

能動的情報収集対話システムを目的とした 対話終了判断モデルの構築

小川 義人^{†1} 村上 真^{†2} 白井 克彦^{†1}

我々はユーザと能動的にコミュニケーションを取り、感想情報のような再利用する価値のある情報を言語情報として所得するシステムの実現を目指している。このようなシステムにおいては、システムがしつこく情報収集を続けてしまうと次回からの情報収集に支障をきたす恐れがあるため、ユーザが対話したくないことをシステムが判断する必要がある。そこで本研究では、音声情報を用いて対話終了を判断するためのモデル構築を行った。

Construction of Decision Model for the Spoken Dialogue System to Close Communication

YOSHITO OGAWA,^{†1} MAKOTO MURAKAMI^{†2}
and KATSUHIKO SHIRAI^{†1}

We aim at developing a robot that actively communicates with a human, and extracts information from the human mind that is rarely expressed as verbal information explicitly. With the spoken dialogue system for information collection, a human dislikes the system when the system makes a too long communication, and he may not cooperate in next communication. So, the system must independently decide whether it may or may not keep communicating with a human. In this paper, we analyze relation between a human's speech features and whether he want to close communication or not, and construct the decision model for the system to close communication.

^{†1} 早稲田大学
Waseda University

^{†2} 東洋大学
Toyo University

1. はじめに

1.1 研究背景

情報は、その情報源により自発的に発信されるものと自発的に発信されにくいものに分けることができる。自発的に発信された情報とは各種のニュースや楽曲の曲名、アーティスト名等で、自発的に発信されにくいものとは個人的な感情や印象等である。近年、情報化が進み、インターネット等で多く情報が利用されるようになってきているが、現在利用されている情報は自発的に発信されたものが多く、自発的に発信されにくい情報はあまり利用されていない。しかし、発信されにくい情報にも利用する価値がある。楽曲検索を例に挙げるならば、現在は曲名やアーティスト名が検索に利用されているが、感想情報を多数のユーザから収集、蓄積し再利用すれば、楽しくなれる曲、というような楽曲の感想で検索するという検索サービスの向上が見込まれる。このような発信されにくい情報を利用するためには、その情報を人から引き出す手段が必要である。

人から情報を引き出す手段としてはインタビュー等の情報収集対話が挙げられるが、情報を再利用するには、多数の人から情報を引き出すことが必要となる。人間が多数の人に逐一インタビューして回ることは非現実的であり、自動でインタビューする仕組みが必要である。インタビューする仕組みには、モニター上しか動けないエージェントよりも、人間の生活空間を自在に動き回れるロボットの方が適している。しかし、自発的に発信されにくい情報を引き出すためには、自発的に発信されにくいという情報の性質上、システムのほうからユーザに働きかけて情報収集対話を開始しなければならないが、従来の音声対話システムであるタスク遂行型音声対話システムは、ユーザがシステムを利用する意図を持ち、システムに働きかけて対話を開始するという対話を想定しており、システムから働きかけて開始する情報収集対話は想定していない。システムから情報収集対話を開始し、発信されにくい情報を収集するためには、タスク遂行型の音声対話システムではなく、情報収集対話に適した新しい音声対話システムを構築する必要がある。我々はこのような音声対話システムを情報収集対話型音声対話システムと定義した。

情報収集対話型の音声対話システムに必要とされる3点の事柄について、タスク遂行型と情報収集対話型の音声対話システムを比較し表1に示した。1点目は対話の目的である。タスク遂行型ではタスクを速やかに、正確に達成することを目的とし、タスク達成の速さやその正確さが重要であるのに対し、情報収集対話型の目的は、多少正確さを犠牲にしても、なるべく多くの情報を収集することである。したがって、如何にユーザから多くの情報

表 1 タスク遂行型と情報収集対話型の音声対話の比較
Table 1 a Comparison of Information Collection Dialogue System
to Task-Oriented Dialogue System

	タスク遂行型 (従来の音声対話)	情報収集対話型
対話の目的	正確に、速やかに タスクを遂行する	多くの情報を得る
対話開始の 判断	ユーザが 行う	システムが 行う
対話終了の 条件	タスクが 遂行されたら 終了	システムが 終了すべきと 判断したら終了

を引き出せるかが重要である。2点目は対話開始時に誰がその判断をするかである。タスク遂行型ではシステムを利用する意図がユーザにあり、ユーザが対話開始の判断をする。そのため、システムが対話を開始するかどうか判断する必要はない。一方、情報収集対話型ではシステムが対話開始の判断をし、このときユーザには対話を開始しようという意図があるとは限らない。よって、ユーザに負担を与えずに対話を開始することが必要となる。3点目は対話終了の判断である。タスク遂行型の目的はタスクを達成することであるため、タスクが遂行されるという明確な終了条件を持つが、情報収集対話型の目的は出来るだけ多くの情報を収集することであり、明確な終了条件を持たない。しかし、多くの情報を引き出そうとするあまりに情報収集対話を引き伸ばしすぎるとユーザに不快感を与え、次回からの情報収集対話実行に悪影響を及ぼしてしまう可能性がある、そうなることのないよう、ユーザが対話をやめたいときには、システムがそれを察知して情報収集対話を終了する必要がある。

1.2 研究の目的

1.1節で述べた情報収集対話型が満たすべき3点の事柄の内、本研究は対話終了の判断を対象とする。システムが対話を適切に終了するためには、ユーザの様子と対話を終了すべきか否かの対応を記述した統計的モデルを構築することが必要であり、本研究ではそのモデルを対話終了判断モデルと定義し、その構築を目的とする。

2. 提案手法

まず、対話終了判断モデルに有効な特徴量を考える。話者とシステムが対話しているという前提があるので、発話内容と発話から抽出できる韻律等の音声情報が利用できる。また、視線、表情、姿勢、動作などの映像情報も利用可能である。

対話を終了したいかどうかを判断するために、対話の様子から抽出した特徴量と対話を継続するか終了するかの対応を記述したモデルを構築する。モデルには混合正規分布 (Gaussian Mixture Model, 以下では GMM と呼ぶ) を使用し、対話を継続すべき状況から抽出した特徴量で対話継続モデルを構築し、対話を終了すべき状況から抽出した特徴量で対話終了モデルを構築する。識別時には識別する対話の特徴量と各モデルとの尤度比を計算し識別を行う。

モデルを構築するには、どのデータが対話を継続すべき状況で、どのデータが対話を終了すべき状況なのかをラベリングする必要がある。ラベリングの際に本来ならば話者本人による評価をもとにラベリングすべきだが、話者の心的状況が全て外に表れるわけではなく、話者の評価とデータの特徴が対応しない可能性がある。そのため本研究では話者ではなくデータを視聴した第三者の評価をもとにしたラベリングを行う。

対話終了判断モデルには言語情報、音声情報、映像情報等の様々な特徴量の利用が考えられるが、本研究では音声情報と表面的な言語情報のみを用いてモデル構築を行う。

3. データ収録

3.1 目的

対話終了判断モデルを構築するための対話データを収録するために、情報収集対話における人間の応答を収録した。

3.2 手法

想定している情報収集対話システムに近付けるため、パーソナルロボット PaPeRo[7] を用いて、WoZ 方式で対話を行った。対話形式は、実験者が PaPeRo のスピーカを通して質問し、その質問に話者が答え、その応答を受けてさらに実験者が質問をするというものである。話題は、日本の観光地、旅行歴についてで、はじめにある観光地に行った事があるかを質問し、行ったことがあればその観光地で行った場所、いつ行ったか、誰と行ったか等の質問をし、行ったことが無ければその観光地に行くとしたらどこに行きたいかを質問した。質問はあらかじめ用意した合成音声に対話内容にそって適宜選択、再生することで行った。収録は PaPeRo と被験者の間に設置したマイクで行った。

対話終了判断モデルを構築するためには、対話を継続しても良い状況の対話データと、対話を終了すべき状況の対話データが必要である。両方の状況のデータを収録するため、「どうして?」、「それでそれで?」などの質問を、実験者が話者が対話を止めたそうだと感じるまで行った。

表 2 データ収録結果
Table 2 Result of Data Collecting

話者	収録で得たデータ数
A	54
B	43
C	48
D	55
合計	200

話者は A, B, C, D の 4 名である。

3.3 結 果

収録したデータのうち応答が短すぎるものを取り除き、計 200 発話のデータを得た。ただし、話者ごとに得たデータ数は異なる。表 2 に話者毎のデータ数を示す。ここで得た 200 データを対象に、4 節に述べる主観評価によるラベル付けを行った。

4. 主観評価によるラベリング

4.1 目 的

3 節で対話を継続しても良い状態と終了すべき状態と感ぜられるデータを収録したが、話者の心的状況が全て発話に表出するとは限らず、話者が対話を終了したいと考えていても第三者にはそう感ぜられない状況があり得る。本来ならそのような状況も適切に判断すべきだが、それは人間であっても困難である。本研究では人間に判断できない状況を判断するのは諦め、人間が対話を聞いたときに下す判断の模倣を目指す。そのため、第三者が各データを聞いたときに下す判断を調べ、それを元にラベリングを行った。

4.2 主観評価手法

収録したデータを第三者に提示し、話者が対話を終了したそうかどうかを 1 から 5 の 5 段階で判定させた。終了したくなさそうなほど小さい数値を、終了したそうなるほど大きい数値を付けさせた。順序効果をなくすため、評価者毎にデータの順序を変更した。ここでの評価者は a, b, c, d, e の 5 名で、いずれも話者 4 人との面識はない。

評価後、各データに対し、評価者による評価の平均をそのデータのラベルとした。ラベルの値ごとのデータ数を図 1 に示す。

4.3 クラス分け手法

対話終了判断モデルを構築するためには、1 から 5 のラベルではなく、対話を継続するか終了するかのラベルが必要である。そこで、データのうちラベルの値が下位 30%のものに

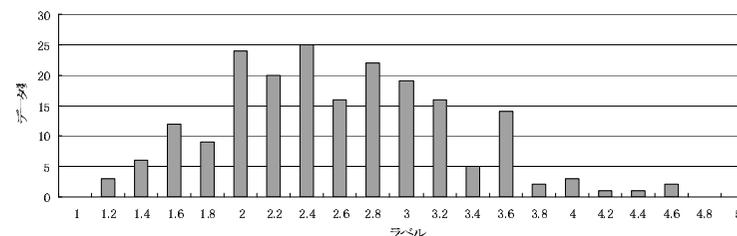


図 1 各ラベル値のデータ数
Fig. 1 Result of Estimation

表 3 クラス分け結果
Table 3 Result of Labeling

終了クラス	中間クラス	継続クラス
74	63	63

継続のラベル、上位 30%のものに終了のラベルを付けた。ただし、ラベルの値が 0.2 刻みの離散値なので実際のデータ数は全データの 30%ではない。以下では継続のラベルが付けられたデータの集合を継続クラス、終了のラベルが付けられたデータの集合を終了クラスと呼ぶ。継続クラス、終了クラスに含まれないデータは中間クラスとした。

各クラスのデータ数を表 3 に示す。

5. 予備実験

5.1 目 的

本研究では GMM を利用して対話終了判断モデルを構築することを考えているが、GMM を利用することが有効かどうかを検証するため仮の対話終了判断モデルを構築し、評価した。

5.2 GMM

本研究では GMM を利用している。GMM とは複数個の正規分布が混合された確率密度分布で、 D 次元空間での GMM の尤度はデータ \mathbf{x} 、混合数 K 、各正規分布の平均 μ_1, \dots, μ_K 、共分散行列 $\Sigma_1, \dots, \Sigma_K$ 、各正規分布の混合重み w_1, \dots, w_K に対し、式 1 で表される。

$$p(\mathbf{x}; \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K, w_1, \dots, w_K) = \sum_{k=1}^K w_k \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp\left\{-\frac{(\mathbf{x}-\mu_k)^T \Sigma_k^{-1} (\mathbf{x}-\mu_k)}{2}\right\} \quad (1)$$

本研究では EM アルゴリズムを用いて GMM のパラメータ推定を行う。EM アルゴリズムは Statistical Pattern Recognition Toolbox for Matlab[8] に含まれるものを利用した。

5.3 GMM を用いた対話終了判断モデル

本研究では GMM を利用して対話終了判断モデルを構築した。このモデルは継続クラスと終了クラスそれぞれを学習した 2 つの GMM からなる。以下では、継続クラスを学習した GMM を対話継続モデル ω_c 、終了クラスを学習した GMM を対話終了モデル ω_e と呼ぶ。このモデルは対話終了を判断すべき音声データが与えられたとき、まず、音声データからモデル構築に用いた特徴量を抽出し、特徴量ベクトル x を作る。次に x と ω_c 、 ω_e それぞれの尤度 $P(x|\omega_c)$ 、 $P(x|\omega_e)$ を計算し、 $P(x|\omega_e)/P(x|\omega_c)$ が 1 以上であれば対話終了、そうでなければ対話継続と判断する。

5.4 特徴量

本節では比較的一般的な音声特徴量を用いて GMM を構築する。特徴量を算出する際には雑音の影響を軽減するため RMS 処理をした後、波形を見て雑音レベルを定め、パワーが雑音レベル以上の範囲を有音区間、雑音レベル以下の範囲を無音区間とした。また、パワーについては周波数の影響を軽減するため RMS 処理をした後に開始部 0.1[s] 間、開始部 0.2[s] 間等の各範囲の内、有音区間の値の平均、分散を求めた。ここでのパワーの数値は-1 から 1 までで正規化された値となっている。以下にここで用いた特徴量を挙げる。

- 発話中有音区間長
応答開始から応答終了までのうち、話者が実際に発話している区間の時間長。
- 発話中無音区間割合
話者の応答の内、無音区間である部分の割合。発話中無音区間長を発話継続長で割って算出した。
- 交代潜時
質問終了から応答開始までの交代潜時。
- F0 の発話全体の平均、分散
STRAIGHT[9] を用いて F0 を算出し、有音区間の F0 の値から算出した。
- パワーの発話全体の平均、分散

表 4 正解条件

Table 4 Condition of Correct Answer

クラス	ラベルの値	正解条件
継続クラス	2.2 以下	継続と判断される
中間クラス 1	2.4	継続と判断される
中間クラス 2	2.6, 2.8	終了と判断される
終了クラス	3.0 以上	終了と判断される

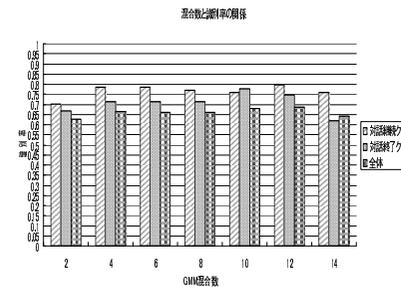


図 2 予備実験 識別率

Fig. 2 Correct Answer Probability in Preparation Experiment

5.5 識別率評価

構築したモデルを識別率を用いて評価する。継続クラス、終了クラスのデータはクラスとモデルによる判断が一致したとき正解とした。中間クラスについては、さらにラベルの値が 2.4 の中間クラス 1 と 2.6, 2.8 の中間クラス 2 に分割し、前者は継続、後者は終了と判断されたとき正解とした。クラス、ラベルの値ごとの正解条件を表 4 に示す。また、識別にはデータ数が少ないため識別には leave-one-out 法を用いた。以上の条件で混合数 2 から 14 まで 2 ずつ増加させて識別率を算出した。

5.6 結果

混合数と識別率の関係を図 2 に示す。終了クラスでは混合数が増えるにつれて識別率が上昇し、混合数 10 で最大になり、それ以降は下降している。それに対して継続クラスでは混合数 12 で最大になるものの、混合数 4 以降はあまり変化していないことが分かる。

5.7 考察

今回は 2 群判別なのでランダムならば識別率は 0.50 である。図 2 を見ると、2 から 14

まで全ての混合数で継続クラス、終了クラス、全体全てで0.50を上回っており、ここからGMMを用いた対話終了判断モデルが有効であると言える。しかし、識別率は最大でも継続クラスで0.797、終了クラスでは、0.778であり、決して高いとは言えない。これは特徴量として比較的単純な情報を用いているためと考えられ、特徴量を吟味することで識別率の改善が見込まれる。

6. 特徴量の分析, 選択

6.1 目的

実際にモデル構築に用いる特徴量を決定するために、継続クラス、終了クラスのデータから特徴量を抽出して分析した。

6.2 分析した特徴量と算出法

音声と話者の心的状況を分析している先行研究には、話者の状態やパラ言語情報を韻律と結びつけたもの、特にF0とパワーに注目したもの[1]、F0とパワー、話速に注目したもの[2]、F0と1モーラあたりの発話時間（以下ではモーラ長と表記する）に注目したもの[4-6]、F0とパワー、モーラ長に注目しているもの[3]があるので、これらを参考に以下の28特徴量を抽出し分析した。ここでも第5節と同様にパワーはRMS処理を施した後に算出している。本研究ではこれらの特徴量毎にヒストグラムを作成し、分布を分析した。

- 交代潜時
第5節と同様。
- 発話継続長
応答開始から応答終了までの時間長
- 発話中無音区間長
話者の応答の内、無音区間である部分の時間長。
- 発話中無音区間割合
第5節と同様。
- 発話モーラ数
発話量を表す特徴量。各音声データを人間の耳で聞いてテキストに起こし、発話しているモーラ数を調べた。発話中有音区間長とほぼ同じと考えられるので、発話モーラ数を分析項目に入れる際に発話中有音区間長を分析項目から外した。
- スピーチレート
1[s]間に何モーラ発話しているかを表す。発話モーラ数を発話中の有音区間長で割るこ

とで算出した。

- 発話に含まれるフィルターの個数
発話中に幾つのフィルターが含まれているかを表す。各音声データを人間の耳で聞いて、含まれているフィルター数を調べた。
- 発話1[s]あたりのフィルターの個数
発話1[s]間に幾つのフィルターが含まれているかを表す。フィルターの個数を発話中の有音区間長で割ることで算出した。
- F0の発話全体の平均
第5節と同様。
- F0の発話全体の分散
第5節と同様。
- パワーの発話全体の平均、分散
- パワーの発話開始部、終了部それぞれ0.1, 0.2, 0.4, 0.8[s]間の平均、分散

6.3 結果

継続クラスと終了クラスの分布に差があった特徴量には図3に示すようなものがあった。継続クラスと終了クラスの分布に差が無かった特徴量には図6に示すようなものがあった。継続クラス、終了クラスで分布に特に差があった特徴量を以下に示す。

- 交代潜時
 - 発話中無音区間割合
 - 発話モーラ数
 - スピーチレート
 - 発話1[s]あたりのフィルター個数
 - F0の発話全体の平均
 - パワーの発話全体の平均
 - パワーの発話開始0.4[s]間の平均
 - パワーの発話開始0.8[s]間の平均
 - パワーの発話全体の分散
 - パワーの発話開始0.4[s]間の分散
 - パワーの発話開始0.8[s]間の分散
- 今回、継続クラス、終了クラスの分布に差がある特徴量の中には、正規分布をするものと別の分布をするものがあった。正規分布をする例としては図4に示したスピーチレートが、

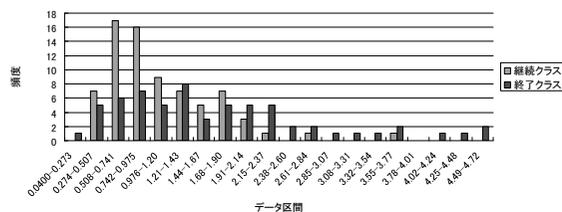


図 3 交代潜時のヒストグラム
Fig. 3 Histogram of Time by Turn Taking

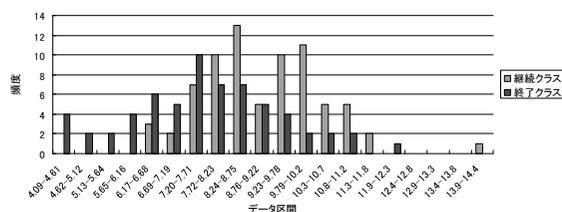


図 4 スピーチレートのヒストグラム
Fig. 4 Histogram of Speech Rate

別の分布となった例としては図 5 に示したパワーの発話開始部 0.8[s] 間の分散がある。

これら 10 特徴量を物理的な意味でまとめると、話速、発話量、無音区間（交代潜時、発話中無音区間割合、発話 1[s] あたりのフィルター個数も間を埋めるものという意味で間に分類）、及び全体、発話開始部の音圧で継続クラスと終了クラスが異なるという結果だった。ただし、発話開始部については 0.1[s], 0.2[s] に関する特徴量では分布に特に差が無かったことから、ある程度の時間長を考慮する必要があると考えられる。一般に発話権の委譲を示すような特徴は発話終了付近のパワーパターンに現れると言われているが、対話そのものの終了したさを示す特徴は発話終了部には現れず、発話開始時等に現れることが分かった。

6.4 特徴量の選択

6.4.1 目的

計 28 特徴量を抽出したが、データ数が少なく、それに比べて 28 次元のモデルでは次元数が大きすぎる。そこで、継続クラス、終了クラスの分布に比較的差がある特徴量を選択し、選択された特徴量を用いてモデルを構築する。6.3 節では研究者の主観によって継続ク

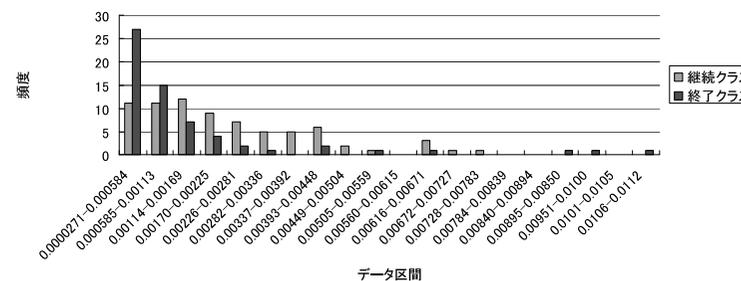


図 5 パワーの発話開始部 0.8[s] 間の分散のヒストグラム
Fig. 5 Histogram of Power Variance between Beginning 0.8[s]

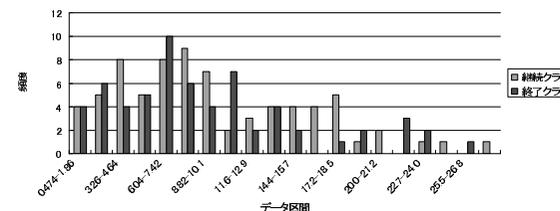


図 6 発話継続長のヒストグラム
Fig. 6 Histogram of Speech Length

ラス、終了クラスの分布の差を分析したが、ここでは客観的な指標によって特徴量を選択することを考える。

6.4.2 手法

各クラスごとに特徴量が正規分布すると仮定し、その過程の元でクラスの分布に差があるほど増加する関数 F を定義し、 F の値によって用いる特徴量を決定する。関数 F を式 2 に示す。ここで μ_1, μ_2 はの継続クラス、終了クラス毎の特徴量の平均、 Σ_1, Σ_2 は継続クラス、終了クラス毎の特徴量の分散である。

$$F(\mu_1, \mu_2, \Sigma_1, \Sigma_2) = \frac{(\mu_1 - \mu_2)^2}{\sqrt{\Sigma_1 \times \Sigma_2}} \quad (2)$$

特徴量毎の関数 F の値を関数 F の値順に有効数字 3 桁で表 5 に示す。表 5 を見ると、関数 F の値でも主観による分析と同じ特徴量が上位に位置することが分かる。そこで、本研

究では関数 F の値の上位 12 特徴量, 即ち主観による分析で継続クラス, 終了クラスの分布に差があったとした特徴量から特徴が重複すると考えられるものを除いた 10 特徴量を用いてモデルを構築することとする. この 10 特徴量を以下に示す.

- 交代潜時
- 発話中無音区間割合
- 発話モーラ数
- スピーチレート
- 発話 1[s] あたりのフィルター個数
- F0 の発話全体の平均
- パワーの発話全体, 発話開始 0.8[s] 間の平均, 分散

これらの特徴量の中には図 5 のように正規分布とは違った分布をする特徴量もあるが, 今回は全ての特徴量が正規分布もしくは混合正規分布をすることでこれら 10 特徴量で 10 次元のモデルを構築する.

7. 対話終了判断モデル構築

第 5 節で GMM を利用した対話終了判断モデルの有効性を確認した. この節では第 6 節で分析, 選択した特徴量を用いて対話終了判断モデルを構築し, その評価を行う.

7.1 対話終了判断モデル

モデルの仕組みは第 5 節の予備実験で構築したものと同一で, 対話継続モデル ω_c と対話終了モデル ω_e からなり, 音声データからモデル構築に用いた特徴量を抽出し, 特徴量ベクトル x を作り, x と ω_c, ω_e それぞれの尤度 $P(x|\omega_c), P(x|\omega_e)$ を計算し, $P(x|\omega_e)/P(x|\omega_c)$ が 1 以上であれば対話終了, そうでなければ対話継続と判断する. 第 5 節では比較的単純な音声特徴量を用いているが, この節では第 6 節で継続クラスと終了クラスで差が見られた 10 特徴量を用いてモデルを構築する.

7.2 識別率評価

構築したモデルを識別率を用いて評価する. 正解条件についても第 5 節 (表??) と同様である. ここでも識別にはデータ数が少ないため識別には leave-one-out 法を用いた. 以上の条件で混合数 1, 2, 4, 8 のときの識別率を算出した.

混合数と識別率の関係を図 7 に示す. 全てのクラスで混合数 2 のときに識別率が最大となった. また, そのときの識別率は継続クラスが 0.878, 終了クラスが 0.810 と比較的高いが, 中間クラス 1 では 0.640, 中間クラス 2 では 0.553 とランダムな判断とあまり差が無

表 5 関数 F の値
 Table 5 Values of Function F

特徴量	関数 F の値
スピーチレート	0.850
モーラ数	0.694
パワーの発話開始 0.8[s] 間の平均	0.684
無音区間	0.606
発話中無音区間割合	0.393
パワーの発話全体の平均	0.326
パワーの発話開始部 0.4[s] 間の平均	0.233
発話 1[s] あたりのフィルターの個数	0.182
パワーの発話全体の分散	0.180
パワーの発話開始部 0.8[s] 間の分散	0.168
F0 の発話全体の平均	0.140
パワーの発話開始部 0.4[s] 間の分散	0.111
パワーの発話開始部 0.1[s] 間の分散	0.0981
パワーの発話開始部 0.2[s] 間の平均	0.0917
パワーの発話開始部 0.1[s] 間の平均	0.0856
発話中無音区間長	0.0850
発話に含まれるフィルターの個数	0.0506
F0 の発話全体の分散	0.0353
パワーの発話開始部 0.2[s] 間の分散	0.0346
パワーの発話終了部 0.1[s] 間の平均	0.0192
パワーの発話終了部 0.4[s] 間の分散	0.00788
パワーの発話終了部 0.1[s] 間の分散	0.00666
パワーの発話終了部 0.8[s] 間の分散	0.00609
発話継続長	0.00530
パワーの発話終了部 0.4[s] 間の平均	0.00110
パワーの発話終了部 0.2[s] 間の分散	0.000745
パワーの発話終了部 0.8[s] 間の平均	0.000384
パワーの発話終了部 0.2[s] 間の平均	0.000289

かった.

8. 考 察

全てのクラスで識別率が最大になることから, このモデルには混合数 2 が最も適していることが分かる. 混合数 2 のときに識別率が最大となるのは, 特徴量選択の際に各特長量ごとに正規分布すると仮定して選択することで, 比較的単純なモデルとなる特徴量が選択されたためだと考えられる. 中間クラスの識別率があまり高くないのは, 中間クラスのデータがはっきりと継続や終了と言えないようなデータであるのに加えて, 継続クラスと終了クラス

を識別出来る特徴量を用いたためと考えられる。

また、今回対話継続クラスでありながら、対話終了クラスであると識別されたデータには、Yes/No クエスチョンや5W の質問など簡潔に答えられる質問に対してに簡潔に答えているものがあつた。このような認識間違いは質問の種類によってモデルを複数構築し、それらを使い分けることで改善できる可能性がある。

逆に、今回対話終了クラスでありながら、対話継続クラスであると識別されたデータには以下のようなものがあつた。

- 発話量が多いが質問内容に関する情報が少ないもの
- 質問に答えず、応答の中で自分で話題を変えているもの

いずれも発話内容に関するものであるため、自然言語処理を行い発話内容の意味を理解するモデルを構築することで改善できる可能性がある。

さらに、今回は全ての特徴量に対してGMM を利用したが、ヒストグラムを見るとパワーの発話開始部0.8[s] 間の分散のように混合正規分布ではない特徴量もあつた。このような特徴量に対しては特徴量に合った分布をモデルに組み込むことが必要である。

9. ま と め

今回、情報収集対話システムを目的とした対話終了判断モデルを構築するため、WoZ 方式でデータを収録した。そのデータに対して主観評価によって話者が対話を終了したそうかのラベリングを行い、振られたラベルに基づいてデータを分割し、継続クラスと終了クラスから特徴量を抽出して対話終了判断に有効な特徴量を分析した。その結果、話速、発話量、

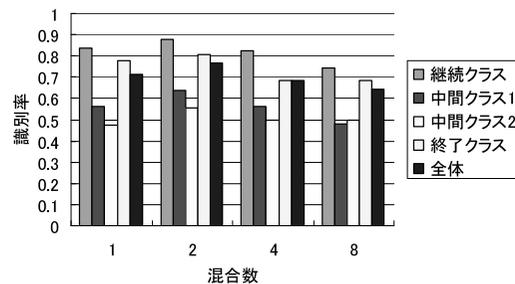


図 7 混合数と識別率の関係

Fig. 7 Relation between Number of Mixture and Correct Answer Rate

無音区間、音圧が有効との結果を得た。さらに継続クラスと終了クラスのデータから対話終了判断モデルを構築して評価した。その結果、混合数2のときに全体は0.765という識別率を得た。全体の識別率は高くはないが、音声情報だけでもある程度の推定は行うことが出来た。また、継続クラス、終了クラスについては0.8以上の識別率を得た。

9.1 今後の課題

さらに識別率を改善するためには、ToBI等を用いて一発話内での特徴量の変動を分析すること、発話内容の意味や映像情報等、今回用いなかった情報をモデルに取り入れるほか、質問の種類に応じたモデルを選択するなどして、1発話での推定をより正確にすること、さらに、人間はその瞬間瞬間だけではなく、それまでの対話履歴からも話者の状態を判断すると思われるので、各特徴量が会話開始時からどの程度変化しているかを特徴量として用いるモデル化や過去数発話をまとめて扱うようなモデル化をすることが有効だと考えられる。

参 考 文 献

- 1) 渋谷 貴紀, 川端 豪: "音声対話システムのための親近感特徴量の探索", 電子情報通信学会技術研究報告, vol.106, no.122, pp.25-30 (Jun. 2006) .
- 2) 中里 央, 澤井 夏美, 樽松 理樹, 羽倉 淳, 藤田ハミド: "感情表現における音声発話特徴の抽出", 日本音響学会聴覚研究会資料, Vol. 37, No.8, H-2007-119.
- 3) 西田 昌史, 小川 純平, 堀内 靖雄, 市川 熹: "対話音声を対象とした韻律情報による発話印象のモデル化", 情報処理学会 研究報告 2005-SLP-59- (14), Vol.2005, No.127, pp.79-84 (Dec. 2005) .
- 4) 藤江 真也, 江尻 康, 菊池 英明, 小林 哲則: "肯定的/否定的発話態度の認識とその音声対話システムへの応用", 電子情報通信学会論文誌, vol.J-88-D-II, no.3, pp.489-498 (Mar. 2005) .
- 5) 前川 喜久雄: "韻律によるパラ言語情報の表出に関する準備的考察", 人工知能学会 言語・音声理解と対話処理研究会資料 Vol.15, pp.31-36, (Oct. 1996) .
- 6) 八木 大三, 藤江 真也, 菊池 英明, 小林 哲則: "韻律情報を用いた肯定的/否定的態度の認識", 日本音響学会 講演論文集, pp.141-142, (Mar. 2004) .
- 7) NEC Personal Robot Research Center
<http://www.incx.nec.co.jp/robot/>
- 8) Statistical Pattern Recognition Toolbox for Matlab
<http://cmp.felk.cvut.cz/cmp/software/stprtool/>
- 9) STRAIGHT
<http://www.wakayama-u.ac.jp/kawahara/STRAIGHTadv/index.j.html/>