

平成21年 5月20日現在

研究種目：基盤研究（C）  
 研究期間：2006～2008  
 課題番号：18500115  
 研究課題名（和文） マルチメディアコーパスと確率モデルを利用したマルチモーダル対話制御に関する研究  
 研究課題名（英文） Studies on Multimodal Dialogue Control using Multimedia Corpus and Probability Model  
 研究代表者  
 遠藤 勉（ENDO TSUTOMU）  
 九州工業大学・大学院情報工学研究院・教授  
 研究者番号：10112294

研究成果の概要：音声発話とジェスチャを併用して対話のできるシステムを目指して、人間の行う対話データやジェスチャデータを収集し、人手による解析（タグ付け）を行った。これらのデータ（コーパス）を利用して、音声認識ソフトウェアの認識精度の向上、対話における発話意図の推定、ハンドジェスチャの認識等を行うプログラムを開発し、画面上で写真を操作するインタフェースや対話型病名診断システムに適用して、その有効性を確認した。

交付額

（金額単位：円）

|        | 直接経費      | 間接経費    | 合計        |
|--------|-----------|---------|-----------|
| 2006年度 | 1,300,000 | 0       | 1,300,000 |
| 2007年度 | 600,000   | 180,000 | 780,000   |
| 2008年度 | 500,000   | 150,000 | 650,000   |
| 総計     | 2,400,000 | 330,000 | 2,730,000 |

研究分野：知能工学

科研費の分科・細目：情報学・知能情報学

キーワード：自然言語処理，マルチモーダル対話，人工知能，ジェスチャ認識，確率モデル，機械学習，音声認識，コーパス

## 1. 研究開始当初の背景

人間と機械の自然なインタフェースの実現を目指して、音声発話、ジェスチャ等を併用したマルチモーダル対話システムに関心が集まっているが、システムの多くは対話手段を複合的に利用できるようにしたものであり、複数のモダリティを統合するための一般的な理論や、対話の流れを制御するためのモデルの確立には至っていない。言語学における語用論の分野では、対話における発話解釈を発話文の意味だけではなく、話し手がその文を用いて聞き手に伝えようとする意図したメッセージを、文脈を参照しながら推論によって把握することであると定義しており、文脈処理が発話解釈の中心的な課題と位置付けられている。しかしながら、機械上でマル

チモーダル対話システムを実現するには、以下の問題点を工学的に解決する必要がある。[1]一連の対話は時間的・空間的に連続したものであり、マルチメディア文書のように明確な解析単位が存在しない。[2]連続するという制約から対話には無意味な表出（フィラーやつなぎ動作等）が存在する。[3]解析単位が明確でないため表出と意味との対応が1対1でなく、意味表現の設定が難しい。[4]使用モダリティの抽象度のレベルに差異があり、統合レベルの設定が難しい。[5]文脈情報の位置付けが明確でなく、一貫した理論に基づく利用法が確立されていない。

## 2. 研究の目的

本研究は、コーパスと確率モデルを利用し

てマルチモーダル対話システムの構築に必要な要素技術と制御技術を開発することにより、上記1. で述べた問題点の解決を図るとともに、対話支援型問題解決システムに適用してその有効性を検証しようとするものであり、具体的に以下の課題に取り組む。

- (1) 初等数学の教示をドメインとする既存のマルチメディアコーパスを整備するとともに、他のドメインのコーパス収集とコーパス管理ツールの開発を行う。
- (2) 対話システムの入力モダリティ(音声発話、ジェスチャ)認識モジュールを開発する。
- (3) コーパスに基づく発話意図推定のアルゴリズムを開発する。
- (4) 音声発話とジェスチャを併用したマルチモーダルインタフェースを構築する。
- (5) 文脈情報を対話システムの基本知識源と位置付け、文脈情報の表現・利用法を検討してその有効性を検証する

### 3. 研究の方法

#### (1) コーパスの整備と管理ツールの開発

##### ①マルチメディアコーパスの整備

小学校1年算数の世界を対象にして、ドリルテキスト並びに教師との対話を相互参照しながらドリルの問題解決を行うシステムの開発過程で、対話コーパスの収集を行った。被験者24名による模擬対話をビデオカメラで撮影し、全72対話2211発話を収集した。その内の27対話は人間対人間による模擬対話、45対話はWizard of OZ法による人間対機械の模擬対話である。

音声データは人手により書き起こし、フィルターや言い淀み、言い直し等話し言葉特有の現象にタグ付けを行った。さらに、各語の品詞情報、語間の係り受け情報(係り、受け、関係ラベル)を付与した。関係ラベルとは、sbj, obj, ins, loc等の格関係及び名詞を修飾する語間の関係のことである。

##### ②Webからの評価文抽出

WeblogやWeb掲示板等の普及により、製品やサービス等に対するレビューを投稿する機会が増えている。レビューは消費者が製品を選ぶとき、企業がマーケティングやリスク管理等を行うときの貴重な情報源となる。レビューから評価文を抽出するには人手による大規模な評価表現辞書やパターンを利用して行うのが普通であるが、多大なコストがかかる。そこで人手により少量の評価文を与え、文章の類似度に基づいて評価文を抽出する方法を試みた。

文章の類似度は以下の手順で算出する。[1]文章を形態素解析及び係り受け解析して、階層非循環有向グラフに変換する。[2]比較する文章のグラフからすべての階層部分属性(品詞情報、単語情報、複合名詞情報等)を抽出する。[3]2つの文章から抽出された階層部分

属性列それぞれについて一致度を計算し、その総和を2つの文章の文節数で除算する。

##### ③タグ付け支援ツールの開発

Web上のゲームレビュー文書を対象にして評価項目タグ(オリジナリティ、グラフィック、音楽、熱中度、満足度、快適度、難易度)と極性タグ(好評/不評)を付与する作業を支援するため、以下の機能を持つツールを開発した。[1]アノテータの負担を軽減させるため、マウス操作でタグ付けを可能にするGUIを構築する。[2]各評価項目と関連度の高い表現を抽出し、強調表示する。[3]タグ付けの揺れを減少させるため、同一文書からのタグ付き事例と別文書からの類似事例を提示する。

#### (2) 対話における入力モダリティの解析

##### ①キーワード抽出に基づく音声理解

音声発話はマルチモーダル対話システムにおける基本的なモダリティであるが、既存の音声認識エンジンの精度は必ずしも高くなく、認識誤りに対する対処が必要である。そこで、音声認識エンジンが出力したN-best候補に対して、上記(1)①のコーパスを利用し、係り受け関係を付与したキーワード列を生成する音声理解モジュールを開発した。処理手順は以下の通りである。[1]音声認識ソフトウェアJulianを用いて、入力発話のスコア付きN-best候補を求める。[2]形態素長と品詞情報を用いて形態素長の正規化を行った後、N-bestにおける出現頻度と認識エンジンのスコアに基づいて各語に対する信頼度(事後確率)を計算する。この中で内容語(動詞、名詞、形容詞、指示詞等)をキーワード候補とする。[3]キーワード候補から入力発話に対するキーワード列候補を求め、キーワード列に属する語の連想確率を計算する。連想確率とは、共起確率に比例し、連想情報量(1語当たりの平均共起確率)で重み付けを行ったものである。キーワード列 $kp_i$ の各語の信頼度の積及び連想確率の積を調和平均した値 $P(kp_i)$ を用いてキーワード列候補をランキングする。[4]キーワード列中の語に対して、日本語の係り受け関係の制約(文末を除く全ての語は文末側のいずれか1つの語に係る、係り受け関係は交差しない)を満たす係り受け関係を全て生成し、以下の式でキーワード列のスコアを求める。

$$SS_{kp_i} = P(kp_i) + \prod_{j=1}^{m-1} w_{k_j, k_i} \cdot dist_j$$

ここで、 $w_{x, y_i}$ は語 $x$ 語 $y_i$ の係り受け関係の重みで、コーパスを用いて次式で計算する。

$$w_{x, y_i} = \frac{freq(x, y_i, label)}{\sum_{j=1}^m freq(x, y_j, label)}$$

ゼロ頻度問題に対応するため、 $x$  または  $y_i$  がコーパス中に存在しない場合は、語の代わりに品詞を用いて重みを計算する。 $dist_j$  は語  $k_j$  と  $k_i$  の距離に応じた重みで、2語が離れるほど小さくなる。 $SS_{k_i}$  を用いて、係り受けラベル付きキーワード列候補を再ランキングする。[5] キーワード列に付与された係り受け関係全体の整合性をチェックするために、FREQT アルゴリズムによりコーパスから頻出係り受けパターンを抽出し、パターンの生起確率を計算する。これを[4]で求めた  $SS_{k_i}$  に加えた値をキーワード列の最終スコアとして再度ランキングし、第1位のキーワード列を求める。

## ② 頭部ジェスチャ認識

「うなずき」、「首振り」、「首かしげ」等の頭部ジェスチャも音声発話を補完する重要な伝達情報である。そこで、以下の手順で頭部ジェスチャを認識するモジュールを開発した。[1] Haar like 特徴を用いて顔（正面、左側面、右側面、30度傾き）、目、鼻、口の領域を抽出する。[2] 連続した20フレームの画像に対して[1]で抽出した各領域のオプティカルフローを求め、「方向無し」も含めて5方向（または9方向）に量子化する。[3] 各フレーム毎に顔の向き、オプティカルフローの大きさと方向、両目の位置のずれ、顔の中心に対する口の横の位置等の18素性（20フレームで360素性）、20フレーム全体に対する顔の向き、オプティカルフローの方向コードの総和等26素性、合計386素性を求める。[4] 上記の素性を基に決定木学習アルゴリズム C4.5 を用いて頭部ジェスチャ「うなずき」、「首振り」、「首かしげ」、「その他」を識別する決定木を生成する。

## (3) 対話システムにおける発話意図推定

対話システムにおいて、ユーザの発話意図を正しく理解し、それに基づいて自然なインタラクション処理を行うことは重要である。そこで、上記(1)①の対話データに意図タグを付与し、意図タグ付き対話コーパスを構築した。対話のドメインである問題解決の教示に関して18種類の意図タグを設定した。その一部を表1に示す。

表1 発話意図タグの例

| 意図   | 属性       | 意味内容        |
|------|----------|-------------|
| 情報提供 | Howto    | 解法に関する情報の提供 |
|      | Confirm  | 解くのに必要な情報提供 |
| 同意表示 | Correct  | 確認要求に対する同意  |
|      | Response | 上記以外の同意表示   |
| 行為要求 | Solve    | 解法に関する行為要求  |
|      | Inform   | ヒントを含む行為要求  |
|      | Other    | 上記以外の行為要求   |

発話意図の推定には2つの手法を試みた。

### ① 発話文の類似度を用いる手法

入力発話文とタグ付きコーパス中の発話事例との表層上の類似度により意図推定を行う。入力発話と発話事例を、形態素解析により得られる形態素を要素とするベクトルで表現し、以下の9個の尺度により類似度を計算する。[1] 形態素の有無に対応した2値ベクトルの内積、[2] 2値ベクトルの余弦、[3] 形態素の出現頻度ベクトルの内積、[4] 出現頻度ベクトルの余弦、[5] 出現頻度を補正した内積、[6] 出現頻度を補正した余弦、[7] Dice 係数、[8] Jaccard 係数、[9] 形態素の一致度。最後に、類似度最大の事例に付与された発話意図タグの中で最も数の多いもの（多数決方式）を推定結果とする。

## ② 最大エントロピー(ME)法を用いる手法

以下の手順で発話意図を推定する。[1] 発話事例を形態素解析して、得られる全形態素を素性とする。[2] 発話事例に対して、発話意図タグと出現する素性をまとめたものを訓練データとする。[3] 発話文の集合を  $B$ 、出力となる発話意図の集合を  $A$  とするとき、発話文  $b \in B$  で出力値  $a \in A$  となる事象  $(a, b)$  の確率分布  $p(a, b)$  を ME 法により求める。素性関数は発話文  $b$  において素性  $f$  が観測され、かつその発話意図が  $a_i$  となるときに1を返す関数とする。素性関数の重みは一般化反復スケール法で推定する。[4] 入力発話を形態素解析して、最大の確率値が得られた発話意図を推定結果とする。

## (4) マルチモーダルインタフェースの構築

ハンドジェスチャと音声発話を併用して、画面上の写真进行操作（カーソルの移動、写真の選択と解除、写真の移動、写真の拡大・縮小）するインタフェースを構築した。操作のイメージを図1に示す。



図1 写真操作アプリケーション

## ① ハンドジェスチャ認識

色の付いたマーカを両手の親指と人差し指に装着してハンドジェスチャを行い、マーカの位置と動きをUSBカメラで撮像して、認識を行う。写真の選択は人差し指と親指を合わせる、移動は選択したまま動かす、拡大は写真を両手で選択した後手を外側に移動

する、縮小は写真を両手で選択し、手を内側に移動する。認識は以下の手順で行う。[1]マーカの認識は色情報で行うが、照明条件の変化、背景や服にマーカと同じ色が存在すること等を考慮して、肌色に接している色領域をマーカと認定することとし、認識開始前に画面の指定の場所に手を合わせて、マーカと肌色の初期設定を行う。[2]背景差分を利用してマーカ領域と肌色領域を抽出し、当該フレームのマーカ位置を決定する。[3]前フレームと現フレームのマーカの位置を比較することによりマーカの動きを検出する。[4]連続DPマッチング法を用いてジェスチャの種類(写真操作)を決定する。

## ②音声認識

音声発話はハンドジェスチャで選択された写真に対するコマンド(ここでは、拡大・縮小)認識のために使用する。従って、使用語彙数は少ない。しかしながら、多人数で写真操作を行う場合は、操作とは直接関係のない対話も音声情報として入力されることになる。そこで、タスクに関連した発話のみの認識が可能な小語彙音声認識器 **Julian** と語彙数は多いが認識精度が高くない大語彙認識器 **Julius** を組み合わせて、両者の認識結果を類似度等に基づき統合することで写真に対するコマンド認識を試みた。

## (5) 文脈情報の定義と対話管理

対話システムは、文脈を考慮して、ユーザ発話を理解し、必要なタスクを実行してシステム発話を生成する。文脈とは、狭義には文間の脈絡、すなわち対話履歴を意味するが、広義には言語表現が使用され、評価される言語的・非言語的環境、すなわち、言語理解、タスク実行及び言語生成に使用する全ての知識として定義される。

本研究では、人間の言語理解を、記憶内に内在する知識構造と入力文中の情報との適切な統合であると考え、エピソード記憶(EM)、談話記憶(DM)、基底記憶(GM)の3階層からなり、知識表現に拡張格フレーム構造を使用した概念ネットワーク(Three Layer Conceptual Network, TLCN)を提案し、これに基づく対話処理アーキテクチャを開発した。基底記憶は対話システムに必要な全ての知識を、エピソード記憶は入出力発話に関する情報を保存する。両記憶の中間に位置する談話記憶は、一つの対話セッションにおける発話文理解とシステムのタスク遂行に必要な文脈情報を保存する。一方、拡張格フレームとは通常の格フレームにおける名詞概念と動詞概念間の格関係に、名詞概念間及び動詞概念間の関係を追加し、さらに名詞概念をドメインに応じたクラスに分類することにより、言語的知識とドメイン知識を統一的に表現できるようにした枠組である。

TLCN に基づく対話処理アーキテクチャ(TLCN アーキテクチャ)は、3つの記憶構造に、その管理に必要な処理手続き(EM:発話文理解・生成, クラス分類等, DM:入力情報とドメイン知識の照合, 応答文生成等, GM:知識の探索等)を付加した3つのモジュール(EMM, DMM, GMM)と、システム全体の制御並びにタスク遂行を司る対話マネージャから構成されており、訓練及び対話の2つの動作モードを持っている。TLCN アーキテクチャの構成を図2に示す。

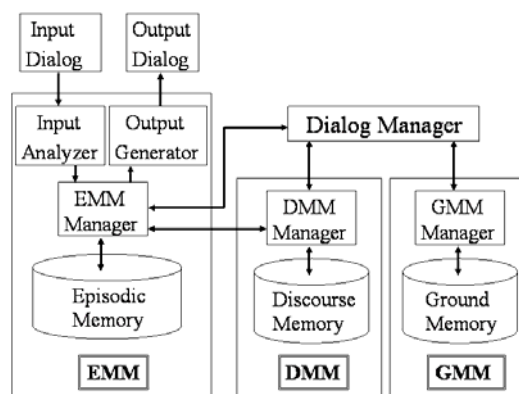


図2 TLCN アーキテクチャの構成

## 4. 研究成果

### (1) コーパスの整備と管理ツールの開発

#### ①Webからの評価文抽出結果

人手により評価文10文を与え、これらとWeb上のレビュー1052文との類似度を総当たりで計算し、用意した評価文1文毎に類似度上位N文を抽出した。類似度計算に使用するパラメータを調整することで、文の構造を重視する場合と、表層を重視する場合の2通りの類似文を抽出した。さらに、構造重視と表層重視のそれぞれで抽出した文集を組み合わせることで、新たな抽出文集を求めた。それぞれの場合の抽出精度(%)を表2に示す。

表2 評価文の抽出精度

|          |      |
|----------|------|
| 構造重視     | 88.9 |
| 表層重視     | 91.5 |
| 組合せ(AND) | 95.7 |
| 組合せ(OR)  | 88.3 |

#### ②タグ付け支援ツールの有効性評価

試作したツールの有効性の確認実験を行った。アノテータは2名で、ケース1:支援ツール無し、ケース2:GUI有り+強調表示有り、ケース3:ケース2+類似事例提示有り、ケース4:ケース3+タグ付き事例有り、の4つの条件で、データを変えて2回実験した。その結果、完全な手作業によるアノテーションと比較し、支援ツールを使うことで作業時間を3~4割削減できた。さらに、タグ付けの一致度( $\kappa$ 値)を1回目で0.350(ケース

1)から0.753(ケース4)に、2回目で0.491(ケース2)から0.687(ケース4)まで向上させることができた。

(2) 対話における入力モダリティの解析

①キーワード抽出に基づく音声理解

5人の被験者の各50発話、計300発話に対してキーワード抽出手法の評価を行った。N-bestにおける、Nは10とした。発話中の平均キーワード数は4.42語である。連想確率の計算には整備したコーパス中の2211発話に加えて、重点領域研究「音声対話」コーパス中の1084発話も利用した。評価は、10-bestリストからどの程度正しいキーワードを抽出できるかを測定するため、尺度として、1位精度=(各発話における全てのキーワードを含んだ第1位のキーワードリストの数)/(発話数)、適合率=(第1位のキーワードリスト中で正しく抽出されたキーワード数)/(第1位のキーワードリスト中のキーワード数)、再現率=(第1位のキーワードリスト中で正しく抽出されたキーワード数)/(発話におけるキーワード数)、を用いた。抽出結果を表3に示す。手法の項目において、A:認識エンジンの出力、B:信頼度を用いた復元、C:(信頼度+連想確率)による復元、D:Cと係り受け関係による復元、である。

表3 キーワードの抽出結果

| 尺度\手法 | A     | B     | C     | D     |
|-------|-------|-------|-------|-------|
| 1位精度  | 43.67 | 34.33 | 46.00 | 51.33 |
| 適合率   | 80.64 | 78.21 | 81.86 | 85.19 |
| 再現率   | 85.14 | 82.58 | 86.43 | 85.44 |

②頭部ジェスチャ認識

頭部ジェスチャ「うなずき」、「首振り」、「首かしげ」、「その他」の動画像を各600種類収集し、その内各500動画像、計2000動画像を学習データとして、C4.5で決定木を作成した。残りの各ジェスチャ100動画像をテストデータとして、認識実験を行った。認識精度を表4に示す。オプティカルフローの方向数は5である。方向数を9にしても精度に大きな違いは見られなかった。

表4 頭部ジェスチャの認識精度

| 入力\結果   | A  | B  | C  | D  |
|---------|----|----|----|----|
| うなずき(A) | 81 | 0  | 0  | 19 |
| 首振り(B)  | 2  | 93 | 0  | 5  |
| 首かしげ(C) | 0  | 0  | 88 | 12 |
| その他(D)  | 4  | 3  | 1  | 92 |

(3) 対話システムにおける発話意図推定

意図タグ付きコーパス中の81対話を用いて意図推定実験を行った。この内1対話をテストデータ、残り80対話を事例データとする交差検定で評価した。類似度による推定

では、9個の尺度の有効性を評価するために予備実験を行った結果、2値の内積(Inn)、Dice係数、Jaccard係数、形態素の一致率(CorM)を用いる手法の正解率が高かった。推定結果を表5に示す。問題1~3は、算数の問題の種類を表し、その問題を解決するときになされた対話の発話意図を推定したものである。いずれの問題についても、ME法を用いる方が高い推定率が得られた。

類似度を用いる手法では、最終的な結果は多数決で判定しているが、類似度最大の候補が複数ある場合(ME方の場合は確率値が最大の候補が複数ある場合)には、その中に1つでも正解が含まれていれば正解とする基準、すなわち候補正解率で評価するための追加実験を行った。ここで、候補正解率=(推定結果の中に正解が1つ以上含まれているテストデータ数)/(テストデータ総数)、である。推定結果を表6に示す。この条件では、2値の内積による類似度法が最も高い候補正解率を得た。

表5 発話意図推定結果(正解率:%)

| 手法  | 発話文の類似度 |       |       | ME法   |
|-----|---------|-------|-------|-------|
|     | Inn     | Dice  | CorM  |       |
| 問題1 | 80.23   | 81.53 | 83.21 | 84.98 |
| 問題2 | 90.43   | 88.72 | 89.11 | 92.32 |
| 問題3 | 89.13   | 88.43 | 89.14 | 92.42 |
| 平均  | 86.60   | 86.22 | 87.16 | 89.91 |

表6 発話意図推定結果(候補正解率:%)

| 手法  | 発話文の類似度 |       |       | ME法   |
|-----|---------|-------|-------|-------|
|     | Inn     | Dice  | CorM  |       |
| 問題1 | 88.74   | 84.38 | 86.19 | 85.21 |
| 問題2 | 96.01   | 92.59 | 93.00 | 92.32 |
| 問題3 | 93.96   | 90.11 | 90.82 | 92.42 |
| 平均  | 92.92   | 89.03 | 90.00 | 89.98 |

(4) マルチモーダルインタフェース

ハンドジェスチャと音声発話を併用して画面上の写真操作を行うインタフェースの認識精度を検証した。被験者は5名の学生A~Eで、3つの操作をハンドジェスチャのみで20回行った結果を表7に示す。選択操作は常時高精度な認識が可能であったが、拡大・縮小の精度は約75%であった。ただし、試行回数を重ねるにつれて認識精度の向上が見られた。簡易なマークを利用することで、高精度な指先の抽出が可能になり、写真の選択や移動、拡大・縮小といった作業を直感的に制御でき、しかも複数のマークを同時に追跡することができるため、マウス等よりも多様な操作が実現可能であることを確認した。

画像が選択された後の拡大・縮小に対して音声発話を併用した結果、応答時間に多少問題を残すものの正確な動作が可能であった。ハンドジェスチャと音声発話を併用する本

方式は、多人数で会話しながら画面操作を行うインタフェースを構築するための指針を与えることになろう。

表7 写真操作の認識精度

|    | A  | B  | C  | D  | E  | 全体 |
|----|----|----|----|----|----|----|
| 選択 | 20 | 20 | 19 | 19 | 19 | 97 |
| 拡大 | 18 | 16 | 12 | 13 | 14 | 73 |
| 縮小 | 18 | 17 | 13 | 12 | 15 | 75 |

#### (5) TLCNに基づく対話システム

ドメインとして医療分野を選び、日常の病気に関して患者と英語で対話しながら病名診断と必要なアドバイスを与えるコンサルテーションシステムを構築した。一般的な20種類の病気に関する医学書やWebサイトの情報から基底記憶を構成し、医者の指導の下でシステムの訓練を行った後、80個の医者と患者の対話例をシミュレートした結果、表8に示す病名診断精度が得られた。この表から分かるように診断精度は、対象とする病気の種類と対話における発話数により変化する。10種類の病気の場合は、9個の発話で59%の精度を得ている。

病名を診断するための情報は対話の進行に応じて変化する。そこで、診断精度を向上させるために、フィードバック結合を持たせ、そのダイナミックな動作によって入出力の時間パターンの変換を実現しようとするリカレントニューラルネットワークを病名診断に適用した。症状、原因、検査等4種類の情報を受理する入力ユニット、7個の隠れユニット、病気の真偽を出力するユニットからなるネットワークを病名毎に構成し、80個の対話例をシミュレートした結果、平均で69%の診断精度を得た。

TLCNアーキテクチャは言語的・非言語的知識の表現と利用を統一的に扱うことを可能にしており、対話システムにおける文脈情報処理のモデルを提供することになろう。

表8 対話型病名診断システムの精度

| 病気の<br>種類 | 患者からの発話数 |     |     |     |     |
|-----------|----------|-----|-----|-----|-----|
|           | 7        | 9   | 11  | 13  | 15  |
| 10        | 56%      | 59% | 62% | 64% | 66% |
| 15        | 49%      | 55% | 58% | 60% | 63% |
| 20        | 46%      | 50% | 53% | 57% | 61% |

#### 5. 主な発表論文等

〔雑誌論文〕(計3件)

- ① Kazutaka Shimada, Daigo Hashimoto, Tsutomu Endo, A Graph-based Approach for Sentiment Sentence Extraction, *New Frontiers in Applied Data Mining*, Vol.5433, pp.38-48, 2009, 査読有.

- ② Kazutaka Shimada, Tsutomu Endo, Sayaka Minewaki, Speech Understanding Based on Keyword Extraction and Relations between Words, *Computational Intelligence*, Vol.23, NO.1, pp.45-60, 2007, 査読有.

〔学会発表〕(計19件)

- ① Md. Khalilur Rhaman, Tsutomu Endo, Recurrent Neural Network Classifier for Three Layer Conceptual Network and Performance Evaluation, The 11th International Conference on Computer and Information Technology, 2008.12.27, Bangladesh.
- ② Md. Khalilur Rhaman, Tsutomu Endo, Three Layer Conceptual Network Dialog Processor, The Twelfth IASTED International Conference on Artificial Intelligence and Soft Computing, 2008.9.2, Spain.
- ③ Kazutaka Shimada, Tsutomu Endo, Seeing several stars: a rating inference task for a document containing several evaluation criteria, The 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2008.5.23, 大阪.
- ④ 橋本大吾, 嶋田和孝, 遠藤勉, 階層非循環有向グラフを用いた文章の類似度に基づく評価文抽出, 言語処理学会第14回年次大会, 2008年3月20日, 東京.
- ⑤ Kazutaka Shimada, Kaoru Iwashita, Tsutomu Endo, A case study of comparison of several methods for corpus-based speech intention identification, The 10th Conference of the Pacific Association for Computational Linguistics, 2007.9.21, Australia.
- ⑥ 田中太喜, 嶋田和孝, 遠藤勉, マルチモーダル対話のための頭部ジェスチャ認識, 情報処理学会火の国情報シンポジウム2007, 2007年3月1日, 鹿児島.

#### 6. 研究組織

##### (1) 研究代表者

遠藤 勉 (ENDO TSUTOMU)  
九州工業大学・大学院情報工学研究院・教授

研究者番号：10112294

##### (2) 研究分担者

嶋田 和孝 (SHIMADA KAZUTAKA)  
九州工業大学・大学院情報工学研究院・助教

研究者番号：50346863