

音声対話における心的状態変化の予測をともなうメタ発話生成機構

菊池 英明^{†,††} 白井 克彦^{†††}

音声対話制御において、コミュニケーションの調整をはかるメタ的な発話の生成は重要な課題である。音声対話では、音声メディアの特性によって情報の授受とコミュニケーションの調整を同じチャネルで行うために時間のリソースを適切に分配する必要があるが、行為遂行中にも対話の状況や相互の心的状態が変化するためリソースの適切な分配は容易ではない。本研究では、ユーザの目標を協調的に達成するための計画立案の枠組みの中で、対話における相互の心的状態を認識・予測しながら、状況に適したメタ発話を生成する機構を提案する。具体的には、まず対話における状況を表現するための多次元心的状態モデルを定義する。多次元にすることにより各次元の独立性を高め、複雑な制御を可能にするとともにメンテナンス効率を向上させる。この多次元心的状態モデルを用いて認識・予測した心的状態を表す。次に、プランニングにおいて心的状態を参照しながらあらかじめ記述されたプランニングルールに従って目標を達成するために最適な行為を決定する機構を実現する。提案した機構を音声対話検索システムに導入し、評価実験を行った結果、対話の円滑さ向上に効果があることが確認できた。本稿では心的状態として発話権の状態およびシステム内部の処理状態を扱い、状態提示発話を適切に出力することを具体的に示したが、本機構では一般的な心的状態表現およびメタ発話行為を扱っている。

Mechanism of Generating Meta-utterance with Predicting Changes of Mental States in Spoken Dialogue

HIDEAKI KIKUCHI^{†,††} and KATSUHIKO SHIRAI^{†††}

In controlling spoken dialogues, it is important to generate meta-utterances which coordinate communication. To realize smooth conversation between human and machine, time resources should be appropriately divided into speakers because exchanges for problem-solving and coordinating communication are done in same channel. This is a feature of the speech dialogue medium. This research will propose the mechanism of generating meta-utterances with predicting changes of mental states in spoken dialogues. At first, we defined the model of multi-dimensional mental states for expressing situation in dialogues. By multi-dimensionality, independency of each dimension of states and efficiency of maintenance of them can be improved. Mental states, which are recognized and predicted, will be expressed by using this model. At next, we described the mechanism, which determine an appropriate action to achieve a goal by planning with the planning rules. These rules, in advance, are described with referring each dimension of mental states. By the preliminary experiments, we confirmed that the proposed model is efficient for improvement of smoothness of conversation and efficiency of problem-solving.

1. はじめに

音声による対話は人間同士の通常のコミュニケーション形態の1つであることから、音声対話処理技術は将来のマンマシンインタフェース技術の1つとして

期待される。マンマシンインタフェースのメディアとしての音声対話では、ユーザが自らの思考を妨げられることなく発話できることが望まれ、そのためには話題展開や発話タイミングなどの自由を保障する必要がある¹⁾。しかし、ユーザの行為に制約を与えなければ、その結果として問題解決が阻害されたり発話が重複してしまったりなど、対話の進行が滞るケースが増える。このような場合に、たとえば、問題解決の状況を説明したり、再発話を要求したり、発話の番を譲ったりするなどの、コミュニケーションの調整をはかるシステムのメタ的な行為が対話の円滑な進行に役立つ^{2),3)}。

† 国立国語研究所

The National Institute for Japanese Language

†† 早稲田大学人間科学部

School of Human Sciences, Waseda University

††† 早稲田大学理工学部

School of Science and Engineering, Waseda University

Bateson はコミュニケーションにおける参加者間のやりとりのうち、こうしたコミュニケーションの調整をはかるレベルのメッセージをメタメッセージと呼び、メッセージとメタメッセージの二重性はコミュニケーションの場につねに存在するとしている⁴⁾。

ここで、音声対話の形態は対話様式や目的などによってさまざまであるが、本研究では、音声対話処理技術の実現の容易さと市場の広さから電話会話のような音声のみによる目的指向対話に焦点をあてる。目的指向対話においてユーザの行為に制約を与えず、なおかつ円滑な進行をはかるには、目的を達成するための情報授受のやりとりを中心として、先に述べたコミュニケーションの調整をはかるメタ的な行為をシステムが適宜行う必要がある。たとえば画面を用いたユーザインタフェースでは、メタ的行為の1つである状態提示行為をメッセージの画面表示などで遂行できる。しかしながら、音声のみによる対話においては、単一のチャネルしか持たないメディアとしての特性により、メタ的行為を発話によって行わなければならない。その結果、情報の授受とコミュニケーションの調整を同じチャネルで行うために時間のリソースを適切に分配しなければならないという問題が生じる(単一チャネルにおける時間リソース分配の問題)。これには、発話行為の遂行に若干の時間を要するという音声対話メディアの別の特性も関わっている。これまでに我々は問題解決行為の合間を埋めるような状態提示発話出力が、全体の発話数や発話時間を増加させても効率評価の向上に有効であることを確認した⁵⁾。本研究ではこの成果をさらに一般化し、目的指向対話におけるさまざまな状況下でメタ発話を適切に出力する機構を構築することを目的とする。

従来、対話参加者の真の目標を協調的に達成するためのプランニングの枠組みがさまざまに提案されており⁶⁾、それらの多くは談話を階層的に構造化する中で発話行為の1つとしてメタ発話を扱う。コミュニケーションの調整をはかるメタ的発話行為を扱うために、行為レベルでの状況の変化に対応して再プランニングを行う従来のモデルを拡張し、発話権移行レベルの頻繁な状況変化に対応してプランニングを行うモデルも提案されている^{7)~9)}。ここで、音声対話でのメタ発話生成においては、先に述べたとおり単一チャネルにおける時間リソース分配の問題がある。発話権の所在が固定されている(あらかじめ時間リソースが対話参加者に分配されている)対話様式であれば、プランニングによって決定されたメタ発話を通常のやりとりの中に埋め込むことが容易である。しかしながら、発話権の移

動が自由な様式の場合には、発話権の所在や内外の状況の変化を予測しなければ通常のやりとりを阻害しないようにメタ発話を出力することが難しい。

本稿では、音声対話における状況変化の予測に応じてプランニングにより状況に適したメタ発話生成を行う機構を提案する。具体的には、対話における状況を複数の独立した次元の心的状態として表現し、その変化の予測をプランニングにあわせて行うことによってより状況に適した行為の決定を可能にする。以下、2章にメタ発話の例とその中の状況変化について示し、提案するメタ発話生成機構で扱う多次元的状態モデルを定義する。3章ではメタ発話生成機構について説明し、実際の音声対話システムに導入して評価を行う。4章では提案機構の一般性を議論する。

2. メタ発話と多次元的心的状態モデル

ここでは対話における状況を表現する心的状態モデルについて定義する。人間同士の会話における発話の機能を社会的にとらえようとする際、複数の側面について考察する手法が用いられる¹⁰⁾。また、人間の思考そのものが複数の心の競合もしくは協調によるものとする考え方がある^{11),12)}。本稿で扱う「メタ発話」は、先に述べた Bateson のメタメッセージを発話に特化したものであり、すなわちコミュニケーションの過程におけるさまざまな思考に関わるものである。したがって、メタ発話を含む発話行為のプランニングにおいて対話の状況をとらえるモデルとして多次元性を持った心的状態のモデルを用いることにする。

以下では、音声対話におけるメタ発話の具体例を示したうえで、本研究で扱う多次元的心的状態モデルを定義する。なお、本稿での議論は、以降、学術論文検索をタスクとした目的指向対話を例として行う。

2.1 メタ発話と心的状態変化

図1に論文検索対話の一例を示す。この例において、メタ発話は太字で示した S4, S5, S6 であり、いずれもシステムの処理状態を提示する役割を持つ。このうち、S5 と S6 を含むやりとりにおいてシステムがとらえるべき心的状態の変化を図2に示す。

図2において、心的状態としては仮に、ユーザ・システムそれぞれから見た発話権の状態(発話権がどちらにあるか)とシステム内部の処理状態を示した。

この図は、図1で示したようにユーザの検索要求にシステムが応じて検索処理を行っている間のやりとりであり、このやりとりが始まるまでの発話権はシステム側にあることを、ユーザ・システムともに共有信念として持っているとは仮定する。また、ユーザ発話 U4

S1: こんにちは
 U1: うん
 U2: 細菌学の論文なんだけど
 S2: はい
 U3: 炭疽菌の治療法についてあるかな
 S3: 炭疽菌と治療法で検索します
 S4: しばらく時間がかかるのでお待ち下さい
 <しばらく待ち状態>
 U4: あの一
 S5: もうすぐ検索が終わります
 <若干の待ち状態>
 S6: 検索が終わりました
 S7: 16件の論文が見つかりました
 S8: これでいいですか
 U5: はいありがとう

図 1 論文検索対話におけるメタ発話の例

(S はシステム, U はユーザの発話, 太字はメタ発話)

Fig.1 Example of meta-utterance in dialogue for academic paper retrieval.

(S: system, U: user, Gothic style: meta-utterance)

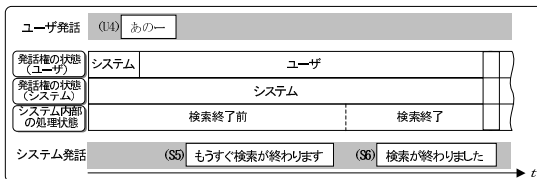


図 2 検索終了前後の状態提示発話

Fig. 2 Utterance for presenting states.

は検索処理の終了を待たずに検索条件を変更しようとするものと仮定する。

まず, ユーザ発話 U4 を発した時点で, ユーザにとっては発話権を獲得しようとしているため, ユーザ側からみた発話権の状態は「ユーザ」に移ったといえる。一方, システム内部の処理状態は図の点線で示す位置が「検索終了」のタイミングとして予測されるとき, U4 開始の時点では, システム内部の処理状態が「検索終了前」と予測されている。したがって, ユーザの大局的な目標達成のためには発話権を譲ることが得策でないと判断し, 代わりにして発話権獲得を撤回させるための状態提示発話 S5 を生成する。なお, この間, システム側からみた発話権の状態は一貫して「システム」のままである「検索終了前」であることを示す状態提示発話 S5 は, 煩わしさを抑えるために通常は行わないものであるとしたとき, この状況下に限ってこの発話を行う制御を実現するには, ユーザ・システムそれぞれからみた発話権の状態およびシステム内

部の処理状態を認識および予測しておく必要があるといえる。なお次に, 実際に検索処理が終了した時点で (図 2 では予測と同じタイミングで検索処理が終了したとしている), その旨を伝える発話 S6 を生成する。

ここでは検索処理終了の予測を例として説明したが, こうした内部状態のほかにも発話権状態, さらにユーザの注意, 積極性などの状態も心的状態の次元の一つとして同様に扱い, 認識および予測を行うことは状況にあったメタ発話の生成にとって有効である。

2.2 多次元心的状態モデルの定義

前節に示したように, 心的状態は複数の次元に存在しうる。対話制御機構において心的状態を多次元的に扱うことによって, 各状態が相互に作用しあうような複雑な制御が可能になるとともに, モデルとしてのメンテナンスが容易になるという利点がある¹³⁾。本研究で提案するメタ発話生成機構では, 複数の次元の心的状態を独立にモデル化し, プランニング時に利用する。以下には, まず多次元心的状態モデルの一般的な定義を行う。

まずはじめに目的指向対話を協調的問題解決の手段と位置づけ, プランニングについて以下の前提をもうける。なおここで示すプランニングは具体的な行為を決定するのではなく, プランを複数の行為の集合と見なし, 状況に応じた適切なプランを決定することを目的とする。

前提 1: 問題解決のゴールは複数のサブゴールに分割され, その構造はあらかじめ規定されない。

$$G = \{G1, G2, \dots, Gn\}$$

G : 問題解決のゴール

(ex. “会議室予約”, “論文検索”, “チケット予約”, ...)

G_i : ゴール G のサブゴール

(ex. G = “論文検索” の場合, $G1$ = “検索条件作成”,

$G2$ = “検索結果提示”, ...)

前提 2: サブゴールは 1 つ以上のプランの実行により達成される。

$$P_i = \{Pi1, Pi2, \dots, Pim\}$$

P_i : 問題解決のサブゴール G_i 達成のためのプラン群

P_{ij} : サブゴール G_i を達成するためのプランの 1 つ

これらの前提 1, 2 から, 1 つのゴールに対してサブゴールの系列は複数あり, さらに 1 つのサブゴールに対してプラン系列も複数ありうるといえる。たとえば, G = “論文検索” の場合, サブゴール系列は [“検索

条件作成”, “検索結果提示”, “検索条件修正”, “検索結果提示”...] のように, 検索条件をあまり吟味せずに結果を取得してから判断する方法と, [“検索条件作成”, “検索条件修正”, “検索結果提示”, ...] のように, 検索条件をよく吟味してから検索を行う方法がありうる. 前提 1 は, このようなサブゴール系列を問題解決の前に決定するのではなく, 変化する状況に応じて修正可能であることを示す. 同様に前提 2 より, あるサブゴールを達成するためのプラン系列も状況に応じて修正可能とすることにより, 1 つのゴールに対するサブゴールおよびプラン系列は状況に応じて柔軟に修正することが可能となる. 具体的なサブゴール系列およびプラン系列の構成方法については, 3.2 節の動作例の説明で述べる.

以下に, 複数次元の心的状態モデルを定義する.

$$S = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_n \end{pmatrix}$$

$$S_p = Sp1, Sp2, \dots, Sp_s$$

S : 多次元心的状態の集合

S_p : p 次元目の心的状態の集合

(ex. “発話権の状態”, “処理状態”, “情報授受の状態”, ...)

Spq : p 次元目の心的状態の 1 つ

(ex. Sp = “発話権の状態” の場合, $Sp1$ = “ON”, $Sp2$ = “OFF”)

なお, 各状態の認識方法はその種類によってさまざまであるが, 他の次元に対して依存関係を持つべきではない. つまり, 何らかの障害により 1 つの次元の状態が認識不能となっても対話を続けることができるようにすべきである.

先に述べたように, 状況に適したメタ発話の生成には, 心的状態変化の予測が必要になる. ここまでの説明で示したように, 心的状態を離散的な状態系列として表現するために, 予測可能な状態変化の過程を離散的な状態の 1 つとして加える方法 (ex. Sp = “発話権の状態” の場合, 発話権譲渡を予測して Spq = “OFF 直前” など) を用いる.

以上から, 複数の行為の集合であるプランはゴール G と, それに対応したプラン群 P および, 多次元心的状態 S を用いたプランニングによって決定されるといえる. 次章では, ここで示した心的状態モデルを導入したメタ発話生成機構を説明する.

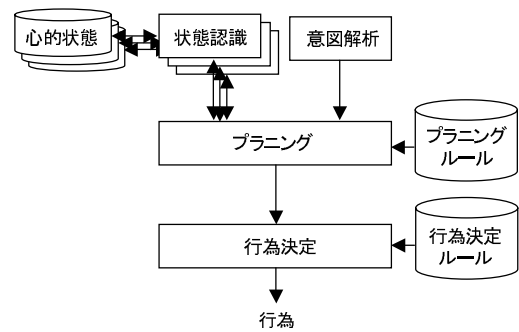


図 3 メタ発話生成機構

Fig. 3 Mechanism of generating meta-utterance.

3. メタ発話生成機構

本章では, 前章で示した多次元心的状態モデルを導入し, 対話参加者の多次元心的状態予測をともなうプランニングによって, 状況に適したメタ発話を生成する機構を説明する.

3.1 構成

提案するメタ発話生成機構の構成を図 3 に示す. 図 3 の各部の処理について以下に述べる.

(1) 意図解析 ユーザ発話についての表層的な意図を解析する. ユーザ発話の音声終端検出を契機に処理が開始される.

(2) 状態認識 次元ごとに状態の変化を認識・予測する. 具体的には, ユーザ発話の音声始終端検出やデータベース検索処理終了など, 心的状態変化に関わるイベントをつねに監視して, 心的状態の値をリアルタイムに更新する. さらに韻律パターン照合によってユーザ発話の音声終了を予測したり, 後述するように内部処理終了を予測したりするなどによって, 状態変化の予測を行う. ここで状態変化の予測を行う点が本機構の特徴である. 状態変化が検出された場合にはプランニング部に信号を送る.

(3) プランニング あらかじめ記述されたプランニングルールを用いて, 意図解析処理の結果あるいは状態認識による状態変化検出信号が得られたときにプランニング処理を開始する. プランニングの結果として, 行為群に相当するプランが決定される.

(4) 行為決定 プランニング処理によるプラン決定を受けて, 対応する行為決定ルールを用いて, 具体的な行為を決定する.

3.2 動作例

以下には, 2 章の図 2 で示したやりとりを実現するためのプランニングルールとそれを用いた処理の説明を行う. 心的状態としては, ユーザ・システムそれぞれ

```

(Rule1)SubGoal("検索条件作成")
&& Act(Process("データベース検索"))
&& Act(Plan("検索結果提示"))
⇒ SubGoal("検索結果提示")
(Rule2)Ready_SubGoal("検索結果提示")
&& [Sprocess = "検索終了"]
&& Act(Plan("検索結果提示"))
⇒ SubGoal("検索結果提示")
(Rule3)[Sprocess = "検索終了前"]
&& Change_State(Susr_turn, "ON", "OFF")
⇒ Ready_SubGoal("検索結果提示")
(Rule4) [Susr_turn = "ON"]
&& [Ssys_turn = "ON"]
&& [Sprocess = "検索終了前"]
&& Act(Plan("状態提示"))
⇒ Change_State(Susr_turn, "ON", "OFF")
    
```

図4 心的状態を参照するプランニングルールの例

Fig. 4 Example of planning rules referring mental states.

の発話権の状態を S_{usr_turn} , S_{sys_turn} とし, システム内部の処理状態を $S_{process}$ として導入する. 発話権の状態は値として "ON", "OFF" をとる. 処理状態は "検索終了", "検索終了前" など, あらかじめ定義された内部処理の離散的な状態を値としてとる.

図4に, 状態提示発話 S_5 を生成するのに必要なプランニングルールを示す. " " の前部が条件部, 後部が結論部であり, "&&" は AND を示す. 各節は命題を示すが, 斜体で示した Act() は手続きをとる命題である. プランニング処理ではあらかじめこれらの標準的な命題に対する振舞いが定義されている. なお, ルールは後ろ向き推論を仮定しており, まずすべてのルールの結論部が評価され, 意図解析により得られたユーザのゴールと一致するルールが選択される. 次に, 選択されたルールの条件部のうち最初の条件から順に評価される. 評価された命題が真でない場合にはその命題を結論部を持つルールが選択される. 条件部の命題がすべて真となった場合に結論部の命題が真となる. なお, 評価されていない命題については便宜上, 偽と同様に扱う.

図4において, Rule1は通常の問題解決に関するやりとりを遂行するためのルールであり, サブゴール "検索結果提示" を達成するためには, サブゴール "検索条件作成" の達成と, 処理 "データベース検索" の実行が必要であることを示している. それに対して Rule2 から Rule4は, 処理 "データベース検索" の終了待ち状態のユーザ割込みに対して状況に応じてメタ発話を出力するためのルールである. Rule2は, サブゴール "検索結果提示" を達成するためには, Ready_SubGoal("検索

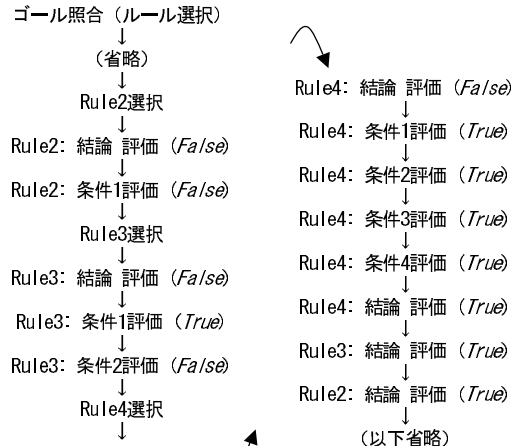


図5 プランニング過程の例

Fig. 5 Example of planning process.

検索結果提示)を達成して, なおかつシステム内部の処理状態 $S_{process}$ が "検索終了" となる必要があることを示す. なお, Ready_SubGoal() はサブゴール達成前の状態に相当する. Rule3は, Ready_SubGoal("検索結果提示")を達成するためには, システム内部の処理状態 $S_{process}$ が "検索終了前" であり, なおかつユーザの発話権の状態 S_{usr_turn} を "ON" から "OFF" にする必要があることを示す. なお, Change.State() は指定された状態の値を変更するという命題である. Rule4は, ユーザの発話権の状態 S_{usr_turn} を "ON" から "OFF" にするには, ユーザとシステムの発話権の状態がともに "ON" であり, システム内部の処理状態 $S_{process}$ が "検索終了前" であり, なおかつ Plan("状態提示")すなわち状態提示プランが遂行される必要があることを示す.

なお, システム内部の処理状態 $S_{process}$ が "検索終了前" となるのは, 状態認識処理における状態変化予測で検索終了時期を推定した結果である. 処理の終了時期の推定は処理時間の統計量に基づいて行う⁵⁾ことを想定している. また, "検索終了前" 状態の始端位置は, その区間に状態提示発話を生成する可能性があるのだから少なくとも発話時間長の分を確保するように決定する必要がある.

以下には図5を用いて, 具体的なプランニング過程を示す. ユーザ発話の意図解析結果からゴールを "特定論文検索" と判断してプランニングによりそのサブゴールとして "検索結果提示" が設定されたとする. 次に, Rule1とRule2により, サブゴール "検索結果提示" を達成するための, 2つのプラン系列が候補として生成される.すでにRule1により生成されたプラン系列

表 1 メタ発話生成機構の評価結果

(S) はシステム発話, (U) は被験者発話を計測の対象としていることを示す)

Table 1 Evaluation result of proposed model.

((S) indicates that the item was counted in system's utterances, and (U) indicates subjects' utterances)

メタ発話制御	総発話 (U)	検索要求 (U)	メタ発話 (S)	重複発話 (割合)	円滑さ	
(1) なし	47	31	0	2 (4.2%)	2.2	
あり	(2) 常時出力	92	49	38	11 (11.9%)	3.1
	(3) 本手法	62	42	35	3 (4.8%)	3.8
計	201	122	73	16 (7.9%)	3.0	

が選択および遂行され, Rule1 に条件として記述された Act が遂行されていても, それが完了しない限りは Rule2 から生成されたプラン系列が選択される。プラン系列の選択過程は図 5 に示したように, Rule2 選択の後, 結論部が評価され False であるため Rule2 の条件 1 の命題 “Ready_SubGoal” が評価される。これが False となるためこれを結論部に持つルール Rule3 が選択される。同様にしてこのルールの条件を満たすようにルールを選択した結果, $[S_{usr_turn} == \text{“ON”}]$, $[S_{sys_turn} == \text{“ON”}]$, $[S_{process} == \text{“検索終了前”}]$ の状態でプラン “状態提示” を遂行することによって, 結果としてサブゴール “検索結果提示” が満たされることになる。このようにして, 論文検索処理実行中にユーザからの発話があった場合, Rule2 から Rule4 で決定されるプラン系列の 1 つとして状態提示プランを遂行できることになる。なお, ここで示したルールは, 論文検索処理終了をユーザが待ちきれずに発話した場合のメタ発話出力を実現するものである。ただしあいづちや無意味語などに対しては持続時間の短さから発話権状態更新を行わないためメタ発話は出力されない。また, ゴールを変更するような発話に対しては発話意図が解析されてプランニングにより新たなサブゴール系列・プラン系列が決定されることになる。

行為決定処理で用いる行為決定ルールについてはイベントをアーク, 行為をノードとする有限状態オートマトンによる記述¹⁴⁾を想定しているが, 本稿の範囲では, 行為としてメタ発話を対象としており, プランと行為の関係がほぼ一定であるため, 具体的な行為そのものの形式は特に言及しない。

3.3 実用性評価

3.1 節で示したメタ発話生成機構と図 4 のルールを含む 17 のプランニングルールを音声対話システム汎用プラットフォーム¹⁴⁾に導入してその実用性評価を行った。具体的には, 12 名の被験者にあらかじめ 3 種類の文献を与え, それぞれの理解に必要な論文をシステムとの音声対話によって検索するよう要求した。なお, 被験者のうち 4 名のみが音声インタフェースの使用経験があり, 念のために全被験者に数回の練習を実験

前に行わせシステムの使い方に慣れさせた。心的状態としては発話権状態とシステム内部の処理状態を導入し, プランニングルールにおいてはシステム内部の処理状態を提示するメタ発話を生成するためのルールを加えた。システム内部の処理としては, 音声認識, 意図解析, データベース検索を対象とした。

各被験者が行う 3 対話において, 各対話によってメタ発話制御方法を, (1) なし, (2) 常時出力, (3) 提案手法と変え (順不同), それぞれの対話における被験者の総発話数, 検索要求数, 重複発話数とシステムのメタ発話数を調べた。なお, (2) の常時出力の場合は提案手法によって出力する可能性のあるすべてのタイミングでメタ発話を出力した。また, 実験後のアンケートにより対話の円滑さを被験者に 5 段階で評価させた。以上の結果を表 1 にまとめる。

表において, 円滑さ評価の平均がメタ発話出力なしの場合の 2.2 から常時出力あるいは本手法による出力によって向上している (t 検定, $p < 0.01$)。この結果から, メタ発話の出力が円滑さの評価向上に有効であるといえる。なお, 表においてメタ発話出力によってユーザ発話数および検索要求数が大幅に増加している。状態提示メタ発話により無駄なやりとりを回避するという当初予想した効果は認められないが, メタ発話によって対話が円滑になった結果, 対話によって目的を達成しようとするやりとりが増えたものと考えられる。

また, メタ発話を常時出力する場合に比べ, 本手法を適用して問題解決の進行を考慮することによって検定による有意差はないものの ($\chi_2(1) = 0.13$) 重複発話の数が大幅に減った。さらに, 発話数や検索要求数が減っているが, 収録対話の観察から, 重複から復帰するための発話や重複により遂行されない検索要求の数が減ったことによると考えられる。

以上から, 音声対話におけるメタ発話生成において, メタ発話の出力によって対話を円滑にし, さらにメタ発話出力方法として本手法を適用することで, つねにメタ発話を出力するよりも効率良く問題解決を進めることが可能になることを確認した。

4. 一般性についての考察

ここでは、これまでに述べたメタ発話生成機構についていくつかの側面から一般性を考察する。

4.1 心的状態

本稿ではユーザ・システムそれぞれの発話権の状態と、システム内部の処理状態を心的状態としてとりあげた。しかし、2章で示した多次元心的状態モデルは特にこれらを規定するものではない。

なお、心的状態の値としては離散的な値だけでなく連続量を取りうる。たとえば、認識した音声の尤度を、受理した発話の信頼度についての心的状態として用いるなどが考えられる。ただし、心的状態の値が連続量である場合、プランニングにおいては状態の値に対して比較演算やマッチング処理などを可能にする必要がある。

4.2 心的状態変化の予測

本稿では、システム内部の処理状態のみを予測の対象としてとりあげたが、発話権の状態についても文脈やユーザモデルによってある程度予測することができると思われる。

なお、状態変化の予測の難しさは、対象とする心的状態の種類によって異なる。現時点では、状態遷移のパターンや状態滞在時間の分布がある程度既知である場合を想定している。

4.3 メタ発話

プランニングにおいて、心的状態を参照して決定するプランとしては状態提示のみを示したが、他のメタ行為についても同様にプランニングの対象とすることができる。たとえば、前述した発話の信頼度についての心的状態を導入することで、再発話要求を促すプランを用意することも可能になる。また、発話権の状態変化を予測することで、ユーザ発話の合間にあいづちのような調整的行為を遂行することも可能になる。

4.4 プランニング

状態認識処理を常時行い、状態変化に応じて再プランニングを行う本機構により、プラン遂行中に、前提としていた状態の値が変化してしまうような場合、その状態に適した再プランニングが実現できる。これによって、対話における頻繁な状況の変化に対応可能になったといえる。

また本稿では、発話生成機構をプランニングと行為決定に分けて、それぞれ異なるルールを参照するようにした。これは、状態変化に応じてプランニングを行うことで、処理の頻度が著しく増加することを考慮して、プランニングの処理を減らすために反射的な制御を行為

決定処理にまかなわせたためである。状態変化に応じてプランニングを行った結果、プランの変更がなく行為決定を必要としないケースは非常に多いため、このように処理を分けることは効率を大幅に向上させる。しかしながら、2種類の独立したルールが同じ心的状態モデルを共有するのは、メンテナンス上の効率を下げる。メンテナンス効率を向上するルール記述形式あるいはアーキテクチャを検討する余地があるといえる。

また、本稿では電話会話のような音声のみによるコミュニケーションの形態を想定して議論してきたが、音声以外のメディアを利用できる環境においては発話の代替となる手段を用いることができる。プランニングにおけるこうした環境変化の考慮は今後の課題といえる。

4.5 システムの汎用性

本研究は目的指向対話におけるさまざまな状況下でメタ発話を適切に出力する機構の構築を目指している。目的指向対話のようにある程度対話の流れが定まっている対話においては、本機構のルールベースの手法は有効である。すなわち、そのような対話においては通常のやりとりを構造化したうえでコミュニケーションの調整行為を位置づけられる。しかしながら、対話の流れの予測がつかない場合、通常のやりとりを構造化できないためにメタレベルのやりとりを位置づけることは困難である。その場合にはむしろプランニングとは独立した枠組みとすることが有効と考える。

5. おわりに

本研究では、ユーザの目標を協調的に達成するためのプランニングの枠組みの中で、対話における相互の心的状態を認識・予測しながら、状況に適したメタ発話を生成する機構を提案した。具体的には、多次元の心的状態モデルを定義し、その状態変化をつねに予測する機能を設け、あらかじめ状態の予測結果を参照するよう記述したルールをプランニングにおいて扱うことを可能にした。提案した機構を音声対話検索システムに導入し、評価実験を行った結果、対話の円滑さと効率の向上に効果があることが確認できた。ここで示した機構は、先の研究⁵⁾における状態提示発話の制御モデルを、一般的なメタ発話制御に拡張したものである。考察の結果、心的状態の定義およびメタ発話の一般性はあるが、状態変化の予測およびプランニング処理についてはいくつか課題が残るといえる。

参考文献

- 1) 菊池英明, 工藤育男, 小林哲則, 白井克彦: 音声

- 対話インタフェースにおける発話権管理による割込みへの対処, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1502-1511 (1994).
- 2) DISCプロジェクト報告, <http://www.disc2.dkslidsなど>
- 3) Bell, L. and Gustafson, J.: Interactive Dialogue in Multi-Modal Systems, *Proc. ESCA workshop on Interactive Dialogue in Multi-Modal Systems (IDS'99)*, Kloster Irsee, Germany (1999).
- 4) Bateson, G.: *Steps to an ecology of mind*, Chicago University Press (1972).
- 5) 菊池英明, 白井克彦: 対話効率の向上を目的とした音声対話制御のモデル化, *ヒューマンインタフェース学会誌*, Vol.2, No.2, pp.145-152 (2000).
- 6) Allen, J.F., Hendler, J. and Tate, A. (Eds): *Readings in planning*, Morgan Kaufmann (1990).
- 7) Cohen, P.R.: The pragmatics of referring and the modality of communication, *Computational Linguistics*, Vol.10, No.2, pp.97-146 (1984).
- 8) Clark, H.H. and Wilkes-Gibbs, D.: Referring as a collaborative process, *Cognition*, No.22, pp.1-39 (1986).
- 9) Novick, D.: Controlling interaction with meta-acts, *Conference on Human Factors in Computing Systems (CHI'91)*, New Orleans, USA (1991).
- 10) Hymes, D.: Models of the interaction of language and social life, *Directions in Sociolinguistics*, Gumperz, J. & Hymes, D. (Eds.), New York, Holt, Reinhart & Winston (1972).
- 11) M. ミンスキー: 心の社会, 産業図書 (1985).
- 12) 山鳥 重: 脳からみた心, NHK ブックス 482, 日本放送協会 (1985).
- 13) 鈴木堅悟, 青山一美, 菊池英明, 白井克彦: 多次元的状態を扱う音声対話システムの構築, 情報処理学会研究報告, 2001-SLP-37, pp.13-18 (2001).
- 14) 青山一美, 平野 泉, 菊池英明, 白井克彦: 音声対話システム汎用プラットフォームの検討, 情報処理学会研究報告, 2000-SLP-30, pp.7-12 (2000).
- (平成 13 年 11 月 19 日受付)
(平成 14 年 4 月 16 日採録)



菊池 英明 (正会員)

1991 年早稲田大学理工学部電気工学科卒業。1993 年同大学大学院修士課程修了。同年(株)日立製作所入社。1997 年同退社。同年早稲田大学大学院後期博士課程入学。2000 年同単位取得退学。同年より国立国語研究所特別奨励研究員。2002 年早稲田大学人間科学部非常勤講師。ヒューマンインタフェース, 音声対話の研究に従事。情報処理学会平成 8 年後期大会優秀賞受賞。人工知能学会, 電子情報通信学会, 日本音響学会各会員。



白井 克彦 (正会員)

1963 年早稲田大学理工学部電気工学科卒業。1968 年同大学大学院博士課程修了。工学博士。同年同大学理工学部講師。1970 年同助教授。1975 年同教授(電気工学科)。1991 年同教授(情報学科)。音声認識・合成技術, 自然言語処理, CAI 等を中心にヒューマンインタフェースの研究に従事。1998 年人工知能学会会長。日本音響学会, IEEE 等各会員。電子情報通信学会フェロー。