

修士学位論文

論文題目

人間同士の対話現象を模倣した
音声対話システムの研究

豊橋技術科学大学大学院 工学研究科
修士課程 情報工学専攻

平成19年 1月 9日

学籍番号 033430

西村 良太

目次

| | | |
|-------|------------------------|----|
| 1 | 緒論 | 1 |
| 2 | 応答タイミング生成 | 4 |
| 2.1 | 応答現象 | 4 |
| 2.2 | 決定木によるタイミング生成 | 6 |
| 2.3 | 決定木構築に用いた素性 | 7 |
| 2.3.1 | ピッチ (F0)・パワーの算出 | 8 |
| 2.3.2 | ピッチ (F0)・パワーの回帰係数 | 8 |
| 2.4 | 決定木の学習 | 9 |
| 2.4.1 | 学習データの作成 | 9 |
| 2.4.2 | 決定木の評価 | 10 |
| 2.4.3 | 追加実験 | 16 |
| 2.4.4 | 発話ごとの評価 | 21 |
| 3 | 韻律変化のモデル化 | 35 |
| 3.1 | 人間同士の対話における話者間の韻律の関係 | 35 |
| 3.1.1 | 対話コーパス | 35 |
| 3.1.2 | 対話中の2話者の基本周波数 (F0) の変動 | 38 |
| 3.1.3 | 対話中の2話者の基本周波数の相関 | 38 |
| 3.2 | 対話の印象と対話現象との関係 | 40 |
| 3.3 | 韻律情報変化のモデル化 | 43 |
| 4 | 音声対話システム | 46 |
| 4.1 | 想定タスク | 46 |
| 4.2 | 対話システムの構成 | 47 |
| 4.2.1 | 音声認識・韻律解析部 | 49 |
| 4.2.2 | 情報収集部 | 49 |
| 4.2.3 | 対話履歴管理部 | 50 |
| 4.2.4 | 特徴量計算部 | 50 |
| 4.2.5 | 応答生成部 | 50 |
| 4.2.6 | 応答選択部 | 52 |
| 4.2.7 | 天気情報 | 52 |
| 4.3 | 出力部 | 52 |
| 4.4 | システムの動作例 | 53 |

| | | |
|----------|------------------------------|-----------|
| 5 | 評価 | 57 |
| 5.1 | 決定木からの応答タイミングの主観評価 | 57 |
| 5.1.1 | 評価方法 | 57 |
| 5.1.2 | 評価結果 | 58 |
| 5.2 | 対話システムとの対話による評価 | 59 |
| 5.2.1 | 評価方法 | 59 |
| 5.2.2 | 評価結果 | 60 |
| 6 | 結論 | 64 |
| 7 | 謝辞 | 67 |
| A | 付録 | 72 |
| A.1 | 発表業績 | 72 |
| A.2 | 音声認識部に用いた文法・単語辞書 | 72 |
| A.2.1 | 音声認識部に用いた文法 | 72 |
| A.2.2 | 音声認識に用いた単語辞書 | 77 |

図目次

| | | |
|------|---------------------------------------|----|
| 2.1 | 学習された決定木の一部 | 7 |
| 2.2 | 決定木によるタイミング生成 | 7 |
| 2.3 | 音声のピッチ・パワーの回帰係数 | 9 |
| 2.4 | 回帰係数の計算に用いるフレーム (100ms) | 9 |
| 2.5 | 評価方法の違い | 21 |
| 2.6 | 正解からのずれ (あいづち) (自動車販売 女性エージェント) | 27 |
| 2.7 | 正解からのずれ (話者交替) (自動車販売 女性エージェント) | 28 |
| 2.8 | 正解からのずれ (あいづち) (自動車販売 男性エージェント) | 29 |
| 2.9 | 正解からのずれ (話者交替) (自動車販売 男性エージェント) | 30 |
| 2.10 | 正解からのずれ (あいづち) (海外旅行計画 女性エージェント) | 31 |
| 2.11 | 正解からのずれ (話者交替) (海外旅行計画 女性エージェント) | 32 |
| 2.12 | 正解からのずれ (あいづち) (海外旅行計画 男性エージェント) | 33 |
| 2.13 | 正解からのずれ (話者交替) (海外旅行計画 男性エージェント) | 34 |
| 3.1 | 各対話におけるオーバーラップ出現頻度のヒストグラム (D02 と D03) | 36 |
| 3.2 | 各対話におけるフィラー出現頻度のヒストグラム (D02 と D03) | 36 |
| 3.3 | 各対話におけるオーバーラップ出現頻度のヒストグラム (D01 と D04) | 37 |
| 3.4 | 各対話におけるフィラー出現頻度のヒストグラム (D01 と D04) | 37 |
| 3.5 | CSJ 対話音声の例 | 38 |
| 3.6 | 各対話におけるフィラー出現頻度のヒストグラム | 39 |
| 3.7 | オーバーラップ頻度と被験者評価値 (盛り上がり) | 43 |
| 3.8 | 対話音声の F0 とモデル出力値 | 44 |
| 3.9 | 人間同士の F0 相関値 vs モデルと正解の相関値 | 45 |
| 4.1 | システムの構成図 | 47 |
| 4.2 | 対話管理部の構成図 | 48 |
| 4.3 | 応答テンプレート | 51 |
| 4.4 | 対話例 1 | 55 |
| 4.5 | 対話例 2 | 56 |
| 5.1 | 被験者実験でのアンケート内容 | 61 |

表目次

| | | |
|-----|----------|---|
| 2.1 | RWC コーパス | 8 |
| 2.2 | | 8 |

| | | |
|------|--|----|
| 2.3 | 決定木性能 (All data) | 11 |
| 2.4 | 決定木性能 (自動車販売 女性エージェント) | 12 |
| 2.5 | 決定木性能 (自動車販売 男性エージェント) | 13 |
| 2.6 | 決定木性能 (海外旅行計画 女性エージェント) | 14 |
| 2.7 | 決定木性能 (海外旅行計画 男性エージェント) | 15 |
| 2.8 | 決定木性能 (全データ) | 16 |
| 2.9 | 決定木性能 (自動車販売 女性エージェント) | 17 |
| 2.10 | 決定木性能 (自動車販売 男性エージェント) | 18 |
| 2.11 | 決定木性能 (海外旅行計画 女性エージェント) | 19 |
| 2.12 | 決定木性能 (海外旅行計画 男性エージェント) | 20 |
| 2.13 | 発話ごとの決定木性能 (自動車販売 女性エージェント) | 23 |
| 2.14 | 発話ごとの決定木性能 (自動車販売 男性エージェント) | 24 |
| 2.15 | 発話ごとの決定木性能 (海外旅行計画 女性エージェント) | 25 |
| 2.16 | 発話ごとの決定木性能 (海外旅行計画 男性エージェント) | 26 |
| 3.1 | CSJ の対話コーパス | 35 |
| 3.2 | 各対話の F0 の相関 | 39 |
| 3.3 | アンケート被験者同士の「盛り上がり」の相関 | 41 |
| 3.4 | 各アンケート項目の被験者間の相関の平均値 | 41 |
| 3.5 | 被験者評価値と対話現象との相関 | 42 |
| 3.6 | モデルと実際の値との相関 | 45 |
| 4.1 | 音声分析条件など | 49 |
| 5.1 | 被験者の主観によるあいづちタイミングの評価 | 58 |
| 5.2 | 被験者の主観による話者交替タイミングの評価 | 58 |
| 5.3 | 被験者の評価 | 63 |

1 緒論

近年，計算機の性能向上と音声認識技術の発達に伴い，音声認識技術が多くの場面で使われるようになってきた．そして，音声認識技術を用いたインターフェースも発展してきている．近年では，観光案内や情報検索，カーナビゲーションシステムへの応用など，様々な対話システムが検討・実用化されている [1,2]．しかし，一般にこれらのシステムにおいては，ユーザがシステムに話しかけた際に，途中でシステムからの反応が全くなく，ユーザ発話終了後にシステムが応答を返すまで，システムがユーザ発話を聞いているのかどうか分からないといった問題があり，これが音声認識を利用した音声対話システムに壁を感じる一因となっている．今後は，音声対話システムがより身近なものになり，生活の中に入り込んでくることが予想されるが，その際には，より自然な対話を実現する必要がある．人間同士の雑談のような対話は，自然な対話のひとつの理想の形であると考えられる．我々はこのような対話に現れる様々な現象をシステムで実現し，自然な対話システムを構築することを目指している．

例えば，人間同士の会話においては，話者は互いにうなづきやあいづちによって相手の発話を理解していることを明示しており，それにより会話がスムーズに進行する．これを音声対話システムにも応用し，ユーザの発話に対して，システムがあいづちなどの反応を返す事が出来れば，ユーザは人間と対話している場合と同じように自然に対話が行えるのではないかと考えられる．

あいづちについては，これまでも多くの研究がなされている．メイナードによればあいづちは「続けて」というシグナル，内容理解を示す表現，相手の意見や考え方に賛成の意思表示をする表現などを表すものであるとしている [3]．また，あいづちを返す場合には，そのタイミングが重要である．対話を自然にするために取り入れたあいづちも，タイミングがおかしければ，かえって対話に支障をきたす原因になりかねない．小磯らは，発話句音声末1モーラ分のピッチ，パワーの変動のパターンが話者交替やあいづちに強く関わっていると分析している [4]．分析では，ピッチの変動が平坦型，平坦下降型，上昇下降型であり，またパワーの変動が平坦型，平坦下降型である場合にあいづちが多く打たれている．野口らは，韻律情報，品詞情報があいづちに与える影響を調べており，局所的な韻律情報のあいづちシグナル性は品詞情報のそれに比べて低いとしている [5]．話者交替については，Sacksらは話者の交替は発話者が質問や確認の発話を行って聞き手に対して返答を求めた箇所で発生するものが望ましいとしている [6]．またこれらにおいては，言語情報の方が効果的であるという結果を報告している研究が多いが，大須賀らは，韻律情報が話者交替の予測に有効にはたらき，発話末まで待たなくても予測できるとしている [7]．

これまでも、リアルタイム応答として、ユーザ発話にあいづちを打つという研究は行われてきている。平沢らは連続音声認識アルゴリズムの中間結果を用いて言語情報からあいづちを打つことを検討している [8]。岡登らは、コーパス上で実際にあいづちが打たれた箇所の直前の韻律特徴からのテンプレートを作成することで、あいづちの生成を行っている [9]。佐藤らは、実際の人間同士の会話でのあいづちの出現頻度からあいづちの生成フローチャートを作成し、あいづちの生成を行っている [10]。Wardらは低ピッチ区間が一定時間続く箇所にあいづちを生成するシステムを構築している [11]。

これまでユーザの使い勝手などを考えてシステムの発話に対して割り込む（バージョン）ことができるシステムは存在するが [8] [12]、ユーザに対するシステム発話のタイミングを考慮した例は少ない。また、これまでのシステムのほとんどは、音声認識結果とポーズの検出によってシステムが応答を返すようなものになっている。この方式では、ポーズの検出の為に少なからず遅延が生じる。またユーザの発話にシステムが割り込むことが出来ないため、発話が重複するような対話の実現ができなかった。しかし、実際の対話においては、あいづちや話者交替が重複して発生することは少なくない。あいづちについては、全体の3割ほどが重複して発生しており、さらに雑談対話においては、半数以上が重複して発生している [13]。このことから、円滑に対話を行う上で、重複発話を無視することは出来ない。

このように、あいづちや話者交替のタイミングは、対話の自然性を考える上で非常に重要であり、あいづちや話者交替を行う場合には、相手の発話に応じて適切なタイミングで応答を返し、時にはそれらをオーバーラップさせることによって、スムーズに会話が進行していく。そこで我々は、音声対話システムにおいてあいづちや、システムからユーザへの割り込み発話など、種々の現象を考慮しそれらを適切なタイミングで行う天気予報を話題とする雑談システムを構築した。

今回構築したシステムでは、逐次的に音声認識と韻律解析を行うことにより、ポーズの検出による遅延の問題を解消している。また、ポーズを検出する必要ないので、ユーザ発話への割り込みが可能になっている。これにより、より実際に近い対話を実現することが可能である。

また、本研究では、円滑に対話を実現するために、タイミングのみに重点を置くのではなく、対話中の韻律の変化も考慮している。実際の人間同士の対話においては、対話が進むにつれて、お互いの韻律（声の高さや、話速など）が同調して進んでいる。垣田 [14] は、簡単な質問応答形式にて話者の基本周波数に関して話者間で関係があるかを実験により調査しており、ほとんどの話者で、一方の話者の基本周波数が高ければ、もう一方の話者の基本周波数も高くなることを指摘している。また、長岡ら [15] は、交替潜時 (Switching pause) が、2話者間で有意な正の相関を示

したと報告している．これらのことから，人間同士の対話の韻律情報には，何らかの関係がある事が予想される．人間同士の対話のように，より円滑に対話を進める為には，音声対話システムにおいても，韻律情報を制御することが必要である．

以上のことをふまえ，本研究では，音声対話システムにおいて，あいづちやその他の応答現象を適切なタイミングで生成できるシステムを提案する．タイミングの決定は，実際の間同士の対話の特徴から学習した決定木を用いることで実現する．

人間同士の対話において，実際にあいづちや，その他の応答現象が起こった部分の情報を元に決定木を作成する．決定木に用いる素性としては，応答現象が起こったときの直前 100ms・500ms の韻律情報の変動，発話長，ポーズ長などを考える．

また，韻律変化については，人間同士の対話音声から，対話者間でどのように韻律情報に相互作用があるかを分析し，そこから，どのような動きがあるかを見出し，モデル化を行う．このモデルを，対話を円滑に進めるための韻律情報制御モデルとして，音声対話システムに組み込んだ．この際，人間同士の対話音声からの情報が，対話を進める上での理想的なものであると仮定し，各種特徴を調べた．

そして，上述のタイミング生成決定木と韻律変化モデルを組み込んだ音声対話システムを実際に構築する．このシステムは，リアルタイムに音声認識をし，また，応答現象選択・タイミング生成を行うので，発話内容だけではなくタイミングも考慮して発話するものとなっている．また応答音声の韻律を変化させることにより，相手への同調を示して対話を進めることが出来るようになっている．

最後に，本対話システムと被験者が対話を行い，提案手法のタイミング生成の妥当性を調べる．また，被験者から得られる回答から現状のシステムの改良点を検討する．

以降，本論文では，2章で応答タイミングのモデル化，3章で韻律のモデル化，4章で本論文の応答手法を組み込んだ音声対話システムについて述べ，5章で被験者実験による評価結果，最後に6章で結論と今後の課題を述べる．

2 応答タイミング生成

人間同士の対話では、発話者の‘完結したターン’に対して、受話者が要求された応答を返したり、関連する次のターンを開始することで、相手の発話内容の受理を示す。発話者の‘継続するターン’に対しては、受け手があいづちなどを打つことで、理解していることを示す [16]。また、発話者の発話に重ねて同じ事を言う場合‘ユニゾン’や、同じ内容の発話をお互いが補完し合って完成させていくもの‘共同補完’がある。また、発話者が一度で喋ると、受話者が記憶できないようなもの、例えば電話番号を伝えたりする場合などには、‘分割して発話’され、受話者は、復唱して内容理解を示す。また、発話者が発話中に‘一部分を確認’したい場合に、その部分を上昇調にすることで、発話が完了する前に、受話者の確認や訂正を得ようとするものがある。

これらの、特定の型には、典型的には、ある特定の型の理解提示が続く。‘完結したターン’には、要求された応答・関連する次のターンが続き、‘継続するターン’には、あいづちが、しばしば相手発話に重複して返される。‘記憶の為の分割提示’には、復唱が続く。また、不完全な発話に対しては、補完が成される。‘一部分を確認する発話’に対しては、肯定または修正にて応答する。

本章では、ユーザの発話に対して、システムからの応答現象を選択し、その応答タイミングを決定する方法について述べる。

2.1 応答現象

まず、人間同士の対話には、どのような応答現象があるのかを考えなければならない。そして、それに対応が出来るような応答現象を扱えなければならない。先行研究 [16] を参考にすると、以下のような種類が挙げられている。

完結するターンは、発話の提示に対して、受け手が要求された応答を産出したり、関連する次のターンを開始したりすることによって、理解を前提することで発話を受理する。

継続するターンは、受け手が「はい」や「うん」などのあいづちを打ったりすることで理解していることを主張する。

分割提示：短いターンは、受け手の記憶容量の範囲内の長さで発話を分割し、その都度確認を取る。

分割提示：長いターンは、返答が長くなる際に、分割提示をし、聞き手からの割り込みを受けずに発話権を確保する。

不完全発話：共同補完などは，発話が終了しないうちに相手の発話の続きを完成させるようなものである．

試行構成素は，発話提示の途中に上昇調にすることで，発話提示完了前に部分的に受け手の確認・修正を得ようとするものである．

‘完結するターン’に対しては，‘Elizaルールによるテンプレート型応答（要求された応答）’にて対応する．‘継続するターン’に対しては，‘あいづち’を扱うことで対応する．また，タイミングは決定木にて生成されるものである．‘分割提示：短いターン’に対しては，‘復唱’をすることで対応する．確認を取らなければならないような重要な言葉を復唱するようにする．‘分割提示：長いターン’に対しては，発話者の休止中に聞き手が‘あいづち’をすることで，相手から発話権を奪うことなく，話し手に対して理解を示すことが出来る．‘不完全発話：共同補完’に対しては，‘共同補完’をシステムに搭載することで対応する．ただし，相手が何を言うかの予測が必要である．‘試行構成素’に対しては，否定・肯定・修正などを行う必要がある．システムでは，対話履歴やスロットを活用して実装することが可能であるが，現段階ではシステムに実装はしていない．

これらのことから，システムで扱う応答現象には，‘要求された応答’，‘あいづち’，‘復唱’，‘共同補完’を考える．

要求された応答は，発話者交替が起こる場合に，ユーザが要求している応答をシステム側が応答するものである．これには，それまでの知識（スロット値や対話履歴）を用いて，Elizaテンプレートを用いて応答文を作成し応答することで実現している．

例文：ユーザ 明日の豊橋の天気はどうか？

システム 明日の豊橋の天気は晴れです．

あいづちは，発話者交替が起こらずに，聞き手が理解などを示す場合に用いられるものである．あいづち応答については，タイミングがより重要になる．先行研究では，ピッチ/パワーの概形が，ある形になっているときに，あいづちが生じるという結果を報告している [17]．

例文：ユーザ こちらのスポーツカーペ にえっと…

システム はい

復唱は，ユーザ発話に重要な言葉が含まれていた場合に，ユーザの発話中に確認の意味で，その言葉を繰り返すものである．これには，重要な言葉がユーザから発話された場合，それを即座に応答に用いることによって実現することが可能である．

例文：ユーザ まあご存知のキャデラック

システム キャデラック ははっ

共同補完は，ユーザ発話と内容が同じである応答を，ユーザにオーバーラップし

て返すものである。システムが，ユーザの残りの発話を予測でき，またそれに賛成であるときに起こる。これは，ユーザが発話している途中に，その内容を，システムが持っている知識（スロット値や Eliza テンプレート）と照らし合わせ，その後に続く内容をオーバーラップして応答することで実現することが可能である。

例文：ユーザ　しかも一番新しい 90 ... 5　　そうですね
 システム そうです　95年

これらを，適切なリズムで応答することによって，自然な対話の流れを実現する。

2.2 決定木によるタイミング生成

人間の応答には，何の考えもなしに簡単に応答できるものから，しばらく考えないといけないものまでの，幾つかのレベルが存在すると考えられる。今回のシステムでは，このことから，応答を幾つかの段階に分け，それぞれのレベルについて同時に応答文を準備し，その中から，適切な応答を状況に応じて選択し出力するというようにモデル化を行った。

要求された応答，あいづち，復唱，共同補完などについて，それぞれ，同時に応答文を用意し，応答するようにする。そして，これらの応答現象（応答文）の中から，どの応答現象を用いるか，そして，どのタイミングで応答するかを決定木で決定する。

ここで，決定木とは，内部節点が質問を表し，葉節点が予測またはクラスを表すような木であり，C4.5 [18] や CART (Classification And Regression Tree) [19] が代表的である。本研究では決定木の生成には帰納学習システム C4.5 を用いる。C4.5 は与えられた学習データから初期決定木を構築し，その後枝刈りを行う。

決定木では，100ms ごとにリアルタイムに決定木に対して素性を入力し，応答を決定している。これまでの音声対話システムでは，ユーザ発話の中にポーズが検出された場合に，それまでの発話に対して応答をするものがほとんどであった（ポーズ駆動）。しかし，本研究では，ポーズに関係なく，逐次ユーザ発話からの素性を決定木に入力し，応答・タイミングを決定している。このことから，ユーザ発話に対してオーバーラップして応答することも可能になっている。

図 2.1 は，実際の決定木の例（一部）である。この場合には，最初に‘前のシステム発話終了時刻からの時間’により判定がなされ，その結果によって，1段階下の判定へと移る。このように，素性の状態によって判定が下されていき，最終的な判定にたどり着き，結果を出力する。

先行研究 [13] においては，システムはポーズ駆動であったため，オーバーラップする応答を返すことはできなかった。また，この以前のシステムでは，応答の種類も

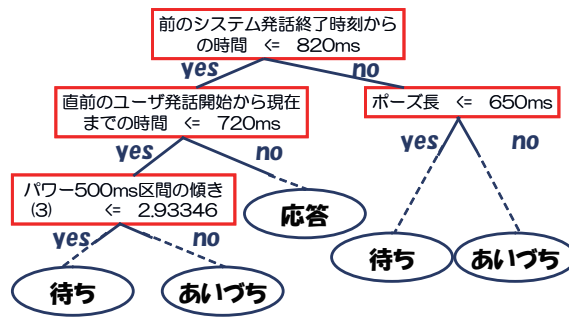


図 2.1: 学習された決定木の一部

少なく、決定木の判定としては‘応答’、‘待ち’、‘あいづち’のみであった。本システムでは、応答の種類も多く、また、それらの応答を統一的な枠組みで実現している。

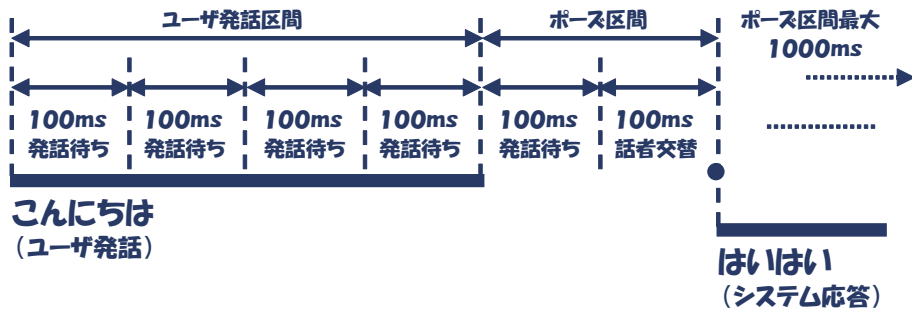


図 2.2: 決定木によるタイミング生成

タイミングの決定は図 2.2 のようになっている。ユーザ発話に対して、100ms ごとに判定をする。図中では、ユーザの発話中では 4 回の判定が行われ、その間はずっと‘発話待ち’が出力されていたので、システムは応答せずに待ち状態になっている。もし、この発話中の判定で応答する判定が出た場合には、オーバーラップ応答になる。ポーズ区間に入っても、決定木による判定は続き、6 回目の判定で‘話者交替’が出力され、システムは応答を返している。

決定木の為に用いたトレーニングデータは、2 タスク 48 対話である。用いたコーパスは RWC 音声対話データベース [20] である。表 2.1 に、コーパスの内訳を示す。対話内容は、自動車販売タスクと海外旅行計画タスクの 2 種類である。

2.3 決定木構築に用いた素性

本提案手法では、各応答現象のタイミング判定を行う決定木構築の素性として以下のものを考える。

- 直前のユーザ発話開始から現在までの時間

表 2.1: RWC コーパス

| | エージェント 性別 | タスク | ファイル数 | データ量 | 時間 |
|--------|--------------|--------|-------|-------|-------|
| ディスク 1 | 女 | 自動車販売 | 12 | 327MB | 87 分 |
| ディスク 2 | 男 | 自動車販売 | 12 | 449MB | 119 分 |
| ディスク 3 | 女 | 海外旅行計画 | 12 | 353MB | 94 分 |
| ディスク 4 | 男 | 海外旅行計画 | 12 | 401MB | 107 分 |

- ポーズ長
- 前のシステム発話終了時刻からの時間
- ピッチ・パワーの 100ms 区間の傾き (3 つずつ)
- ピッチ・パワーの 500ms 区間の傾き (5 つずつ)

これらはすべて連続値である．ここで，ピッチ，パワーの変動は図 2.3 に示すように，音声の 100ms 区間を 3 分割し，それぞれの区間の回帰係数を計算することで求めた．500ms 区間の傾きについては，100ms ごとに区切り，合計 5 つの区間について回帰係数を計算した．回帰係数計算法については後述する．なお，ここでは便宜上“ピッチ”とは基本周波数のことを意味し対数値を使用している．

2.3.1 ピッチ (F0) ・パワーの算出

決定木学習に用いるために算出したピッチ・パワーは，praat [21] を用いて音声を解析して得られたものである．ピッチ・パワー共に，フレームシフトは 5ms である．

2.3.2 ピッチ (F0) ・パワーの回帰係数

2.3.1 節で得られたピッチ・パワー値から，それぞれの回帰係数を算出する．100ms の区間については図 2.4 のように，21 フレームのデータから，500ms の区間については，100 フレームのデータから，回帰係数を計算する．

(2.1) 式は回帰係数を求める式である．ここで， $c(x)$ は x フレーム目の特徴 (ピッチ，パワー) 値を表す． $\Delta c(l)$ は，第 l 区間の回帰係数を示している．

表 2.2:

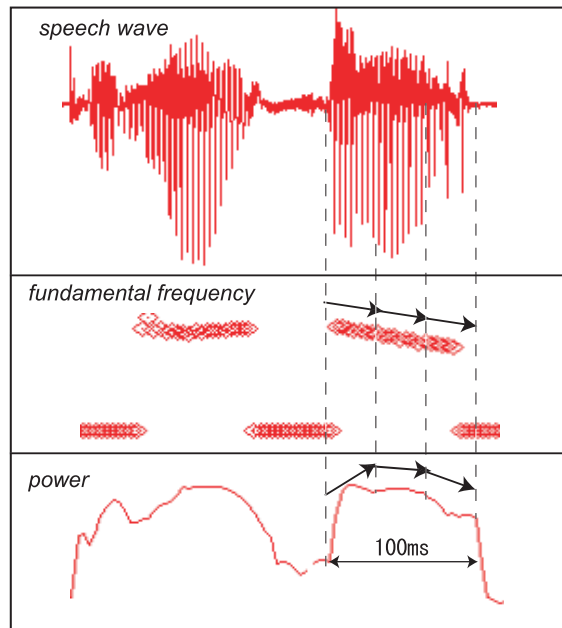


図 2.3: 音声のピッチ・パワーの回帰係数

$$\Delta c(l) = \frac{\sum_{k=-K}^K c(Kl + k)}{\sum_{k=-K}^K k^2} \tag{2.1}$$

回帰係数を計算する範囲を表す K は、100ms 区間を 3 つに分割した場合には、1 つの区間が 11 フレームであるため、 $K = 5$ として計算を行い、500ms 区間を 5 つに分割した場合には、1 つの区間が 21 フレームであるため、 $K = 10$ として計算を行う。

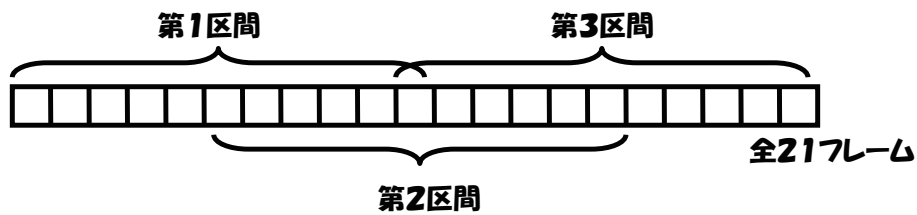


図 2.4: 回帰係数の計算に用いるフレーム (100ms)

2.4 決定木の学習

2.4.1 学習データの作成

今回の決定木では、100ms ごとに応答を選択し出力することを考える。オーバーラップして発生した応答については、その部分の元話者の発話音声の情報と応答内

容を正解データとして用意する．また，ポーズに入ってから話者交替が発生した場合には，ポーズ開始点から，ポーズ終了点（話者交替点）までを 100ms ごとに区切り，‘発話待ち’を正解としたデータを用意する．そして，実際に話者交替が起こった点では，その応答内容を正解としたデータを用意した．

‘発話待ち’以外の，応答を示すデータにおいては，応答タイミングの± 50ms の点にも，同じ正解データを用意した．これは，正解点をぼかして，決定木の過学習による正解率の低下を防ぐものである．

また，各正解タグ（‘発話待ち’，‘要求された応答’，‘あいづち’，‘復唱’，‘共同補完’）について，データ数にバラツキがある．これを解決するために，データ数が少ないものを，データ数が多いものに合わせるように，データを増やして（複製して）対応した．

2.4.2 決定木の評価

2.3 節で述べた素性を持つ決定木を，各タスク（自動車販売，海外旅行計画 各タスクでエージェント性別ごとに 計 4 種類）で学習し，各タスクのテストデータで評価を行った．表 2.3～表 2.7 に，オープンテストによる人間同士の対話における各対話現象との一致度合いを示す．

テスト結果として，N-fold で行った結果の平均を示してある．

また，評価尺度としては

$$\text{再現率} [\%] = \frac{\text{対話コーパス中の応答現象（正解）と被験者の応答現象の一致回数}}{\text{対話コーパス中の応答現象の総数}} \times 100$$

$$\text{適合率} [\%] = \frac{\text{対話コーパス中の応答現象と被験者の応答現象の一致回数}}{\text{被験者の応答現象の総数}} \times 100$$

を用いた．

人間同士の対話で，実際に応答されていない箇所であっても，必ずしも応答することが不自然であるという場所とは限らないので，人間同士の対話との一致は一つの目安に過ぎないことに注意を要する．

表 2.4 は，すべてのコーパスデータで学習した決定木である．表 2.4，表 2.5 は，自動車販売タスクでの，エージェントが女性のもので学習した決定木である．表 2.5，表 2.6 は，自動車販売タスクでの，エージェントが男性のもので学習した決定木である．表 2.6，表 2.7 は，海外旅行計画タスクでの，エージェントが女性のもので学習した決定木である．表 2.7，表 2.8 は，海外旅行計画タスクでの，エージェントが男性のもので学習した決定木である．

結果を見てみると、クローズドテストでは、非常に高い精度が出ているが、その反面、オープンテストでは、精度が低くなってしまっている。特に、相対的に見て頻度が少ない‘復唱’や‘共同補完’は、オープンテストでは、ほとんど正解をすることが出来ていない。これは学習データの数が少ないことが原因である。他のクラスについても、十分な精度が得られていない為、精度をより高くするために、より沢山のデータを用意することや、決定木の素性の見直し（表層的な言語情報も含める）などのことも考えなければならない。

表 2.3: 決定木性能 (All data)

(a) オープンテスト

| テスト：オープン | | | | | |
|----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 663 | 547 | 447 | 19 | 23 |
| 交替 | 573 | 1407 | 1696 | 22 | 33 |
| 継続 | 979 | 3486 | 62019 | 64 | 97 |
| 復唱 | 15 | 33 | 13 | 4 | 0 |
| 補完 | 19 | 25 | 28 | 0 | 1 |
| 再現率 | 0.390 | 0.377 | 0.931 | 0.062 | 0.014 |
| 適合率 | 0.295 | 0.256 | 0.966 | 0.037 | 0.006 |
| F 値 | 0.336 | 0.305 | 0.948 | 0.046 | 0.009 |

(b) クローズドテスト

| テスト：クローズド | | | | | |
|-----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 1699 | 0 | 0 | 0 | 0 |
| 交替 | 1 | 3730 | 0 | 0 | 0 |
| 継続 | 410 | 1954 | 64207 | 44 | 30 |
| 復唱 | 0 | 0 | 0 | 65 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 73 |
| 再現率 | 1.000 | 1.000 | 0.963 | 1.000 | 1.000 |
| 適合率 | 0.805 | 0.656 | 1.000 | 0.596 | 0.709 |
| F 値 | 0.892 | 0.792 | 0.981 | 0.747 | 0.830 |

表 2.4: 決定木性能 (自動車販売 女性エージェント)

(a) オープン (N-fold) テスト

| テスト：オープン | | | | | |
|----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 128 | 139 | 105 | 7 | 3 |
| 交替 | 133 | 304 | 417 | 4 | 4 |
| 継続 | 251 | 861 | 12546 | 12 | 12 |
| 復唱 | 8 | 7 | 3 | 0 | 0 |
| 補完 | 4 | 3 | 4 | 1 | 0 |
| 再現率 | 0.335 | 0.353 | 0.917 | 0.000 | 0.000 |
| 適合率 | 0.244 | 0.231 | 0.960 | 0.000 | 0.000 |
| F 値 | 0.283 | 0.279 | 0.938 | 0.000 | 0.000 |

(b) クローズドテスト

| テスト：クローズド | | | | | |
|-----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 382 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 862 | 0 | 0 | 0 |
| 継続 | 61 | 433 | 13186 | 3 | 0 |
| 復唱 | 0 | 0 | 0 | 18 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 12 |
| 再現率 | 1.000 | 1.000 | 0.964 | 1.000 | 1.000 |
| 適合率 | 0.862 | 0.666 | 1.000 | 0.857 | 1.000 |
| F 値 | 0.926 | 0.799 | 0.982 | 0.923 | 1.000 |

表 2.5: 決定木性能 (自動車販売 男性エージェント)

(a) オープン (N-fold) テスト

| テスト：オープン | | | | | |
|----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 142 | 191 | 142 | 5 | 9 |
| 交替 | 157 | 400 | 626 | 3 | 10 |
| 継続 | 302 | 1289 | 19133 | 13 | 21 |
| 復唱 | 3 | 7 | 3 | 0 | 0 |
| 補完 | 5 | 8 | 10 | 0 | 0 |
| 再現率 | 0.290 | 0.334 | 0.922 | 0.000 | 0.000 |
| 適合率 | 0.233 | 0.211 | 0.961 | 0.000 | 0.000 |
| F 値 | 0.259 | 0.259 | 0.941 | 0.000 | 0.000 |

(b) クローズドテスト

| テスト：クローズド | | | | | |
|-----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 489 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 1196 | 0 | 0 | 0 |
| 継続 | 126 | 724 | 19887 | 5 | 16 |
| 復唱 | 0 | 0 | 0 | 13 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 23 |
| 再現率 | 1.000 | 1.000 | 0.958 | 1.000 | 1.000 |
| 適合率 | 0.795 | 0.623 | 1.000 | 0.722 | 0.590 |
| F 値 | 0.886 | 0.768 | 0.979 | 0.839 | 0.742 |

表 2.6: 決定木性能 (海外旅行計画 女性エージェント)

(a) オープン (N-fold) テスト

| テスト: オープン | | | | | |
|-----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 167 | 127 | 137 | 9 | 12 |
| 交替 | 143 | 230 | 446 | 8 | 7 |
| 継続 | 297 | 811 | 15169 | 19 | 21 |
| 復唱 | 6 | 10 | 5 | 1 | 3 |
| 補完 | 7 | 6 | 8 | 0 | 1 |
| 再現率 | 0.369 | 0.276 | 0.930 | 0.040 | 0.045 |
| 適合率 | 0.269 | 0.194 | 0.962 | 0.027 | 0.023 |
| F 値 | 0.312 | 0.228 | 0.946 | 0.032 | 0.030 |

(b) クローズドテスト

| テスト: クローズド | | | | | |
|------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 489 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 1196 | 0 | 0 | 0 |
| 継続 | 126 | 724 | 19887 | 5 | 16 |
| 復唱 | 0 | 0 | 0 | 13 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 23 |
| 再現率 | 1.000 | 1.000 | 0.958 | 1.000 | 1.000 |
| 適合率 | 0.795 | 0.623 | 1.000 | 0.722 | 0.590 |
| F 値 | 0.886 | 0.768 | 0.979 | 0.839 | 0.742 |

表 2.7: 決定木性能 (海外旅行計画 男性エージェント)

(a) オープン (N-fold) テスト

| テスト: オープン | | | | | |
|-----------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 139 | 124 | 102 | 4 | 7 |
| 交替 | 114 | 324 | 390 | 3 | 8 |
| 継続 | 221 | 831 | 14801 | 9 | 25 |
| 復唱 | 3 | 3 | 3 | 0 | 0 |
| 補完 | 5 | 4 | 6 | 0 | 1 |
| 再現率 | 0.370 | 0.386 | 0.932 | 0.000 | 0.063 |
| 適合率 | 0.288 | 0.252 | 0.967 | 0.000 | 0.024 |
| F 値 | 0.324 | 0.305 | 0.949 | 0.000 | 0.035 |

(b) クローズドテスト

| テスト: クローズド | | | | | |
|------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 376 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 839 | 0 | 0 | 0 |
| 継続 | 121 | 494 | 15250 | 10 | 12 |
| 復唱 | 0 | 0 | 0 | 9 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 16 |
| 再現率 | 1.000 | 1.000 | 0.960 | 1.000 | 1.000 |
| 適合率 | 0.757 | 0.629 | 1.000 | 0.474 | 0.571 |
| F 値 | 0.861 | 0.773 | 0.980 | 0.643 | 0.727 |

2.4.3 追加実験

前節での決定木学習では、オーバーラップに関する負のデータを用意しておらず、負のデータとしては、ポーズ中のデータしか使用していなかった為、決定木の性能があまり高くなかった。そこで、オーバーラップしている部分についても負のデータを用意し決定木学習を行った。用いた素性は、前節と同じように 2.3 節に示したものである。結果を表 2.8~表??に示す。

結果としては、負のデータを学習していない決定木と比較して、“あいづち”や“話者交替”に対しては精度が良くなっている。

以前に構築した、オーバーラップを考慮しない決定木での実験結果では、あいづち、話者交替ともに、F 値で 0.2 前後の値であった。それと比較をすると、今回の結果は、F 値で 0.5 前後となっており、精度が良くなっている。

表 2.8: 決定木性能 (全データ)

(a) オープン (N-fold) テスト

| | テストデータ open | | | | |
|-----|-------------|-------|--------|-------|-------|
| | クラス分け | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 904 | 629 | 96 | 37 | 33 |
| 交替 | 569 | 2198 | 901 | 31 | 32 |
| 継続 | 290 | 2013 | 211330 | 23 | 33 |
| 復唱 | 27 | 28 | 6 | 1 | 3 |
| 補完 | 35 | 33 | 3 | 1 | 1 |
| 再現率 | 0.532 | 0.589 | 0.989 | 0.015 | 0.014 |
| 適合率 | 0.495 | 0.448 | 0.995 | 0.011 | 0.010 |
| F 値 | 0.513 | 0.509 | 0.992 | 0.013 | 0.011 |

(b) クローズドテスト

| | テストデータ closed | | | | |
|-----|---------------|-------|-------|-------|-------|
| | クラス分け | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 1699 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 3730 | 0 | 0 | 1 |
| 継続 | 164 | 1053 | 21242 | 1 | 8 |
| 復唱 | 0 | 0 | 0 | 65 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 73 |
| 再現率 | 1.000 | 1.000 | 0.945 | 1.000 | 1.000 |
| 適合率 | 0.912 | 0.780 | 1.000 | 0.985 | 0.890 |
| F 値 | 0.954 | 0.876 | 0.972 | 0.992 | 0.942 |

表 2.9: 決定木性能 (自動車販売 女性エージェント)

(a) オープン (N-fold) テスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 206 | 143 | 14 | 11 | 8 |
| 交替 | 146 | 465 | 237 | 4 | 10 |
| 継続 | 48 | 562 | 45964 | 10 | 1 |
| 復唱 | 9 | 6 | 3 | 0 | 0 |
| 補完 | 5 | 4 | 0 | 3 | 0 |
| 再現率 | 0.539 | 0.539 | 0.987 | 0.000 | 0.000 |
| 適合率 | 0.498 | 0.394 | 0.995 | 0.000 | 0.000 |
| F 値 | 0.518 | 0.455 | 0.991 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 382 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 862 | 0 | 0 | 0 |
| 継続 | 23 | 308 | 46250 | 4 | 0 |
| 復唱 | 0 | 0 | 0 | 18 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 12 |
| 再現率 | 1.000 | 1.000 | 0.993 | 1.000 | 1.000 |
| 適合率 | 0.943 | 0.737 | 1.000 | 0.818 | 1.000 |
| F 値 | 0.971 | 0.848 | 0.996 | 0.900 | 1.000 |

表 2.10: 決定木性能 (自動車販売 男性エージェント)

(a) オープン (N-fold) テスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 241 | 191 | 40 | 8 | 9 |
| 交替 | 146 | 633 | 398 | 7 | 12 |
| 継続 | 83 | 956 | 60372 | 0 | 8 |
| 復唱 | 6 | 7 | 0 | 0 | 0 |
| 補完 | 8 | 14 | 1 | 0 | 0 |
| 再現率 | 0.493 | 0.529 | 0.983 | 0.000 | 0.000 |
| 適合率 | 0.498 | 0.351 | 0.993 | 0.000 | 0.000 |
| F 値 | 0.495 | 0.422 | 0.988 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 489 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 1196 | 0 | 0 | 0 |
| 継続 | 53 | 506 | 60850 | 0 | 10 |
| 復唱 | 0 | 0 | 0 | 13 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 23 |
| 再現率 | 1.000 | 1.000 | 0.991 | 1.000 | 1.000 |
| 適合率 | 0.902 | 0.703 | 1.000 | 1.000 | 0.697 |
| F 値 | 0.949 | 0.825 | 0.995 | 1.000 | 0.821 |

表 2.11: 決定木性能 (海外旅行計画 女性エージェント)

(a) オープン (N-fold) テスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 234 | 164 | 24 | 20 | 10 |
| 交替 | 138 | 403 | 276 | 8 | 9 |
| 継続 | 53 | 594 | 48350 | 14 | 17 |
| 復唱 | 15 | 7 | 3 | 0 | 0 |
| 補完 | 14 | 5 | 2 | 1 | 0 |
| 再現率 | 0.518 | 0.483 | 0.986 | 0.000 | 0.000 |
| 適合率 | 0.515 | 0.344 | 0.994 | 0.000 | 0.000 |
| F 値 | 0.517 | 0.402 | 0.990 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 452 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 834 | 0 | 0 | 0 |
| 継続 | 43 | 324 | 48648 | 10 | 3 |
| 復唱 | 0 | 0 | 0 | 25 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 22 |
| 再現率 | 1.000 | 1.000 | 0.992 | 1.000 | 1.000 |
| 適合率 | 0.913 | 0.720 | 1.000 | 0.714 | 0.880 |
| F 値 | 0.955 | 0.837 | 0.996 | 0.833 | 0.936 |

表 2.12: 決定木性能 (海外旅行計画 男性エージェント)

(a) オープン (N-fold) テスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 211 | 126 | 36 | 0 | 3 |
| 交替 | 124 | 441 | 264 | 3 | 7 |
| 継続 | 91 | 568 | 55960 | 28 | 10 |
| 復唱 | 2 | 4 | 1 | 0 | 2 |
| 補完 | 4 | 10 | 1 | 1 | 0 |
| 再現率 | 0.561 | 0.526 | 0.988 | 0.000 | 0.000 |
| 適合率 | 0.488 | 0.384 | 0.995 | 0.000 | 0.000 |
| F 値 | 0.522 | 0.444 | 0.991 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 376 | 0 | 0 | 0 | 0 |
| 交替 | 0 | 839 | 0 | 0 | 0 |
| 継続 | 46 | 326 | 56267 | 6 | 12 |
| 復唱 | 0 | 0 | 0 | 9 | 0 |
| 補完 | 0 | 0 | 0 | 0 | 16 |
| 再現率 | 1.000 | 1.000 | 0.993 | 1.000 | 1.000 |
| 適合率 | 0.891 | 0.720 | 1.000 | 0.600 | 0.571 |
| F 値 | 0.942 | 0.837 | 0.997 | 0.750 | 0.727 |

2.4.4 発話ごとの評価

2.4.3 節で構築した決定木はポーズを検出してから各 100ms ごとに振舞いを決定するものであった。つまり、ある発話に対する応答が、タイミングは違うが決定木の判定は正しいという場合でも、タイミングを示すフレーム数（100ms 単位）が正解ラベルと一致していなければ正解とはならない評価方法であった。そこで、判定位置（フレーム位置）は関係なく、応答の種類（あいづち、話者交替、発話待ちというようなもの）が一致すれば正解とする評価方法で評価した結果を以下に示す。また、正解であった場合の決定木の出力と実際のタイミング（コーパスデータのタイミング）とのずれも示す。

図 2.5 に、評価法の詳細を示す。図の上が、これまでのセグメント単位での評価法である。そして、下が発話ごとの評価方法である。

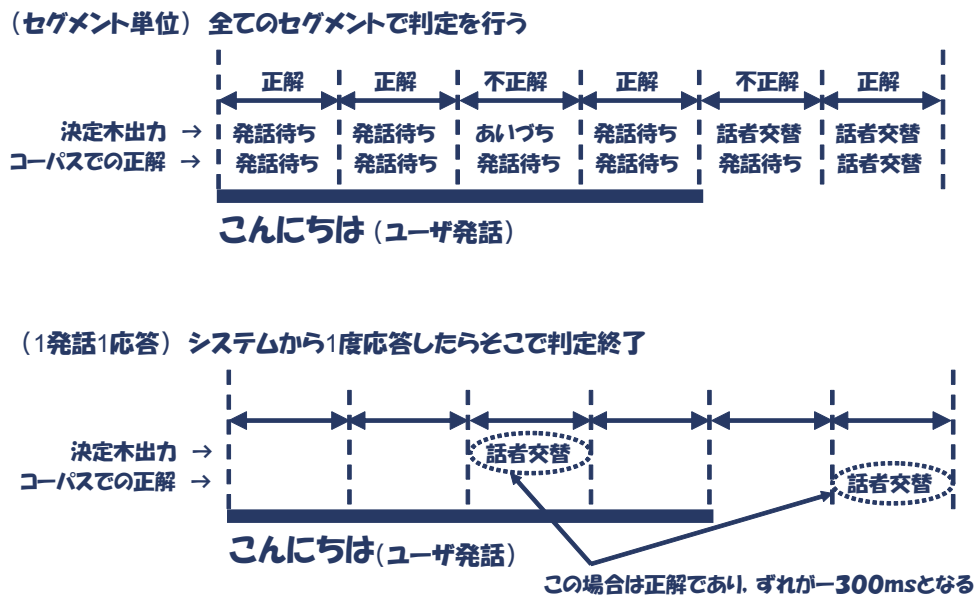


図 2.5: 評価方法の違い

実際のシステムでは、ユーザからの入力に対して、1 回の応答を返す（あいづちと復唱は 1 回以上も可能）。この評価法は、実際のシステムの動作を評価することに近い。

評価方法の違いから、成績に違いが出る。発話ごとの評価では、応答のタイミングがずれていても出力がコーパスと同じ応答現象なら正解となる為、その点に関しては評価が甘くなっている。しかし、一度応答してしまうと、それ以降のタイミングで応答される応答に関しては評価されないため、その部分で成績が落ちてしまうことも考えられる。

ここで、データ全体から学習した決定木は、不具合によって実験が完了できない

ため結果を掲載していない。また、オープンテストは、学習で使用したデータのタスクとは別のタスクのデータを用いてテストを行っている。テストデータの選び方は、学習に用いたデータのタスクと別のタスクで、エージェントの性別が同じものを用いた。例を挙げると、学習データに自動車販売・女性エージェントのものを用いた場合には、テストデータには、海外旅行計画・女性エージェントのものを用いたということである。結果は、表 2.13～表 2.16，図 2.6～2.13 に示す。

結果は、オープンテスト・クローズドテスト共に、あいづちの精度が低い値になっている。実際にはあいづちを返すところで、話者交替をしたり、無反応であったりしている。また、実際には話者交替の部分や、無反応である部分であいづちを返してしまっている部分がある。そして、話者交替・無反応（話者継続）は数が非常に多く、その中であいづちと間違えたものの割合は小さい。しかし、実際にシステムが返すあいづちが適切かどうかは、被験者実験を通して評価をしないと分からないので、この決定木についても被験者実験を行い、評価をしなければならない。

応答のずれに関しては、クローズドテストのずれが少なく、オープンテストは、ずれが広範囲に分布している。このことから今回の決定木も過学習していることが分かる。オープンテストとクローズドテストの結果が似たものになることが望ましいので、過学習にならないように工夫が必要である。また、あいづち・話者交替共に、実際の正解よりも少し遅めに応答を返すことが多いようである。

表 2.13: 発話ごとの決定木性能 (自動車販売 女性エージェント)

(a) オープンテスト

| | | テストデータ open | | | | |
|-----|--|-------------|-------|-------|-------|-------|
| | | クラス分け | | | | |
| | | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | | 34 | 165 | 253 | 0 | 0 |
| 交替 | | 27 | 592 | 215 | 0 | 0 |
| 継続 | | 26 | 753 | 1585 | 6 | 0 |
| 復唱 | | 2 | 10 | 13 | 0 | 0 |
| 補完 | | 1 | 7 | 14 | 0 | 0 |
| 再現率 | | 0.075 | 0.710 | 0.669 | 0.000 | 0.000 |
| 適合率 | | 0.378 | 0.388 | 0.762 | 0.000 | 0.000 |
| F 値 | | 0.125 | 0.501 | 0.712 | 0.000 | 0.000 |

(b) クローズドテスト

| | | テストデータ closed | | | | |
|-----|--|---------------|-------|-------|-------|-------|
| | | クラス分け | | | | |
| | | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | | 24 | 97 | 261 | 0 | 0 |
| 交替 | | 25 | 616 | 219 | 2 | 0 |
| 継続 | | 14 | 492 | 1594 | 1 | 0 |
| 復唱 | | 0 | 7 | 9 | 2 | 0 |
| 補完 | | 0 | 4 | 8 | 0 | 0 |
| 再現率 | | 0.063 | 0.715 | 0.759 | 0.111 | 0.000 |
| 適合率 | | 0.381 | 0.507 | 0.762 | 0.400 | 0.000 |
| F 値 | | 0.108 | 0.593 | 0.760 | 0.174 | 0.000 |

表 2.14: 発話ごとの決定木性能 (自動車販売 男性エージェント)

(a) オープンテスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 65 | 160 | 151 | 0 | 0 |
| 交替 | 64 | 665 | 110 | 0 | 0 |
| 継続 | 136 | 430 | 2816 | 0 | 2 |
| 復唱 | 1 | 6 | 2 | 0 | 0 |
| 補完 | 1 | 11 | 4 | 0 | 0 |
| 再現率 | 0.173 | 0.793 | 0.832 | 0.000 | 0.000 |
| 適合率 | 0.243 | 0.523 | 0.913 | 0.000 | 0.000 |
| F 値 | 0.202 | 0.630 | 0.871 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 82 | 141 | 266 | 0 | 0 |
| 交替 | 63 | 927 | 205 | 0 | 1 |
| 継続 | 142 | 354 | 2480 | 0 | 0 |
| 復唱 | 0 | 5 | 8 | 0 | 0 |
| 補完 | 1 | 8 | 13 | 0 | 1 |
| 再現率 | 0.168 | 0.775 | 0.833 | 0.000 | 0.043 |
| 適合率 | 0.285 | 0.646 | 0.834 | 0.000 | 0.500 |
| F 値 | 0.211 | 0.705 | 0.834 | 0.000 | 0.080 |

表 2.15: 発話ごとの決定木性能 (海外旅行計画 女性エージェント)

(a) オープンテスト

| テストデータ open | | | | | |
|-------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 41 | 84 | 249 | 7 | 1 |
| 交替 | 63 | 516 | 266 | 16 | 1 |
| 継続 | 30 | 564 | 1507 | 0 | 0 |
| 復唱 | 5 | 2 | 11 | 0 | 0 |
| 補完 | 0 | 4 | 7 | 1 | 0 |
| 再現率 | 0.107 | 0.599 | 0.717 | 0.000 | 0.000 |
| 適合率 | 0.295 | 0.441 | 0.739 | 0.000 | 0.000 |
| F 値 | 0.157 | 0.508 | 0.728 | 0.000 | 0.000 |

(b) クローズドテスト

| テストデータ closed | | | | | |
|---------------|-------|-------|-------|-------|-------|
| クラス分け | | | | | |
| | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | 97 | 97 | 250 | 7 | 1 |
| 交替 | 47 | 605 | 169 | 9 | 4 |
| 継続 | 50 | 563 | 1757 | 0 | 0 |
| 復唱 | 5 | 2 | 15 | 3 | 0 |
| 補完 | 1 | 5 | 14 | 0 | 2 |
| 再現率 | 0.215 | 0.725 | 0.741 | 0.120 | 0.091 |
| 適合率 | 0.485 | 0.476 | 0.797 | 0.158 | 0.286 |
| F 値 | 0.298 | 0.575 | 0.768 | 0.136 | 0.138 |

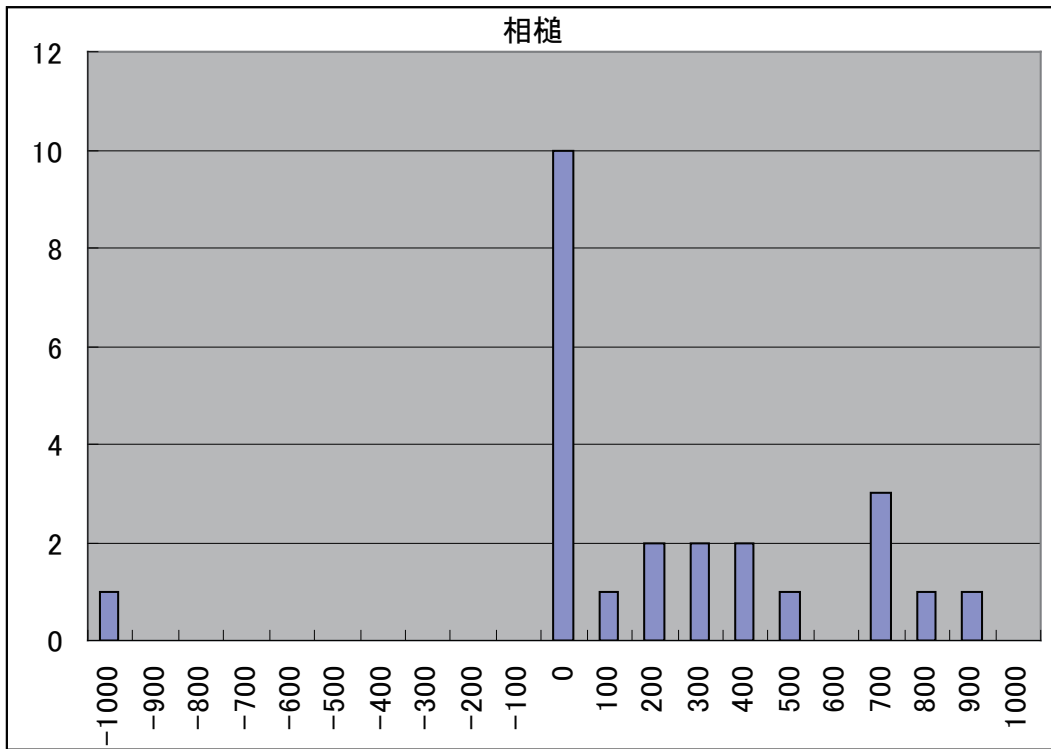
表 2.16: 発話ごとの決定木性能 (海外旅行計画 男性エージェント)

(a) オープンテスト

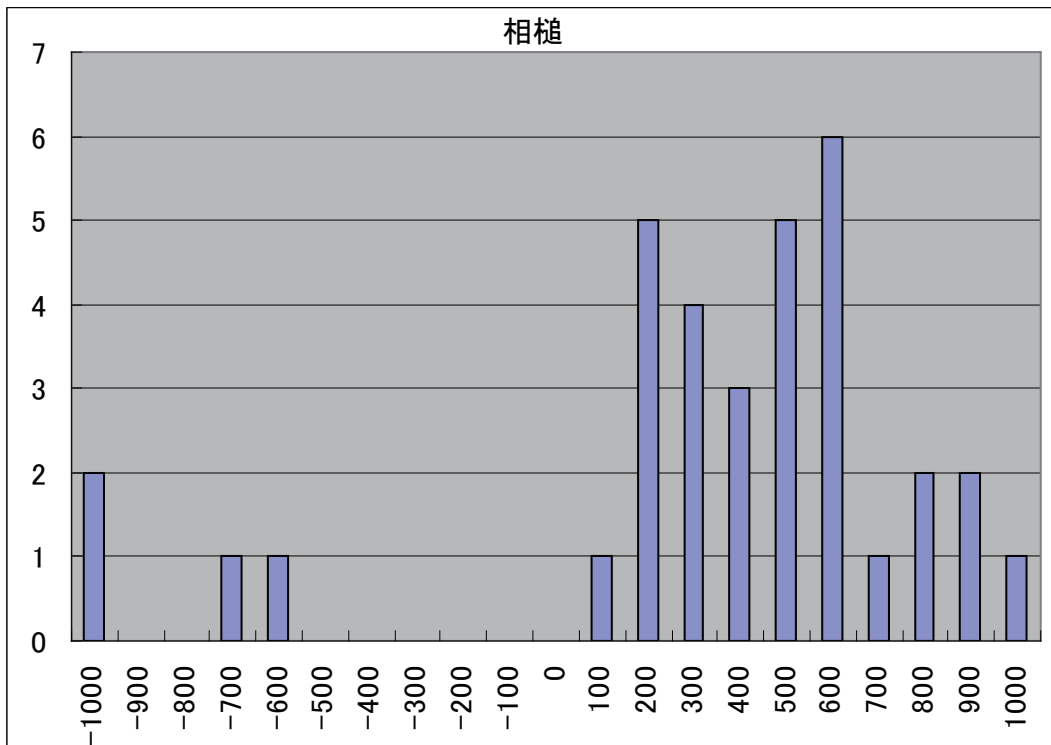
| | | テストデータ open | | | | |
|-----|--|-------------|-------|-------|-------|-------|
| | | クラス分け | | | | |
| | | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | | 68 | 175 | 227 | 3 | 16 |
| 交替 | | 132 | 811 | 239 | 2 | 12 |
| 継続 | | 12 | 322 | 2591 | 10 | 41 |
| 復唱 | | 2 | 2 | 9 | 0 | 0 |
| 補完 | | 3 | 7 | 12 | 0 | 1 |
| 再現率 | | 0.139 | 0.678 | 0.871 | 0.000 | 0.043 |
| 適合率 | | 0.313 | 0.616 | 0.842 | 0.000 | 0.014 |
| F 値 | | 0.193 | 0.645 | 0.856 | 0.000 | 0.022 |

(b) クローズドテスト

| | | テストデータ closed | | | | |
|-----|--|---------------|-------|-------|-------|-------|
| | | クラス分け | | | | |
| | | 相槌 | 交替 | 継続 | 復唱 | 補完 |
| 相槌 | | 97 | 128 | 146 | 1 | 4 |
| 交替 | | 71 | 673 | 90 | 0 | 5 |
| 継続 | | 10 | 287 | 3044 | 6 | 37 |
| 復唱 | | 0 | 5 | 2 | 2 | 0 |
| 補完 | | 2 | 7 | 4 | 0 | 3 |
| 再現率 | | 0.258 | 0.802 | 0.900 | 0.222 | 0.188 |
| 適合率 | | 0.539 | 0.612 | 0.926 | 0.222 | 0.061 |
| F 値 | | 0.349 | 0.694 | 0.913 | 0.222 | 0.092 |

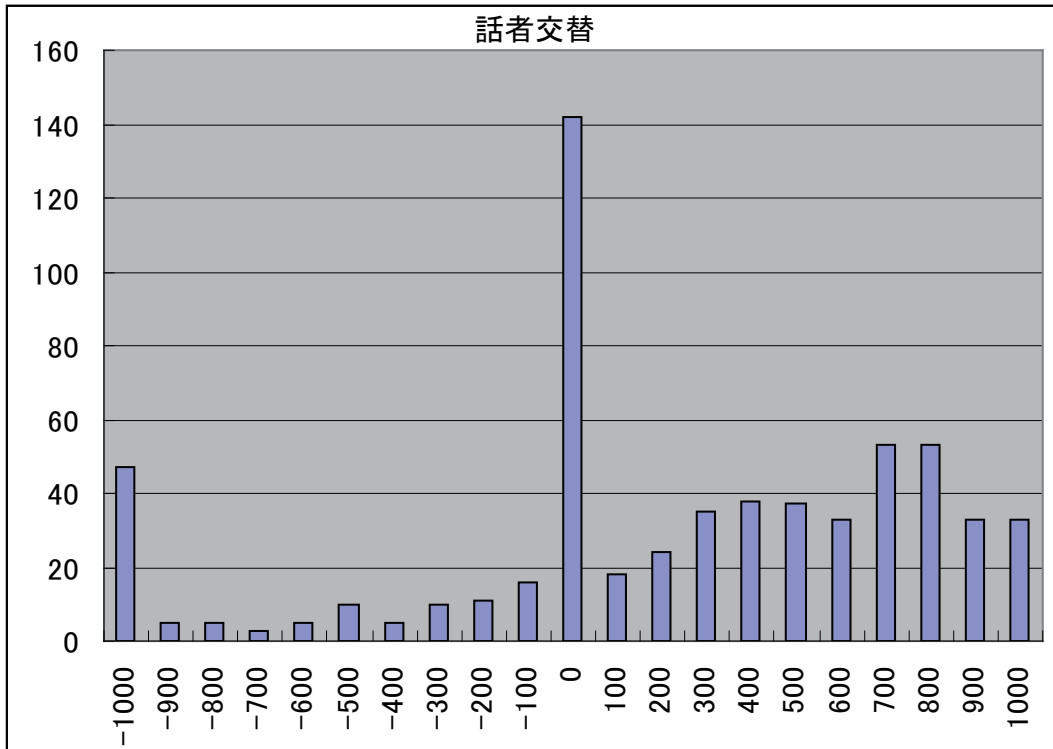


(b) クローズドテスト

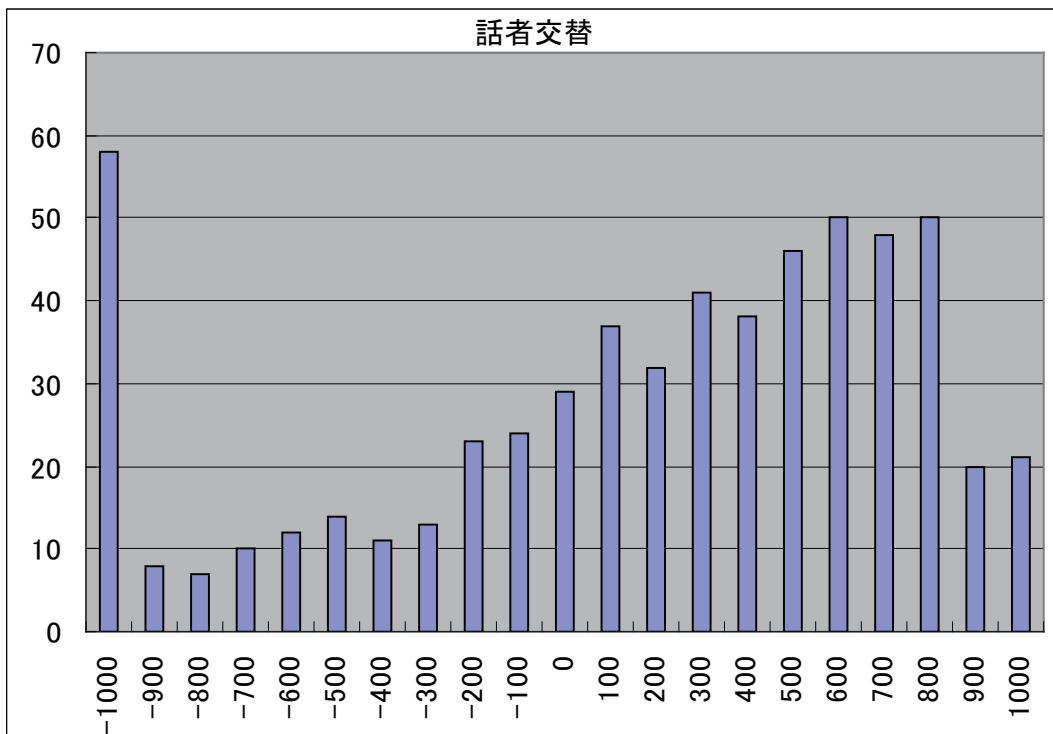


(a) オープンテスト

図 2.6: 正解からのずれ (あいづち) (自動車販売 女性エージェント)

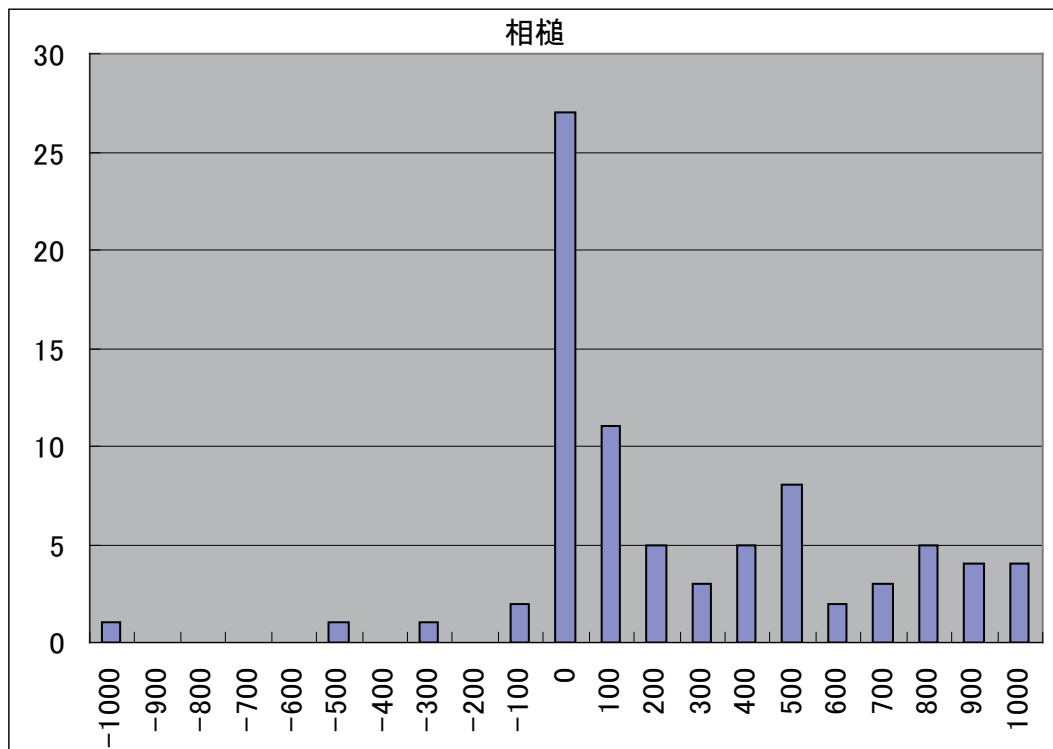


(b) クロードテスト

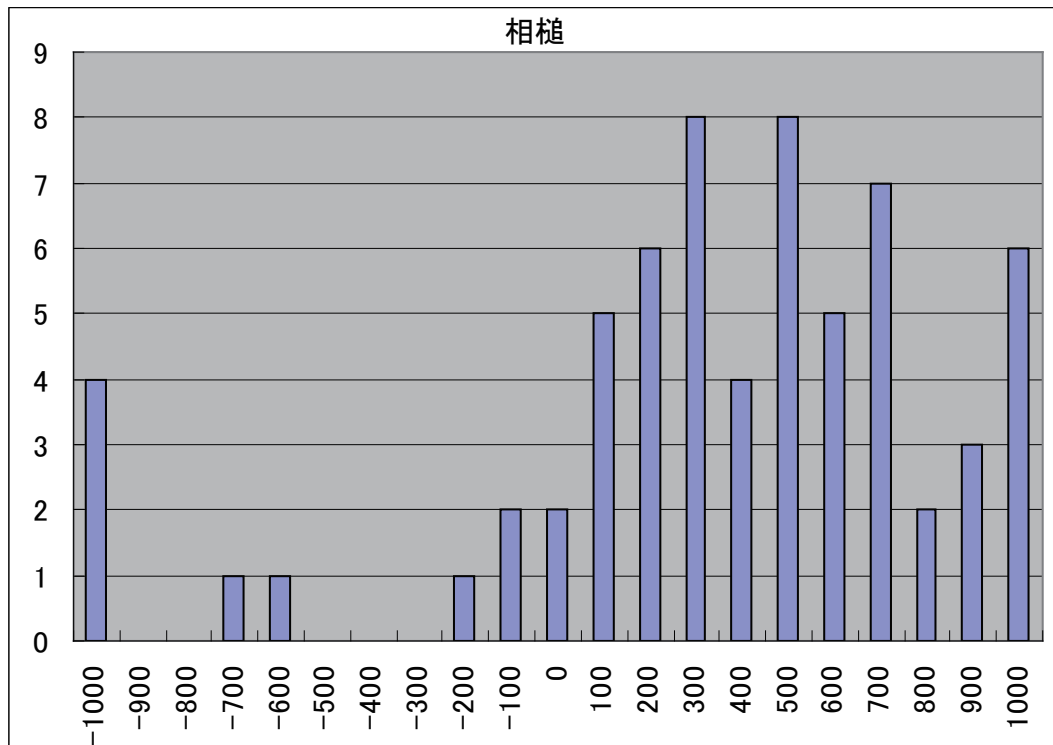


(a) オープンテスト

図 2.7: 正解からのずれ (話者交替) (自動車販売 女性エージェント)

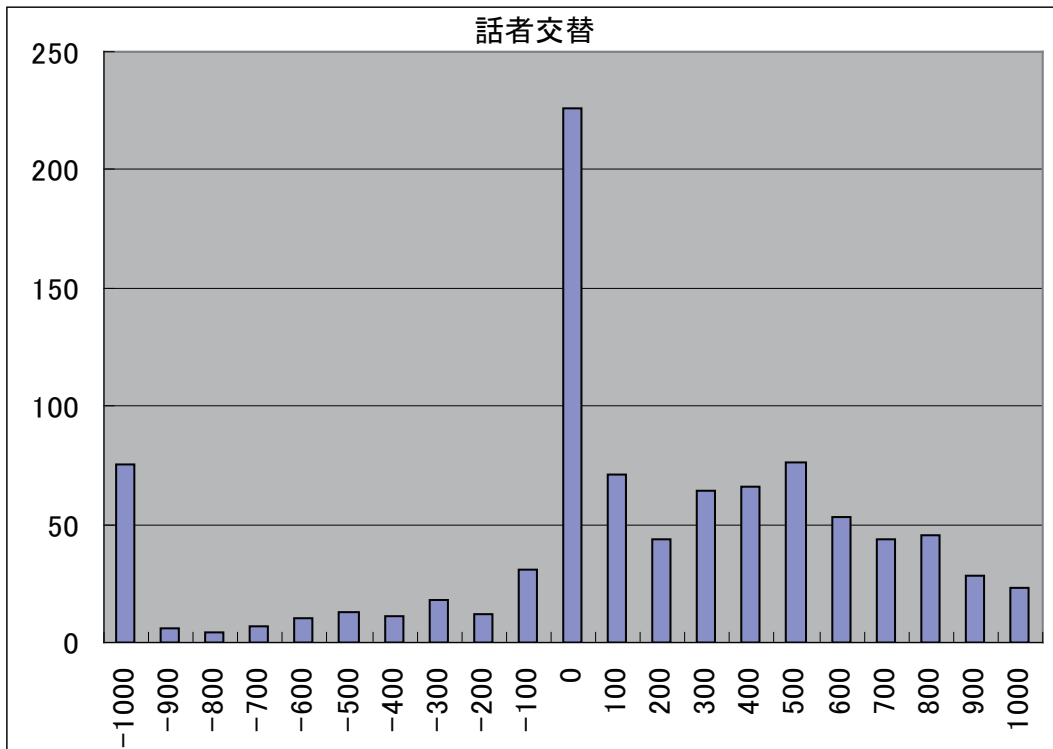


(b) クローズドテスト

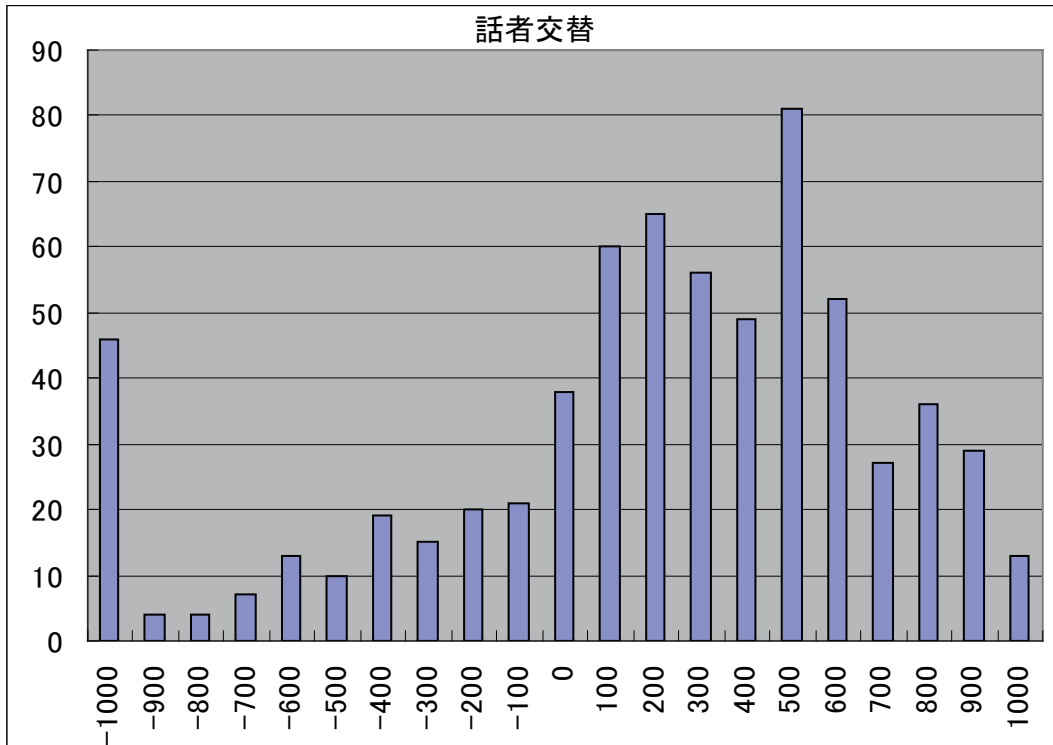


(a) オープンテスト

図 2.8: 正解からのずれ (あいづち) (自動車販売 男性エージェント)

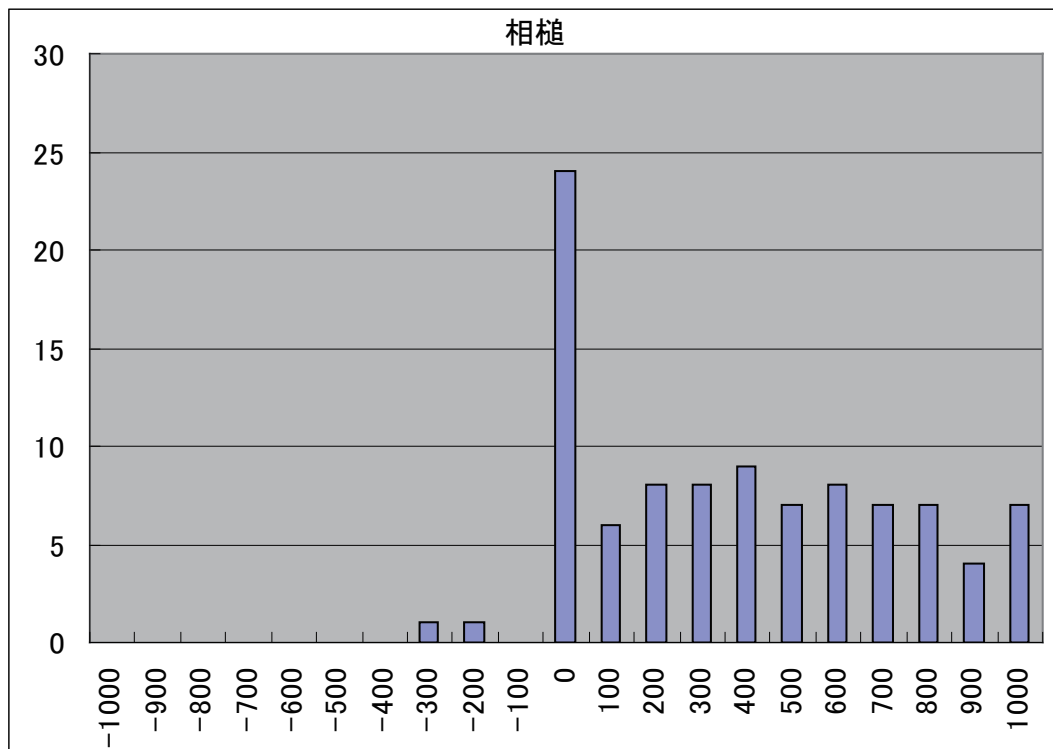


(b) クローズドテスト

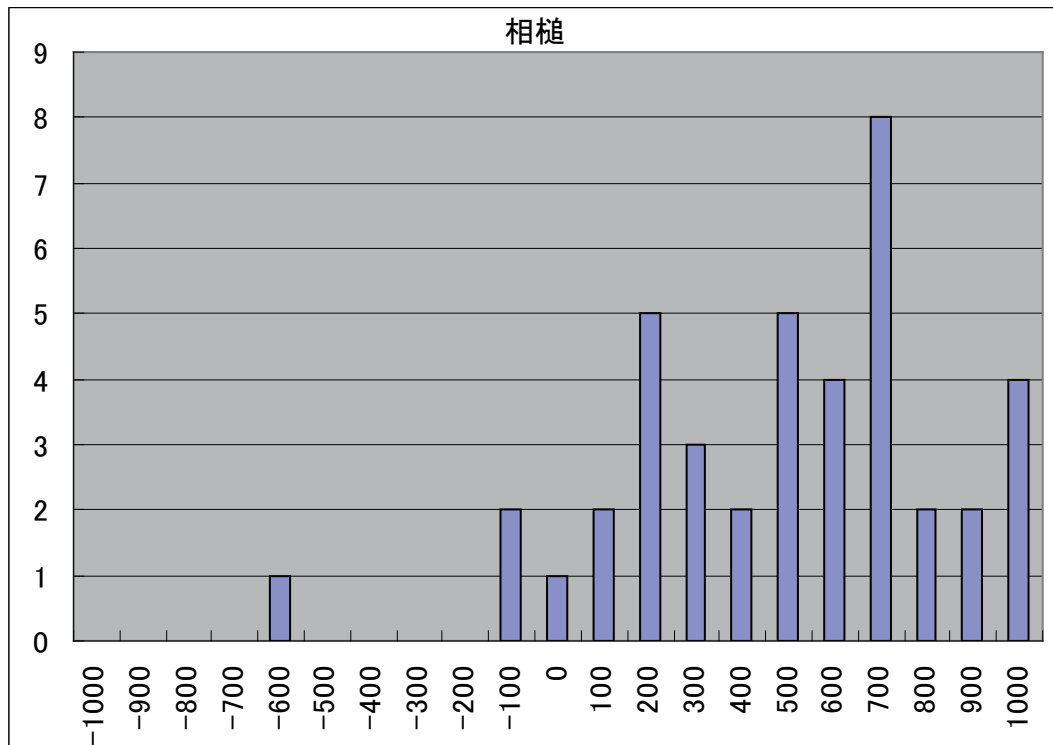


(a) オープンテスト

図 2.9: 正解からのずれ (話者交替) (自動車販売 男性エージェント)

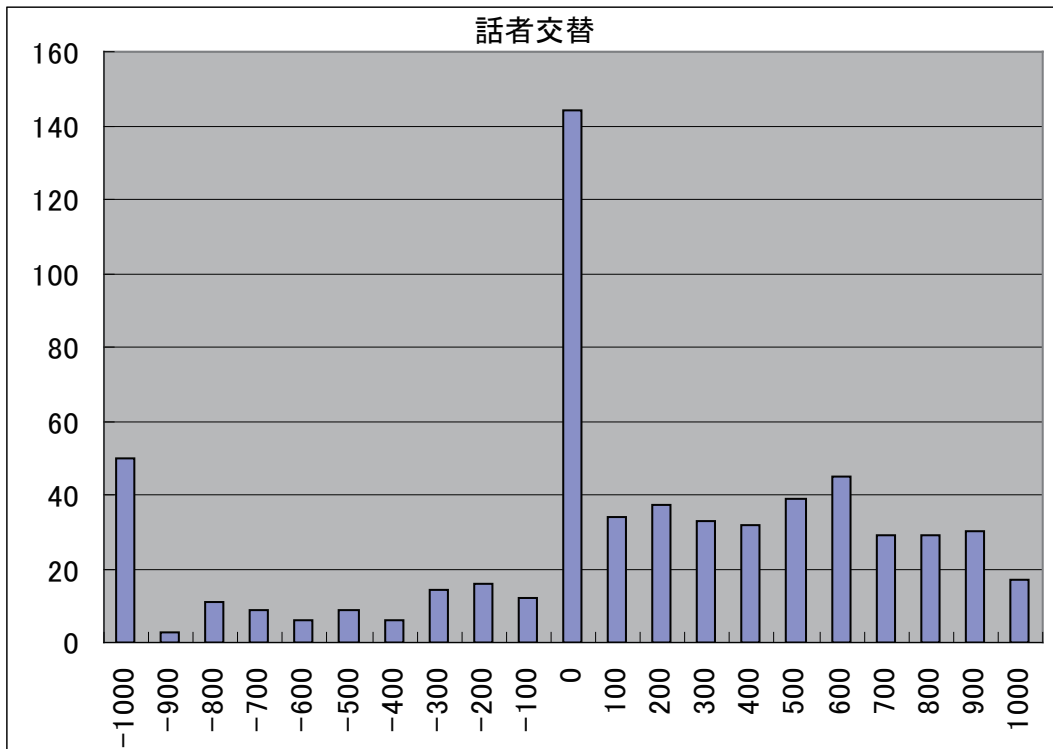


(b) クローズドテスト

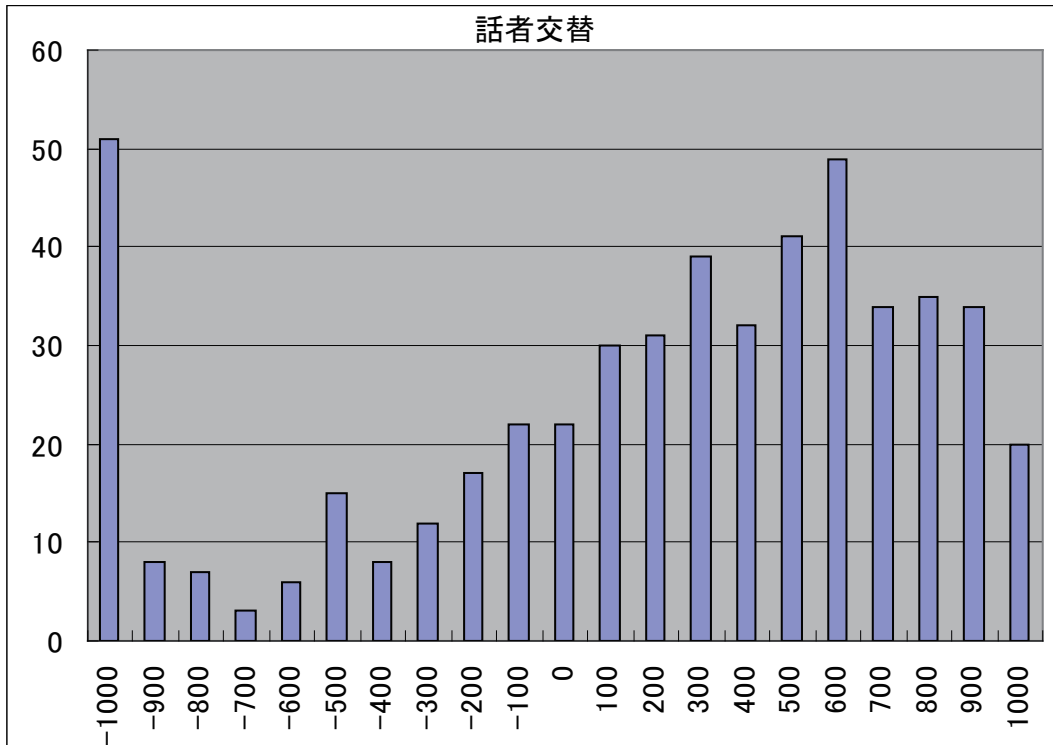


(a) オープンテスト

図 2.10: 正解からのずれ (あいづち) (海外旅行計画 女性エージェント)

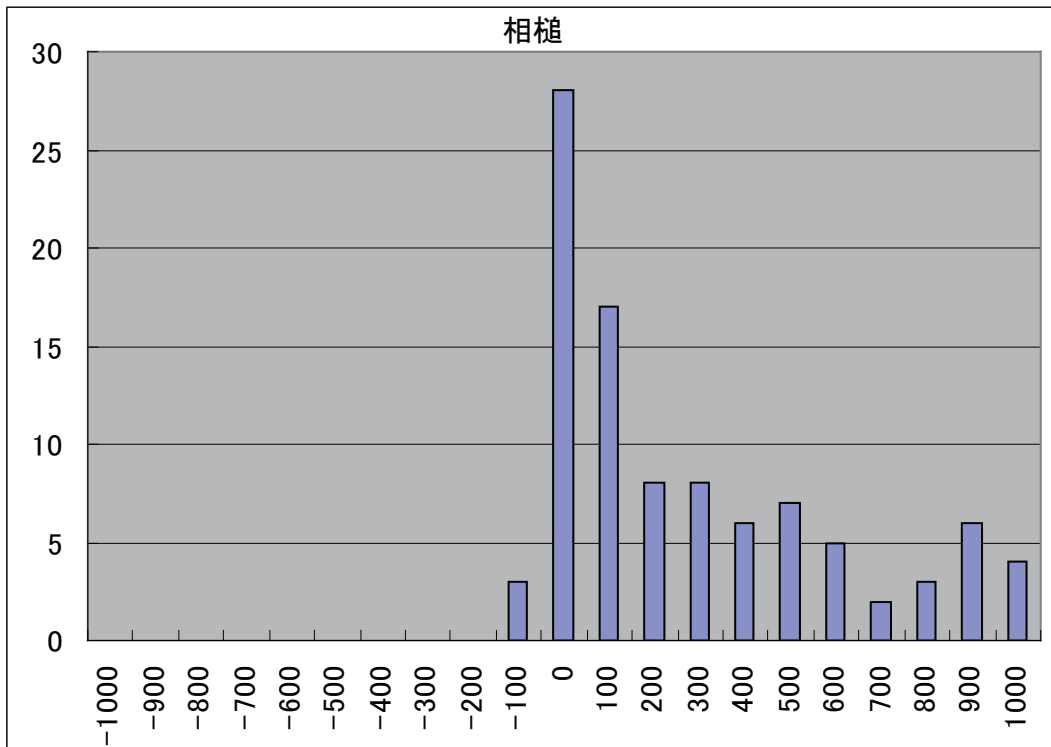


(b) クローズドテスト

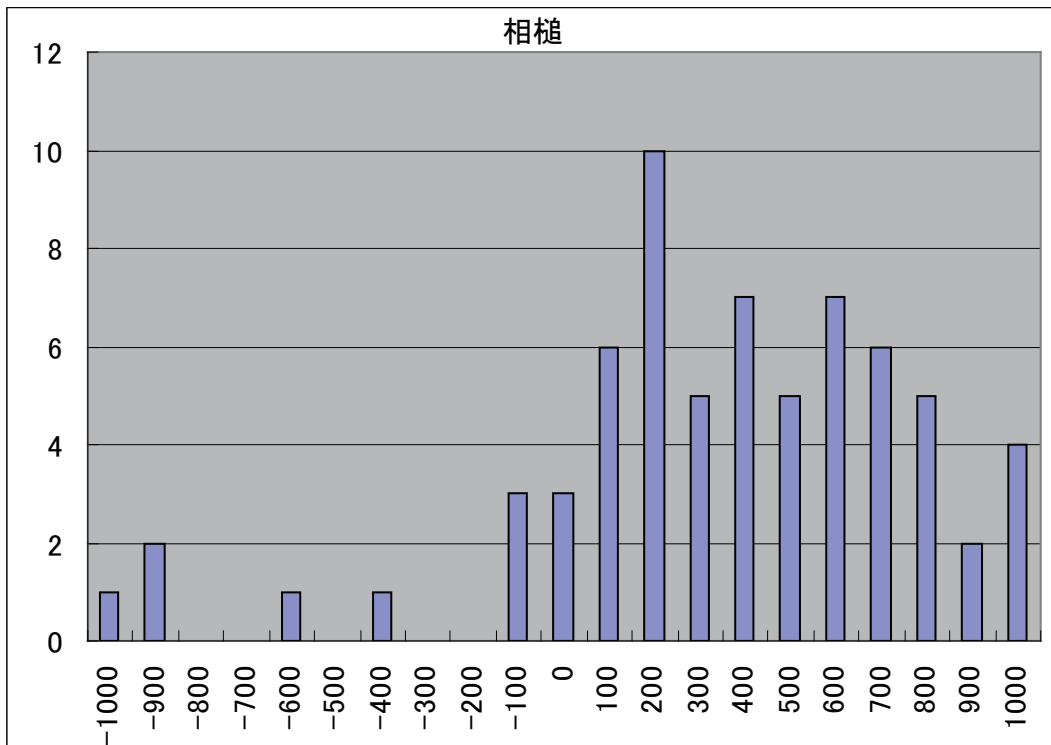


(a) オープンテスト

図 2.11: 正解からのずれ (話者交替) (海外旅行計画 女性エージェント)

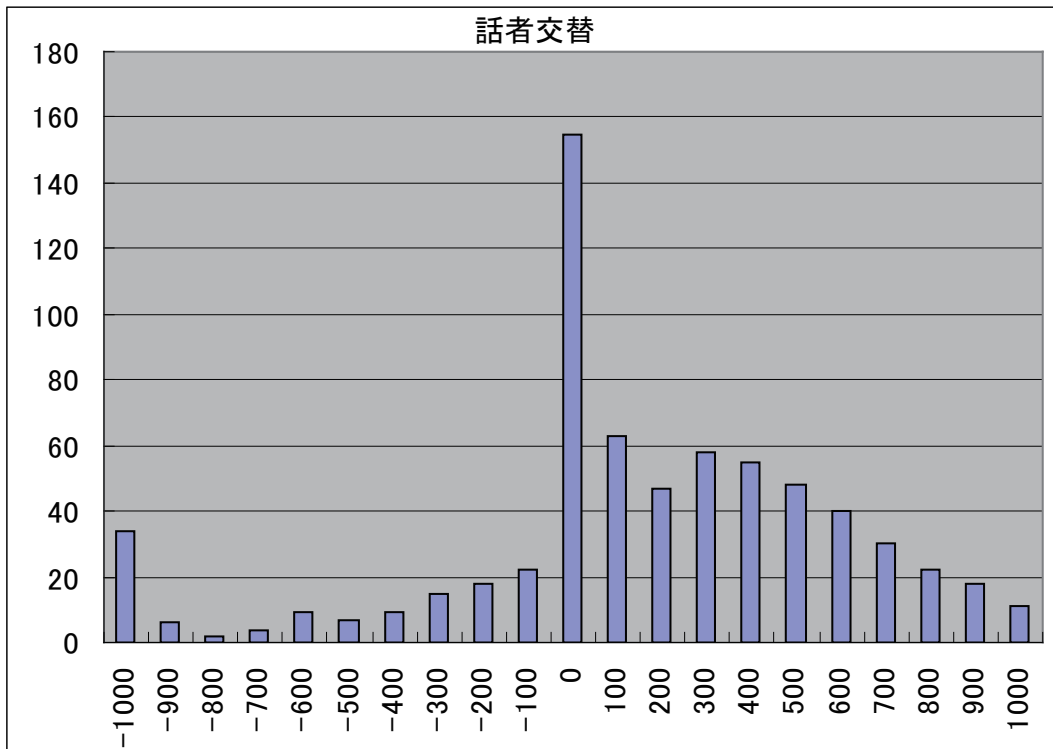


(b) クローズドテスト

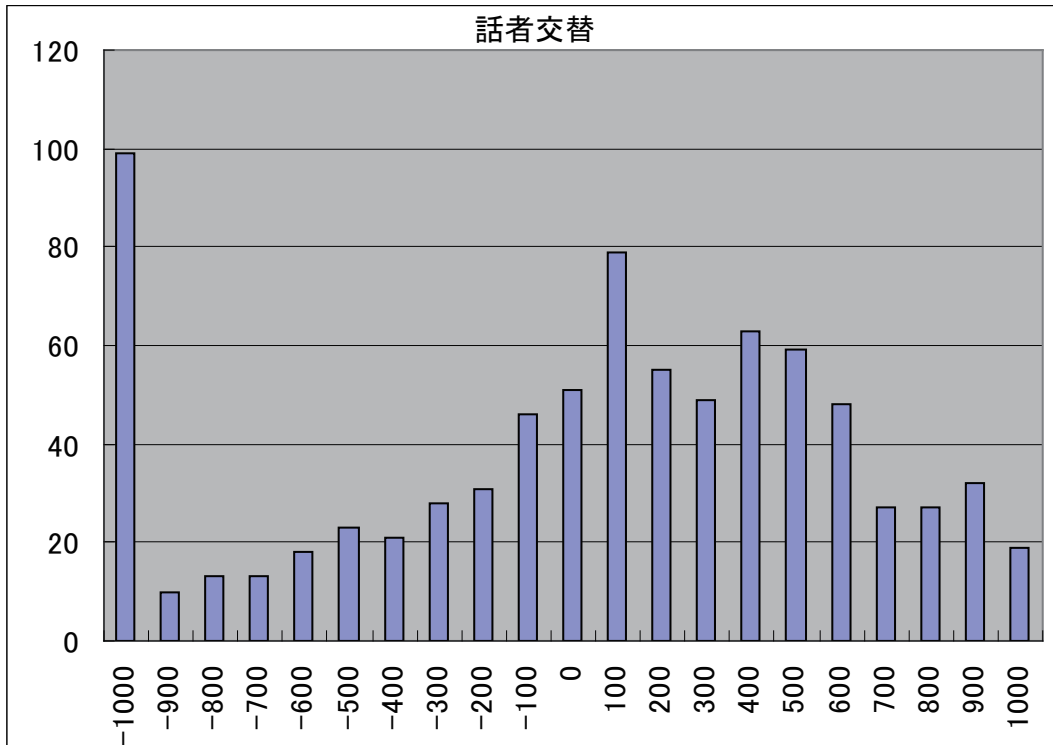


(a) オープンテスト

図 2.12: 正解からのずれ (あいづち) (海外旅行計画 男性エージェント)



(b) クローズドテスト



(a) オープンテスト

図 2.13: 正解からのずれ (話者交替) (海外旅行計画 男性エージェント)

3 韻律変化のモデル化

3.1 人間同士の対話における話者間の韻律の関係

3.1.1 対話コーパス

応答タイミング・韻律変化のモデル化に際して，人間同士の対話を調査・分析した．調査・分析に用いたコーパスは，国立国語研究所から提供されている「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese;CSJ)である [22]．CSJ コーパスは，現代日本語の自発音声を種々の研究用付加情報とともに大量に格納したデータベースであり，語数にして 750 万語，時間にして 660 時間の音声が含まれている．CSJ コーパスは，自動音声認識の研究リソースとして活用することが想定されているため，ほとんどが学会講演のような独話（モノローグ）音声であるが，対話音声としては，表 3.1 に示す内容・時間数が存在する．

表 3.1: CSJ の対話コーパス

| 種類 | 内容 | ファイル数 | 時間 |
|-----|------------|-------|-----|
| D01 | 模擬講演インタビュー | 16 | 3.4 |
| D02 | 課題指向対話 | 16 | 3.1 |
| D03 | 自由対話 | 16 | 3.6 |
| D04 | 学会講演インタビュー | 10 | 2.1 |

今回の対話コーパスの調査は，これらの対話音声に対して行った．対話は，1つ10分～20分程の長さであり，全体で 58 対話，12 時間である．

模擬講演インタビューと学会講演インタビューは，16 名により事前に行われた学会講演ないし模擬講演（10 名は両方，6 名は模擬講演のみ）に関してインタビュワーが様々な質問を発し，講演者がこれに答える形式の対話である．発話の大半は，質問に対する回答によって占められている．インタビュワーは 20 代と 30 代の女性各 1 名のいずれかが勤めている．インタビュワーは，学会講演インタビューに関しては事前に予稿集論文に目を通したうえで，また，インタビュー対象の模擬講演については，その講演の収録現場で講演を聴取したうえで，インタビューに臨んでいる．

課題指向対話では，インタビューとの対比のため，参加者 2 名（上記インタビューと同一ペア）の発話量が等しくなりやすい課題が選定されている．具体的には，実在の芸能人に講演を依頼した場合の謝礼（ギャラ）の額を想像し，その額が高い順に，芸能人 9 ないし 10 名をソートするタスク（ギャラ・タスク）である．対話開始時点で各話者に手渡されている人名リストは，わざと一致しないように作成してあ

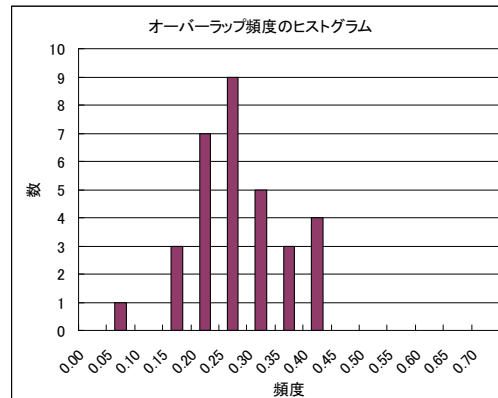


図 3.1: 各対話におけるオーバーラップ出現頻度のヒストグラム (D02 と D03)

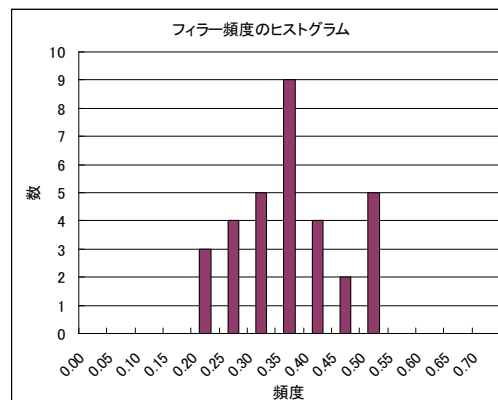


図 3.2: 各対話におけるフィラー出現頻度のヒストグラム (D02 と D03)

るので、謝礼額の推定に先立って（あるいは同時に）、推定対象となる芸能人の完全なりストを作成するための対話も必要とされる。

自由対話では、課題の制約なしに、10分程度、自由に対話を行っている。これらの4種類の対話音声は、同一の話者ペア（講演者とインタビュアー）によって発話されている。

全対話には女性対女性、女性対男性の対話があり、男性対男性対話はない。いずれも女性（インタビュアー）が、男性または女性と対話をするものである。音声収録には、対面ブースを用いており、話者はそれぞれ独立した防音ブースに入るが、ガラス越しにお互いの顔が見えるようになっている。お互いの音声は、イヤホンを通して聞くようになっている。

このコーパスの各対話に含まれるオーバーラップ頻度のヒストグラムは、図??、図??に示す。また、フィラー頻度については、図??、図??に示す。D02 と D03 は、雑談に近いタスクであり、D01toD04 は比較的形式的なタスクであるため、それぞれでヒストグラムを作成した。図を見ると、フィラーの頻度は、雑談タスクよりも、形式的なタスクの方がより多くなっている。オーバーラップに関しては、頻度に差は

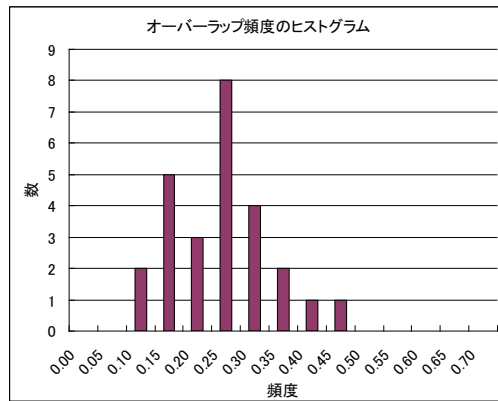


図 3.3: 各対話におけるオーバーラップ出現頻度のヒストグラム (D01 と D04)

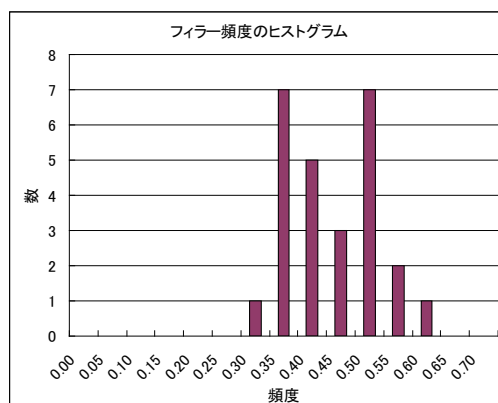


図 3.4: 各対話におけるフィラー出現頻度のヒストグラム (D01 と D04)

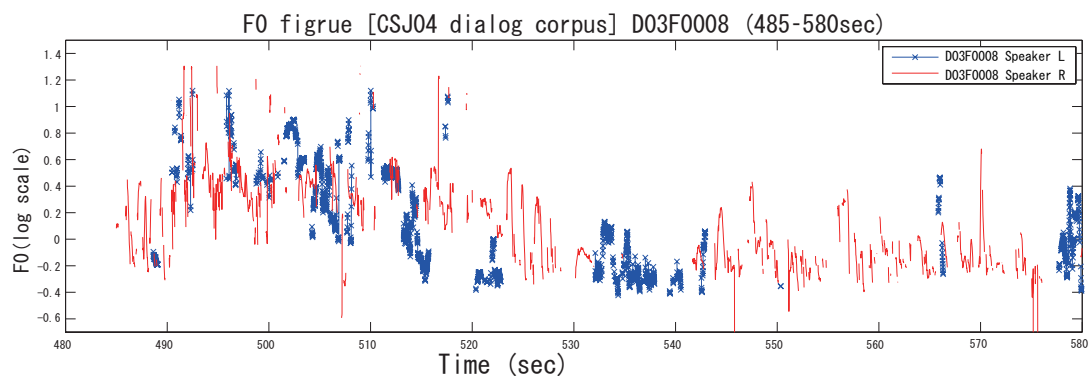


図 3.5: CSJ 対話音声の例

見られなかった。

3.1.2 対話中の 2 話者の基本周波数 (F0) の変動

コーパス中の実際の対話の音声の基本周波数 (F0) をプロットしたものを図 3.5 に示す。線の違いは話者の違いを示している。値は対数值 ($\log F0$) で、各話者の全体の平均値を揃えてある。

この図中の対話の話題は、前半が「子供に将来の夢を聞いたら“子猫ちゃん”だった」というもので、後半が「子供の言語獲得について」である。前半は、面白いエピソードであるため、二人から「笑い」も起きており、双方がよく発言し、盛り上がっている。後半は、まじめな話に入っていく、二人とも落ち着いて話している。

図を見ると、盛り上がっているところでは、F0 が高くなっており、ダイナミックレンジも大きくなっている。それに対して、落ち着いている部分では声の高さも、それぞれの平均値からあまり変化しておらず、ダイナミックレンジも小さい。

このように、対話のはずみ度合い、盛り上がり、基本周波数をはじめとする韻律には強い関係があることが予想される。それは 2 話者が互いに影響しあって強制的に変動しているものだと考えられる。

3.1.3 対話中の 2 話者の基本周波数の相関

対話をしている 2 人の基本周波数の相関をこれらのコーパスで調べた。

対話中の基本周波数は、図 3.6 に示すように、各発話中の基本周波数を発話時間で割った平均値を代表値として、1 つの発話ごとに 1 つの値で表した。また、その点は、発話開始時刻と発話終了時刻の間 (中央) にとった。2 話者の基本周波数の相関を求める際には、データ数を揃えなければならないので、一方の話者の各発話の代表点の時刻に対して、もう一方の話者の二つの代表点から補間値を求めて、デー

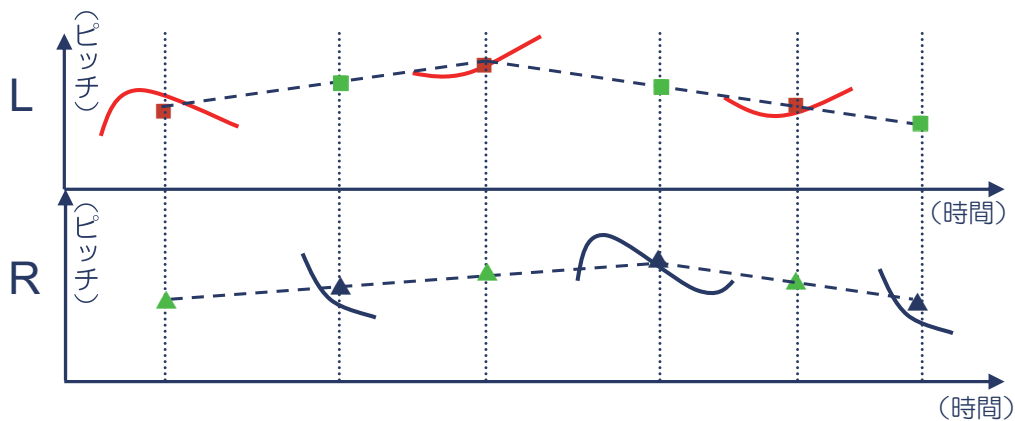


図 3.6: 各対話におけるフィラー出現頻度のヒストグラム

タ数を揃えた。

表 3.2: 各対話の F0 の相関

| 種類 | 最大値 | 最小値 | 平均 |
|-----|-------|--------|-------|
| D01 | 0.382 | 0.070 | 0.195 |
| D02 | 0.477 | -0.002 | 0.222 |
| D03 | 0.521 | 0.012 | 0.234 |
| D04 | 0.206 | -0.005 | 0.085 |
| 平均 | 0.397 | 0.019 | 0.184 |

2 話者間の対数基本周波数 $\log F_0$ の相関値は表 3.2 のようになった。一般に相関値は大きく、対話中での 2 話者のお互いの F_0 値には関連があると言える。また、58 対話中 4 対話を除いて正の相関を示しており、対話において、声の高さは相手に合わせて変化していくと考えられる。そして、対話の内容によって、相関値に違いが見られた。比較的自由的な形式の対話 (D2, D3) は、インタビュー形式のもの (D1, D4) に比べて相関が高い。つまり、自由的な形式の対話では基本周波数が同調する傾向が高いということを示している。また、性別による違いもあり、相関値が高いものには、女性同士の対話が多かった。一方、女性と男性の対話は、相関値の低いものが多かった。コーパスの対話 ID の [M,F] は、性別を表している。

実際に音声聞いて比べてみると、相関が高い対話は、盛り上がる場所や落ち着いた場所が良く一致している。あいづちも、相手の声の調子に応じてうまく変化させており、それによって対話がスムーズになっている感じが感じとれる。

このことから、音声対話システムにおいても、韻律情報 (F_0) を相手に合わせて変化させることで、対話を盛り上げることが出来る可能性がある。どのように変化させれば対話を盛り上げることが出来るかは、実際にコーパスを調べて、モデルを作成することでシステムに適應させる。

相関が低い対話は、両方が淡々と喋っている感じがある対話や、片方の人の喋り方が特徴的で、もう一方の人が単調に喋っているときに、それに対して抑揚を付けて応答している対話であった。

そして、対話を実際に聞くと、対話している二人が同年代であったり、立場的に近いような場合（音声を聞いてそのように感じた場合）には、よく盛り上がり、相関が高くなっているように感じられる。それとは逆に、どちらかの立場が上であるような場合には、二人が大きく盛り上がるということが少なく、相関も低くなっているのではないかと思われる。

3.2 対話の印象と対話現象との関係

CSJ コーパスの対話音声を実際に人間が聞いた場合の各対話の印象と、コーパス中の現象（2話者の logF0 相関値、オーバーラップ頻度、フィラー頻度）との関係を調べた。4名に対話音声を聞いてもらい、各対話について、以下の各項目について5段階のアンケートをとった。

- 相手との親しさ
 - 親しみを持っている (5 4 3 2 1) 遠慮している
- 盛り上がり
 - 良い (5 4 3 2 1) 盛り上がってない
- 相手の意見に同意しながら対話
 - 同意しながら対話 (5 4 3 2 1) 意見を戦わせながら対話
- 立場の違い
 - 目上 (5 4 3 2 1) 目下
- L話者 (インタビュアー) のフランクさ
 - 気を使っている (5 4 3 2 1) くだけている
- R話者 (インタビュイー) のフランクさ
 - 気を使っている (5 4 3 2 1) くだけている
- L話者 (インタビュアー) の表現
 - 敬語ばかり (5 4 3 2 1) 敬語をつかっていない
- R話者 (インタビュイー) の表現
 - 敬語ばかり (5 4 3 2 1) 敬語を使っていない

アンケートは、この形式のアンケート用紙に丸を付けて回答する形式にした。

「相手との親しさ」、「盛り上がり」に関しては、対話中の2話者からの印象の平均を考えて評価値を付けるようにした。「立場の違い」については、インタビュアー (L話者) 側から見て相手が目上か目下かというところを評価して値を付けてもらった。また、「フランクさ」は言い方の違いを示すものであり、声の調子などから受ける印象を評価してもらった。「表現」は文字上でどのようになっているかを示してお

り、語彙的に敬語を用いているかどうかを評価してもらった。

被験者は4名（男性1名、女性3名）であり、実験に際して、「正式な聞き取り実験の前に、10分間程度評価サンプルからランダムに対話サンプルを聞いて雰囲気をつかむこと」と「各対話について、解答する際に全体をかいつまんで5分以上は聞くこと」を注意事項として伝えた。

ここで、アンケート結果の被験者間での違いを見る為に、被験者間での結果の相関を調べた。「盛り上がり」に対するアンケート結果の相関値は表3.3のようになった。平均値は、0.470であった。

表 3.3: アンケート被験者同士の「盛り上がり」の相関

| 被験者 | 相関値 |
|-------------|-------|
| 被験者1 - 被験者2 | 0.523 |
| 被験者1 - 被験者3 | 0.538 |
| 被験者1 - 被験者4 | 0.630 |
| 被験者2 - 被験者3 | 0.314 |
| 被験者2 - 被験者4 | 0.342 |
| 被験者3 - 被験者4 | 0.476 |
| 平均 | 0.470 |

この結果から、「盛り上がり」に関しては各被験者間で相関があることが分かる。このことから、各対話の盛り上がりの指標として、このアンケート結果を用いることとする。また、その他のアンケート結果については、表3.4のようになった。「話者の表現」への回答が比較的バラツキがあったようである。

表 3.4: 各アンケート項目の被験者間の相関の平均値

| アンケート項目 | 相関値の平均 |
|-----------|--------|
| 親しさ | 0.444 |
| 盛り上がり | 0.470 |
| 同意・否定 | 0.387 |
| 立場 | 0.478 |
| L話者のフランクさ | 0.399 |
| R話者のフランクさ | 0.384 |
| L話者の表現 | 0.300 |
| R話者の表現 | 0.262 |

アンケート結果の評価値と、各対話音声の情報「対話中のオーバーラップ頻度」、
「対話中の2人のF0相関値」、「対話中のフィラー頻度」との相関を表3.5に示す。こ

ここで、「対話中のオーバーラップ頻度」とは、対話中に話者交替がオーバーラップして起こった回数を、対話中の話者交替の総数で割ったものである ((3.2) 式)。また、「対話中のフィラー頻度」は、対話中にフィラーが発話された回数を、全発話数で割ったものである ((3.3) 式)。

$$\text{対話中のオーバーラップ頻度} = \frac{\text{オーバーラップした話者交替数}}{\text{対話中の話者交替の総数}} \quad (3.2)$$

$$\text{対話中のフィラー頻度} = \frac{\text{フィラーの発話数}}{\text{対話中の発話の総数}} \quad (3.3)$$

表 3.5: 被験者評価値と対話現象との相関

| | F0 相関値 | オーバーラップ 頻度 | フィラー 頻度 |
|-----------|--------|---------------|------------|
| 親しさ | 0.348 | 0.627 | 0.072 |
| 盛り上がり | 0.350 | 0.718 | 0.127 |
| 同意・否定 | 0.279 | 0.638 | 0.090 |
| 立場 | -0.282 | -0.098 | 0.267 |
| L 話者フランクさ | -0.283 | -0.417 | 0.108 |
| R 話者フランクさ | -0.266 | -0.637 | 0.047 |
| L 話者の表現 | -0.340 | -0.238 | 0.265 |
| R 話者の表現 | -0.182 | -0.404 | 0.289 |
| 年の差 | -0.483 | 0.068 | -0.411 |
| 実際の年の差 | -0.301 | -0.171 | -0.002 |

表 3.5 中の‘年の差’は、‘立場’の目上・目下の評価値を、目上か目下かに関係なく基準(レベル3)からどれだけ離れているかという値にしたものである。つまり、[目上 5 - 4 - 3 - 2 - 1 目下]を、絶対値としての年齢差として、それぞれ[3 - 2 - 1 (基準) - 2 - 3]という風に変換したものである。そして、‘実際の年の差’は、コーパスデータに記録されていた、実際の話者間の年齢差を数値として持ったものである。

表 3.5 を見てみると、オーバーラップ頻度は、全体的に高い相関値を示している。「親しさ」「盛り上がり」「同意・否定」は、オーバーラップの頻度が高いと、評価値も高くなっている。つまり、オーバーラップがたくさん起こっている対話では、親しさがあったり、盛り上がっていたり、同意して対話が進んでいるということである。「フランクさ」に対しては、負の相関が高いので¹、くだけていると感じた対話にて、より多くオーバーラップが起こったということである。

¹フランクさの評価値のつけ方がまずく、評価点を逆にしたほうがよかった。

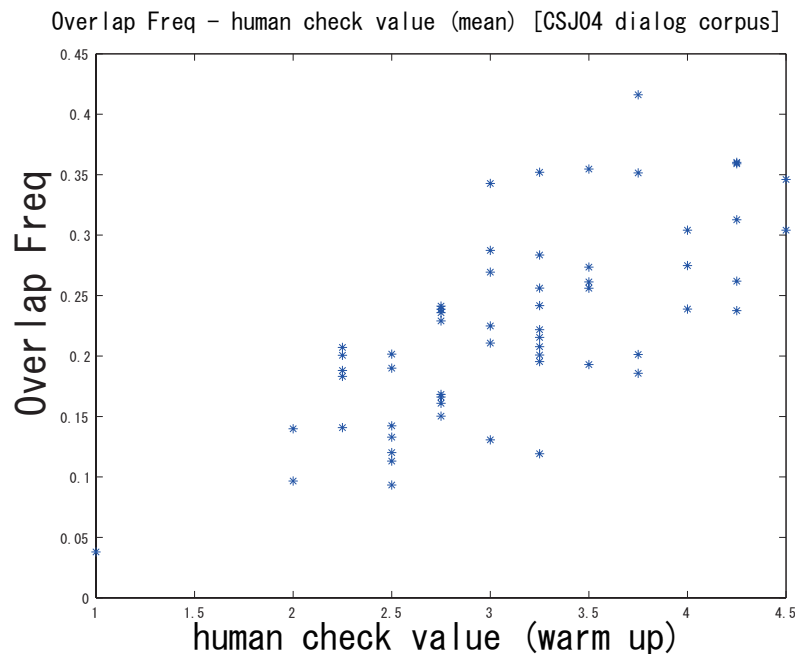


図 3.7: オーバーラップ頻度と被験者評価値 (盛り上がり)

F0 相関値に関しては、「親しさ」「盛り上がり」「同意・否定」と正の相関関係がある。つまり、基本周波数が 2 話者間で同調して進んでいる対話は、親しさがあつたり、盛り上がっていたり、同意して対話が進んでいるということである。また、F0 相関値が高いとくだけた対話の印象になると言える。

フィラー頻度に関しては、「立場」と「表現」にて正の相関がある。これは、相手が目上であつたり、敬語を使っていたりする場合には、フィラーが起こりやすいことを示している。

‘年の差’の項目を見てみると、‘F0 相関値’‘フィラー頻度’のそれぞれについて、高い相関が出ている。これは「年齢差があると、F0 の同調性が無くなり、またフィラーの回数が少なくなる」ということである。また、‘実際の年の差’は、‘フィラー頻度’の相関がなく、アンケート結果と違っている。

図 3.7 は、オーバーラップ頻度と被験者の評価値 (盛り上がり) の関係をプロットしたものである。この図を見ても、オーバーラップ頻度と盛り上がりとの間に、相関があることが分かる。

3.3 韻律情報変化のモデル化

韻律情報として、基本周波数 (F0) のモデル化を考える。実際の人間同士の対話では、お互いの F0 の変化には相関があることが分かったので、そのことを踏まえ、

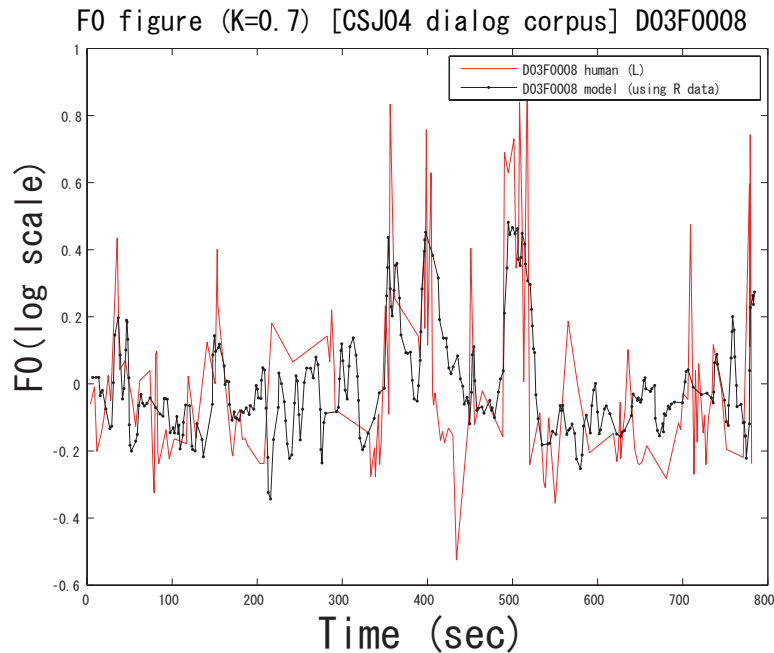


図 3.8: 対話音声の F0 とモデル出力値

相手の F0 の変化をみて、その変化に合わせてシステム側の F0 を変動させるようにモデル化を行った。

モデルは、目標値に対して時定数を持って変動するように (3.4) 式を用いた。

$$\begin{aligned}
 M(t) &= \mu_{sys} + \alpha_{sys}(t) \\
 \alpha_{sys}(t) &= \alpha_{sys}(t-1) + K(\alpha_{usrN\mu} - \alpha_{sys}(t-1))
 \end{aligned}
 \tag{3.4}$$

ここで、 $M(t)$ は対話ターン t におけるモデルの値である。 μ_{sys} は、時間によって変化しない、システムの標準値 (平均値) を表す。 $\alpha_{sys}(t)$ は、対話ターン t での、システムのオフセット値である。 $\alpha_{usrN\mu}$ は、ユーザの直前 N 発話のオフセット値の平均である。これが目標値である。 $\alpha_{sys}(t-1)$ は、 t の 1 ターン前のシステムのオフセット値である。 K は、時定数を表す。

ここで、現在は $K = 0.7$ 、 $N = 3$ を用いている。最適値の設定は今後の課題である。

このモデルの確認のため、人間同士の対話のうちの一方向の話者の韻律を (3.4) 式に当てはめて得られた値と実際のもう一方の話者の韻律とを比較した。これを図 3.8 に示す。両方の値は対数値であり、それぞれ平均値を引くことで正規化している。この図の場合には、話者 R の値をモデルに入力して得られた値が示されており、この結果は、システムが話者 L 側になったことを想定して出力した結果であるので、正解として実際の話者 L の値を描画してある。

CSJ コーパスの対話音声とモデル値の相関は、表 3.6 のようになった。モデルの

出力値と実際の値との相関は，正の相関を示しており，また，実際の2話者間の基本周波数の相関値（表 3.2）と同程度になっており，対話中の韻律の変動をモデル化することが出来ていると考えられる．

表 3.6: モデルと実際の値との相関

| 種類 | 最大値 | 最小値 | 平均 |
|-----|-------|--------|-------|
| D01 | 0.427 | -0.105 | 0.136 |
| D02 | 0.478 | -0.025 | 0.205 |
| D03 | 0.553 | -0.016 | 0.208 |
| D04 | 0.257 | -0.212 | 0.080 |
| 平均 | 0.429 | -0.090 | 0.157 |

また，「人間同士の F0 相関値」と「モデルと正解との相関値」の相関を見るために，お互いの関係を図にプロットし，相関を求めた．図 3.9 に，プロット図を示す．x 軸が人間同士の相関であり，y 軸がモデルと正解との相関である．

相関値は，0.859 であった．図を見ても分かるとおり，非常に強い相関が見られる．これは，人間同士の F0 相関値が高い場合には，今回のモデルが良く働いているということである．

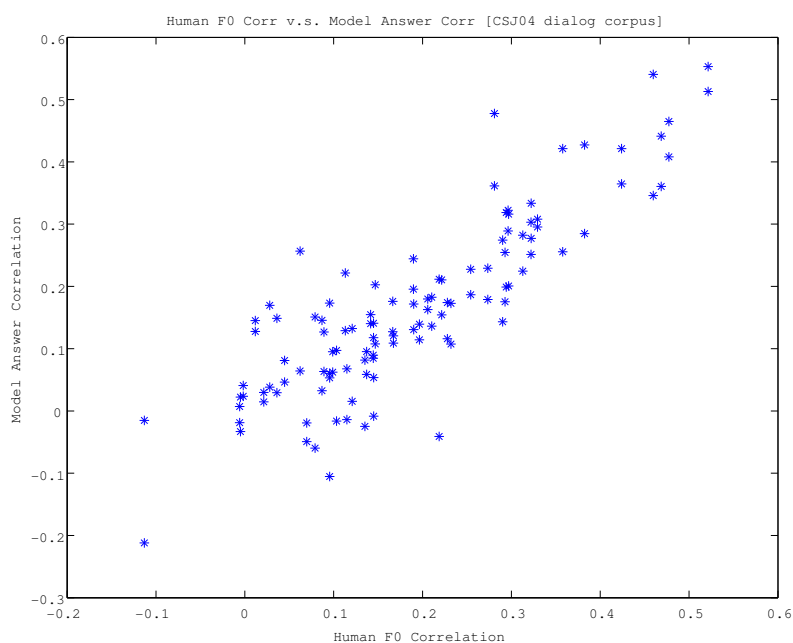


図 3.9: 人間同士の F0 相関値 vs モデルと正解の相関値

4 音声対話システム

現状の音声対話システムは、発話してもシステム側からの応答がない、発話途中でシステムに入力されているかがわからないといった問題がある。そこで本研究では、前述のリアルタイム応答生成手法を、対話システムの応答タイミング生成部に応用し、自然な対話の中でリアルタイムにあいづちや話者交替などのタイミングを検出して応答できるシステムの構築を行った。本システムは、ユーザの発話に適度にあいづちを打ち、システムの発話タイミングを自動で検出して、ときにはユーザの発話に対して割り込むことが可能である。

4.1 想定タスク

本システムは、システムと人間とがより円滑に対話することを目的とする為、システムの想定タスクを天気に関する話題の対話とした。天気に関する話題は、対話の中でユーザが天気を尋ねるだけでなく、その地域が暑いかなど話題が発展することが容易であり、しかも、特定の語彙に限定しやすいタスクであると考えられる。また、WWW から最新のデータを更新できる。

天気案内を行うシステムの例としては、堂坂ら [23] が気象情報案内を行う対話システム「飛遊夢(ひゅーむ)」の構築を行っている。この対話システムでは気象情報を案内する機能の他に、ユーザが「東京に大雨警報はありますか？」という発話を行い、かつ全国に大雨警報が発令されていないとき、もし東京という地名情報がとれない場合でも、システム発話として「警報はありません」と返せるような対話戦略を採用している。また、Zibert ら [24] は、スロベニア語とクロアチア語のどちらの言語でも対話できる天気情報検索システムを構築している。また、マサチューセッツ工科大学と NTT コミュニケーション科学基礎研究所では、ユーザがシステムと電話で対話を行うことで天気情報を照会できる気象情報電話対話システム Jupiter (日本語名 Mokusei) を開発している [25]。しかしながらこれらの対話システムはいずれも目的指向の対話システムであり、本研究で扱うような自然な対話ではない。

ここで、今回想定しているタスクは雑談に近いが、雑談とは岡田ら [26] によれば、「伝えようとして伝わること」と「結果として伝わること」に分類できるとされる。つまり、前者はある「プラン」や「ゴール」から議論を始めることが多いとし、後者は「行為の結果として後から出現する」ものとしている。本システムの中での雑談は、これらの雑談の特徴の両者を考慮したものを考える。それはユーザは事前に大体の対話プラン(聞きたい天気情報など)は持ちながらも、そのプランを達成することに固執せず、システムが時折振ってくる話題に対して楽しく会話ができると

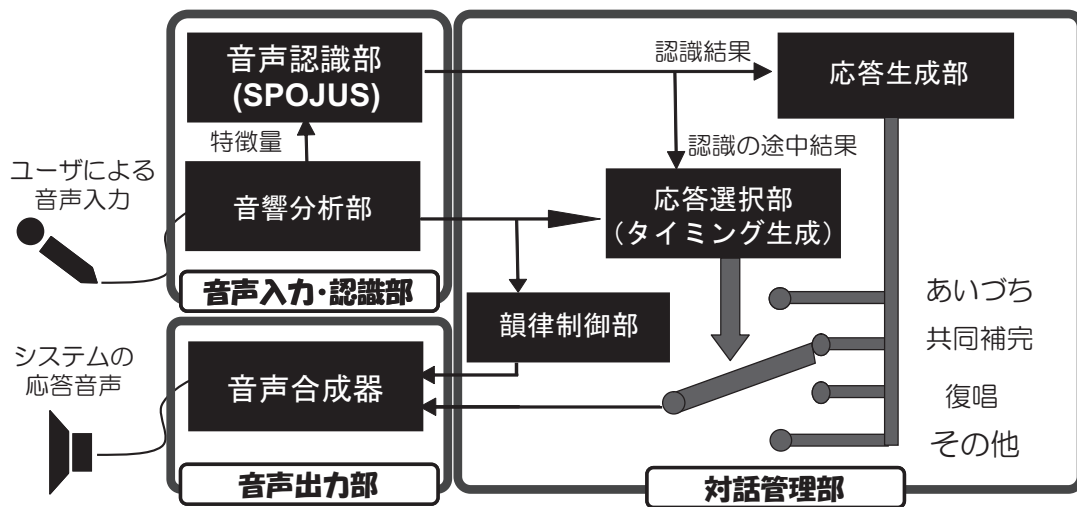


図 4.1: システムの構成図

いうもの为目标とする。雑談タスクはユーザ主導でもシステム主動でもなく、混合主動で対話を自由に進められるので、発話タイミングが重要な要素となるタスクである。

4.2 対話システムの構成

3.2節より、話者の韻律（基本周波数）やオーバーラップ発話是对話のスムーズさや盛り上がりと強い関係があることが分かった。そこで、韻律の変動やオーバーラップといった現象を対話システムが実現するための仕組みを考察し、実装して構築した我々の音声対話システムの概略を説明する。

システムの構成は、図 4.1 に示すように、大きく分けて 3 つの部分からなっている。音声入力・認識部は、入力された音声を解析し、音声認識とピッチ（基本周波数）・パワーの計算をする。対話管理部は、認識結果と韻律情報から応答文と応答タイミングを生成する。音声出力部は、応答文を音声にて出力する。この流れに沿って、リアルタイムに音声認識と韻律解析を行う。ポーズの検出をしてから応答をする従来のシステムと違い、ポーズを検出する必要がないため、応答の遅延の問題を解消している。また、これにより、システムはユーザ発話への割り込み（オーバーラップ）応答が可能になっている。また、韻律の制御も実装した。図中の韻律制御部は、人間（ユーザ）の韻律の変動にしたがって、システムの韻律を制御して出力することができる。すなわち、3.2節で分かったような対話者間での同調が実現可能となっている。

対話管理部は、さらに図 4.2 のように 5 つの部分に分かれている。音声認識・韻律

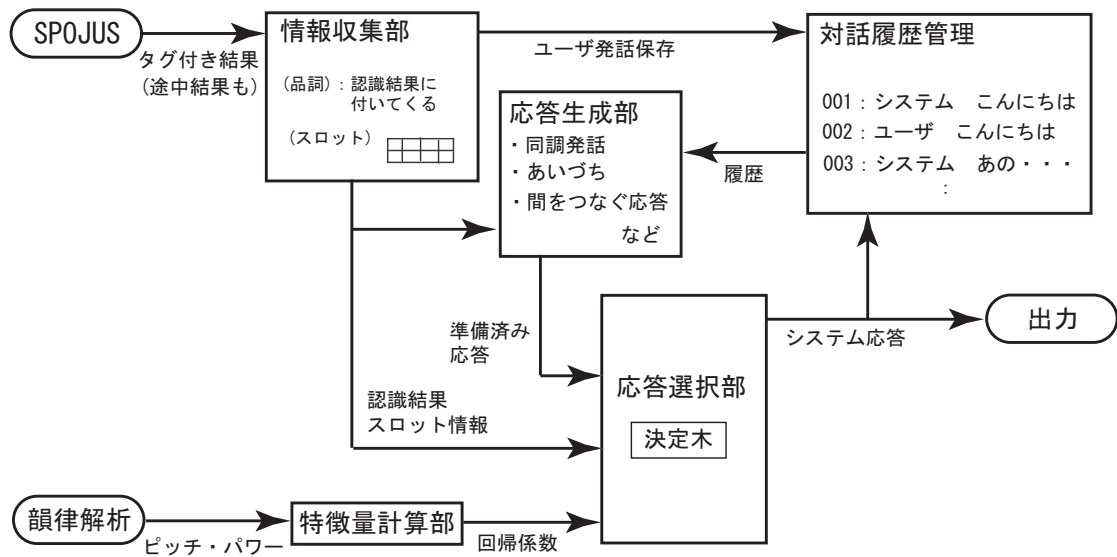


図 4.2: 対話管理部の構成図

解析部より送られてくる認識結果と韻律解析データを、管理部内の各部分で順に処理していき、応答文を生成する。応答文を適切なタイミングで出力するために、決定木によってどの応答をどのタイミングで出力するかを決定する。韻律情報であるピッチ・パワーについては、回帰係数によって傾きを求め、その値を決定木の素性として用いる。

音声認識器の出力である認識途中結果や認識結果は、情報収集部に送られ、そこで必要な情報を収集し、それを情報スロットに格納しておく。また、収集された情報は応答生成部にも送られ、そこで応答文を生成する。応答文は複数用意され、生成された応答文の中から適切なものを決定木により選択し出力する。応答文の生成には、認識結果と情報スロットを元に ELIZA [27] 方式にて生成する方法を採用している。応答文のテンプレートは、各雑談現象ごとに用意されており、認識結果と情報スロットを見て、それぞれに対して応答文が生成される。ユーザ発話とシステム応答は、対話履歴管理部にて保存される。認識結果、情報スロットの値、応答文の準備状況を応答選択部の決定木に入力することで、応答タイミングと応答に用いる応答文を決定する。決定木により選択された応答文は出力部へ送られる。

タスクとしては、天気情報案内を想定しているため、天気情報を対話に取り入れる必要がある。その為の天気情報は、Web よりダウンロードする。

以下に、それぞれの部分を詳しく記載する。

4.2.1 音声認識・韻律解析部

音声認識・韻律解析部は，ユーザがマイクにより入力した音声を音声認識器にて認識している．それと同時に韻律解析器 [28] [29] にて解析し，ピッチ (F0)・パワーを解析する．音声認識には本研究室で開発された SPOJUS [30] [31] を用い，300 語程度の単語辞書を用いて認識を行っている．言語モデルとしては文脈自由文法で書かれたものを用いている．なお，使用した単語辞書，文脈自由文法をそれぞれ付録 A.2.1，A.2.2 に示す．音声認識部で用いる音声分析条件，音響モデルは，表 4.1 に示すとおりである．

表 4.1: 音声分析条件など

| | |
|-----------|---|
| 音声特徴量 | 38 次元 MFCC12 次元+ ケプストラム+ ケプストラム パワー+ パワー |
| サンプリング周波数 | 16 kHz |
| 分析窓 | Hamming 窓 |
| フレーム長 | 25 ms |
| フレーム周期 | 10 ms |
| 音響モデル | 5 状態 4 出力分布 (4 混合ガウス分布、全共分散行列) 連続出力分布型 HMM, 114 音節 |

認識の途中結果をリアルタイムに出力することも可能であり，今回のシステムにおいても，認識途中結果を用いている．また，単語辞書に予め品詞をつけておくことで，認識途中や最終結果に形態素解析をかけなくとも品詞情報が得られる．ピッチ・パワーは，フレームシフトを 5ms，フレーム幅を 64ms として計算をしており，音声認識・韻律解析のどちらも，用いている音声のサンプリング周波数は，16kHz である．

4.2.2 情報収集部

情報スロットは，ユーザ発話に含まれる重要なキーワードを格納するために用意されている．今回は，天気情報タスクであるので，天気に関するキーワードを格納するようになっている．具体的には，天気の話題であることを確認するために，ユーザが天気に関することを発話した場合には，そのキーワードを格納しておくためのスロットがあり，ユーザ発話中に“天気”や“空模様”といった言葉が含まれている

場合には、その言葉をスロットに格納する。また、“今日、明日、明後日”のように、日時が示された場合にそれらを格納するためのスロットが用意されており、“県名、地域名、都市名”を格納しておくためのスロットや、都市名のスロットが埋まった際に、その都市の天気を格納しておくためのスロットなどが用意されている。応答文を生成する際には、4.2.5 節に示すように、これらのスロットの中身を対話に織り交ぜることが可能である。

4.2.3 対話履歴管理部

対話履歴管理部は、これまでの対話の履歴を全て保存している。保存される内容は、発話者（ユーザ、システム）、発話内容、応答雑談現象番号、発話開始時刻、発話終了時刻である。現在のシステムでは、これらの対話履歴を利用していないが、今後は、履歴を用いることで、対話全体のリズムを捉えて応答を返すことが出来るようにしたい。

4.2.4 特徴量計算部

特徴量計算部は、入力されたピッチ・パワーの傾きを計算している。100ms(決定木判定の間隔)ごとに、100ms 区間について3つの傾きを、500ms 区間について5つの傾きを計算しており、計算には回帰係数を用いている。100ms 区間については、回帰係数のフレーム長は、55ms であり、フレームシフト時のオーバーラップは30ms である。500ms 区間については、回帰係数のフレーム長は、105ms であり、フレームシフト時のオーバーラップはない。

あいづちや、話者交替は、相手の発話のピッチやパワーの変動に誘発されているという知見があり [4,7]、これを利用してタイミング生成器にて応答タイミングを決定するために、ピッチ・パワーの情報を時々刻々と求め、タイミング生成器に送信するようにしている。

4.2.5 応答生成部

応答生成には、ELIZA 方式を用いている。この方法は、ユーザ発話の中のキーワードから、システムがそれに応じた応答をテンプレートを用いて生成するものである。

応答生成は、各応答現象ごとに行われ、それぞれに対して応答文が用意される。現時点では応答現象としてあいづち、復唱、共同補完を扱っている為、それぞれの応答現象ごとに、別の応答を用意する。応答文は各応答現象について、1つずつ保持され、どの応答現象を用いて応答するかは、応答選択部にて決定される。応答文は、

```

@ 今日 (.*) 天気 (.*) (だ|です)
= tenki: .+, tenki_1day: .+, $0 は、$1 ですよ。
= tenki: .+, city: (), 場所はどこですか？
@ 明日 (.*) どう
= tenki: .+, city: .+, tenki_2day: .+, $1 の$0 は、$2 のようですよ。
= tenki: .+, city: (), 場所はどこですか？ : どの$0 が知りたいんですか？

```

図 4.3: 応答テンプレート

ユーザ発話の認識結果を受けて、それに応じて生成している。しかし、認識途中や認識後に、どの応答現象においても応答が準備出来ていないという場合がある。この場合は、応答現象には関係なく返す為の応答を用意しておき、それをを用いることとする。例えば、「心配」というキーワードが発話されていれば「心配事があるんですか？」と応答する。キーワードが入力されていなければ「よく聞こえませんでした。」などと応答するようにする。これによって、発話者が交代するタイミングに全く何も応答がないということが避けられる。あいづちについては、認識結果によらずどのような場合にも対応ができる“はい”や“うん”を準備している。

今回のシステムの場合には、対話中に天気の情報をつなぎ交ぜて、対話を進めていくという戦略をとっている。ユーザとの対話で、今の話題が天気についてであると考えられる場合には、地名などが入力されると、それを記憶しておき、実際の天気予報のデータを対話につなぎ交ぜて生成し出力する。

応答生成に用いる具体的なデータ形式は、図 4.3 のようになっている。これは“@行”にキーワードもしくは正規表現を含んだ文を用意し、直下の“=行”には、“@行”に対応する応答文のセットを用意する。このまとまりをルールセットとする。ルールにない発話が見れた場合には、「何が知りたいのですか？」「意味がわかりません。」といった応答を返すこととした。図 4.3 で、“=行”は、スロット名とスロット値とのマッチング用正規表現、および応答文が記述されている。2 行目では、tenki スロットの値が空ではなく、今日の天気スロット (tenki_1day) も空ではない時に、応答文を返す。応答文の中にある“\$0”は、左側に記述されたスロットの値を参照するためのもので、左側から 0 番, 1 番 …, と名前がついている。下から 4 行目の“=行”では、tenki スロット, 都市名スロット (city), 明日の天気スロット (tenki_2day) が用いられている。認識結果が“@行”とマッチし、さらにスロットの値が“=行”で示されたようにマッチした場合に、その行の応答文を返すようになっている。

ここで、ユーザが「明日はどう？」と発話した場合に応答を返すためには、天気スロットが埋まっている必要がある。天気スロットは、ユーザ発話から“天気”、“

空模様”というキーワードを格納するようになっている。それに加えて、ユーザがこれまでに地名も発話していた場合には、その土地の天気に応答文として出力される。例えば、天気スロットに“空模様”，都市名スロットに“豊橋”が入っており、また豊橋の明日の天気が“晴れ”であった場合には、システム応答は“豊橋の空模様は、晴れですよ。”となる。

最後の“=行”では、2種類の応答文がコロンで区切られて記述されている。複数の応答を用意する場合にはこのように記述しておくことで、この中からランダムに応答を用意することができる。

地名を発話していない場合には、都市名スロットと明日の天気スロットが空であるので、一番下の「場所はどこですか？」という応答文が出力されることになる。

4.2.6 応答選択部

応答選択部では、決定木により、応答文候補の中から応答に用いる応答文と、応答のタイミングを決定する。決定木の素性には、2.3節に示した韻律情報を用いる。100ms毎に各素性を決定木へ入力すると、決定木が何らかの応答を返すタイミングであると判断すれば、どの応答現象を用いて応答するかも決定し、出力されるので、応答生成部はそれに応じた応答を出力する。応答を返すべきでないタイミングにおいては、決定木からは、待ち状態であるとする結果が出力され、応答は出力されない。また、応答は1ユーザ発話につき1応答としている。ただしあいづち、復唱についてはその制限はなく、何度でも打つことが出来る。

4.2.7 天気情報

天気情報は、国際気象海洋株式会社天気案内 WebPage [32] より自動でダウンロードしている。このページには、全国主要都市の今日・明日・明後日の3日分の天気情報と、今日・明日の降水確率、気温が記載されている。このページから、全国の天気情報を取得しプログラム内に読み込んでおく。ユーザが地名を示した場合には、その土地の3日分の天気情報を天気スロットに格納しておくことで、それ以降のシステム応答に天気情報を盛り込むことが可能になる。

4.3 出力部

出力部は、システムが生成した応答を、音声にて出力する。出力音声には、音声合成を用いるものと、録音音声を用いるものの、2種類を用意した。音声合成方式

には、擬人化音声対話エージェントのツールキット Galatea Toolkit [33] に含まれる音声合成器の GalateaTalk を用いている。この音声合成器は、発話者タイプ（男女など）の変更や抑揚、話速を自由に変更が出来るため、本システムの韻律モデルにより、抑揚や話速を変化させることもできる。録音音声方式には、実際に録音した人間の声を用いている。声の高さや話速の変更は、事前にそれらを変更した音声を用意しておき、再生することで、再合成にかかる時間をなくしている。

韻律の制御は、3章で述べた韻律制御モデルを用いて、ユーザに追従するように制御する。

4.4 システムの動作例

実際のシステムとユーザの対話例を図 4.4 と図 4.5 に示す。図においては、上がユーザ発話であり、下がシステム応答である。図 4.5 において、システム応答の波形に、ユーザ発話区間と対応した振幅の小さい波形が見られるが、これは収録環境によるノイズであり、合成音ではない。

現段階のシステムでは、1 ユーザ発話に対しては、1 システム応答としている。ただし、あいづちと復唱に関しては、その制限は外している。対話の進行を、システムの動作と共に解説していく。まず図 4.4 を見る。システムの呼びかけに対してのユーザ発話「こんにちは」という挨拶に対しては、システムはまず挨拶を返し、そして、今日の天気を応答する。これは、ユーザの居場所が予め分かっており、それがスロット値として参照可能になっている事で実現される。また、天気を言うことで、話題づくりをしている。次のユーザ発話「最近雨ばかりだよ」に対しては、“最近”と“雨”というキーワードと今現在天気についての話題であるということから、“多いですよ”という共同補完を、オーバーラップして応答している。これは、ユーザ発話からのスロット値とシステム知識のスロット値を用いて、Eliza テンプレートにて生成している。次のユーザ発話「浜松の天気はどうですか？」に対しては、地名“浜松”を重要語句とし、システムはすぐに復唱している。そして、ユーザの発話終了後には、要求された応答として「雨ばかりですよ」と応答している。

次に、図 4.5 を見る。最初のユーザ発話「えっと明日の豊橋の天気を教えて」では、最初の部分に「えっと」というフィラーが入っている。そして、この発話に対して、システムは「はい」というあいづちを返している。そのあと、システムは、ユーザ発話に対して応答をしようとしているが、ユーザが途中でシステム発話に対して割り込んだ為、システムは応答を停止し、ユーザに発話権を移した。ユーザの割り込み発話である「やっぱり明日の浜松の天気を教えて」に対して、システムはあいづちを打ち、そして、天気の情報も返すことが出来ている。次のユーザ発話は、「明

後日の名古屋はどうか？」という発話であるが、ユーザが言い淀みをした為に、システムは2つの文として認識をした。そして、ユーザ発話の前半部分である「明後日の名古屋は...」に対して、システムは「明後日は、何かあるんですか？」という応答を返そうとするが、ユーザの発話が継続したため、システムは途中で応答を停止している。ユーザ発話後半の「どうですか？」に対しては、これまでの対話により、「天気」「名古屋」「明後日」という情報がスロットに入っているために、“明後日の名古屋の天気はどうか？”という意味であると理解し、正しい応答を返すことが出来ている。このように、ユーザの割り込みや、言い淀みがあっても、頑健に応答することが出来ている。

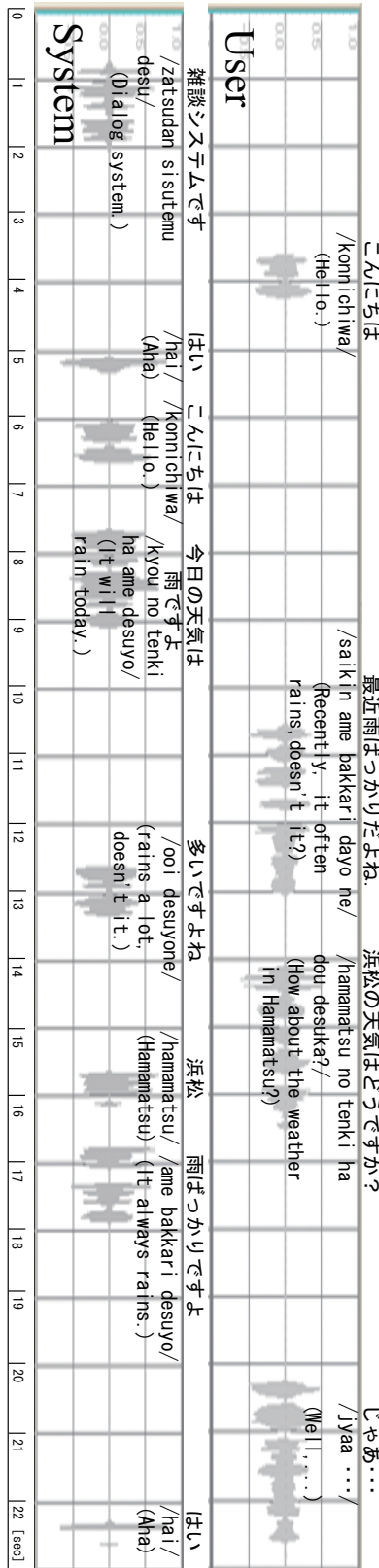


図 4.4: 対話例 1

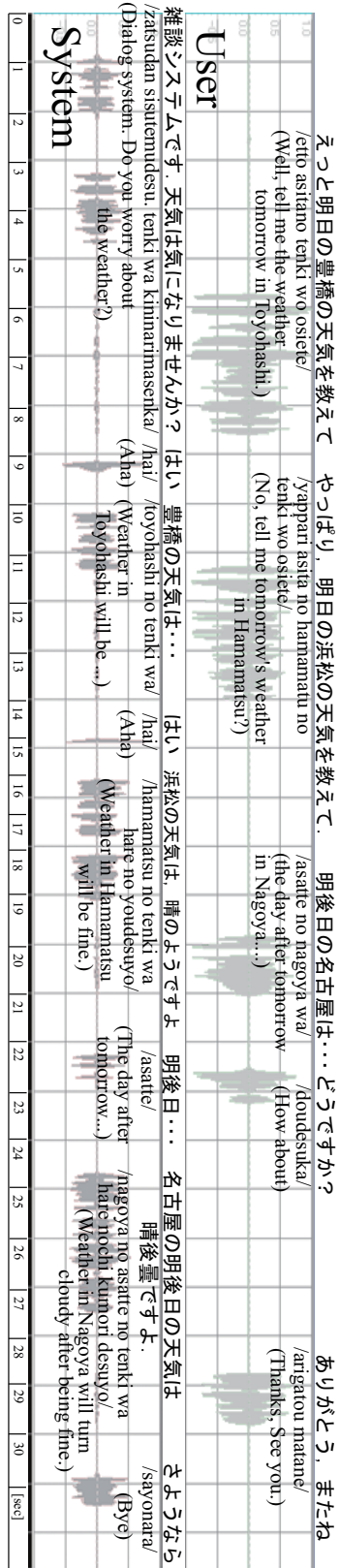


図 4.5: 対話例 2

5 評価

決定木のタイミングの主観評価と、システム全体の主観評価を被験者実験にて行う。対話システムは、いくつかのモジュールを統合して出来ているため、全体を評価する方法は非常に難しく、評価法についても色々研究がなされている [34,35]。また、システムを評価する前の事前知識（システムの説明をどのように行うか）によっても、システムの評価が変わる [36]。対話システムの評価を行う際には、被験者に対する説明方法にも十分に注意が必要である。

5.1 決定木からの応答タイミングの主観評価

5.1.1 評価方法

決定木によって生成されたあいづちや話者交替などの応答タイミングの自然性を調べるため、被験者による主観的な評価を行った。

ここで、評価の対象を‘あいづち’と‘話者交替’とする。今回のシステムでは、種々の応答現象を扱っているが、出現回数が多いあいづちは評価できるものの、その他の現象は、タスクの制限やシステムの学習不足によって出現しにくくなっており、評価が出来ない。

評価をするときには、あいづちや話者交替のタイミングだけを判定し、言語的なつながりや韻律的なつながりは無視してもらうようにする。

評価の為のサンプル音声は、テストコーパス中の発話に対して決定木が生成したあいづちタイミング箇所、同じ話者が別の箇所で発声したあいづち音声を挿入した。話者交替に関しても同様に、決定木が話者交替と判定した箇所に対して、同じ話者のフィラー音声をつけた音声を作成した。この意味の無い音声を挿入することで、話者交替時の言語的なつながりの影響をなくすることが可能である。用意したサンプルの数は、それぞれ 20 個である。

比較対象として、同じコーパス中で実際に発声したあいづち、話者交替を、別音声で置換し、人間のタイミングのみを再現したサンプルも用意した。ここでは、前述の音声を置換した理由に加えて、コーパス中の対話の音声的なつながりの影響をなくすことで、決定木が生成したタイミングと人間のタイミングを公平に被験者に判断させるという狙いもある。用意したサンプルの数は、それぞれ 20 個である。

また、コーパスの音声をそのまま聞かせる実験も行った。これの目的は、タイミングのみを判定してもらうという今回の実験で、タイミング以外の要素（韻律的なつながりや、言語的なつながり）が無意識のうちに判定結果に影響を及ぼしているかどうかを判断するものである。用意したサンプル数は、あいづち、話者交替でそれぞれ

れ 20 個である。

このサンプル音声を被験者 5 人が聴取し，1：早すぎる，2：早い，3：よい，4：遅い，5：遅すぎる，6：論外という 6 つの評価基準で評価を行った．‘論外’はそのあいづちの出現自体が不自然である場合に選択するよう教示した．

5.1.2 評価結果

主観評価結果を表 5.1，表 5.2 に示す．表中の「自然率」は，全体に対する「適切」の数の割合である．

表 5.1: 被験者の主観によるあいづちタイミングの評価

| タイミング | 早い | やや早い | 適切 | やや遅い | 遅い | 論外 | 自然率 |
|-------|----|------|----|------|----|----|------|
| システム | 23 | 38 | 27 | 0 | 0 | 12 | 27.0 |
| コーパス | 4 | 26 | 65 | 3 | 0 | 2 | 65.0 |
| 元音声 | 4 | 15 | 73 | 3 | 0 | 5 | 73.0 |

表 5.2: 被験者の主観による話者交替タイミングの評価

| タイミング | 早い | やや早い | 適切 | やや遅い | 遅い | 論外 | 自然率 |
|-------|----|------|----|------|----|----|------|
| システム | 29 | 31 | 17 | 9 | 6 | 8 | 17.0 |
| コーパス | 5 | 26 | 54 | 4 | 7 | 4 | 54.0 |
| 元音声 | 2 | 15 | 72 | 4 | 6 | 1 | 72.0 |

あいづちのタイミングの評価（表 5.1）において，システムが出力したタイミングに対しては，タイミングが適切であると評価されたものは全体の 27.0%であった．評価の内訳を見てみると，タイミングが「遅い・遅すぎる」という評価はなく，論外と評価されたもの以外は，すべてタイミングが「適切・早い・早すぎる」と評価されている．また，タイミングが「やや早い」と評価されたものがとても多かった．このことについては，決定木が，相手の発話中に「発話待ち」（つまり応答しない）を，もっと長い区間にわたり出力しないといけないうところで，あいづちを返してしまっているということである．これについては，相手の発話中の「発話待ち」を上手く学習できるように工夫をする必要がある．実際の人間のタイミング（コーパスタイミング）では，適切であると評価されたものは，65.0%であった．「遅い・やや遅い」と評価されたものもいくつかあったものの，ほとんどが「適切・やや早い」であった．このことから，やはりシステムタイミングの出力が早すぎることが分

かる．そして，元音声では，コーパスタイミングと同じタイミングであるにも関わらず，適切と評価されたものの割合が73.0%に上がっている．これは，やはりタイミングのみを評価するようにしても，人間は無意識のうちに，韻律的なつながりや言語的つながりなどの，何らかの要因に影響されていると考えられる．

話者交替のタイミングの評価（表 5.2）において，システムが出力したタイミングに対しては，タイミングが適切であると評価されたものは全体の17.0%であった．評価の内訳を見てみると，あいづちのタイミングの評価と同様に「早い・やや早い」と評価されたものがとても多かった．実際の人間のタイミング（コーパスタイミング）では，適切であると評価されたものは，54.0%であった．また，元音声で適切であると評価されたものは，72.0%であった．この傾向は，あいづちに対するものと同様であるが．コーパスタイミングと元音声の評価の差が，話者交替の方が大きくなっている．この原因については，今後調査・分析する必要があるが，考察としては，比較的短いあいづちよりも，比較的長い話者交替後の発話の方が，先行発話に対するつながりが強く，それが評価に影響を与えたのではないかと考えられる．

あいづち，話者交替のどちらも「早い・やや早い」に評価がかたまっていることから，応答のタイミングをもう少し遅くしなければならないと考えられる．これは，システムが「発話待ち」をもう少し長い間出力し続ける必要があるということである．実際の人間の対話コーパスから，上手く発話待ちの学習ができるように工夫をしなければならない．

5.2 対話システムとの対話による評価

5.2.1 評価方法

応答現象とそのタイミングを評価するため，本システムと被験者が実際に対話を行い，被験者が評価する実験を行った．

実験は，オーバーラップ応答がある場合と無い場合，出力音声の質による印象の違い（合成音声，人間の声の録音）の組み合わせで，計4通り行う．

各実験の後，被験者にはアンケートに答えてもらい評価を行う．被験者には，応答内容よりも，応答のタイミングの評価に重点を置くように説明を行い，また，オーバーラップ応答をする場合としない場合の違いや，対話に与える影響を特に注意して評価してもらった．1対話につき，10発話程度のやりとりを行った．

被験者は5人で，全員大学院修士課程の学生である．また，性別は全員男性である．対話システムの経験の有無については，ばらばらである（経験あり2人，経験なし3人）．

すべての実験が終わった後で、図 5.1 に示す内容のアンケートに答えてもらった。

5.2.2 評価結果

結果を表 5.3 に示す。

表を見ると、話しやすさについては、オーバーラップが無いほうがよく、また録音音声の方が良いという結果になっている。「オーバーラップなし」の方が良いと答えた人の理由としては、「間が悪い」「喋っている途中で割り込まれた感が強い。」というふうに、タイミングが良くないことから来る問題が多かった。また、肯定的な意見としては、「ちゃんと入力されていることが分かるのは良かった。」「正しく入力されている気になった。」「リズムが良くなった。」等があった。これらのことから、オーバーラップは、入力がされているかどうか分かりやすくなる反面、タイミングが悪いと、悪い印象を与えてしまうことが分かる。

出力音声の違いに対する意見としては、録音音声に対しては「人間の声の方が抵抗が少ない」「人間の声の方が会話をしている気分になれた」「人間音声の方が親しみを感じる」という意見が挙がっていた。合成音声に対しては「機械音声はぎこちなく、発言を理解するのに時間がかかる」という意見が挙がっていた。やはり、合成音声よりも、人間の声の録音音声の方が、普段聞きなれている声であり、自然性もあるので、そこで生じる違和感も無くなり、被験者からの印象が良かった。

また、システムとの親しみやすさについては、オーバーラップがあるものの方が、若干良くなっている。しかし、大きく改善されたということでもなく、オーバーラップのありなしに関係なく低い評価になっている。理由としては「人間サイドの意図を汲み取ることができていない」「返答がおかしなところがある」「正しく応答してくれない」などの意見が挙がった。これらは、応答内容に関する部分が大きく、親しみやすさについては、応答内容や、応答の正確さが関連していると考えられる。

応答回数については、オーバーラップの回数、それ以外の応答回数の両方とも「適切」あるいは「やや少ない」という評価であった。応答の回数については、あいつち、復唱以外のものについては、1回のユーザ入力に対して1回の応答に限定している。今回の結果から、応答の回数をもう少し増やす必要がある。

応答のタイミングについては、オーバーラップについては「普通」という評価が多く、それ以外の応答タイミングについては「やや悪い」という評価が多かった。オーバーラップにおいては「普通」という評価が付いているにも関わらず、上述した「あいつちのある・なし」に関わるアンケートにおいては、「割り込まれている感じが強い」という否定的な意見が出ていた。これは、オーバーラップのタイミング自体は普通であるが、オーバーラップの仕方や頻度に問題があるのではないかと考えられる。

- 対話システムを使った感想（自由筆記）
- どちらのタイプが話しやすかったか（オーバーラップあり，オーバーラップなし）
- どちらのタイプが話しやすかったか（人間の声，合成音声）
- 人間の応答音声と合成音声のどちらがよいと思うか（人間の声，合成音声，どちらでもよい）
- 応答音声の違いによってどのように感じたか（自由筆記）
- システムとの対話は親しみやすいものであったか？（全く，あまり，どちらともいえない，少し，かなり）
- （親しみやすさについて）なぜそのように感じたのか（自由筆記）
 - オーバーラップがある場合のみ -
 - オーバーラップの回数（少なすぎる，やや少ない，適量，やや多い，多すぎる）
 - オーバーラップのタイミング（悪い，やや悪い，普通，やや良い，良い）
 - オーバーラップがあることによって，発話がちゃんと入力されている気はしたか（全くしなかった，あまりしなかった，どちらともいえない，結構した，かなりした）
 - オーバーラップがあることによって，親しみやすさはどうなったか（全く，あまり，どちらともいえない，少し，かなり）
 - オーバーラップがある対話にどんな印象をもったか（自由筆記）
- システムのオーバーラップ以外の返答のタイミングはどうだったか（悪い，やや悪い，普通，やや良い，良い）
- 対話全体でシステムの返答回数はどうだったか（多すぎる，やや多い，適量，やや少ない，少なすぎる）
- システムの返答内容はどうだったか（不自然，やや不自然，どちらともいえない，やや自然，自然）
- （不自然と回答したときのみ）なぜ不自然と感じたか（同じことを言うことが多い，タイミングが早いもしくは遅い，言ったことに答えてくれない，その他）
- どうすれば自然になると思うか（もっといろんなことを言えるようにする，もっとタイミングよく返答するようにする，言ったことに適切に答えられるようにする）

図 5.1: 被験者実験でのアンケート内容

オーバーラップがあると、ちゃんと入力されているように感じるかどうかという質問に対しては、意見はばらばらであった。かなり良いと評価した人もいれば、あまり良くないと評価した人もいる。被験者が言ったことが、きちんと入力されているかどうかを発話中に確認することができたような、比較的上手く対話中にオーバーラップ応答が働いた場合には、オーバーラップ応答は好印象だったようである。逆に、確認が必要でない場合などに復唱などがあると、かえって応答が邪魔に感じてしまう。必要な箇所をしっかりと見極めて応答することが出来れば、効果があるということである。また、オーバーラップ音声の音量が大きすぎて不自然に感じるので、応答内容に応じて、音量を調節（オーバーラップは音量を小さくするなど）して対応するようにしたい。

システムの応答内容については、全員が「やや不自然・不自然」を選択し、自然であると感じた人はいなかった。これは、応答生成や応答選択などの対話戦略の部分が弱いことを示している。また、なぜ不自然と感じたかという問いに対しては、「同じことを言う」「言ったことに答えない」という意見が多かった。「同じことを言う」ということに関しては、応答のバリエーションを増やす必要がある。しかし、ただ増やせばよいと言うものではなく、対話戦略も考えて、各状況に応じた応答を返すことが出来るようにしなければならない。また、「言ったことに答えない」ということに関しては、音声認識の精度も関連している。ある地域の天気を尋ねたい場合に、目的の地名を認識することが出来なければ、システムはその地域の天気を返答することができない。これらの問題に対して、「どこを改善すれば自然になると思うか」という質問に対しては、応答のバリエーションを増やすことよりも、言ったことに答えられるようにするということが重要であるという意見が多かった。ユーザの発話に対して、正しく応答ができるシステムでなければ、対話も進まず、自然に感じることはできない。ユーザが何を求めているか、そしてそれに対して正しく答えることが重要である。

表 5.3: 被験者の評価

| 評価項目 | 評価結果 | | | |
|-------------------------------------|--------------|---------------|----------------|-----------------------|
| | あり 1 | あまり 2 | どちらともいえない 3 | なし 4 |
| オーバーラップあり・なし どちらが話しやすかったか | 5 | 0 | 0 | 4 |
| 録音音声・合成音声 どちらが話しやすかったか | 5 | 0 | 0 | 0 |
| システムとの親しみやすさ オーバーラップあり | 0 | 2 | 2 | 1 |
| システムとの親しみやすさ オーバーラップなし | 0 | 4 | 0 | 1 |
| オーバーラップの回数 | 多すぎる 0 | やや多い 1 | 適量 2 | やや少ない 2 |
| システム全体での 応答回数 | 多すぎる 0 | やや多い 0 | 適量 3 | やや少ない 2 |
| オーバーラップの タイミング | 悪い 1 | やや悪い 0 | 普通 3 | やや良い 1 |
| オーバーラップ以外の タイミング | 悪い 0 | やや悪い 3 | 普通 1 | やや良い 0 |
| オーバーラップがあると ちゃんと入力されていると 感じるか | 全く 0 | あまり 2 | どちらともいえない 1 | けっこう 1 |
| システムの 応答内容 | 不自然 2 | やや不自然 3 | どちらともいえない 0 | やや自然 0 |
| なぜ不自然と感じたか (やや・不自然と回答した場合) | 同じことを言う 3 | タイミングが早い 0 | タイミングが遅い 1 | 言ったことに答えない 5 |
| どうすれば自然になると思うか | 1 | 1 | 1 | 5 |
| | | | | かなり 1 |
| | | | | 自然 0 |
| | | | | その他 0 |
| | | | | 言ったことに答えられるようにする 5 |

6 結論

本研究では、人間同士の対話に現れる種々の応答現象の応答種類決定、応答タイミングの検出を、リアルタイムにすることができる対話システムの構築を行った。

1章では、本研究の背景を述べた。その中で人間同士の対話では、話者は互いにあいづちやうなずきで、相手の発話内容の理解を明示し、また、そのタイミングが重要であることを示した。そして、応答のタイミングだけではなく、応答するときの韻律情報も重要であることを示した。

2章では、韻律情報を素性に用いた決定木を構築し、あいづち / 話者交替 / 話者継続 / 復唱 / 共同補完のいずれかに分類するルールの生成を行った。この決定木は、素性として、音声の直前 100ms のピッチ・パワーの傾き、直前 500ms のピッチ・パワーの傾き、ユーザの発話長、ポーズ長、以前のシステム出力からの時間を使用する。決定木は、これらの素性を 100ms ごとに逐次的に解析し、どの応答現象を用いて応答すべきかを出力する。この決定木を人間同士の対話での応答現象が起こった箇所と比較し、再現率、適合率を用いて評価したところ、オープンテストデータで、あいづちの分類において再現率 39.0 %、適合率 29.5 %、話者交替の分類において再現率 37.7 %、適合率 25.6 % の結果を得た。これらの出現現象は、確率的なゆれがある点を考慮すれば比較的上手く動作していると言える。しかし、コーパスでの出現回数が少なく、十分に学習がなされていない復唱、共同補完については、復唱の分類において再現率 6.2 %、適合率 3.7 %、共同補完の分類において再現率 2.4 %、適合率 0.6 % であった。

3章では、協調的な音声対話システムを実現するために、人間同士の対話における韻律的な同調と対話としての盛り上がりとの関連を分析し、そのモデル化を試みた。また、そのモデルを音声対話システムに実装した。盛り上がっている対話においては、基本周波数 (F0) やダイナミックレンジも同調しており、詳しく話者間の基本周波数の相関を調べると、強い相関がある対話とそうでない対話があった。それらについて、人間が実際に音声を聞いた場合の印象を付けてもらい、その印象との相関を調べた。すると、盛り上がっている対話や、親しみを持って対話をしているものや、同意して対話を進めているものでは、基本周波数の相関が高い。これらのことから、基本周波数を相手に追従して変化させるモデルを提案し、モデル化を行って、システムに実装した。

4章では、構築した決定木のルールを応答選択 (タイミング生成) 部とした対話システムの詳細について述べた。この対話システムは、ユーザとシステムの混合主導で対話を進めることができる。また、話題としては「天気に関する話題」を採用している。システムは、逐次的に応答すべきタイミングであるかを見ており、それ

ゆえに、オーバーラップした応答も返すことが可能になっている。また、韻律制御部にて、応答音声の韻律をユーザに同調させていくことにより、ユーザに対して同調性を示して、対話がより円滑に進むようにしている。

5章では、被験者実験による、システムの評価を行った。決定木からの応答タイミングの主観評価では、システムの出カタイミングは、実際の人間のタイミングよりも早いという結果が出た。システムのタイミングで「適切」であると評価された割合は27.0%であった。人間の対民部で「適切」であると評価された割合は65.0%であった。システムは、実際には応答を返してはいけない段階（発話待ち）のところで、応答を返してしまうため、実際の応答よりも早い段階で応答してしまい、その為、応答のタイミングが早いと評価された。これに対しては、ユーザの発話中の待ち状態の学習方法や、素性、ルールなどを検討しなければならない。対話システムとの対話による評価では、出力音声に対しては録音音声よりも自然性があると評価され、また、オーバーラップ応答については、無いほうが良いと評価された。オーバーラップ発話が無いほうが良いと評価された理由としては「間が悪い」「割り込まれたという感覚が強い」という意見があった。しかし、「正しく入力されている気になった」「リズムが良くなった」という意見もあり、オーバーラップ応答自体が否定されているということではなかった。このことから、オーバーラップ応答は、適切なタイミングで返すことが出来れば、対話に良い効果を与えるが、タイミングが悪ければ、かえって悪い印象を与えてしまうということである。

今後の課題としては、次のようなものがある。まず、はじめに“雑談対話制御”がある。現在のシステムは、タスクや語彙も小さく、雑談ができるというものではないが、より広く雑談ができるような対話制御を行う。次に、システムがユーザの言ったことに対して“知らないふり”をするようにしたい。これによって、システム側から対話を制御し、より発展した対話を実現することが可能になると思われる。例えば、ユーザから「豊橋-東京間ってどれくらいの距離かなあ？」と言われ、システムがその情報を持っていたとしてもすぐには提示せず、「どれくらいですかね？(システム発話)」と知らないふりをし、ユーザが「200km ぐらいかなあ～」と言ったら、システムから「もっと遠いと思うよ。」(実際は300km程度)と対話を続けるというふうにすることで、対話をより楽しめるようにするものである。次に、現在はユーザ対システムの1対1対話であるが、これを、“性格の異なる2つのエージェント(システム)との3人対話”に拡張する。エージェント間では、実際に発話した内容以外にも、すべての情報が共有できるので、様々な応用が考えられる。最後に、“視線追跡による対話焦点の分析”をしたい。これは、視線が対話の流れにどのような影響を及ぼしているかということ进行分析し、それを活用したいと考えている。

これらのことを調査・分析し、システムをより使いやすく、また自然な対話がで

きるようにしていきたい。

7 謝辞

本研究を進めるにあたり，熱心に御指導いただいた豊橋技術科学大学情報工学系の中川 聖一教授，名古屋大学の北岡 教英助教授に深く感謝致します。

対話システムの構築の際に，多くの助言をいただいた静岡大学工学部システム工学科の甲斐 充彦助教授に深く感謝致します。

そして，本研究にて韻律解析器として使用しているプログラムには，産業技術総合研究所の後藤真孝先生が提案されたアルゴリズムを用いて，早稲田大学の藤江真也先生が実装されたものを利用させて頂いております。ここでお礼を申し上げます。

また，すばらしい研究の環境を提供してくれた中川研究室の皆様，被験者実験に御協力いただきました皆様に心より感謝の意を表します。

参考文献

- [1] 小暮悟, 中川聖一: 音声対話システムにおける頑健な意味理解と対話システムの移植性に関する研究, 豊橋技術科学大学大学院 博士学位論文 (2002.2).
- [2] 西崎博光, 中川聖一: 音声文書を対象とした音声入力型情報検索システムに関する研究, 豊橋技術科学大学大学院 博士学位論文 (2003.2).
- [3] メイナード. 泉子.K.: 会話分析, くろしお出版 (1993).
- [4] 小磯花絵, 堀内靖雄, 土屋俊, 市川薫: 先行発話断片の終端部分に存在する次発話者に関する言語的・韻律的要素について, 電子情報通信学会技術報告, NLC95-72, pp. 25-30 (1996).
- [5] 野口広彰, 片桐恭弘, 伝康晴: 心理実験を用いたあいづち応答の手がかり特徴の検証, 人工知能学会研究会資料, SIG-SLUD-A002-13 (2000).
- [6] Sacks, H., Schegloff, E. and Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation, *Language*, 50, pp. 696-735 (1974).
- [7] 大須賀智子, 堀内靖雄, 西田昌史, 市川薫: 音声対話での話者交替/継続の予測における韻律情報の有効性, 人工知能学会誌, Vol. 21, No. 1, pp. 1-8 (2006).
- [8] 平沢純一, 川端豪: 音声対話システム Noddy - ユーザ発話途中でのうなずき・相槌生成 -, 情報処理学会, 音声言語研究会報告, SLP-20-4, pp. 51-52 (1998).
- [9] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一: 韻律情報を用いた相槌の挿入, 情報処理学会論文誌, Vol. 40, No. 3, pp. 469-477 (1999).
- [10] 佐藤康将, 井上貴雄, 目良和也, 相沢輝昭: 自然言語対話システムのための多様なあいづち生成手法の改良, 言語処理学会第8年次大会発表論文集, pp. 248-251 (2002).
- [11] Ward, N. and Tsukahara, W.: Prosodic features which cue back-channel responses in English and Japanese, *Journal of Pragmatics*, 32, pp. 1177-1207 (2000).
- [12] C.Kamm., S.Narayanan., D.Dutton. and R.Ritenour: Evaluating spoken dialogue systems for telecommunication services, Eurospeech-97, Rhodes, Greece, pp. 2203-2206 (1997).

- [13] Kitaoka, N., Takeuchi, M., R., N. and S., N.: Response Timing Detection Using Prosodic and Linguistic Information for Human-friendly Spoken Dialog Systems, *人工知能学会論文誌*, Vol. 20, No. 3, pp. 220–228 (2005).
- [14] 垣田邦子: 簡単な“質問-答”形式の対話における F0 の話者間相互作用, *日本音響学会講演論文集*, No. 2-P-2, pp. 305–306 (1995).
- [15] 長岡千賀, 小森政嗣, Maria, D. R., 中村敏枝: 2 者対話における好意の表出 ~ 交替潜時を分析指標として ~, *日本心理学会第 65 回大会発表論文集*, No. 0341 (2001).
- [16] 石崎雅人, 伝康晴: 言語と計算- 3 談話と対話 (2001).
- [17] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs, *Language and Speech*, Vol. 41, No. 3-4, pp. 291–317 (1998).
- [18] J. Quinlan, R.: C4.5: Programs for machine learning, *Morgan Kaufmann* (1992).
- [19] Breiman, L., Friedman, j. H., Olshen, R. A. and Stone, C. J.: Classification and Regression Trees, *Wadsworth, Belmont* (1984).
- [20] 田中, 速水, 山下, 鹿野, 板橋, 岡: RWC 計画における音声対話データベースの構築, *情報処理学会音声言語情報処理 11 - 7* (1996).
- [21] Boersma, P. and Weenink, D.: Praat: doing phonetics by computer (Version 4.3.14) [Computer program], *from <http://www.praat.org/>* (2005).
- [22] K., M., H., K., S., F. and H., I.: Spontaneous speech corpus of Japanese, *Proceedings of the Second International Conference of Language Resource and Evaluation*, Vol. 2, pp. 947–952 (2000).
- [23] 堂坂浩二, 安田宜仁, 宮崎昇, 中野幹生, 相川清明: システム知識制限下における効率の対話制御, *情報処理学会, 音声言語研究会報告*, 2000-SLP-33, pp. 49–54 (2000).
- [24] Zibert, J., Sanda Martincic-Ipsic, M. H., Ipsic, I. and Mihelic, F.: Development of a bilingual spoken dialog system for weather information retrieval, *Eurospeech-2003, Geneva, Switzerland*, pp. 1917–1920 (2003).

- [25] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T. J. and Hetherington, L.: JUPITER: A Telephone-Based Conversational Interface for Weather Information, *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 1, pp. 100–112 (2000).
- [26] 岡田美智男, 鈴木紀子, 石井和夫, Altman, E.: 雑談の構成的な理解に向けて, 情報処理学会, 音声言語研究会報告, 97-SLP-17, pp. 39–44 (1997).
- [27] Weizenbaum, J.: ELIZA - A computer program for the study of natural language communication between man and machine, *communications of the ACM*, Vol. 9, No. 1, pp. 36–45 (1965).
- [28] 後藤真孝, 伊藤克亘, 速水悟: 自然発話中の有声休止箇所のリアルタイム検出システム, 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No. 11, pp. 2330–2340 (2000).
- [29] 藤江真也, 福島健太, 柴田大輔, 小林哲則: FST と韻律情報を用いた相槌・復唱機能を有する対話ロボット, 人工知能学会研究会資料 SIG-SLUD-A401-03 (2004).
- [30] 甲斐充彦, 中川聖一: 日本語連続音声認識システム SPOJUS-SYNO の改良と評価, 電子情報通信学会技術報告, SP93-20 (1993).
- [31] 豊橋技術科学大学情報工学系中川研究室: 日本語連続音声認識システム SPOJUS-SYNO, <http://www.slp.ics.tut.ac.jp/SPOJUS/>.
- [32] 国際気象海洋株式会社: IMOC Weather, <http://www.imocwx.com/>.
- [33] 嵯峨山茂樹, 川本真一, 下平博, 新田恒雄, 西本卓也, 中村哲, 伊藤克亘, 森島繁生, 四倉達夫, 甲斐充彦, 李晃伸, 山下洋一, 小林隆夫, 徳田恵一, 広瀬啓吉, 峯松信明, 山田篤, 伝康晴, 宇津呂武仁: 擬人化音声対話エージェントツールキット Galatea, 情報処理学会研究報告 (2002-SLP-45-10) (2003).
- [34] R.López-Cózar, la Torre, A., J.C.Segura, A.J.Rubio and V.Sánchez: Assesment of dialog systems by means of a new simulation technique, *Speech Communication*, Vol. 40(3), pp. 387–407 (2003).
- [35] M.A.Walker, D.Litman, C, A. and A.Abella: PARADISE: A general framework for evaluating spoken dialogue agents, *Proceedings of the ACL*, pp. 271–280 (1997).

-
- [36] Jokinen, K. and Hurtig, T.: User Expectations and Real Experience on a Multimodal Interactive System, *Proceedings of ICSLP-2006*, pp. 1049–1052 (2006).

A 付録

A.1 発表業績

[学部]

2004年3月4日

電子情報通信学会東海支部 卒業研究発表会

西村良太「音声対話システムのための韻律情報を用いた自然な発話タイミングの生成」

[修士]

2006年3月3日

人工知能学会 言語・音声理解と対話処理研究会（第46回）

西村良太，北岡教英，中川聖一："応答タイミングを考慮した雑談音声対話システム"，人工知能学会研究会資料，SIG-SLUD-A503-05，2006．

2006年11月16日～17日

人工知能学会 言語・音声理解と対話処理研究会（第48回）

西村良太，北岡教英，中川聖一："対話における韻律変化・タイミングのモデル化と音声対話システムへの適用"，人工知能学会研究会資料，SIG-SLUD-A602-07，2006．

2007年3月13日～15日（予定）

日本音響学会 2007年春季研究発表会

西村良太，北岡教英，中川聖一："音声対話システムにおける対話中の韻律変化のモデル化と適用"

講演番号：1-9-3

A.2 音声認識部に用いた文法・単語辞書

A.2.1 音声認識部に用いた文法

SSSS NOINP FILLER SP SENT NOINP

NOINP

NOINP noi

SP

SP noc
SENT
SENT TOSHI1 TOSHI2
SENT DAY TIME TOSHI1 DOSHI1
SENT TOSHI1 DAY TIME DOSHI1
SENT DAY TIME TOSHI2 DOSHI1
SENT TOSHI2 DAY TIME DOSHI1
SENT TOSHI1 TOSHI2 DAY TIME DOSHI1
SENT DAY TOSHI1 DOSHI2
SENT NOUN DOSHI3
SENT DAY NOUN DOSHI3
SENT JANAKU
SENT JANAKU no DOSHI1
JANAKU TOSHI1 ja_nakute
JANAKU TOSHI1 ja_nakute TOSHI1
JANAKU TOSHI2 ja_nakute TOSHI2
SENT DAY ADJE
SENT DAY ADJE desuka
SENT DAY TOSHI1 ADJE
SENT DAY TOSHI1 ADJE desuka
SENT DAY TOSHI2 ADJE
SENT DAY TOSHI2 ADJE desuka
SENT DAY TOSHI1 DOSHI5
SENT DAY TOSHI2 DOSHI5
SENT ADJE DOSHI6
SENT saikin no tenki ni_tuite nan da kedo
SENT saikin AMARI tenki ga yokunai DESUNE
SENT saikin k_tenki bakari DESU YONE
SENT city qte k_tenki bakari DESUNE
SENT DAY SORA ha DOUKANA
SENT DOUKANA
SENT tenki desuka
SENT AMARI YARANAI YO
SENT KEKKOU YARU YO
KEKKOU

KEKKOU kekkou
YARANAI yattenai
YARANAI sitenai
YARU yaru
YARU siteru
SENT AMARI mottenai YO
SENT KEKKOU motteru YO
YO
YO yo
YO kedo
SORA
SORA no tenki
DOUKANA dou
DOUKANA dou desuka
DOUKANA dou kana
AMARI
AMARI amari
AMARI anmari
DESUNE DESU
DESUNE DESU YONE
DESUNE YONE
DESU desu
DESU da
YONE yo ne
YONE yo
YONE ne
FILLER
FILLER filler
FILLER filler filler
FILLER DOSHI4
ADJE atsui
ADJE samui
ADJE suzushi
ADJE attakai
DAY

DAY day
DAY day no
DAY day ha
NOUN what
NOUN what wo_ga
NOUN what ni_tuite
NOUN tenki ni_tuite
TIME
TIME time
TIME time no
TIME time goro no
TOSHI1
TOSHI1 PREF
TOSHI1 PREF no
TOSHI1 PREF ha
TOSHI1 tokyo k_to
TOSHI1 tokyo k_to no
TOSHI1 tokyo k_to ha
TOSHI1 tokyo k_to ni
TOSHI1 osaka k_fu
TOSHI1 kyoto k_fu no
TOSHI1 kyoto k_fu ha
TOSHI1 kyoto k_fu ni
TOSHI1 pref_ken k_ken
TOSHI1 pref_ken k_ken no
TOSHI1 pref_ken k_ken ha
TOSHI1 pref_ken k_ken ni
TOSHI2
TOSHI2 city
TOSHI2 city ha
TOSHI2 city no
TOSHI2 city c_shi no
TOSHI2 city c_shi ha
TOSHI2 city c_shi ni
PREF pref_ken

PREF tokyo

PREF osaka

PREF kyoto

PREF hokkaido

PREF pref

DOSHI1

DOSHI1 tenki ha

DOSHI1 tenki wo oshiete

DOSHI1 tenki wo oshiete kudasai

DOSHI1 tenki ha k_tenki

DOSHI1 tenki ha dou

DOSHI1 tenki ha dou desuka

DOSHI1 tenki ha k_tenki desuka

DOSHI1 tenki ha k_tenki DESU YONE

DOSHI1 k_tenki DESU YONE

DOSHI1 tenki wo_ga shiritai

DOSHI1 tenki wo_ga shiritai desu

DOSHI1 tenki wo onegai

DOSHI1 kousui wo oshiete

DOSHI1 kousui ha

DOSHI1 kousui wo_ga shiritai

DOSHI2 kasa ha hitsuyo

DOSHI2 kasa ga hitsuyo

DOSHI2 kasa ha hitsuyo desuka

DOSHI2 kasa ga hitsuyo desuka

DOSHI3 kikeru

DOSHI3 kikeru no desuka

DOSHI3 dekiru

DOSHI3 dekiru no desuka

DOSHI3 hanaseru

DOSHI3 hanaseru no desuka

DOSHI4 meirei_verb

DOSHI5 iku

DOSHI5 ikutsumori

DOSHI6 no ha sukidesu

DOSHI6 no ha kiraidesu

A.2.2 音声認識に用いた単語辞書

noi . [inte:1][noi] # NOI
noc . [inte:1][noc] # NOC
#unknown * # %
filler えっと [inte:3][filler] # e Q to
filler えーと [inte:3][filler] # e e to
filler えーっと [inte:4][filler] # e e Q to
filler ええ [inte:2][filler] # e e
filler はい [inte:2][filler] # ha i
filler うん [inte:2][filler] # u N
filler そうか [inte:3][filler] # so o ka
filler そうです [inte:4][filler] # so o de su
filler そうそう [inte:4][filler] # so o so o
filler そうですね [inte:5][filler] # so o de su ne
filler そうですか [inte:5][filler] # so o de su ka
filler こんにちは [inte:5][filler] # ko N ni ti wa
filler こんばんは [inte:5][filler] # ko N ba N wa
filler おい [inte:2][filler] # o i
filler どうもありがとう [inte:8][filler] # do o mo a ri ga to o
filler ありがとう [inte:5][filler] # a ri ga to o
filler さようなら [inte:5][filler] # sa yo o na ra
filler おはよう [inte:4][filler] # o ha yo o
filler またね [inte:3][filler] # ma ta ne
filler じゃあ [inte:2][filler] # zya a
filler やっぱり [inte:4][filler] # ya Q pa ri
filler もういいよ [inte:5][filler] # mo o i i yo
meirei_verb 違う [verb:3][meirei_verb] # ti ga u
meirei_verb 違うって [verb:5][meirei_verb] # ti ga u Q te
meirei_verb 違います [verb:5][meirei_verb] # ti ga i ma su
meirei_verb 聞いている [verb:4][meirei_verb] # ki i te ru
iku 行く [verb:2][iku] # i ku
ikutsumori 行くつもり [verb:5][ikutsumori] # i ku tu mo ri

what 何 [noun:2][what] # na ni
goro 頃 [noun:2][goro] # go ro
goro 頃 [noun:2][goro] # ko ro
day 今日 [noun:2][day] # kyo o
day 明日 [noun:3][day] # a si ta
day 明日 [noun:2][day] # a su
day 明後日 [noun:4][day] # a sa Q te
time 午前 [noun:3][time] # go ze N
time 午後 [noun:2][time] # go go
time 1時 [noun:3][time] # i ti zi
time 2時 [noun:2][time] # ni zi
time 3時 [noun:3][time] # sa N zi
time 4時 [noun:2][time] # yo zi
time 5時 [noun:2][time] # go zi
time 6時 [noun:3][time] # ro ku zi
time 7時 [noun:3][time] # si ti zi
time 8時 [noun:3][time] # ha ti zi
time 9時 [noun:2][time] # ku zi
time 10時 [noun:3][time] # zyu u zi
time 11時 [noun:5][time] # zyu u i ti zi
time 12時 [noun:4][time] # zyu u ni zi
saikin 最近 [noun:4][saikin] # sa i ki N
kekko 結構 [noun:4][kekko] # ke Q ko o
siteru してる [noun:3][siteru] # si te ru
mottenai 持ってない [noun:5][mottenai] # mo Q te na i
yaru やる [noun:2][yaru] # ya ru
sitenai してない [noun:4][sitenai] # si te na i
yattenai やってない [noun:5][yattenai] # ya Q te na i
motteru 持ってる [noun:4][motteru] # mo Q te ru
qte って [noun:2][qte] # Q te
no の [part:1][no] # no
ni に [part:1][ni] # ni
ni_tuite について [part:4][ni_tuite] # ni tu i te
wo を [part:1][wo] # o
ga が [part:1][ga] # ga

wo_ga が [part:1][wo_ga] # ga
wo_ga を [part:1][wo_ga] # o
ha は [part:1][ha] # ha
tenki 天気 [noun:3][tenki] # te N ki
tenki 空模様 [noun:3][tenki] # so ra mo yo o
kiraidesu きれいです [verb:5][kiraidesu] # ki ra i de su
sukidesu すきです [verb:4][sukidesu] # su ki de su
oshiete 教えて [verb:4][oshiete] # o si e te
onagai お願いします [:7][onagai] # o ne ga i si ma su
kikeru 聞ける [verb:3][kikeru] # ki ke ru
dekiru できる [verb:3][dekiru] # de ki ru
hanaseru 話せる [verb:4][hanaseru] # ha na se ru
k_tenki 晴れ [noun:2][k_tenki] # ha re
k_tenki 曇り [noun:3][k_tenki] # ku mo ri
k_tenki 雨 [noun:2][k_tenki] # a me
k_tenki 雪 [noun:2][k_tenki] # yu ki
k_tenki みぞれ [noun:3][k_tenki] # mi zo re
bakari ばかり [noun:4][bakari] # ba Q ka ri
bakari ばかり [noun:3][bakari] # ba ka ri
shiritai 知りたい [:4][shiritai] # si ri ta i
desuka ですか [verb:3][desuka] # de su ka
kana かな [verb:2][kana] # ka na
dou どう [verb:2][dou] # do o
ja_nakute じゃなくて [conj:4][ja_nakute] # zya na ku te
ja_nakute ではなくて [conj:5][ja_nakute] # de wa na ku te
kousui 降水確率 [noun:8][kousui] # ko o su i ka ku ri tu
kasa 傘 [noun:2][kasa] # ka sa
hitsuyo 必要 [:4][hitsuyo] # hi tu yo o
desu です [verb:2][desu] # de su
da だ [verb:1][da] # da
kedo けど [verb:2][kedo] # ke do
nan なん [:2][nan] # na N
yo よ [verb:1][yo] # yo
ne ね [verb:1][ne] # ne
kudasai ください [:4][kudasai] # ku da sa i

amari あまり [:3][amari] # a ma ri
anmari あんまり [:4][anmari] # a N ma ri
yokunai 良くない [verb:4][yokunai] # yo ku na i
k_ken 県 [noun:2][k_ken] # ke N
k_fu 府 [noun:1][k_fu] # hu
k_to 都 [noun:1][k_to] # to
c_shi 市 [noun:1][c_shi] # si
atsui 暑い [adje:3][atsui] # a tu i
samui 寒い [adje:3][samui] # sa mu i
attakai 暖かい [adje:5][attakai] # a ta ta ka i
attakai 暖かい [adje:5][attakai] # a Q ta ka i
suzushi 涼しい [adje:4][suzushi] # su zu si i
tokyo 東京 [noun:4][tokyo] # to o kyo o
osaka 大阪 [noun:4][osaka] # o o sa ka
kyoto 京都 [noun:3][kyoto] # kyo o to
hokkaido 北海道 [noun:6][hokkaido] # ho Q ka i do o
pref_ken 青森 [noun:4][pref_ken] # a o mo ri
pref_ken 秋田 [noun:3][pref_ken] # a ki ta
pref_ken 岩手 [noun:3][pref_ken] # i wa te
pref_ken 宮城 [noun:3][pref_ken] # mi ga gi
pref_ken 山形 [noun:4][pref_ken] # ya ma ga ta
pref_ken 福島 [noun:4][pref_ken] # hu ku si ma
pref_ken 茨城 [noun:4][pref_ken] # i ba ra ki
pref_ken 栃木 [noun:3][pref_ken] # to ti gi
pref_ken 群馬 [noun:3][pref_ken] # gu N ma
pref_ken 埼玉 [noun:4][pref_ken] # sa i ta ma
pref_ken 千葉 [noun:2][pref_ken] # ti ba
pref_ken 神奈川 [noun:4][pref_ken] # ka na ga wa
pref_ken 長野 [noun:3][pref_ken] # na ga no
pref_ken 山梨 [noun:4][pref_ken] # ya ma na si
pref_ken 静岡 [noun:4][pref_ken] # si zu o ka
pref_ken 愛知 [noun:3][pref_ken] # a i ti
pref_ken 岐阜 [noun:2][pref_ken] # gi hu
pref_ken 三重 [noun:2][pref_ken] # mi e
pref_ken 新潟 [noun:4][pref_ken] # ni i ga ta

pref_ken 富山 [noun:3] [pref_ken] # to ya ma
pref_ken 石川 [noun:4] [pref_ken] # i si ka wa
pref_ken 福井 [noun:3] [pref_ken] # hu ku i
pref_ken 滋賀 [noun:2] [pref_ken] # si ga
pref_ken 兵庫 [noun:3] [pref_ken] # hyo o go
pref_ken 奈良 [noun:2] [pref_ken] # na ra
pref_ken 和歌山 [noun:4] [pref_ken] # wa ka ya ma
pref_ken 岡山 [noun:4] [pref_ken] # o ka ya ma
pref_ken 広島 [noun:4] [pref_ken] # hi ro si ma
pref_ken 島根 [noun:3] [pref_ken] # si ma ne
pref_ken 鳥取 [noun:3] [pref_ken] # to to ri
pref_ken 徳島 [noun:4] [pref_ken] # to ku si ma
pref_ken 香川 [noun:3] [pref_ken] # ka ga wa
pref_ken 愛媛 [noun:3] [pref_ken] # e hi me
pref_ken 高知 [noun:3] [pref_ken] # ko o ti
pref_ken 山口 [noun:4] [pref_ken] # ya ma gu ti
pref_ken 福岡 [noun:4] [pref_ken] # hu ku o ka
pref_ken 大分 [noun:4] [pref_ken] # o o i ta
pref_ken 長崎 [noun:4] [pref_ken] # na ga sa ki
pref_ken 佐賀 [noun:2] [pref_ken] # sa ga
pref_ken 熊本 [noun:4] [pref_ken] # ku ma mo to
pref_ken 宮崎 [noun:4] [pref_ken] # mi ya za ki
pref_ken 鹿児島 [noun:4] [pref_ken] # ka go si ma
pref_ken 沖縄 [noun:4] [pref_ken] # o ki na wa
pref 久米島 [noun:4] [pref] # ku me zi ma
pref 大東島地方 [noun:9] [pref] # da i to o zi ma ti ho o
pref 宮古島地方 [noun:8] [pref] # mi ya ko zi ma ti ho o
pref 石垣島地方 [noun:9] [pref] # i si ga ki zi ma ti ho o
pref 与那国島地方 [noun:9] [pref] # yo na gu ni zi ma ti ho o
pref 伊豆諸島 [noun:5] [pref] # i zu syo to o
pref 小笠原諸島 [noun:8] [pref] # o ga sa wa ra syo to o
pref 東部富士五湖 [noun:7] [pref] # to o bu hu zi go ko
city 稚内 [noun:4] [city] # wa ka na i
city 旭川 [noun:5] [city] # a sa hi ka wa
city 留萌 [noun:3] [city] # ru mo i

city 札幌 [noun:3][city] # sa po ro
city 岩見沢 [noun:5][city] # i wa mi za wa
city 倶知安 [noun:3][city] # ku tya N
city 網走 [noun:4][city] # a ba si ri
city 北見 [noun:3][city] # ki ta mi
city 紋別 [noun:4][city] # mo n be tu
city 釧路 [noun:3][city] # ku si ro
city 根室 [noun:3][city] # ne mu ro
city 帯広 [noun:4][city] # o bi hi ro
city 室蘭 [noun:4][city] # mu ro ra N
city 浦河 [noun:4][city] # u ra ka wa
city 函館 [noun:4][city] # ha ko da te
city 江差 [noun:3][city] # e sa si
city 青森 [noun:4][city] # a o mo ri
city 八戸 [noun:4][city] # ha ti no he
city むつ [noun:2][city] # mu tu
city 深浦 [noun:4][city] # hu ka u ra
city 弘前 [noun:4][city] # hi ro sa ki
city 秋田 [noun:4][city] # a ki ta
city 鷹巣 [noun:4][city] # ta ka no su
city 横手 [noun:3][city] # yo ko te
city 盛岡 [noun:4][city] # mo ri o ka
city 宮古 [noun:3][city] # mi ya ko
city 大船渡 [noun:5][city] # o o hu na to
city 二戸 [noun:3][city] # ni no he
city 一関 [noun:5][city] # i ti no se ki
city 仙台 [noun:4][city] # se N da i
city 石巻 [noun:4][city] # i si ma ki
city 古川 [noun:4][city] # hu ru ka wa
city 白石 [noun:4][city] # si ro i si
city 山形 [noun:4][city] # ya ma ga ta
city 酒田 [noun:3][city] # sa ka ta
city 新庄 [noun:4][city] # si n zyo o
city 米沢 [noun:4][city] # yo ne za wa
city 福島 [noun:4][city] # hu ku si ma

city 小名浜 [noun:4][city] # o na ha ma
city 若松 [noun:4][city] # wa ka ma tu
city 白河 [noun:4][city] # si ra ka wa
city 郡山 [noun:5][city] # ko o ri ya ma
city 相馬 [noun:3][city] # so o ma
city 田島 [noun:3][city] # ta zi ma
city 水戸 [noun:2][city] # mi to
city 土浦 [noun:4][city] # tu ti u ra
city 宇都宮 [noun:5][city] # u tu no mi ya
city 大田原 [noun:5][city] # o o ta ha ra
city 前橋 [noun:4][city] # ma e ba si
city 水上 [noun:4][city] # mi zu ka mi
city 熊谷 [noun:4][city] # ku ma ga ya
city さいたま [noun:4][city] # sa i ta ma
city 秩父 [noun:3][city] # ti ti bu
city 東京 [noun:4][city] # to o kyo o
city 大島 [noun:4][city] # o o si ma
city 父島 [noun:4][city] # ti ti zi ma
city 八丈島 [noun:5][city] # ha ti zyo zi ma
city 千葉 [noun:2][city] # ti ba
city 銚子 [noun:3][city] # tyo o si
city 館山 [noun:4][city] # ta te ya ma
city 横浜 [noun:4][city] # yo ko ha ma
city 小田原 [noun:4][city] # o da wa ra
city 長野 [noun:3][city] # na ga no
city 松本 [noun:4][city] # ma tu mo to
city 諏訪 [noun:2][city] # su wa
city 飯田 [noun:3][city] # i i da
city 軽井沢 [noun:5][city] # ka ru i za wa
city 甲府 [noun:3][city] # ko o hu
city 河口湖 [noun:5][city] # ka wa gu ti ko
city 静岡 [noun:4][city] # si zu o ka
city 浜松 [noun:4][city] # ha ma ma tu
city 御前崎 [noun:5][city] # o ma e za ki
city 三島 [noun:3][city] # mi si ma

city 網代 [noun:3][city] # a zi ro
city 石廊崎 [noun:5][city] # i ro o za ki
city 名古屋 [noun:3][city] # na go ya
city 豊橋 [noun:4][city] # to yo ha si
city 岐阜 [noun:2][city] # gi hu
city 高山 [noun:4][city] # ta ka ya ma
city 津 [noun:1][city] # tu
city 尾鷲 [noun:3][city] # o wa se
city 四日市 [noun:4][city] # yo ka i ti
city 上野 [noun:3][city] # u e no
city 新潟 [noun:4][city] # ni i ga ta
city 長岡 [noun:4][city] # na ga o ka
city 高田 [noun:3][city] # ta ka da
city 相川 [noun:4][city] # a i ka wa
city 湯沢 [noun:3][city] # yu za wa
city 津川 [noun:3][city] # tu ga wa
city 富山 [noun:3][city] # to ya ma
city 伏木 [noun:3][city] # hu si ki
city 金沢 [noun:4][city] # ka na za wa
city 輪島 [noun:3][city] # wa zi ma
city 福井 [noun:3][city] # hu ku i
city 敦賀 [noun:3][city] # tu ru ga
city 大野 [noun:3][city] # o o no
city 大津 [noun:3][city] # o o tu
city 彦根 [noun:3][city] # hi ko ne
city 京都 [noun:3][city] # kyo o to
city 舞鶴 [noun:4][city] # ma i zu ru
city 大阪 [noun:4][city] # o o sa ka
city 神戸 [noun:3][city] # ko u be
city 豊岡 [noun:4][city] # to yo o ka
city 洲本 [noun:3][city] # su mo to
city 姫路 [noun:3][city] # hi me zi
city 奈良 [noun:2][city] # na ra
city 風屋 [noun:3][city] # ka ze ya
city 和歌山 [noun:4][city] # wa ka ya ma

city 潮岬 [noun:6] [city] # si o no mi sa ki
city 岡山 [noun:4] [city] # o ka ya ma
city 津山 [noun:3] [city] # tu ya ma
city 広島 [noun:4] [city] # hi ro si ma
city 呉 [noun:2] [city] # ku re
city 福山 [noun:4] [city] # hu ku ya ma
city 庄原 [noun:4] [city] # syo o ba ra
city 松江 [noun:3] [city] # ma tu e
city 浜田 [noun:3] [city] # ha ma da
city 西郷 [noun:4] [city] # sa i go o
city 鳥取 [noun:3] [city] # to to ri
city 米子 [noun:3] [city] # yo na go
city 徳島 [noun:4] [city] # to ku si ma
city 池田 [noun:3] [city] # i ke da
city 日和佐 [noun:3] [city] # hi wa sa
city 高松 [noun:4] [city] # ta ka ma tu
city 松山 [noun:4] [city] # ma tu ya ma
city 新居浜 [noun:4] [city] # ni i ha ma
city 宇和島 [noun:4] [city] # u wa zi ma
city 高知 [noun:3] [city] # ko o ti
city 室戸岬 [noun:6] [city] # mu ro to mi sa ki
city 足摺 [noun:4] [city] # a si zu ri
city 下関 [noun:5] [city] # si mo no se ki
city 山口 [noun:4] [city] # ya ma gu ti
city 柳井 [noun:3] [city] # ya na i
city 萩 [noun:2] [city] # ha gi
city 福岡 [noun:4] [city] # hu ku o ka
city 八幡 [noun:3] [city] # ya ha ta
city 飯塚 [noun:4] [city] # i i zu ka
city 久留米 [noun:3] [city] # ku ru me
city 大分 [noun:4] [city] # o o i ta
city 日田 [noun:2] [city] # hi ta
city 中津 [noun:3] [city] # na ka tu
city 佐伯 [noun:3] [city] # sa e ki
city 長崎 [noun:4] [city] # na ga sa ki

city 佐世保 [noun:3][city] # sa se bo
city 巖原 [noun:4][city] # i zu ha ra
city 福江 [noun:3][city] # hu ku e
city 佐賀 [noun:2][city] # sa ga
city 伊万里 [noun:3][city] # i ma ri
city 熊本 [noun:4][city] # ku ma mo to
city 阿蘇乙姫 [noun:6][city] # a so o to hi me
city 牛深 [noun:4][city] # u si bu ka
city 人吉 [noun:4][city] # hi to yo si
city 宮崎 [noun:4][city] # mi ya za ki
city 延岡 [noun:4][city] # no be o ka
city 都城 [noun:6][city] # mi ya ko no zyo o
city 油津 [noun:4][city] # a bu ra tu
city 高千穂 [noun:4][city] # ta ka ti ho
city 鹿児島 [noun:4][city] # ka go si ma
city 阿久根 [noun:3][city] # a ku ne
city 枕崎 [noun:5][city] # ma ku ra za ki
city 鹿屋 [noun:3][city] # ka no ya
city 西之表 [noun:5][city] # i ri o mo te
city 名瀬 [noun:2][city] # na ze
city 沖永良部 [noun:6][city] # o ki no e ra bu
city 那覇 [noun:2][city] # na ha
city 名護 [noun:2][city] # na go
city 久米島 [noun:4][city] # ku me zi ma
city 南大東島 [noun:9][city] # mi na mi da i to o zi ma
city 宮古島 [noun:5][city] # mi ya ko zi ma
city 石垣島 [noun:6][city] # i si ga ki zi ma
city 与那国島 [noun:6][city] # yo na gu ni zi ma