様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月 １４ 日現在

機関番号：６２６１８
研究種目：基盤研究(C)（一般）
研究期間：2015～2018
課題番号：１５Ｋ００３９０

研究課題名（和文）音声アシスタントとの円滑な話者交替を実現する音声言語特徴の解明

研究課題名（英文）Explication of acoustic and linguistic features for realization of smooth turn-taking in conversation with voice assistant

研究代表者

石本　祐一（Ishimoto, Yuichi）

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・特任助教

研究者番号：５０４０９７８６

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：本研究の知覚実験により、発話の終わりを特徴付ける音響的特徴であるとされてきた発話末のF0の下降（Final lowering）は、聞き手の発話末知覚にほとんど影響を与えていないことがわかった。また、発話末予測モデルを構築して統語情報と韻律情報の組み合わせの効果を調べたところ、係り先未定文節数の差分・発話末要素の有無・文節の平均F0・文節の平均モーラ長といった情報が発話末の文節およびその直前の文節の検出に寄与することが示された。これらの結果から、発話末予測には単一の特徴が利用されるのではなく、統語情報と韻律情報が密接に関係していることが示唆された。

研究成果の学術的意義や社会的意義
インタラクション研究の一分野である会話分析では、発話末付近に次話者が割り込みにならずに話しだすことができる話者移行適格場が存在するとしている。しかし、人間がどのような情報を用いて話者移行適格場を認知しているかは明らかになっておらず、情報システムへの活用にも至っていない。これまで話者移行適格場は会話分析専門家の直感によって認定されていたが、話者移行適格場を客観的な指標で検出することにより、発話中で話者交替が可能な箇所を人間の直感に頼らずに捉えることができる。また、ここで得られた音声・言語特徴を用いることで音声対話システムにおける円滑な話者交替の実現が期待できる。

研究成果の概要（英文）：In this research project, we first conducted perceptual experiments to investigate the influence on cognition of the end of the utterance using Japanese utterances with modified F0s at the middle or end. The result showed that subjects were able to detect the end of the utterance even in the absence of final lowering, and that placing an F0 downstep in the middle of the utterance to simulate final lowering did not affect the responses. This suggests that hearers do not make use of final lowering to perceive the end of an utterance, although the appearance of final lowering is more likely in the presence of certain syntactic factors.
Next, we focused on combinations of syntactic and prosodic features. We constructed a statistical model that estimates the position of a bunsetsu in an utterance from the features. The model achieved high performance. Accordingly, the combination of syntactic and prosodic features is effective in predicting end-of-utterance.

研究分野：音声工学

キーワード：発話末予測　自発発話　韻律情報　統語情報　話者移行適格場

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９、ＣＫ－１９（共通）

１．研究開始当初の背景

人間同士の自然な会話では現話者の発話終了から次話者の発話開始への話者移行時間の平均はおよそ 70ms と非常に短く、現話者の発話末に重複するように次話者が話し始めることも頻繁に生じる。これは話し手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示している。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 280-340ms であることを考えると、このような発声タイミングは話し手の発話終了の位置を予測していなければなしえない。すなわち、円滑な会話を実現するために、次話者となる聞き手は話し手の発話末を予測しつつ、話し出すタイミングをはかっているといえる。

音声認識技術や人工知能技術の発展により、日常生活のなかに置かれ音声対話により人々の支援を行う音声アシスタントサービスが一般に用いられるようになってきた。しかし、これらの音声対話システムのほとんどにおいて自然な話者交替を伴う円滑なコミュニケーションを図っているものはなく、利用者は相手がコンピュータであることを強く意識した発声を行い、注意深く返答を待つことになる。これは、従来の音声対話システムでは発話の終わりを無音区間の存在により検出しているためであり、音声認識や自然言語処理による発話理解が今後どれほど高速化しても、話者交替時に不自然な間が空くことは避けられない。また、従来研究には利用者の視線やジェスチャ等も含めたマルチモーダル情報によって話者交替箇所を推定する試みもあるが、円滑な話者交替が実現できているとは言い難い。さらに、スマートフォンやスピーカー型の音声アシスタントではそのようなマルチモーダル情報を利用することが困難である

一方、人間は視線やジェスチャがない状況でも自然な話者交替が可能である。インタラクション研究の一分野である会話分析では、発話末付近に次話者が割り込みにならずに話しだすことができる話者移行適格場が存在するとしている。しかし、人間がどのような情報を用いて話者移行適格場を認知しているかは明らかになっておらず、情報システムへの活用にも至っていない。

２．研究の目的

本研究の目的は、人間とコンピュータ対話システムの自然なコミュニケーションを将来実現するために、人間同士の話者交替に関する性質の解明を目指し、音声・言語の両面から人間に発話末を予測させる特徴について明らかにすることである。これまでインタラクション研究において、前述の話者移行適格場は会話分析専門家の直感によって認定されていた。この話者移行適格場を客観的な指標で検出できれば、発話中で話者交替が可能な箇所を人間の直感に頼らずに捉えることができる。また、ここで得られた音声・言語特徴を用いることで音声対話システムにおける円滑な話者交替の実現が期待できる。

３．研究の方法

(1)発話末の音響的特徴のひとつとして、平叙文末尾で基本周波数(F0)が局所的に下降して発話の終了を示す現象(Final lowering)が生じることが指摘されている（Pierrehumbert and Beckman, 1988)。また、自発性の独話を対象にした研究において Final lowering が発話の最終アクセント句末にのみ生じるのではなく最終アクセント句全体にわたって生じることが明らかにされている（前川, 2018)。また、我々も自然会話を対象としてアクセント句単位での発話内の韻律変化について調査し、話者交替が起こりうる発話の最終アクセント句で F0 が基底値に達する、モーラ時間長が発話末に向かって短くなっていき最終アクセント句で伸張する、パワーが最終アクセント句で急激に低下する、という現象が現れることを示した（Ishimoto et al., 2011)。すなわち、発話の最終アクセント句に顕著な韻律変化が生じており、その変化が発話末予測の手がかりとなっている可能性がある。そこで、Final lowering に着目し、F0 の下降の有無が発話末の認知に与える影響について知覚実験により調査する。

(2)発話末を表す統語的指標としては、これまでに助動詞や終助詞等からなる発話末要素の存在が挙げられている。しかし、自発発話では発話末要素が存在しない発話、例えば名詞や動詞が発話末となる発話が頻繁に生じるため、発話末要素の出現の検知だけでは発話末の判別には不十分である。そこで、発話末を投射する統語情報として発話末要素に加えて文節係り受け構造を取り入れることにする。さらに韻律情報として発話速度・F0・パワーを考慮し、これらの統語・韻律情報が組み合わさって発話中から発話末位置が投射されることによって発話末予測がなされると考えた。検証のため、統語・韻律情報から発話末の位置を発話末到来前に予測できるか発話末予測モデルを構築して調べる。

４．研究成果

(1)アクセント句の F0 下降が人間の発話末の検知に影響を与えるかどうかについて検証した。データとして千葉大学 3 人会話コーパスに収録されている自発会話音声を用い、①発話末の F0 が下降しない場合、②発話中の F0 が下降する場合の 2 種類の発話末知覚実験を行った。

① 実験刺激の基としてコーパス中の強い感情や意図が含まれていない発話のうち、A. 発話末要素あり、B. 発話末要素なしの発話を選んだ。さらに、発話末で F0 下降がない発話音声を音声分析合成システム STRAIGHT で作成した。具体的には、図 1 に表すように最終アクセント句について

1. 元音声の F0
2. 平坦な（下降しない）F0

を持つ音声刺激を用意した。

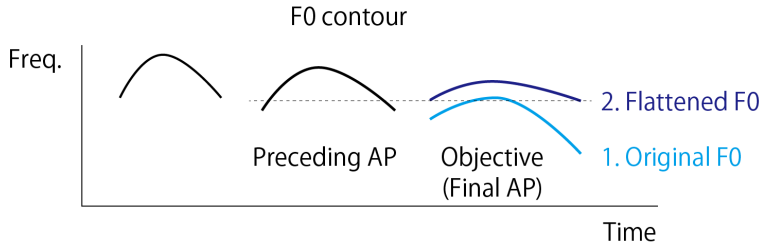

図 1 発話末のアクセント句に対する F0 操作の概要

被験者に対しこれらの音声を提示し、発話末であると感じた瞬間にボタンを押させる知覚実験を行った。音声刺激の終わりから被験者がボタンを押すまでの時間の平均を図 2 に示す。発話末要素が存在する音声の方が反応が遅くなる傾向はあるものの、F0 の操作による違いは見られなかった。

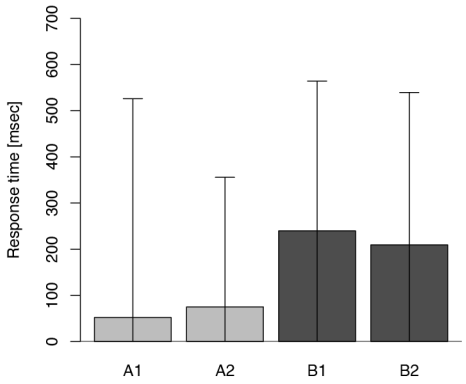

図 2 音声終了時刻からの反応時間

② 実験刺激の基としてコーパス中の強い感情や意図が含まれていない発話のうち、統語的に発話末を示さない C. 節単位弱境界、D. 節境界なし（格助詞）を含む発話を選び、それぞれ C, D の箇所以降を削除した音声を用意した。すなわち、これらの音声は発話の途中で途切れた状態になっている。さらに、その音声の最後のアクセント句に対して、図 3 に示すように

1. 元音声の F0
3. F0 下降を強調させた F0

を持つように操作を施した。

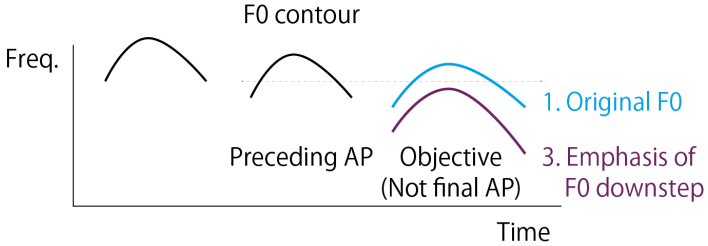

図 3 発話中のアクセント句に対する F0 操作の概要

作成した音声刺激に対して、①と同様の被験者に発話末であると感じた瞬間にボタンを押させる知覚実験を行った。音声終了時刻を基準とした反応時間の平均を図 4 に示す。多重比較検定の結果、C1 と D1、C3 と D1 の間に有意な差が見られた。すなわち、被験者は格助詞で音声が終わるよりも節単位弱境界で終わる方が発話の終了性を感じているといえ、発話末知覚には統語的な要素が関わっていることが見てとれる。一方で F0 下降を強調した C3 と D3 に関しては反応時間が遅れは見られなかった。これは、F0 の下降が発話末らしさに影響を与えていないことを示している。

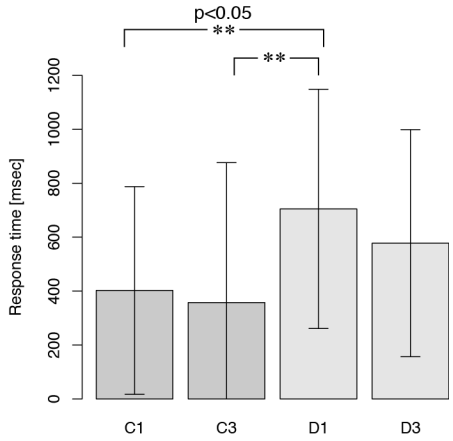

図 4 音声終了時刻からの反応時間

①,②の結果は、発話末の Final lowering は聞き手の発話末予測に利用されていないことを示唆している。F0 下降が発話末要素で顕著に現れることを考えると、単なる F0 下降が発話末を表すのではなく、発話末要素出現の結果として F0 が下降している可能性が考えられる。

(2) 上述の知覚実験結果より、人間は発話末予測において F0 下降のような韻律の変化を単に見ているのではなく、音響特徴と統語要素と組み合わせていることが考えられる。

そこで、まず統語情報として「係り先未定文節数」を基に、文節ごとに係り先未定文節数の差分を求め発話末への統語的近さの指標として扱うことにした。係り先未定文節数とその差分の概要を図 5 に示す。

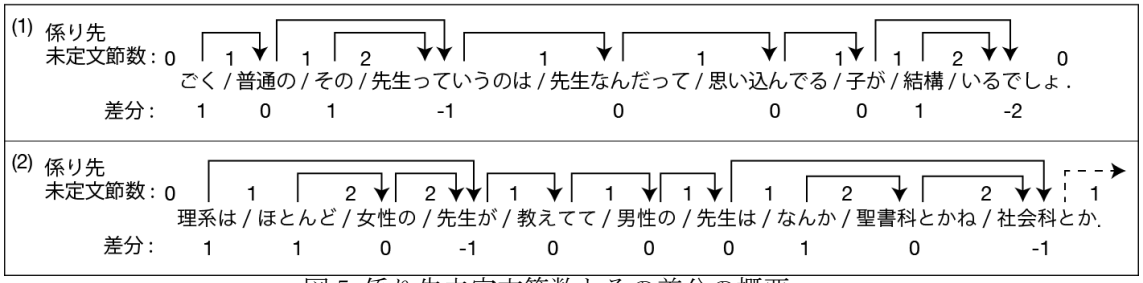

図 5 係り先未定文節数とその差分の概要

係り先未定文節数はある文節の直後において『いくつの文節の係り先がまだ生起していないか』を示しており、文の進行に伴い漸進的に決定される。例えば図 5 の例文(1)において、文節「その」の後の文節境界の時点では、「その」に加えて直前の「普通の」の係り先も未定であることから、係り先未定文節数は 2 となる。直後の「先生っていうのは」が出現することにより「普通の」および「その」の係り先が決定され、これらの係り先未定文節数 2 は解消される。ただし、「先生っていうのは」の係り先が新たに未決定になり係り先未定文節数は 1 となる。最終的に係り先文節が全て出現することで文末では係り先未定文節数が 0 になる。しかし、自発発話においては係り先が現れていない状態で発話が終了することもありうる。例文(2)において、最後の文節「社会科とか」では係り先が現れずに発話が終了しており、係り先未定文節数は発話末で 1 となっている。このことから、係り先未定文節数のみで発話末を決定することはできない。そこで、前後の文節境界について係り先未定文節数の差分を求め、発話末への統語的近さを示す指標として扱うこととする。これは、係り先未定文節数の差分値の減少は係り受けの解決を意味し、負値の時点で統語的な発話末らしさが現れていると考えられるためである。

韻律を表す音響特徴量として、文節全体を対象として平均 F0・平均パワー・平均モーラ長を求めた。各特徴量は性差および個人差の影響を減ずるために、各話者の発話全体の平均値と標準偏差を用いて Z-score に変換することで標準化した。また、F0 は事前に対数化を行なっている。

以上の統語情報および韻律情報から各文節が発話末付近であるかどうかを予測するロジスティック回帰の階層ベイズモデルを発話者ごとに構築した。モデルの説明変数は前述の統語情報 2 変数および韻律情報 3 変数とし、目的変数は発話末付近の文節（発話の最終文節およびその直前の文節）か否かを表す 2 値データとした。推定結果の 1 例として、ある発話者のデータにおける MCMC サンプルから各パラメータの事後分布と 95%ベイズ信頼区間を図 6 に示す。

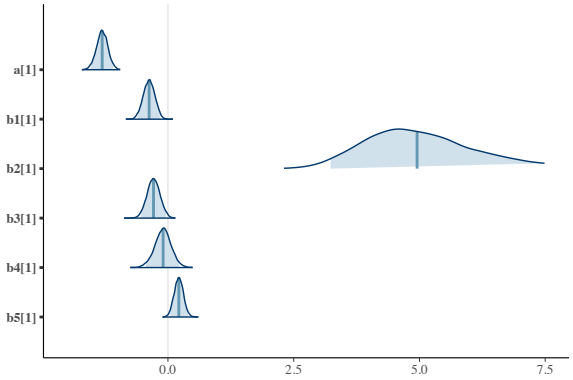

図 6 本モデルの各パラメータの事後分布と 95%ベイズ信頼区間（発話者 1 名分）

図から平均パワー（b_4）以外の係り先未定文節数の差分・発話末要素の有無・平均 F0・平均モーラ長といった統語・韻律情報が発話末付近の文節であるかどうかの判定に寄与していることがわかる。また、発話末要素の有無の係数（b_2）が他の係数に比べ大きく、発話末か否かの判定に大きな役割を担っていることが見てとれる。しかし、前述のように自発発話においては発話末要素が存在しない発話もあり、そのような発話においては今回取り上げた統語・韻律情報が発話末予測に役立つと考えられる。次に、本モデルの ROC 曲線を図 7 に示す。

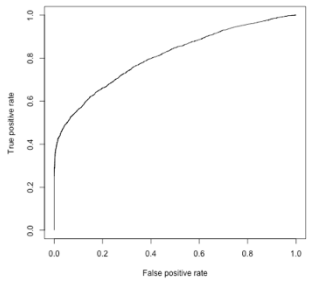

図 7 本モデルによる発話末予測における ROC 曲線

AUC の値は 0.807 であり、本モデルは比較的良い分類性能を示している。

以上の結果から、係り先未定文節数の差分・発話末要素の有無・平均 F0・平均モーラ長といった統語情報および韻律情報の組み合わせが発話末の文節およびその直前の文節の検出に寄与することがわかった。

５．主な発表論文等

〔学会発表〕（計 12 件）

① 石本祐一，寺岡丈博，榎本美香，“話者移行適格場予測のための発話内文節位置推定モデルの構築，” 日本音響学会 2019 年春季研究発表会，2019.

② Yuichi Ishimoto, Takehiro Teraoka, Mika Enomoto, “A Prediction Model for End-of-Utterance Based on Prosodic Features and Phrase-Dependency in Spontaneous Japanese,” APSIPA Annual Summit and Conference 2018, 2018.（査読あり）

③ Yuichi Ishimoto, Tomoko Ohsuga, “Spontaneous Speech Resources in Japan,” LREC 2018 Special Speech Sessions: Speech Resources Collection in Real-World, 2018.

④ 石本祐一，寺岡丈博，榎本美香，“言語情報と韻律情報に基づく自発発話終了位置の統計的予測モデルの構築，” 日本音響学会 2018 年春季研究発表会，2018.

⑤ Yuichi Ishimoto, Takehiro Teraoka, Mika Enomoto, “End-of-Utterance Prediction by Prosodic Features and Phrase-Dependency Structure in Spontaneous Japanese Speech,” Interspeech2017, 2017.（査読あり）

⑥ 石本祐一，寺岡丈博，榎本美香，“統語情報と韻律情報を用いた発話頭からの漸進的発話末予測の検討，” 日本音響学会 2017 年秋季研究発表会，2017.

⑦ 石本祐一，榎本美香，“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，” 日本認知科学会第 34 回大会，2017.

⑧ 石本祐一，寺岡丈博，榎本美香，“韻律情報と文節係り受け構造を用いた発話末予測モデルの構築，” 日本音響学会 2017 年春季研究発表会，2017.

⑨ Yuichi Ishimoto, Takehiro Teraoka, Mika Enomoto, “A Study on Prediction of End-of-Utterance by Prosodic Features and Phrase-Dependency Structure in Spontaneous Speech,” 5th Joint Meeting of the Acoustical Society of America and

the Acoustical Society of Japan, 2016.
⑩ Yuichi Ishimoto, Mika Enomoto, “Experimental Investigation of End-of-utterance Perception by Final Lowering in Spontaneous Japanese,” Oriental COCOSDA 2016, 2016.（査読あり）
⑪ 石本祐一, 寺岡丈博, 榎本美香, “韻律情報と文節係り受け構造を用いた発話末予測モデルの検討,” 日本音響学会 2016 年秋季研究発表会, 2016.
⑫ 石本祐一, 小磯花絵, “日本語話し言葉コーパスに基づく自発発話の継続・終了に関わる韻律情報の分析,” 日本音響学会 2016 年春季研究発表会, 2016.

６．研究組織

(1)研究分担者

研究分担者氏名：榎本 美香

ローマ字氏名：Mika Enomoto

所属研究機関名：東京工科大学

部局名：メディア学部

職名：講師

研究者番号（8 桁）：10454141

研究分担者氏名：寺岡 丈博

ローマ字氏名：Takehiro Teraoka

所属研究機関名：拓殖大学

部局名：工学部

職名：助教

研究者番号（8 桁）：30617329

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。