

漫才の動的構造の分析 – 間の合った発話タイミング制御を目指して –

川嶋 宏彰^{*1} スコギンズ リーバイ^{*1} 松山 隆司^{*1}

Analysis of the Dynamic Structure of Manzai
– Toward a Natural Utterance-Timing Control –

Hiroaki Kawashima,^{*1} Levi Scoggins^{*1} and Takashi Matsuyama^{*1}

Abstract – In human conversation, there exists non-verbal information that has a higher purpose than simply the expression of words; a purpose that includes the expression of desire or the creation of an exciting atmosphere. For example, in Japanese stand-up comedy, the timing of the punchline can completely change the impact of the routine on the audience. In other words, if an interactive speech system were to be able to dynamically control its responses in an autonomous way, it would be considered to be interacting naturally and proactively such as a human dialog. To acquire detailed understanding of what are the key features of deciding utterance timing, we analyze the dynamic structure of Japanese stand-up comedy, which is a dialog that the utterance timing becomes prominent, with respect to various contexts based on dialog acts.

Keywords : utterance-timing control, manzai dialog corpus, dialog act, speech dialog, backchannel and non-backchannel

1. 序論

人間同士の対話は、単に言語的な意味を伝えるだけでなく、意欲を高める、楽しい雰囲気を作るなどといった、より上位の目的・意図を持つことが多く、その伝達には非言語的な情報が大きな役割を担っている。例えば、人はその成長の過程で、抑揚や発話速度、発話開始のタイミングなどの非言語的なパターンに接しながら対話のプロトコルを学習しており、わずかなパターンの変化から相手の状態を推定するとともに、自分の意図を伝えるための適切な発話方法を決定している。

これに対し、従来の音声応答システムにおける音声認識や音声合成では、言語的な情報をいかに精度よく入力し提示するかという点に重点がおかれ、キーボードやディスプレイを音声に置き換える役割としては適している。ところが、ユーザの入力(コマンド)が初めて応答を返すというリアクティブなインタラクションになっているため、使用時にもどかしさを感じることがある。

近年、システム自身が自律的に人間の行動や意図を理解・推察し、適切な情報提示を自発的に行うといった、双方向かつプロアクティブ (proactive) なインタラクションシステムの設計が期待されている^[1]。音声対話システムにおいてもこのような設計は重要であり、人間同士が行っているような自然な対話を、人工的な

対話システムによって実現するためには、システムが、常にユーザの話し方から相手の状態や意図を推定し、ユーザの発話終了前、さらにはユーザの発話がない場合であっても、対話や発話の目的に応じて適切なタイミングで情報提示(発話や相槌動作など)を行っているのが望ましい。

本論文では、このような非言語的な対話のプロトコルとして重要な要素の1つである「間」の取り方に焦点を当てる。一方の発話終了時に対する他方の発話開始時までの間、すなわち「発話移行区間長」に関しては、どのような長さの間が生成されやすいかといった知見が得られており^{[2],[3]}、実際の対話システムにおける間合いを制御するという検討もなされている(詳細は2.章参照のこと)。しかしながら、多くは音声のピッチやパワーなどの韻律情報と、相槌や発話交替のタイミングについて分析しており、発話移行区間長が、文脈や発話内容とどのような関係にあるかについては、十分な知見が得られているとはいえない。これは、音声対話には、対話に参加する人数やその内容(議論、雑談、交渉、相談、協調など)、話者のスキルや性格によって様々なものがあり、統計的な分析がしばしば困難になることに原因があると考えられる。

そこで本論文では、発話された文の、語彙や文法に基づいた意味をそのまま用いるのではなく、対話における発話内行為(dialog act, DA)、すなわち発話の持つ目的・機能に注目した分類を行い、この発話内行為の種類と発話移行区間長との関係に注目する。これ

*1: 京都大学情報学研究科

*1: Graduate School of Informatics, Kyoto University

によって、

- 発話された語彙の情報だけでは読み取れないような発話の役割を明確にすることができ、
- 発話内容のバラエティを吸収しながら、統計的な解析が可能になる

という利点がある。さらに、分析対象として漫才を用いる。これは、漫才が

- 発話タイミングの影響が顕著に現れる
- 話者の役柄が明確である（ボケおよびツッコミ）
- あらかじめシナリオが決まっている

といった、発話移行区間長の分析にとって適した特徴を持つからである。特に漫才では、言語的には同じ発話内容であっても、間の取り方や発話速度が異なれば、聞き手へ伝わる意味合いが大きく変化するため、最適な間合いの取り方というコツが凝縮されていると期待できる¹。また、発話速度やテンポは比較的速いものの、言語的には通常の雑談に近いいため、より一般的な対話との関係が深いと考えられる。

本論文では、この分析のために漫才の対話コーパスを整備し、この分析を通じて、今後プロアクティブな対話システムを設計し評価する上での有用な知見を得ることを目的とする。

まず、次章では発話移行区間に関する関連研究について紹介する。3. 章では、漫才を分析するために作成した漫才対話コーパスについて述べる。作成した漫才対話コーパスを用いた分析結果について 4. 章で示し、そこから得られる知見について 5. 章で論じる。

2. 関連研究と本研究の位置づけ

実際の対話における間の取り方やその影響を分析した研究として、市川らは、実際の自由対話において、発話権の維持を示す発話や、キーワードとなる単語が現れてから約 400msec 以内で相槌が打たれることが多いという結果を示した^[2]。一方、長岡らは、テレフォンショッピングにおけるオペレータと客の間の対話において、両者の発話移行区間長（交替潜時）の長さの組み合わせを 6 通りに操作して（第三者である）被験者に提示し、発話移行区間長が、オペレータの接客態度の印象に影響を与えていることを示した^[3]。

相槌などを適切に制御するための実際のシステムとして、岡登らは、会話全体の韻律情報を解析することによって、ピッチパターンのテンプレートを作成し、ユーザの発話終了時に対する相槌挿入のタイミングを、テンプレートマッチングによって推定する方法を提案し

た^[4]。北岡らは、ユーザの発話が終了したあとに相槌を入れるシステムを、決定木を用いて開発しており、ユーザの発話長や最後の品詞、発話区間末の F0 やパワーに基づいて、発話後の各フレーム（100msec 間隔）で相槌を入れるか否かを決定している^[5]。

一方で、人間同士の対話では、ポーズを常にとった対話が行われることはまれであり、相手の発話終了を待たずにオーバーラップするような発話も多く見られる。藤江らは、このようなオーバーラップを伴う相槌発話や復唱を行うロボットを開発しており、発話のタイミングは、F0 やパワーのような韻律情報の変化に基づいて生成している^[6]。

これら多くの従来研究では、発話に対する相槌に注目して、そのタイミングの生成方法を検討している。しかし、ポーズやオーバーラップの度合いは、発話目的（相槌、質問、意見 など）や、発話の相手や第三者にどのような印象を与えたいかといった意図に強く依存していると考えられる。

本論文では、オーバーラップを含んだ発話の移行を、1 章で述べた DA による分類を用いながら分析することで、発話の目的や文脈によって、発話のタイミングがどのように影響を受けるかを分析する。発話のタイミングは、実際に相手や第三者に伝達される意図や印象を大きく変化させる要因になることから、(1) 人間の発話からその意図を精度よく推定するシステムや、逆に (2) 警告や説得といった目的・意図をより効果的にユーザに伝える対話システムの実現につながると期待できる。

3. 漫才対話コーパスの作成

漫才とは「二人で滑稽な問答を中心に演じる寄席演芸」（大辞林 第二版より）であり、ボケ役およびツッコミ役の対話によって進められる。ただし、漫才師が対話を進める上では、客の反応（笑い声など）も非常に重要であるため、ボケ役とツッコミ役の発話および客の反応から構成されるやり取りを漫才対話と呼ぶことにする。

漫才対話を分析するに当たり、大量の対話コーパスが必要になる。そこで、本研究では市販されているビデオおよび放送映像を用いて漫才対話コーパスを作成した。最近では、従来の漫才の枠組みに当てはまらない新しい形の漫才も認知されているが、本論文で対象とする漫才師は、その対話様式の統制のために、若手の実力派（20 代から 30 代前半）の中からボケ役とツッコミ役の役割分担がはっきりしているものを選んだ。

以下、3.1 節で漫才対話の参加者が持つ役割について説明し、3.2 節で漫才の台本（ネタ）について述べる。3.3 節では、漫才対話における発話区間および発

1: この「聞き手」とは、話者に対するもう一方の漫才師（発話相手）および客である。漫才は客に聞かせる演芸であるため、漫才師の発話は、発話相手よりもむしろ客に伝わる印象を意識して間合いを制御しているとも考えられる。ただし本論文では、誰を意識した発話かによらず、発話の目的・機能と間合いの間には普遍的な構造があると仮定する。

話移行区間の構造を説明し、最後に 3.4 節で漫才対話コーパスの作成方法について述べる。

3.1 漫才対話の参加者の役割

本研究では、ボケ役、ツッコミ役および客が、対話に対して以下のような役割を持つと考える。

ボケ役: 冗談を言い出す担当であり、個々の発話の言語的（語彙や文法的）意味が重要になる。

ツッコミ役: ボケ役の冗談に対して毒舌なコメント（いわゆるツッコミ）を入れる担当であり、言語的意味よりも、むしろ発話のタイミングや抑揚をはじめとする非言語的情報が重要になる。

客: ボケ役とツッコミ役の対話が滑稽である場合に笑いや拍手で反応を行い、漫才の評価およびツッコミ役との同調を担当している。

3.2 漫才のネタとサブネタ

1つの対話の台本をネタと呼ぶことにすれば、ネタは実際の寄席に先立ってあらかじめ作成されることが多く、通常1つの話題（「怖い」など）を設定する。さらに、漫才のネタの中には、共通の話題を発展させた短い話題（「怪談」「怖い番組」など）が存在する。その短い話題をここではサブネタと呼ぶことにする。いくつかのサブネタをつなげることで1つのネタになる。

漫才の典型的な対話の流れは以下の通りである。まず、漫才師の2人が現れ、漫才師のコンビ名を紹介する。続いて、客の雰囲気を知るため、最近の話題に関し短いやり取りをし、その後、サブネタを複数披露する。最後のサブネタを、ネタ全体のオチを含めて落とすことで演技が終了する。

個々のサブネタにおいては、ボケ役の発話に対し、ツッコミ役が質問やツッコミを入れる。このやり取りを繰り返しながら対話を盛り上げ、通常1,2分程度で次のサブネタに移る。次のサブネタに移る際には話題が切り替わるが、漫才師やネタによって、ボケ役が主導する場合やツッコミ役が主導する場合がある。本論文で用いた対話は、ツッコミ役がサブネタの話題を導入するケースが多くみられた。

3.3 発話区間と発話移行区間

3.1 節で述べたように、漫才対話には、ボケ役、ツッコミ役、客が参加している。本論文では、このうちボケ役とツッコミ役の発話について分析を行うものとする²。

ボケ役とツッコミ役の発話を時間軸上に発話単位で並べると図1のようになる。発話 B と発話 T はそれぞれボケ役とツッコミ役の発話である。ボケ役の発話

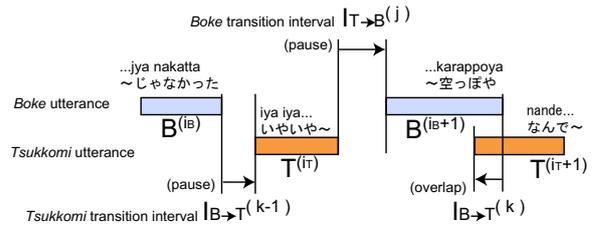


図1 漫才の発話区間の構造
Fig. 1 Structure of manzai utterances.

B の終了時刻からツッコミ役の発話 T の開始時刻までの間を発話移行区間 $I_{B \rightarrow T}^{(k)}$ とする。添字 k は、最初の発話ペアから数えて k 番目のペアであることを表す。同様に、ツッコミ役の発話 T の終了時刻からボケ役の発話 B の開始時刻までの発話移行区間 $I_{T \rightarrow B}^{(j)}$ を考えることができる。通常、 k と j は対話が進むにつれて交互にインクリメントされていくが、両者の発話開始が同期するなどの状況によって、添え字 k と j がずれてくるため、ここでは添え字を分ける。なお、何番目の発話ペアであるかが重要でない場合、添え字を省略して、 $I_{B \rightarrow T}$ や $I_{T \rightarrow B}$ のように表すものとする。また、区間の長さを表す際も、同じ記号を用いるものとする。この発話移行区間長が正であれば、直前の相手の発話に対して「ポーズ」をとった発話（図1左側の発話ペア参照）となり、逆に負もしくは0であれば「オーラップ」した発話（図1右側の発話ペア参照）となる。

3.4 漫才対話コーパスの作成

3.4.1 発話区間のセグメンテーション

市販のビデオなどにおける漫才の音声データでは、ボケ役とツッコミ役の発話が多重的に重なる場合が多く、信号処理による自動的なセグメンテーションが困難である。そこで、アノテーションソフト Praat³ を用いて、音声ファイルに含まれる発話区間のセグメンテーションを手で行った。これは、Praat によって音声波形やパワー変化を画面に表示させるとともに、セグメンテーション作業者が、音声を繰り返し聞きながら、発話の始点・終点を画面上で指定することで行われた。セグメンテーションを行った対話の数は5つであり、作業は3人によって分担して行われた。また、セグメンテーションの個人差を調べるために、このうち1つの対話に関しては2人で行った。

発話区間のセグメンテーションは、以下のような基準でボケ役、ツッコミ役、および客ごとに行った。

ボケ・ツッコミ役: 発話区間と無音区間に分ける。

客: 笑い区間と無音区間に分ける。

さらに、セグメンテーションによって得られたボケ役

2: 客の役割については、笑うタイミングとツッコミ役の発話開始との同期に関して興味深いパターンが見出されており、客は単に第三者的立場にあるというよりは、積極的に対話に参加していると考えられることもできるが、本稿では単純化のため扱わないものとする。

3: <http://praat.org>

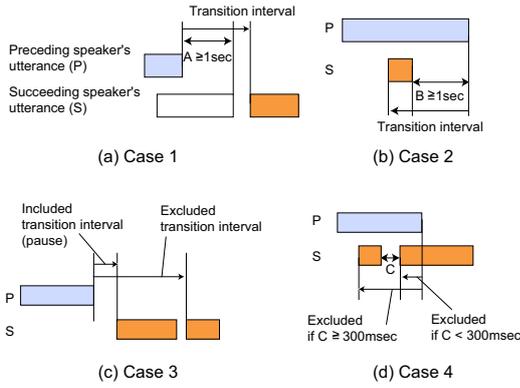


図2 発話移行区間から除外したケース
Fig.2 Excluded cases of transition intervals.

およびツッコミ役の発話に対しては、その発話内容の書きおこしをローマ字によって行なった。

3.4.2 発話移行区間のラベリング

セグメンテーションされたボケ役およびツッコミ役の発話区間を用いて、両者の間での発話移行区間 $I_{B \rightarrow T}$ と $I_{T \rightarrow B}$ をラベリングした。ラベリングの際の発話ペアの決定方法について説明するために、まず、ボケ役の発話を $B^{(i_B)}$ 、その始点と終点をそれぞれ $b_B^{(i_B)}$, $e_B^{(i_B)}$ と表す。同様に、ツッコミ役の発話 $T^{(i_T)}$ の始点と終点をそれぞれ $b_T^{(i_T)}$, $e_T^{(i_T)}$ のように表す。ここで、 i_B および i_T は、ボケ役、ツッコミ役それぞれの発話区間が、対話の始まりから何番目に出現したかを表す。

発話移行区間 $I_{B \rightarrow T}$ のラベリングの際には発話ペアを次のように決める。各ツッコミ役の発話 $T^{(i_T)}$ に関して、 i_B を対話の始めからインクリメントしながらはじめに $e_B^{(i_B)} \geq b_T^{(i_T)}$ を満たすボケ役発話 $B^{(i_B)}$ を見つけ、これがさらに $b_B^{(i_B)} < b_T^{(i_T)}$ を満たす場合は、発話 $T^{(i_T)}$ と $B^{(i_B)}$ を発話ペアとし、満たさない場合は、発話 $T^{(i_T)}$ と $B^{(i_B-1)}$ を発話ペアとする。このとき、それぞれの発話ペアは、オーバーラップ ($I_{B \rightarrow T} \leq 0$)、ポーズ ($I_{B \rightarrow T} > 0$) の発話移行区間長を持つ。発話移行区間 $I_{T \rightarrow B}$ でも同様であり、上記の説明において B と T を入れ替えることになる。

ただし、発話終了ではなく発話の中盤で相手が発話(相槌など)を入れる場合がある。これは図2のCase 1,2において、それぞれ区間長 A, B が大きい場合である。本論文で解析対象とするのは、一方の発話終了に対するもう一方の発話開始であるため、経験的に、 A, B それぞれ 1sec 以上の場合には発話ペアから除外した。また、Case 3,4 は1つの先行発話に対して2つ以上の後続発話に対応する場合であるが、Case3のように一番目の後続発話がポーズならば、二番目以降の後続発話は発話ペアから除外した。一方Case4は一番目の後続発話がオーバーラップする場合である。もし後続発話

同士が非常に近い(経験的に区間長 C が 300msec 未満とした)ならば、後続発話同士が意味的につながっているとみなして二番目以降の後続発話を発話ペアから除外し、逆に後続発話同士が離れているならば、1つ目の後続発話を除外した。以上の処理の後、発話の対応が意味的に妥当であることを手動で確認した。

3.4.3 発語内行為のタグ付与

対話の持つ構造の解析を行うために、発話された言葉の役割(陳述や警告など)に基づいてタグを振ることがしばしば行われる。Austin は、発話も1つの行為の遂行であると考え、発話の持つ意味を発語行為、発語内行為、発語媒介行為の3つの段階に分け、発語内行為を分析することの重要性を述べた[7]。ここで発語行為とは、言及対象を伴い文法的に正しい言葉を発する行為、発語内行為は、警告や約束など、その言葉を発することの目的や機能に注目した行為である。またその意味は、話し手や聞き手の持つ事前知識や文脈に応じて慣習的規則が働いて定まるとしている。さらに、発語媒介行為は、聞き手に結果的に現れる反応や影響(驚かすなど)に注目した行為であるとしている。

この発語内行為は、Searl らによって発話行為(speech act)の理論として発展し[8]、発話を行為としてとらえる分析が多く行われている。複数人の対話については、相槌などの対話独特の行為も含めて dialog act (DA) という呼称が用いられる。多くの対話に関する研究では、対話の種類(雑談、相談、共同作業)などに応じて、発語内行為の種類を分類するためのタグ集合を定義し、対話のアノテーションや分析を行うといった方法がとられる。

DA のタグは、Allen と Core のマニュアル[9]に基づいて定義されることが多く、Jurafsky ら[10]はこれを拡張することで42個のタグを定義している。会議の対話では文献[11]のマニュアルなどが用いられる。本論文でも、他のDAの研究で得られている知見との比較や参照を容易にするために、基本的には文献[10]~[12]を参考にする。ただし、以下で述べるように、漫才対話に頻出するタグに絞ると共に、漫才に特有のタグを追加する。本論文で用いるDAを表1に示す。

漫才では相槌だけでなく「なんでやねん」「どないやねん」などの質問、もしくは意見に近い意味を持った発話を、強い語意を伴って行う場合がある。これらは、しばしば発話された文の意味よりも、対話相手(通常ボケ役)の冗談的な要素を持つ発話を受け止め、対話を安定化させる役割を持つため、本研究では tsukkomi というDAを定めた(使用例を表2に示す)。

対話の発話権を持つ人の発話をノンバックチャンネルの発話、発話権を持っていない人の発話をバックチャンネルの発話と定義する。DAは大きくはこのバック

表 1 本論文で用いる Dialog act のタグ
Table 1 Dialog act tags.

Dialog act tag (省略)	機能
statement (st)	陳述する
opinion (op)	意見を述べる
question (qs)	質問する
command (cm)	依頼・命令する
continuer (cn)	相槌を入れる(「うん」など)
tsukkomi (ts)	冗談を受け止める(質問や意見に近い強い語気)
backchannel-question (bq)	発話権が移行しない質問の相槌を入れる(「ほんまに?」など)
incipient (in)	話題を転換し発話権を取得する(「あのさあ」「ええと」など)
agreement (ag)	陳述や意見に対し同意する
disagreement (dg)	陳述や意見に対し拒否・反駁する
answer (an)	質問に対する回答を行う

表 2 *tsukkomi* Dialog act の例
Table 2 Example of *tsukkomi* dialog act.

話者:Dialog act	発話
B: statement	風呂場帰ってきたらもう (after leaving the bath)
B: statement	… バスタオル持ってるって (... already had the towel)
T: tsukkomi	どないやねん! (how is that!?)

クチャンネルとノンバックチャンネルに分けることができ、表 1 のうち、statement, opinion, question, command はノンバックチャンネル発話に、continuer, tsukkomi, backchannel-question, incipient はバックチャンネル発話に、残りはいずれの発話にも現れるものとする。このうち incipient は、現在の話題を変えて自らが発話権を得るための「あのさあ」「ほんならあの」などから始まる発話であり、発話を行う時点では発話権を譲渡されていないとしてノンバックチャンネルに分類した。

以上のタグを、3.4.1 節のセグメンテーションによって得られた、ボケ役およびツッコミ役の各発話区間に人手で付与した。いずれの DA にも分類できない擬音語や喚声などはタグ xx を付与した。4. 章以降では、この DA のタグを用いて分類を行いながら、発話移行区間長との関係について分析を行う。

3.5 コーパス作成者が異なる場合のセグメンテーションの時間差評価

セグメンテーションの作業者によって、実際には発話移行区間の区切り位置に時間差が生じる。この時間差がどの程度の分布をもつかについて評価を行った。2 人の作成者に、同じ漫才師の約 5 分間の対話を 3.4 節で述べた手法でセグメンテーションしてもらった。

セグメンテーションされた発話区間の開始時間の差については、平均 62msec (標準偏差 56msec)、終了時間の差については平均 122msec (標準偏差 127msec)



図 3 サーストンの一対比較法による主観的な発話移行区間長(ここではポーズ)の尺度化
Fig.3 Scaling result of subjective transition interval (pause) using Thurstone's method of paired comparisons.

であった。個々の発話における音声強度は、開始時の立ち上がりで鋭く、それに対し終了時の下り方は比較的緩やかになるという特徴がある。したがって、発話終了時のセグメンテーションには、発話の開始時に比べ、作業者間のばらつきが大きくなると考えられる。

次に、人間がどれだけの発話移行区間長を聞き取れるかという時間解像度を測定するために、10 人の評定者に対するサーストンの一対比較法^[13]によって、人間が主観的に弁別可能な、発話移行区間長 $I_{B \rightarrow T}$ (この実験では正、すなわちポーズのみ) の尺度を構成した。発話移行区間長としては、7 種類の異なる長さ (10, 20, 40, 80, 160, 320, 640msec) を用意し、この中からペアを作成して評定者にランダムに提示した。提示する音声は、漫才対話コーパス中から切り出した音声から作成した発話対であり、ボケ役の冗談に対するツッコミ役の tsukkomi DA の発話とした。

構成された尺度を図 3 に示す。この図より、10msec から 80msec までの距離は互いに近く、これ以上ポーズが長くなると、距離が大きくなっていることが分かる。したがって、80msec までのポーズの大小関係は人間にとって聞き分けにくい、80msec を超えた発話移行区間長になると、その聞き分けが容易になるといえる。

4. コーパスに基づく漫才対話の動的構造の分析

発話移行区間長は、発話内容や相手の直前の発話、連続する発話に現れるリズムなど、様々な要因に影響される。そこで本章では、漫才の持つ動的構造を、以下の視点に分けて分析する(図 4 参照のこと)。

1. 発話移行区間長の 1 次元分布: 対話全体における発話移行区間長の分布を、発話者の役割(ボケやツッコミ役)、DA タグに分けて調べることで、発話の目的・機能が、間のとり方に与える影響について分析する(図 4(a), 4.1 節)。
2. 相手の発話移行区間長との関係: 発話の短期的な文脈を考慮すれば、一方の間の取り方が、他方の間の取り方に何らかの影響を与える可能性がある。そこで、ある発話移行区間長と直前の相手の発話移行区間長の相関関係について分析を行う(図 4(b), 4.2 節)。

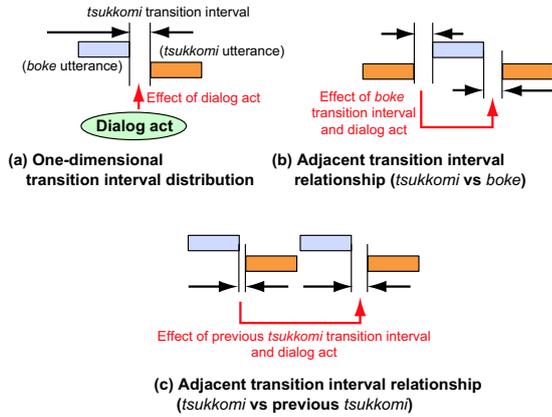


図 4 4. 章の分析の流れ
Fig. 4 Flow of analyses in Chapter 4..

表 3 漫才対話コーパスのネター一覧
Table 3 Netas in manzai dialog corpus.

対話	長さ [sec]	ボケ役発話数	ツッコミ役発話数
a1	267	92	133
b1	295	162	145
c1	323	158	155
c2	306	175	183
d1	458	131	174

3. 自身の連続する発話移行区間に現れる特徴:
対話のリズムやペースなどが存在すれば、同じ発話者の連続する発話移行区間の間には何らかの特徴的な構造があると考えられる。そこで、隣接する発話移行区間長の相関関係について分析を行う(図 4(c), 4.3 節, 4.4 節)。

以下, 3. 章で作成した漫才コーパスに含まれる, 5 つの対話を利用して分析を行う。各対話を区別するための記号, 長さ, 発話数を表 3 に示す。c1 と c2 については同じ漫才師であるが, 異なるネタを演じた対話である。

4.1 発話移行区間長の 1 次元分布の分析

4.1.1 役柄の違いが与える影響

発話におけるオーバーラップの割合がどの程度であるかについて, まずは役柄ごとの発話移行区間長の 1 次元分布の特徴を調べる。

図 5(a) はボケ役が発話を開始する際の発話移行区間長 $I_{T \rightarrow B}$, 図 5(b) はツッコミ役が発話を開始する際の発話移行区間長 $I_{B \rightarrow T}$ の頻度分布であり, コーパス中の全ての発話ペア(ただし DA_{xx} が関与するものを除く)を用いた。図より, ボケ役の発話移行区間長の分布はほぼ 0msec を中心として左右対称に広がっているが, ツッコミ役では負の領域に分布が偏っていることが特徴的な違いとして見える。

各対話におけるオーバーラップ発話の割合を表 4 に示す。対話 c2 を除く全ての対話で, ツッコミ役のオーバーラップの割合は, ボケ役よりも大きいことが分かる。

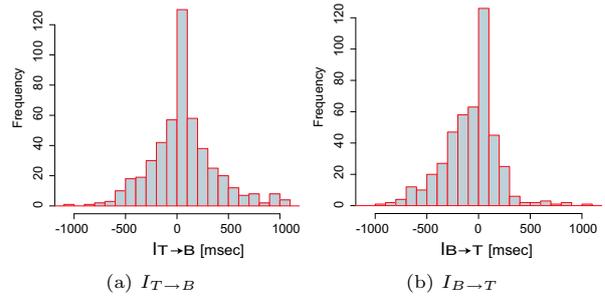


図 5 発話移行区間長 $I_{T \rightarrow B}$, $I_{B \rightarrow T}$ の頻度分布
Fig. 5 Histogram of transition interval $I_{T \rightarrow B}$ and $I_{B \rightarrow T}$.

表 4 各役柄のオーバーラップ発話の割合と発話移行区間長の中央値, および各役柄の中央値の差(ボケ役 - ツッコミ役) [msec]

Table 4 Ratio of pause utterance and median of transition interval for each role, and difference of the median values from tsukkomi to boke role [msec].

対話	$I_{T \rightarrow B} \leq 0$ の割合	$I_{B \rightarrow T} \leq 0$ の割合	中央値の差
	$I_{T \rightarrow B}$ の中央値	$I_{B \rightarrow T}$ の中央値	
a1	28% (17/60) 66	65% (40/62) -229	295
b1	13% (13/101) 130	39% (39/99) 27	103
c1	43% (51/118) 11	53% (62/116) -51	62
c2	51% (52/102) -24	46% (48/104) 21	-45
d1	51% (40/78) 0	75% (55/73) -113	113

さらに, 発話移行区間長の中央値の差に注目すれば, c1, c2 を除いた対話では, ツッコミ役はボケ役よりも 100msec 以上発話移行区間長が負に偏っていることが分かる。これは, 3.1 節で述べたようなボケ役とツッコミ役の役割の違いに起因していると考えられ, 実際, ツッコミ役ではボケ役の発話途中に割り込むような発話が多い。一方, 対話 c1 と c2 は同じ漫才師によるものであるが, ボケ役とツッコミ役それぞれの発話移行区間長の中央値の差は, いずれの対話も 65msec 未満であり, この漫才師のコンビでは他のコンビほど, ボケ役とツッコミ役での間の取り方の差が顕著ではないといえる。

本論文では, オーバーラップを含んだ発話の移行を, DA の分類によって調べることが主な目的であるため, 以下の節では, ツッコミ役が決定するような発話移行区間 $I_{B \rightarrow T}$ に焦点を当てて解析を行う。また, 図 5 から分かるように, 発話移行区間長の絶対値はほぼ 1sec 未満に分布する。そこで以下の解析では, 外れ値による影響を避けるために, 絶対値が 1sec 以上の発話移行区間長は解析対象から除外した(1 発話ペアが該当)。

4.1.2 Dialog act の違いによる影響

発話の持つ目的や機能と発話移行区間長の関係を調べるために、ツッコミ役の発話が各 DA (表 1 参照) を持つ場合の、発話移行区間長 $I_{B \rightarrow T}$ の特徴を調べる。ここでは 5 対話に現れる発話ペアを混ぜて用いる。

表 5 (最右列除く) は、ツッコミ役が各 DA の発話を行う際の、出現頻度、発話移行区間長 $I_{B \rightarrow T}$ の中央値、およびオーバーラップ数を示したものである。表より、continuer, tsukkomi, incipient などのバックチャネル DA は、他の DA に比べてオーバーラップの傾向が強いことが分かる。一方で、backchannel-question については、バックチャネル DA であるにもかかわらずポーズになる傾向がある。また、ノンバックチャネル DA の question に関しても、他の DA と比べてポーズの割合が多い。したがって、質問に関する DA はポーズに偏る傾向があるといえる。さらに、相手に同意と反駁のいずれを行うかでも特徴的な違いがあることが、agreement の場合は非常にオーバーラップが多く起こるのに対し、disagreement ではポーズとオーバーラップの割合が等しいことから分かる。

頻度の傾向をより詳細に調べるために、DA 毎の発話移行区間長 $I_{B \rightarrow T}$ の頻度分布を図 6 に示す。図より、statement, opinion, continuer, tsukkomi などの多くの DA では 0msec 付近の頻度が高いことが分かる。そこで、 $I_{B \rightarrow T}$ の絶対値が 80msec 以下の発話数とその割合を、表 5 最右列に示す (閾値は 3.5 節の結果を利用)。表より、特に incipient や question では 0msec 付近の発話の割合が低いが、incipient ではオーバーラップ傾向が、question では逆に数 100msec のポーズをとる傾向があることが、それぞれの中央値から分かる。一方、agreement, disagreement ではそれぞれ負、正に 100 から 200msec 外れてピークが見られた。

表 5 DA 毎の発話移行区間長 $I_{B \rightarrow T}$ の中央値、オーバーラップ ($I_{B \rightarrow T} \leq 0$) の割合、0msec 付近 ($|I_{B \rightarrow T}| \leq 80$ msec) の割合

Table 5 Median, the ratio of overlap, and the ratio of nearly 0msec of transition interval $I_{B \rightarrow T}$ for each DA.

DA tag	出現頻度	中央値 [msec]	オーバーラップ数	0msec 付近の数
st	64 (14%)	42	23 (36%)	29 (45%)
op	30 (6.6%)	55	8 (27%)	11 (37%)
qs	10 (2.2%)	118	1 (10%)	2 (20%)
cm	12 (2.6%)	-70	7 (58%)	5 (42%)
cn	51 (11%)	-59	27 (53%)	26 (51%)
ts	175 (39%)	-111	117 (67%)	60 (34%)
bq	19 (4.2%)	66	4 (21%)	6 (32%)
in	11 (2.4%)	-222	9 (82%)	2 (18%)
ag	36 (7.9%)	-154	28 (78%)	28 (22%)
dg	28 (6.2%)	9	14 (50%)	21 (25%)
an	17 (3.8%)	28	6 (35%)	11 (35%)

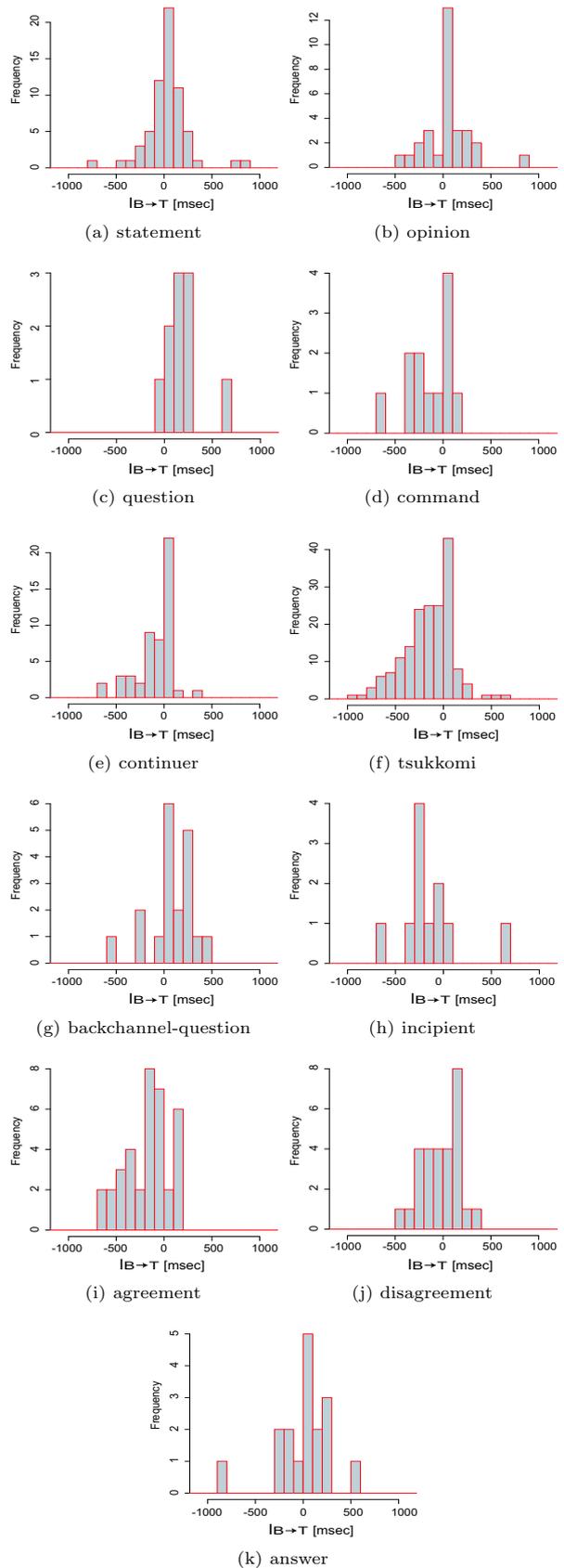


図 6 DA 毎の発話移行区間長 $I_{B \rightarrow T}$ の頻度分布
Fig. 6 Histogram of $I_{B \rightarrow T}$ for each DA.

4.2 ツッコミ役の発話移行区間長と直前のボケ役の発話移行区間長との関係

ツッコミ役の発話移行区間と直前のボケ役の発話移行区間の関係について調べる。

表 6 は、隣接する発話移行区間長 $I_{T \rightarrow B}$ と $I_{B \rightarrow T}$ の相関、および帰無仮説「相関係数=0」に対する検定（無相関検定）の有意確率（0.10 以下を太字で示す）であるが、DA によって傾向が分かれた⁴。ツッコミ役の発話が incipient である場合、高い相関（相関係数は 0.77）があることが分かる。一方、continuer, answer の場合でも、残りの DA に比べるとやや相関が認められた（相関係数はそれぞれ 0.38, 0.45）。なお、tsukkomi は他の DA と比べてサンプル数が大きく異なることから比較対象外とした。

以上より、incipient, continuer, answer では、直前の相手の発話に何らかの影響を受けている可能性がある。参考までに、相関係数の絶対値が高く、かつ検定結果が低い有意確率であった incipient および continuer について、 $I_{B \rightarrow T}$ と直前の $I_{T \rightarrow B}$ 散布図を、図 7(a), 7(b) にそれぞれ示す。

一方で、statement をはじめとするノンバックチャネル DA ではほとんど相関が見られない（図 8(a) に statement の散布図を示す）。ただし、どのような DA を持ったボケ役の発話に対して、ツッコミ役が発話を行うかを見ることで、特徴的な傾向が得られる場合がある。表 7 は、ボケ役とツッコミ役の隣接発話対に現れる DA 組のうち、出現頻度の高い上位 3 組の相

表 6 隣接する発話移行区間長 $I_{T \rightarrow B}$ と $I_{B \rightarrow T}$ の相関分析結果（ボケ役に対するツッコミ役の発話の DA 毎）

Table 6 The result of correlation analyses between adjacent utterance transition intervals $I_{B \rightarrow T}$ and $I_{T \rightarrow B}$ (for each DA of the tsukkomi utterance in response to the boke utterance).

DA tag	出現頻度	相関係数	有意確率
st	62 (15%)	0.14	0.263
op	28 (6.8%)	0.21	0.288
qs	9 (2.2%)	-0.18	0.635
cm	9 (2.2%)	-0.03	0.937
cn	48 (12%)	0.38	0.008
ts	152 (37%)	0.23	0.004
bq	16 (3.9%)	0.08	0.765
in	11 (2.7%)	0.77	0.006
ag	32 (7.8%)	0.04	0.808
dg	27 (6.6%)	-0.07	0.714
an	16 (3.9%)	0.45	0.079

4: 出現頻度が表 5 と若干異なるのは、解析対象外としている分類不可の発話（タグ xx）や長さの絶対値が 1sec 以上の発話移行区間の存在が、連続するどの発話移行区間対を有効とするかに影響を与えるためである（後出の表 8 でも同様）。

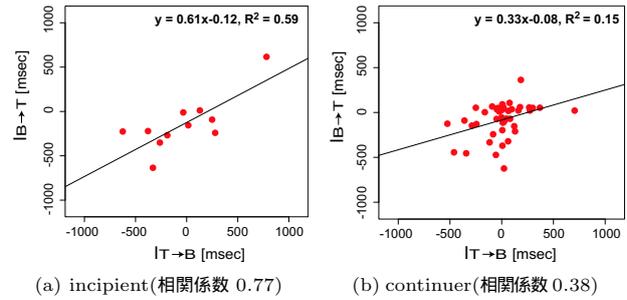


図 7 $I_{B \rightarrow T}$ 対 $I_{T \rightarrow B}$ の散布図および回帰直線
Fig. 7 Scatter plot of $I_{B \rightarrow T}$ v.s. $I_{T \rightarrow B}$ with the regression lines.

表 7 ボケ役とツッコミ役の隣接発話対に現れる DA 組と発話移行区間長の相関（出現頻度の上位 3 組）

Table 7 DA pairs appeared in adjacent boke and tsukkomi utterances, and correlation of those transition intervals (top three in respect to the appeared number).

$B \rightarrow T$	出現頻度	相関係数	有意確率
st \rightarrow ts	120 (29%)	0.17	0.065
st \rightarrow cn	32 (7.8%)	0.25	0.164
cn \rightarrow st	22 (5.4%)	-0.49	0.021

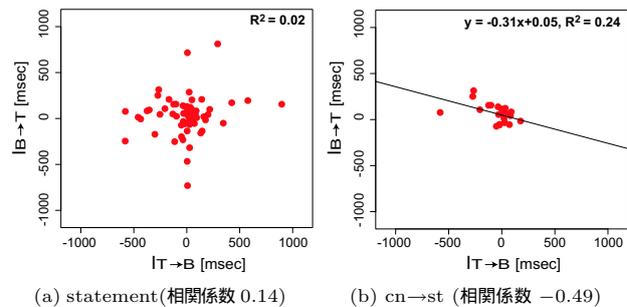


図 8 $I_{B \rightarrow T}$ 対 $I_{T \rightarrow B}$ の散布図および回帰直線
Fig. 8 Scatter plot of $I_{B \rightarrow T}$ v.s. $I_{T \rightarrow B}$ with the regression line.

関、および無相関検定の有意確率を示したものである。表 7 のうち、上位 2 つの組に関しては目立った傾向が見られない。一方で、ツッコミ役の DA が statement であるときに、直前のボケ役の DA を continuer に特定した場合は負の相関（-0.49）が見られた。図 8(b) はその散布図であり、図 8(a) の一部を取り出したものになっている。

4.3 ツッコミ役の発話移行区間長と 1 つ前の自身の発話移行区間長との関係

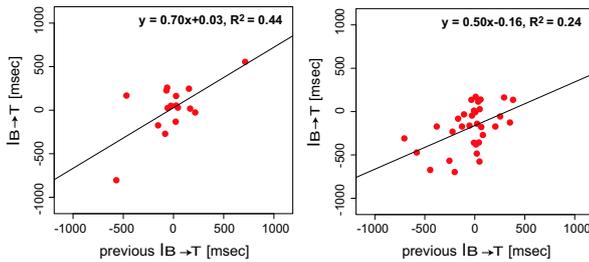
ツッコミ役の発話移行区間長と、1 つ前のツッコミ役の自身の発話移行区間長との関係について検討する。

表 8 は、隣接する発話移行区間長 $I_{B \rightarrow T}^{(k)}$ と $I_{B \rightarrow T}^{(k-1)}$ の相関、および無相関検定での有意確率であるが、表 6 と同様、DA によって傾向が分かれた。特に、answer,

表 8 発話移行区間長 $I_{B \rightarrow T}^{(k-1)}$ と $I_{B \rightarrow T}^{(k)}$ の相関分析結果 (2 番目のツッコミ役の発話 DA 毎)

Table 8 The result of correlation analyses between utterance transition intervals $I_{B \rightarrow T}^{(k-1)}$ and $I_{B \rightarrow T}^{(k)}$ (for each DA of the second tsukkomi utterance).

DA tag	出現頻度	相関係数	有意確率
st	62 (15%)	0.13	0.327
op	28 (6.8%)	0.05	0.816
qs	9 (2.2%)	-0.48	0.192
cm	9 (2.2%)	-0.10	0.794
cn	48 (12%)	0.33	0.024
ts	155 (38%)	0.24	0.003
bq	16 (3.9%)	0.22	0.417
in	11 (2.7%)	0.35	0.298
ag	32 (7.7%)	0.49	0.005
dg	27 (6.5%)	0.01	0.951
an	16 (3.9%)	0.66	0.005



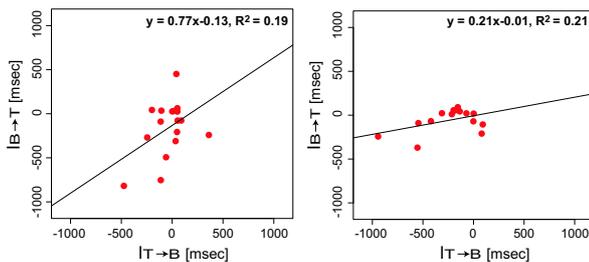
(a) answer (相関係数 0.66) (b) agreement (相関係数 0.49)

図 9 $I_{B \rightarrow T}^{(k)}$ 対 $I_{B \rightarrow T}^{(k-1)}$ の散布図および回帰分析
Fig. 9 Scatter plot and regression analysis of $I_{B \rightarrow T}^{(k)}$ v.s. $I_{B \rightarrow T}^{(k-1)}$.

表 9 $I_{B \rightarrow T}^{(k-1)}$ と $I_{B \rightarrow T}^{(k)}$ のツッコミ役発話に連続して現れる DA 組と発話移行区間長の相関 (出現頻度の上位 4 組)

Table 9 DA pairs appeared in tsukkomi role's utterance of $I_{B \rightarrow T}^{(k-1)}$ and $I_{B \rightarrow T}^{(k)}$, and their correlation for each DA pairs (top four in respect to the appeared number).

$T \rightarrow T$	出現頻度	相関係数	有意確率
ts \rightarrow ts	89 (22%)	0.18	0.087
st \rightarrow st	29 (7.0%)	-0.08	0.665
cn \rightarrow ts	17 (4.1%)	0.43	0.082
ts \rightarrow cn	15 (3.6%)	0.46	0.088



(a) cn \rightarrow ts (相関係数 0.43) (b) ts \rightarrow cn (相関係数 0.46)

図 10 高頻度で現れる DA 組における $I_{B \rightarrow T}^{(k)}$ 対 $I_{B \rightarrow T}^{(k-1)}$ の散布図および回帰分析

Fig. 10 Scatter plot and regression analysis of $I_{B \rightarrow T}^{(k)}$ v.s. $I_{B \rightarrow T}^{(k-1)}$.

agreement ではそれぞれ 0.66, 0.49 と比較的高い相関が見られた (散布図を図 9(a), 9(b) にそれぞれ示す)。また, continuer でもやや相関が認められた (相関係数は 0.33)。一方, backchannel-question と incipient の相関係数はそれぞれ 0.22, 0.35 であるが有意とはいえず, tsukkomi は他の DA と比べてサンプル数が大きく異なることから, 前節と同様に比較対象外とした。

以上より, answer, agreement, continuer では, 自分の対話のリズムを一定に保ちながら発話のタイミングを制御している可能性がある。question については比較的高い負の相関 (-0.48) がみられたが, 10%の有意水準でも有意な相関とはいえず, 今後サンプル数を増やして検証する必要がある。

次に, 前節と同様に発話の DA を限定することで発話移行区間長の関係を見る。表 9 は, $I_{B \rightarrow T}^{(k-1)}$ と $I_{B \rightarrow T}^{(k)}$ に現れるツッコミ役の発話の DA 組のうち, 頻度の高い上位 4 組を示したものである。このうち, 比較的高い相関を持つのは, ツッコミ役の発話に continuer \rightarrow tsukkomi および tsukkomi \rightarrow continuer のような順序で DA タグが現れた場合である。

このときの $I_{B \rightarrow T}^{(k-1)}$ と $I_{B \rightarrow T}^{(k)}$ の散布図を, 図 10(a), 10(b) にそれぞれ示す。相関係数はそれぞれ 0.43, 0.46 でありそれほど差は無いが, 図 10(a) の散布図からは, 直前の発話 DA が continuer である場合, tsukkomi の発話移行区間長 $I_{B \rightarrow T}^{(k)}$ は, 直前の発話移行区間長 $I_{B \rightarrow T}^{(k-1)}$ と同程度か, より短くなる傾向が見える (ただし回帰直線の傾きは, $I_{B \rightarrow T}^{(k-1)}$ で約 300msec のポーズを持つサンプルの影響によって 1 よりも小さくなっている)。一方で, 図 10(b) の散布図および回帰直線の傾きからは, 直前の発話 DA が tsukkomi であり, さらにそれが大きくオーバーラップしたとしても, 次の発話 DA が continuer である場合は, その発話移行区間長の多くは 0msec から -300msec に分布し, 比較的 0msec に近くなることが分かる。これは, tsukkomi の発話を行った後は, 相手の発話終了に同期して相槌 (continuer) を入れていると考えられる。

4.4 相手の発話終了時の予測に基づく発話開始

前節の表 9 より, $I_{B \rightarrow T}^{(k-1)}$ と $I_{B \rightarrow T}^{(k)}$ のツッコミ役発話に連続して現れる DA 組のうち最も高い頻度となるのは tsukkomi \rightarrow tsukkomi であり, これら発話移行区間長の散布図を図 11(a) に示す。この図からは, 直前の発話移行区間長 $I_{B \rightarrow T}^{(k-1)}$ と次の発話移行区間長 $I_{B \rightarrow T}^{(k)}$ の間に目立った傾向は見られない。しかし, 散布図としてはちょうど図 10(a) と 10(b) を合わせたものと見ることできる。そこで, tsukkomi DA が続けて現れる場合, 直前の発話移行区間長に対して, 次の発話移行区間長を 0msec に近づける (相手の発話終了に自分の発話開始を同期させる) 場合と, 大幅に変化させな

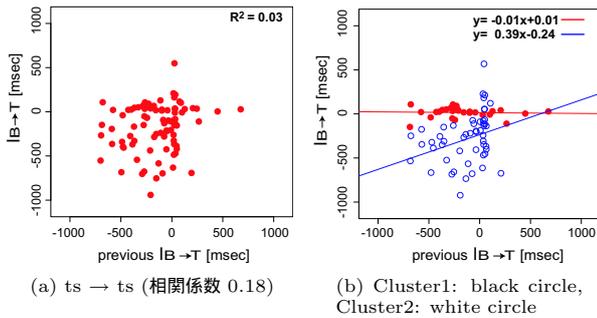


図 11 $I_{B \rightarrow T}^{(k)}$ 対 $I_{B \rightarrow T}^{(k-1)}$ の関係
 Fig. 11 Relation of $I_{B \rightarrow T}^{(k)}$ v.s. $I_{B \rightarrow T}^{(k-1)}$.

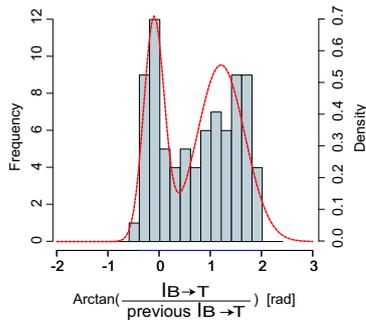


図 12 $\arctan(I_{B \rightarrow T}^{(k)} / I_{B \rightarrow T}^{(k-1)})$ の頻度分布および
 当てはめられた混合ガウス分布
 Fig. 12 Histogram of $\arctan(I_{B \rightarrow T}^{(k)} / I_{B \rightarrow T}^{(k-1)})$
 and the fitted Gaussian mixture.

い(テンポを保つ)場合の二通りの方策がとられている」という仮説を立てる。

この仮説を検証するために、散布図の各点と原点とを結んだ直線と、図の x 軸(横軸)とがなす角度を $\arctan \left\{ \frac{I_{B \rightarrow T}^{(k)}}{I_{B \rightarrow T}^{(k-1)}} \right\}$ として、 $[-\pi/4, 3\pi/4]$ の範囲で求めた。全ての点に対する直線の角度について、頻度分布を用いて分布の形状を確認した際に、もし約 $0[\text{rad}]$ および約 $\pi/4[\text{rad}]$ に平均を持つような 2 つの山を持った分布が得られるならば、その結果は上記の仮説を支持する。

図 12 は、得られた各直線の角度に関する頻度分布である。これに、expectation-maximization アルゴリズム^[14]を用いて混合正規分布を当てはめた結果、BIC が最も高くなるのはガウス分布が 2 個のときであり、その平均は $-0.1[\text{rad}]$ および $1.2[\text{rad}]$ であった。図 12 にこの密度分布を示す。したがって、2 つの傾向が存在するといえる。

このとき、それぞれの正規分布に属する確率の大小に基づいて、散布図の点のクラスタリングを行った結果を図 11(b) に示す。一方(黒丸)はほぼ傾き 0 の直線上に並ぶが、もう一方(白丸)は第三象限に分散しており、わずかな正の相関を持つに留まった。したがって、仮説のうち「相手の発話終了に自分の発話開始を

同期させる」という傾向はあるが、一方で「テンポを保持する」という傾向があるとはいえない。むしろ、ツッコミ役が、ボケ役の発話に対して tsukkomi DA の発話を開始するタイミングは、ボケ役の言語的な情報を重視している可能性が高い。つまり、ボケ役の発話が途中であっても、冗談(ボケ)として解釈する上で言語的に十分な情報、もしくは何らかのキーワードが得られた時点で、ボケ役の発話をさえぎるようにツッコミ役が発話を開始しており、その後、ボケ役の発話は文を完結させるまで持続する。この、キーワードなどの言語的な情報が現れる位置は、発話された文の文法的構造に強く依存し、単に直前の発話移行区間長との相関からでは予測できないと考えられる。

なお、以上は tsukkomi DA が連続する場合の結果であるが、4.1 節で述べたように、図 5 の頻度分布において 0msec 付近に鋭いピークがある DA が多いことから、tsukkomi 以外の DA でも、相手の発話終了へ同期した発話開始が頻繁に行われていると考えられる。

5. 討議

5.1 DA と発話移行区間長の関係

ここで、4. 章の分析結果によって得られた知見を以下にまとめる。

1. オーラップを伴う発話の割合は無視できないほど高く、ツッコミ役の発話移行区間長は下記のように偏る傾向がある(4.1 節参照):
 - 質問に関する DA(question, backchannel-question) では正
 - backchannel-question (質問的相槌) を除くバックチャネル DA (incipient, tsukkomi, continuer) では負, command (命令) を除くノンバックチャネル DA では正
 - agreement では負, answer では正
2. ツッコミ役が発話が incipient, continuer や answer の DA を持つとき、その発話移行区間長は、直前のボケ役の発話移行区間長と正の相関がある(4.2 節参照)。
3. ツッコミ役が発話が answer, agreement や continuer の DA を持つとき、その発話移行区間長は、1 つ前の自分の発話移行区間長と正の相関がある(4.3 節参照)。
4. 発話移行区間長が 0msec 付近の場合とそれ以外の場合では、発話移行区間長を決定するのに異なる要因が存在する(4.4 節参照)。

さらに、各 DA の定量的な特徴を、表 10 の 4 項目に関する星型図として図 13 にまとめる。それぞれの項目の値は、各項目の最大・最小値に基づいてスケールリングされた部分円の面積で表している。

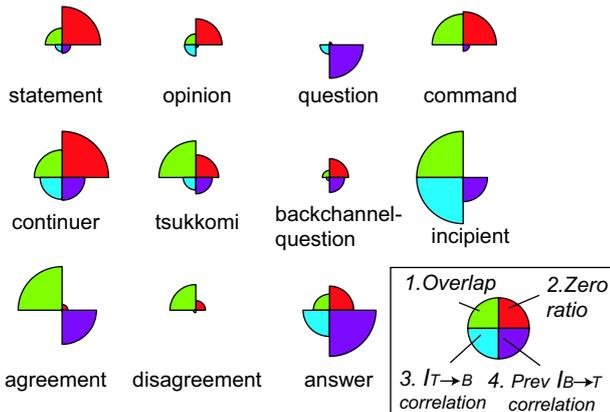


図 13 各 DA の特徴．部分円の面積（半径）は、各項目の最大・最小値でスケールされた度合いを表す（表 10 参照のこと）

Fig. 13 Feature of each DA. The size of each partial circle represents a scaled degree of each feature based on its maximum and minimum (see Tabular 10).

表 10 図 13 における 4 項目

Table 10 Four types of features for Fig.13.

項目	算出方法	参照
1. オーラップの度合い	$I_{B \rightarrow T} \leq 0$ の割合	表 5 の 4 列目
2. 相手の発話終了時に発話開始する度合い	$ I_{B \rightarrow T} \leq 80\text{msec}$ の割合	表 5 の 5 列目
3. 相手の発話移行区間長との相関	$I_{B \rightarrow T}$ と直前の $I_{T \rightarrow B}$ との相関係数の絶対値	表 6 の 3 列目
4. 直前の自分の発話移行区間長との相関	隣り合う $I_{B \rightarrow T}$ の間の相関係数の絶対値	表 8 の 3 列目

以上の結果より、発話の DA はその発話移行区間長の決定に密接に関わっていることが分かる。例えば、質問や回答 (answer) に関連する発話でポーズが多いのは、発話を行うために考える時間をとっているか、もしくは意図的にそう見せかけるためであることが、一方、多くのバックチャネル DA や agreement では、オーバーラップによって肯定的な意味合いや語気を強めることが考えられる。さらに continuer や answer では直前の発話移行区間長と正の相関があることから、対話のテンポを保つことが重要であると考えられる。

したがって、実際の対話システムを実装するには、DA の種類ごとに、発話タイミングを生成するモデル (決定木や回帰モデルなど) を用意・学習しておき、実際の対話時には、発話の DA によってこれらモデルを切り替えるといった機構が有効であると予想される。

5.2 DA 以外の要因による影響

4. 章で得られた相関係数や、4.4 節の考察からも分かるように、DA や周囲の発話移行区間長の情報だけでは、発話タイミングを決定するには不十分である。そこで、関連研究 (2 章参照) で用いられているような発話の韻律情報 (ピッチやパワーなど) や、発話の言語的情報 (キーワードなど) を、本論文で用いた DA

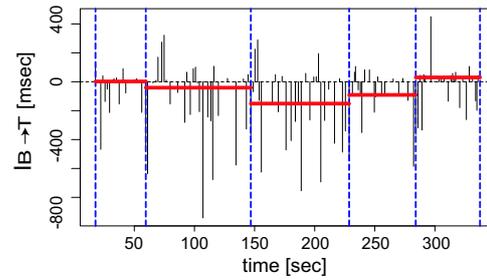


図 14 各サブネタにおけるツッコミ役発話移行区間長の中央値の変化 (横線)

Fig. 14 Median of tsukkomi role's transition interval for each subneta region (horizontal lines).

表 11 各対話のサブネタ毎の中央値 [msec]

Table 11 Median for each subneta region of each dialog [msec].

対話	サブネタ						
	1	2	3	4	5	6	7
a1	28	-279	-224	-243			
b1	12	59	35	21			
c1	2	-41	-151	-91	30		
c2	-160	7	13	73			
d1	7	-165	-47	-163	-66	-276	-53

や文脈などの情報と共に用いることが、適切な発話のタイミングを決定するためには不可欠である。

一方、対話の盛り上がりなどの長期的な流れの影響も無視できない。図 14 は、対話 c1 に関して、サブネタの切り替わる位置を縦の破線で、各サブネタ内の発話移行区間長の中央値をサブネタ毎に横線で、各 $I_{B \rightarrow T}$ の大きさを、縦の細い実線で示している。この対話では、ネタの中盤にかけて発話移行区間長が負に偏っていき、その後終盤になるにつれオーバーラップの度合いが減少することが分かる。このような中央値のパターンは、表 11 に示すように、漫才師やネタによって異なるものの、対話全体の盛り上げ方や雰囲気制御の上では考慮していく必要がある。

6. 結論

本論文では、音声対話における「間」の取り方が、各発話の持つ目的・機能に依存すると考え、各発話を 11 種類の dialog act (DA) に分類することによって分析を行った。分析対象としては、通常の対話よりも間合いの変化のさせ方が顕著に現れる漫才を選び、漫才対話コーパスの作成、およびその統計的な解析を行った。その結果、DA と発話移行区間長の関係について、5.1 節で述べたような一定の知見を得ることができた。

ただし、本論文の結果は DA のタグ集合の定義によって異なってくると考えられる。特に、{agreement, answer, continuer}, {disagreement, tsukkomi, opinion} のそれぞれの組に関しては、言語的には同じ内容の場合もあり、これらを別の DA のタグとして分類すべき

か否か、分類するとしたら厳密な定義をどのように行うかが常に議論される^[11]。したがって、トップダウン的にDAを定義し、人手でアノテーションを行うだけでなく、コーパスそのものからDAの統計的な学習を行う方法^[12]、もしくはDAを機械的にクラスタリングする方法などを検討する必要がある。

本論文の分析では、発話移行の特徴が顕著に現れる対話として、漫才というジャンルを取り上げた。一方で、DAの出現頻度には漫才特有の偏りが生じており、さらに、漫才自体が客に聞かせる演芸であるため、通常の二者間会話とは異なる間合いを持つ可能性がある。したがって、本研究を通じて得られた知見が、雑談、相談、説得や議論といった普段の対話に対して、どの程度普遍的であるかについて、より多くの対話ジャンルのコーパスによる分析、および実際の対話システムを用いた被験者実験などを通じて評価することが不可欠であり、これらを今後の課題とする。

謝辞

本研究の一部は、科学研究費補助金 18049046 の補助を受けて行なった。

参考文献

- [1] 松山隆司, 川嶋宏彰, 鷺見和彦: 人間と共生する情報システムの実現を目指して; 情報処理-特集: シンビオティック・システムの実現に向けて (2006).
- [2] 市川薫, 佐藤伸二: 対話理解に対する抑揚情報の役割; *IPSJ SIG-SLP*, pp. 51-58 (1994).
- [3] 長岡千賀, Maria Draguna, 小森正嗣, 中村敏枝: 音声対話における交替潜在時が対人認知に及ぼす影響; ヒューマンインタフェースシンポジウム, pp. 1431-1434 (2002).
- [4] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一: 韻律パターンの認識を用いた相槌挿入とその評価; 情報処理学会研究報告, pp. 33-38 (1996).
- [5] Kitaoka, N., Takeuchi, M., Nishimura, R., and Nakagawa, S.: Response timing detection using prosodic and linguistic information for human-friendly spoken dialog; *Journal of The Japanese Society for Artificial Intelligence*, Vol. 20, No. 3 SP-E, pp. 220-228 (2005).
- [6] Fujie, S., Fukushima, K., and Kobayashi, T.: Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system. *Proc. EUROSPEECH*, pp. 889-892, (2005).
- [7] Austin, J. L. (坂本百台訳): 言語と行為 (How to do things with words, Oxford Univ. Press 1962, 1975); 大修館書店 (1978).
- [8] Searle, J. R. (阪本百大, 土屋俊訳): 言語行為 - 言語哲学への試論 (Speech Acts, an Essay in the Philosophy of Language, Cambridge Univ. Press 1969); 勁草書房 (1986).
- [9] Allen, J. and Core, M.: Draft of DAMSL: Dialog act markup in several layers; (1997).
- [10] Jurafsky, D., Shriberg, E., Fox, B., and Curl, T.: Lexical, prosodic, and syntactic cues for dialog acts; *Proc. ACL/COLING-98: Workshop on Dis-*

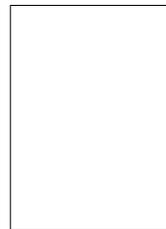
course Relations and Discourse Markers, pp. 114-120 (1998).

- [11] Dhillon, R., Bhagat, S., Carvey, H., and Shriberg, E.: Meeting recorder project: Dialog act labeling guide; *ICSI Technical Report TR-04-002* (2004).
- [12] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech; *Computational Linguistics*, Vol. 26, No. 3, pp. 339-371 (2000).
- [13] 大串健吾, 中山剛, 福田忠彦: 画質と音質の評価技術; 昭晃堂 (1991).
- [14] Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm; *J. R. Statist. Soc. B*, Vol. 39, pp. 1-38 (1977).

(2002年1月1日受付, 1月1日再受付)

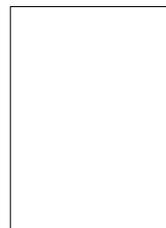
著者紹介

川嶋 宏彰 (正会員)



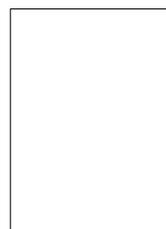
2001年京都大学大学院情報学研究科修士課程修了。2002年同大学院博士課程中退。2002年より同大学院助手。京都大学 博士(情報学)。時系列パターン認識, メディア統合, ハイブリッド・ダイナミカル・システム, 実世界インタラクションの研究に従事。2004年FIT論文賞, 2005年船井ベストペーパー賞。

スコギンズ リーバイ



2005年京都大学大学院情報学研究科修士課程修了。同年(株)コーエー入社。在学中は対話の解析に関する研究に従事。

松山 隆司



1976年京大大学院修士課程了。京大助手, 東北大助教授, 岡山大教授を経て1995年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究科知能情報学専攻教授。2002年学術情報メディアセンター長, 京大評議員。2005年情報環境機構長。工博。画像理解, 人工知能, 分散協調視覚, 3次元ビデオ, 実世界インタラクションの研究に従事。1980年情報処理学会創立20周年記念論文賞, 1990年人工知能学会論文賞, 1993年情報処理学会論文賞, 1994年電子情報通信学会論文賞, 1995年第5回国際コンピュータビジョン会議 Marr Prize, 1999年電子情報通信学会論文賞。IAPR, 情報処理学会, 電子情報通信学会フェロー。