

## 音声対話における韻律を用いた話題境界検出

大久保 崇<sup>†</sup> 菊池 英明<sup>††</sup> 白井 克彦<sup>†</sup>

<sup>†</sup> 早稲田大学 理工学部

<sup>††</sup> 早稲田大学 人間科学部

あらまし 音声対話の話題による自動分割を目指して、話題の切れ目の深さと韻律情報の関係を分析した。これまでも、談話構造と韻律情報の関係を分析した研究は行われてきたが、本研究ではより詳細な分析を行うため、発話の開始アクセント句と終了アクセント句における韻律情報に関する20種類のパラメータを用いた。その結果、幾つかの韻律情報について話題の切れ目の深さとの相関が得られた。次に、決定木学習を利用した話題の切れ目の深さを判別する実験を行った。判別実験の精度は今後の検討を要するものであったが、得られた決定木から、話題の切れ目の判別に寄与する韻律情報についての考察が得られた。

キーワード 音声対話, 韻律情報, 談話セグメント, 決定木

## Detection of Topic Boundary by Prosodic Information in Spoken Dialogues

Takashi OKUBO<sup>†</sup>, Hideaki KIKUCHI<sup>††</sup>, and Katsuhiko SHIRAI<sup>†</sup>

<sup>†</sup> School of Science and Engineering, WASDEA University

<sup>††</sup> School of Human Science, WASEDA University

**Abstract** We analyzed relationship between topic segments and prosodic information in spoken dialogues, for the purpose of realization of automatic topic segmentation. We introduced twenty types of parameters of prosodic information 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 constructed decision tree model to judge strength of topic boundaries and measured its accuracy. As the result, we get some considerations about relationship between prosodic features, discourse markers and speaker-changes.

**Key words** spoken dialogue, prosodic information, topic segment, decision tree

### 1. はじめに

本研究では、人間同士の会話を自動的、又は半自動的に話題で分割する技術の実現を目的とする。この技術は、会議の議事録の自動生成や、会話の内容による動画像のシーン分割に応用できる。また、リアルタイムで話題の分割が行えれば、複数のタスクにまたがって自由度の高い会話をするような対話システムにも応用が可能である。

話題境界の検出には、会話の内容理解が有効であることは言うまでもないが、談話構造と音声の韻律情報の関係を分析した研究 [1] [2] から、韻律情報の利用も有効であると考えられる。[1]では、対話における話題の切れ目の深さについて、基本周波数とパワーとの間に相関が見られることが報告されている。また、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た [2] では、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の関係をより詳細に分析し、その分析結果に基づいて決定木を用いた判別実験を行う。

本稿ではまず、対話における話題の切れ目の深さと韻律情報の各パラメータとの関係を分析する。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、話題の切れ目との前後関係に着目する。そして、分析によって得られた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目指す。以下では、2章で分析方法や分析データについて述べ、3章に分析の結果を示す。そして、4章で決定木を用いた判別実験を行い、5章でまとめとする。

## 2. 話題境界と韻律の各パラメータの比較

### 2.1 分析に使用するデータ

#### 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 [3] を使用した。このコーパスは2名の話者間で行われた課題遂行対話を収録したものであり、「談話タグワーキンググループ」の活動の中で検討された様々なタグ [4] が付与されている。

#### 2.1.2 TBI

本研究では、対話中の話題の切れ目を、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) [5] に基づいて決定する。TBI は 1, 2 の二段階で話題の切れ目の深さを表し、話題の遷移が大きい場合に 2 を、そうでない場合は 1 をタグ付け作業の主観で評価し付与する。また、完全に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。

#### 2.2 分析方法

まず、コーパスから 5 つの対話 (タスクはクロスワード、旅館予約、会議室予約、地図課題) を選び、対話中のすべての発話から話題の遷移に関係ないと思われる「相槌」「フィラー」「言い淀み」のみで発話を構成しているもの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 の内訳を示した。

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

次に、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 [6] 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 (図 1 参照) を抽出することによって、分析対象を発話の開始部分と終了部分に限定した。この時、話者によっては多くの発話が「えー」や「あー」といったフィラーから始まる場合があり、分析対象の多くがフィラーになってしまう事を防ぐため、最初のアクセント句がフィラーの発話については、フィラー以外の最初のアクセント句を発話の開始アクセント句として抽出した。そして、抽出した区間の韻律情報を ESPS/waves+ を用いて測定した。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、話者ごとの平均と標準偏差を用いて標準化した値を使用した。扱う韻律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)

以上のような韻律パラメータを用いて、TBI との関係について分析を行う。図 1 に、分析の際に発話のどの部分を対象しているかを示す。図中の対象発話は TBI の付与されている発話であり、その直前の発話を先行発話と定義する。

まず、各 TBI の対象発話における韻律的な特徴を調べる。次

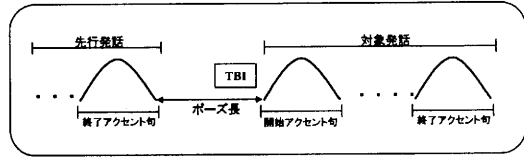


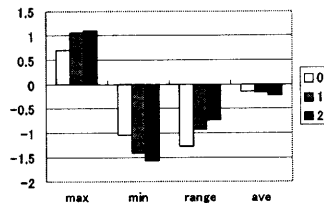
図 1 発話中の分析に用いる部分

に、対象発話の開始アクセント句と先行発話の終了アクセント句との差分をとり TBI との関係进行分析する。また、各韻律パラメータだけでなく、図中に示した対象発話と先行発話間のポーズ長と TBI の関係や、話者交代の有無と TBI の関係についても分析する。

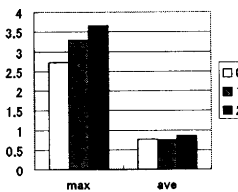
## 3. 分析結果

### 3.1 対象発話の終了部分の韻律と TBI の関係

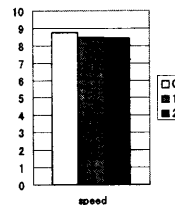
図 2 に、TBI 毎に対象発話の終了部分における各韻律パラメータを平均した結果を示した。このグラフから、まずパワーの最大について TBI と正の相関が見られる。また、基本周波数の最大とレンジについても正の相関が見られる。これらは、TBI が 1 や 2 の発話に、疑問形や半疑問形のイントネーションで終わる発話が多いことが影響していると考えられる。



(a) 基本周波数



(b) パワー

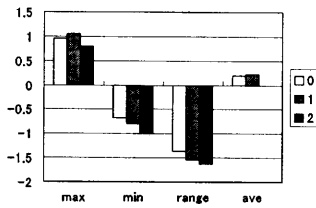


(c) 発話速度

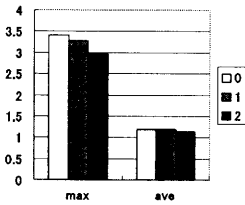
図 2 発話の終了部分の韻律パラメータと TBI の関係

### 3.2 対象発話の開始部分の韻律と TBI の関係

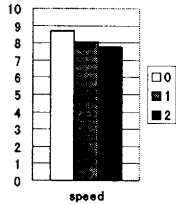
図 3 に、TBI 毎に対象発話の開始部分における各韻律パラメータを平均した結果を示した。結果から開始アクセント句の各パラメータに関して考察する。発話速度に関して、TBI が大きくなると発話速度が遅くなるという傾向が見られる。このことから、話題の切れ目が深くなるほど、最初の部分をゆっくり話す傾向があり、これは、話者が新しい話題の開始を意識して、わかり易くゆっくり話し出す傾向がある事を示していると考えられる。しかし、基本周波数やパワーについては、TBI とは負の相関が見られる。従来の研究では、話題を変える発話の開始部分では、基本周波数やパワーが他の発話に比べて大きくなる傾向が見られていたことを考慮して、発話データの内容からその要因を考えてみたところ、談話標識が関係している可能性が考えられた。そこで、次節で談話標識と TBI の関係を述べ、談話標識を考慮して発話の開始部分の分析を行う。



(a) 基本周波数



(b) パワー



(c) 発話速度

図 3 発話の開始部分の韻律パラメータと TBI の関係

### 3.3 発話の開始部分と談話標識

談話標識とは「話題の始まり、転換、途切れた会話の再開など、談話同士の対応付けの機能を持つもの」であり、本研究で用いた対話コーパスに対して談話標識の認定を行った研究 [7] では、図 4 に示す語を談話標識と認定している。

これらの語は TBI が 1 や 2 の発話の先頭に多く現れる事が

「じゃ」「では」「さいしょ(に)」「まず」「つぎ(に)」「こんど」「それでは」「で」「そういえば」「そしたら」「それで」「さいご(に)」「それじゃ」「そうすると」「でしたら」「ま」「ようするに」「つまり」

※ ( ) 内の語は省略可

図 4 談話標識の候補語

確認されている [7]。図 5 に、TBI 毎に談話標識を含む発話の割合を示した。TBI が 0 の発話には談話標識がほとんど現れず、逆に 2 の発話では約半分の発話が談話標識から始まっていることがわかる。また、図 4 からわかるように、いずれの談話標識も 1 モーラから 4 モーラ程度の短い語が多く、韻律的な特徴が現れにくいと考えられる。また、談話標識には機能語としての特性から、通常の発話の開始アクセント句とは韻律的な特徴が異なることも考えられる。これらのことから、開始アクセント句が談話標識である場合、それを考慮に入れた分析を行う必要があると考えた。そこで、発話の開始アクセント句に談話標識がある発話については、その次のアクセント句（それが談話標識やフィラーの場合はその次のアクセント句）を発話の開始アクセント句として分析の対象とした。図 6 に TBI 毎に各韻律パラメータの平均を求めた結果を示す。

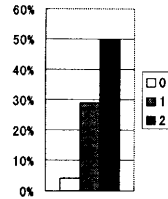
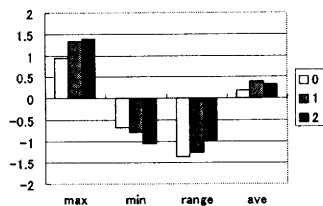


図 5 談話標識から始まる発話の割合と TBI の関係

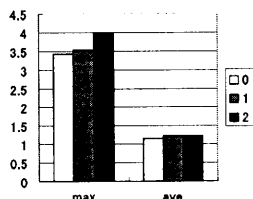
図 6 より、基本周波数のレンジやパワーの最大値に TBI との相関が見られることがわかる。この事から話題の切れ目が深くなるほど、発話の開始部分の声が大きくなる傾向があることが考えられる。基本周波数のレンジについては、アクセントの有無を考慮に入れる必要があるのではいえないが、話題の切れ目の深さに比例して基本周波数も大きくなるという事も考えられる。また、発話速度については、談話標識を除外することで、よりはっきりと TBI との負の相関を確認することができた。以上の事から、談話標識を考慮することで、TBI 毎の韻律の差異をより明確にとらえることが考察できる。

### 3.4 先行発話との差分と TBI の関係

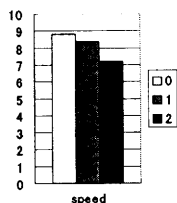
次に、発話の前後関係を考慮に入れるため、TBI 対象発話の開始アクセント句における各パラメータと、先行発話の終了アクセント句における各パラメータの差分を計算し平均を算出した。この時の開始アクセント句は、談話標識を除いたものを採用する。結果を図 7 に示す。



(a) 基本周波数



(b) パワー



(c) 発話速度

図6 談話標識除外をした発話の開始部分の韻律パラメータとTBIの関係

図7より、直前発話との差分をとることで、基本周波数の平均、最大値とパワーの最大値、平均値において、TBIとのより大きな相関を確認することができた。このことから、話者が話題を転換させる場合、直前の発話よりも声を大きく、高くする傾向があることがわかる。

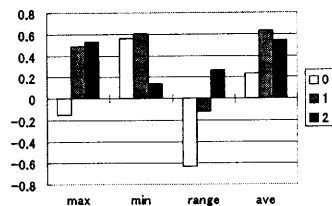
### 3.5 発話間のポーズ・話者交代とTBIの関係

発話間のポーズ長や話者交代の有無についてもTBIとの関係を分析する。図8にTBI毎のポーズ長を平均した結果を、図9にTBI毎の話者交代の起こる割合を示した。図8から、対象発話と先行発話の話題の切れ目が最も深いとき、ポーズ長が最も長くなることがわかる。図9より、TBIが0の時に最も話者交代が起こる割合が多い。

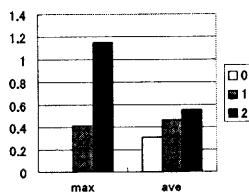
以上、話題の切れ目の深さと韻律の関係を実験した。次章では、これらの分析結果に基づいて、決定木を用いた話題の切れ目の深さの判別実験を行う。

## 4. 決定木学習を用いた話題の切れ目の深さの判別

本章では決定木学習を利用して分析対象である200発話を、TBIが0か1か2の3カテゴリに分類する。決定木の作成にはC4.5アルゴリズム[8]を利用する。また、冗長な木ができる



(a) 基本周波数



(b) パワー

図7 先行発話との差分とTBIの関係

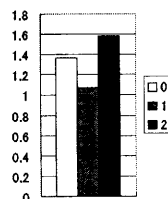


図8 発話間のポーズ長とTBIの関係

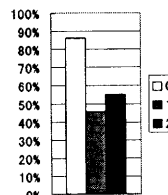


図9 話者交代の起こる割合とTBIの関係

のを防ぐため、枝刈りを信頼度1%で行う。用いるパラメータについては次節で改めて説明する。

まず、クローズトなデータに対する決定木の評価を行うため、分類対象となる200発話すべてを訓練データとして決定木を作成し、木の精度や分類に有効なパラメータを考察する。次に、10分割の交差検定を行い、判別率などから決定木学習による判別精度を評価する。

### 4.1 決定木作成に用いるパラメータ

決定木作成には、3章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。これらの韻律パラメータについて、決定木中の表記方法の一覧を表2に示した。表の

1 行目にある，“開始”は開始アクセント句を，“終了”は終了アクセント句，“差分”は TBI 対象発話の開始アクセント句と先行発話の終了アクセント句の差分を表す。

表 2 韻律パラメーター一覧

		開始	終了	差分
基本周波数 (f0)	最大	F-f0_max	L-f0_max	S-f0_max
	最小	F-f0_min	L-f0_min	S-f0_min
	レンジ	F-f0_rang	L-f0_rang	S-f0_rang
	平均	F-f0_ave	L-f0_ave	S-f0_ave
パワー (pw)	最大	F-pw_max	L-pw_max	S-pw_max
	平均	F-pw_min	L-pw_min	S-pw_min
発話速度		F-speed	L-speed	

また、表に示した 20 種の韻律パラメータの他に、3 章の分析で取り上げた以下の 3 つの要素もパラメータとして用いる。

- pause：発話間のポーズ長
- 話者交代：話者の交代が起こったかどうか
- 談話標識：発話の先頭が談話標識であるかどうか

以上の 23 種類のパラメータから決定木を作成する。パラメータについて、話者交代と談話標識は“有”か“無”の 2 値の変数とし、それ以外のパラメータは実数である。

#### 4.2 クローストな決定木の評価と考察

全 200 発話を訓練データとして用いて決定木を作成した。図 10 は、作成された決定木の概要である。図中の楕円には分岐に使われたパラメータを表記し、その両脇には分岐の条件を記した。談話標識と話者交代のパラメータに関しては、分岐条件は“有”か“無”のみであり、その他のパラメータに関しては、C4.5 アルゴリズムによって算出された閾値より大きい小さいかである。図では、右の分岐に閾値より大きいデータを、左の分岐に閾値より小さいデータが分類されるように示した。また、決定木の葉にあたる部分には、分類された TBI の結果を示した。

決定木の精度評価は、決定木の判別率と判別の信頼度を示す  $\kappa$  値を用いる。表 3 の上段に、図 10 に示した決定木の判別率と  $\kappa$  値を示した。この判別率と  $\kappa$  値から、作成された決定木による分類は精度が高く信頼できるものと考えられる。そのため、作成された決定木に分岐条件から TBI の分類に寄与するパラメータを考察することには十分意味があるといえる。そこで、決定木に分岐に用いられるパラメータを考察していく。

図 10 から、まず、談話標識の有無や話者交代の有無といった、発話の韻律以外の情報を用いて分類が行われ、その後幾つかの韻律情報を用いて TBI 毎に分類されていることがわかる。また、分岐に使われる韻律情報を見ると、発話の開始部分の情報が多く使われていることがわかる。このことから、発話の開始部分の韻律情報が TBI の判別に寄与していることがわかる。また、韻律以外の情報によって、判別に寄与する韻律情報が変化することも考察される。例えば、話者交代が行われない場合は、その発話の開始部分の韻律を分類に使っているが、話者が交代する場合は、直前の発話との差分、つまり相手の発話との比較を分類に用いていることがわかる。

表 3 決定木による判別の精度

	判別率	$\kappa$ 値
クローストな木の評価	81 %	0.7
交差検定による評価	56.5 %	0.301

#### 4.3 交差検定による判別の評価と考察

次に 200 発話に対して 10 分割の交差検定を行い、決定木を用いた TBI の判別精度を調べる。交差検定による判別精度も、さきほどの決定木の評価と同様に、判別率と  $\kappa$  値に基づいて評価する。交差検定による判別率は表 3 の下段に示した。判別率をみてみると、60%未満であり判別の精度がよいとは言えない。また  $\kappa$  値についても 0.3 と低い値しか得られなかった。しかし人手による TBI の付与実験を行った結果 [1] においても、 $\kappa$  の平均は 0.385 とそれほど高くない。このことから、決定木における判別の  $\kappa$  がとりわけ低いわけではないことがうかがえる。

決定木による判別の精度がよくなかった原因として、データ量の不足やパラメータの不足が挙げられるが、根本的な判別の精度の向上には韻律情報の特性や談話構造における話題の特性を考慮することが必要だと考えられる。そもそも韻律情報には、談話構造に関する情報以外に、発話に含まれるイントネーションや感情といった情報も含まれている。これらを考慮せずに、韻律パラメータから話題の切れ目の判別を行おうとしたため、イントネーションや感情といったその他の情報の影響を受け、判別がうまくいかなかったといえる。図 10 の決定木では、談話標識や話者交代といった情報に関する分岐に始まり、その分岐毎に異なる韻律パラメータを用いて話題の切れ目が判別されている。談話標識は話題の始まりなどに関係する言語情報であり、話者交代は対話の構造に関する情報である。このように、話題の遷移に関する情報と、それに深い韻律情報を用いて判別を行っていくことが、判別精度の向上につながると思われる。

## 5. ま と め

本稿では、基本周波数やパワーを中心とした韻律情報と話題の切れ目の関係を分析し、決定木による判別を検討した。

分析の結果、いくつかの韻律パラメータに TBI との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に相関が見られた。次に、発話の開始部分に関しては、TBI が大きくなるほど談話標識が現れる割合が多くなる傾向が見られた。また、その談話標識は 1 モーラから 4 モーラの短い語であり、韻律的な特徴が現れ難いため、談話標識から始まる発話は、談話標識の次の部分を発話の開始と考えることで、韻律と TBI との間に相関が見られることがわかった。また、発話の開始部分と先行する発話の終了部分との差分を取ることで、発話の前後の韻律と TBI との関係性を分析したが、話題の切れ目が深くなるほど、先行する発話より基本周波数やパワーを大きく発話する傾向があることもわかった。

このように、TBI と韻律の間には判別に有効と思われるいくつかの関係が見られた。そして、実際に決定木を使って判別実

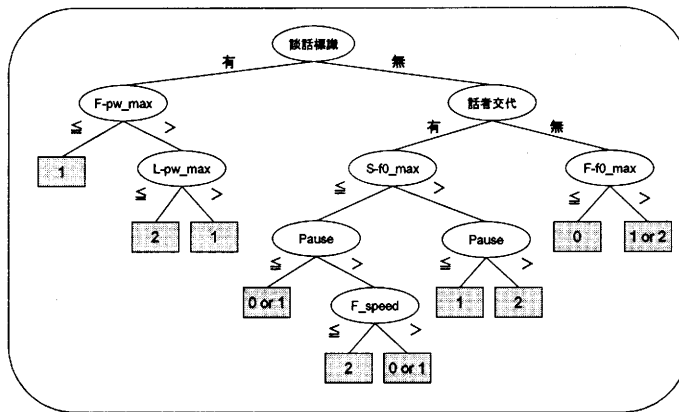


図 10 全 200 発話から作成した決定木

験を行ってみたところ、分析において TBI との相関が見られたパラメータが分類にも寄与していることがわかった。また、分類のされ方を順に見ていくと、最初に談話標識の有無や話者交代の有無で分岐され、その分岐毎に異なる韻律情報を使って TBI が分類されていることが得られた。このように、話題の遷移に関係する言語的な情報や対話構造の情報にともない、話題の切れ目に寄与する発話中の韻律情報が変化することが考察できた。

今後は、話題の遷移との関係が深いと思われる対話の構造や言語的な情報と、それに関係した韻律的な特徴を分析することで、話題の切れ目と韻律のより詳細な分析を行っていく。そして、その結果に基づいて話題境界の判別精度の向上を目指す。

## 謝 辞

本研究の一部は、早稲田大学理工学総合研究センターの研究課題「マルチモーダル情報空間における統合的ヒューマンインタフェースに関する研究」によるものである。ここに記して謝意を表する。

## 文 献

- [1] 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", 第 28 回 SIG-SLUD, pp37-44, 1999
- [2] 小磯花絵, 米山聖子, 榎洋一, "「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, 2003.3. 9
- [3] 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」1999 年度版音声対話コーパス, [CD-ROM], 1999
- [4] 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", 第 28 回 SIG-SLUD, pp19-24, 1999
- [5] 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 84-91, 1999
- [6] 前川喜久雄, 菊池英明, 五十嵐陽介, "X<sub>J</sub>T<sub>O</sub>B<sub>I</sub>: 自発音声の韻律ラベリングスキーム", 情報処理, SLP-39-23, pp25-30, 2002
- [7] 中里取, 田本真詞, 菊池英明, 吉村隆, "課題遂行対話における対話潤滑語の認定", 人工知能学会誌, Vol.14, No.5, pp.900-906, 1999.9.
- [8] J.Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers