

## エージェントアーキテクチャに基づいた 音声対話支援システムの構築

城塚 音也 桑田 喜隆 安地 亮一 小泉 宣夫  
NTTデータ通信株式会社  
情報科学研究所

要約: コンピュータが人間同士の音声対話を理解し、リアルタイムに対話の支援を行う、知的な Computer Mediated Communication (Intelligent CMC: ICMC) の検討を行っている。その実現例として、金融相談端末上の相談業務における音声対話を知的に支援する実験システムを構築した。本稿では、我々がサイバーコミュニケーションと呼ぶ知的 CMC の枠組みを説明し、エージェントアーキテクチャによって実現した実験システムについて述べる。また、コンピュータエージェントによる自律的な支援情報の提示の有効性を調べる実験の結果について報告する。

## Intelligent dialog supporter based on agent architecture

Otoya SHIROTSUKA, Yoshitaka KUWATA, Ryoichi ANCHI and Nobuo KOIZUMI

Laboratory for Information Technology

NTT DATA CORPORATION

e-mail: {sirotuka, kuwata, anchi, nkoizumi}@lit.rd.nttdata.co.jp

Abstract: An experimental system for the support of financial consulting service was built in order to investigate an Intelligent Computer Mediated Communication (ICMC) system with spoken dialogue understanding function which supports human-human communication in spoken language by augmenting helpful information into their work space. In this paper, the framework of ICMC, which we call "Cyber Communication", is explained and the experimental system based on multi agent architecture is described. An experimental result is reported which is performed to evaluate the effectiveness of output function of support information by computer agents.

### 1 はじめに

人間同士のコミュニケーションと人間-コンピュータ間のインタラクションがシームレスにつながった新しいコミュニケーション環境の実

現を目標に、人間同士のコミュニケーションをコンピュータにより知的に支援する方法の検討を行っている。我々はこのようなコミュニケーション環境をサイバーコミュニケーションと呼び、その実現例として年金相談業務を対象に相

談対話をコンピュータが理解し、相談業務の支援を行う相談業務支援システム「サイバーコンサルタント」を試作した<sup>1)</sup>。本稿では、従来の対話支援研究について述べ、試作した相談業務支援システムについて説明する。また、試作したシステムの支援機能がどの程度対話の支援となるかを調べるための評価実験の結果を報告する。

## 2 従来のコミュニケーション支援研究

コンピュータを介したコミュニケーションは Computer Mediated Communication (CMC) と総称されているが、その内容は電子メール、テレビ会議システム、CSCW ツールと雑多である。これらは広義のコンピュータによって支援されたコミュニケーションといえる。CMC とサイバーコミュニケーションの違いは、コンピュータが人間のコミュニケーションを仲介する道具として受動的に存在するのでなく、自律的なエージェントとして人間同士のコミュニケーションを理解し、コミュニケーションを促進するための情報処理をアクティブに行う点にある。

リアルタイムの人間同士の対話をコンピュータに理解させ、対話支援を行わせる試みとしては、たとえば西本ら<sup>2)</sup>は、キーボード対話ではあるが、人間同士の対話をコンピュータが理解し、お互いの対話中に現れるキーワード群に基づいて話題の空間や個人の視点を可視化し、言及されていない話題空間に属する、見逃されている関連情報を提示することにより対話の促進を図っている。

また人間同士の音声対話をモニターし、なんらかの決められたアプリケーション処理を行うシステムとしては、伊藤ら<sup>3)</sup>が家のレイアウトの決定をタスクに人間同士の対話をコンピュータが理解し、対話内容に反映してレイアウトの変更をおこなうマルチモーダルシステム、有馬ら<sup>4)</sup>が会議室の予約の受け付け係と利用者という人間同士の音声対話から情報抽出を行い、対話の要約をリアルタイムに生成するシステムを構築した。

われわれは将来のコミュニケーション環境において、西本らの対話促進機能と伊藤、有馬らの音声対話理解に基づいたアプリケーション処

理機能を併せ持つことが必要と考え、今回構築した相談業務支援システムにおいて、小規模ながらこれを実現している。

## 3 相談業務支援システム

### 3.1 相談業務の支援形態

試作した年金相談をタスクとした実験システムはPCによる遠隔会議システムをベースとしている。図1に実験システムの形態を示す。

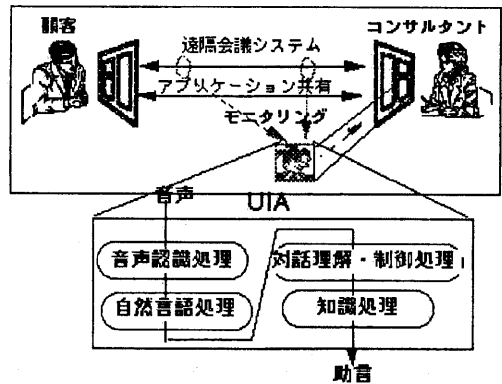


図1 サイバーコンサルタントの形態

相談業務のためのしくみとして相談員（コンサルタント）側には年金、顧客に関する知識データベースと年金の受給額等を計算するためのアプリケーションが用意されている。これらを使用して得られる結果は相談員側の端末画面に表示されるが、相談員の操作により同じ内容を顧客側に表示させることができる（情報の共有機能）。相談業務の知的支援は、各端末においてそれぞれの音声認識され、その音声の内容に応じて適宜支援情報を相談員の画面に表示すること（音声のモニタリングに基づく支援機能）により行われる。

相談員と顧客という関係上、その対話形態、共同作業の形態は両者同質とはなっていない。現在のサービスモデルでは、顧客はコンピュータによる共同作業に慣れていないことを前提としている。そのため、顧客の情報入力方法は音声と画像のみに制限されており、情報の共有機能のみ相談員主導の下、一方通行で行われる非

対称な作業空間となっている。この支援形態は、顧客から見た場合、相談員との対話に際して、対話をモニターして情報のアシストを行っている他者の存在に気づかない。相談員から見ると、支援が必要と感じない場合は提示される情報を無視することにより顧客との対話を邪魔されることを避けることができる。そのため、文献2、3と比べて人間同士の対話を妨げることの少ない支援形態であるといえる。

また、コンピュータが誤った支援情報を提示したとしても、相談員が、その提示された支援情報を取捨選択して、顧客に提示したり、内容を口頭で説明することができるのでサービスの質が低下することを防ぐことができ、誤った支援情報の出力に対してロバストな支援形態であるという特徴がある。

### 3.2 支援の種類

システムは年金相談をおこなう相談員のために以下の3種類の支援を行う。

1. 対話内容のアドバイス  
顧客に話すべき対話の内容や話の進め方を提示する。
2. リファレンス情報の表示  
対話の内容に関連した情報を知識データベースから取得し、編集して提示する。
3. ユーザ情報の取得  
対話内容から相談に必要なユーザ情報（保険の種類や職業等）を取得し、必要

に応じて情報の加工（年金の受給額計算等）を行う。

現在の実験システムでは、1については、ごく基本的な知識（例：年金の受給額を計算するためには年金の加入期間と標準報酬月額を聞かなければいけない）しか使用しておらず、個人の状況や対話のコンテキストに応じた対話内容の推定と推定された発話内容に基づいたアドバイスを作成できるようにはなっていない。2については、個人の状況（加入年金の種類等）に応じて、支援内容が編集される。現在はあらかじめ用意された知識ベースのみを使って支援内容を作成しているが、サービスの種類によっては、より大規模なデータベースの検索（インターネット上のサーチエンジン等）を利用して、リファレンス情報を収集することも有効であろう。

### 3.3 システム構成

サイバーコンサルタントは、図2のような構成で実現されている。技術的特徴としては、

- ワードスポッティングを用いた音声認識
- ルールベースの知識処理
- マルチエージェントアーキテクチャ

が挙げられる。

サイバーコンサルタントは年金相談に必要な知識をハイパーテキストの形式で持っている。知識ハイパーテキストは、それぞれ「国民年金」や、「繰り上げ受給」等のトピック毎に詳細に分類され、インデックスが付けられている。

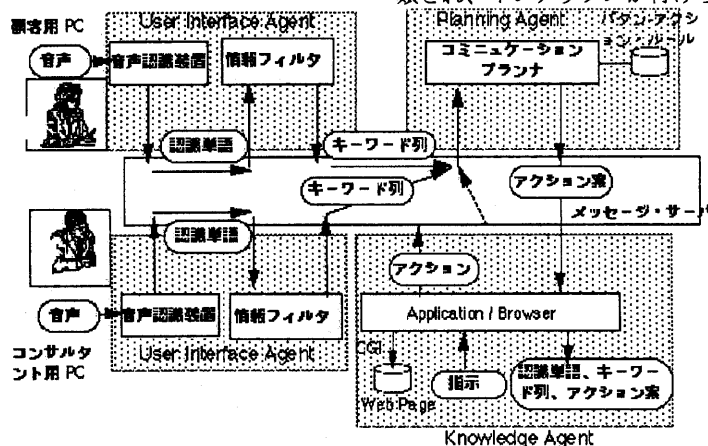


図2 サイバーコンサルタントの構成

## 音声認識

自由発話を対象にした音声認識は語彙を限定することができず、また、極端にドメインを限定しないかぎり良好な認識精度が得られないという問題がある。また、遠隔会議システムの場合、相談員と顧客の端末の間でおたがいに相手側に伝送する音声を認識しようとする、圧縮ひずみや回線ひずみ等による認識率の低下が予想される。

そこで音声認識は、顧客側の相談端末とオペレータ側双方で動作させることとし、それぞれ年金相談に必要なキーワードのみを認識対象語彙として登録しておき、発話中のキーワードのみを検出するキーワードスポッティングを採用した。ユーザが発話を中断した時、認識処理は一時的に中断され、それ以前に認識されたキーワードがメッセージサーバに送られ、新たな発話待ち状態に入る。これらの操作はすべて自動的に行われるため、オペレータと顧客共、音声認識が作動していることはまったく意識せず自然な相談対話を続けることができる。

現在の実験システムは「厚生年金」や「繰り上げ受給」などの年金に関するキーワード、「いくらですか」や、「教えて下さい」等の相談に用いられる一般的な言い回しが、認識対象語彙として登録されている。認識対象語彙数は62である。

## 知識処理

音声認識によって対話から抽出されたキーワードを基に対話支援のためのプランニングを行う。キーワードの誤抽出や抽出もれに対してロバストな言語処理を実現するために複数のキーワードの出現組み合わせパターンとそれに対応した支援計画案からなる支援ルールを複数用意し、抽出されたキーワード群に対してルールとのパターンマッチングをおこなうことにより、行うべき支援の計画案を決定している。以下に支援ルールの例を示す。正規表現で示されたパターンにマッチする様なキーワード列がメッセージサーバから送られたとき、その第3引数のようなコマンドをメッセージサーバに返送する。

# menu01

```
[“(厚生年金|共済年金).*違い”, menu01],  
[“(厚生年金.*共済年金”, menu01],  
[“(年金.*しくみ”, menu01],
```

```
[“(年金|遺族).*もらえる”, menu01],  
[“(年金.*種類”, menu01],  
[“(自営業|公務員).*年金”, menu01 ],  
[“(年金.*引続き”, menu01 ],
```

# menu04

```
[“(年金.*(受給|増やしたい)”, menu04],  
[“(受給.*増やしたい”, menu04],  
[“(年金.*(割増|受給)”, menu04],  
[“(割増.*受給”, menu04],
```

# menu05

```
[“(標準報酬月額”, menu05 ],  
[“(保険料.*(どのように|決まる)”,  
menu05],
```

## マルチエージェントアーキテクチャ

本システムで用いているマルチエージェントアーキテクチャは、ブラックボードモデルを用いており、メッセージサーバを介して、機能毎に存在する複数のエージェントが通信をおこない協調動作をおこなうアーキテクチャである<sup>6</sup>。

音声認識部からメッセージサーバに送られたキーワードは、知識処理部（コミュニケーションプランナ）に転送される。知識処理部は、あらかじめ知識から生成されたルールを元に、キーワードに対してパターンマッチングを試みる。その発話がどの知識に最も近いものか、どの知識を提示すれば対話の参考となるかのプランを立て、知識のインデックスをメッセージサーバに送信する。メッセージサーバに転送された知識インデックスは、コマンドとしてアプリケーション部に転送される。コマンドを受けたアプリケーション部は、知識データベースを検索し、相当する知識画面をオペレータ端末上に提示する。知識画面の例を図3に示す。さらに、対話の流れから提示された情報を顧客に提示することをオペレータが判断した場合は、アプリケーション共有によって顧客側画面にも知識画面が表示される。

メッセージサーバを用いたエージェントアーキテクチャは、単純なプロトコルで通信を行うため、実装が容易で機能拡張がしやすいという利点がある。また、役割毎にプロセスが分けられており、同一マシン上で複数のエージェントを動作させることも、地理的、ネットワーク的に分散した環境でも柔軟に対応できる。

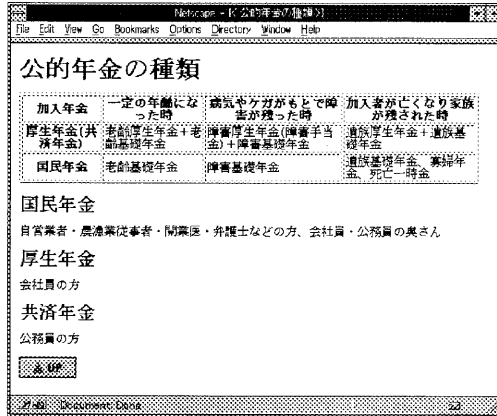


図 3 知識画面の例 (リファレンス)

#### 4 支援機能評価のための予備実験

サイバーコンサルタントの支援形態が、人間同士の対話の支援にどの程度効果があるのかを調べる必要がある。しかし実際に対話音声の認識理解を行いながらリアルタイムに支援情報を出力するには、処理速度的にも、音声対話理解の精度的にも問題がある。そこで、純粋に支援形態の有効性を調べる有効な方法として Wizard of Oz 法による評価実験を行った。また、評価をおこなう人間 (NTT データの社員) が被験者となった場合、年金相談をタスクにすると、その知識レベルがあまりにも低く、支援が有効に働かないことが予想される。そこで基本的な知識がかならずあると思われる会社の福利厚生、会社組織に関する相談をタスクに評価を行うこととした。6 種類の相談事例 (米国支店の住所の問い合わせ、病気休暇中の手当申請の手続き方法、会社、組合等がサポートしている住宅ローンの借り入れ限度額等) を用意し、6 人の被験者に相談員役を演じてもらい、のべ 36 対話の相談を行った。顧客の役はすべて特定の 1 人が演じている。

支援情報の提示は、被験者にはコンピュータが対話をモニターして、自動的に提示していると説明しているが、実際は相談対話をモニターしている人間が手動で支援情報を選択、提示することにより行う。

それぞれの相談内容に対して 3 種類の支援条件を設定してお互いを比較した。

- 支援なし
- 遅れない提示タイミング
- 約 5 秒ほど遅れのある提示タイミング

相談員が使用できる支援情報として、NTT データの社内ネット上の総務部のホームページに公開されている情報を用いた。相談員は World Wide Web (WWW) ブラウザを使用して、必要な情報を探し出すことができる。支援の遅れを約 5 秒としたのは、WWW のハイパーリンク先の情報を表示する時の遅延が 5 秒未満であることから、より大きな値として設定した。

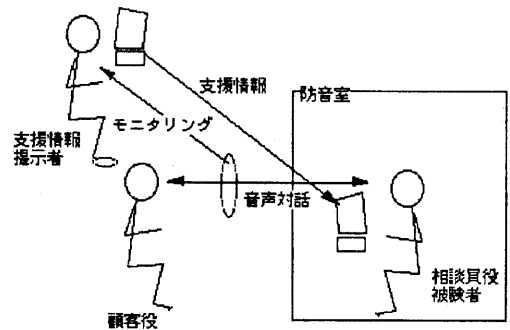


図 4 実験の構成

図 4 に実験の構成を示す。各対話に対して、その対話の達成に対する所要時間を計測し、それぞれの対話音声を録音した。また、相談員役の被験者の WWW ブラウザの操作の様子を録画するとともに、被験者に対してアンケートを行った。

評価実験の結果を図 5 に示す。

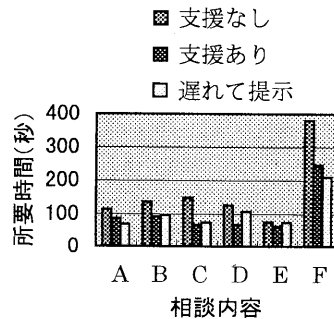


図 5 相談達成のための所要時間

結果から以下の2つの傾向が伺える。

- 支援情報の提示により約11%~51%、平均33%、支援なしの場合より対話達成にかかる時間が短縮された
- 支援情報の提示タイミングが遅らせた場合の達成時間に対する影響は伺えない。アンケートの結果からも遅れに気づかなかったというコメントが多数をしめた。

相談員役を行った被験者に対して行ったアンケートによると、支援機能が便利であることは肯定されたが、一方、以下のような支援のしくみの問題が指摘された。

- 提示されたWWWページのどの部分が答えるべき内容なのかわからず、戸惑ってしまう。
- 支援情報と相談員の情報検索結果を同じ画面で表示しているため、コンピュータと相談員のどちらが操作して出た情報なのかわからない。
- 唐突に支援情報が出てくると作業を邪魔されている感じがする。こちらから情報が必要ときだけ支援が提示されるしくみでもいいのではないか。
- コンピュータにより提示された情報が、ハイパーリンクで整理された知識のどの部分から出てきたのかわからず不便。提示された情報と関連する情報としてどのようなものがあるかを知りたい。

現在のサイバーコンサルタントによる年金相談でも、これら指摘された問題と同様の問題が生じると思われる。これらの問題を解決するためには、提示情報をその用途に応じてリアルタイムに編集する必要がある。また、支援情報を画面への提示の方法、そのタイミング等についても考慮する必要がある。

今回の実験では、コンピュータの対話理解の失敗により誤った支援情報がなされた場合、支援の効用がどの程度損なわれるかについては調べなかった。これについては次回の実験において、支援情報の誤り率を変えて評価を行う予定である。

## 5 まとめ

人間同士の対話を知的にサポートするしくみを持った相談業務支援システム、サイバーコンサルタントは、音声認識のワードスポッティング技術と、簡単なパターンマッチングによる知識処理を組み合わせることでその支援機能を実現している。自由発話からのキーワード音声認識の性能はかなりのキーワードの抽出もれや誤抽出を含むため、パターンマッチングを用いて情報を抽出する事で、ロバストな情報提示を図っている。

支援機構の有効性を調べる評価実験をWizard of OZ法により行い、平均33%の対話の短縮を確認した。また、被験者のアンケートからは本支援機構が快適な支援となっていることが確かめられたが、支援の形態、提示方式については改良する必要があることが分かった。

今後はこれらの問題点に対応すると共に、現在の対話の表層的な特徴からの支援手法を拡張し、対話プラン認識に基づいた支援および、マルチモーダル対話の支援を検討する予定である。

### 謝辞

音声対話技術の応用について議論を頂いた当研究所の管村所長に感謝いたします。

<sup>1</sup> 安地、城塚、桑田，“相談業務を対象とした対話支援環境の構築”，情報処理学会第69回ヒューマンインタフェース研究会，（1996，11）

<sup>2</sup> 西本、角、間瀬，Argumented Informative Discussion Environment “AIDE”，第2回知能情報メディアシンポジウム予稿論文集，pp259-266（1996，12）

<sup>3</sup> 伊藤、木山、関、小島、張、岡，“同時複数話者の会話音声およびジェスチャのリアルタイム統合理解によるNovel Interface System”，情報処理学会研究報告，95-SLP-7，pp17-22，（1995.7）

<sup>4</sup> 有馬、亀山、河合“自動対話要約システムの構築”，音響学会講論集，pp157-158，（1996，3）

<sup>5</sup> 磯部、森島、小泉，“ガベージモデルを用いたキーワード抽出”，音響学会講論集，pp31-32，（1996，9）

<sup>6</sup> 桑田、城塚、安地、小泉，“マルチエージェントアーキテクチャに基づく音声対話支援システム”，情報処理学会第54回全国大会，7H-04，（発表予定）