TR-IT-0161

# 韻律を用いた発話アクトの識別
# Identification of Communicative Acts using Prosody

藤尾 茂
Shigeru Fujio

ニック キャンベル
Nick Campbell

樋口 宜男
Norio Higuchi

1996.3.29

対話型応答システムの制御や音声翻訳を行なう上で、発話アクトや発話意図を理解することが重要である．文字列を認識するだけで発話アクトや発話意図を識別できることもあるが、同じ文字列で数種類の意図が考えられる場合は韻律や文脈などを利用する必要がある．我々は同一表記で複数の発話アクトの可能性がある発話に対して韻律を用いて発話アクトの識別を行なう方法を提案し，その方法を用いた発話アクトの識別実験を行なった．本報告ではその結果について述べる．

## 1 はじめに

対話型応答システムの制御や音声翻訳を行なう上で、発話アクトや発話意図を理解することが重要である．文字列を認識するだけで発話アクトや発話意図を識別できることもあるが、同じ文字列で数種類の意図が考えられる場合は韻律や文脈などを利用する必要がある．例えば，「そうですか」という文字列が発声された時，文末で基本周波数が上昇していれば，疑問を意図しており，下降していれば，話し手に対するあいづちを意図したものである．対話型応答システムや音声翻訳で，ユーザが「そうですか」と発声した場合，発話アクトが疑問であれば発話権は移動するが，そうでない場合には必ずしも発話権の移動が起こるとは限らない．このように，発話アクトを識別することにより，対話型応答システムや音声翻訳の談話管理を円滑に行なうことができる．我々は発話アクトの識別における韻律利用の可能性を調べるために，同一表記で複数の発話アクトの可能性がある発話の生起頻度の分析を行ない，それらの韻律的な違いの分析を行なった．また，これらの結果を基に任意の文字列の発話に対して適用可能な韻律を用いた発話アクトの識別方法を提案し，その方法を用いた発話アクトの識別実験を行なった．本報告ではその結果について述べる．

## 2 発話アクト

本報告ではＡＴＲで使用されている 27 個の対話行為 (Communicative Acts, CA) ラベル [1] を用いて発話アクトを分類した．本報告中で使用されている発話アクトの説明とそれらに対して考えられるシステムの動作を表 1に示す．

## 3 韻律を用いた発話アクトの識別

日常の会話において人間は同一表記で複数の意図が考えられる発話でも問題なく聞き分けていると思われる．これは，人間がこれらの発話の韻律的な特徴の違いや文脈の流れなどを使って識別しているからだと考えられる．我々はこれらの情報の内の韻律の違いのみに着目して研究を進める．韻律を用いた発話アクトの識別を考える場合，まず，韻律パラメータの発話アクト間での違いを分析し，次に分析の結果得られた違いを用いて識別を行なう必要がある．本報告では聴取実験により韻律全体の発話アクト間での違いの調査を行なうとともに，個々の韻律パラメータについては基本周波数と時間長のみに着目して違いの分析を行なった．

表 1 発話アクト

| 発話アクト名 | 説明 | システム動作 |
|---|---|---|
| Inform | 話し手は聞き手に情報を与える | 応答 / 待機 |
| Wh-question | 話し手が聞き手にいつ，どこ，だれが，いかに，だれにを尋ねる | 応答 |
| YN-question | 話し手が聞き手に「はい」「いいえ」で答えられる質問をする | 応答 |
| Confirmation -question | 話し手が聞き手に確認をする | 応答 |
| Confirmation -question-to-self | 話し手が知り得たことを１人つぶやく | 待機 |
| Acknowledge | 聞き手が話し手に談話を継続するためにあいづちをうつ | 応答 / 待機 |
| Yes | YN-question の肯定応答 | 応答 / 待機 |
| Temporizer | 話し手が躊躇を示す | 待機 |

## 4 複数の CA ラベル候補を持つ発話の分析

### 4.1 複数のＣＡラベル候補を持つ発話の生起頻度

自由発話データベース [2] の 51 会話中の複数の CA ラベル候補を持つ表現とその表現を持つ発話の出現頻度を調べた．この結果，次のような表現の発話が該当した．括弧内の数字は出現回数を示す．

- 文末に「ですか」を持つ発話
  1. そうですか (49)
     a) Acknowledge(38)
     b) Temporizer(11)
  2. ～ですか（そうですかを除く）(66)
     a) Wh-question(26), YN-question(25), Confirmation-question(8)
     b) Confirmation-question-to-self(7)
- 文末に「ですね」を持つ発話
  1. そうですね (30)
     a) Acknowledge(4), Yes(4)
     b) Temporizer(22)
  2. ～ですね（そうですねを除く）(78)
     a) Confirmation-question(59)
     b) Inform(19)

Wh-question, YN-question, Confirmation-question の 3 つの CA ラベルと Acknowledge, Yes の 2 つの CA ラベルはこれらの発話に対するシステムの動作は同様であり，発話意図も類似しているので本報告では 1 つのグループとして扱い，それぞれ Question, Acknowlege とする．なお，「そうですか」においては YN-question，「そうですね」においては Confirmation-question が CA として考えられるが本データベースには出現しなかった．

### 4.2 聴取実験による CA ラベルの識別

程度の正確さで識別できるかを調べるために，単独の発話を聞いてＣＡラベルを識別する実験を行なった．被験者は 10 人（男性 3 名，女性 7 名）である．結果を表 2 に示す．「そうですね」以外はどれも 70 ～ 80% 程度は正しく CA ラベルを識別することができ，聴覚的に区別可能な違いがあることがわかった．なお，「そうですね」は全体的に CA ラベルを Acknowledge/Yes と識別する傾向にあり，Temporizer の正解率が低いが，この Temporizer の中にも正解率が 70% を越えるものが約 20% 含まれており，これらには聴覚的に区別可能な違いがあると考えられる．

表 2 聴覚による CA ラベル識別の正解率

| | 正解率 | | |
|---|---|---|---|
| そうですか | 82.2 % | Acknowledge | 80.5 % |
| | | Temporizer | 91.2 % |
| ～ですか | 76.8 % | Wh/YN/Conf-question | 90.0 % |
| | | Conf-question-to-self | 61.8 % |
| そうですね | 49.0 % | Acknowledge/Yes | 82.0 % |
| | | Temporizer | 38.0 % |
| ～ですね | 76.8 % | Conf-question | 90.5 % |
| | | Inform | 76.7 % |

## 4.3 音響分析による CA ラベル間の韻律パラメータの比較

「ですか」や「ですね」を文末に持つ発話について各 CA ラベル間の違いを調べるために基本周波数 $F_0$ と時間長の分析を行なった．時間長の分析には正規化された値である z-score[3] を使用した．

### 4.3.1 そうですかの分析

「そうですか」の分析の結果，図 1に示す例のように Temporizer の場合はためらいの気持ちがあるために Acknowledge に比べて /o,u/ と /a/ の時間長が長くなることがわかった．発話速度が遅くなったために /o,u/ と /a/ の時間長が長くなることも考えられるので /o,u/，/a/ それぞれの z-score から CA ラベル間で違いがなかった /e/ の z-score を差し引いた値を /o,u/，/a/ の伸長を表すパラメータとした．これらのパラメータの各ラベルごとの分布を図 4に示す．

### 4.3.2 「ですか」を持つ発話の分析

「ですか」を文末に持つの発話（「そうですか」を除く）の分析の結果，図 2に示す例のように Conf-q-to-self は他の CA に比べて /k/ の時間長が長くなることがわかった．図 4に /k/ の z-score から /e/ の z-score を引いた値の各ラベルごとの分布を示す．また，文末が無声化して抽出不可能な発話が多くデータが非常に少ないので，参考程度ではあるが Conf-q-to-self は他の CA に比べて，文末の $F_0$ が下降する傾向があった．これは，疑問の意図の有無によるものだと思われる．

### 4.3.3 そうですねの分析

「そうですね」の分析の結果，CA ラベル間であまり有意な違いは見つからなかったが，聴覚による識別実験では Temporizer と回答された割合と /ou/，「ね」の /e/ の時間長に相関が見られた．

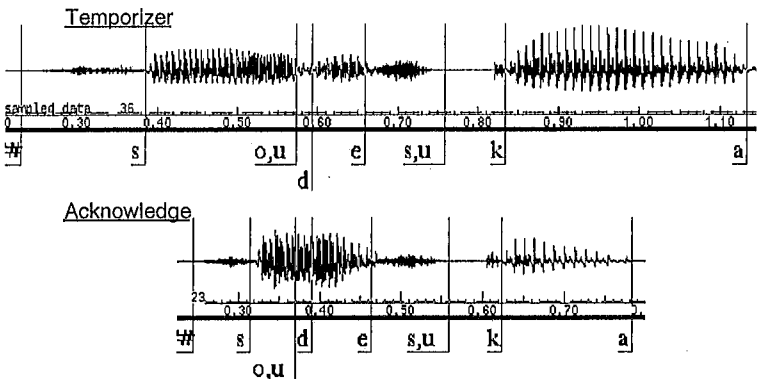



図 1 そうですかの CA の違いによる時間長の違い

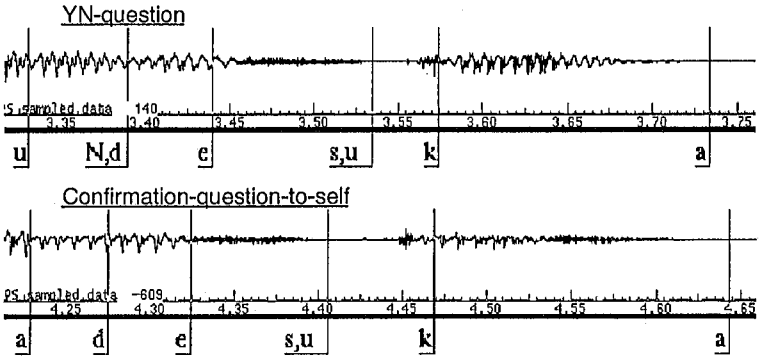



図 2 ですかの CA の違いによる時間長の違い

### 4.3.4 「ですね」を持つ発話の分析

「ですね」を文末に持つ発話（「そうですね」を除く）の分析の結果，図4に示すようにConf-questionは「で」の /e/ の中心の $F_0$ から「ね」の /e/ の中心の $F_0$ を差し引いた値がInformに比べて小さくなることがわかった．これはConf-questionの場合，疑問の意図があるため図3に示す例のように文末で $F_0$ が上がる傾向があるためである．

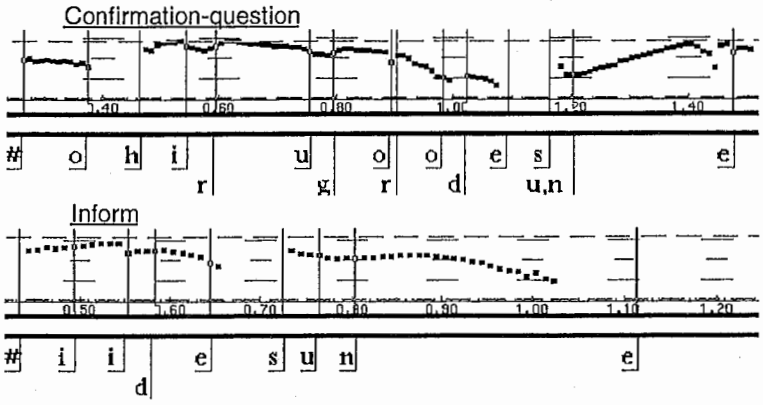
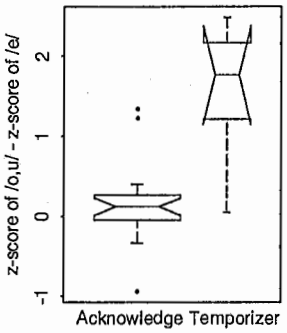
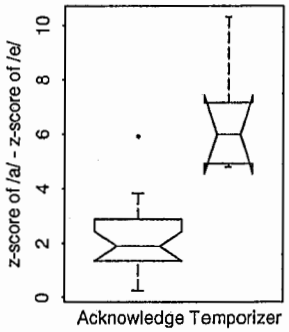


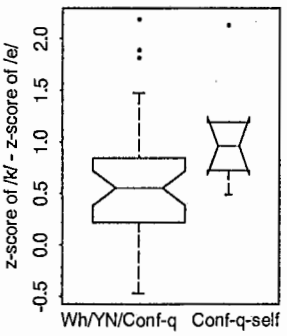

図 3 ですねのCAの違いによる $F0$ パターンの違い



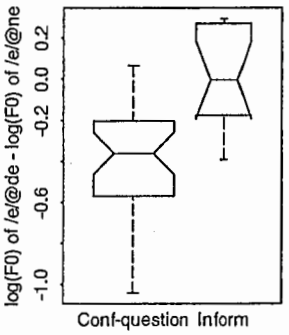

「そうですか」の /o,u/ の時間長のz-scoreから /e/ の時間長の z-scoreを引いた値の分布の違い




「そうですか」の /a/ の時間長のz-scoreから /e/ の時間長の z-scoreを引いた値の分布の違い




「ですか」の /k/ の時間長の z-scoreから /e/ の時間長の z-scoreを引いた値の分布の違い




「ですね」の「で」の /e/ の log($F_0$) から「ね」の /e/ の log($F_0$)を引た値の分布の違い

図 4 各CAラベル間の分布の違い

## 5 テキスト非限定な発話アクトの識別

前節までの結果を基にした「そうですか」という文字列に対してのTemporizerの識別方法，文末に「ですか」や「ですね」を持つ発話に対してのQuestionの識別方法が考えられる．しかし，これらの方法では同じ発話アクトを扱う場合であっても任意の文字列の発話に適用できない．対話型応答システム等のアプリケーションへの適用を考えた場合，限られた文字列だけでなく任意の文字列に対応できることが必要となる．したがって，文字列に制限のない発話アクトの識別方法を確立する必要がある．そこで，任意の文字列からなる発話に対してTemporizerとQuestionを検出する手法を検討する．

### 5.1 Temporizerの識別方法

Temporizerの「そうですか」の場合には，「そう」の/ou/と「か」の/a/の時間長が長くなることが確認されたことから，躊躇やためらいがある発話には時間長の長い音素が出現する傾向があると考えられる．また，経験的に，伸長される音素は文字列によって異なっていると思われる．Temporizerには非常に多くの文字列の発話が考えられるので，伸長される音素とそれが伸長しているか否かの閾値をすべての文字列の発話に対して個別に求めることは不可能である．

そこで，本報告ではテキスト音声合成システムによって予測された時間長を標準として各音素毎に標準との時間長の比を求め，その最大値を入力として判別分析によりTemporizerの識別を行なう．テキスト音声合成システムとしてはATR $\nu$-Talk音声合成システム[4]を用いる．ただし，ATR $\nu$-Talkの時間長の予測値は朗読調の場合のものなので，全発話の各音素ごとに実際の発話の平均値とATR$\nu$-Talkの予測値の平均値が一致するように係数を掛けて補正を行なう．また，Temporizerは躊躇やためらいの発話であり，長い発話は出現しないことが予備的検討で確認されたため，発話の長さが10モーラ以下の場合についてのみ識別を行なうこととした．

### 5.2 Questionの識別方法

我々は日常会話において文字列から明らかに疑問とわかる場合を除き，文末の基本周波数の上昇などの韻律情報や文脈の流れなどを用いて疑問か否かを識別している．また，文末に「ですか」や「ですね」を持つ発話の分析の結果から，文末の基本周波数パタンの傾きがQuestionの識別に有効な要因であることも，既に確認されている[?]．一般に「です，ます」調の発話ではこのような終助詞が付くことが多いが，くだけた発話では同一の文字列を用いて，韻律の変化によって異なる発話アクトを表す傾向が見られる．

文末の韻律を表すパラメータとして以下のものを用いる．

- 文末の2音素の区間内で抽出された基本周波数の回帰直線の傾き
- 文末の1音素の時間長のz-score[3]
- 文末の1音素の区間内で抽出されたパワーの回帰直線の傾き

## 6 発話アクト識別実験

提案するTemporizerおよびQuestionの識別方法の有効性を確認するために識別実験を行なった．

### 6.1 使用データ

実験には2種類のデータを用いた．1つは自由発話データベース[2]（以下，旅行会話とする）であり，もう1つは，くだけた発話の収集を目的として新たに作られたものである．後者は2人の親近度の高い話者（関東地区出身）に雑談する時のような口調で構わないと指示を与えた上で，表裏にまたがる迷路の表側のみの情報を片方の話者に，裏側のみの情報をもう一方の話者に与え，お互いに情報交換しながら迷路を解いてもらい収録した（以下，迷路会話とする）．

旅行会話にはTemporizerの発話は出現するが，話者間の親近度が低いために終助詞を伴った丁寧な表現の疑問の発話しか出現しなかった．一方，迷路会話にはTemporizerの発話は出現しないが，終助詞を伴った疑問だけでなく終助詞を含まない様々な疑問の発話が出現した．したがって，Temporizerの識別実験には旅行会話を，Questionの識別実験には迷路会話を用いた．

## 6.2 Temporizer の識別

Temporizer の識別実験の結果を表 3に示す．「そうですね」の場合には 32.4% しか識別されていないが，人間による文脈無しでの Temporizer 識別の聴取実験の正解率を考慮すると決して悪い数字とはいえない．

なお，表 3からもわかるように使用したデータベースには「そうですか」「そうですね」以外に Temporizer は無かった．このため，他の文字列の Temporizer への適用可能性を検証する必要があり，今後引続きデータ収集を行なう予定である．

## 6.3 Question の識別

まず，先に示したパラメータについて，各パラメータ単一での Question の場合とそれ以外の場合の分布の違いの分析を行なった．図 5に示すように文末の 2 音素の基本周波数の回帰直線の傾きに違いが見られたが，他の 2 つのパラメータにははっきりとした違いが見られなかった．また，文末の 2 音素の基本周波数の回帰直線の傾きを入力として判別分析で識別したところ，約 70% の Question の発話を識別できた．

表 3 Temporizer の識別結果

| 正解率 | |
|---|---|
| Temporizer | Temporizer以外 |
| そうですか 84.0 % (21/25)<br>[ 91.2 % ] | 78.6 % (246/313) |
| そうですね 32.4 % (11/34)<br>[ 38.0 % ] | |

(注) [ ] 内の数字は人間による文脈なしでの Temporizer識別の聴取実験における正解率

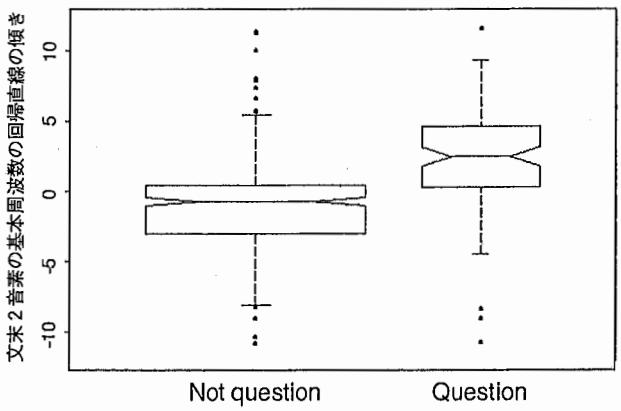



図 5 文末の 2 音素の基本周波数の回帰直線の傾きの分布の違い

次に，全パラメータを用いてtreeを作成し，Questionの識別を行なった．なお，treeの学習に際してはQuestionのサンプル数が他に比べて少なかったので，このサンプル数の差が学習に悪影響を及ぼさないようにデータに重みをつけた．この時，文末のアクセント型（文末の1モーラで上がるLH型，変わらないLL型またはHH型，下がるHL型）の違いによる識別精度への影響を調べるために，アクセント型で分類した場合と分類しない場合の両方について識別を行なった．表4に示す結果のように約80%のQuestionの発話を識別できており，全パラメータを用いたことによる精度の向上が確認された．また，アクセント型により3種類に分類して識別した場合と，分類せずに識別した場合での精度の差はあまりなかった．この結果から，Questionの文末の基本周波数パタン等に文末のアクセント型が大きな影響を与えていないと考えられる．したがって，Questionの識別過程において文末のアクセント型による分類は必要ないと思われるが，これは2名の話者のデータでの実験結果であるので，さらに話者を増やして分析および実験する必要がある．

## 7 むすび

複数の発話アクトの候補を持つ発話について聴覚的および音響的な分析を行なった．この結果，「ですか」，「ですね」を文末に持つ発話は複数の発話アクトを持ち韻律的な違いがあることがわかった．そして，これらの結果を基に任意の文字列に適用可能な発話アクトの識別方法を検討し，TemporizerおよびQuestionの識別方法を提案した．これらの方法を用いた識別実験を行なった結果，音素の時間長の伸長度がTemporizerの識別のために有効であり，これを用いることにより約8割の精度で識別可能であることが確認された．また，文末の基本周波数パタンの傾きがQuestionの識別のために有効であり，これを用いることにより約7割の精度で識別可能で，さらに，文末の音素の時間長および文末のパワーの傾きを入力として加えることにより，識別の精度が約8割に向上することが確認された．しかし，Temporizerの識別については「そうですか」と「そうですね」以外の文字列の発話に対する有効性の検証をする必要があり，Questionの識別については使用データの話者数を増やした実験を行なう必要がある．

## 参考文献


[1] M.Tomokiyo :"Segmentation and Aggregation of Utterances by Using Speech Act Labels", 信学技報, NLC95-23, 1995

[2] T.Morimoto *et al.* :"A speech and language database for speech translation research",Proc. ICSLP94, vol.IV, pp.1791-1794(1994)

[3] W.N.Campbell and S.D.Isard :"Segment durations in a syllable frame",Journal of Phonetics vol.19,pp.37-47,1991

[4] 匂坂, 海木, 岩橋, 三村: "ATR $\nu$-Talk 音声合成システム," 情報処理学会「音声言語情報処理と音声入出力装置」研究グループ，電子情報通信学会「音声認識の実用化をめざす新手法」時限研究専門委員会 研究会資料 (1992-10).


表4 Questionの識別結果

| 文末のアクセント型の区別 | アクセント型 | 正解率 | |
|---|---|---|---|
| | | Question | Question以外 |
| あり | HL | 76.9 % (10/13) | 98.1 % (154/157) |
| | LL or HH | 85.2 % (52/61) | 84.8 % (195/230) |
| | LH | 82.8 % (24/29) | 92.3 % (36/39) |
| | 平均 | 83.5 % (86/103) | 90.4 % (385/426) |
| なし | —— | 89.3 % (92/103) | 86.4 % (368/426) |

# 付録

## A　迷路会話収録に用いた迷路

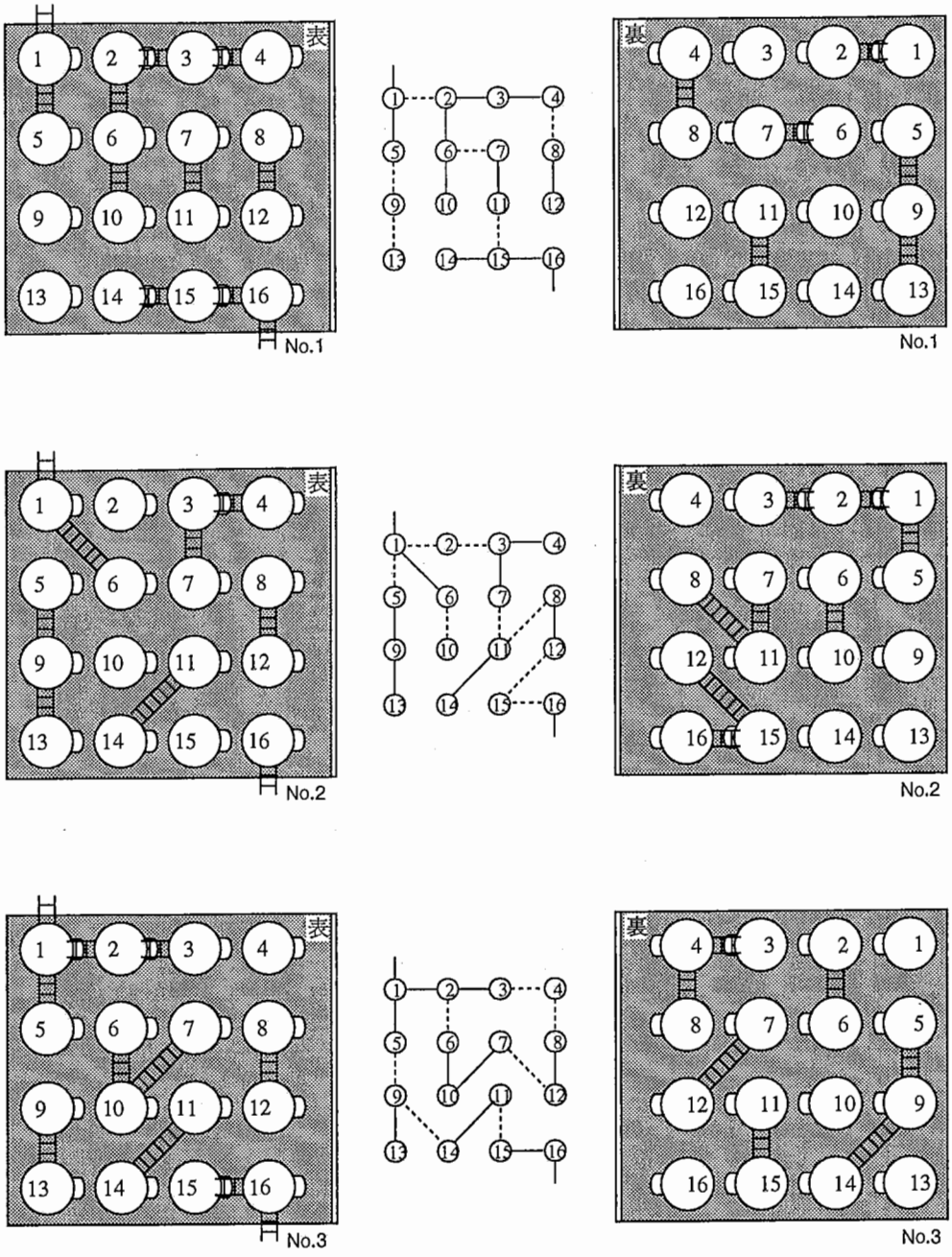



図 6 迷路会話収録用迷路（その 1）

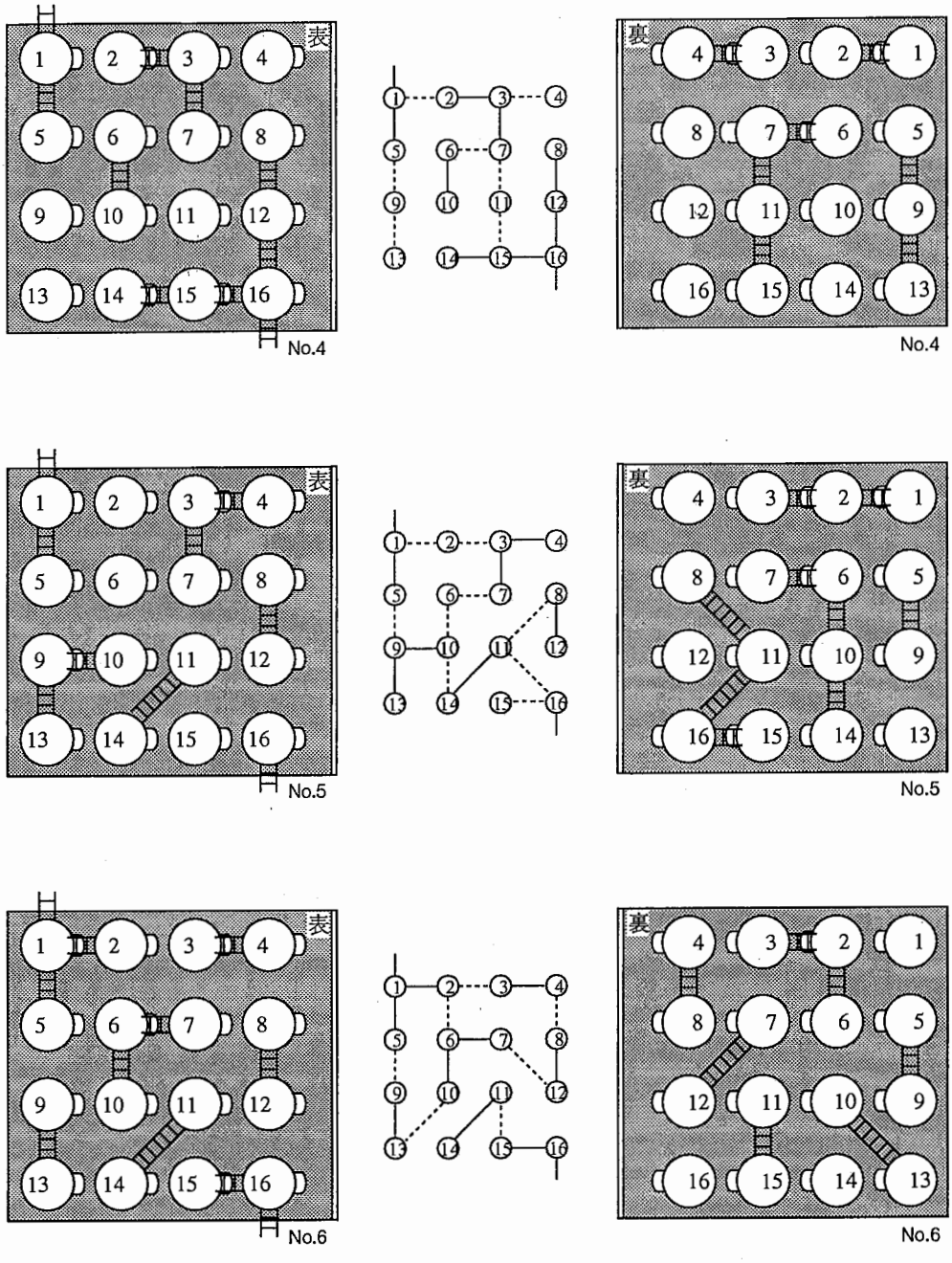



図 7 迷路会話収録用迷路（その 2）

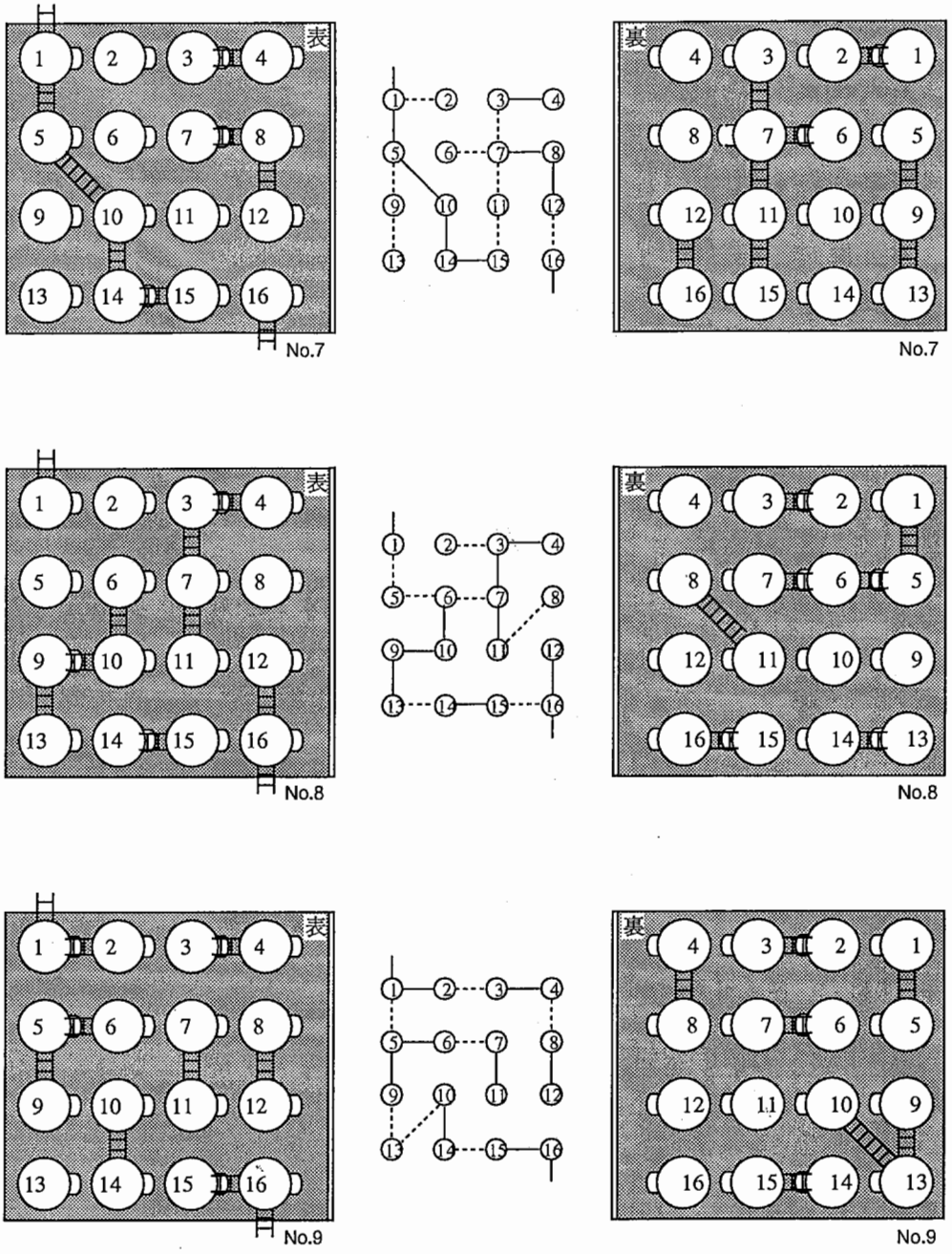



図 8 迷路会話収録用迷路（その 3）