

# 落語の役柄交替における視覚的「間合い」の解析

川 嶋 宏 彰<sup>†</sup> 西 川 猛 司<sup>†</sup> 松 山 隆 司<sup>†</sup>

円滑な会話を進めるうえでは、参与者間で自然な間合いの発話権受け渡し（話者交替）が行われる必要がある。従来研究では、一方の発話終了からもう一方の発話開始までの時間（発話移行区間長）がしばしば解析されてきたが、実際の対面会話においては、相手話者の視線や口元の動きなどの視覚的に観察可能な身体動作も、発話権受け渡しの「間合い」を支えていると考えられる。そこで本研究では、話者が1人であるにもかかわらず複数の役柄間での円滑な話者交替を表現している落語に着目し、演者の頭部動作の開始と発話開始・終了の時間的構造の解析を行った。その結果、先行発話終了に対する頭部動作の開始タイミングは、二者間会話の漫才における後続話者の発話開始タイミングと類似しているという知見が得られた。これは、視覚的な刺激の制御が、話者交替において受け手が感じる冗長な間合いを補間可能であることを示唆し、実際に被験者実験によってその評価を行った。

## Analysis of Visual Timing Structure in *Rakugo* Turn-taking

HIROAKI KAWASHIMA,<sup>†</sup> TAKESHI NISHIKAWA<sup>†</sup>  
and TAKASHI MATSUYAMA<sup>†</sup>

Coordination of turn-taking with natural timing is important to realize a smooth conversation among multiple speakers. The existing studies often analyze an utterance transition interval, which is the interval from the end of the previous speaker's utterance to the beginning of the succeeding speaker's utterance. However, the effect of visually observable motions (e.g., gazing and lip motions) is also important to coordinate timing of turn-taking in face-to-face interaction. In this paper, we therefore focus on Japanese traditional *Rakugo* performances, in which smooth conversations among multiple roles are displayed by only one performer's act based on the control of the head motion timing during switching the roles. The result of analysis shows that the dynamic structure of intervals from the end of the previous utterance to the beginning of the head motions in *Rakugo* performances is similar to the structure of utterance transition intervals in *Manzai* dialogs, which are two-speaker conversations. The result suggests that the control of visual stimuli have the capability of filling redundancies in the pauses of turn-taking. We therefore examined the effectiveness based on psychological experiments.

### 1. はじめに

会話の話者交替における「間合い」は、意思伝達や協調の成否を左右する重要な要因である。近年、伝送遅延の生じる遠隔会議システムでの会話や、音声対話システムとの会話に代表されるように、会話の行われる状況が多様化しているが、これら新たな会話状況では、日常会話で用いられる間合いが保たれず、会話の円滑性が失われることがある。たとえば、発話の意図を伝えるためには、相手の発話終了を待たずにオーバーラップをともなう発話を行うことや、相手の発話終了から比較的短いポーズで発話することが重要となる<sup>1),2)</sup>。しかし、先に述べた新たな会話状況では、伝

送や処理遅延の影響によって冗長な間合いができるため、本来の発話意図を十分に伝達できない。また、先行話者の発話に対して後続話者の応答発話が遅れた場合、先行話者は不安感から再発話を行い、しばしば後続話者の発話とのぶつかりが生じる。

話者交替、すなわち話者間における発話権の受け渡しでは、先行話者の発話権解放から後続話者の発話権取得までの時間に、話者にとって自然に感じる時間長が存在すると考えられる。この自然な間合いについては、従来研究において、先行話者に対する後続話者の発話タイミングが注目され、会話の場面や文脈との関連が報告されてきた（2章参照のこと）。しかし、発話権の解放・取得の認識や、その間合いが自然であったか否かの評価は、あくまで会話参与者や第三者（聴衆、観客など）の各自の主観の中で、意識的、無意識的に行われている。したがって、音声という単一のモ

<sup>†</sup> 京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

ダリティではなく、話者の視線や口元の動きなどの視覚的手段をはじめとするさまざまなモダリティの信号も、話者交替の間合いを表現、認識し、評価する際に関与している可能性がある。

そこで本論文では、「身体動作のような視覚情報を相手に提示することによっても、音声発話と同様に自然な話者交替の間合いを表出できる」という仮説を立て、会話映像の解析を通じてその妥当性を検証することを目的とする。この仮説が妥当であるならば、遠隔会議システムや、表示装置をともなう音声対話システムにおいて、音声のみでは冗長な間合いが生じた場合に、話者への提示映像に視覚的な刺激を人工的に生成・付加することで、話者の主観における間合いの冗長さを解消できるような、新たな会話支援システムのデザインにつながると期待できる。

このとき、実際の会話支援システムにおいて再現可能な行動単位を、解析の対象として選ぶべきである。特に、伝達される具体的な意味内容と動作との関係を調べる場合、どのような動作を解析対象とするかが重要となる。しかし、音声ではその波形から得られる情報（ピッチなど）以上に、その開始タイミングが意図伝達と関連があるという知見が得られていることから<sup>3)</sup>、実際の動作（刺激）が何であるかよりも、何らかの刺激が適切なタイミングに現れること自体が重要とも考えられる。したがって、ここで想定する支援システムでは、解析対象と同じ動作を表示するのではなく、どのような視覚刺激であったとしても、その開始タイミングを再現できればよいと考える。ゆえに、解析対象とする動作は、開始タイミングが明確であり、話者交替においてつねに積極的に利用されるものが望ましい。

ところが、実際の会話では、会話の性質（議論、雑談など）、実験状況の自然さ、話者の性格などのさまざまな要因が影響して動作を決定するため、話者交替に現れる一般的な行動を定義することは困難である。ここでは、他者に知覚される視覚刺激の開始タイミングが、話者交替に利用されている実例を調べたいのであり、一般の対話ではなく、むしろ人工的にデザインされた視覚刺激を解析することも意義がある。

そこで、我々はこの会話映像の解析の困難さを解決するために、解析対象として、日本の寄席演芸の1つである落語に着目する。この理由の1つは、寄席演芸における演者の動作や発話の時間関係（タイミング構造）は、発話意図を正確に伝え、かつ観客を楽しませるためにデザインされ、最適化されていると期待できるからである。さらに重要なのは、落語においては、1人で複数の役柄を演じ分けながら会話情景を表現す

るため、役柄交替を発話だけで行った場合、一方の発話終了からもう一方の発話開始までの時間（発話移行区間長）は必ず正となり、発話内容によってはしばしば冗長となりうる。このとき落語では、演者の頭部動作のタイミング制御によって、発話の冗長な間合いを解消し、自然な話者交替の間合いを演出している可能性がある。本論文ではこの検証のために、各役柄の発話開始・終了に対する、役柄交替の頭部動作開始のタイミングが、二者間会話での発話タイミングと比較してどのような類似性を持つかについて解析を行う。

まず次章では、関連研究に対する本研究の位置づけを述べ、3章では、各話者の主観における発話権の受け渡しがどのように行われるかの話者交替モデルについて述べ、このモデルに基づいて、落語の役柄交替円滑化の仮説を述べる。実際に落語を解析するために、4章では、本研究で作成した落語解析用コーパスについて述べ、5章にその解析結果を示す。6章では実際に被験者に刺激を提示することでその心理的效果を調べ、最後に7章で本論文の結論を述べる。

## 2. 関連研究と本研究の位置づけ

会話における話者交替のメカニズムは、社会心理学をはじめさまざまな分野の関心を集めてきた。Duncanらは、話者交替において話し手と聞き手が視線を合わせることから、話者間で共有される種々の合図（シグナル）によって話者交替の意図が伝達されると考え、このプロトコルをモデル化した（シグナルモデル<sup>4)</sup>。一方、小磯らは、聞き手は話し手の発話に現れる文法的切れ目などから話者交替が可能となるタイミングを自律的に認識し、発話開始を行うとするモデルを提案した（自律モデル<sup>5)</sup>。

これらは、聞き手が話し手の何を手がかりに、発話権の解放時点を認識するかに関するモデルであり、このどちらが妥当であるかを論じることは本論文の目的ではない。むしろ、多人数の会議ではシグナルモデル、雑談では自律モデル、などのように、会話の状況（参加者の人数・立場、目的の有無など）に依存して、いずれが優位になる場合もあると考えられる。本論文では、後続話者が、先行話者の発話権解放時点を自律的もしくはシグナルにより認識したとして、その解放時点に対する実際の発話権取得までの時間差（発話権移動区間長）に着目する。

話者交替における発話のタイミング制御に関しては、

本論文では、数学的に厳密な意味ではなく、ある時刻から別の時刻までの時間差の意味で「区間長」を用いる。そのため、区間長は負になる場合もあるとする。

先行話者の発話終了から後続話者の発話開始までの時間（交替潜時、本論文では発話移行区間長と呼ぶ）を解析した研究がある。市川らは、実際の自由対話において、相手の発話の維持やキーワードが現れてから約 400 msec 以内で相槌が打たれることが多いという結果を示した<sup>6)</sup>。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、両者の交替潜時の長さの組合せを操作して（第三者である）被験者に提示し、その長さが、オペレータの接客態度の印象に与える影響を解析した<sup>7)</sup>。リーバイらは、寄席演芸の 1 つである漫才演目の解析を通じて、発話移行区間長と会話の文脈や発話目的・意図との関連を調べた<sup>8),9)</sup>。さらに藤原らは、一般のタスク指向対話における発話意図と対話リズムの関係について分析を行った<sup>3)</sup>。

上述の音声発話に関する解析では、発話の目的や文脈によって発話タイミングが適切に制御されるという知見が得られている。本論文で解析する頭部動作のタイミング制御においても、発話タイミングと同様の傾向が得られるか、異なるのはどのような場合かを調べることは重要になる。そこで、対話の発話内行為（dialog act, DA）<sup>10)</sup>を用いて発話目的（陳述、質問など）を分類し、発話の DA ごとの傾向の違いを二者間会話と比較する。比較対象は、落語と同じく観客に聞かせる寄席演芸が望ましく、本論文では漫才を用いる。

3. 会話における間合いのタイミング構造

1章で述べたように、話者交替における発話権解放・取得の表現や認識は、発話だけでなく、相手の動作などのさまざまなモダリティの信号によって行われるとする。そのうえで、まずはじめに発話権移動タイミングのモデル化を行う（3.1 節）。次に、このモデルに基づいて、発話権の取得に発話を用いる場合（3.2 節）、および動作を用いる場合について述べる（3.3 節）。以上をふまえ、落語における発話と動作のタイミング構造およびその効果に関する仮説を述べる（3.4 節）。

3.1 発話権移動タイミングのモデル

先行話者（preceding speaker）の発話権解放（Release）を  $R^p$ 、後続話者（succeeding speaker）の発話権取得（Acquisition）を  $A^s$  で表し、発話権解放時刻  $T(R^p)$  に対する発話権取得時刻  $T(A^s)$  までの区間  $I(R^p, A^s)$  を発話権移動区間と呼ぶことにする（図 1）。実際には、発話権の解放や取得が直接相手に伝わることはなく、後続話者が「先行話者が発話権を解放した」と認識（aware）する時刻  $T(W_R^s)$  は、先

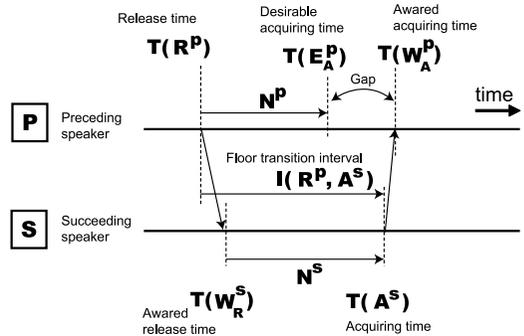


図 1 発話権移動区間  
Fig. 1 Turn-taking interval.

行話者自身が解放したと思う時刻  $T(R^p)$  とずれることがあります。逆に、先行話者が「後続話者が発話権を取得した」と認識する時刻  $T(W_A^p)$  も、後続話者自身が取得したと思う時刻  $T(A^s)$  と必ずしも一致しない。

このとき、発話権移動区間長（区間そのものと同じ記号  $I(R^p, A^s)$  を用いる）には、先行話者と後続話者が最も自然と感じる長さが文脈や発話内容などに応じて存在すると仮定する。先行話者と後続話者それぞれが自然であると感じる長さを  $N^p, N^s$  とすれば、以下のような間合いの認識・評価・調整が行われる。先行話者の間合いの認識と評価 先行話者は時刻  $T(R^p)$  に発話権を解放するとともに、時刻  $T(W_A^p)$  に後続話者の発話権取得を認識する。このとき、次の式が成り立つときに最も自然な間合いを感じる。

$$I(R^p, W_A^p) \triangleq T(W_A^p) - T(R^p) = N^p \quad (1)$$

後続話者の発話権取得を最も自然であると感じる時刻は  $T(E_A^p) \triangleq T(R^p) + N^p$  であり、 $T(E_A^p) = T(W_A^p)$  ならば式に一致する。なお、先行話者が待てる間合いには上限  $N_{max}^p$  があり、 $T(R^p) + N_{max}^p$  以降も後続話者が発話権を取得しないと認識する場合は、その間合いの冗長さから不安感を感じ、再発話が生じるとする。

後続話者の間合いの調整 後続話者は時刻  $T(W_R^s)$  で先行話者の発話権解放を認識し、以下の式が成り立つように発話権取得の時刻  $T(A^s)$  を調整する。

$$I(W_R^s, A^s) \triangleq T(A^s) - T(W_R^s) = N^s \quad (2)$$

なお、必ずしも意識的に調整するのではなく、文脈や発話内容によっては、思考時間などの要因が関与して自然とこの間合いになりうるとする。

第三者の間合いの認識と評価 会話に直接参与しない第三者（other）も話者交替の間合いを各自感じる。ある第三者が認識する発話権解放と取得の時

間を  $T(W_R^o), T(W_A^o)$ , 自然と感じる発話権移動区間長を  $N^o$  とすれば,

$$I(W_R^o, W_A^o) \triangleq T(W_A^o) - T(W_R^o) = N^o \quad (3)$$

の成立を評価することになる。

式 (1), (2), (3) の間合いから離れるほど, 各自が感じる話者交替の不自然さは増大する。ここで, この不自然さが話者にどの程度許容されるかは別途議論が必要であるが, 本論文では簡単化のために, ちょうど式 (1), (2), (3) が成立する場合を主に考える。

たとえば,  $N^p = N^s = N$  のように両話者の間で自然であると感じる発話権移動区間長が共有されるとする。さらに, 後続話者と先行話者における, 先行話者の発話権解放時刻と後続話者の発話権取得時刻の認識がそれぞれ正確であれば

$$T(W_R^s) = T(R^p), \quad T(W_A^p) = T(A^s)$$

$$\therefore I(R^p, A^s) = T(A^s) - T(R^p) = N$$

が成立することから, 両話者の主観ではいずれも自然な発話権受け渡しが行われたと感じることになる。また, 両話者で自然な間合いが共有されず,  $N^p = N^s = N$  でない場合でも, 発話権解放・取得の認識のずれによって, 式 (1) と式 (2) がともに成立する場合があります。

以下 3.2, 3.3 節では, 会話の参与者 (先行, 後続話者) に注目し, 発話権の解放・取得が具体的にどのように行われるかについての仮定を述べる。第三者の場合, これらの仮定で,  $T(W_R^s)$  を  $T(W_R^p)$ ,  $T(W_A^p)$  を  $T(W_A^s)$  に置き換えることになる。

### 3.2 発話による発話権移動タイミングの制御

二者間会話において音声発話 (Utterance) に着目したとき, 話者交替は, 先行話者の発話 (先行発話), および後続話者の発話 (後続発話) から構成される (図 2)。ここで, 先行発話の開始, 終了を  $U_b^p, U_e^p$ , 後続発話の開始, 終了をそれぞれ  $U_b^s, U_e^s$  と表す。

通常, 先行話者は自身の発話に含まれる重要なキーワードをすべて発信した後に発話権を解放する。また, 後続話者は, 先行発話中の文法的・意味的区切りや韻律情報によって先行話者の発話権解放を認識する。したがって, 先行話者の発話権解放時刻  $T(R^p)$  および後続話者に認識される時刻  $T(W_R^s)$  は, 先行話者の発話終了  $T(U_e^p)$  と同時かそれよりも前になる。一方で, 後続話者は発話開始  $U_b^s$  によって先行話者に発話権取得を伝達し, 時間差なく先行話者に認識されると考えるのが自然である。したがって, 以下の式が成立する。

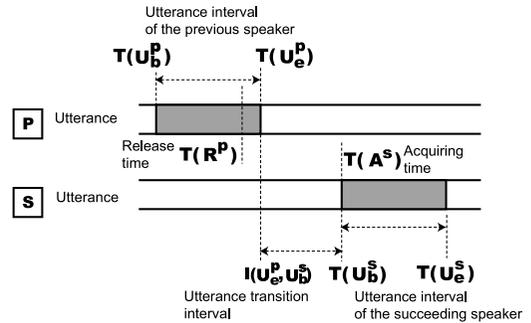


図 2 話者交替における発話のタイミング構造  
Fig. 2 Utterance timing structure in turn-taking.

$$T(R^p) \leq T(U_e^p), \quad T(W_R^s) \leq T(U_e^p) \quad (4)$$

$$T(A^s) = T(W_A^p) = T(U_b^s) \quad (5)$$

ただし, 実際に客観的な解析を行う際には,  $T(R^p)$  や  $T(W_R^s)$  の正解がなく, その推定も困難であるため, 本論文では, 話者交替に関する多くの研究と同様に, 先行発話終了時刻  $T(U_e^p)$  に対する後続発話開始時刻  $T(U_b^s)$  までの時区間  $I(U_e^p, U_b^s)$  (発話移行区間) を調べる。このとき式 (4) より, 先行話者の発話を後続話者がオーバーラップして発話開始し, 発話移行区間長  $I(U_e^p, U_b^s) \leq 0$  となる場合も生じる。二者間会話では, 言語的な発話内容が同じでも, オーバーラップ ( $I(U_e^p, U_b^s) \leq 0$ ) やポーズ ( $I(U_e^p, U_b^s) > 0$ ) を適切に用いて話者交替を行うことで, 発話の裏にある意図を伝える場合が多い<sup>1)</sup>。

### 3.3 動作による発話権移動タイミングの制御

実際の対面会話においては, 身体動作 (以下, 単に動作と呼ぶ) が話者交替の意思伝達手段として用いられると考えられる<sup>11)</sup>。そこで, 前節の音声発話による発話権移動を, 動作を用いた場合に拡張する。

会話中に表出される動作はさまざまであるが, 発話との関係により以下 2 つに大別する。

- 発話付随動作: 発話に付随する動作 (発声器官, 顔形状の動きなど)。発話音声は, 発声器官の物理的運動の結果, 空気振動として生成される。
- 発話独立動作: 発話と独立になりうる動作 (手・眼球の動き, 表情・姿勢の変化など)。発話音声の有無にかかわらず表出できる。

このうち唇や顔形状などの発話付随動作や, 発話独立動作のうち相手に対する顔きなどは, 話者交替の自然さを支えるうえで重要な役割を担っていると考えられる。これらは, しばしば発話開始時に発話音声よりも早い時点から表出される。そこで, 後続発話時における発話付随動作の開始時刻を  $T(M_b^s)$ , 発話独立動作の開始時刻を  $T(m_b^s)$  とすると, 発話権解放の伝達

各人 (会話参与者や聴衆) が感じる自然な間合いは, 普段のコミュニケーションから学習されるためにおおよそ共有されると考えられ, この場合  $N^p$  と  $N^s$  は同程度となる。

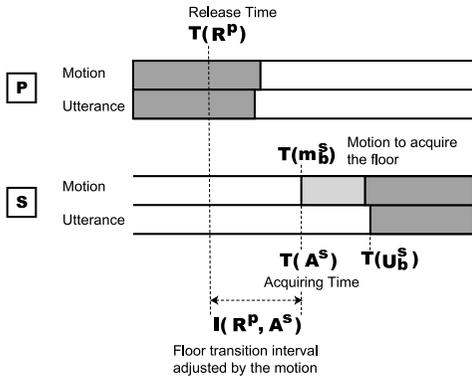


図 3 動作による発話権取得  
Fig. 3 Acquisition of floor using motion.

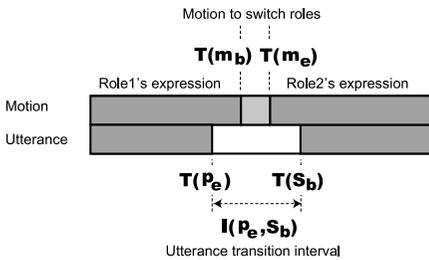


図 4 落語における役柄交替タイミング構造  
Fig. 4 Turn-taking timing structure of Rakugo.

時には、式 (5) の代わりに

$$T(A^s) = T(W_A^p) = \min\{T(M_b^s), T(m_b^s)\} \quad (6)$$

が成立すると仮定する。すなわち、後続発話の開始前に、しばしば動作によって発話権取得の表現・認識がなされると考える (図 3)。

### 3.4 落語の役柄交替タイミング構造

落語は、演者 (落語家) 1 人が複数役柄を演じ分けることによって、滑稽な会話情景を表現する芸である。ここで、落語家によって同時に演じられる役柄はただ 1 つであるため、話者交替においては、演じる役柄の切替え (役柄交替) が生じる。落語における役柄交替は、二者間会話の話者交替に相当する。

落語家が二者間会話を演じる場合、一方の役柄は、頭部が右方向に向けられることによって表現され、他方の役柄は、頭部が左方向に向けられることによって表現される。したがって、先行話者に相当する役柄 (先行役柄) から後続話者に相当する役柄 (後続役柄) への役柄交替は、頭部を左から右もしくは右から左に振るような動作によって表現される<sup>12)</sup>。本論文では、この頭部動作を役柄交替動作と呼び、動作開始時刻を  $T(m_b)$ 、終了時刻を  $T(m_e)$  と表す。また、役柄交替動作と発話とのタイミング構造を役柄交替タイミング構造と呼ぶ (図 4 に例示)。

落語家によって同時に演じられる役柄はただ 1 人であるため、先行役柄の発話 (先行発話) と後続役柄の発話 (後続発話) は、時間的に重複することはない。したがって、先行発話終了時刻を  $T(p_e)$ 、後続発話開始時刻を  $T(s_b)$  と表すと、落語演目中の役柄交替においては、以下の式 (7) がつねに成立する。

$$I(p_e, s_b) > 0 \quad (7)$$

ここで、時区間  $I(p_e, s_b)$  は二者間会話における発話移行区間  $I(U_e^p, U_b^s)$  に対応するため、 $I(p_e, s_b)$  も同様に発話移行区間と呼ぶことにする。

さて、落語の観客や視聴者が、先行役柄の発話権解放と後続役柄の発話権取得を認識する時刻を、3.1 節と同様にそれぞれ  $T(W_R^o)$ 、 $T(W_A^o)$  で表す。すると、まず発話権解放に関して、式 (4) は

$$T(R^p) \leq T(p_e), \quad T(W_R^o) \leq T(p_e) \quad (8)$$

となる。ここで  $T(R^p)$  は、落語家が演じる先行役柄の発話権解放時刻である。一方、落語における後続役柄の発話権獲得を、後続発話開始にて表現しているならば、式 (5) は

$$T(A^s) = T(W_A^o) = T(s_b) \quad (9)$$

となる。ここで  $T(A^s)$  は、落語家が演じる後続役柄の発話権取得時刻である。しかし、式 (9) では次の問題が生じる。

ある発話の文脈で、非常に小さな発話権移動区間長  $I(W_R^o, W_A^o)$  が観客に望まれるとする。これを満たすには、式 (8)、(9) より、しばしば 3.2 節で述べたような発話のオーバーラップ ( $T(s_b) \leq T(p_e)$ ) が必要となる。ところが、式 (7) の制約から発話のオーバーラップは表現不可能であり、音声のみでは発話目的や意図が観客に十分伝達されない恐れがある。また、1 人で役柄を切り替える際に短い発話移行区間長を用いたのでは、観客に伝わる会話の情景が異なってくる。そのため、落語では全般に通常よりも冗長な間合いで発話が行われる。このとき、落語家は話速の変化などいくつかの方法を用いて、二者間会話の自然な話者交替の情景を演出していると考えられるが、本論文では、特に役柄交替動作、すなわち頭部動作を行うタイミングに着目し、次の仮説を立てる。

落語の話者交替円滑化仮説 落語においては、先行役柄の発話終了に対する役柄交替の頭部動作開始のタイミング  $I(p_e, m_b)$  を、二者間会話における発話移行区間長  $I(U_e^p, U_b^s)$  と同様に制御することで、自然な発話権移動区間を演出している

すなわち、落語における後続役柄の発話権取得は、しばしば式 (9) の代わりに

$$T(A^s) = T(W_A^o) = T(m_b) \quad (10)$$

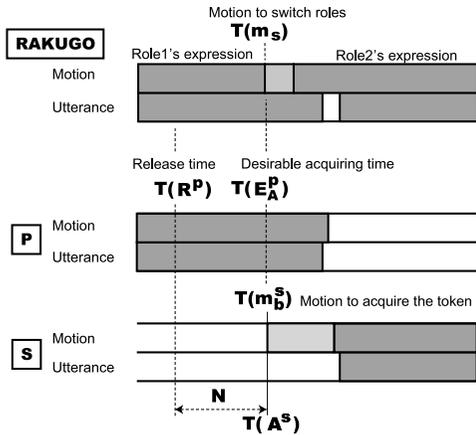


図5 役柄交替動作による発話権移動区間長の制御

Fig. 5 Control of a turn-taking interval based on head motion.

を用いて表現・伝達されると考える。

役柄交替動作は、3.3節の発話独立動作と同様に、発話の時間と同期させる必要はない。したがって、落語家は式(7)の制約下においても、役柄交替動作の開始タイミングを用いて、自然な発話権移動区間長に調整できることになる。たとえば図5の状況においては、落語家は、役柄交替動作の開始時刻  $T(m_b)$  を早めることで、先行発話終了時刻  $T(p_e)$  の前にも発話権取得を表現することができる。

以下の章では、先行発話終了時刻  $T(p_e)$ 、後続発話開始時刻  $T(s_b)$ 、役柄交替動作開始時刻  $T(m_b)$  に現れるタイミング構造を、 $I(p_e, m_b)$  を中心に解析することで、上述の仮説を検証する。

#### 4. 落語役柄交替コーパスの作成

役柄交替タイミング構造の解析にあたり、役柄交替を記録したコーパスが必要となる。そこで、役柄交替動作、および先行・後続発話区間を、市販ビデオ中の映像と音声からそれぞれ抽出した。演者はいずれも十分な経験がある落語家である。

コーパス作成に用いた演目

本研究では、表1に示す3演目から役柄交替コーパスを作成した。得られた知見が1人の落語家の個性によるものであることを防ぐために、2人の落語家の演目を選定した(演目名のRおよびSが落語家の識別子)。また、得られた知見が演目に依存しないよう、同一落語家の2つの演目を選定した(R1およびR2)。

ラジオ落語のように、もちろん音声のみでも聞き手は落語の面白さを感じられるが、頭部動作をはじめとする視覚情報によって、役柄の意図や状況がより詳細に伝達されると考えられる。

表1 コーパス作成の対象とした落語演目  
Table 1 Rakugo performances in the corpus.

演目	演目の時間	役柄交替場面数
R1	22分15秒	106
R2	20分58秒	64
S1	18分23秒	59

なお、これら演目は以下の基準によって選定した。

- 役柄交替が比較的多い。
- 二者間会話を主とする。
- 役柄交替動作が明らかである。

#### 4.1 コーパスの作成方法

役柄交替動作の開始時刻  $T(m_b)$  と終了動作  $T(m_e)$ 、先行、後続発話の開始時刻  $T(p_b)$ 、 $T(s_b)$ 、および先行、後続発話の終了時刻  $T(p_e)$ 、 $T(s_e)$  の抽出方法について述べる。

#### 役柄交替動作の区間抽出

本研究では、役柄交替動作が発生する時区間  $I(m_b, m_e)$  を、ビデオ映像から手動抽出した。役柄交替動作は、頭部を左右にふりむける動作であるため、水平方向への頭部の動作として抽出できる。しかし、落語演目においては、役柄交替の場面でなくても、水平方向への頭部の動作がしばしば生じる。そのため、役柄交替動作の自動抽出は困難である。 $I(m_b, m_e)$  の手動抽出においてはアノテーションソフト Anvil<sup>13)</sup> を用いた。

手動抽出においては、ビデオ映像中の  $I(m_b, m_e)$  であると見なされたフレーム列に対して、役柄交替動作を表すタグを付加し、その他のフレーム列に対しては、タグを付加しなかった。このタグにより、解析において、 $T(m_b)$ 、 $T(m_e)$  に相当するフレーム時刻を特定できる。

#### 先行発話・後続発話の区間抽出

本研究では、先行発話と後続発話が発生する時区間  $I(p_b, p_e)$ 、 $I(s_b, s_e)$  を、ビデオ音声から抽出した。まずは自動抽出を行い、音強が閾値以上である区間を発話区間と見なした。ただし発話音声の音強は、発話中であっても著しく閾値の上下を変動する。また、観客の笑い声が落語家の発話音声に混入する場面や、極端に短い発話移行区間長が存在する。そこで、自動抽出後に手動により発話区間を修正した。さらに、抽出した発話区間の中から、先行発話・後続発話を手動により特定した。自動抽出および手動修正においてはアノテーションソフト Praat を用いた。

ここで、特定した時区間に対して、先行発話あるい

表 2 落語における DA の分類  
Table 2 Dialog act categories for *Rakugo*.

DA タグ (省略形)	機能
statement (st)	陳述する
opinion (op)	意見を述べる
calling (cl)	相手をよびかける
do-question (dq)	質問する (Do 質問)
wh-question (wq)	質問する (WH 質問)
reply (re)	よびかけに対して返事する
yes-answer (ye)	Do 質問に対して肯定する
no-answer (no)	Do 質問に対して否定する
wh-answer (wa)	WH 質問に対して回答する
ambiguous answer (am)	質問・依頼に対してあいまいな返答をする
agreement (ag)	陳述や意見に対して同意する
disagreement (dg)	陳述や意見に対して反駁する
command (cm)	依頼・命令する
tsukkomi (ts)	冗談を受け止める
sudden reaction (sd)	相手の発話に対する突発的反應をする(「あっ!」「えっ!」)

は後続発話を表すタグを付加した。なお、その他の時間に対しては、タグを付加しなかった。

発語内行為のタグ付加

発話音声と発話目的・意味との関連を解析する手法として、対話の発語内行為 (dialog act, DA) による分類があげられる。発語内行為とは、Searl らによって提案された、発話によって達成される行為 (主張, 質問, 命令など) を指す<sup>14)</sup>。DA はこの対話版であり、行為の分類方法については、対話の状況に応じてさまざまなものが提案されている<sup>10), 15)~17)</sup>。

本コーパスで用いた DA の分類を表 2 に示す。抽出した先行発話と後続発話に対して、これら DA をタグとして付加した。DA タグの定義においては、リーバイらによって定義された漫才対話 DA<sup>8), 9)</sup> を参考にし、落語にあわせて拡張した。発話時間が長い場合、1 つの発話中に複数の DA がみられる場合がある。この場合、先行発話については発話終了時にみられる DA をタグとして付加し、後続発話については発話開始時にみられる DA をタグとして付加した。なお、表 2 中の突発的反應とは、相手の発話に対する驚きや気づきなどから短い発話(「えっ!」など)を行うことを表す。

4.2 漫才対話コーパス

比較対象の二者間会話として漫才を用いる。漫才はボケ役、ツッコミ役と呼ばれる 2 人の演者による対話演芸であり、従来研究には、漫才における発話移行区間長を調べた研究がある<sup>8), 9)</sup>。本論文ではこの文献の漫才対話コーパスをベースとした(全 5 対話。表 3 参照こと)。 $\{a-d\}$  は漫才師を表し、c に関しては異なる

表 3 漫才対話コーパスにおける各対話  
Table 3 Dialogs in *Manzai* dialog corpus.

対話	対話の時間	ツッコミ役 → ボケ役	ボケ役 → ツッコミ役
a1	4 分 27 秒	64	67
b1	4 分 55 秒	109	107
c1	5 分 23 秒	122	121
c2	5 分 6 秒	117	121
d1	7 分 38 秒	88	88

表 4 漫才における DA の分類  
Table 4 Dialog act categories for *Manzai*.

DA タグ (省略形)	機能
statement (st)	陳述する
opinion (op)	意見を述べる
question (qs)	質問する
answer (an)	質問に対して回答を行う
agreement (ag)	陳述や意見に対して同意する
disagreement (dg)	陳述や意見に対して反駁する
command (cm)	依頼・命令する
tsukkomi (ts)	冗談を受け止める
continuer (cn)	相槌を入れる(「うん」など)
backchannel-question (bq)	質問的相槌を入れる(「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)

話題の 2 対話がある。各対話において、ボケ役およびツッコミ役の発話区間  $I(U_b^p, U_e^p)$ ,  $I(U_b^s, U_e^s)$  が手動で切り出され、発話ペアがラベリングされている。発話ペアのラベリングは、基本的には一方の発話終了付近に対して他方が発話を行う場合であるが、相手の発話中に相槌を行うなどの場合が若干含まれるため、両者の話者交替回数に違いが生じている。

漫才では落語に比べて役柄の違いが非常に明確であり、多くの発話では、ボケ役が冗談をいい出し、それに対してツッコミ役が強い語気でコメントを入れる。話題の導入部を除けば、ボケ役の発話に対してツッコミ役が応答する形で対話が進んでいく漫才師が多い。

漫才対話コーパスで用いられている DA の分類を表 4 に示す。表 2 における落語の DA とは分類が異なるため、比較のため両表に水平の罫線を用いている。これは、漫才が二者間会話であり、落語においては出現頻度が低い相槌などの発語内行為が存在すること、また漫才対話コーパスよりも落語役柄交替コーパスの方が、質問や回答の回数が多く、分類を細かくしているためである。

ただし漫才師によっては両役柄の違いが明確でないものや、各役柄の役割・対話のスタイルが異なるものもある。ここでは本文で述べるようなスタイルの漫才師を選んだ。

5. 落語の役柄交替タイミング構造の解析

本章では、3.4 節で述べた、落語の話者交替円滑化仮説を検証するために、4 章のコーパスを用いて、落語の役柄交替における頭部動作開始タイミング  $I(p_e, m_b)$  と、二者間会話（漫才）の発話移行区間長  $I(U_e^p, U_b^s)$  とを比較する。この解析における視点を以下に示す。なお、以下では時区間  $I(p_e, m_b)$  を役柄交替始動区間と呼ぶことにする。

- (1) 役柄交替始動区間長の分布の解析 (5.1 節): 役柄交替始動区間長  $I(p_e, m_b)$  と、二者間会話における発話移行区間長  $I(U_e^p, U_b^s)$  の頻度分布の形状や、オーバーラップ ( $I(p_e, m_b) \leq 0, I(U_e^p, U_b^s) \leq 0$ ) の割合を比較しながら、両者の類似性について調べる。
- (2) Dialog act と役柄交替始動区間長との関係 (5.2 節): 発話移行区間長は、会話の文脈や発話目的・意図によって異なった傾向が見られる (2 章)。そこで、落語の役柄交替始動区間長に關しても同様の特徴が見られるかを、特に後続発話を DA ごとに分類することで調べる。

なお、以下に示す統計量およびグラフは、演目名の明示がないかぎり、全演目のデータを用いている。

5.1 役柄交替始動区間長の分布の解析

図 6 に、役柄交替始動区間長  $I(p_e, m_b)$  の頻度分布を表す。 $I(p_e, m_b)$  の分布は、0 を中心としておよそ  $\pm 500$  msec の範囲で左右対称に広がる特徴を持つことが分かる。表 5 (上) は、個人差や演目差の影響を見るために、3 演目それぞれにおいて役柄交替動作の開始が先行発話にオーバーラップする ( $I(p_e, m_b) \leq 0$ ) 割合を示したものであるが、いずれにおいても約 30-50% であった。

一方、図 7 は役柄交替における発話移行区間長

$I(p_e, s_b)$  の頻度分布であるが、0 付近で最も頻度が高く、時間が長くなるに従い低くなる。なお、3.4 節の式 (7) の制約から、 $I(p_e, s_b)$  は負の値を持つことはない。役柄交替始動区間長  $I(p_e, m_b)$  に比べると、発話移行区間長  $I(p_e, s_b)$  は 500 msec を大きく超え、最大でおよそ 7,000 msec にまで分布する (ただし、数秒を越える場合は、観客の笑いを待つ、しぐさ・表情によって情景を伝えるなどの特殊な状況であった)。

次に、落語における役柄交替始動区間長  $I(p_e, m_b)$  と、漫才における発話移行区間長  $I(U_e^p, U_b^s)$  とを比較する。漫才では、ボケ役が後続話者になる場合、およびツッコミ役が後続話者になる場合があるため、それぞれの発話移行区間長の分布を図 8 (a), (b) に示す。ツッコミ役が発話タイミングを制御する際は、ボケ役に比べてよりオーバーラップをとまなう発話が多い

表 5 落語の役柄交替始動区間長および漫才の発話移行区間長が負または 0 となる割合 (カッコ内は頻度/総数)  
Table 5 The ratio of the overlaps of motion-beginning interval lengths in Rakugo and that of utterance transition intervals in Manzai (with the overlapped number / the total number).

落語		
演目	オーバーラップ割合 (頻度/総数)	
R1	37%	(39/106)
R2	31%	(20/64)
S1	49%	(29/59)
総和	38%	(88/229)

漫才			
対話	ツッコミ役 → ボケ役	ボケ役 → ツッコミ役	
a1	28% (17/60)	65% (40/62)	
b1	13% (13/101)	39% (39/99)	
c1	43% (51/118)	53% (62/116)	
c2	51% (52/102)	46% (48/104)	
d1	51% (40/78)	75% (55/73)	
総和	38% (173/459)	54% (244/454)	

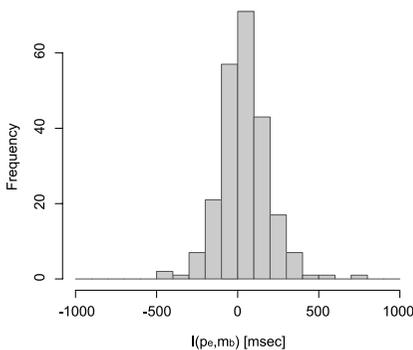


図 6 落語における役柄交替始動区間長  $I(p_e, m_b)$  の頻度分布  
Fig. 6 Histogram of the motion-beginning interval lengths  $I(p_e, m_b)$  in Rakugo.

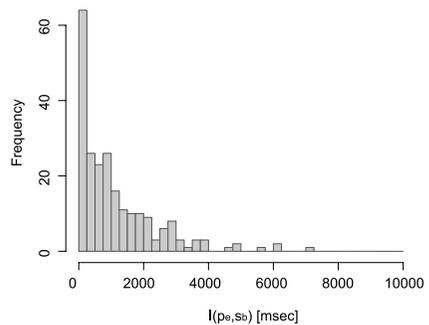
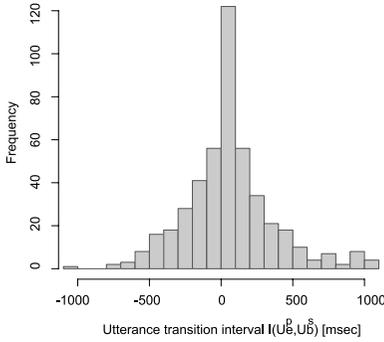
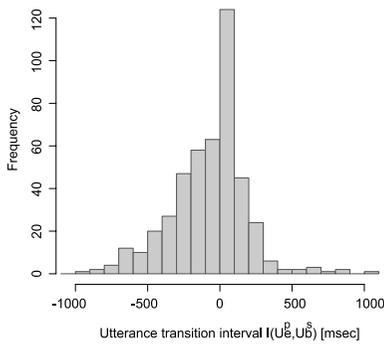


図 7 落語における発話移行区間長  $I(U_e^p, U_b^s)$  の頻度分布  
Fig. 7 Histogram of the utterance transition interval lengths  $I(U_e^p, U_b^s)$  in Rakugo.



(a) ツッコミ役発話終了に対するボケ役発話開始



(b) ボケ役発話終了に対するツッコミ役発話開始

図 8 漫才における発話移行区間長の頻度分布

Fig. 8 Histogram of utterance intervals in *Manzai*.

(表 5 (下)). 一方、ボケ役が後続話者である場合は、落語と同様、0 を中心に左右対称に広がる分布が見られた。このような役柄による違いがあるにせよ、落語の役柄交替始動区間長と漫才の発話移行区間長では、動きと音声という異なるモダリティの信号であるにもかかわらず、いずれもおよそ  $\pm 500$  msec の範囲に分布が集中している。

知見 (A)

本節の結果より、落語における役柄交替始動区間長には、漫才の発話移行区間長と同様に以下の特徴があるといえる。

- 役柄交替始動区間長はおよそ  $\pm 500$  msec の範囲に分布が集中する。
- 役柄交替始動区間長が負になる割合は無視できないほど高い (約 30–50%)。

さらにこれらの特徴は、解析対象としたデータの範囲内では、若干の演目差や演者の個人差はあるものの、大きな傾向としては変わらなかった。

考察 (A)

落語では、役柄交替始動区間長は 500 msec を超えることがほとんどないことから、発話のみでは冗長と

なるような長い発話移行区間長を、頭部動作による役柄切替えで解消している可能性がある。逆に、このような動作を入れることで、後続発話開始時刻  $T(s_b)$  を大きく遅らせたような、ゆったりとした会話情景の演出が可能になると考えられる。

先行発話中に頭部動作をオーバーラップさせることは、二者間対話における発話のオーバーラップと同様に、何らかの意図を伝えている可能性がある。しかし、頭部動作のタイミングが実際に発話意図と関連があるかどうかは、本節の結果だけでは不明である。そこで続く節では、DA ごとでの役柄交替始動区間長の傾向 (発話意図との関係) について解析を行う。

5.2 Dialog act と役柄交替始動区間長との関係

各役柄交替における後続発話を、表 2 の発話内行為 (DA) によって分類したときの、各 DA の出現頻度および役柄交替始動区間長  $I(p_e, m_b) \leq 0$  の割合を、表 6 の 1, 2 列目にそれぞれ示す。

まず各 DA の出現頻度に関しては、意見 (op)、つつこみ (ts) が多く、また、先行発話に対する受け答え (re, ye, no, wa, am) は、総和をとると最も多い。一方、役柄交替始動区間長は DA によって傾向が異なり、肯定的回答 (ye) や返事 (re) などの回答ではオーバーラップの割合が高く、否定的回答 (no)、あいまい回答 (am)、および反駁 (dg) ではポーズの傾向がある。同意 (ag) は、反駁 (dg) に比べればオーバーラップする割合が高い。質問に関する DA {dq, wq} は比較的オーバーラップの割合が低く、つつこみ (ts) はさらに低い。

落語における役柄交替始動区間長と DA との関係を、漫才における発話移行区間長と DA との関係と比較するために、まず DA の再分類を行う。これは、漫才と落語の DA が単に異なるということ以上に、DA 付与作業者の基準差の影響を減らすためである (4.2 節)。DA の分類が細かすぎる場合、DA タグを付与する作業者の影響が大きくなる。そこで、作業者間での一致率が高いタグ体系として、文献 16) の標準化案タグセットを参考にし、これよりもさらに粗いタグセットとするために、以下のように分類し直した。

まず、落語の {cl, dq, wq} を漫才の質問 (qs) に対応させた。次に、落語の回答は肯定とそれ以外で傾向が異なるため、コーパス中の漫才の回答 (an) を、肯定的回答 (an-p) と非肯定的回答 (an-np) に再分類し、それぞれに対して落語の {re, ye} と {no, wa, am} を

呼びかけ (cl) は質問ではないが、相手の返答を要求する発話として、ここでは質問に併合した。

表 6 各 DA の特徴 (出現頻度, オーバラップ ( $m_b \leq p_e$ ), ( $s_b \leq m_b$ ) の割合,  $I(p_e, s_b)$  中央値 [msec])  
 Table 6 Feature of each DA (the total number, the overlap ratio of ( $m_b \leq p_e$ ) and ( $s_b \leq m_b$ ), and the median of  $I(p_e, s_b)$ [msec]).

落語 DA (頻度)	$m_b \leq p_e$ の割合 (数)	$s_b \leq m_b$ の割合 (数)	$I(p_e, s_b)$ の中央値	漫才 DA (頻度)
st (9)	44% (4)	22% (2)	840	st (64)
op (37)	41% (15)	0% (0)	840	op (30)
cl (6)	50% (3)	0% (0)	924	qs (10)
dq (10)	30% (3)	0% (0)	1,460	
wq (13)	38% (5)	0% (0)	1,760	
re (13)	77% (10)	0% (0)	776	an-p (11)
ye (12)	75% (9)	0% (0)	516	an-np (6)
no (5)	0% (0)	0% (0)	1,264	
wa (27)	41% (11)	0% (0)	536	
am (9)	11% (1)	0% (0)	1,224	ag (36)
ag (18)	50% (9)	0% (0)	444	dg (28)
dg (14)	29% (4)	0% (0)	1,396	cm (12)
cm (8)	38% (3)	0% (0)	692	ts (175)
ts (34)	9% (3)	41% (14)	81	なし
sd (14)	57% (8)	7% (1)	348	

対応させた (表 6 の 5 列目参照) . 依頼・命令 (cm) , つっこみ (ts) はそのままとした . さらに, 文献 16) では陳述 (st) と意見 (op) が明確に分類されているわけではないため, 以下の解析では, 落語および漫才の {st, op} はいずれも 1 つにまとめるものとする .

図 9 に, 落語の役柄交替始動区間長  $I(p_e, m_b) \leq 0$  の割合と, 漫才の発話移行区間長  $I(U_e^p, U_b^s) \leq 0$  の割合を, 再分類した DA ごとに比較して示す . ここで, 比較対象とする漫才の発話移行区間長としては, 先行話者がボケ役, 後続話者がツッコミ役となるものを用いた . これは, 4.2 節で述べたように, 漫才では落語に比べて役柄間の違いが大きく, 特に発話の意味的なつながりを考えると, ボケ役の発話に対してツッコミ役が応答する発話ペアが多いからである . 以下に, この図より得られる知見を述べる .

知見 (B)

- an-p と an-np, および ag と dg を比較すれば, 漫才においても, 肯定的な受け答えである肯定的回答 (an-p) が非肯定的回答 (an-np) に比べて先行発話をオーバラップする傾向が強く, 同意 (ag) も反駁 (dg) に比べて同様の傾向がある (特徴 1) .
- {st+op, qs} を比較すると, 漫才の質問 (qs) は

文献 16) では「肯定・受領」, 「否定・拒否」に分類されているが, 質問に対する回答か, 陳述・意見に対する同意・反駁かの区別はなかった . これらは明確に区別できそうであり, 本論文では別の DA としてまとめた .

依頼・命令 (cm) は, 漫才では命令の傾向が強く, 一方落語では依頼であり質問に近い意味合いがある . 依頼と命令を DA として分けるべきであったが, 今回は考察対象外とした .

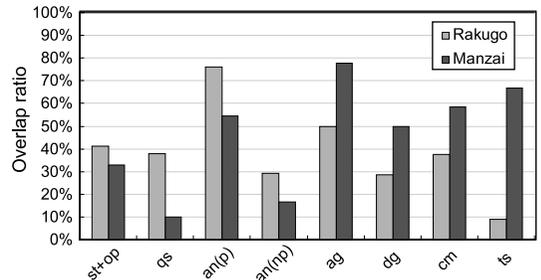


図 9 落語と漫才の DA ごとでのオーバラップ割合の比較  
 Fig. 9 Comparison between Rakugo and Manzai in terms of the overlap ratio of each DA.

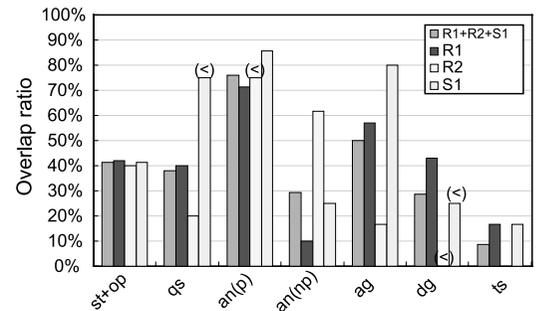


図 10 落語の 3 演目間の比較 . (<) は発話頻度が 5 未満のデータ  
 Fig. 10 Comparison between the three Rakugo performances. (<) denotes the utterance data which sample size is smaller than five.

比較的遅いタイミングで反応をしている . 落語では同様の傾向がわずかに見られる (特徴 2) .

- つっこみ (ts) は, 漫才では非常にオーバラップの割合が高く, 落語では逆に極端に低い (特徴 3) .

ただし, 上記の特徴は, コーパスサイズを考慮すれば十分な信頼性を持つものとはいえない . さらに, これには個人差や演目による差が存在する可能性もある . 図 10 は, 各演目ごとに, 上記で注目している DA でのオーバラップ割合を示したものであるが, たとえば S1 における質問 (qs) のオーバラップ割合は比較的高いため特徴 2 と異なる (ただし qs 自体の頻度が非常に小さい) . 一方, 特徴 1 や 3 についてはそれぞれの演目内でも同様の傾向が見られた .

つっこみ (ts) に関しては, 絶対数が多いにもかかわらず, 落語と漫才とでタイミング構造が大きく異なる . そこで, 落語において, 先行発話終了に対する役柄交替動作開始 (役柄交替始動区間長)  $I(p_e, m_b)$  ではなく, 役柄交替動作開始に対する後続発話開始  $I(m_b, s_b)$  に注目し, 後続発話が発話交替動作をオーバラップする  $I(m_b, s_b) \leq 0$  の割合を表 6 の 3 列目に示す . 表より,  $I(m_b, s_b) \leq 0$  となる DA は, ほぼつっこみ (ts) に限られることが分かる . さらに, DA ごとでの発話

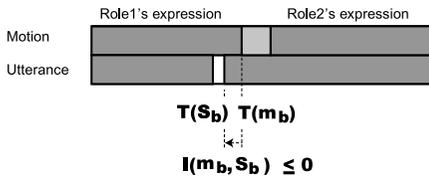


図 11 つっこみ (ts) の特殊な役柄交替タイミング構造

Fig. 11 Special turn-taking timing structure of tsukkomi (ts).

移行区間長  $I(p_e, s_b)$  の中央値 (表 6 の 4 列目) から、つっこみ (ts) においては他の DA に比べてほとんど発話の間をあけていないことが分かる。これらの結果より、つっこみ (ts) では、図 11 に示すような特徴的なタイミング構造を用いていることになる。

### 考察 (B)

特徴 1 の解釈としては、前向きな受け答えにおいて返答が遅れたときに、聞き手には回答に迷っていると解釈されるため、落語家は素早い返答を演出する必要があること、一方、否定的な受け答えにおいては、即時的に返答したときに熟慮していないと解釈されるため、間をおいた返答を演出する必要があることが考えられる (これは文献 3) の考察とも一致する)。特徴 2 に関しては、個人差や演目差があるものの、質問を考える思考時間を演出するために、ややタイミングを遅らせている可能性がある。これらのタイミング制御を、漫才では後続話者の発話開始によって、落語では頭部動作の開始によって行っていると考える。ただし、落語のつっこみ (ts) 発話では、頭部動作ではなく発話開始を動作開始よりも早めることで、相手の発話を途中でさえぎる様子演技していると思われる。

なお、今回得られた知見は限定的な数のデータから得られたものであり、今後より大量のデータに基づいた詳細な解析が必要である。

## 6. 心理実験による視覚刺激の効果の検証

5 章の考察 (B) のような、頭部動作の開始タイミングと発話意図との関連性は、現時点では予測の段階にとどまっている。一方、考察 (A) のように、頭部動作という視覚刺激を用いて間合いの冗長性を解消している可能性は高い。ただし、心理的に実際に期待する効果を与えるかどうかは、コーパスを用いた解析だけでは難しい。そこで、遠隔対話で伝送遅延が生じている状況を想定し、間合いを補間するための視覚刺激を話者に提示するような被験者実験を行うことでその効果を検証する。

### 6.1 手続き

本実験では、二者間会話における 1 度の話者交替を



図 12 視覚刺激の例 (被験者が見る相手話者の映像に対して重畳表示)

Fig. 12 An example of visual stimuli (overlaid on the video to which each subject speaks).

想定した。被験者は、椅子に座った状態でディスプレイに表示された相手話者に発話を行うものとし、その被験者の発話に対して、相手話者が返答を行うものとする。このとき、被験者の発話終了時刻に対する相手話者の発話開始時刻 (発話移行区間長) について、被験者に話者交替の不自然さを評価させた。

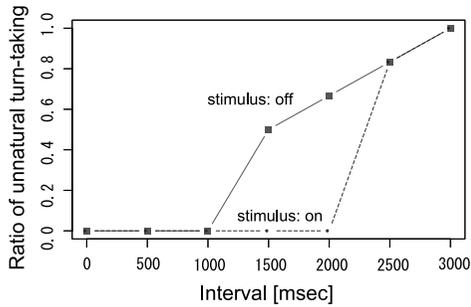
このとき、相手話者の映像に対して視覚刺激を挿入し、挿入しない場合との被験者の評価を比較した。用いる視覚刺激は、開始タイミングが話者に意識しやすいものである必要がある。そこで、回転動作するアイコンを映像に挿入することとした (図 12)。

評価方法としては恒常法を用いた。すなわち、被験者と相手話者との 1 度の話者交替を反復し、反復のたびに発話移行区間長を変動させ、被験者に不自然さを評価させた。発話移行区間長は、0, 500, 1,000, 1,500, 2,000, 2,500, 3,000 msec の 7 段階とした。さらに、1,500 msec 以上の 4 つの段階については、視覚刺激を挿入する場合を条件に加えた。以上の 11 条件について、各条件が 6 回ずつランダムに提示されるように、話者交替を計 66 回反復した。各反復において、被験者は「もっと早く返答してほしいと思ったか否か」についての二者択一により評価した。

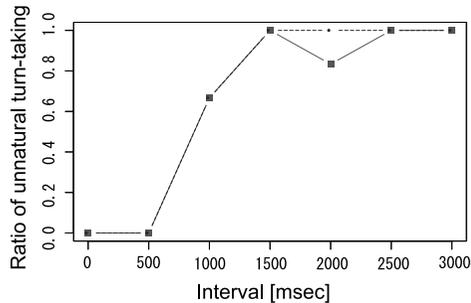
なお、視覚刺激の挿入する時間は、被験者の発話終了時刻から 1,000 msec 経過した時点から、相手話者の発話開始時刻までとした。また、同じ条件の試行を反復するために、相手話者の映像には、あらかじめ撮影した動画を用い、その再生開始タイミングを変化させた。

### 6.2 条件

被験者は全部で 7 人であり、このうち 5 人に対し、視覚刺激の効果の評価する実験を行った (6.3 節に結果を示す)。次に、7 人中 4 人に対し、視覚刺激の種類を変えた場合の比較評価を行った (6.4 節)。被験者はすべて映像中の相手話者と親しい人間とした。これは、あらかじめ被験者と相手話者で、間合いの共有が



(a) 被験者 A



(b) 被験者 E

図 13 恒常法による評価結果 (5 人中 2 人の結果を示す)

Fig. 13 Evaluation by using the constant method (the results of two subjects).

行われているという仮定を設けるためである。

被験者と相手話者の発話内容を以下に示す。

- 被験者：休暇中に行ったことの報告
- 映像中の相手話者：「へえ、そうなんだ」

発話内容によっては、発話移行区間長の変化により、相手話者の発話の意味や目的が、被験者に異なって理解される可能性がある（たとえば「それはいいね」などの肯定、否定がある場合）。そのため本実験では、相手話者の発話への印象が、発話移行区間長の変化にそれほど依存しないと考えられる、単に報告を受理するような中立的な発話内容を選択した。

### 6.3 視覚刺激の効果

被験者各自について、相手話者の発話に対して早く返答してほしいと思った割合を算出したところ、5 人中 4 人の被験者では、視覚刺激の提示によって、早く返答してほしいと思った割合は低下した。その例を図 13(a) に示す。

グラフ中の横軸は発話移行区間長、縦軸は被験者が「早く返答してほしい」と評価した割合、すなわち話者交替の不自然さを表す。四角および丸（それぞれ実線および破線でつないでいる）は、視覚刺激の提示の有、有の条件に対する値をそれぞれ表す。ただし、視覚刺激の提示は 1,500 msec 以上の条件について行っ

表 7 視覚刺激の挿入による不自然さの低下度（自然さの改善度の評価）

Table 7 The reduction of unnaturalness rate when the stimulus is inserted (evaluation of the improvement of naturalness).

被験者	Interval [msec]			
	1,500	2,000	2,500	3,000
A	0.50	0.67	0	0
B	0.33	0.50	0	0
C	0	0	0.83	0.50
D	0.33	0.33	0	0
E	0	-0.17	0	0

ているため、1,000 msec 以下の条件の数値は、提示の有無にかかわらず同じ値である。

ここで、視覚刺激提示がない条件での、話者交替が不自然であった（早く返答してほしいと思った）割合から、提示がある場合の割合を引いたものを、便宜上、低下度と呼ぶことにし、不自然さが減少した（自然さが改善した）度合いを表す値として用いることにする。表 7 は、被験者ごとの、1,500 msec 以上の条件における不自然さの低下度を表す。

表 7 より、被験者 A, B, D の 3 人については、1,500 msec と 2,000 msec の 2 つの条件について、不自然さに低下が見られる。とくに、1,500 msec の条件において、不自然さは 0 になった。一方、2,500 msec 以上の条件においては、低下度は 0 であり、不自然さは 1 に近い値を示した。また、被験者 C については、2,000 msec 以下の条件において、視覚刺激提示の有無によらず、不自然さは 0 であった。一方、2,500 msec 以上の条件において不自然さは増加したが、視覚刺激の提示によって 0.5 ポイント以上低下した。

一方、被験者 E のみ不自然さに低下が見られない。ここで、被験者 E の評価結果を図 13(b) に示す。被験者 E の不自然さは、1,000 msec 以上の条件において 0 から上昇しているが、本実験では、被験者の先行発話終了時刻から 1,000 msec の時点で視覚刺激を提示している。したがって、被験者 E は、刺激提示の開始時刻もしくはそれ以前において、すでに不自然さを感じていたことになる。

### 6.4 視覚刺激の種類の影響

アイコンの回転動作を突発的に開始するのではなく、常時、点滅や回転をさせておき、点滅や回転のスピードをそれぞれ速めるような視覚刺激についても、低下度に基づいて定量的に評価を行った。被験者としては、映像刺激の効果があつた 2 人（被験者 A, D）を残し、あと 2 人（被験者 F, G）を新たに追加して同様の実験を行った。

この結果を表 8 に示す。被験者 G を除けば、突発

表 8 各映像刺激による不自然さの低下度

Table 8 The reduction of unnaturalness rate when each of the three stimuli is inserted.

被験者	映像刺激	Interval [msec]			
		1,500	2,000	2,500	3,000
A	突発的	0.50	0.67	0.17	0
	常時回転	0.33	0	0	0
	常時点滅	0.50	0.67	0.17	0
D	突発的	0	0.5	0.17	0
	常時回転	-0.17	0	0	0
	常時点滅	0	0.5	0	0
F	突発的	0	0.67	1.00	0.5
	常時回転	0	0.33	0.67	0.17
	常時点滅	0	0.50	0.83	0.17
G	突発的	1.00	0	0	0
	常時回転	0.67	0.17	0	0
	常時点滅	0.50	0.17	0	0

的な開始、点滅スピードの変化、回転スピードの変化の順に、不自然さの低下度は大きかった。被験者 G では 2,000 msec の条件で突発的な刺激提示の効果がなくなっているが、他の刺激ではやや効果があった。

### 6.5 考 察

7人中6人の被験者では視覚刺激の提示によって不自然さの低下（自然さの改善）が見られたが、効果の持続時間はおよそ1秒程度であった。これは、視覚刺激の提示によって解消できる間合いの長さには上限があることを示唆している。発話の遅れが大きいときには、数秒ごとに動きや刺激を切り替えるなど、視覚刺激を適切にデザインする必要があるといえる。

7人中1人の被験者については視覚刺激による不自然さの低下が見られなかった。これは、この被験者は視覚刺激の生成時刻（1,000 msec）においてすでに不自然さを感じていたことが原因といえる。不自然さを感じる間合いについては個人差や文脈の影響が大きく、視覚刺激の提示タイミングを適応的に決定する方法を検討する必要がある。

### 7. おわりに

本論文では、会話においては身体動作のような視覚情報を相手に提示することによっても、音声発話と同様に自然な発話極移動の間合いを表出できる、という仮説を検証する1つの具体的方法として、演者が1人であるにもかかわらず、自然な役柄交替を演じている落語の解析を行った。その結果、落語役柄交替コーパス、漫才対話コーパスのサイズ（演目数、演者数、発話数）についてはいずれも改善の余地があるものの、特に、落語の役柄交替における頭部動作のタイミング制御と、二者間会話での発話によるタイミング制御には、5章の知見(A)および知見(B)で述べた類似性が

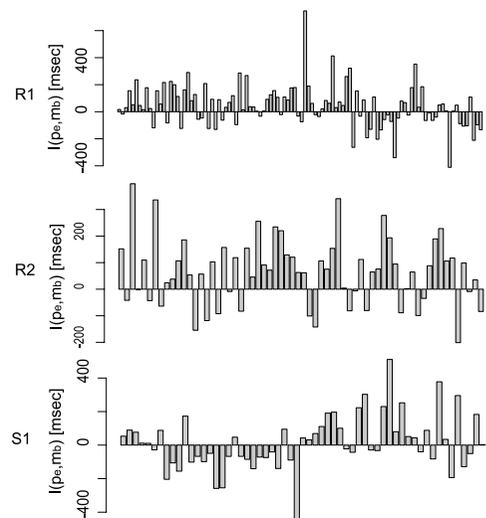


図 14 各演目内での役柄交替始動区間長の長期的変動

Fig. 14 Long-term change of motion-beginning intervals during each performance.

見られ、それぞれ考察を行った。

今回用いたコーパスサイズであっても、考察(A)で予測したように、視覚刺激が間合いの冗長さを補間している可能性は高い。そこで、恒常法に基づく心理実験を7人の被験者で行った。その結果、相手の発話タイミングが遅れる状況で視覚刺激を提示することで、多くの被験者で「もっと早く返答してほしい」という間合いの不自然さを軽減することができ、視覚刺激の効果を確認することができた。一方、発話意図との関係に関する知見(B)は、コーパスサイズを考慮すればまだ十分な信頼性があるとはいえず、その考察(B)は現時点では予測に近い。

さらに、今回の解析で扱わなかった観点として、会話の流れ（盛り上がりなど）がある。漫才では、発話移行区間長が長期的に変動するという知見があるが<sup>9)</sup>、落語においても、図14に示すように、役柄交替始動区間長  $I(p_e, m_b)$  が正に偏る場面や負に偏る場面がある（縦軸と横軸は、それぞれ役柄交替始動区間長と、役柄交替場面の順序を表す）。

今後は、コーパスを充実させながら今回得られた知見をより詳細に分析していくとともに、DAの分類・アノテーション方法の検討や、長期的変動といった別の視点からの解析、通常の会話との比較、実際の会話支援システムを用いた評価などを行う必要がある。

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

## 参 考 文 献

- 1) Shirberg, E.: Spontaneous Speech: How People Really Talk, and Why Engineers Should Care, *Proc. EUROSPEECH* (2005).
- 2) 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).
- 3) 藤原敬記, 伊藤敏彦, 荒木健治: タスク指向対話における発話意図の対話リズムへの影響, 情報処理学会研究報告 (2007-SLP-66), pp.37–42 (2007).
- 4) Duncan, S.J. and Fiske, D.W.: *Face-to-Face Interaction*, chapter 11, Lawrence Erlbaum (1977).
- 5) 小磯花絵, 伝 康晴: 円滑な話者交替はいかにして成立するか—会話コーパスの分析に基づく考察, 認知科学, Vol.7, No.1, pp.93–106 (2000).
- 6) 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).
- 7) 長岡千賀, Draguna, M., 小森正嗣, 中村敏枝: 音声対話における交替潜在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).
- 8) スコギンズ・リーバイ, 川嶋宏彰, 松山隆司: 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, *インタラクション 2005*, D-404 (2005).
- 9) 川嶋宏彰, スコギンズ・リーバイ, 松山隆司: 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, *ヒューマンインタフェース学会論文誌*, Vol.9, No.3, pp.379–390 (2007).
- 10) 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).
- 11) Kendon, A.: Some Function of Gaze-direction in Social Interaction, *Acta Psychologica*, Vol.26, pp.22–63 (1967).
- 12) 野村雅昭: *落語の言語学*, 平凡社 (2002).
- 13) Kipp, M.: *Gesture Generation by Imitation – From Human Behavior to Computer Character Animation*, Ph.D. Thesis, Boca Raton (2004).
- 14) J.R. サール(著), 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).
- 15) Allen, J. and Core, M.: Draft of DAMSL: Dialog Act Markup in Several Layers (1997).
- 16) 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人: 発話単位タグ標準化案の作成, *人工知能学会論文誌*, Vol.14, No.2, pp.251–260 (1999).
- 17) Dhillon, R., Bhagat, S., Carvey, H. and Shriberg, E.: Meeting Recorder Project: Dialog Act Labeling Guide, *ICSI TR-04-002* (2004).

(平成 19 年 4 月 2 日受付)

(平成 19 年 9 月 3 日採録)



川嶋 宏彰

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。同大学院助手を経て 2007 年より同大学院講師。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。FIT 論文賞 (2004), 船井ベストペーパー賞 (2005), FIT2006 ヤングリサーチアワード (2007)。



西川 猛司

2007 年京都大学大学院情報学研究科修士課程修了。在学中は視覚的問合いの解析と有効性の検証に関する研究に従事。



松山 隆司 (フェロー)

1976 年京都大学大学院修士課程修了。京都大学助手, 東北大学助教授, 岡山大学教授を経て 1995 年より京都大学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, 京都大学評議員。2005 年情報環境機構長。工学博士。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, 1996 年国際パターン認識連合 Fellow, 1999 年電子情報通信学会論文賞。情報処理学会フェロー。