

間の合った発話タイミング制御を目的とした漫才の動的構造の分析*

スコギンズ リーバイ

川嶋 宏彰

松山 隆司†

京都大学大学院情報学研究科‡

I. はじめに

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と考えられる。例えば漫才では、冗談を言い出すタイミングにより、客に伝わる面白さがまったく変わってくる。そこで、本研究では漫才の動的な対話構造の分析を行ない、その結果を用いて、間の合った発話タイミング制御を実現するためのシステムデザインを提案する。

発話タイミングに注目した研究としては、発話長と韻律テンプレート及び発話ピッチにより相槌を挿入するシステム [1, 3] が挙げられる。しかし、相手の発話終了時からのポーズのとり方のみを扱っており、オーバーラップを含んだ発話を行なうことができない。さらに、全ての対話を通じてタイミングの生成の仕方が固定されている。実際の（人間同士の）対話では、発話のオーバーラップやポーズのとり方を動的に変化させることで、対話のトーンを生み出していると考えられる。我々のシステムデザインは、対話中の構造を用いることで、発話のオーバーラップとポーズのとり方を動的に制御する。

II. 漫才対話コーパスに基づく分析

漫才とは二人の漫才師が滑稽な問答を中心に演じる寄席演芸である。漫才は、通常の対話と比較すると (1) 発話のペースが速く、発話タイミングの取り方で対話のトーンが大きく変化する、(2) 発話者の役割がボケ役とツッコミ役に分かれており、個々の発話の持つ役割がはっきりしている、(3) 事前に用意された台本と、客を笑わせるという明確な目的が存在する、という特徴を持つ。

いつ発話を開始するかというタイミング（発話タイミング）は、対話の目的や発話内容、前後の文脈などの様々な状況に依存して動的に制御される。本論文では、個々の発話の持つ発話内行為 (Dialog Act, DA) とそのパターン、および対話中の話題の切り替わりという2つのレイヤーに注目し、これらの動的な構造によって、発話のオーバーラップやポーズがどのように制御され、対話のトーンにどのような影響を与えるかについて検討を行う。

1. 漫才対話コーパス

一般に使用されている対話コーパスは、電話での対話や部屋の中での二人の会話などが多く、漫才の対話コーパスは存在しない。そのため、本研究では分析を行なうにあたり、漫才対話コーパスを作成した。市販の漫才ビデオよりボケ役・ツッコミ役の発話区間および客の笑い声区間をセグメンテーションし、発話内容を書きおこした。本論文では約 45 分・2709 個発話の漫才コーパスの中から二つの漫才対話（合計 10 分 48 秒・761 個発話）を用いて分析を行なう。

2. 漫才対話の構造

漫才の中の発話は時間軸上で図 1 のような構造を持つ。本論文では以下の記号を定義する：

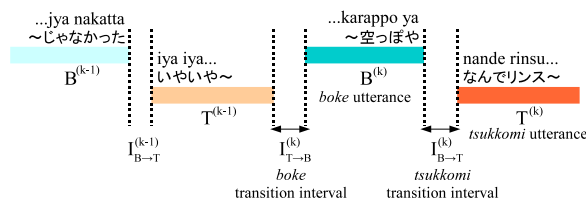
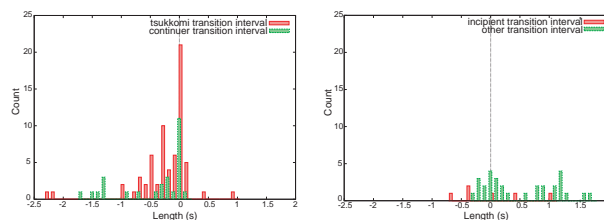


図 1: 漫才の発話構造



(a) tsukkomi, continuer の DA

(b) incipient 等の DA

図 2: DA 別の $I_{B \rightarrow T}$ 長ヒストグラム

- ボケ役の発話を B 、ツッコミ役の発話を T と記述する。一つのやり取りとして対応する発話ペアには $B^{(k)}$ と $T^{(k)}$ のように同じ添え字をふる。
- 一方の発話終了時から他方の発話開始時までの間を移行区間 (I) と呼ぶ。例えば、 B から T の移行区間を $I_{B \rightarrow T}$ と記述する。

移行区間 I が負であればこれをオーバーラップと呼ぶ。一方で、 I が正であればこれをポーズと呼ぶ。通常の漫才においては主にツッコミ役が漫才のペースやトーンを決めていると考えられるため、ツッコミ役が制御する移行区間である $I_{B \rightarrow T}$ が、時間的にどのような変化をするかに注目して以下の分析を行う。

3. 漫才対話全体の発話特徴

A. Dialog Act の特性

まず、漫才対話の早いペースやり取りに関係していると考えられる $I_{B \rightarrow T}$ の符号を調べることで、発話オーバーラップの割合が 57% と発話ポーズの割合が 43% であることが分かった。対話の半分以上の発話が重なっていることから、これを無視すると対話のトーンが大きく異なってくると考えられる。

次に、発話のオーバーラップとポーズの役割を調べるために、Jurafsky ら [2] による DA のラベル { continuer (相槌), backchannel-question (質問的相槌), incipient (発話権の獲得), statement (陳述), agreement (同意), answer (回答), backchannel-opinion (質問的意見) } を漫才対話中の各発話にふた。これに加え、漫才にしか存在しない tsukkomi (ツッコミ) というラベルを追加した。tsukkomi とは相槌と似ており相手が喋っている間挿入するが、「うん」や「はい」ではなく「なんでやねん」、「どないやねん」などの疑問形であり、強い語意を持つ DA であると定義する。DA 別のヒストグラムを図 2 に示す。図 2(a) の tsukkomi, continuer は図 2(b) の incipient 等より発話のオーバーラップの割合が大きい。こ

* Analysis of the Dynamic Structure of *Manzai* for the Purpose of Good Timing Control

† Levi SCOGGINS, Hiroaki KAWASHIMA and Takashi MATSUYAMA

‡ Graduate School of Informatics, Kyoto University

DA の遷移	頻度	割合
tsukkomi → tsukkomi	35/66	53.0%
tsukkomi → backchannel-question	11/66	16.7%
tsukkomi → continuer	5/66	7.6%

表 1: ツッコミ役の DA の遷移割合 (上位三位)

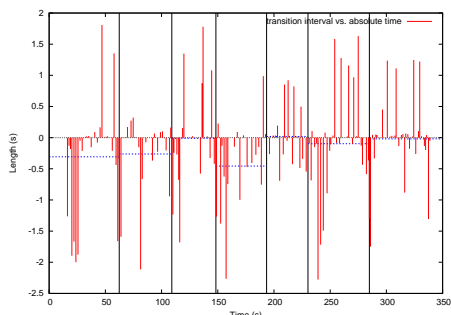


図 3: サブネタ区間変化

れはツッコミ役が tsukkomi, continuer の DA を通じて漫才対話の緊張感や躍動感を与ようとしていると考えられる。

B. Dialog Act パターン

前節で DA に分類したデータを時間的に並べると, DA の特徴的な連続パターンが存在することが分かった。例えば, ツッコミ役の DA のパターンには表 1 に示すような代表的な遷移が存在する。さらに, continuer の後に tsukkomi が複数表れるなどの DA パターンも存在する。この DA パターンをつなぐことによってツッコミ役は漫才対話の独自のテンポを生み出していると考えられる。

C. サブネタ区間

漫才対話コーパスのデータを人間に聞かせると, 対話の中にいくつかの話題が現れることが分かった。漫才では一つの台本に沿った対話をネタと呼ぶことから, この一つ一つの話題をサブネタと呼ぶことにする。漫才対話はサブネタにどのような影響を受けているのかを検討した。図 3 中の黒い縦線はサブネタの切り変わる位置を表している。発話毎の $I_{B \rightarrow T}$ 長を赤い縦線の長さで, サブネタ毎の平均 $I_{B \rightarrow T}$ 長を青い横の点線で表している。図 3 により, 発話オーバーラップが多いサブネタと少ないサブネタが交代に表れることが分かる。この特徴は他の漫才のネタでも確認でき, ツッコミ役はサブネタ毎にタイミングの取り方を切りかえることで対話の流れに緩急をつけ, 人の注目を引きつける効果を与えていると考えられる。

III. 間の合った発話タイミング制御

本研究で行なった分析により, 発話のオーバーラップと DA パターン, 及びサブネタ毎の発話タイミング特性は漫才対話のトーンを決める上で重要な役割を担っていることが分かった。これを音声対話システムで実現するためには, 我々はマルチレイヤーのシステムが有効であると考え。各レイヤーには以下のような機能を持たせる(ただし, 以下ではシステムがツッコミ役を行うことを想定している):

決定木レイヤー システムが発話していないという状態において, フレーム毎 (50-100ms 程度) に, 発話を開始するか否かを決定する。入力直前までのボケ役の発話から抽出された韻律情報 (エネルギー, ピッチ変化など) であり, 竹内ら [3] のシステムを参考にしているが, ボケ役の発話時にも常に処理を

連鎖長	頻度	DA パターン
2	12	ct
2	11	tq
4	10	tttt
3	10	ctt

表 2: Prediction suffix tree による最多 DA パターン (c は continuer, t は tsukkomi, q は backchannel-question を表している)

行うことで, 相手の発話終了時を予測しながら発話のオーバーラップを行うことが可能となる。

可変長マルコフ連鎖レイヤー 特徴的な DA パターンを状態として持つような, 可変長マルコフ連鎖を, あらかじめ対話コーパスより学習しておく。さらに, 各 DA に対してそれぞれ異なる決定木をあらかじめ求めておく。システムの実行時には, 各発話終了毎にこの状態を遷移させることで, DA の系列を生成していき, これに伴って決定木を動的に切り替えていく。可変長マルコフ連鎖は, 予測接尾木 (Prediction Suffix Tree, PST) を用いることで学習することが可能である。実際に PST を用いて抽出された, 出現頻度の高い DA パターンを表 2 に示す。

サブネタレイヤー 決定木レイヤー によって発話開始が決定された後に, 実際にシステムが発話を開始するまでの待ち時間を, サブネタ毎に切り替えていく。これにより, 漫才の 1 つのネタの中で, 緩急のある対話を実現することが可能となる。

IV. おわりに

本論文では, 特に発話間のタイミングに注目し, 具体的な対象として漫才を取り上げることで, 対話の持つ動的な構造について分析を行った。漫才対話コーパスの作成と分析により, 発話のオーバーラップ, DA パターンとサブネタ毎の発話タイミング特性は漫才対話のトーンを決める上で重要な役割を持つことが分かった。ただし, 今回の分析結果の解釈は直感的なものに留まっており, 今後はより定量的な分析とシステムの評価を行なう必要がある。現在, 漫才コーパスの拡充, システムの実装と評価, 人間の持つ時間解像度の主観的評価を行なっており, 稿を改めて報告する予定である。

機械を人間の環境にコピキタス的に置こうとすれば, 人間同士の対話に表れるトーンや緊張感を生成・理解することが不可欠である。本研究で提案するデザインをベースとしたシステムは, プロアクティブな要素を持ったマンマシン・インタラクションシステムの実現に向けての第一歩となる。

謝辞: 本研究の一部は, 科学研究費補助金 13224051 及び 16700175 の補助を受けて行なった。

参考文献

- [1] 市川薫, 佐藤伸二: 対話理解に対する抑揚情報の役割, 情報処理学会 音声言語情報処理 研究報告, 94-SLP-2, pp. 51-58, 1994.
- [2] Jurafsky, D., Shriberg, E., Fox, B. and Curl, T.: Lexical, prosodic, and syntactic cues for dialog acts. *Proc. ACL/COLING-98: Workshop on Discourse Relations and Discourse Markers*, pp. 114-120, 1998.
- [3] Takeuchi, M., Kitaoka, N. and Nakagawa S.: Generation of natural response timing using decision tree based on prosodic and linguistic information. *EUROSPEECH2003*, pp. 609-612, 2003.