韻律情報を利用した対話状態・心的状態の推定
Estimation of Dialogue State and Mental State Using Prosodic Information

早稲田大学理工学部

School of Science and Engineering, Waseda University

白井 克彦

SHIRAI Katsuhiko

＜研究協力者＞

早稲田大学人間科学部

School of Human Sciences, Waseda University

早稲田大学大学院理工学研究科

Graduate School of Science and Engineering, Waseda University

菊池　英明

KIKUCHI Hideaki

大久保崇

OKUBO Takashi

小林 季実子

KOBAYASHI Kimiko

This paper describes the some trials of estimation of dialogue state and mental state using prosodic information for controlling dialogue in spoken dialogues. At first, we analyzed relationship between topic boundaries and prosodic information in spoken dialogues. We introduced twenty types of prosodic parameters in initial and final accentual phrases of utterances, and investigated correlation between those parameters and topic boundaries. As the result, it was confirmed that some prosodic features correlated with topic boundaries strongly. Next, we tried to construct models for discrimination of misunderstanding, politeness, and emotion. It turned out that introducing appropriate prosodic parameters especially related with prosodic phrase makes these models practical. After that, we have developed the platform for generalization of system architecture to make maintenance ease. This paper particularly shows the architecture of the platform and an example of process of outputting backchannels in a dialogue.

Key Words: prosody, mental state, topic boundary, misunderstanding, politeness, emotion.

1. 研究の目的

対話音声における韻律には、同音異義語の区別や文体の区別、文構造の明示、強調や感情伝達の他に、心理状態の表現、発話権制御など対話特有の様々な役割がある[1]。対話における韻律の役割を解明するあるいは対話の構造を韻律によって説明しようとする試みがこれまでにもなされている[2][3][4][5]。音声対話システムにおいて自然な対話制御を実現するためには、そうした対話における韻律の様々な役割をモデル化して利用することが重要である。本研究では、特に対話の状態とユーザの心的な状態を推定して対話制御に利用する試みを行った。

対話の状態としては、話題境界に焦点を当て、その自動検出を試みた。発話や対話全体の理解には話題の境界を正確に検出することが重要である。話題境界の検出に韻律情報の利用が有効であることは既に報告されているが、ここでは対話音声を対象として、ハンドラベリング結果を利用することによってより多くの韻律的特徴を導入することの有効性を評価した。詳細を2章に述べる。

人間同士の対話では、対話の進行とともに対話参加者の感情や心理状態が様々に変化し、多かれ少なかれ互いに相手の状態を認識しながら対話を行っている。音声対話システムが状況に適した行為を振舞うためには、発話内容や相手の声の調子からユーザの感情や心理状態を推測することが必要となる。本研究ではユーザの心的状態として、システム側の音声誤認識に起因する誤解状態、発話の丁寧さ、感情を扱い、それぞれの自動判別を試みた。詳細をそれぞれ3章、4章、5章に述べる。

こうした状態推定のモデルを対話制御に利用す

るために、音声対話システム汎用プラットフォーム上に韻律情報を用いた対話制御を実現した。詳細を6章に述べる。

2. 話題境界の検出

話題境界の検出に韻律情報の利用が有効であることは既に報告されている。[6]では、対話における話題の切れ目の深さについて、基本周波数やパワーとの間に強い相関が見られることが報告されている。また、[7]では、より詳細な韻律情報と独話における話題の切れ目の深さとの関係を分析し、話題の切れ目の前後の発話について、発話間のポーズ、発話末と発話開始位置のアクセントにおけるピッチレンジリセットの程度などが関係を持つことが報告されている。そこで我々は、対話における話題の切れ目の深さと韻律情報の関係をより詳細に分析した。特に基本周波数やパワーに関しては、発話における位置や、話題の切れ目と韻律の関係を考慮して、決定木を用いた話題の切れ目の判別を試みた。

2.1 音声対話における話題境界と韻律の関係

まず対話における話題の切れ目の深さと韻律情報の関係を分析した結果を述べる。

2.1.1 対話データ

分析には、人工知能学会「談話・対話研究におけるコーパス利用研究グループ」によって作成されたコーパス[8]を使用した。本研究では、コーパスに付与された「談話セグメントタグ」(Topic Break Index: 以下 TBI)[9]に基づいて対話中の話題の切れ目を決定する。TBI は1,2の二段階で話題の切れ目の深さを表し、話題の変化が大きい場合に2を、そうでない場合は1をタグ付け作業者の主観で評価し付与する。また、話題が連続している発話には TBI は付与されておらず、本稿ではこのような発話を TBI が0の発話として定義する。

分析には、TBI が付与された5つの対話（タスクはクロスワードパズル、旅館予約、会議室予約、地図課題）を用いる。なお、話題の遷移に関与しないと思われる相槌、フィラー、言い淀みのみで構成される発話を分析対象から除外した。表1に分析対象となる発話の数と TBI の内訳を示す。

表1. 分析に用いるデータの TBI 毎の発話数

TBI	0	1	2	計
発話数	72	90	38	200

2.1.2 分析方法

発話の開始部分と終了部分に注目するために、分析対象となる発話に対して、日本語話し言葉音声の韻律ラベリングスキーム X-JToBI[11]を用いて韻律ラベルを付与し，そのラベル情報をもとに開始アクセント句と終了アクセント句の区間を抽出した。発話速度を除いた韻律情報の各パラメータについて，話者の違いによる絶対的な差を考慮し，話者ごとに標準化した値を使用した。扱うパラメータは以下の通りである。

- 基本周波数の最大(max)・最小(min)・レンジ(range)・平均(ave)
- パワーの最大(max)・平均(ave)
- 発話速度(speed)

2.1.2 分析結果

分析の結果，話題の切れ目の深さといくつかの韻律パラメータとの間に相関が見られることがわかった。まず，発話の終了部分に関しては，パワーの最大や基本周波数のレンジとの間に正の相関が観察された。発話の開始部分に関しては，談話標識を除外すれば終了部分と同様の相関関係が見られた。また，発話の開始部分と先行発話の終了部分との差分について観察したところ，話題の切れ目が深くなるほど，後続発話の基本周波数やパワーが大きくなる傾向が見られた。

2.2 決定木による話題の切れ目の深さの判別

次に、決定木を用いた話題の切れ目の深さの判別についての予備的な実験結果を示す。判別対象は2.1.1に示した対話データの200発話とし，発話毎に TBI(0/1/2)を判別する。決定木の学習には C4.5アルゴリズム[16]を利用する。なお，冗長な木ができるのを防ぐために，枝刈りを信頼度1%で行う。決定木学習には，前節の分析に用いた基本周波数，パワー，発話速度の各種韻律パラメータを用いる。 また，この他に，前節の分析で大きな影響が観察された発話間のポーズ長、話者交代の有無、談話標識かどうかを判別のパラメータとして導入する。

オープンな条件とクローズトな条件で学習した決定木によるそれぞれの判別精度を表2に示す。人手による TBI の付与実験を行った結果[6]においても κ の平均値は0.385とそれほど高くないが、それにも劣る結果が得られた。この原因として学習データ量の不足が第一に考えられるが、パラメータについてもさらに検討する必要がある。

表2. 決定木による判別精度

	判別率	κ 値
クローズト	81.0%	0.70
オープン	56.6%	0.30

3. 誤解状態の検出

音声対話システムと人間のコミュニケーションにおいて、現状では音声の誤認識・誤理解は避ける事ができない大きな問題である。円滑なコミュニケーションを実現するために、システムが自分自身の誤解を認識し解消したうえで、誤解が引き起こしたユーザの状態の変化を推測することで、状況にふさわしい対応をとることが重要である。本研究では、まずシステムの誤解状態を意図的に発生させた対話を収録し、対話中でのシステムの確認発話に対するユーザ応答分析を行い、ユーザの感情や心理状態の認識に役立つ特徴が得られるかを検証した。以下には言語情報・パラ言語情報の分析結果を報告する。

3.1 対話データ

誤解発生状況として、システムからの確認発話時に限定し、Wizard of OZ 方式で対話収録実験を行った。被験者とは別室にいる操作者が、確認方法（直接・間接）と誤解の有無を制御する。対話のタスクは「チケット予約」とし、システム側が主導権をもってユーザの発話を促す形で対話を進める。被験者にはあらかじめチケット予約一覧を渡し、コンピュータと対話して一覧に記載されているチケットを予約するよう指示した。被験者一人あたり20〜30対話、学生4名、他4名の計8名の被験者により、合計196対話（約211分）を収集、確認発話に関するシステムとユーザのやりとり1310個を収集した。図１に収録した対話の例を示す。

```
S:いらっしゃいませ、チケットを予約します。
S:ご希望の公演名をお願いします。
U:吉田拓郎をお願いします。
S:吉田拓郎ですね。
U:はい、そうです。[Explicit:Right:Yes]
S:ご希望の公演会場はどこですか？
U:大宮ソニックシティです。
S:東京国際フォーラムホール A ですね。
U:いいえ、大宮ソニックシティです。
[Explicit:Miss:Correct]
S:大宮ソニックシティ大ホールですね。
U:はい、そうです。[Explicit2:Right:Yes]
S:ご希望公演日をお願いします。
U:8月10日です。
S:8月9日ですね。
U:いいえ、違います。[Explicit:Miss:No]
S:ご希望公演日をお願いします。
…
(以下省略)
```

図1. 収録対話例
（S はシステム、U は被験者の発話を示す。
[]内はそれぞれの発話の分類を示すタグ。）

3.2 言語情報の分析

表3に確認発話に対する被験者応答の分類とその割合を示す。システムの確認発話に対する被験者応答の発話内容を分析した結果、システムの誤った確認に対しての応答にいくつかの傾向が見られた。1つは、否定のみの発話より訂正発話が多く行われていることである。特に今回の実験では、被験者がシステムの誤解を指摘すると、システムは同じ質問を繰り返した。そのため、被験者は自然にシステムの次の発話を予測して訂正発話を行い、繰り返される質問を避けて効率を上げようと試みているのではないかとも考えられる。

また、図2に示すように対話回数の増加に伴う被験者の対応は様々であり、慣れていくスピードにも差が見られたが、最終的にはほとんどの被験者がシステムの動作や発話に慣れ、効率のよい対応を取るようになった。この結果からも、システムの振る舞いに対するユーザの慣れとともに変化する応答方法からユーザの状態を的確に推測できれば、より円滑に対話が進められるのではないかと考えられる。

表3. 確認発話に対する被験者応答の分類

システム確認の分類		総合	総数（割合[%]）
直接確認	正解	肯定のみ	234(100.00)
	誤解	否定のみ	127(56.19)
		否定訂正	7(3.10)
		訂正のみ	90(39.82)
		その他	2(0.89)
間接確認	正解	次の値	233(96.28)
		その他	9(3.72)
	誤解	否定のみ	45(18.22)
		否定訂正	14(5.67)
		訂正のみ	79(31.98)
		訂正+次	10(4.05)
		次の値	99(40.08)

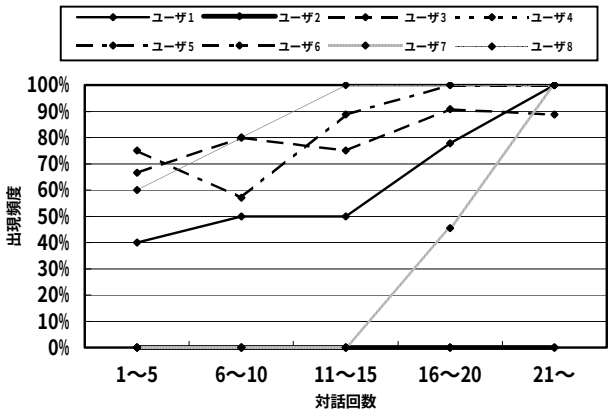

図2. 間接確認における訂正発話の出現頻度の推移

3.3 パラ言語情報の分析

システムの確認発話に対する被験者応答のパラ言語情報を分析した結果、確認の種類やその正誤による変化の他、対話回数、対話進行状況による各特徴量の変化が見られた。典型的には、対話回数の増加に伴いシステムの振る舞いに対する慣れが増すためか応答が早くなる（図3に一例を示す）他、システムの初めての誤解に対しては戸惑いのためかパワーやピッチの増加がより大きくなる（図4に一例を示す）、などの傾向が見られた。こうした慣れや戸惑いなどのユーザの状態は、話者による量的な違いはあってもパラ言語情報として表出されており、システムがこれを抽出して利用するに値するといえる。

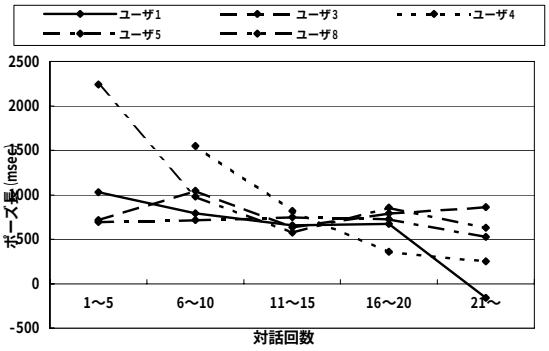

図3. 誤った間接確認に対する訂正発話のポーズ長

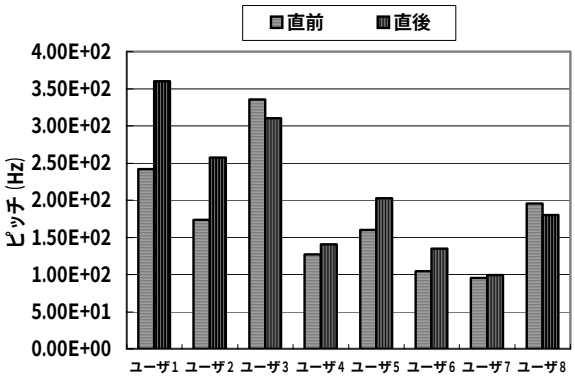

図4. システムの最初の誤った確認に対するユーザ応答のピッチ（最大値）の変化

3.4 まとめ

本研究では、システムの誤解状態を意図的に発生させた対話において、被験者の応答の種類、応答の早さ、応答発話のパワー・ピッチなどについて分析した結果について述べた。応答の種類としては、対話回数の増加に伴い訂正発話が増えていく傾向が見られた。応答の早さについて、全般的に誤解時に遅れが生じる傾向があるが、特に初めての誤解の際にその傾向が顕著にあらわれることがわかった。また応答発話のパワーおよびピッチについても、初めての誤解の際に最大値が通常より大きくなる傾向が見られた。

4. 丁寧さの推定

人間同士の自然な対話において、相手や状況に応じて発話スタイルは柔軟に変化する。音声対話システムにおいても話者の発話スタイルの変化を検出し、それを利用することが自然な対話制御の実現に有効と考えられる。発話スタイルの定義は様々に考えられるが、ここでは発話の音声的な丁寧さに注目し、丁寧かぞんざいかの二値で捉えるよう単純化して扱う。以下では、人間同士の音声対話データを対象として発話単位の丁寧さ評価を行なって丁寧さのタグを与え、韻律パラメータとの関係を判別分析により調べる。

4.1 分析方法

分析には、早稲田大学白井研究室旅行プランニング対話[10]を用いた。この対話は親しい学生同士による旅行計画をタスクとしたものであり、自発性が極めて高い対話である。このうち任意に抽出した男性2名による268発話を分析の対象とした。分析に先立ち、各発話について丁寧さの聴取評価を行った。対話者とは異なる3名の被験者が、発話毎に丁寧さを五段階で評価し、2名以上が4以上の評価を与えたものを丁寧な発話、同じく2名以上が2以下の評価を与えたものをぞんざいな発話と認定した。丁寧さの認定結果の内訳を表4に示す。

判別分析の説明変数としては、基本周波数・パワー値の最大値・最小値・範囲（最大値と最小値の差分）・平均値・分散、発話速度などのパラメータを発話単位で算出して用いた。なお、発話中の韻律的な単位をさらに詳細にして分析するために、日本語音声韻律ラベリングスキーム X-JToBI[11]によって韻律ラベリングを行い、アクセント句を認定した。その結果を用いてアクセント句単位でも上記パラメータを算出して分析に用いた。

4.2 分析結果

まず始めに前述の各パラメータを説明変数とした時の「丁寧」「ぞんざい」「どちらでもない」の2群（真偽）および3群の正判別率を求めた結果を表5に示す。この結果から、発話の丁寧さの判別にはF0の最大値、パワーの最大・最小値が有効であるといえるが、いずれも判別精度は低い。

次に、発話先頭のアクセント句単位で同様の判別分析を行った結果を表6に示す。この表では、発話速度による判別精度が若干良い。

表 4. 発話の丁寧さを認定した結果の内訳[発話数]

話者	丁寧	ぞんざい	どちらでもない	合計
A	3	31	74	108
B	17	40	103	160
合計	20	71	177	268

表 5. 発話単位での正判別率[%]

説明変数		丁寧	ぞんざい	どちらでもない	全体
全パラメータ		45.0	46.5	22.6	30.6
F0	最大	42.9	7.1	71.9	52.8
	最小	66.7	30.0	10.1	19.7
	範囲	47.6	12.9	67.4	51.7
	平均	5.0	57.7	41.2	42.9
	分散	100.0	2.8	0.0	9.3
Pwr	最大	14.3	34.3	68.0	55.0
	最小	0.0	27.1	72.5	55.0
	範囲	9.5	40.0	69.1	56.9
	平均	45.0	66.2	1.7	22.0
	分散	35.0	1.4	75.1	52.6
発話速度		55.0	56.3	9.6	25.4

表 6. 発話先頭アクセント句単位での正判別率[%]

説明変数		丁寧	ぞんざい	どちらでもない	全体
全パラメータ		25.0	35.3	64.2	53.7
F0	最大	37.5	17.6	76.5	57.7
	最小	87.5	29.4	17.3	25.2
	範囲	50.0	35.5	69.1	58.5
	平均	25.0	50.0	65.4	58.5
	分散	12.5	52.9	64.2	57.7
Pwr	最大	50.0	44.4	11.0	22.5
	最小	71.4	40.7	8.1	21.1
	範囲	0.0	61.1	40.4	43.1
	平均	0.0	35.2	68.4	54.9
	分散	0.0	35.2	63.2	51.5
発話速度		65.7	76.5	54.4	60.8

表 7. 発話末アクセント句単位での正判別率[%]

説明変数		丁寧	ぞんざい	どちらでもない	全体
全パラメータ		47.3	50.1	53.9	50.9
F0	最大	55.6	63.3	68.4	65.8
	最小	62.4	59.7	65.5	63.6
	範囲	73.4	77.6	68.9	70.4
	平均	43.3	42.9	54.3	48.8
	分散	45.4	43.2	52.3	46.5
Pwr	最大	50.0	49.9	53.2	51.1
	最小	43.2	53.3	60.1	57.4
	範囲	49.9	67.8	72.2	69.8
	平均	52.6	86.4	65.4	73.4
	分散	51.2	44.3	59.7	56.4
発話速度		27.8	50.5	48.7	49.2

次に、発話末のアクセント句単位で同様の判別分析を行った結果を表7に示す。この表では、F0の範囲やパワーの平均値による判別精度が良い。

以上の結果から、発話の丁寧さについて「丁寧」「ぞんざい」「どちらでもない」の3群判別を行うには発話末のアクセント句に注目して F0とパワーの範囲を利用するのが有効といえる。また、丁寧かどうかの2群判別には発話単位での F0の分散値による結果が最良となっているが、「丁寧」と認定された発話自体が20発話のみであるため、この結果の信頼性は低い.

5. 感情の判別

音声対話では話し手の感情が話し方や音色に変化を与え、聞き手はそこから話し手の感情を感じ取り、適切に対応していると考えられる。音声対話システムの実現において、話者の感情を判別することは音声対話の自然さを向上するうえで非常に重要である。本研究では、そのような音声対話システムの実現を目指し、話者の感情と韻律情報の関係を明らかにする。本稿では特に感情表現能力に長けた落語家に注目し、どのような韻律情報が感情の判別に有効かを判別分析により調べる。なお、落語家音声の韻律的特徴については武田ら[12]により報告されている。本稿では音声対話システムでの応用に向けて感情判別のアルゴリズムを検討するための韻律的特徴の予備的な分析結果を報告する。

5.1 分析方法

まず始めに落語音声を被験者に聴取させ発話毎に該当する感情をあらかじめ用意した6感情より選択させた。聴取の対象としたデータは落語音声の中でも比較的有名な演目である「船徳」を3人の噺家が演じたものであり、そのうち演技がほぼ共通していると見られる部分を抜粋して使用した。被験者は男性6名、女性2名の計8名であり、用意した6感情は基本6感情のうちデータに現れにくい「悲しみ」「嫌悪」を除き、変わりに「平静」「あきれ」を加えたものとした。以降の分析では半数以上の被験者が同じ感情を選択した発話のみを対象とした。表8に分析対象データの内訳を示す。また、被験者により選択された感情の内訳を表9に示す。

判別分析には線形判別関数とマハラノビスの汎距離を用い、説明変数としては4章と同じものを用いた。

5.2 分析結果

まず、全てのパラメータを用いた判別分析の結果に基づき、寄与の少ないパラメータを除いた。さらに判別精度を向上するために、6感情をグループ分けした。具体的には感情表現が豊かになりやすい

「喜び」「驚き」「怒り」と、「平静」「恐れ」「あきれ」の2グループに分け、判別の第一段階でこれらのグループ判別を行うこととした。

線形判別関数とマハラノビスの汎距離による判別の的中率をそれぞれ表10、11に示す。これらの結果から、「喜び」「驚き」はどの噺家においても高い精度で判別可能であり、それ以外についてもある程度の判別精度が得られているがその傾向は噺家によって異なることがわかる。なお、判別に寄与するパラメータは噺家毎に大きく異なっていた。感情判別をシステムで利用するためには感情表現方法の話者類型化や自己学習などの方策が必要であろう。

表 8. 分析対象とした落語音声データ[sec (発話)]

噺家	総演技	共通箇所	分析対象
A	1287.0	309.6 (116)	199.2 (80)
B	1593.0	185.4 (85)	137.4 (63)
C	1227.0	682.2 (142)	261.0 (112)

表 9. 被験者により選択された感情の内訳[発話]

噺家	平静	喜び	驚き	怒り	恐れ	あきれ
A	36	5	6	19	7	7
B	35	3	2	13	8	2
C	56	5	6	25	7	13

表 10. 線形判別関数による判別的中率[%]

噺家	平静	喜び	驚き	怒り	恐れ	あきれ
A	90.0	80.0	99.3	72.0	86.0	86.0
B	88.9	94.4	83.3	88.9	88.9	97.8
C	97.2	91.7	86.1	81.6	94.7	86.8

表 11. マハラノビス汎距離による判別的中率[%]

噺家	平静	喜び	驚き	怒り	恐れ	あきれ
A	83.3	73.3	83.3	62.0	58.0	74.0
B	83.3	94.4	88.9	80.0	66.7	100.0
C	100.0	94.4	86.1	78.0	86.8	77.6

6. 音声対話システム汎用プラットフォーム

これまで述べた対話状態や心的状態の推定とその結果を利用した対話制御を実現することを目的として、音声対話システム汎用プラットフォーム上で韻律情報を用いた対話制御の実現を試みた。具体的には、あらかじめ用意した心的状態モデルの更新に発話時間や発話終了後の無音時間、発話速度の変化、基本周波数パタンなどの要素を用いることを可能にし、それを参照するように対話制御ルールを記述する。システム設計者は実験やアプリケーションの目的に応じて心的状態更新の要素を調節するとともに、システムの振る舞いをルールによって記述することができる。

6.1 音声対話システム汎用プラットフォーム

本プラットフォームは音声対話による問題解決を行う上で必要な知識について、記述多様性と記述容易性のトレードオフを考慮した形で蓄え、知識の内容に依存しない形で利用できるように実装されている。具体的には、システムの実際の行動レベルの記述の上に一段階抽象化した計画レベルを用意する手法を提案した[13]。上位レベルとなる計画レベルでは、システムの問題解決戦略を記述するプラニングルールをプロダクションルールで表現する。また、下位レベルとなる行動レベルの記述、すなわち実際に行う発話や画面表示、アプリケーションコマンドなどを決定する行動決定ルールを有限状態オートマトンによって表現する。各ルールを最大限にパッケージ化しておくことによって、問題解決およびその手段としての対話の多様性と対話記述の容易性を実現する。また、システムを構成するモジュールを並列に動作させて多様な対話制御方法に対応し、プラニングや行為決定の要素となる心的状態を多次元にする[14]ことによってユーザとシステムの多様なコミュニケーション形態を実現する。

プラットフォームの構成を図5に示す。図中の単語辞書、音声認識用文法やプラニングルール、行為決定ルールなどのデータはシステム設計者が用意すべきデータである。先に述べた設計方針に従って対話制御部において二段構成のルール群（プラニングルールと行為決定ルール）を読み込んで対話制御に用いる。

6.2 韻律情報を用いた対話制御例

上述のプラットフォーム上での韻律情報を用いたあいづち出力制御の実現例を示す。なお、ここではあいづちを「相手の発話権継続を促すための合図」と位置付け、発話終了前に発話権がユーザから一時的に譲渡された場合にあいづちの出力を行うことを目指す。そのために、対話における発話権についての状態を心的状態としてとらえ、ユーザに発話権がある状態から一時的にシステムに譲渡された状態への変化を認識し、その変化に応じてあいづちの出力を行うことが必要になる。発話権についての心的状態（以降、発話権状態）のモデル化を行ったうえで、さらに、タスク指向対話を速やかに進行するために必要な対話基本状態のモデルとの統合を行って、行為決定ルールの形に変換する。データベース検索タスクにおける検索条件取得プランを例として、オートマトンの形で表現された対話基本状態と発話権状態のモデルと、それらが合成された行為決定ルールを図6に示す。

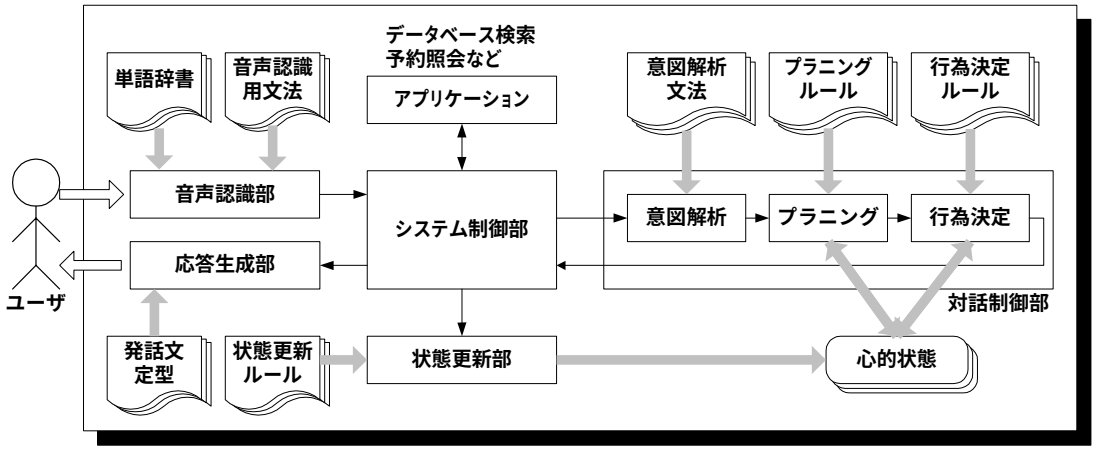

図5．音声対話システム汎用プラットフォームの構成

図6の”U_発話停止”や”U_発話終了”といったイベントは、状態更新部において判断される。これらの更新を無音時間長のみによって判断するように設定してもよいが、F0の句末音調を併用して判断すればより適切に状態を更新できる。

あいづち出力制御に有益な韻律情報[17][18]を以下に列挙する。

無音時間長 – しきい値設定により一時的な発話権譲渡か完全な譲渡かを判断できる。

F0 句末音調 – 下降調もしくは上昇-下降調、半疑問調の場合、一時的な譲渡と判断でき、下降調の場合にはその度合いによっては完全な譲渡と判断できる。

発話速度変化 – 単位末で速度変化が緩やかになれば発話権譲渡と判断する。

これらの情報を利用可能とするために、あらかじめ無音時間計測や F0句末音調照合などの手続きを用意している。システム設計者はこれらの手続きの結果に応じて状態更新が行われるよう状態更新ルールを記述すればよい。

図7に、論文検索対話における韻律情報を用いたあいづち出力の例を示す。ユーザ発話の「西川さんの論文で」の句において F0句末音調が上昇-下降調であり、なおかつポーズ長が設定されたしきい値を超えているために、図6で示した心的状態モデルにおける”検索条件待ち”状態から”発話権一時譲渡”状態に移行される。その結果、” 発話権一時譲渡”状態での行為として指定されているあいづち行為をシステムが実行する。

ここではあいづち行為の詳細については触れなかったが、多様なあいづち表現や画面表示による代替的なあいづち行為を、他の次元の心的状態を参照して適切に選択する枠組みについても本プラットフォームにおいて搭載することができる。

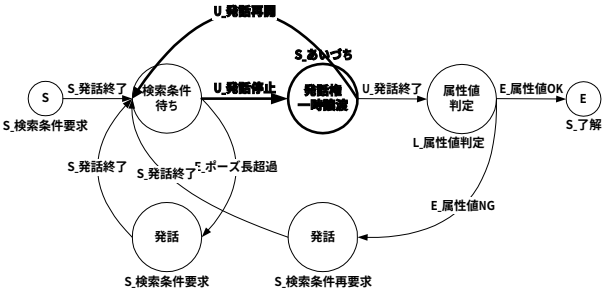

図 6. オートマトンにより表現された行為決定ルール(データベース検索タスクにおける検索条件取得プラン)

図7．韻律情報を用いたあいづち出力の例

7. まとめ

対話制御において重要な対話状態・心的状態推定を目的として、韻律情報を利用したモデルを構築し、システムに実装するための枠組みを検討した。まず、話題境界の判別において、韻律情報におけるアクセント句単位でのパラメータを用いて統計的なモデルを学習し、オープンデータに対しても人間と同程度の判別精度が得られることを確認した。さらに、システムと人間のタスク指向対話を収録し、特に心理的変化が顕著に現れると考えられるシステムの誤解発話に対するユーザ応答の分析を行ったところ、ユーザの戸惑いがポーズ長、パワー、基本周波

数増加に影響を与えていることが分かった。こうした戸惑いや慣れなどのユーザの心的な状態をユーザ発話の韻律情報から推測するモデルを構築することの見通しを立てた。韻律を利用した心的状態推定モデルをより精密に構築するために、アクセント句を人手で解析してラベル情報として利用することを試みた。これにより、喜びや怒りなどの基本感情の識別、発話における丁寧さの判別において精度が向上することを確認した。また、こうして得た様々なモデルを実装するための音声対話システムの汎用的なプラットフォームを開発した。

本研究を通して、音声の韻律情報を利用することで音声対話システムとユーザのコミュニケーションをより自然で円滑なものにする見通しを得た。今後は、本稿で述べたものやそれ以外の心的状態を多次元的に扱うことによってより自然なコミュニケーションの実現を目指す。また、状態推定モデルを精緻化するとともに様々な応用分野[15]に役立てていきたい。

参考文献

[1] 市川熹, “対話理解に対する抑揚情報の利用,” *情報処理学会研究報告*, SLP-2-8, pp.51-58 (1994).

[2] E. Couper-Kuhlen, M. Selting (Eds.), “Prosody in conversation,” Cambridge University Press, (1996).

[3] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y. “An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogues,” *Language and Speech*, Vol.41, pp.295-321 (1998).

[4] 垣田邦子, “簡単な"質問ー答"形式の対話におけるF0の話者間相互作用,” *日本音響学会講演論文集*, 2-P-2, pp.305-306 (1995).

[5] 野口広彰, 片桐恭弘, 伝康晴, “あいづち促進・抑制に対する韻律と品詞の影響,” *日本認知科学会第17回大会発表論文集*, pp.240-241 (2000).

[6] 村井美智代, 山下洋一, “談話セグメントと韻律情報の関連について,” *人工知能学会研究会資料*, 第28回 SIG-SLUD, pp.37-44 (1999).

[7] 小磯花絵, 米山聖子, 槙洋一, “「日本語話し言葉コーパス」を用いた談話構造と韻律との関係に関する一考察,” 人工知能学会研究会資料, SIG-SLUD-A203-P17, pp.139-144 (2003).

[8] 人工知能学会談話・対話研究におけるコーパス利用研究グループ, “様々な応用研究に向けた談話タグ付き音声コーパス”, *人工知能学会研究会資料*, 第28回 SIG-SLUD, pp.19-24 (1999).

[9] 山下洋一, 小磯花絵, 堀内靖雄, “音声対話に対する談話セグメントタグ方式の検討,” *人工知能学会誌*, Vol.14, pp.84-91 (1999).

[10] 岩野裕利, 杉田洋介, 松永美穂, 白井克彦, “対面および非対面における対話の違い ～頭の振りの役割分析～,” *情報処理学会研究報告*, SLP-15-19, (1997).

[11] 前川喜久雄, 菊池英明, 五十嵐陽介, “X-JToBI：自発音声の韻律ラベリングスキーム,” *情報処理学会研究報告*, SLP-39-23, pp.25-30 (1997).

[12] 武田昌一, 海老義人, 鈴木修平, 舟木克年, “落語音声の韻律的特徴の解析,” *日本音響学会秋季研究発表会講演論文集*, 1-P-4, pp.289-290 (1996).

[13] 青山一美, 平野泉, 菊池英明, 白井克彦, "音声対話システム汎用プラットフォームの検討," *情報処理学会研究報告*, 2000-SLP-30, pp.7-12, (2000.2).

[14] 鈴木堅悟, 青山一美, 菊池英明, 白井克彦, "多次元心的状態を扱う音声対話システムの構築," *情報処理学会研究報告*, 2001-SLP-37, pp.13-18, (2001).

[15] 大久保崇, 菊池英明, 白井克彦, “韻律情報を利用した文章入力システムのための韻律制御モデル,” *日本音響学会秋季研究発表会講演論文集*, (2004) (発表予定).

[16] J.Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers.

[17] N., Ward, “Using prosodic clues to decide when to produce back-channel utterances,” *Proc. of ICSLP1996*, pp.1728-1731 (1996).

[18] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一, “韻律情報を用いたあいづちの挿入,” *情報処理学会論文誌*, Vol.40, No.1, pp.469-478 (1999).