音声対話における韻律を用いた話題境界検出

菊池 英明† **大久保 崇**‡ **白井 克彦**‡

†早稲田大学人間科学部
埼玉県所沢市三ヶ島 2-579-15
†早稲田大学理工学部
東京都新宿区大久保 3-4-1
kikuchi@waseda.jp, {okubo,shirai}@shirai.info.waseda.ac.jp

概要

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.

1. はじめに

対話音声における韻律には，同音異議語の区別や文体の区別をはじめ，文構造の明示，強調や感情伝達，心理状態の表現など対話特有の様々な役割があり [1]，音声対話システムにおける韻律情報の利用は重要な課題である．これまでに我々は音声対話システムにおける様々な局面での韻律情報の利用を検討してきた．本稿では談話構造の解析を目的とした，韻律情報による話題境界検出の試みについて報告する．

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．[2] 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 [3] 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．さらに，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．

2. 音声対話における話題境界と韻律

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 [4] 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 [5] が付与されている．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）[6] 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 をタグ付け作業者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．

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

2.2. 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[7] 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（図 1 参照）を抽出した．そのうえで，抽出した区間の韻律情報を ESPS/waves+により測定した．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)

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にしているかを示す．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．

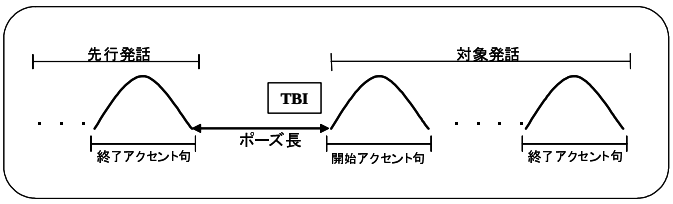


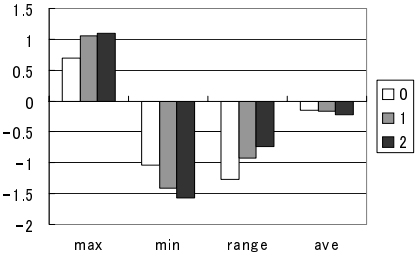
図 1. 発話中の分析対象

3. 話題境界と韻律の関係

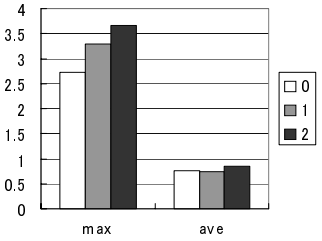
本章では，前章に述べた方法によって話題境界と韻律の関係を分析した結果を述べる．

3.1. 対象発話について

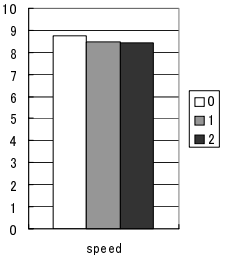
まず，対象発話について TBI と韻律の関係を調べる．対象発話の終了部分と開始部分について，TBI 毎の各韻律パラメータの平均をそれぞれ図 2 と図 3 に示す．



(a) 基本周波数



(b) パワー

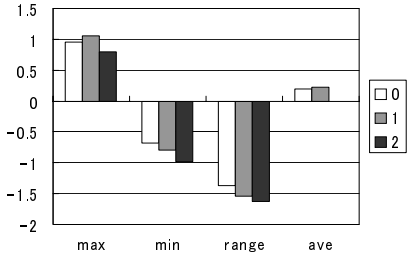


(c) 発話速度

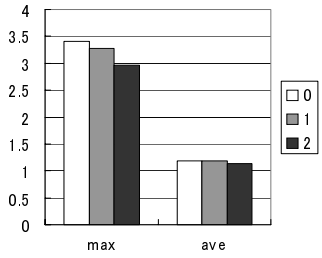
図 2. 発話の終了部分の韻律パラメータと TBI の関係

まず発話の終了部分において，基本周波数のレンジとパワーの最大値について TBI との間に正の相関が見られる．これは，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．

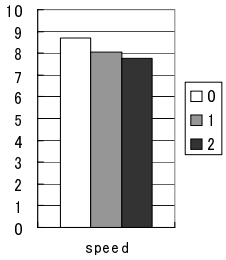
一方，発話の開始部分において，基本周波数のレンジやパワーの最大値について，TBI との間に負の相関が見られる．これに対し，[2] などでは，話題を変える発話の開始部分では，基本周波数やパワーが他の発話に比べて大きくなる傾向が報告されている．発話の内容を観察したところ，談話標識の存在が関与していると考え，談話標識を除外して分析を行なった．対象発話の開始部分について談話標識を除外した後の TBI 毎の各韻律パラメータの平均を図 4 に示す．なお，談話標識とは「話題の始まり，転換，途切れた会話の再開など，談話同士の対応付けの機能を持つもの」であり，あらかじめ特定の語が談話標識として認定されている [8].



(a) 基本周波数

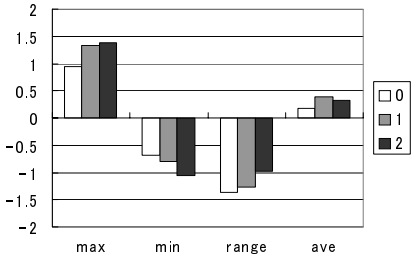


(b) パワー

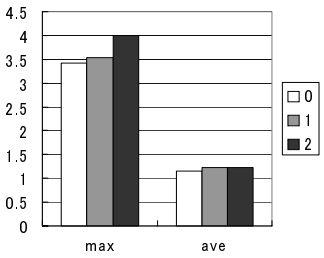


(c) 発話速度

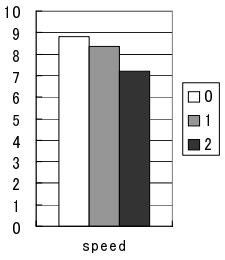
図 3. 発話の開始部分の韻律パラメータと TBI の関係



(a) 基本周波数



(b) パワー



(c) 発話速度

図 4. 談話標識を除外した発話の開始部分の韻律パラメータと TBI の関係

図 4 より，談話標識を除外すれば，発話の開始部分についても基本周波数のレンジやパワーの最大値と TBI との間に正の相関が見られることがわかる．また，発話速度について，談話標識を除外することで，よりはっきりと TBI との間に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 毎の韻律の違いをより明確にとらえることができるといえる．

3.2. 先行発話との差分について

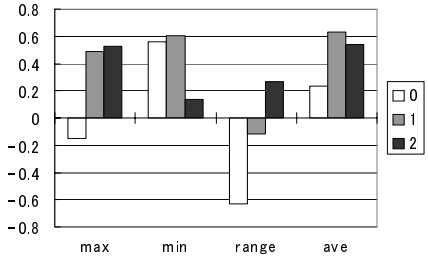
次に，発話の前後関係を考慮にいれるために，TBI 対象発話の開始アクセント句における各パラメータと，先行発話の終了アクセント句における各パラメータの差分を求め，その平均と TBI の関係を分析した．この時の開始アクセント句は，談話標識を除いたものを採用する．結果を図 5 に示す．

図 5 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 との間により強い相関を確認することができた．この事から，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といえる．

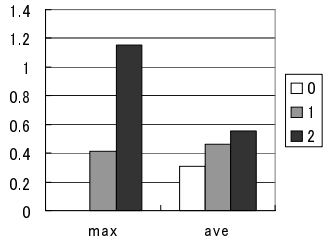
3.3. 発話間のポーズ・話者交代と TBI の関係

次に，発話間のポーズ長や話者交代の有無について TBI との関係を調べる．図 6 に TBI 毎のポーズ長を平均した結果を，図 7 に TBI 毎の話者交代の起こる割合を示した．図 6 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 7 より，TBI が 0 の時が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分析した．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．



(a) 基本周波数



(b) パワー

図 5. 先行発話との差分と TBI の関係

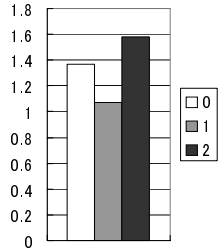


図 6. 発話間のポーズ長と TBI の関係

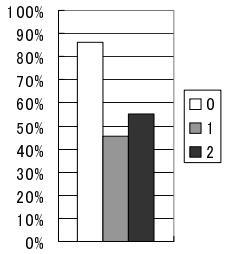


図 7. 話者交代の起こる割合と TBI の関係

4. 決定木による話題の切れ目の深さの判別

本章では，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 章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 [9] 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．用いるパラメータについては次節で改めて説明する．

4.1. 決定木学習に用いるパラメータ

決定木学習には，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

4.2. クローズ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学習した．図 8 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．分岐条件は，談話標識と話者交代のパラメータに関しては”有”か”無”であり，その他のパラメータに関しては，算出された閾値との大小関係である．図には，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 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κ値を用いる．表 3 の上段に，図 8 に示した決定木の判別率とκ値を示した．この判別率とκ値から，作

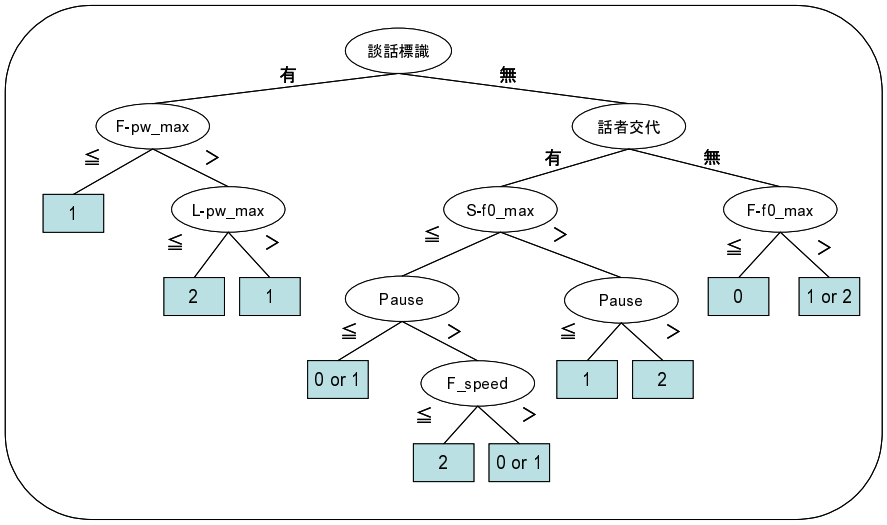


図 8. 全 200 発話から作成した決定木

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であると考えられる．

以下に，作成された決定木の分岐条件から TBI の分類に寄与するパラメータを考察する．

図 8 から，まず，談話標識有無や話者交代有無といった，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．この事から，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	κ 値
クローズトな木の評価	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 κ 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い難い．また κ 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 [2] においても，κ の平均は 0.385 とそれほど高くない．この事から，決定木における判別の κ 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，感情や心理状態などの情報も含まれている．話題の切れ目の深さのより正確な判別には，こうした情報をも考慮した総合的なモデルを構築することが必要であろう．

5. まとめ

本稿では，韻律情報と話題の切れ目の関係を分析したうえで，決定木による話題の切れ目の深さの判別を試みた．

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．因すると考える．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．

こうした分析結果に基づき，決定木を学習して判別実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．

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．

6. 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] 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, pp.37-44, 1999.

[3] 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, 2003.3.

[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」1999 年度版音声対話コーパス , [CD-ROM], 1999.

[5] 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.

[6] 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 84-91,199

[7] 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”, 情処研報, SLP-39-23, pp25-30, 2002.

[8] 中里収, 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.

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

[10] 大久保崇, 菊池英明, 白井克彦,”音声対話における韻律を用いた話題境界検出,” 信学技報, Vol.103, No.519, pp.235-240, 2003.