

特集 車載マルチメディアにおける対話エージェントについて*

Interactive Agent in In-Vehicle Multimedia System

笹木美樹男 浅見克志
Mikio SASAKI Katsushi ASAMI

This paper presents a concept of interactive media communication under mobile environment, especially for car drivers. We propose In-Vehicle Profiles, which can describe various dynamic situations of users. Moreover, we have proposed a new architecture of system including interactive agent, who communicates with drivers so as to retrieve and display every desired content.

Key words : In-Vehicle Profile, Agent, Speech Dialogue, Scenario Interpreter

1. はじめに

近年、カーナビに代表される車載情報機器のユーザインタフェースにおいて音声入力が増加しつつある。しかし、現状で扱える情報メディアは限定されており、多様なデータベースへの動的検索要求を運転中に安全かつ素早く抽出することは難しい。そこで本稿では、車室内ユーザの環境・状況・要求・状態を記述する車載プロフィールを定義し、エージェントとの円滑な対話によりメディアを通信する構成を提案する。

2. メディアの検索

本稿で考えた次世代システムをFig. 1に示す。ここでは、車載端末中の対話エージェントがあたかも助手席のナビゲータのごとくユーザの環境・状況・要求・状態を把握してプロフィールに収め、センターに自動送信する。これによりユーザの要求は新たな発話やスイッチ操作を行う以前に概略がセンターに伝わっている。ここでセンターとは通信で接続するものに限らず、特にリアルタイム情報を要求しなければDVDのように車室内の大容量記憶を制御する装置であってもよい。

一方で、対話エージェントはユーザとの対話で獲得した口コミ情報（映像、音声を含む）も位置や時間の情報と共にプロフィールに格納して送信する。これを各ユーザの対話エージェントが実行することにより、センター上のデータベースは増殖する。

以上の二つの特徴により、ユーザは即時的な入力作業に煩わされることなく、津々浦々の映像・音声を車室内で安全・快適に獲得することが可能になる。これを対話型メディア通信と呼ぶことにする。対話型メ

* 日本人間工学会の了解を得て、シンポジウム「カーナビ・携帯電話の利用性と人間工学(2000.2)」より一部加筆して転載

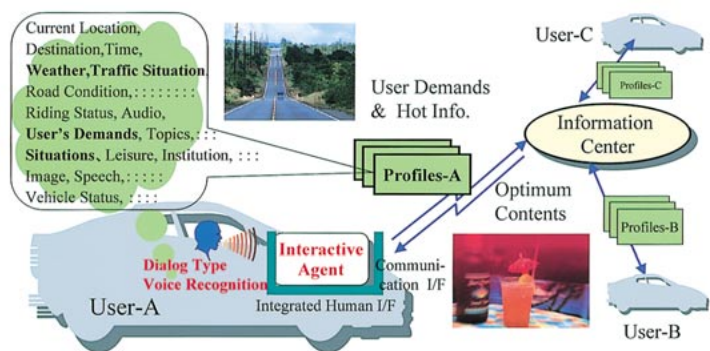


Fig.1 Interactive media communication

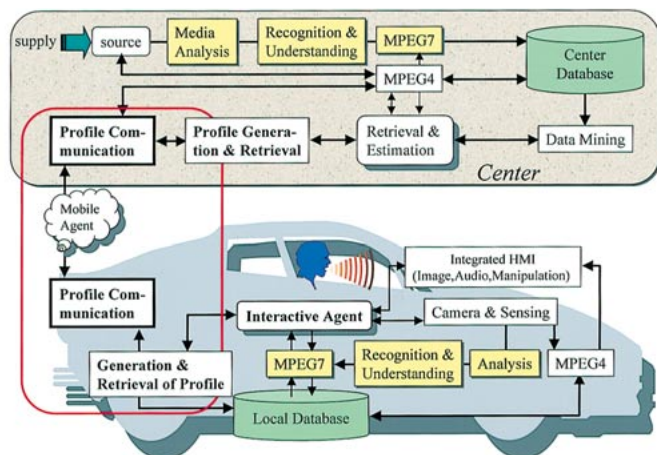
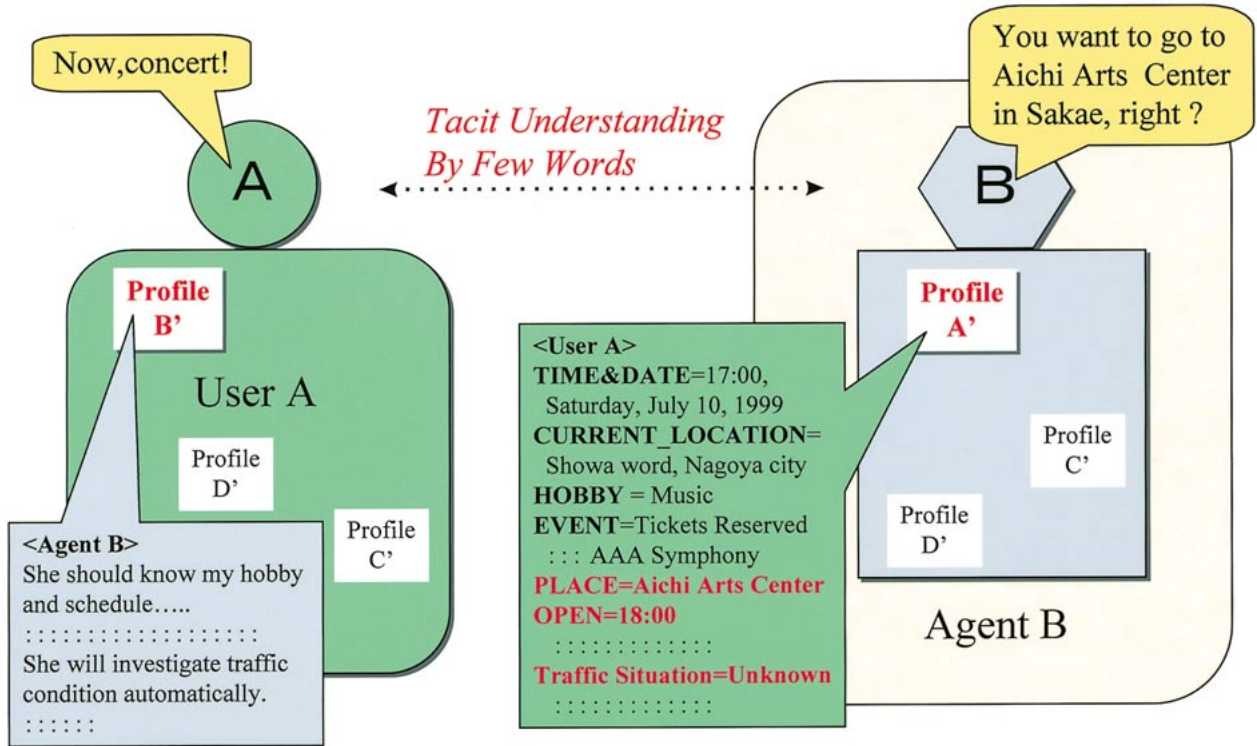


Fig.2 Interactive media communication system

ディア通信システムをより詳細に記述するとFig. 2のようになる。

このような対話エージェントはユーザと以心伝心の関係にあることが望まれる (Fig. 3)。すなわち、対話エージェントが秘書のようにユーザの要求を推定す



Agent tries to get the accurate profile of user, and activates necessary portions adaptively.

Fig.3 Interactive agent

ることができれば、ユーザは一言で意図をシステムに伝えることができる。

例えば、ユーザAが対話エージェントBに「今からコンサート！」と言いつたときに、エージェントBが「栄の芸術文化センターですね」と確認の発話をして、渋滞状況の検索を始める。この以心伝心の関係を構築するには、ユーザAの現在位置、時間、スケジュールなどのダイナミックな情報と、住所や勤務先、趣味、その他のスタティックな情報のまとめ（プロフィール）を把握しておく必要がある。

一方、ユーザAも「エージェントBは自分の趣味やスケジュールを把握しており、必要な情報は自動的に検索してくれるはずだ」というエージェントBのプロファイルが頭の中にあるがゆえにそのような発話ができる。そこで、「互いに相手のプロフィールを正確に獲得し、必要部分を適応的に活性化している状態」が究極のヒューマンインタフェースの条件と考え、これを可能にする対話エージェントを開発する。

3. 車室内の対話生成モデル

車室内の対話生成を我々は、Fig. 4のようにモデル

化した。すなわち、ユーザの発話に対するエージェントの応答は、車外環境・状況、時間・空間、コンテンツ情報、車室内の状況、ユーザの要求・状態を把握することで対話進行プロセスの中で決定される。

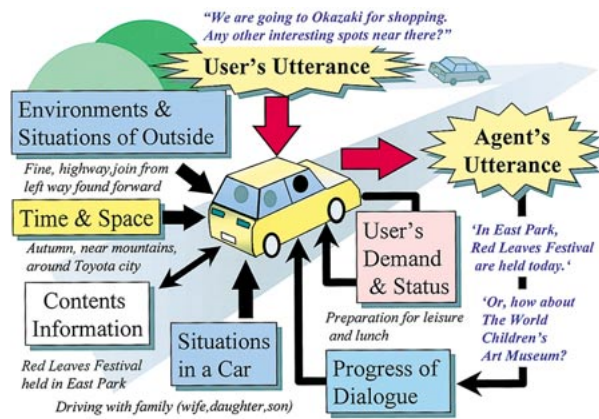


Fig.4 Model of generating dialogue in a car

例えば、ユーザの要求が目的地付近での行楽と食事の準備であるとわかっていれば、エージェント発話はおのずから限定される。人間は物理的に状態が変化するものであり、長時間ドライブの後には当然疲労し、

何も食べなければ空腹になる。また、一人でドライブする場合と家族でドライブする場合では、当然経路地や目的地での要求は変わってくる。さらには同じ時間空間情報でもそれを季節や曜日・時間帯として解釈したり、山や海の近くなのか、街中なのかを把握することでエージェントはユーザの発話意図を容易に理解できる。

このようにドライバに適応して対話進行や発話内容を変えるには、動的に個人情報や記述するデータ記憶構造が必要となる。一方で、発話対を含むこれらの入出力変数は状態更新を伴うシステムで表現される (Fig. 5)。すなわち、あるユーザ発話 $X\$$ は話題、時間・位置、環境・状況、要求・状態などの付帯条件 US のもとでなされる。そこで、対話エージェント内では $X\$$ がトリガーとなって US が同定され、エージェント発話、制御出力、予測要求・状態などが出力される。この関係を以下の式で表現する。

$$AO = F(US, X\$) \quad (1)$$

但し、

$$US = (Topic, T, P, E, S, D_n, Q_n) \quad (2)$$

$$AO = (Y\$, C_n, D_{n+1}, Q_{n+1}) \quad (3)$$

である。ここで、

AO : システム出力

US : 付帯条件

$Topic$: 話題

T : 時間

P : 位置

E : 環境

S : 状況

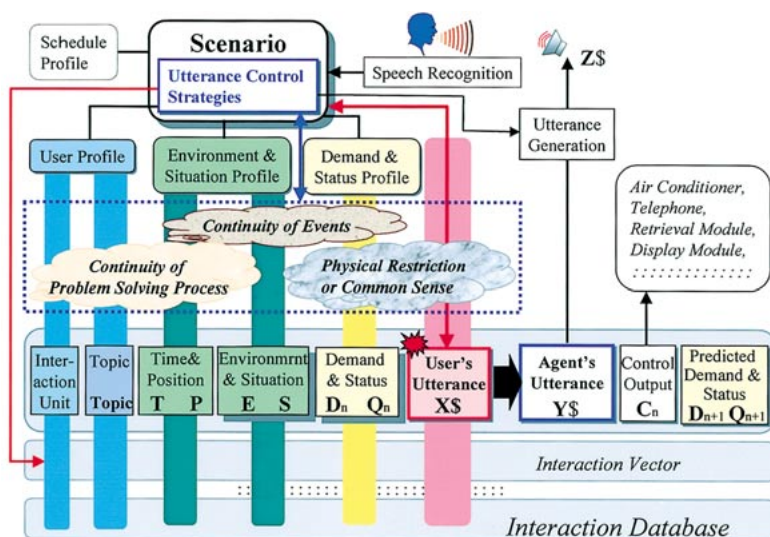


Fig.5 Interaction database and scenario

$X\$$: ユーザ発話

$Y\$$: エージェント発話

C_n : 時刻 n における機器制御出力

D_n : 時刻 n におけるユーザ要求の推定値

Q_n : 時刻 n におけるユーザ状態の推定値

n : 時間順序によるユーザ発話の通し番号

この一連の入出力変数群 ($US, X\$, AO$) を対話ベクトルと呼ぶことにする。対話ベクトルは US の様々な組み合わせに対して車室内といえどもほぼ無限に発生しうするため、これに対応する対話データベースを機械上にすべて持つておくことは困難である。そこで、この対話データベースを縮約化するために以下のことを考えた。

- (1) 様々な現実的制約を規則として用い、付帯条件、ユーザ発話およびシステム出力の諸属性で構成される多次元空間をクラスタリング (代表的な有限個の部分空間に分割) できる対話ベクトルの表現形式を車載プロフィールという形で定義する。
- (2) 対話ベクトルの軌道を制御する手段としてシナリオを定義する。
- (3) ユーザプロフィールやスケジュールなどの個人情報を積極的に活用する。

4. 車載プロフィール

Fig. 6に車載プロフィールの構成を示す。このうち、車室内ユーザモデルに対応するドライバの記述は下記の三つで構成される。

(1) 環境・状況プロフィール

現在地、目的地、天候、交通状況、道路状態、行楽情報、施設案内、映像・音声、検索要求、乗車状態などが記述される。これらは中期記憶または短期記憶に相当し、動的かつ不足項目を多く含む。

(2) 要求・状態プロフィール

ユーザの動的な要求や状態の推定結果を記述するが、対話エージェント内のユーザモデルの正確さに影響されるため誤同定を多く含む。中期記憶または短期記憶に相当し、動的に変化する属性を記述する。

(3) ユーザプロフィール

ユーザの比較的安定した個人情報を記述する。長期記憶に相当し、静的に近い定常的属性を記述する。属性値の

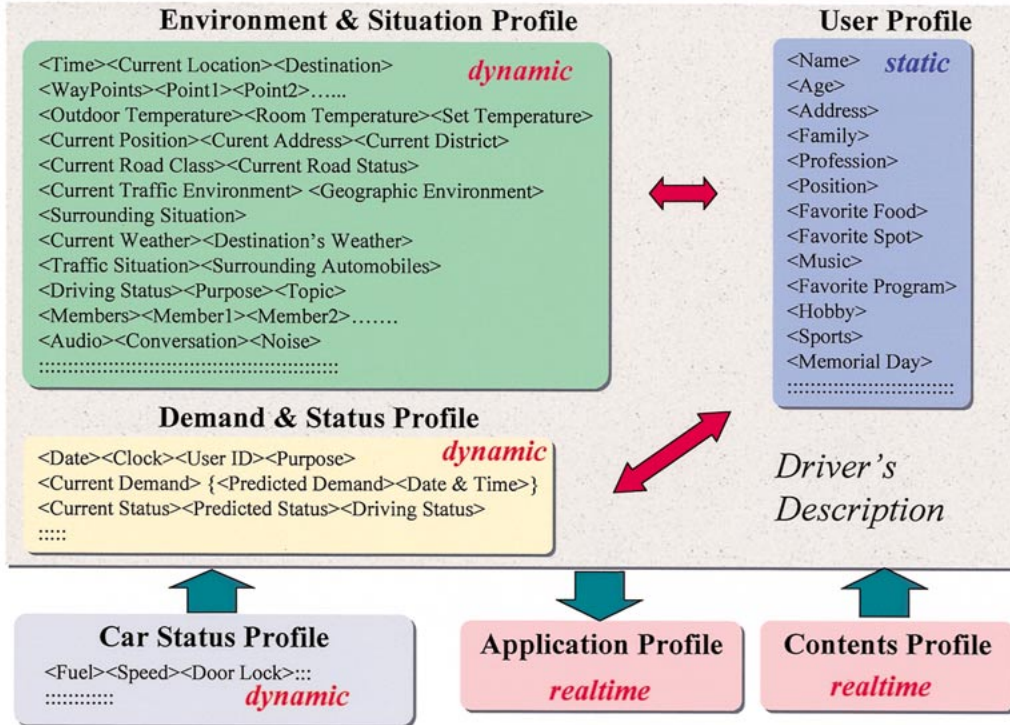


Fig.6 Configuration of car profiles

誤りは少ないが、常にUSと相関が高いわけではない。

これらのプロフィール中の項目はすべてが充填されていることはまれであり、実際上は大半がblankのままでは対話が始まる。話題から活性化された項目で未充填のものは、次の手段を駆使して充填することができる。

- a) ユーザへの問い合わせ
- b) 過去のプロフィールを参照
- c) 他ユーザのプロフィールを参照
- d) 事象の連続性
 - 例) 走行中の車室内で人の増減はない。
- e) 問題解決プロセスの連続性
 - 例) レストラン検索に際して、検索結果を提示するまで話題を終了しない。
- f) 物理的・常識的制約
 - 例) 車は海を走ったり、空を飛んだりしない。
 - 例) 休まずに長時間運転を続けると疲労する。
 - 例) ひどい渋滞ならば快適ではない。
- g) 時間の周期性
 - 例) スキーは冬、花見は春がシーズンである。

- 例) 朝夕には通勤ラッシュがある。
- h) 空間構造の類似性
 - 例) 駅前には店や銀行、公共施設が多い。
 - 例) 高速道路の入出にはインターチェンジや料金所を通る。

ユーザの時空間上の軌道と各プロフィールの関係を Fig. 7 に概念的に示す。これをもとに、上記のd) ~ h) に関する予測・推定機構は次の形で設計することができる。

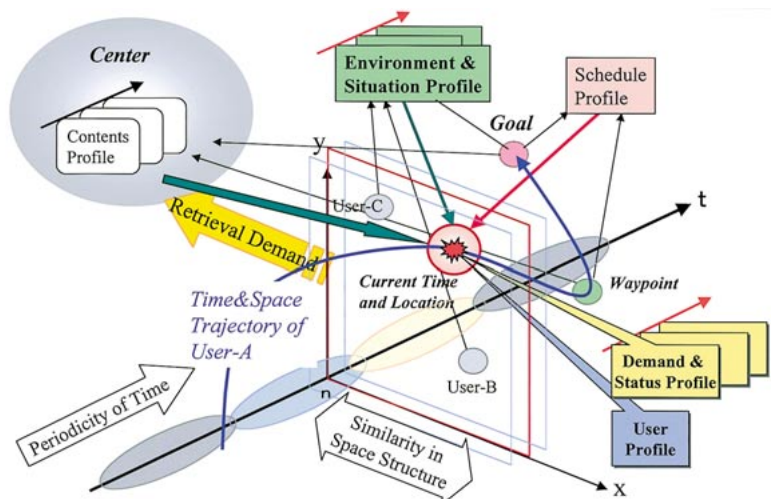


Fig.7 Restriction of profiles on user trajectory

異なるプロフィール間の制約
あるプロフィールについての時間的連続性
プロフィール内部の項目間の制約

これらは特に車室内ユーザモデルにおける状態変化とシーンの推定を行う部分で演算される。現在、車載に適した意味ネットワーク、実世界記述、規則データベースとして整備を進めている。

5. 要求推定

車室内において対話エージェントが現在のユーザ要求を推定し、これからの要求を予測できれば、下記の機能が飛躍的に使いやすくなる。

コンテンツ検索(地図, 施設, レストラン, ゴルフ場, JR, スーパー, 公園, 他)

機器制御(エアコン, ヒータ, ワイパ, ドアロック他)

メディア制御(カーナビ画面の各種操作, 画面遷移, コマンド, オーディオ, TVその他)

Fig. 8に要求推定の動作例を示す。今、ユーザが「岡崎, めし, インド料理」というキーワード列を入力した際、<場所, 要求キーワード, 要求細分化の語彙>という組み合わせで認識され、「岡崎にあるインド料理屋で食事したいが、どこがよいか」という意味で解釈され、記憶媒体や通信手段によりデータベースを検索する。このとき、該当するインド料理屋のみならず、環境・状況(季節, 日時, 天候, 現在位置, 周辺状況)や目的地(インド料理屋)にいたる経路に応じて、渋滞情報, 目的地周辺のイベント情報やショッピングゾーン, 公園, 行楽地などを紹介する。

ユーザの現在状態が空腹であるならば、食事後は満腹になると予測できる。そこで、この満腹の状態に対

応する要求として公園や買い物などが予測されて出てくる。さらには、ユーザプロフィールに基づいて「そういえば、明日は娘さんのお誕生日でしたね。」と話しかけ、予測要求に対する反応を求める。

これらは以下のタイプの複合で生成される。

- (1) 一次要求
ユーザが発話した要求キーワードから推定した要求。
- (2) 関連要求
ある要求に関連してユーザ状態を介さずに導出される要求。
- (3) 二次要求
推定された現在状態に推論規則を適用して導出される要求。
- (4) 状態予測要求
現在状態から予測した未来状態に推論規則を適用して導出される要求。
- (5) 状況推定要求
現在の環境・状況から推定される要求。異常事態の検出などもこれに含まれる。
例) 雨天でワイパーが作動していない場合、音声メッセージで案内する。同時に関連語彙が音声認識辞書の上位に設定される。
- (6) 状況予測要求
未来の環境・状況から推定される要求。スケジュールやイベント、時間的周期性に基づく要求予測もこれに含まれる。
例) 昼食の時間帯になると最寄りのレストランの案内を表示し、音声でガイドする。
例) 「ジュース」と一言いえば、コンビニ、自動販売機などの最寄りスポットを地図表示する。

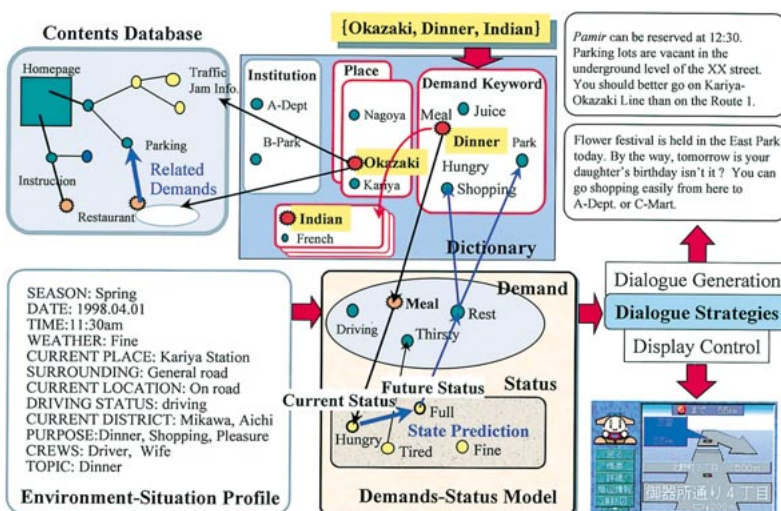


Fig.8 Example behavior of demand estimation

6. 実験システムの構築

Fig. 9に実験システムの構成を示す。複数モジュール（音声認識、音声合成、表示系、操作系、対話エージェント、通信インタフェースなど）、シナリオ、プロフィール群を独自に考案した自律分散型のメッセージ交換環境HPS [HMI Profile System (HMI: Human Machine Interface)]で結合した。

映像・音声・操作系を統合化した対話動作はシナリオと呼ぶ外部データで記述し、対話エージェントにおけるシナリオインタプリタで解釈・実行するという構成を取った。これにより、話題やアプリケーションの拡張に対して柔軟に対処できる。将来的には、通信で最新の話題の対話シナリオを配信することも考えられる。

シナリオはシナリオ記述関数や対話ユニット、他のシナリオのコール、各種制御メッセージなどで記述される。対話ユニットは問い合わせや無

発話待機、確認、提案、問い返しなどの基本対話単位で構成される。

音声認識から出力される認識候補は単語列推定において単語列として選別され、シナリオインタプリタにテキストデータとして入力される。これをもとに対話ユニットを参照して発話出力を決定し、音声合成で

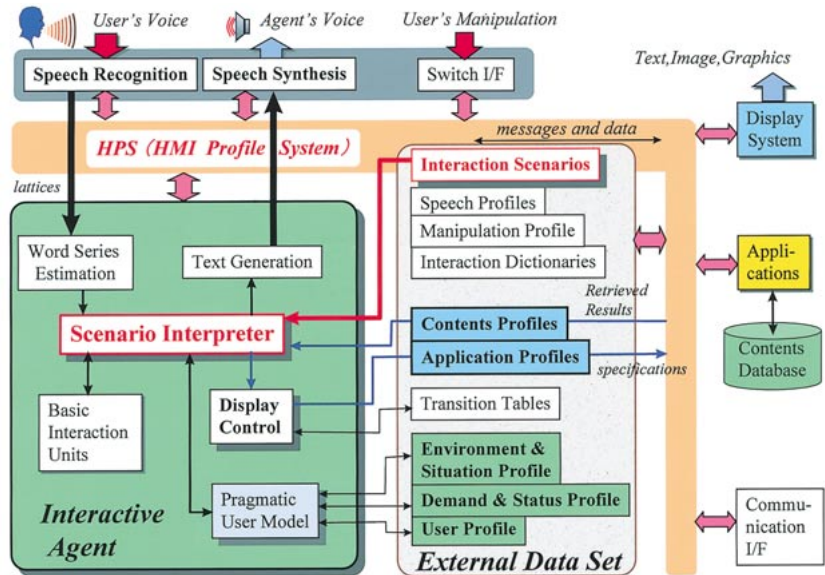


Fig.9 System architecture

Table 1 Demand-Status Table

I'm hungry.

Demand keyword

Primary Demand	Estimated State	Secondary Demand	Action	Expected State After Estimation				
For-Meal	—	Move	Move	Hungry	—	—	—	—
Meal	Hungry	Shopping	Meal	Full	—	—	—	—
—	Hungry	—	No Meal	Hungry	Sleepy	—	—	—
—	Enough	—	—	Toilet	Sleepy	Rest	Smoking	—
Drink	Thirsty	—	Sintfinn	Normal	—	—	—	—
Res.	Rest	—	Park	Comfort	Fine	Representative	—	—
—	Rest	—	Resr	Normal	—	—	—	—
Rest	Rest	Drink	Rest	Normal	—	—	—	—
Rest	Sleepy	Drink	Rest	Normal	—	—	—	—
Rest	Toilet	Drink	Rest	Normal	—	—	—	—
Toilet	Toilet	Rest	Rest	Normal	—	—	—	—
Shopping	Normal	Pleasure Resort	Depart	Tired	Thirsty	Satisfied	Hungry	—
Event	—	Shopping	Event	Satisfied	Tired	—	—	—
Pleasure Resro	—	SHopping	Event	—	—	—	—	—
MOVE	—	—	Congest	Irritated	Tired	—	—	—
Move	Fine	Shopping	No con	Satisfied	—	—	—	—
Move	Fine	Rot	Conjest	Normal	—	—	—	—
Play	Fine	Play	Sports	Hot	Tired	Thirsty	Satisfied	—
Ski	Fine	Hot Spring	Long Drilve	—	—	—	—	—
Golf	Fine	—	Short Draive	—	—	—	—	—
—	—	—	Long Draive	Hungry	Sleepy	Tired	Toilet	—
Driving	Very Fine	—	No Congest	Comfort	—	—	—	—

音声に変換する。この出力決定過程で表示系制御を介してアプリケーション（レストラン検索、景観映像検索など）に検索要求などの指令を送り、データベースの検索結果を表示する。一方で表示内容がコンテンツプロファイルとしてシナリオインタープリタに返され、次の応答を決定する。

なお、車室内ユーザモデルは環境・状況プロファイル、要求・状態プロファイル、ユーザプロファイルを参照して駆動され、発話戦略を起動する。

7. 動作機構

動作例として問い返すシナリオの動作機構を説明する (Fig. 10)。

まず、「岡崎でインド料理を食べたい」という発話がなされたとき、要求キーワードを最優先して検出し、「食べたい」を認識する。この結果、＜食事＞の話題を認識し、残りの音声区間および次の発話に対して食事に関連する話題の認識辞書を設定することで誤認識を低減させる。一方で、車室内で考えられるすべての属性項目をあらかじめレベル分けしておき、その内、＜レストラン検索＞に関する項目群がユーザフレーム上で活性化される（現在地、目的地、好きな料理、よくいく店、同乗者、渋滞状況、他）。この中で特に検索に必須であるレベル1の項目が充填されていない場合、これを検索実行における不足条件とみなして検出し、ユーザに問いかける発話戦略を駆動する。

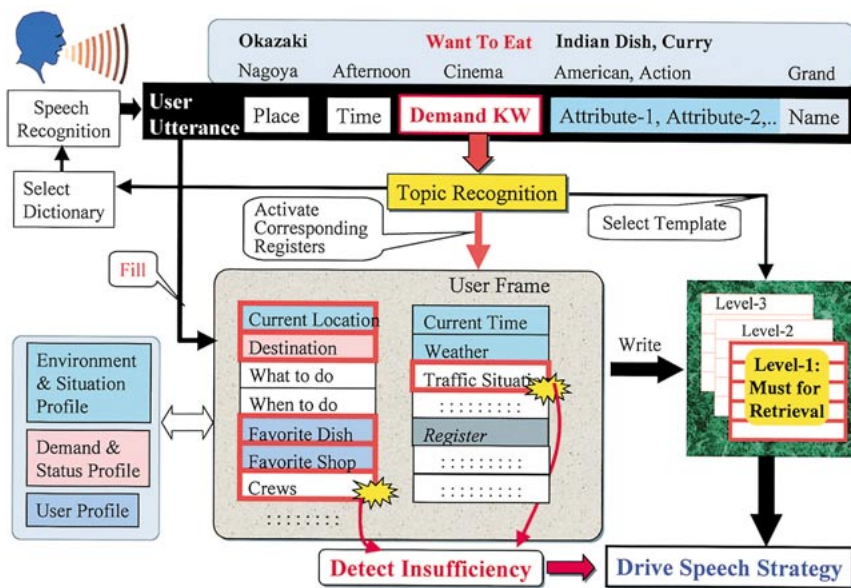


Fig.10 Scenario for asking again

ここで、このユーザフレームのレジスタには、それまでの対話で獲得したキーワードに加えて、環境・状況、要求・状態、ユーザプロファイルからの情報が充填される。逆にユーザフレーム上の活性化データをもとに推論規則を適用して新たな項目値が推定された結果、プロファイルにその推定結果が格納される場合もある。

8. 実験結果

二つの話題（レストラン検索と交差点映像検索）を扱う対話シナリオで検索実験を行った (Fig. 11)。「ご用件があれば...」から始まり、要求キーワードを含む単語列（図中のもの以外の同義語も多数用意。食事の要求は30語、料理名



Fig.11 Demonstration of speech dialogues

は46種類, 有名チェーン店20件, 映像の要求は約6語, 交差点名約50個, 愛知県下の代表的地名75個, 対話語彙約30語)を音声認識し, レストラン検索と交差点映像検索の二つの話題を自在に遷移できる.

また, 音声とキー入力は随時切り替え可能である. 車載プロファイルの効果により, 従来のステップ式の入力に比べて1/3程度の対話回数で話題の遷移やメディア検索が行えることがわかった (Table 2).

以上は2000年4月時点のプロトタイプである. 現在は, インターネット検索, 地図表示, 旅行情報検索, 無線映像検索・配信などが音声対話(英語にも対応)で行え, プロファイルによる個人適応化なども一部可能になった. また, データベース側には国際標準案 ISO / MPEG-7 を用いたコンテンツ記述や独自の高速検索手法を実装している.

Table 2 Examples of dialogues for media retrieval

	<i>Restaurant Retrieval</i>	<i>Retrieval of Crossing Images</i>
Dialogue Generation using In-Vehicle Profiles and Word Series Estimation	"French restaurant in Okazaki." 'Found one.' (Display) "How about Italian?" 'Found three.' (Display)	"Traffic situation in Ootemachi" 'No images for Komaki in database.' "No! Ootemachi, Chiyoda-ku, Tokyo" 'Image of Ootemachi, Chiyoda-ku, Tokyo will be displayed.'
Conventional Step typed Dialogue Generation	"Restaurants retrieval" 'Where do you like to go ?' "Okazaki, Aichi" 'What do you like to eat ?' "French" 'Found one.' (Display) "Restaurants retrieval" 'Where do you like to go ?' "Okazaki, Aichi" 'What do you like to eat ?' "Italian" 'Found three.' (Display)	"Traffic images retrieval" 'Where do you like to display ?' "Ootemachi" 'No images for Komaki in database.' "Traffic images retrieval" 'Where do you like to display ?' "Ootemachi, Chiyoda-ku, Tokyo" 'Image of Ootemachi, Chiyoda-ku, Tokyo will be displayed.' (Display)

9. まとめ

カーナビなどの車載マルチメディアにおいて安全かつ快適に所望の情報を検索する方式として, 車載プロファイルと対話エージェントを提案した. 今後はメディアデータベースの知的表現形式として期待されている ISO / MPEG 7 (2001年国際標準化予定) と整合できる車載プロファイルの確立をめざす.

< 参考文献 >

- 1) 笹木, 浅見 " 対話型メディア通信について ", IMPS'99, pp. 87-88, Sept. 1999
- 2) M.Sasaki, K.Asami, " A Proposal for In-Vehicle Profiles ", ISO/IEC/JTC1/SC29/WG11/M5308
- 3) M.Sasaki, K.Asami, " Interactive Agent using In- Vehicle Profile ", ISO / IEC / JTC1 / SC29 / WG11 / M5309

< 著 者 >



笹木 美樹男
(ささき みきお)

基礎研究所 第2研究室
人工知能型HMIおよびそれに関連するメディア処理システムの研究開発に従事.



浅見 克志
(あさみ かつし)

基礎研究所 第2研究室 (現在, ATR 音声言語通信研究所に出向中) 人工知能型HMIで用いる対話エージェント, 特に発話意図抽出手法の研究開発に従事.