

コミュニケーション支援のための個人情報公開システム(PIP)

- 音声とキー入力を用いたマルチモーダル対話の検討 -

福井美佳、芝崎靖代、笹氣光一、竹林洋一

(株)東芝 研究開発センター

本報告では、個人間のコミュニケーションの円滑化と情報共有を促進する個人情報公開システムPIP(Personal Information Provier)について述べる。本システムでは、キーボード入力による日本語口語文と独り言音声を用いて、マルチモーダル対話によりユーザの意図と感情を認識する。さらに対人関係に基づいて個人が所有する情報をエージェントを介してリアルタイムで公開する。評価実験により、本システムの有効性が示された。

Multimodal Personal Information Provider Using Natural Language and Emotion Understanding from Speech and Keyboard Input

Mika FUKUI, Yasuyo SHIBAZAKI, Koichi SASAKI

and Yoichi TAKEBAYASHI

Research & Development Center, Toshiba Corporation 1,
Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan

Abstract. We have developed a multimodal Personal Information Provider (PIP), for enhancing information/knowledge sharing and closer human relations among groups. This system employs natural language and emotion understanding from speech and keyboard input, with user-initiative dialogue manager and multimodal response generator. The system runs in real time on a personal computer with an interface agent to open user's stored information to others with the user's permission. Experimental results have shown the effectiveness of the proposed system.

1.はじめに

近年、コンピュータやネットワーク環境の整備が進み、個人が多量の情報を能動的・受動的に取得し、所有することが可能になった。ネットワーク中から必要な情報を集めてくる情報収集エージェント[1]や、個人の興味に従ってデータベースの検索を行うインタフェース・エージェントなども実現されている[2][3]。また、マルチデータベース検索などを行う自然言語インタフェース[4][5]や音声対話システム[6]、人間の代わりに秘書業務や道案内などを行うやマルチモーダルインタフェースシステムが試作されている[7][8]。しかし、個人で所有している多量の情報を、他人と共有し有効活用するための支援についての研究は十分になされていない。個人間のコミュニケーションには、情報の交換に伴う時間のロスに加えて、情報の欠落・誤解、感情的な応酬による人間関係のもつれなどの問題が発生しやすい。これを嫌うユーザが多いため、莫大な情報が死蔵されていると考えられる。

これらの問題を解決するため、我々は、個人が所有する情報を所有者の代理人として対話的に公開する個人情報公開システム PIP(Personal Information Provider)を提案する。本システムは、個人が持つ情報や知識を個人の対人関係に基づいて相手に提供することを特徴とする。また、独り言音声とキーボード入力された口語文から感情を認識し、柔軟な対応を行うことにより、対人関係を悪化させない機能を持つ。

本稿では、個人情報公開システムの構想とプロトタイプシステムの開発について述べる。2章では、個人情報公開システムの目的、特徴を述べる。3章では、独り言とキー入力による感情認識手法について述べる。4章では、公開する情報をスケジュールに限定したプロトタイプシステムの構成と実装について述べる。5章で評価実験の結果を示し、最後にまとめと今後の課題を述べる。

2.個人情報公開システム PIP

2.1.PIPのインタフェース

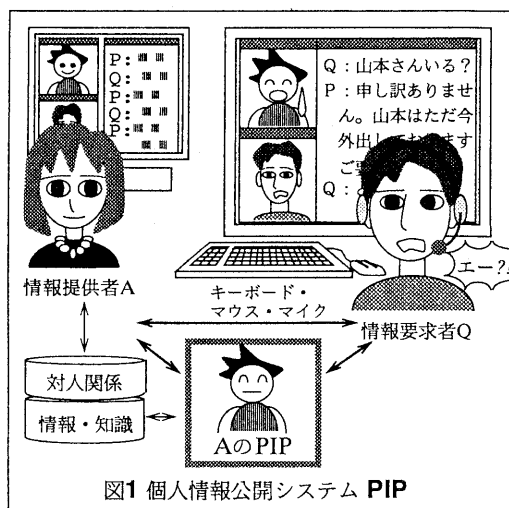
図1に、本システムの概念図を示す。左の情報提

供者Aが留守や多忙の際、右の情報要求者(ここではユーザと呼ぶ)がAの所有する情報を得るため、Aの個人情報公開システムPIPと対話を行う。ユーザとPIPの対話とそれぞれの画像は、ユーザのディスプレイに表示されている。

PIPは、ユーザの独り言やキーボード入力文から意図と感情を抽出する。情報を要求するユーザの負担を軽減するため、検索用のqueryではなくキーボードを介した自然言語の口語文を受け付ける。対話を通じて得たユーザの要求に従い、検索や伝言などを実行するが、そのユーザへの開示が許されていない情報は呈示しない。システムがユーザの要求を満たせずユーザの感情が悪化した場合、そのままでは情報提供者とユーザの関係までが悪化する危険性があるため、情報提供者が直接対応を行う機能を持たせた。ユーザの感情や、情報提供者とユーザの対人関係(親しさなど)、情報提供者の状況(不在、多忙、暇など)を考え併せて、情報提供者との直接対話を許すべきかを判断する。

2.2. 対人関係

集団には、地位や年齢、住居といった社会的な要因から決定されるフォーマル構造と、実際のつきあいによって形成される心理的關係によってできるインフォーマル構造がある[9]。本システムでは前者を社会的関係、後者を個人的な関係とし、あらかじめ登録した以下の対人関係情報を利用する。



社会的.... 部下、同僚、上司など

個人的.... 親密度、短気度、分野(仕事・専門分野・趣味・休暇)ごとの心理的距離

「短気度」は情報提供者が主観的に評価した相手の性格であるが、個人的な対人関係をあらわすものとして登録し、感情認識に利用する。

3.感情認識

3.1.感情空間モデル

感情を何種類かの基本的な感情に分類する試みは多数行われている[10]。Ortonyらは、感情をpositive-negative傾向をもつものと定義し、eventの望ましさを誘因とする満足、不満など、agentの行動への評価を誘因とする感嘆、非難など、そしてobjectの魅力誘因とする好意、嫌悪などの感情の3つに分類している[11]。

感情認識手法としては、一文あるいは一発話や表情などから、瞬間の感情を抽出する手法が提案されている[12]。金澤等の研究は、音声信号のピッチやアクセント等非言語情報である音声律情報に注目して、怒りや喜び等の意図や感情情報を理解、生成する[13]。また、物語などの文章から感情を表す言語表現や状況を抽出し、登場人物の感情状態を推論する試みがある[14][15]。重要なgoalが成功/失敗した際に、happinessやsadnessなどの疑似的な感情が生起されるエージェントやコンピュータモデルも開発されている[16][17]。

本システムでは「依頼」に伴う感情認識に限定し、Ortonyらの研究を基に以下の拡張を行った。前節の感情を引き起こす誘因にあげたobjectのかわりに、余裕、焦りなどの感情を引き起こすと思われるstate(ユーザの状況)を採用した。また、感情状態は以上3つの誘因の組み合わせによって表されると考え、図2のような感情空間上に感情状態を定義した。ユーザの入力する対話文と独り言音声に含まれる感情表現と、対話の遷移状態に基づく感情遷移モデルに従って、例えば、システムがユーザの意図をなかなか理解できないと、ユーザは徐々に不快になるといったユーザの感情状態を推定する。

3.2.キー入力からの感情表現の抽出

電子掲示板(chat)を介して図3のような対話データ2423文を収集し、表1に示すような410の感情表現を抽出した[18]。頻度1位の「^^」はフェイスマークと呼ばれ、キーボード対話で多用される感情表現である。以下、終助詞、間投詞、前置詞、接続詞などの頻度が高く、「困る」「嬉しい」などの動詞や形容詞による直接的な感情表現の頻度は低い。

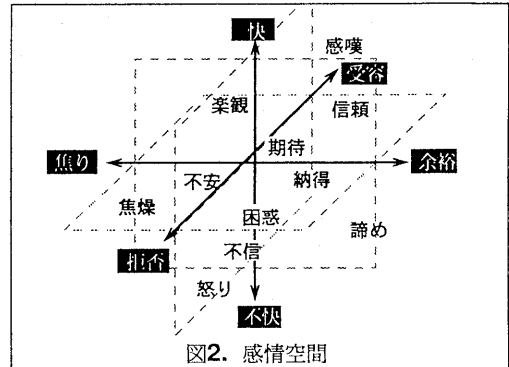


図2. 感情空間

A>こんにちは。
 B>こんにちは。ねえ、夜は何時までだったら電話していいの? 昨日質問したいことがあったのだが。
 A>そーねえ、10時頃までにしてくれるとありがたい。ゆーべは11時には寝たし。^^
 C>おお、はい!
 B>じゃ、昨日は無理だったか。ディスクの入れ替えしてて遅かったから。
 A>たいいてい12時頃まで起きてるけどね。

:感情表現

図3. キーボード対話例

表1 感情表現の例

出現回数	感情表現	品詞	表現される感情状態
34	^^;	f.	不安、焦燥、期待
33	なあ	post.	不安、期待、納得
24	かなあ	post.	不安
22	かも	post.	不安、期待
22	うん	int.	納得
21	でも	prep.	諦め
19	でも	conj.	焦燥、怒り
18	^^	f.	感嘆、期待
16	だな	post.	不安、納得
15	まだ	adv.	楽観、焦燥
15	ちゃんと	adv.	焦燥
15	ごめん	int.	諦め、納得

f. フェイスマーク post. 終助詞
 int. 間投詞 conj. 接続詞
 adv. 副詞 prep. 前置詞

3.3. 独り言音声からの感情抽出

計算機の操作中に感情を表す独り言が発声されることが多いのではないかと考え、計算機業務に従事する研究者26人を対象に、独り言についてのアンケートを行った。アンケートの結果に基づき頻度の高い独り言13語を選択した。

非言語音声認識処理によりこれらの独り言を認識する[13][19]。音響分析と始終端検出の後、非言語音声の特徴ベクトルが抽出され、認識辞書とのパターン照合が行われる。1語彙あたり150データ(30回×5人分)を収集し、認識辞書を作成した。収集した独り言の非言語音声特徴ベクトルの例を表2に示す。データ収集にあたり、日常発声する独り言を選択させ、事前に発声テストを行って抑揚や長さが揃うよう配慮した。

認識性能の評価のため、別途5人分13語彙1060データを収集し認識性能を求めた。5人中2人は、認識辞書にデータを提供した特定話者、3人は辞書にデータを提供していない不特定話者である。その結果、13語彙では特定話者は78%(584/751)、不特定話者は62%(190/309)、9語彙758データでは、特定話者は88%(463/525)、不特定話者では78%(181/233)だった。認識精度が低い原因として、語彙数が多いことと、各被験者から全語彙のデータを収集したため、日常発声しない不自然な独り言のデータまで含まれている可能性が考えられた。そこで、各被験者の日常発声する独り言5語彙に限定して認識性能を求めたところ、特定話者の平均は97%(277/286)、不特定話者で、93%(129/138)となり、各被験者の日常発声する独り言に関しては十分な認識性能が得られた。

表2 独り言音声の時間周波数スペクトル(特徴ベクトル)

発話	感情	特徴ベクトル	発話	感情	特徴ベクトル
えー?	怒り		なんで	不信 困惑	
あーあ	諦め 焦燥		なるほど	期待 信頼 納得	
はぁー (溜め息)	諦め 焦燥		そーかあ	感嘆 納得	
あれ?	不安		おー	喜び 感嘆	

4. プロトタイプシステムの開発

4.1. システム構成

図5にプロトタイプシステムの構成図を示す。各プロセスは主にCで開発しsparc stationでリアルタイムに稼働している。約2秒で1応答文を生成、表示する。

ユーザの入力を解析するのは、独り言の音声認識部(図5右上)と、自然言語と画像の入出力を行うインタフェースプロセスである。サーバプロセス(図5左上)は、ユーザとの対話管理を行う対話管理部と、感情認識を行いユーザと情報提供者の直接対話のタイミングを決定する感情認識部からなる。検索部(図5左下)はスケジュール情報の検索を行う。

感情認識部は、対話の履歴およびインタフェースプロセスから渡される感情表現抽出結果、システムの応答結果と、ユーザの入力に対する意図抽出および感情語抽出結果からユーザの感情を決定する。情報提供者とユーザとの直接対話の指示が出ると、情報提供者のディスプレイにも入出力用のウィンドウが表示され、ユーザのインタフェースプロセスと通信を行う。

4.2. 対話例

仮想的な情報提供者「山本かえで」を想定し、スケジュール情報と、「山本かえで」の上司や同僚、友人などの仮想の対人関係情報を複数登録した。図6に実際の対話画面を示す。左上に情報公開エージェントの画像、左下にユーザ自身の画像

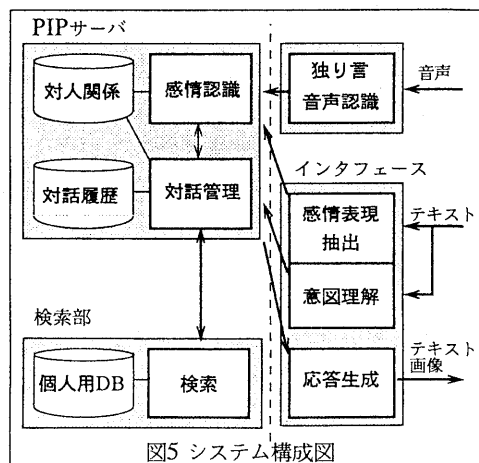


図5 システム構成図

(ユーザの似顔絵)が表示され、右のウィンドウでユーザとエージェントの対話をテキストベースで行う。検索されたスケジュールがグラフで表示されている。図7に他の対話例を示す。

4.3. 意図抽出

意図抽出部は口語入力文を対象とするため、構文解析は行わず、形態素解析とキーワード抽出の結果からユーザの意図と、スケジュール検索に必要な日付、場所などを抽出する。抽出する意図は、挨拶、要求、肯定、否定、表明とし、要求の種類はスケジュール情報の公開、伝言、情報提供者との直接対話、終了に限定して、口語入力文の解析を可能に

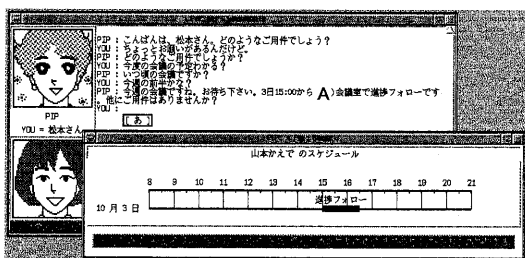


図6 対話画面

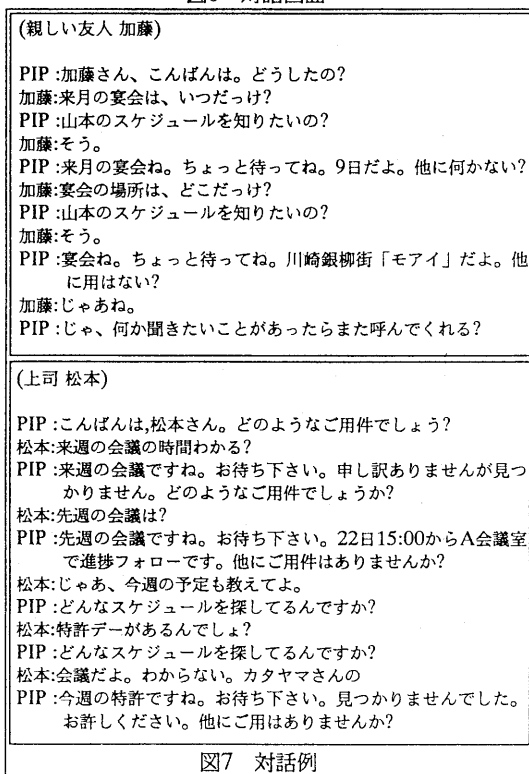


図7 対話例

した。キーワード辞書の見出しは806語。システムの利用形態を想定した180の対話対を含む対話例30を作成し、発話意図の特定や付加情報を伴う言語表現を抽出、発話意図や付加情報の種類によってキーワードを意味カテゴリに分類して用いた。

4.4. 対話管理

対話管理部は、スケジュール検索、情報提供者の状態、伝言管理、直接対話の4つのサブネットワークを持ち、意図抽出部から渡されたユーザの発話意図によって遷移し、応答表現を生成する。

4.5. 応答生成

応答文の生成はスロット法にて行った。応答文例は885登録されている。応答の意図としては、承諾、拒絶、解答、相づち、確認、混乱、終了の宣言、挨拶、要求、謝罪、提案、感謝がある。要求の種類は、スケジュール、ユーザ同士の直接対話、伝言、終了などがある。

対話管理部から送られる応答表現中の、システムの意図、要求種類、文型、その他の情報や、応答に表情を与えるための態度や親密度などの表情情報が合致する文例を探し、必要な情報を埋めて表示する。また、図6のエージェントの画像の表情は複数用意しておき、応答表現で指定された発話意図の種類と親密度や態度によって画像を変更している。

5. 評価実験

現状の対話機能とマルチモーダル感情認識の性能評価実験を行った。被験者は独り言などを利用した感情認識を行う事をあらかじめ説明したA群(8名)と説明しなかったB群(7名)で行った。課題は3つ用意し、比較的容易な課題1と、必ず失敗する課題2、複数の検索や伝言のタスクを実行する課題3の順に行い、各課題終了後にアンケートを行って課題の実行開始時、実行中、終了時の感情状態について、positive-negativeの7段階で記入してもらった。

タスク達成率は、必ず失敗する課題2をのぞけば、1回目に成功した対話は22/30=73%、2回目を合わせれば87%となった。感情認識の性能は課題1の実行開始時、実行中、終了時のすべての感情を認

識できたのが53%、終了時の感情に限れば73%であった。今回は1回のタスク実行に伴う感情認識手法を用いており、特に課題3のような複数のタスクを実行する際の感情遷移状態の認識は対象外だったせいか、課題2、3におけるB群の認識性能が落ち込んだ。一方、A群は独り言やキー入力による感情表現がB群より多く見られ、操作に慣れるにしたがって感情認識率が上がり、課題3のすべての感情で75%、終了時に限れば87%となり、十分な性能を得ることができた。

タスク達成率は悪くはなかったが、意図理解の性能不足や応答文のわかりにくさなどにより対話が長引くケースが見られ、感情認識の性能を落とす要因となった。感情認識の性能をあげるためにも、まずは対話の性能を向上させる必要がある。

6.おわりに

個人間の知識共有とコミュニケーション支援を目的とした個人情報公開システムを提案し、プロタイプシステムを開発した。スケジュール情報の公開に限定し、形態素解析とキーワード抽出により自然言語入力文からユーザの意図を抽出、あらかじめ登録された対人関係に基づくスケジュール情報公開機能を実現した。また、独り言13語の非言語音声認識辞書を作成し、評価実験の結果、各被験者の日常発声する独り言に関しては十分な認識性能を得られた。キーボード入力文と非言語音声認識手法を用いた感情認識機能を実現し、独り言や感情表現の含まれる対話における有効性が示された。

今後は評価実験の結果に基づき対話と感情認識の性能をあげていくと共に、個人の知識を簡単に登録し公開できるようにしていく。また、対人関係情報の自動抽出手法を開発していく。

参考文献

- [1] P. Maes : "Agents that Reduce Work and Information Overload", CACM, Vol.37, No.7, pp.31-40 (July, 1994).
- [2] T. Oren et al.: "Guides: Characterizing the Interface", The Art of Human-Computer Interface Design, Addison-Wesley, pp.367-381 (1990).
- [3] B・ローレル著、遠山峻征訳："劇場としてのコンピュータ", アジソン ウェスレイ・トッパン (1992).
- [4] 住田一男 他："マルチデータベース日本語インタフェースの試作-日本語入力文解析処理-", 情報処理学会第51回全国大会, 4, pp.67-68 (1995).
- [5] 浮田 他："自然言語入力による機器操作案内システム", 信学技報 OS88-18, pp.13-18 (1988)
- [6] 島津明 他："対話の分析 - 間投詞的応答に注目して -", 信学技報, NLC93-9, pp.65-71, (1993-05).
- [7] 竹林洋一, "音声自由対話システムTOSPURG II - ユーザ中心のマルチモーダルインタフェースの実現に向けて -", 電子情報通信学会論文集, Vol.J77-D-II, No.8, pp.1417-1428 (1994).
- [8] 長尾確："マルチモーダル・ヒューマンコンピュータインタラクション", 人口知能学会研究会資料, SIG-SLUD-9303 (1993).
- [9] 狩野素朗："個と集団の社会心理学", ナカニシヤ出版 (1992).
- [10] 福井泰之："感情の心理学", 川島書店.
- [11] Ortony, A., Clore, G. L. and Collins, A. The Cognitive Structure of Emotions, Cambridge University Press, (1988).
- [12] 森島繁生 他："自然な表情アニメーションのための感情空間の構成", NICOGRAPH論文集, pp.17-25, (1993).
- [13] 金澤博史 他："計算機との対話のための非言語音声の認識と合成", 電子情報通信学会論文集, Vol.J77-D-II, No.8, pp.1512-1521 (1994).
- [14] Lehnert et al : "The Role of Affect in Narrative Structure", Cognition and Emotion, pp.299-322 (1987)
- [15] M. G. Dyer : "Emotions and their Computations: Three Computer Models", Lawrence Erlbaum Associates Limited (1987)
- [16] Bates, J. Loyal, A. B. and Reilly, W. S. An Architecture for Action, Emotion, and Social Behavior, in Proc. 4th European Workshop on Modeling Autonomous Agents in a Multi-Agent World, MAAMAW '92, 1992, pp. 55-68.
- [17] Swagerman, J. The Artificial Concern REalization System ACRES - A computermodel that has emotions and can name them.
- [18] 芝崎靖代、福井美佳、竹林洋一："キーボード対話からの感情抽出に関する検討", 情報処理学会第51回全国大会, 2, pp.17-18 (1995).
- [19] 橋本秀樹 他："ワークステーションにおける音声認識インタフェースの検討", IPSJ-HI, HI-46-3, pp. 17-24, (1992).