話者移行適格場の到来を予測させる発話中の韻律変化の解明
Prosodic Changes Leading to Transition Relevance Place in Spontaneous Utterance

石本 祐一†，榎本 美香‡
Yuichi Ishimoto, Mika Enomoto
† 国立国語研究所，‡ 東京工科大学
National Institute for Japanese Language and Linguistics, Tokyo University of Technology
yishi@ninjal.ac.jp, menomoto@stf.teu.ac.jp

Abstract

In this paper, we investigated prosodic features that project transition relevance places in Japanese spontaneous conversation to clarify acoustic features for predicting the ends of utterances. Results showed that the fundamental frequency and intensity at the beginning of the final accentual phrase indicate whether the utterance includes utterance-final elements, which are the syntactic cue for detecting the end-of-utterance. In addition, the mora duration lengthened in the final accentual phrase. Then, we carried out perceptual experiments with Japanese utterances missing the utterance-final elements. The result showed that subjects distinguished whether an UFE follows at the verb before the appearance of the UFE. That is, by using acoustic features, hearers feel that the end of the utterance arrives soon before the syntactic features appears.

Keywords — End-of-Utterance, Utterance-Final Element, Accent Phrase, Prosody

1. はじめに

実会話データの話者交替を見ると，隣接発話の話者移行間隔は 70 ms 付近に集中している．これは，前話者が話し終わるや否や，あるいは話し終わる寸前に次話者による発話が開始されていることを意味する．発話を計画してから実際に発声するまでの時間（発話潜時）が 280–340 ms である [1] ことを考えると，このような次話者の発声は話し手の発話終了位置を予測していなければなしえない．すなわち，会話において円滑な話者交替を実現するために，次話者となる聞き手は空隙が生じないように何らかの方法で話し出すタイミングを適切にはかっていると考えられる．

会話分析研究においては，Sacks ら [2] によりターン構成単位 (turn constructional unit; TCU) と呼ばれる発話単位を基にした話者交替システムが提案されている．このシステムではターン (会話の参与者たちが話す権利，発話権) は少なくともひとつ以上の TCU で構成され，TCU 末尾に存在する話者移行適格場 (transition relevance place; TRP) が聞き手が話し出すことができる適切な位置 (タイミング) であるとした．すなわち，聞き手はこの TRP を知覚もしくは予測することができれば，円滑に次話者として話し出すことが可能となる．

語順の拘束力が強く文冒頭に発話の形式を決める要素が現れる英語のような言語では，統語情報により TCU 冒頭から徐々に TRP の位置の投射がなされていくとされている [3]．一方，日本語においては構造が語順によって決定されにくい上，文の後方に様々な語が追加されることもあり，TCU 冒頭から TRP の位置を予測することは容易ではない．Tanaka[3] は日本語において TRP の開始の役割をする語として助動詞「です・ます」や終助詞「ね・よ」などを取り上げ，発話末要素 (utterance-final element) と呼んでいる．発話末要素は発話末に存在し，アスペクトやモダリティなどの付属的な意味を発話に付与する統語的要素である．この理論を榎本 [4] はさらに詳細に検討し，日本語では発話末要素の出現が TRP の開始点であることを実験により示している．しかし，会話においては発話末要素が存在しない発話も頻繁に生じる．そのような発話において，発話末要素の出現を待ち，存在しないことを確認して次発話を計画していては発話潜時により TRP が過ぎてからの発声となってしまう．つまり，聞き手は予め発話末要素がその発話に付与されるのか否かを予測した上で次発話開始のタイミングを決定していることになる．

英語において TRP は語用論的情報や統語情報，韻律情報のどれかひとつで決定されるのではなく，これらの複数の要因の統合により投射されることが Ford and Thompson[5] により指摘されている．同様に日本

語でも，発話末要素だけではなくその他の要因も組み合わさることで発話末予測が行われると考えられる．日本語の発話末における音響的特徴に関しては，小磯ら [6] が無音区間で区切られた間休止単位の末尾について言語的・韻律的特徴を調査している．その結果，実際に話者交替が起こったか否かで，間休止単位の最終モーラの時間長や基本周波数 (F_0) といった韻律的特徴が異なる傾向を示した．また，Campbell[7] は朗読音声について，文末のモーラ長が平均よりも伸縮すること (final lengthening) を示している．しかし，このような最終モーラの情報では発話末の予測には遅すぎるため，聞き手が利用しているとは言い難い．その他，発話全体における韻律情報に関して，発話頭から発話末にかけて基本周波数 (F_0) が自然下降すること (F0 declination) や，F_0 の変化幅が発話末に向かって段々と狭まっていくカタセシスと呼ばれる現象が知られている [8]．また，発話全体の下降傾向とは別に，発話末近くで final lowering と呼ばれる F_0 の局所的な下降が見られることも指摘されている [9]．日本語の韻律研究の多くは朗読音声の観察によるものが多く，自発発話とは特性が異なる可能性があるものの，これらの現象は最終モーラより前の韻律情報が TRP を決定する手がかりとなることを示唆している．

本研究は，自発発話において発話末の予測にかかわる音響的特徴を明らかにすることを目的とする．本稿ではまず分析 1 として，自発発話の韻律情報の時間的な変化の分析を行ない，韻律変化が発話末予測の手がかりとなりうるかどうかを検討する．次に分析 2 として，統語情報である発話末要素と韻律情報との関係を調べる．さらに，発話末要素部分を除去した音声を被験者に聴かせ，発話末予測の度合いを測る知覚実験を行った結果について報告する．

2. 分析 1: アクセント句単位の韻律変化

2.1 分析資料

自発発話のデータとして，千葉大学 3 人会話コーパス (Chiba3Party)[10, 11] に収録されている 12 会話を用いた．これらの会話は親近性のある 3 人がテーマを決めて自由に雑談を行う内容であり，1 会話あたり約 10 分，話者数は 36 名である．

発話単位として TCU を適用することが理想的であるが，TCU の認定は会話分析に関する詳細な知識を必要としコーパス内のデータ全体に付与することが容易ではないため，本研究では統語論的・語用論的な境界で会話を区切るために提案されている長い発話単位 (Long Utterance Unit; LUU)[12] に着目し，LUU を

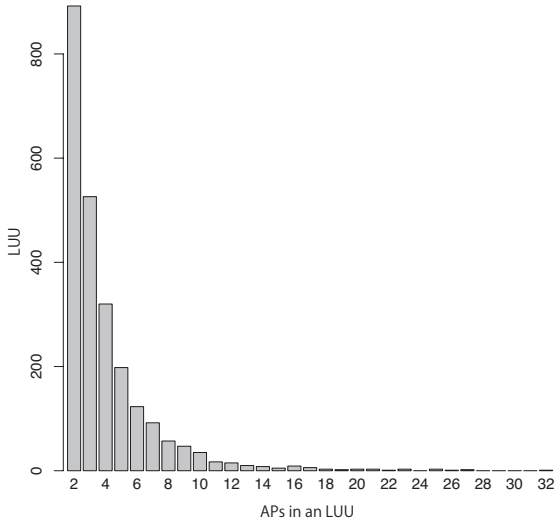

図 1　発話 (LUU) 数と LUU 内に含まれるアクセント句数

TCU の代わりの発話単位として扱う．さらに，分析対象として，話者交替が生じたか否かを示す話者移行型タグ [13] を基に，円滑な話者交替が起こった時の先行発話のみを選択した．なお，発話末尾が上昇調となる疑問形の発話は，発話末の韻律特徴が明らかであるために本研究の対象からは除外する．

2.2 分析手順

発話はひとつ以上のアクセント句 (Accent Phrase; AP) を含んでおり，Chiba3Party に付与された X-JToBI[14] から AP の境界を表す BI 層ラベルを用いて，LUU を AP 単位に分割することができる．Chiba3Party における LUU と LUU 内の AP 数との関係を図 1 に示す．AP 数が 11 以上の LUU は少数であるため，本分析では 2–10 個の AP を持つ LUU を分析対象とした．

韻律情報として AP 単位の声の高さ，強さ，速さに関わる音響的特徴量を抽出し，発話内での変化について調査した．声の高さを表す音響特徴量として，F_0 を用いた．F_0 は 1ms ごとに抽出した後に対数変換し，性差および個人差の影響を取り除くために，各 LUU の平均 F_0 を引くことで簡易的な正規化を行った．なお，抽出エラーによる異常値の影響を取り除くため，抽出された F_0 のうち上位 10%と下位 10%の点を取り除いている．この F_0 値を用い，AP ごとの F_0 の平均値 (F_0mean) を求めた．

声の強さを表すパワーについても，LUU ごとに平均パワーを求め，それを各 LUU の基準音圧として用

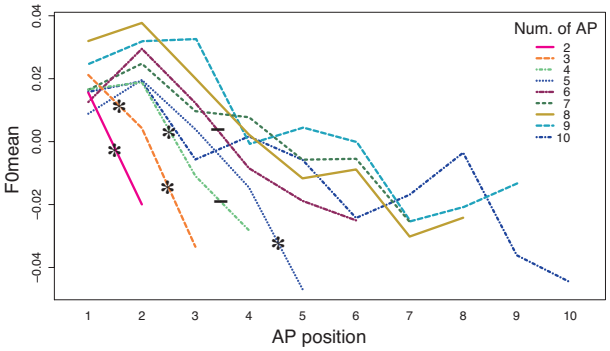

図 2　アクセント句ごとの平均 F_0(F_0mean) の変化

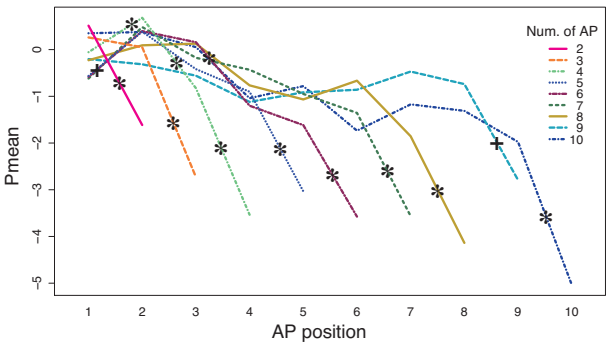

図 3　アクセント句ごとの平均パワー (Pmean) の変化

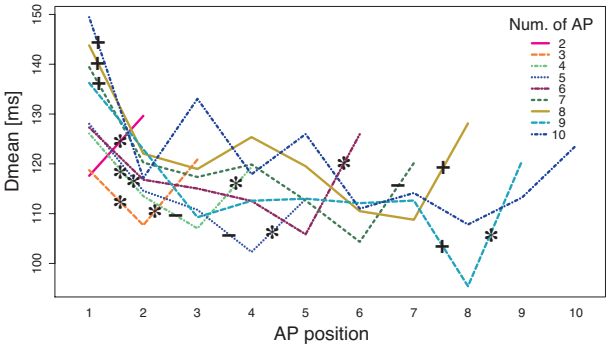

図 4　アクセント句ごとの平均モーラ長 (Dmean) の変化

いることで，録音環境の違いや個人差を小さくする正規化を行った．正規化後に，AP 全体の平均パワー (Pmean) を求めた．

声の速さを表す特徴量としては，AP ごとの時間長から，AP 内のモーラ数を用いて平均モーラ時間長 (Dmean) を算出した．平均モーラ時間長は値が小さいほど早く発話していることを意味する．

2.3　結果

図 2-4 は LUU 内の AP の位置で特徴量がどのように変化するかを，AP の位置ごとに F_0mean, Pmean, Dmean を平均することで示した図である．図中の “*”, “+” および “–” は隣接 AP 間の有意差を表しており，それぞれ有意水準 1%，5%，10%である．

図 2 をみると，LUU 内の AP 数が少ない場合には F_0 は急激に減少しているが，AP 数が多い LUU の場合は隣接した AP 間で緩やかに変化している．さらに，発話末に向けて緩やかに下降し，発話の終わり付近では LUU 内の AP 数に関わらずほぼ同じような値となっている．これは，AP 数が 2-5 個からなる発話を調査し，最終 AP はそれ以外の AP よりも常に低い F_0 となることを示した Maekawa[15] の結果とも合致している．すなわち，発話中の F_0 の低さは発話末に近づいていることを示していると言える．また，4 つ以上の AP からなる LUU では，2 番目の AP(第 2AP) でやや上昇したのち下降が始まっている．これは，話し手の F_0 が第 1AP と第 2AP の間で下降するか否かで，LUU が 3 つ以下の AP からなる発話なのかどうか，すなわち短い発話なのかどうかを予測できる可能性を示している．

図 3 から，パワーは LUU 内の最終 AP で大きく減少することがわかる．特に，AP 数が多い LUU では，最終 AP の手前の AP までは緩やかに変化しているが，最終 AP ではパワーが著しく減少している．これにより，パワーの減少は発話末の接近を表すといえる．図 4 から，最終 AP の時間長が長くなる傾向があることがわかる．発話末に向けて 1 モーラあたりの時間長は短くなっていくが，最終 AP では逆に時間長は長くなる．

これらのことから，聞き手は F_0 の下降やパワーの減少，モーラ長の伸長を認識することで，その AP が LUU の最終 AP であるかどうか，すなわち，発話末であるかどうかを予測できる可能性がある．

3.　分析 2: 発話末要素の有無と韻律変化

前節の分析において，発話の最終 AP 付近で韻律変化が顕著であることが明らかになった．そこで，本節では，最終 AP およびその直前の AP に焦点を絞り，さらに詳細に韻律変化が生じる箇所を調べる．発話末には発話末要素が置かれることが多く，この統語要素が韻律変化に何らかの影響を与えていることも考えられるため，発話末要素が存在する場合としない場合についても考慮する．

3.1　分析手順

前節で扱った LUU を発話末要素が存在するものとしないものに分類する．さらに，AP 内の統語的構造

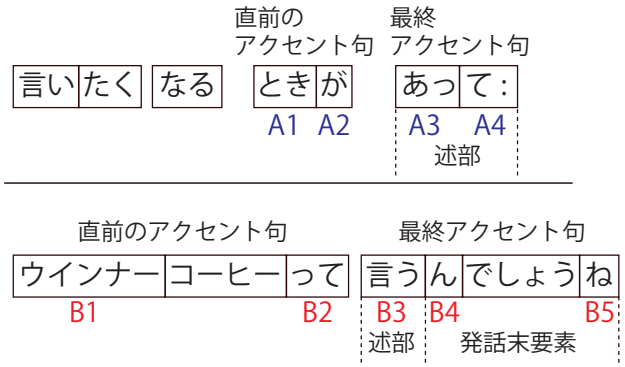

図 5　最終アクセント句およびその直前のアクセント句の分析箇所

を統制するために，最終 AP が述部で始まるもののみを対象とする．ここでいう述部とは，発話末要素がない場合は最終文節，発話末要素がある場合は発話末要素の直前の文節を意味する．さらに，AP 数が 2 以上の LUU に対して，AP 内を単語 (品詞) 単位で区切り，最終 AP とそのひとつ前の AP の中で音響的特徴に変化が現れる箇所を調べることとする．そこで，最終 AP と直前の AP のうち，発話末要素がない群では

A1 直前のアクセント句の句頭
A2 直前のアクセント句の句末
A3 最終アクセント句の句頭 (述部の開始位置)
A4 最終アクセント句の句末 (発話末)

にあたる単語を分析箇所とした．同様に，発話末要素がある群では

B1 直前のアクセント句の句頭
B2 直前のアクセント句の句末
B3 最終アクセント句の句頭 (述部の開始位置)
B4 発話末要素の句頭 (発話末要素の開始位置)
B5 最終アクセント句の句末 (発話末)

にあたる単語を分析箇所とした (図 5)．

以上の選別により，発話末要素がない発話の数は 265，発話末要素がある発話の数は 368 となった．

音響的特徴量として，前節と同様に，平均 F_0(F_0mean)，平均パワー (Pmean)，平均モーラ長 (Dmean) を分析箇所ごとに求めた．

3.2　結果

各分析箇所の F_0mean, Pmean, Dmean の平均値および Tukey 法による多重比較の結果を図 6–8 に示す．図の下部の直線は多重比較によって有意差 ($p<0.01$) が見られた区間を示している．

図 6 より，F_0 については

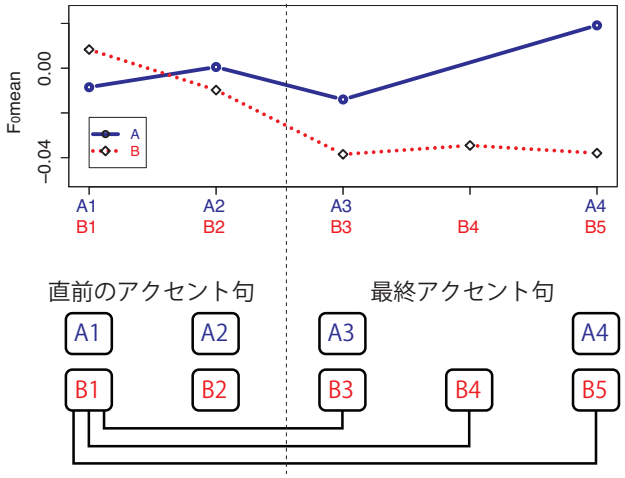

図 6　最終アクセント句およびその直前のアクセント句内の平均 F_0(F_0mean)

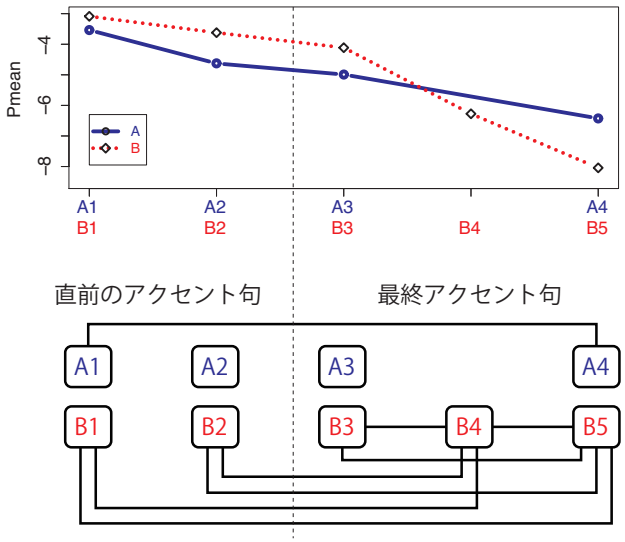

図 7　最終アクセント句およびその直前のアクセント句内の平均パワー (Pmean)

- 発話末要素がない場合は直前のアクセント句から最終アクセント句にかけて F_0 の差が見られない
- 発話末要素がある場合は最終アクセント句は直前のアクセント句頭よりも低い F_0 で始まり，最終アクセント句内ではほぼ同じような低い値となる

となった．すなわち，F_0 の低下が発話末要素のある最終アクセント句を特徴づけており，聞き手に発話末要素の出現を予測させる要因として働きうると考えられる．一方，発話末要素がない場合では，最終アクセント句であってもその直前のアクセント句からほとんど変化がみられないことから F_0 下降がないことが発話末要素の不在を示していると考えられる．

図 7 より，パワーについては次のことが読み取れる．

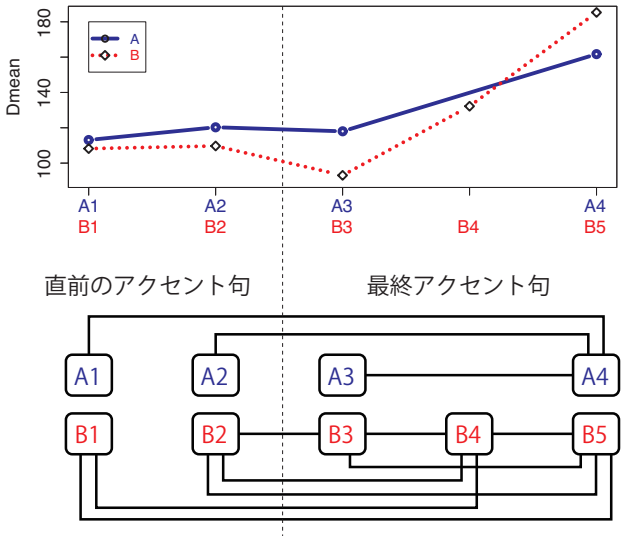

図 8　最終アクセント句およびその直前のアクセント句内の平均モーラ長 (*D*mean)

- 発話末要素がない場合は直前のアクセント句頭から最終アクセント句末まで緩やかにパワーが低下する
- 発話末要素がある場合は発話末要素からパワーが大きく低下し始める

パワーは発話末要素の有無に関わらず，発話末要素の出現前には変化が見られない．発話末要素がない場合は，述部の開始から発話末までが短いために最終アクセント句末までパワーをあまり低下させずに発声されているが，発話末要素がある場合は発話末が近づいた発話末要素の時点でパワーが顕著に低下している．すなわち，パワーの急激な低下が発話末要素の存在を特徴づけているといえる．

図 8 より，モーラ長については

- 発話末要素の有無に関わらず最終アクセント句内でモーラ長が伸長し始め，発話末では最も長いモーラ長となる
- 発話末要素がない場合は述部の終わりでモーラ長が充分に伸長しきるが，発話末要素がある場合は述部の終わりから発話末へ向けてさらにモーラ長が伸びる

となった．つまり，発話末ではモーラ長がある程度の長さまで伸びることになっているが，発話末要素がある場合は述部の終わりに到達してもまだ充分に時間長が伸長していない．すなわち，発話末要素の有無によって述部の時点でのモーラ時間長の伸長の度合いが異なると考えられる．

以上のことから，最終アクセント句での韻律変化は発話末要素の存在によるところが大きいといえる．これは，発話末要素は統語的にも韻律的にも TRP として特徴付けられていることを示唆している．

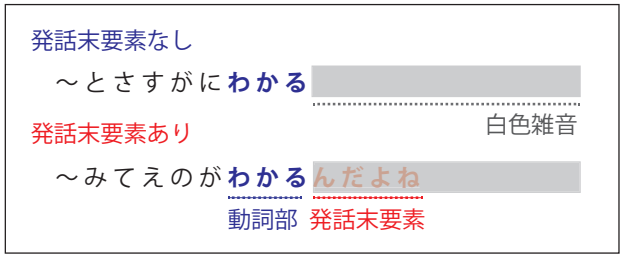

図 9　刺激音声の概形

4.　知覚実験: 発話末予測

前節において，発話末要素が統語的・韻律的に TRP を特徴付けていることが示された．では，ヒトは発話末要素を聴くことで，発話末の到来を予測しているのだろうか．それとも，発話末要素の出現前であっても，発話末予測が可能なのだろうか．本節では，発話末要素を除去した音声から発話末予測が行えるのか知覚実験を行い，この疑問について考察する．

4.1　刺激

前節までと同様，Chiba3Party に収録されている会話を用い，次の条件にあてはまるものから発話末要素がない 15 発話と発話末要素がある 25 発話を選んだ．

- 感情や疑問などの強い意図を含まない
- 動詞部分が終止形

発話の平均時間長は，発話末要素なしが 3.17 秒，発話末要素ありが 2.65 秒で，発話末要素部分は 0.31 秒であった．

次に，発話末要素がある発話に対し，発話末要素部分を除去する処理を行った．動詞部が終止形であるため，発話末要素がある発話では発話末要素部分をすべて取り除くと，発話末要素のない発話と同形となる．さらに，除去によって生じる音声の途切れ感を軽減するために，除去部分以降に発話との SNR が 30 dB 程度の白色雑音を付与し，全体で 8 秒の刺激となるようにした．同様に，発話末要素のない発話についても発話末直後に白色雑音を付与した．作成した刺激の概形を図 9 に示す．

4.2　実験手順

上述の刺激をランダムに提示して発話の開始から被験者の反応が開始されるまでの時間を測定した．被験

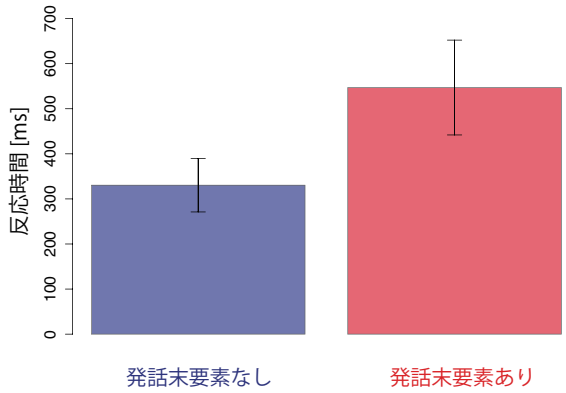

図 10　動詞部終了時刻を基準とした反応時間

者に与える教示は「発話が終了すると思うところでできるだけはやくボタンを押してください．発話の終盤に雑音が流れますが，発話が終わるところを予想してボタンを押してください．」とした．被験者は正常な聴力を有する 10〜40 代の男女 7 名である．

4.3　結果

発話末要素が元よりなかった発話（発話発話末要素なし）と発話末要素を除去した発話（発話末要素あり）に対する動詞部終了時刻を基準とした反応時間の平均と標準偏差を図 10 に示す．刺激は全て終止形で発話が終わっており，言語的には発話末要素が後続するか否かは判断できない．反応時間の t 検定を行ったところ，平均の差は有意であった（両側検定 $t(32)=-8.54$, $p< 0.01$）．すなわち，発話末要素がもともとなかった発話に比べ，発話末要素が後続するはずの発話では，動詞部までしか聴き取れなくとも被験者は反応を遅らせていることになる．これは，被験者は動詞部までの情報から発話末要素が後続するか否かを識別できていることによると考えられる．

本実験の結果は，発話末要素の後続を聞き手は言語情報によらず音響的に知りうることを示している．すなわち，聞き手は発話末要素が後続するか否かをそこまでの発話の音調から予測し，後続する場合は発話末要素を聴いてから次話者として話し，発話末要素がない場合は発話末要素の出現を待つことなく述部の終わりを発話の終わりと判定していることになる．

ただし，発話末要素が TRP として働くことで話者交替の機会を次話者に与えているのだとしても，発話末要素がない発話では発話末要素の出現を待つことはできない．発話末要素の有無に関わらず発話末から平均 70 ms 程度で次発話が開始されているという実際の会話を鑑みると，述部の終わりをもなんらかの方法で予測していることになる．発話末要素の出現前にどのような情報から発話末の到来を予測し，間断ない話者交替を実現しているのかはまだ不明であり，引き続き調査していく必要がある．

5.　おわりに

会話の聞き手が話者交替可能な位置を予測するために使用する韻律情報を明らかにすることを目的として，自発発話内の AP ごとの F_0 とパワー，モーラ長の分析を行った．その結果，発話の最終 AP 付近で韻律が顕著な変化を見せることがわかった．また，この韻律変化は統語要素である発話末要素と密接な関係にあることを示した．さらに，発話末要素を除去した音声による発話末予測の知覚実験を行なったところ，発話末要素の出現前に発話末要素の有無を判別できることがわかった．このことから，発話末要素が存在する発話の場合は，発話末要素の出現を予測することで発話末予測を行っている可能性が考えられる．しかし，発話末要素が存在しない発話の場合における発話末予測の手がかりはまだ不明であり，さらなる調査が必要である．

謝辞　本研究は JSPS 科研費 24700109, 15K00390 の助成を受けたものです．

参考文献

[1] 藤原彰彦, 正木信夫, (1998) “調音位置および調音様式の発話潜時への影響,” 電子情報通信学会技術研究報告, SP97-88, pp. 1–8.
[2] Harvey Sacks, Emanuel A. Schegloff, Gail Jefferson, (1974) “A simplest systematics for the organization of turn-taking for conversation”, *Language*, Vol. 50, No. 4, pp. 696–735.
[3] Hiroko Tanaka, (1999) *Turn-taking in Japanese conversation: a study in grammar and interaction*, John Benjamins Publishing.
[4] 榎本美香, (2009) 日本語における聞き手の話者移行適格場の認知メカニズム, ひつじ書房.
[5] Cecilia E. Ford, Sandra A. Thompson, (1996) “Interaction units in conversations: Syntactic, intonational, and pragmatic resources for the management of turns,” *Interaction and grammar*, Eds. E. Ocks *at el.*, Cambridge University Press, pp. 134–184.
[6] Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, Yasuharu Den, (1998) “An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs,” *Language and speech*, Vol. 41 No. 3–4, pp. 295–321.
[7] Nick Campbell, (1992) “Segmental elasticity and timing in Japanese speech,” *Speech perception, production and linguistic structure*, Eds. Tohkura *at el.*, Ohmsha, pp. 403–418.
[8] 田窪行則, 前川喜久雄, 窪薗晴夫, 本多清志, 白井克彦, 中川聖一, (2004) 言語の科学 *2* 音声, 岩波書店.

[9] Janet B. Pierrehumbert, Mary E. Beckman, (1998) *Japanese tone structure*, MIT Press.
[10] Yasuharu Den, Mika Enomoto, (2007) "A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation," *Conversational informatics: An engineering approach*, Ed. T. Nishida, John Wiley & Sons, pp. 307–330.
[11] http://research.nii.ac.jp/src/Chiba3Party.html
[12] Yasuharu Den, Hanae Koiso, Takehiko Maruyama, Kikuo Maekawa, Katsuya Takanashi, Mika Enomoto, Nao Yoshida, (2010) "Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme," *Proc. LREC2010*, pp. 2103–2110.
[13] 東山英治, 伝康晴, (2010) "3 人会話における発話開始時の聞き手の視線配布," 人工知能学会研究会資料, SIG-SLUD-A902-03, pp. 15–20.
[14] 前川喜久雄, 菊池英明, 五十嵐陽介, (2001) "X-JToBI : 自発音声の韻律ラベリングスキーム," 電子情報通信学会技術研究報告, SP2001-106, pp. 25–30.
[15] Kikuo Maekawa, (2010) "Final lowering and boundary pitch movements in spontaneous Japanese," *Proc. DiSS-LPSS Joint Workshop 2010*, pp. 47–50.