照行為の達成をより細かいゴ

*5 談話的要因として提案されているものとして、例えば、Saliency [Alshawi 87] は、ある対象が談話中に現れた際に統語的要因を含む様々な要因によってその値が定まり、その後、談話の進行にともなってその値が徐々に減衰していくとされており、その対象が談話のその時点においてどの程度顕現的であるかを示すとされる。また、談話構造の一要素である Attention stack [Grosz 86] は、談話のある時点での対話者の注視構造を反映し、その中で対象物の位置が、代名詞等の参照物となりえる優先順位に関係するとされている。

ルの列として実現し、それらを順次達成していく形で参照を進めている。その他の特徴も同様の傾向を示唆している。つまり、音声状況では、参照の過程で確認を取る場合が多く、明確な差は得られなかったものの傾向として、補足表現による断片的な詳細化が多い。

このような特徴について、キーボード対話、音声のみによる対話、音声に視覚情報を伴ったマルチモーダル対話とを比較した場合、キーボード対話とマルチモーダル対話が近い傾向を示し、音声対話と区別されることは、利用できるモードの効率ということだけを考えると、一見、直感に反する。モードの伝達効率という点では、キーボード < 音声 < マルチモーダルと考えられるためである。一方、道順説明のタスクにおいて、同様の傾向が初心者からの質問や確認についても見られることが報告されている [Ishikawa 94]。そこでは、対話の特徴はそこで利用できるモードの伝達効率だけでなく、対話参加者が利用できる情報の量にも影響されるというモデルが提案されている。また、対話の冗長性に着目して、対話者は資源制約された主体であるために、様々な要因によって、対話方略を変更しその効率を高めているというモデルも提案されている [Walker 94]。ゴール達成に関する特徴がモードの伝達効率だけからでは説明しにくいという今回観察された現象を説明する助けとして、これらのモデルが参考になると考える。

6. おわりに

参照もしくは参照物同定要求が、どのような発話形式によって行われ、その際どのような情報が伝達されるかについて、マルチモーダル対話状況と音声対話状況とを比較した。その結果、指差し行為が情報伝達行為でなく、対象を活性化させる効率という観点でそれと比較できるだけであるという捉え方と、対象を活性化するために必要十分な行為を行うという参照行為のモデルが示唆された。また、音声対話においては、より細かいゴールの達成の連続という形で参照が進められていることを示した。今後の課題として、提案したモデルをより詳細化具体化してシステムとして実装し、モデルを評価検証することと、今回の結果が、今回対象とした装置の初期設定の場面だけでなく、地図上の場所や道順の説明など、指差し行為が重要である他の場面で成立するかを検証することが挙げられる。

謝辞

本研究の機会を与えて下さった坂間保雄氏、東田正

信氏、大山芳史氏に感謝いたします。また、本研究の内容は様々な場での議論を通じて洗練されてまいりました。特に、CMC-95とSIG-SLUDの場で議論していただき、貴重なコメントをくださった多くの方々に感謝いたします。

◇ 参考文献 ◇

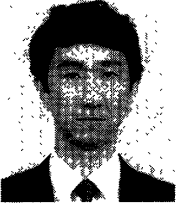
- [André 94] André, E. and Rist, T. "Referring to World Objects with Text and Pictures", *Procs. of COLING '94*, pp.530-534, 1994
- [Alshawi 87] Alshawi, H. "Memory and Context for Language Interpretation", Cambridge Univ. Press, 1987
- [Appelt 85] Appelt, D.E. "Planning English Referring Expressions", *Artificial Intelligence*, 26, pp.1-33, 1985
- [Clark 90] Clark, H.H. and Wilkes-Gibbs, D. "Referring as a Collaborative Process" in "Intentions in Communication" eds. Cohen, P. R., Morgan, J. and Pollack, M.E. The MIT Press, pp.463-493, 1990
- [Claassen 92] Claassen, W. "Generating Referring Expressions in a Multimodal Environment" in "Aspects of Automated Natural Language Generation" eds. Dale, R., Hovy, E., Rosner, D. and Stock, O., Springer-Verlag, pp.247-262, 1992
- [Cohen 84] Cohen, P.R. "The Pragmatics of Referring and the Modality of Communication", *Computational Linguistics*, Vol.10, No.2, pp.97-146, 1984
- [Cohen 95] Cohen, P.R. "Empirical Methods for Artificial Intelligence", The MIT Press, 1995
- [Dale 91] Dale, R. and Haddock, N. "Content Determination in the Generation of Referring Expressions", *Computational Intelligence*, Vol.7, No.4, pp.252-265, 1991
- [Feiner 90] Feiner, S.K. and McKeown, K.R. "Coordinating Text and Graphics in Explanation Generation", *Procs. of AAAI-90*, pp.442-449, 1990
- [Grosz 86] Grosz, B.J. and Sidner, C.L. "Attention, Intention, and the Structure of Discourse", *Computational Linguistics*, Vol.12, No.3, pp.174-204, 1986
- [Hoel 76] Hoel, P.G. "Elementary Statistics, 4th edition", John Wiley & Sons, Inc., 1976
「初等統計学」浅井晃, 村上正康 訳, 培風館, 1985
- [Ishikawa 94] Ishikawa, Y. "Communicative Mode Dependent Contribution from the Recipient in Information Providing Dialogue", *Procs. of ICSLP '94*, pp.959-962, 1994
- [Kato 95] Kato, T. and Nakano, Y.I. "Referent Identification Requests in Multi-Modal Dialogs" in *Procs. of Int. Conf. on Cooperative Multimodal Communication*, Vol.2, pp.175-191, 1995
- [Maybury 93] Maybury, M.T. "Planning Multimedia Explanations Using Communicative Acts" in "Intelligent Multimedia Interfaces", The AAAI Press / The MIT Press, pp.60-74, 1993
- [Neal 91] Neal, J.G. and Shapiro, S.C. "Intelligent Multimedia Technology" in "Intelligent User Interfaces", eds. Sullivan, J.W. and Tyler, S.W., ACM Press, pp.11-43, 1991
- [Oviatt 91] Oviatt, S.L. and Cohen, P.R. "Discourse Structure and Performance Efficiency in Interactive and Noninteractive Spoken Modalities", *Computer Speech and Language*, Vol.5, No.4, pp.297-326, 1991
- [Wahlster 91] Wahlster, W., André, E., Graf, W. and Rist, T. "Designing Illustrated Texts: How Language

Production Is Influenced by Graphics Generation”,
 Procs. of EACL '91, pp.8-14, 1991
 [Walker 94] Walker, M.A. “Experimentally Evaluating

Communicative Strategies: The Effect of the Task”,
 Procs. of AAAI '94, pp.86-93, 1994

〔査読者：竹澤寿幸〕

著者紹介



加藤 恒昭(正会員)

1959年生まれ。1981年東京工業大学電子電気工
 学科卒業、1983年東京工業大学総合理工学研究科
 電子システム専攻修士課程修了。同年、日本電信
 電話公社に入社。自然言語理解及び対話処理に関
 する研究に従事。現在、NTT情報通信研究所主幹
 研究員。電子情報通信学会、情報処理学会、ACL
 各会員。 <kato@nttnly.isl.ntt.co.jp>



中野 有紀子(正会員)

1965年生まれ。1988年東京女子大学文学部心理
 学科卒業、1990年東京大学教育学研究科教育心理
 学専攻修士課程修了。同年、日本電信電話(株)に
 入社。現在、同情報通信研究所に勤務。自然言語
 処理、対話処理、心理言語学の研究に従事。情報
 処理学会、日本心理学会、日本認知科学会、ACL
 各会員。 <yukiko@nttnly.isl.ntt.co.jp>