

# 遠隔会議を対象にした音声対話モニタリングによる 対話支援システム

城塚 音也<sup>†</sup> 桑田 喜隆<sup>†</sup>  
安地 亮一<sup>†</sup> 小泉 宣夫<sup>†</sup>

計算機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 (CMC) を提案する。提案する CMC は、人間同士のコミュニケーションと人間-計算機のインタラクションをシームレスに連結することにより、人間の知的活動を促進する。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。システムは対話中のキーフレーズを音声認識することにより対話の内容を推定し、支援ルールによって支援計画を立案し、支援内容の作成を行う。作成された支援情報は相談員に提示され、相談業務の円滑化を促進する。システムの自律的な対話支援のコンセプトの評価実験および、対話モニタリングによる対話支援性能の評価実験について報告し、実験結果について考察する。

## Intelligent Remote Conference Assistant by Spoken Dialogue Monitoring

OTOYA SHIROTSUKA,<sup>†</sup> YOSHITAKA KUWATA,<sup>†</sup> RYOICHI ANCHI<sup>†</sup>  
and NOBUO KOIZUMI<sup>†</sup>

An Intelligent Computer Mediated Communication (CMC) is proposed which enables computers to provide support information to people who are in conversation by monitoring their spoken dialogue. An experimental system for the support of pension consulting service was built in order to investigate proposed CMC system. The system estimates the topic of the dialogue from detected key phrases by speech recognition technology and designs support information with assist rules. The efficiency of the consulting service can be improved by the support information which is presented automatically by the system. The concept of Intelligent CMC was evaluated by the reference indication to human-human dialogue and the performance of the generating support information function was measured by several experimental conditions. The experimental results are discussed.

### 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計算機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知的な Computer Mediated Communication (CMC) 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を試作した<sup>1)</sup>。本論文では、2章において、従来の対話支援研

究と提案する知的対話支援について述べ、3章において、試作したマルチエージェントアーキテクチャに基づく相談業務支援システムの構成および動作について説明する。4章では、人間同士の対話に応じた自律的な支援提示という支援コンセプトの評価実験とワードスポットティング型の音声認識とルールベースの支援計画立案に基づいた支援機能についての評価実験の結果を報告し、考察を行う。

### 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人間同士のコミュニケーションは 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。我々は、CMC に属す

<sup>†</sup> NTT データ通信株式会社情報科学研究所  
Laboratory for Information Technology, NTT Data Corporation

る新しいコミュニケーション形態として、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はなく、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知的CMCを提案する。

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、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、何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行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付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。

新しい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合わせ持つことが必要と考える。今回構築した相談業務支援システムでは、これら2種類の対話支援機能を実現している。

また、人間同士の即時型コミュニケーションにおいては、音声言語メディアの使用が必須であるため、計算機による人間同士のコミュニケーションをモニタリングするためには、音声認識をフロントエンドとして使用することが必要である。しかしながら自由発話の音声対話を高精度で認識、理解する手法は確立されていない。構築した対話支援システムでは、人間同士の音声対話中のキーフレーズをワードスポッティング型の音声認識により抽出し、対話内容を推定する表層的な音声対話理解のアプローチをとることにより、処理のロバスト性を確保している。

### 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に実験システムの支援モデルを示す。

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

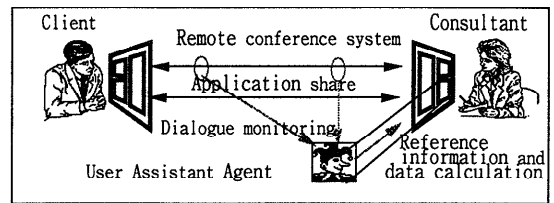


図1 相談業務の支援モデル

Fig. 1 Support model for consulting service.

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。キーボード等の入力により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のキーフレーズをUser Assistant Agent (UAA)が認識し、そ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。現在の支援モデルでは、顧客は計算機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、顧客の能動的な情報入力方法は音声のみに制限されており、情報の共有機能は相談員主導のみ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支援を行っている他者の存在に気づかないため、相談員の信頼を損なわない。

また、相談員から見た場合、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。

音声モニタリングによる支援情報の自律的提示は、計算機から支援情報を得るために特別に操作を行う必要がない。また、必要な支援情報の提示がなされない場合、相談員は能動的にキーボード等で検索することができ、提示された支援情報に、対話の内容と関連のない情報が提示されても、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。そのため、本システムの支援形態は、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

UAAは年金相談を行う相談員のために、以下の3

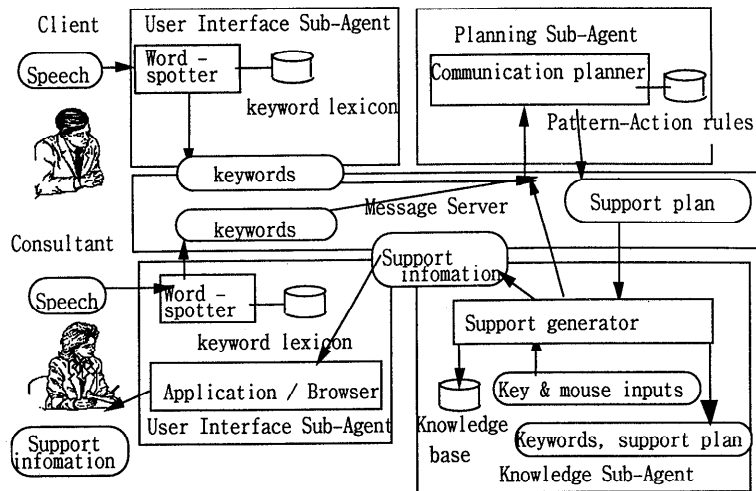


図 2 相談業務支援システムの構成

Fig. 2 Configuration of support system for consulting service.

種類の支援情報を出力する。

- (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顧客情報（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、(1)の対話内容のアドバイス機能は、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)のリファレンス提示機能は、顧客の状況（加入年金の種類等）に応じて、顧客に役立つように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)のユーザ情報の取得機能は、銀行や役所による年金相談サービスでは、個人情報の方が事前に登録されていることから補助的な機能であり、データベース中に入らない加入保険種類等の情報取得を主な対象としている。情報の取得は、モニターしている音声

中に、顧客情報を含んだフレーズ（例：厚生年金に加入→顧客データ [加入保険種別=厚生年金]）を検知することにより行う。対話のモニタリングにより推測された顧客データは誤りを含むことが予想されるので、相談員は内容を確認し、誤りを修正し、欠けている情報を追加することにより顧客データを完成させる。

### 3.3 UAAの構成

UAA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、

- (1) ワードスポッティングを用いた音声認識<sup>5)</sup>
- (2) ルールベースの知識処理
- (3) マルチエージェントアーキテクチャがあげられる。

#### 3.3.1 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互いに相手側に伝送する音声を認識しようとするとき、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フレーズのみを認識対象語彙として登録しておき、発話中のキーフレーズのみを検出する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とき、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、顧客とも、音声認識が作動していることを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 UAA には「厚生年金の受給額」や「繰り上げ受給制度」等の年金に関するキーワード、「いくらですか」や、「もらえますか」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。

### 3.3.2 知識処理

UAA は年金相談に必要な知識をタグづけされたテキスト形式で持っている。タグの種類は「国民年金の受給資格」や、「繰り上げ受給」等の支援タグ、「厚生年金加入者」、「昭和 24 年以前の出生」等の個人情報タグがあり、音声モニタリングにより立案された支援と、その時点で得られている顧客の個人情報に基づいて、該当するタグの部分のテキストを選択、編集し、ハイパーテキストとして提示している。

知識処理による支援種類の決定は、音声認識により対話から抽出されたキーワードに基づいて行う。キーワードの誤抽出や抽出もれに対してロバスタな話題推定を実現するために、複数のキーワードの出現組合せパターンとそれに対応した支援カテゴリからなる支援ルールを複数用意し、検出されたキーワード群に対して各ルールのキーワードパターンとのパターンマッチングを行い、マッチしたルールの支援カテゴリから支援の内容を生成する。

以下に支援ルールの例を示す。

- ```
# menu01 年金の種類について
["(厚生年金|共済年金).*違い", menu01],
["年金.*しくみ", menu01],
["年金.*種類", menu01],
["(自営業|公務員).*年金", menu01],
# menu05 年金受給額の算定アプリケーション起動
["年金受給額.*計算", menu05],
["保険料.*(いくらもらえる|受給額)", menu05],
```

例中、#menu01 年金の種類についての最初のルールは、「『厚生年金』または『共済年金』という単語に続いて、0 個以上の任意の数の文字が出現した後に、『違い』という単語が出現するパターンが、入力されたキーワード群とマッチした場合、menu01 という支援カテゴリに属する支援を出力する」というルールである。

本システムの出力は、誤りが含まれていても、最終的に相談員が取舍選択することができるので、できるだけ広範囲の支援情報を出力する、カバレッジ重視の支援方針をとっている。そのため、現在のシステムでは、複数のルールがマッチした場合は生成されるすべての支援を出力している。また、キーワードどうしの

出現順序はパタンの記述の順序によって区別されているが、その距離は、考慮されておらず、現在のところルールの確信度等には用いていない。

### 3.3.3 マルチエージェントアーキテクチャ

UAA は、機能ごとに存在する複数のサブエージェントがメッセージサーバと呼ぶブラックボードを介して、非同期に通信を行うことにより協調動作を達成するマルチエージェントアーキテクチャで構成されている<sup>6)</sup>。

User interface sub-agent からメッセージサーバに送られたキーワードは Planning sub-agent に転送される。Planning sub-agent 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。その結果、行われている対話に対する支援の種類が決定され、行うべき支援のタグ情報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支援タグは、Knowledge sub-agent に転送され、コマンドを受けた Knowledge sub-agent は、知識データベースを検索し、相談員に提示する支援内容を作成する。User interface sub-agent は、作成された支援内容を端末上に提示する。支援内容の例を図 3 に示す。さらに、相談員に提示された情報を顧客にも提示することが有益と相談員が判断した場合、会議システムのアプリケーション共有機能により顧客側画面にも支援内容が表示される。

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、単純化された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。3 人以上の対話参加者を扱うための拡張や、画像認識等を行う User interface sub-agent の追加が容易であ

| 加入年金       | 一定の年齢になった時    | 病気やケガがもとで障害が起きた時     | 加入者が亡くなり家族が残された時  |
|------------|---------------|----------------------|-------------------|
| 厚生年金(共済年金) | 老齢厚生年金+老齢基礎年金 | 障害厚生年金(障害手当金)+障害基礎年金 | 遺族厚生年金+遺族基礎年金     |
| 国民年金       | 老齢基礎年金        | 障害基礎年金               | 遺族基礎年金、寡婦年金、死亡一時金 |

**国民年金**  
自営業者・農漁業従事者・開業医・弁護士などの方、会社員・公務員の奥さん

**厚生年金**  
会社員の方  
共済年金  
公務員の方

図 3 支援内容の例 (リファレンス)

Fig. 3 An example of support information (reference).

り、役割ごとにプロセスが分けられており、同一マシン上で複数のサブ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に対しても柔軟に対応できる。

#### 4. 評価実験

提案した知的CMCシステムの支援コンセプトの評価を行った。これは人間同士の対話に対する自律的な情報提供によるサポートが相談業務にどの程度役立つかを調べることを目的としている。また、音声モニタリングに基づいたUAAの支援出力機能について性能評価を行い、キーワードの誤認識や話題の誤推定等に起因する誤った支援情報（ノイズ）について調査した。

支援コンセプトの評価に関しては現実のシステムに含まれるノイズを含まないベースラインの条件で実験を行った。評価結果についてはすでに報告を行っている<sup>7)</sup>。本論文では4.1節でその概略を説明し、4.2節以降において音声モニタリングに基づいた支援出力機能の性能評価について述べる。また、両実験とも、3つのシステムの支援機能のうち最も基本的な機能であり、支援のコンテンツが充実しており、支援精度が期待できると考えられるリファレンスの提示機能のみを評価した。

##### 4.1 支援コンセプトの評価

会社の福利厚生、会社組織に関する相談をタスクに Wizard of Oz 法により支援方式の評価実験を行った。6種類の相談事例（例：病気休暇中の手当申請の手続き方法）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支援がない場合とある場合を比較した。

相談員が使用できる支援情報として、我々の所属する会社の社内ネット上に設けられた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、必要な情報を採り出すことができる。

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録音した。また、相談員役の被験者のWWWブラウザを操作する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。

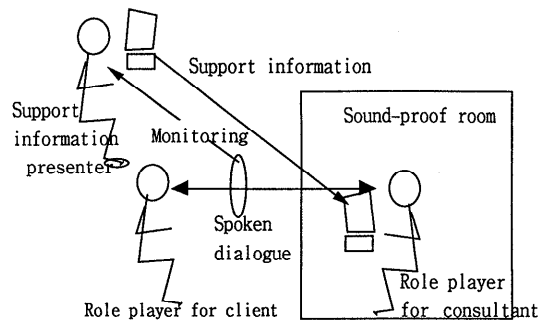


図4 実験の構成

Fig. 4 Configuration of evaluation experiment.

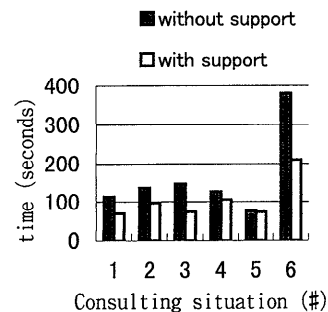


図5 相談達成のための所要時間

Fig. 5 Consumption time for task achievement.

評価実験の結果を図5に示す。

結果から、支援情報の提示により約11~51%、平均33%、支援なしの場合と比較して対話達成にかかる時間が短縮されており、本支援方式が相談対話の効率化に貢献していることが分かる。

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、相談員が自ら情報を検索する画面に支援情報を挿入したことについて、唐突に支援情報が出てくるので作業を邪魔されている感じがする、また、計算機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が分からず不便といった意見が寄せられた。

年金相談タスクでも、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、相談員自らの情報検索行為を妨げない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ことが分かった。

##### 4.2 支援出力機能の性能評価

次に、音声モニタリングによる対話支援情報の出力機能の評価について述べる。タスクは相談システムの対象である年金相談である。年金相談の対話テキスト

表 1 相談状況、対話ペアおよび支援ルールの例

Table 1 Examples of consulting situation, dialogue pairs and support rules.

|                                                                            |
|----------------------------------------------------------------------------|
| 1. 相談状況：                                                                   |
| 前提：64歳の夫の妻。夫はまだ基礎年金の受給をしていない。<br>事件：夫が死亡。<br>質問：妻の私は、夫に代わって年金を受け取れるか？      |
| 2. 発話テキスト（顧客）：                                                             |
| うちの主人は、64だったので、来年から年金もらえるって思っていたら、突然死んでしまったんです。代わりに年金わたしが受け取れないかしら？        |
| 3. 発話テキスト（相談員）                                                             |
| つまり、亡くなられたご主人に代わって、配偶者の方が年金を受け取れるか、というご質問ですね。                              |
| 4. 支援ルール                                                                   |
| 顧客：[“代わりに年金.*受け取れ”，MENU 遺族年金の受給資格]                                         |
| 相談員：[“*配偶者の方.*年金を受け取”，MENU 遺族年金の受給資格]                                      |
| 5. 認識用キーフレーズ                                                               |
| 代わりに年金 [かわりにねんきん]<br>受け取れ [うけとれ]<br>配偶者の方 [はいぐうしゃのかた]<br>年金を受け取 [ねんきんをうけと] |

データとその音声データを擬似的に作成し、実際にシステムに入力することにより、評価を行った。まず、実験に使用した相談状況、作成された対話テキストおよび支援ルールの例を表 1 に示す。

対話テキストデータの作成手順は以下のとおりである。

- (1) 15種類の顧客の相談状況とそれに対して出力すべき正解支援を用意する。また、発話テキストデータ作成者には事前に年金に関する基礎知識を与える。
- (2) 15種類の相談状況に対して発話テキストデータ作成者が、顧客の発話を考え、その発話テキストを作成する。1つの発話テキストは1つ以上の文から構成されている（1つの状況に対して複数の発話テキストの作成を許した）。
- (3) 作成された顧客の発話テキストに対して別の発話テキストデータ作成者が顧客の発話を確認するための相談員の復唱発話を考え、その発話テキストを作成する。復唱発話は顧客の発話の内容を簡潔に要約したものである。こうして作成された15種類の相談状況に関する顧客と相談員の発話テキストの対を対話テキスト対とよび、その集合を対話テキストセットとよぶ。  
以上ステップ 2, 3の作業を異なる5名の対話テキストデータ作成者が行い、5セットの対話テキストセットを作成した（のべ138対話テキスト対）。
- (4) 対話テキストセット中の発話テキストから、キーフレーズを抽出し、支援ルールを作成する。

表 2 実験条件

Table 2 Experimental conditions.

実験 A:

- 2対話テキストセット（30対話テキスト対）により支援ルールおよび認識用キーフレーズを作成
- 2対話テキストセットのテキスト入力および音声入力により評価

|     | 支援ルール数 | 認識用キーフレーズ数 |
|-----|--------|------------|
| 顧客  | 34 個   | 35 単語      |
| 相談員 | 36 個   | 41 単語      |

実験 B:

- 他の3対話テキストセット（108対話テキスト対）により支援ルールおよび認識用キーフレーズを作成
- 2対話テキストセットのテキスト入力および音声入力により評価

|     | 支援ルール数 | 認識用キーフレーズ数 |
|-----|--------|------------|
| 顧客  | 137 個  | 120 単語     |
| 相談員 | 120 個  | 101 単語     |

音声認識仕様

|        |                               |
|--------|-------------------------------|
| サンプリング | 8 kHz, 16 bit                 |
| 音声特徴   | 13 次ケプストラム,<br>13 次 Δ ケプストラム  |
| 認識モデル  | 混合連続型音素 HMM<br>(ガーベージ音素モデル使用) |

(5) 抽出されたキーフレーズを入力音声の中から検出するための音声認識用の認識フレーズ辞書を作成する。

### 4.3 評価手順

対話テキストセットのテキスト入力および音声入力により支援性能の評価を行った。入力音声は男性1名が入力データセットをオフィス内で読み上げることにより収録した。評価の基準は出力される支援と事前に設定した正解支援を比較し、その適合率と再現率で評価した。基準の定義は以下のとおりである。

$$\text{適合率} = \text{出力正解支援数} / \text{出力支援数}$$

$$\text{再現率} = \text{出力正解支援数} / \text{正解支援数}$$

2つの実験 A, Bを行った。実験の条件を表 2 に示す。対話テキストセット 5セットから2セットを選び出して実験 A, B 共通の評価データセット（30対話テキスト対）とした。データセット中の対話テキスト対が、特定の対話状況に偏らないように、各話題状況に対してそれぞれ1つずつ対話テキスト対が作成されている対話テキストセットを選択した。

実験 A では、音声認識誤りの影響のみを観察するため、評価データセットに対して、評価データセット自身から作成された支援ルールと認識用キーフレーズ語彙を使用して支援を出力した。また実験 B では現実に近い実験条件下の性能を見るため、評価データセットに対して評価データセット以外の3セット（108対話テキスト対）から作成された支援ルールセットと認識キーフレーズ語彙を使用して実験を行った。入力音声は顧客音声に関しては1文ずつ句点で分割して入

表 3 評価結果  
Table 3 Evaluation results.

| テキスト入力: |     |      |      |
|---------|-----|------|------|
|         |     | 実験 A | 実験 B |
| 顧客      | 適合率 | 96.8 | 100  |
|         | 再現率 | 100  | 26.7 |
| 相談員     | 適合率 | 100  | 90   |
|         | 再現率 | 100  | 60   |

| 音声入力:        |     |      |      |
|--------------|-----|------|------|
|              |     | 実験 A | 実験 B |
| 顧客<br>w=7000 | 適合率 | 34.8 | 25   |
|              | 再現率 | 76.7 | 70   |
| 顧客<br>w=6000 | 適合率 | 50   | 35.1 |
|              | 再現率 | 53.3 | 43.3 |
| 相談員          | 適合率 | 76.7 | 70   |
|              | 再現率 | 90   | 73.3 |

w はガーベージ音素モデルに対する重み

力しているが相談員音声に関しては一括して入力している。

キーフレーズの認識制御に関しては、顧客の発話に対しては、複数(0から無制限)のキーフレーズが検出できるよう、また相談員音声に対しては必ず1個のキーフレーズが検出されるよう制御を行った。これは、簡潔に復唱を行っている相談員音声はほとんどの場合、キーフレーズを含んだ短い発話であるのに対して、顧客の発話は冗長な部分を含んでおり、一意にキーフレーズを特定しにくく、複数のキーフレーズ仮説を用意する必要があるためである。そのため、入力される一つの顧客音声に対して、複数の支援が一括して出力される。

表 3 に実験結果を示す。

#### 4.4 考 察

テキスト入力の結果から、学習データから得られたキーフレーズだけでは顧客の発話の26.7%の再現率が得られておらず、学習データが不足していることが伺える。それに対して相談員の発話は60.0%と良い。これは相談員としての簡潔な表現(例:~に関するご質問ですね)を行った結果、表現のバラエティが顧客の発話に比べて少なくなったためと考えられる。

音声入力(実験B)に関しては、キーフレーズの認識仮説の出力しやすさの度合いを制御するための、ガーベージ音素モデルに対する重み付け(w)を変えた2種類の結果を比較した。図6に示すとおり、キーフレーズの認識仮説を多数出力させた場合(w=7000)、多数の支援候補が出力されたため、70%の再現率を得ることができた。これはテキストベースの再現率26.7%をはるかに超えている。この原因は音声入力の場合、入力音声に音響的に類似した認識辞書中のキーフレーズが、意味的にも類似している傾向があるため、(例:年

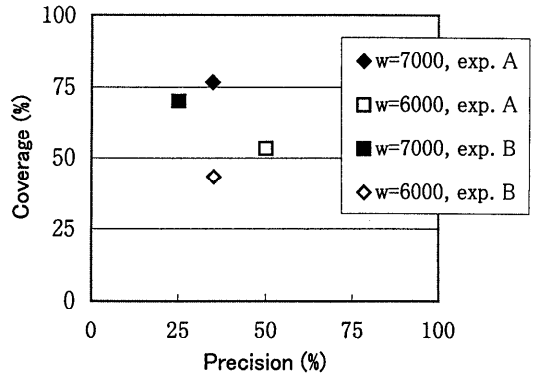


図 6 w の値による顧客発話音声に対する支援性能の変化  
Fig. 6 Performance change on clients' utterances by values of 'w'.

金の額と年金額, 受領額と受給額), テキスト入力ではマッチしなかった支援ルールが、正しいキーフレーズに認識される傾向があるからである。

またキーフレーズの検出精度を上げた場合(w=6000)、適合率は25%から35.1%に向上するが、その反面、再現率は70%から43.3%に低下している。

支援の性能という観点からは顧客音声に比べて相談員の復唱音声からの支援が優れていることがこれらの実験結果から分かる。そのため相談員音声のみのモニタリングによる支援方法も考えられる。しかし、顧客の音声による支援に比べて支援のタイミングが遅れるという問題がある。また、相談員の復唱は現実の相談業務で行われているが、完璧に復唱が行われていることは保証されていない。以上の理由および両者の音声の使用により、出力される支援情報のカバレッジが向上することを考慮すると、両者の音声に基づいた支援が有効であると考えられる。今後、対話音声をどのように支援に使用するかは検討する必要がある。

また、実験により得られた出力支援の性能を参考に、4.1節の実験形態により支援に含まれるノイズが、どのようにシステムのユーザビリティに影響するかを検証し、システム全体のユーザビリティを評価する必要がある。

## 5. 結 論

人間同士の対話を知的にサポートする知的CMCを提案し、その支援機構を備えたマルチエージェント型の年金相談業務支援システムを構築した。本システムは、ワードスポッティング型の音声認識技術と、パターンマッチングによる知識処理を組み合わせた、音声モニタリングによる自律的な支援機能を実現している。本システムは支援のための特別な操作が不要

であり、また誤った支援の出力に対してロバスタなコミュニケーション形態であるという特徴を持つ。

Wizard of Oz 法により我々の提案する自律的な対話支援のコンセプトを評価し、対話の効率化に寄与できることを確認した。また、ワードスポッティングを用いたキーフレーズの抽出による支援出力性能の評価実験により、顧客と相談員それぞれの発話に対する支援出力の性能を検証した結果、顧客の発話と比較して、相談員の復唱の発話は簡潔な表現になることにより、良好な支援精度が得られることが判明した。

謝辞 音声対話処理技術の応用について議論をいただいた当研究所の管村所長に感謝いたします。

### 参考文献

- 1) 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).
- 2) 西本, 角, 間瀬: Augmented 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, pp.259-266 (1996).
- 3) 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, pp.17-22 (1995).
- 4) 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, pp.157-158 (1996).
- 5) 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, pp.31-32 (1996).
- 6) Kuwata, Shiotsuka, Anchi and Koizumi: Multi-Agent Architecture for Consulting Support, *Proc. Practical Application of Agent and Multi-Agent Technology (PAAM-97)*, pp.61-70 (1997).
- 7) 城塚, 桑田, 安地, 小泉: 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, 情報処理学会研究報告, 97-SLP-15, pp.99-104 (1997).

(平成 9 年 6 月 30 日受付)

(平成 10 年 1 月 16 日採録)



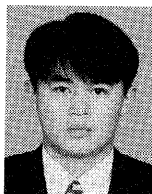
城塚 音也 (正会員)

1988 年東京大学文学部言語学科卒業。同年日本電信電話株式会社入社。以来音声言語処理の研究に従事。1990~1991 年 SRI インタナショナル客員研究員。現在 NTT データ通信 (株) 情報科学研究所にて音声言語および知識処理技術の研究開発を担当。音響学会会員。



桑田 喜隆 (正会員)

1986 年群馬大学大学院工学研究科修士課程修了 (電子工学)。同年日本電信電話公社 (現 NTT) データ通信事業本部に入社, エキスパートシステムの研究開発を行う。引き続き, 生産計画, リアルタイム AI, プランニング, エージェントの研究に従事。この間 (1991~1993 年) マサチューセッツ大学計算機科学科客員研究員。現在 NTT データ通信 (株) 情報科学研究所において知識処理技術の研究開発を担当。



安地 亮一

1995 年筑波大学理工学研究科修士課程修了 (物理工学)。同年 NTT データ通信 (株) 入社, 情報科学研究所にてヒューマンコミュニケーションの研究開発に従事。電子情報通信学会会員。



小泉 宣夫 (正会員)

1975 年京都大学大学院工学研究科修士課程修了 (精密工学専攻)。同年日本電信電話公社 (現 NTT) 電気通信研究所に入社, 音響振動, 拡声通話機器の研究開発を行う。引き続き, ヒューマンインタフェース研究所にて音場創成の研究に従事。この間 (1988~1989 年) マサチューセッツ工科大学機械工学科客員研究員。現在 NTT データ通信 (株) 情報科学研究所において音声処理技術, 知識処理技術の研究開発を担当。工学博士 (静岡大学)。