語りに傾聴を示す応答の生成に関する研究

伊藤 滉一朗

概要

人間には，自らの考えや気持ち，状況や出来事などを伝えたいという欲求が存在する．この欲求を満たす方法の 1 つは，聴き手に語ることである．語りによって聴き手の理解や共感が得られれば，語り手の自己肯定感を高める効果や，孤独感やストレスを和らげる効果が期待できる．語ることは人間にとって重要な活動であり，あらゆる場面で行われている．ただし，語るためには，聴き手の存在が不可欠である．しかし，現代社会では，孤独化の進行に伴う聴き手不在の生活シーンが増加し，日常生活において聴き手がいることは，当たり前ではなくなりつつある．このような状況においては，人が語れる機会を増やすことは重要な課題であるといえる．

上述の課題に対して，コミュニケーションロボットやスマートスピーカーなどの会話エージェントが語りを聴く役割を担うことが考えられる．これらが聴き手として認められるには，語りを熱心に聞いていること，すなわち，傾聴していることを語り手に伝達する機能を備える必要がある．このための明示的な手段は語りに応答することであり，頷きなどのジェスチャが有効な手段の 1 つである．また，語りを傾聴していることを伝える手段としては，発話の表出も有力である．本論文では，傾聴を示す目的で語りに応答する発話を傾聴応答と呼ぶ．傾聴応答の代表例は,「はい」や「ええ」などの相槌であるが，感心や評価，繰り返しなど，相槌以外の応答も存在する．傾聴応答を適切に生成できれば，語り手の語る意欲を促進するなどの効果が期待できる．

傾聴応答は，任意のタイミングで生成すればよいわけではなく，適切なタイミングで生成する必要がある．また，傾聴応答には,「はい」「そうですね」「すごいですね」など，様々な応答表現が存在し，語りの内容に合わせた適切な応答表現を選定する必要がある．傾聴応答の生成タイミングや応答表現が不適切である場合には，かえって話し手の語る意欲を減退させかねない．そのため，語りの聴き手を担う会話エージェントによる傾聴応答の生成を実現するには，傾聴応答タイミングの検出と傾聴応答表現の選定を解決できる必要がある．そこで，本論文では，以下を目的とする．

(1) 傾聴応答の生成に利用可能な傾聴応答コーパスの構築

(2) 傾聴応答の表出タイミングとしての適切さの推定手法の提案

(3) 傾聴応答の一種である不同意応答の生成に利用可能なコーパスの作成，並びに，その生成の実現性の検証

本研究では，主に自然言語処理の側面から，上述したそれぞれの目的の達成を目指す．

本論文は全 5 章から構成される．第 1 章は本論文の序論であり，傾聴応答の生成に関する課題及び研究動向を示すとともに，本論文の位置づけとアプローチを述べる．

第 2 章では，本研究が目指す傾聴応答の生成に利用可能な傾聴応答コーパスの構築について述べる．本研究では，語りに対する傾聴応答を収集するために，事前に収録された語りの音声に対し，表出するに相応しい傾聴応答の表現と表出タイミングを作業者が付与するという収集方式を採用する．この収集方式には，作業者は聴き手として傾聴応答の付与に集中できるため，本研究の想定に沿った傾聴応答を収集できるといった利点がある．さらに，傾聴応答が語りに影響しないため，単一の語りに対して複数の傾聴応答を収集できる．この収集方式によって，単一の語りに対する聴き手 11 名分の傾聴応答を 148,962 個収集した．多頻度性，多様性，網羅性，自然さの観点から収集した応答を評価することで，本方式により自然で多様な応答を大量かつ網羅的に収集できることを確認した．最後に，収集した応答を利用して，傾聴応答の生成タイミングの検出実験を実施した．

第 3 章では，適切なタイミングでの傾聴応答の自動生成に向けて，傾聴応答の表出タイミングとしての適切さの度合いを推定する手法を提案する．具体的には，この適切さを表す指標として，所与のタイミングにおいて傾聴応答を表出する聴き手の割合である表出率を導入し，その推定を行う．表出率は，語りを傾聴する会話エージェントが傾聴応答を表出するか否かを判断するための材料として利用できる．例えば，推定された表出率をもとに，語り手の嗜好や会話エージェントの個性などを加味して，表出するか否かを柔軟に決定することが考えられる．本手法では，Transformer ベースの手法でエンコードされた語りの音響情報と言語情報を用いて，表出率を推定する．本手法の推定性能を評価するために，本研究で構築した傾聴応答コーパスを用いて実験を行い，本手法の有効性を確認した．また，表出率を利用して応答タイミングを検出する実験を行い，その有用性を確認した．

第 4 章では，語りの傾聴における不同意応答の生成について述べる．語りの傾聴では，語り手の発話を受容することが聴き手の基本的な応答方略となる．しかし，語りには，自虐や謙遜などの発話が含まれることがある．この場合，語り手の発話に同意しないことを示す応答，すなわち，不同意応答を確実に表出できることが求められる．本論文では，語りを傾聴する会話エージェントによる不同意応答生成の実現性を示すために，不同意応答の生成に利用できる応答コーパスを作成できること，並びに，応

答コーパスを用いて不同意応答を適切に生成できることを検証する．まず，不同意応答の生成に利用できる応答コーパスを作成するために，時間制約のない環境で語りデータに不同意応答のタイミングと表現をタグ付けし，応答コーパスを作成した．評価の結果，不同意応答タイミングを網羅的に，不同意応答表現を安定的にタグ付けできることを確認した．続いて，事前学習済みの Transformer ベースのモデルに基づく，不同意応答タイミングの検出手法，及び，不同意応答表現への分類手法を実装し，実験により応答コーパスを用いた不同意応答生成の実現性を確認した．

最後に，第 5 章で本論文をまとめ，今後の研究課題，及び，将来の展望について述べる．

目 次

表 目 次

図 目 次

第1章　まえがき

1.1　語りに傾聴を示す応答の生成

人間には，自らの考えや気持ち，状況や出来事などを伝えたいという欲求が存在する．この欲求を満たす方法の 1 つは，聴き手に語ることである．語りによって聴き手の理解や共感が得られれば，語り手の自己肯定感を高める効果や，孤独感やストレスを和らげる効果が期待できる．また，語るという行為は，心理療法の分野でも注目されており，Rogers によるアクティブリスニングやパーソン・センタード・アプローチ [1, 2]，ナラティブセラピーやナラティブアプローチ [3] といった治療法などにも取り入れられている．語ることは人間にとって重要な活動であり，あらゆる場面で行われている．

語るためには，聴き手の存在が不可欠である．しかし，日常生活において聴き手がいるという状況は，当たり前のことではなくなりつつある．日本における国民生活基礎調査 [4] によると，高齢者を中心に，単独世帯数が増加傾向にある．また，内閣官房孤独・孤立対策担当室による孤独・孤立の実体把握に関する全国調査 [5] では，孤独感を抱えた人々の存在が報告されている．イギリスで実施された統計調査 Community Life Survey 2021/22 [6] でも，同様の結果が報告されている．孤独化に伴い，聴き手不在の状況が増えつつある現代社会においては，人が語れる機会を増やすことは重要な課題であるといえる．

上述の課題に対して，コミュニケーションロボットやスマートスピーカーなどの会話エージェントが語りを聴く役割を担うことが考えられる． このような会話エージェントが実現できれば，人間の聴き手の有無によらず，語りたいときに自由に語ることが可能となる．語りの利活用に関しては，語りに基づく認知症の自動検出に関する試み [7, 8, 9] や鬱病の自動検出に関する試み [10, 11, 12] が存在している．語りを聴く役割を担う会話エージェントが実現できれば，これらの試みと組み合わせることで，認知症や鬱病などの疾病の早期検出なども期待できる．

会話エージェントが聴き手として認められるには，語りを傾聴していることを語り手に伝達する機能を備える必要がある．傾聴とは，耳を傾けて熱心に聞くこと [13]，

表 1.1: 傾聴応答の例とその種類

語り	傾聴応答
地方に行った時そこにある美術館にはなるべく行くようにしています	**そうなんですね【感心】**
わたくしのこんにちまでの仕事はえーライターです	**ライター【繰り返し】**
書道も好きで総理大臣賞も頂いたりして	**凄いですね【評価】**
五千歩歩くということはなかなか難しいことで	**そうですね【同意】**
年をとると逆に力が入らなくなってうまくいくこともあるんですよね	**なるほど【納得】**
歯医者さんに行かずにいたら右側の差し歯が三本抜けちゃいました	**えー【驚き】**
十二時前から食事しながらゆっくりして	**のんびりと【言い換え】**
時間もとれるようになりましたので四国の八十八か所のお遍路を	**巡りたい【補完】**

聞き漏らすまいとして熱心に聞くこと [14] を指す．例えば，聴き手が腕組みをした姿勢で，終始無表情であれば，語り手は語りづらさを感じるものと考えられる．語りを傾聴する姿勢に関する指針としては，Squarely（相手と真っ直ぐに向き合う），Open（開いた姿勢で接する），Lean（相手の方に上体を少し傾ける），Eye Contact（相手と適切に視線を合わせる），Relaxed（適度にリラックスする）の 5 つの頭文字をとった，SOLER が提唱されている [15, 16]．このような姿勢に加えて，首を縦に振る頷きも傾聴態度を示すための手段として挙げられる．これまでに，頷きの移動範囲や反復回数などの物理的特徴に関する分析が進められており [17, 18, 19, 20, 21, 22]，語りの音声や語り手の視線などから頷きのタイミングを予測する手法も提案されている [23, 24, 25, 26]．また，笑いの表出によっても傾聴態度を示すことができる．語り手の笑いに対して聴き手も誘われるように笑うという，同調的笑いや共有笑いと呼ばれる現象を対象とした，笑いの生成に関する研究も進められている [27, 28]．

語りを傾聴していることを伝える手段としては，発話の表出も有力である．以降では，傾聴を示す目的で語りに応答する発話を**傾聴応答**と呼ぶ．一般に，傾聴応答にはいくつかの種類があり，「はい」や「ええ」などの相槌のほか，感心，繰り返し，評価，同意，納得，驚き，言い換え，補完などがある [29]．表 1.1 に，傾聴応答の例をそ

の種類と共に示す.【 】で囲んだ文字列が傾聴応答の種類である．傾聴応答を適切に表出できれば，語り手の語る意欲を促進するなどの効果が期待できる [16]．会話エージェントによる傾聴応答の生成を実現するには，下記の 2 つの問題を解決できる必要がある．

(1) 応答タイミングの検出
傾聴応答は，任意のタイミングで表出すればよいわけではない．例えば，語り手の語りを遮るようなタイミングでの傾聴応答の表出は，かえって語る意欲を減退させることに繋がりかねない．そのため，語りを傾聴する会話エージェントの実現のためには，傾聴応答の生成に適したタイミングを検出できる必要がある．

(2) 応答表現の選定
傾聴応答の代表例は相槌であり,「はい」や「ええ」などが代表的な応答表現として挙げられる．ただし，表 1.1 でも示した通り，傾聴応答には相槌以外にもいくつかの種類が存在しており，その応答表現は多様である．また，同一の種類の傾聴応答であっても，とりうる応答表現は 1 つではない．例えば，語りに含まれる語句の一部を用いて応答する形式の発話である繰り返し応答は，語りに合わせて表現が大きく変化する．表出される応答表現が語りの内容に適さない場合には，語る意欲を高める効果を期待できない．そのため，語りを傾聴する会話エージェントの実現のためには，語りの内容や状況に応じて適切に生成する応答表現を選定できる必要がある．

これらの問題を解決するため，これまでいくつかの研究が行われている．次節以降では，まず，これまでの傾聴応答の生成に関する研究動向を概観する．次に，本論文の目的と内容について述べる．

1.2 語りに傾聴を示す応答の生成に関する研究動向

本節では，前節で挙げた各問題ごとに，傾聴応答の生成に関する研究動向を示す．

1.2.1 傾聴応答タイミングの検出

傾聴応答タイミングの検出，すなわち，所定のタイミングに傾聴応答が可能であるかどうかの判定[1]については，傾聴応答の代表例である相槌の生成タイミング検出に関する研究が多数存在する [30, 31, 32, 33, 34]．これらの研究では，語りから抽出されるピッチやパワーなどの音響情報が，表出タイミングの検出に有効であると報告されている．また，語りに含まれる単語や文構造などの言語情報も，その有効性が示されている．これまでに，これらの情報を入力とした検出手法として，ルールベースによる手法 [30, 35]，n-gram モデルによる手法 [36]，有限状態トランスデューサによる手法 [37]，決定木による手法 [38, 31]，CRF による手法 [39]，SVM による手法 [32, 40] などが提案されている．近年では，LSTM や CNN, Transformer などのニューラルネットワークを用いた手法 [41, 33, 42, 43, 34, 44] も提案されている．

語りに含まれる語句の一部をそのまま用いて応答する形式の発話を繰り返し応答と呼ぶ．繰り返し応答の生成タイミングを検出するには，繰り返されるに相応しい語句が語りに含まれてるか否かを判定できる必要がある．これまでに，繰り返される語句についての分析が進められており，出現が珍しい語句や，固有表現，フィラーの直後の語句が，繰り返されやすい傾向にあることが知られている [45, 46]．繰り返し応答タイミングの検出を試みた研究には，以下のようなものがある．

- 繰り返し応答を生成するか否かを判断するための手法として，語りに含まれる焦点語の抽出結果に基づく手法が提案されている [47, 48]．これらの手法では，焦点語と,「どんな」「どの」「なんの」「どこの」「いつの」「だれの」などの疑問視との n-gram 確率を算出し，その確率が閾値を超えていた場合には，繰り返し応答ではなく，焦点語に関する掘り下げ質問を生成する．なお，焦点語の抽出には，文末にもっとも近い名詞または形容詞とするルールベースによる抽出手法 [48] や，語りに含まれる語句の品詞などの特徴量を入力とする CRF による抽出手法 [49] が用いられている．

- 繰り返し応答は語り手の発話の一部を用いて応答する発話であるため，音声認識誤りの影響を受けやすい．この点を踏まえて，音声認識結果の信頼度に基づいて，繰り返し応答の生成タイミングを決定する手法も存在する [50]．語り手の発話の最終述語に対する音声認識結果の信頼度が高い場合に，繰り返し応答の生成が可能であると判断する．

[1]以降では，特に断りがない限り，傾聴応答のタイミング検出とは，所定のタイミングに傾聴応答が可能であるかどうかを判定することを意味する．

- 語りに含まれる語句の極性に着目した，繰り返し応答の生成タイミングの検出手法も提案されている [51]．語りに含まれるネガティブな語句を繰り返すと，語り手は心地悪さを感じ，語りを継続する意欲が低下してしまう可能性が指摘されている [52]．この点を踏まえて，繰り返すに相応しい語句が語りに含まれているか否かを判定する際に，極性に基づく制約を設けている．

- 近年は，事前学習済みの BERT [53] を用いて，繰り返すに相応しい語句が含まれているか否かを判定する手法も提案されている [54]．

語りの内容を評価するような傾聴応答を評価応答と呼ぶ．これまでに，語りの発話の極性（ポジティブ/ネガティブ）に基づく，評価応答の生成タイミングの検出手法が提案されている [47, 48]．これらの手法では，語りの発話の極性を判定するための辞書 [55, 56, 57] を用いて，語りの発話に含まれる各語句の極性を計算し，各語句の極性をまとめあげた結果がポジティブまたはネガティブである場合（つまり，ニュートラルではない場合）には，評価応答の生成が可能なタイミングとする．

相槌の生成タイミングの検出に関する研究は，多数存在する．一方で，相槌以外の傾聴応答については，繰り返しや評価を中心に，いくつかの研究が存在しているものの，生成タイミング検出に関する研究は限られている．

1.2.2 傾聴応答表現の選定

傾聴応答表現の選定に関する試みとして，以下のような研究が挙げられる．

- 山口ら [58] は，相槌の応答表現のバリエーションが常に同じである場合や，応答表現をランダムに変化させる場合には，会話のリズムに単調さや不自然さが残ることを問題点に挙げ，文脈に応じた相槌の多様な応答表現の生成手法を提案している．相槌の繰り返し回数に着目し，相槌の応答表現を，応答系 1 回（「うん」など），応答系 2 回（「うんうん」など），応答系 3 回（「うんうんうん」など），感情表出系（「はー」など）の 4 つにカテゴリ化して，それぞれの生成に相応しいタイミングを検出する．

- 繰り返し応答についても，その応答表現の選定に関する研究が行われている．繰り返しの応答表現は，事前に検出された繰り返すに相応しい語りの語句（繰り返し対象語句）に基づいて生成される．繰り返し対象語句に「ですか」を連接するなどの，応答表現のフレームを事前に用意しておき，繰り返し応答表現

を生成する手法が存在する [47, 48]．繰り返し対象語句を含む発話における述語と格の関係に基づいて，繰り返し応答表現を生成する手法も提案されている [50, 51]．近年は，事前学習済みの T5 [59] を用いた繰り返し応答表現の生成手法も提案されている [60]．この手法では，語りに含まれる語句に対して，繰り返し対象語句としての適切さを推定し，その推定結果を踏まえて，繰り返し応答表現を選定する．

- 評価応答についても，その応答表現の選定に関する研究が存在する．評価応答表現は，ポジティブな表現とネガティブな表現に大別される．評価応答の対象となる発話の極性ごとに，事前に生成する応答表現を定義しておき，定義に従って応答表現を選定するという手法が存在している [47, 48]．これらの手法では，評価応答の対象となる発話の極性がポジティブであれば「いいですね」や「素敵ですね」，ネガティブであれば「大変ですね」や「残念でしたね」という応答表現を生成するように定めている．

- 特定の傾聴応答の種類に限定せず，encoder-decoder モデルによって，傾聴応答表現を選定する手法も存在する [44, 61]．村田ら [61] は，応答タイミング直前の語りの発話を入力として，LSTM を用いた encoder-decoder モデルによって，生成するに相応しい応答表現の選定を試みている．室町ら [44] は，GPT2 [62] ベースの話者交代予測モデルである TurnGPT [63] を拡張し，生成モデルによる応答表現の選定手法を提案している．

傾聴応答の生成に関する研究は，その生成タイミングの検出を中心に進められており，井上ら [48] も指摘しているように，応答表現の選定に関する研究は限られている．

1.3 本論文の目的

1.1 節で述べた通り，会話エージェントが語りの聴き手として認められるには，語りを傾聴していることを伝達できる必要がある．そのための手段として，語りを聴く姿勢や表情，頷き，笑いなどの生成のほか，傾聴を示す目的で語りに応答する発話である傾聴応答の生成が挙げられる．本研究では，そのうち，傾聴応答の生成に焦点を当てる．傾聴応答の生成を実現するためには，傾聴応答タイミングの検出と傾聴応答表現の選定を解決できる必要がある．そこで，本論文では，以下を目的とする．

(1) 傾聴応答の生成に利用可能な傾聴応答コーパスの構築

(2) 傾聴応答の表出タイミングとしての適切さの推定手法の提案

(3) 傾聴応答の一種である不同意応答の生成に利用可能なコーパスの作成，並びに，その生成の実現性の検証

本研究では，主に自然言語処理の側面から，上述したそれぞれの目的の達成を目指す．

傾聴応答の生成の実現においては，語りに対する傾聴応答のデータを用いた，分析や観察，統計モデルの獲得が重要となる．そのためには，まず，語りに対する傾聴応答のデータが必要となる．そこで本研究では，語りに対する傾聴応答を収集することで，傾聴応答コーパスを構築することを目的とする．収集対象の応答は，従来研究の主な研究対象である相槌に限定せず，多様な傾聴応答の収集を目指す．傾聴応答の収集は，事前に収録された語りデータに同期して，作業者が傾聴応答の表出タイミングと表現を付与することで行う．本研究では，同一の語りデータに対して，複数名の作業者が傾聴応答の付与を行う．さらに，収集した傾聴応答の評価と分析，および，その利用例も示す．

傾聴応答は，任意のタイミングで表出すればよいというわけではなく，適切なタイミングで表出する必要がある．これまでに，相槌を中心に，傾聴応答タイミングの検出に関する研究が行われてきた．人間同士の対話における傾聴応答を抽出することで，傾聴応答タイミング検出のためのデータを作成し，作成したデータに基づき検出手法を開発することが主流であった．これらの検出手法は，与えられたタイミングを傾聴応答の生成に適するか否かの二値に分類するものであり，その結果に従って傾聴応答の生成を決定することを想定している．一方，本論文では，適切なタイミングでの傾聴応答の自動生成に向けて，傾聴応答の表出タイミングとしての適切さの度合いを推定する手法を提案する．この適切さを表す指標として，所与のタイミングにおいて傾聴応答を表出する聴き手の割合である表出率を導入し，その推定を行う．表出率は，語りを傾聴する会話エージェントが傾聴応答を表出するか否かを判断するための材料として利用できる．例えば，推定された表出率をもとに，語り手の嗜好や会話エージェントの個性などを加味して，表出するか否かを柔軟に決定することが考えられる．表出率の予測手法の実現には，同一の語りに対して複数名の聴き手の傾聴応答が付与されたデータが必要となる．本研究では，表出率の予測手法の開発にも利用可能な傾聴応答コーパスを構築する．

本研究では，傾聴における不同意応答に関して，その生成に利用可能な応答コーパスの作成と，その生成の実現性の検証に取り組む．傾聴応答の生成に関する多くの研究では，典型的な応答である相槌が研究対象とされてきた．語りを傾聴する聴き手の

基本的な応答方略は，語り手の発話を受容することである．例えば，傾聴応答の生成に関する主な研究対象である相槌は,「語りを続けて」というシグナルや内容理解を示す機能を持っており [64]，これも，語りを受容していることを伝えるものといえる．一方で，語りでは，時として自虐や謙遜などの発話が行われることがある．この場合，その発話内容を否定することなくそのまま受容することは必ずしも適切ではなく，語り手の発話に同意しないことを示す応答，すなわち，不同意応答を積極的に表出することが求められる．このように，語りの傾聴を担う会話エージェントが不同意を示すべき発話を検出し応答できることは不可欠な機能であるものの，傾聴応答生成に関する従来研究において，不同意応答の生成に関する試みは行われていない．そこで本研究では，不同意応答の生成に関して，不同意応答のタイミングと応答表現が付与された応答コーパスを作成する．さらに，作成した応答コーパスを用いて，不同意応答タイミングの検出と不同意応答表現の分類の実現性を実験的に示す．

なお，本研究では，上述した取り組みを実施するにあたり，日本語の語りデータに傾聴応答が付与されたコーパスを作成及び利用する．日本では特に，独居高齢者を中心に単独世帯が増加しつつあり，人間に代わって語りの聴き手を担う会話エージェントが望まれる状況にある．ただし，言語ごとに有効な傾聴応答の生成方略は異なる可能性がある．例えば，傾聴応答の代表例である相槌については，日本語の会話における使用頻度が，アメリカ英語の会話における使用頻度よりも高いことが指摘されている [64, 65, 66]．本研究では，言語間における傾聴応答の有効な生成方略の違いなどの検討については，今後の課題とする．また，日本語の語りデータとしては，高齢者のナラティブコーパスを用いる．独居高齢者が増加している日本においては，高齢者は本研究が目指す聴き手を担う会話エージェントが想定する主要なユーザであると考えられる．世代に応じて，語られる内容や口癖，用いられる語句などに差異があるものと考えられるが，高齢者以外の語りに対する傾聴応答の生成については今後の課題とする．

1.4 本論文の内容

本論文ではまず第一に，語りの聴き手を担う会話エージェントの実現を目的に，傾聴応答の収集と評価を行う．本研究では，傾聴応答を収集するために，事前に収録された語りの音声に対し，表出するに相応しい傾聴応答の表現と表出タイミングを作業者が付与するという収集方式を採用した．この収集方式には，作業者は聴き手として傾聴応答の付与に集中できるといった利点がある．さらに，傾聴応答が語りに影響

しないため，単一の語りに対して複数の傾聴応答を収集可能である．この収集方式によって，単一の語りに対する聴き手 11 名分の傾聴応答を 148,962 個収集した．多頻度性，多様性，網羅性，自然さの観点から収集した応答を評価することで，本方式により自然で多様な応答を大量かつ網羅的に収集できることを確認した．最後に，収集した応答を利用して，傾聴応答の生成タイミングの検出実験を実施した．

第二に，適切なタイミングでの傾聴応答の生成に向けて，傾聴応答の表出タイミングとしての適切さの度合いを推定する手法を提案する．具体的には，この適切さを表す指標として，所与のタイミングにおいて傾聴応答を表出する聴き手の割合である表出率を導入し，その推定を行う．本手法では，Transformer ベースの手法でエンコードされた語りの音響情報と言語情報を用いて，表出率を推定する．本手法の推定性能を評価するために，本研究で収集した応答データを用いて実験を行い，本手法の有効性を確認した．また，表出率を利用して表出タイミングを検出する実験を行い，その有用性を確認した．

第三に，語りの傾聴における不同意応答の生成について述べる．語りの傾聴では，語り手の発話を受容することが基本だが，語りには自虐や謙遜などの発話が含まれることがある．このような発話に対しては，不同意応答を確実に表出できることが求められる．そこで本論文では，語りを傾聴する会話エージェントによる不同意応答生成の実現性を示すために，不同意応答の生成に利用できる応答コーパスを作成できること，並びに，応答コーパスを用いて不同意応答を適切に生成できることを検証する．まず，不同意応答の生成に利用できる応答コーパスを作成するために，時間制約のない環境で語りデータに不同意応答のタイミングと表現をタグ付けし，応答コーパスを作成した．評価の結果，不同意応答タイミングを網羅的に，不同意応答表現を安定的にタグ付けできることを確認した．続いて，事前学習済みの Transformer ベースのモデルに基づく，不同意応答タイミングの検出手法，及び，不同意応答表現への分類手法を実装し，実験により応答コーパスを用いた不同意応答生成の実現性を確認した．

1.5 本論文の構成

本論文の構成は以下の通りである．

第 2 章では，語りの聴き手を担う会話エージェントの実現を目的とした，傾聴応答の収集と評価について述べる．まず，本研究が目指す傾聴応答の生成要件について論じる．次に，傾聴応答の収集方針と収集方法を説明する．本研究では，事前に収録された語りの音声に対し，表出するに相応しい傾聴応答の表現と表出タイミングを作業

者が付与するという収集方式を採用する．続いて，収集された応答を多頻度性，多様性，網羅性，自然さの観点から評価する．最後に，収集した応答を用いて，傾聴応答タイミングの検出実験を実施し，その結果について報告する．

第 3 章では，傾聴応答の表出タイミングとしての適切さの度合いを推定する手法を提案する．まず，傾聴応答の表出タイミングとしての適切さについて論じる．本研究では，この適切さを表す指標として，所与のタイミングにおいて傾聴応答を表出する聴き手の割合である表出率を導入する．次に，表出率の推定手法を提案し，その推定性能を評価するための実験を通して，本手法の有効性を確認する．最後に，傾聴応答の表出タイミングの検出を例に，表出率の有効性を実験的に検証する．

第 4 章では，語りの傾聴における不同意応答の生成について述べる．まず，言語学の視点も交えつつ，不同意応答の定義と分類について論じた後，本研究が対象とする不同意応答の定義を説明する．次に，不同意応答生成のための応答コーパスの作成について述べる．本研究では，時間制約のない環境で語りデータに不同意応答のタイミングと表現をタグ付けする．最後に，作成したコーパスを用いて，不同意応答タイミングの検出及び不同意応答表現の分類の実験を通して，不同意応答生成の実現性を示す．

最後に 5 章では，本論文のまとめと残された課題，将来の展望について述べる．

第2章　語りに傾聴を示す応答コーパスの構築

2.1　はじめに

現代社会では，独居高齢者の増加など，社会の個人化が進行している．これに伴い，聴き手不在の生活シーンが増加しており，人が語れる機会が失われつつある．そこで，コミュニケーションロボットやスマートスピーカーなどの会話エージェントが語りを聴く役割を担うことが考えられる．これらが聴き手として認められるには，傾聴していることを示す目的で語りに応答する発話である傾聴応答を，適切に生成できる必要がある．

傾聴応答に関しては，その代表である相槌を中心に，データの収集と分析，生成法の提案が行われている [30, 31, 32, 33, 34]．ただし，傾聴応答は相槌以外にも存在している．相槌に限定されない傾聴応答の生成方式については，実際の傾聴応答を幅広く収集し，データの観察と分析を通じた今後の検討が待たれる状況にある．

本章では，語りの聴き手を担う会話エージェントの実現を目的に，傾聴応答の収集と評価，その利用について述べる．本研究が目指す会話エージェントは，聴き手に徹することを想定しており，人に代わって聴き手を担うことで，語る機会を増やすことを期待できる．本研究では，傾聴応答を収集するために，事前に収録された語りの音声に対し，表出するに相応しい傾聴応答の表現と表出タイミングを作業者が付与するという収集方式を採用した．この収集方式には，

- 作業者は聴き手として傾聴応答の付与に集中できるため，本研究の想定に沿った傾聴応答を収集できる
- 傾聴応答が語りに影響しないため，単一の語りに対して複数の傾聴応答を収集できる

といった利点がある．この収集方式によって，単一の語りに対する聴き手 11 名分の傾聴応答を 148,962 個収集した．多頻度性，多様性，網羅性，自然さの観点から収集

した応答を評価することで，本方式により自然で多様な応答を大量かつ網羅的に収集できることを確認した．最後に，収集した応答を利用して，傾聴応答の生成タイミングの検出実験を実施した．

本章の構成を以下に示す．2.2 節では，本研究が目指す傾聴応答の生成要件と傾聴応答のデータについて述べる．次に，2.3 節で本研究における傾聴応答の収集方針と収集方法，収集結果を説明した後，収集された傾聴応答を分類する．2.4 節では，収集された傾聴応答を多頻度性，多様性，網羅性，自然さの観点から評価する．2.5 節では，収集された傾聴応答データの利用例として，傾聴応答の生成タイミングの検出について述べる．最後に，2.6 節で本章のまとめを行う．

2.2 傾聴応答の生成

傾聴応答の生成には，語り手の語る意欲を高める効果を期待できる [16]．ただし，単に傾聴応答を生成すればよいというわけではない．例えば，傾聴応答の表現が語りに適さない場合には，かえって語る意欲を下げかねない．表現については，単調であることや，バリエーションが乏しいことも好ましくない．また，傾聴応答を生成する際には，語りに相応しいタイミングで生成することが重要となる．さらに，そのようなタイミングでは積極的に応答を生成することで，傾聴態度を深く伝えることができる．したがって，

- 自然でかつ多様な応答表現であること
- 生成の頻度が高くかつ自然なタイミングであること

が，傾聴応答生成の要件となる．上記の要件を満たす生成方式が解明されていない現状では，傾聴応答の実データを集め，観察・分析を通して有用な知見を蓄積することが肝要である．

傾聴応答に関する研究の多くでは，傾聴応答のタイミングや表現が付与されたデータを用いて，分析や評価，生成法の提案をしている．そのデータは，人間どうしの双方向のやり取りを記録し，聴き手の発話から傾聴応答を取り出すことで作成されている．一方で，既存の音声データに対して相槌タイミングをタグ付けることによる，データ作成も行われている [40]．タグ付け作業者は，聴き手として作業に集中できるため，相槌に適したタイミングが網羅的にタグ付けられたデータの作成を期待できる．

2.3　傾聴応答データの収集

語りに対する傾聴応答の自動生成に向け，本研究では，応答表現が自然でかつ多様であり，表出頻度が高くかつ自然なタイミングであるような傾聴応答データを収集する．また，生成の頻度が高くかつ自然なタイミングであるという傾聴応答生成の要件を踏まえて，応答の表出が自然であるタイミングでは網羅的に応答が表出されているデータの収集を目指す．

2.3.1　収集の方針

本研究では，作業者が事前収録済みの語りの音声を聴きながら，語りに合わせて傾聴応答を表出するという収集方式を採用する．2.2 節で述べた通り，傾聴応答の収集方式としては，話し手と聴き手による双方向のやり取りを記録し，聴き手の発話から傾聴応答を取り出す方式も考えられる．しかし，本研究では，会話エージェントが聴き手に徹するコミュニケーション形態を想定している．そのため，語りの音声に対して傾聴応答を付与する本方式では，作業者は聴き手として傾聴応答の表出に集中できるため，想定に沿った傾聴応答の効率的な収集を期待できる．さらに，傾聴応答のタイミングだけでなく応答表現も付与することで，相槌に限定されない多様な傾聴応答の収集も期待できる．

また，双方向のやり取りから傾聴応答を取り出す方式では，聴き手の反応が話し手の振舞いに影響を及ぼす可能性があり，同一の語りに対する複数の聴き手による応答の収集ができない．一方，本方式は，同一の語りに対する複数の聴き手による応答の収集が可能である．同一の語りに対して複数の聴き手による応答を収集できれば，応答の多様性をさらに高められる．そこで本研究では，複数名の聴き手役の作業者が，同一の語りデータに対して独立に傾聴応答を表出することで，その収集を行う．表出された傾聴応答については，その文字化データ及び時刻データを語りデータに注釈付ける．

2.3.2　収集の方法と結果

語りのデータとして，高齢者のナラティブコーパス JELiCo [67] を使用した．コーパスには，30 名の高齢者による 1 人約 20 分の語りの音声が収録されている．全高齢者共通の 10 個の質問に対し，その回答を語るという収録形式が採用されている．本

表 2.1: 語りデータと応答データの規模

話者数	30	作業者数	11
合計発話時間	8:43:55	合計発話時間	22:16:34
発話単位数	13,238	発話単位数	148,962
形態素数	66,897	形態素数	232,651

研究では，語りの内容の適切さと応答の収録作業の適切さの観点から，269 個の語り音声に対する傾聴応答を収集した．

傾聴応答の表出は，接客の訓練を受けた経験を持つ作業者 11 名が担当した．11 名の作業者がすべて同一の語りデータに対して独立した環境で応答を表出した．作業者は，再生された語り音声に対しリアルタイムに応答を表出した．語り音声は途中で停止することなく，通しで 1 回だけ再生される．応答音声は接話マイクを通して収録した．

使用した語りデータと収集した応答データの規模を表 2.1 に示す．なお，人が知覚できるポーズで分割された音声を発話単位と定めている．語りと応答をそれぞれ形態素解析し，各形態素に発声時間を付与した．形態素解析には MeCab[68] を，時間の付与には Julius[69] の音素セグメンテーションキット[1]をそれぞれ用いた．形態素解析の辞書には，語りデータについては IPADIC neologd [2] を，応答データについては UniDic (Ver. 2.1.2) [3]をそれぞれ用いた．図 2.1 に，収録したデータの一部を示す．

2.3.3 応答タイプの付与

傾聴応答には様々な形態や役割が存在する．本研究では，収録したすべての傾聴応答に対して，その応答のタイプを人手で付与した．応答タイプは，文献 [29] を参考に相槌や感心など，16 種類定めた．表 2.2 にタイプ名とその説明を記す．

収録データにおける応答タイプにおいては，傾聴応答の代表である相槌の占める割合が，全体の 67.96%と最も大きかった．相槌以外の応答タイプは，全体の 32.04%を占めており，感心，繰り返し，評価の順に多く出現している．図 2.2 に全応答タイプの出現割合を，図 2.3 に相槌以外の応答タイプの出現割合をそれぞれ示す．以下に，感心，繰り返し，評価について，語りと応答の例を示す．

[1] http://julius.osdn.jp/index.php?q=ouyoukit.html
[2] https://github.com/neologd/mecab-ipadic-neologd
[3] https://ja.osdn.net/projects/unidic/releases/58338

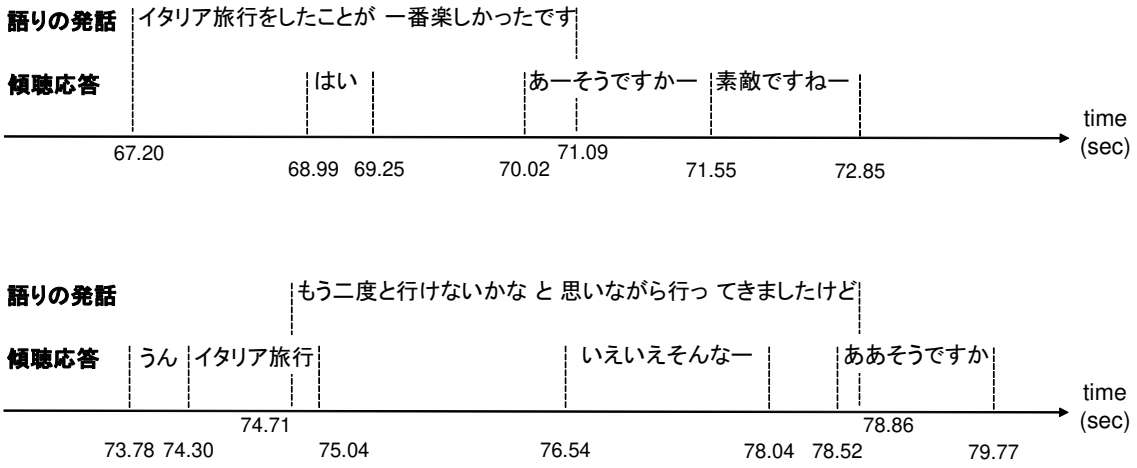


図 2.1: 収録データの例

- 感心
  【語り】地方に行った時そこにある美術館にはなるべく行くようにしています
  【応答】そうなんですね

- 繰り返し
  【語り】わたくしのこんにちまでの仕事はえーライターです
  【応答】ライター

- 評価
  【語り】書道も好きで総理大臣賞も頂いたりして
  【応答】凄いですね

2.4 収集した応答データの分析と評価

収集した応答の分析により，収録データの多頻度性，多様性，網羅性，自然さを評価した．

2.4.1 応答の多頻度性

収録データにおける傾聴応答の出現頻度を集計した．収録データには，語りに対する作業者 11 名の傾聴応答が含まれている．傾聴応答の発生間隔，すなわち，何秒

表 2.2: 傾聴応答のタイプと役割

タイプ名	役割
相槌	聞き取りに成功したことを示す
感心	相手の発話内容に対して，感心，驚き，気づき等の態度を示す
評価	相手の発話内容が示す事態に対しての態度を示す
同意	相手の発話内容に対する同意の態度を示す
不同意	相手の発話内容が同意できないものであることを示す
繰り返し	相手の発話内容を理解したことを示し，相手に安心感を与える
言い換え	相手の発話内容を理解して共有しようとする態度を示す
納得	相手の発話内容が納得できることを示す
驚き	相手の発話内容に対して，純粋に驚きのみを表す
驚きといぶかり	相手の発話内容に対して，驚くとともに怪しむ気持ちを表す
意見	自分の個人的な体験や意見，感情を表す
補完	相手の発話を熱心に聴いていることを示す
あいさつ	相手の存在の承認と好意的に関わろうとする意志を示す
想起	相手の発話内容から，かつての記憶が呼び起こされたことを示す
考え中	相手の発話内容について，なにかを考えている最中だということを表す
その他	上記に該当しない

に 1 度傾聴応答を行うかを表 2.3 の上部に示す．$w_1 \sim w_{11}$ は，11 人の作業者を表す．$w_1 \sim w_{11}$ の発生間隔は，各作業者の傾聴応答の発生間隔の平均値を表し，表 2.1 で示した語りの発話時間を，各作業者の傾聴応答の個数で割ることで算出されている．all の発生間隔は，本研究で収集した全ての傾聴応答の発生間隔の平均値を表し，語りの発話時間の 11 倍を，傾聴応答の総数で割ることで算出されている．収録データでは，全応答データで 2.32 秒に 1 度という高い発生間隔であった．参考データとして，日本語雑談会話を収録した代表的な音声言語資源である名大会話コーパス [70] における，傾聴応答に相当する発話の発生間隔を調査したところ，12.4 秒であった．これらのこ

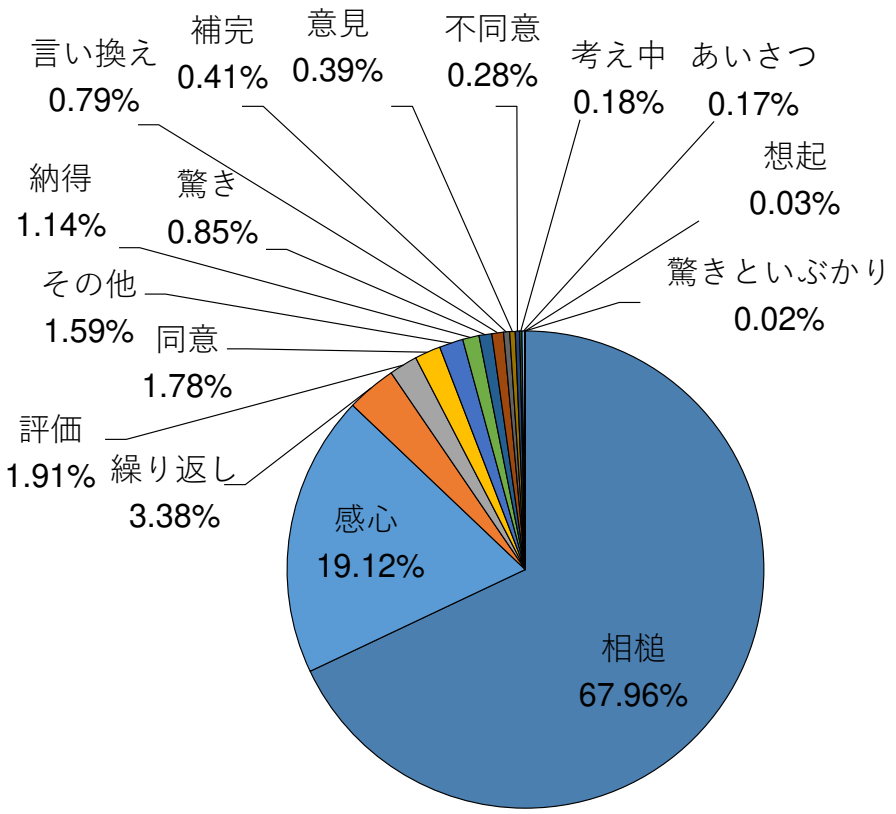


図 2.2: 応答タイプの内訳

とから，収録データにおける傾聴応答の多頻度性を確認した．

2.4.2 応答の多様性

傾聴応答の多様性を評価するために，応答の種類（文字列の異なり）に関する多様度指数を測定した．指数としては，応答あたりのエントロピーを採用した．すなわち，式 (2.1) を測定した．

$$H = -\sum_{i=1}^{S} p_i \log p_i \tag{2.1}$$

ここで S は収録データ内の応答の種類数，p_i は応答 i の出現数が全出現数に占める割合である．応答あたりのエントロピーを表 2.3 の下部に示す．$\mathrm{w}_1 \sim \mathrm{w}_{11}$ は，11 人の作業者を表す．$\mathrm{w}_1 \sim \mathrm{w}_{11}$ のエントロピーは，各作業者が表出した応答あたりのエントロピーである．all のエントロピーは，本研究で収集した全応答データから計算したエントロピーである．収録データでは，全応答データでのエントロピーは 5.11 であっ

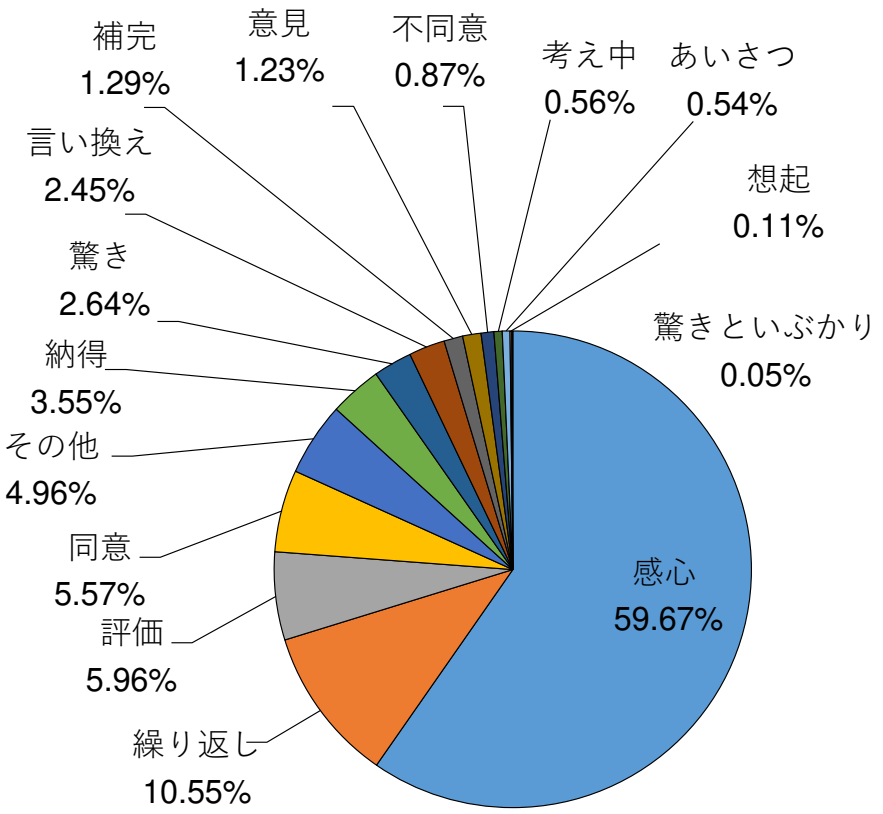


図 2.3: 応答タイプの内訳 (相槌を除く)

表 2.3: 傾聴応答の発生間隔（秒）とエントロピー

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	all
発生間隔	1.85	2.82	2.49	2.26	3.22	2.72	3.30	1.82	1.91	1.60	3.15	2.32
エントロピー	5.19	4.02	3.62	5.30	3.53	4.87	2.78	4.08	5.25	4.44	5.61	5.11

た．名大会話コーパスに対しても同様に測定したところ，エントロピーは 4.86 であった．これらのことから，収録データにおける応答が十分多様であることを確認した．

2.4.3 応答の網羅性

本データの収録では，語りデータに 11 名の作業者が独立に応答を付与している．収集された傾聴応答データが，傾聴応答の表出タイミングを網羅しているかを評価した．

本データでは，高齢者の語りの音声を聴きリアルタイムで作業者が応答を表出しているため，同一の語りの部分に対して表出された応答であっても，発話開始時間が完

全に一致することはない．本研究では，文節境界を応答開始可能なタイミングの候補とし，実際に表出された応答を，語りと応答の各発話時間を利用して対応付けた．具体的には，応答をその発話開始時刻と最も近い文節の終端境界と対応付けた．

語りの文節への分割は，以下の手順で実施した．

(1) 語りの形態素列から，フィラー，感情表出系感動詞，言い直し，言い淀みなどに対応する形態素を取り除いた．

(2) CaboCha[71] を用いて (1) で得られた形態素列を文節列に分割した．

(3) (2) の文節列に (1) で取り除いた形態素を挿入した． 挿入先が文節境界である形態素はそれを 1 つの新しい文節とした．ただし，同じ文節境界に同種の形態素が連続して挿入される場合はまとめて 1 つの文節とした．

以降では，文節境界を表出タイミング候補と呼ぶ．

語りと応答の対応付けの結果，語りデータに含まれる 29,846 個の表出タイミング候補のうち，25,444 個，つまり全体の 85.25%でいずれかの作業者が傾聴応答を表出していた．以降では，いずれかの作業者が傾聴応答を表出していたタイミングの集合を $T(\mathrm{w_{all}})$ と表記する．傾聴応答の表出に適したタイミングを網羅できているかを，以下の手順で評価した．

(1) 11 人の作業者から 1 人を選択し，その作業者の応答表出タイミングを求め，$T(\mathrm{w_{all}})$ に占める割合を算出する．

(2) まだ選択されていない作業者から 1 人を選び，その作業者の応答表出タイミングを求める．

(3) 選択済みの全ての作業者の応答表出タイミングの和集合を計算し，$T(\mathrm{w_{all}})$ に占める割合を算出する．

(4) 11 人全員の作業者が選択されるまで，(2) と (3) を繰り返す．

ただし，11 人の作業者を選択する順番は，全部で $11!(= 39,916,800)$ 通り存在する．そのため，11!通りの選択順について上述の手順を実行し，(1) および (3) で算出される割合の平均値を計算して，網羅性の評価に用いた．

図 2.4 に，網羅性の分析結果を示す．11 人の作業者のうち 7 人の作業者を用いた時点で，$T(\mathrm{w_{all}})$ のうち平均で 90%以上を占めることを確認した．また，10 人の作業者を用いると，平均で 98.32%を占めることを確認した．すなわち，11 人目の作業者を

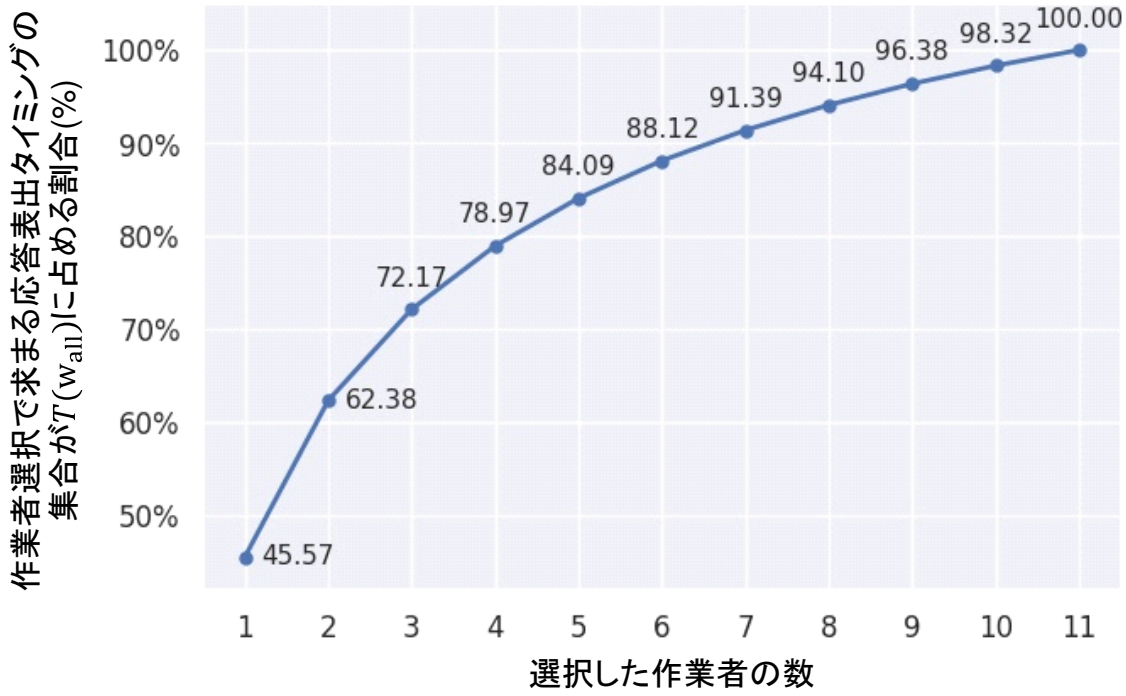


図 2.4: 網羅性の分析結果

追加することで新たに見つかるものは，平均してわずか 1.68%であった．したがって，新たに聴き手を用意して傾聴応答を追加で収録したところで，新たに見つかる傾聴応答の表出に適したタイミングはごくわずかであると考えられる．このことは，本データが傾聴応答の表出に適したタイミングのほとんどを網羅できていることを示しており，本データが高い網羅性を備えていることを確認した．

2.4.4 応答の自然さ

収集した傾聴応答の表現及び表出タイミングの自然さを評価するために，被験者実験を実施した．本研究では，269 個の語り音声に対する 11 名の作業者の応答音声を収集した．すなわち，本研究で収集したデータには，語り音声とそれに対する作業者 1 名による応答音声が合計で 2,959 $(= 269 \times 11)$ 個含まれている．被験者実験では，このうち，ランダムに選択した 53 個のステレオ音声を使用した．このステレオ音声には，全部で 2,191 個の傾聴応答が含まれていた．被験者は，20 歳代の学生計 5 名であり，各応答に対し自然か否かを判定した．その結果，自然でないと判定された応答は，被験者平均で 47.60 個であり，全体の 97.75%を自然な応答が占めた．このことから，

収集された傾聴応答の自然さを確認した．

2.5　収集した応答データの利用

人間はそれぞれ独自の個性を備えており，コミュニケーションにおける振る舞いに影響を与える．聴き手として語りを聴く際にも，聴き手の個性が応答を表出するか否かの判断に影響する．しかし，語りを傾聴する聴き手としては，聴き手自身の個性に基づくタイミングよりも，標準的なタイミングでの応答表出の方が好ましい．

本節では，収録した傾聴応答データの利用例として，聴き手の個性への依存度が低い，標準的な傾聴応答の生成タイミングの定義とその検出実験について述べる．収録データには，1 つの語りに対して作業者 11 名の傾聴応答が含まれている．作業者 11 名の傾聴応答を利用することで，作業者個人の個性の影響を弱め，標準的な応答生成タイミングを定義できる．

2.5.1　実験設定

本実験では，2.4.3 節で述べた表出タイミング候補を傾聴応答を生成するか否かを決定するタイミングとする．語りにおける表出タイミング候補と傾聴応答の対応付けの結果を利用して，標準的な応答生成タイミング検出のための実験データを作成する．収集した応答データには，聴き手 11 人分の傾聴応答が含まれている．したがって，表出タイミング候補に対応付けられた傾聴応答の表出者数は，0 人から 11 人の 12 通り考えられる．本実験では，表出タイミング候補のうち，N 人以上の聴き手の応答が対応付けられたタイミングを正解の応答生成タイミングとして検出対象とする．N は，聴き手個人の応答表出をどの程度尊重するかを表すパラメータとみなせる．N が小さいほど聴き手 1 人あたりの応答表出の影響が大きくなる．

2.5.2　実験データ

表出タイミング候補における正解の応答生成タイミングは，N の値によって異なる．本実験では，人間の傾聴応答の表出頻度を踏まえて N を設定した．まず，本研究で収集した 11 人の作業者の応答表出割合を調査した．図 2.5 に，表出タイミング候補における傾聴応答が対応付けられたものの割合を，11 人の作業者ごとに示す．$\mathrm{w}_1 \sim \mathrm{w}_{11}$ は，11 人の作業者を表す．11 人の作業者の中で最も低い割合は w_{11} の 28.79%で，最も

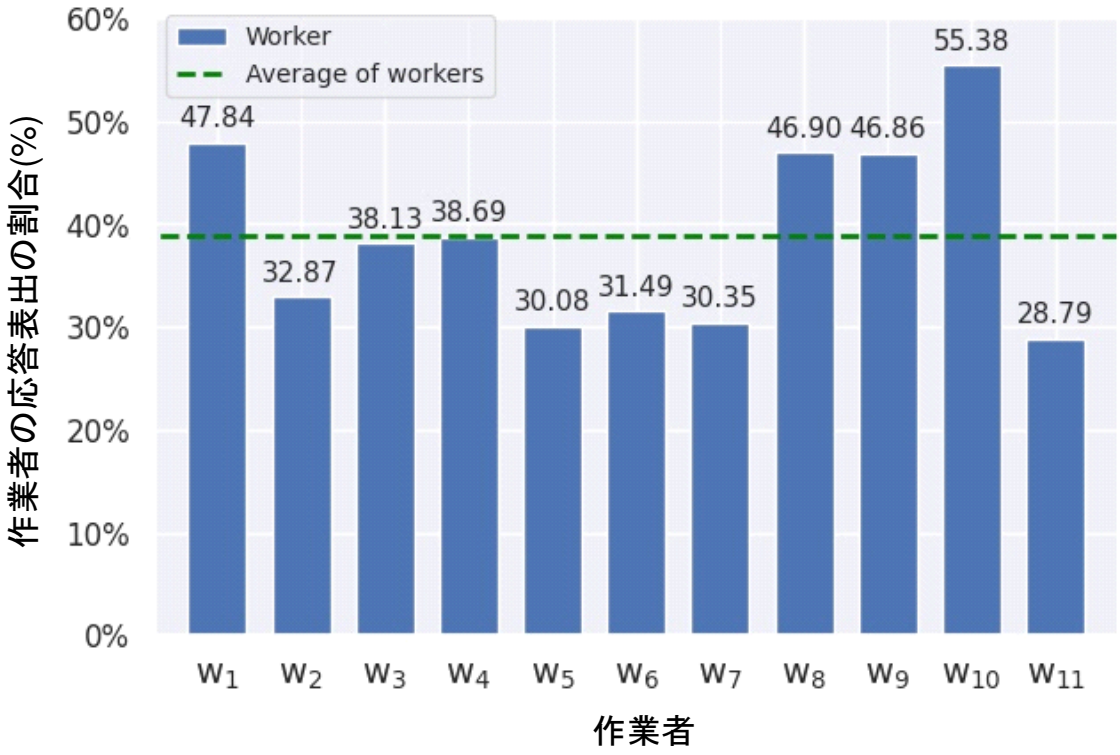


図 2.5: 表出タイミング候補における各作業者の応答表出割合

高い割合は w_{10} の 55.38%であった．また，11 人の作業者間の割合の平均値は，38.85%であった．

次に，表出タイミング候補に占める N ごとの正解の応答生成タイミングの割合を調査した．図 2.6 に，表出タイミング候補に占める N ごとの正解の応答生成タイミングの割合を，上述した作業者間の最低値，平均値，最高値とともに示す．N の値が小さいほど聴き手個人の応答表出を尊重するため，正解の応答生成タイミングが多い．$N \in [4,5,6,7]$ の 4 つについては，正解の応答生成タイミングの割合が作業者間の最低値と最高値の間に位置しており，実際の人間による傾聴応答の表出割合がとりうる値とみなせる．そこで，$N \in [4,5,6,7]$ の 4 つの設定について，応答生成タイミングの検出実験を実施することとした．

本実験データにおける正解の応答生成タイミングは，作業者個人の個性の影響を弱めた，標準的な応答生成タイミングである．このデータを用いてシステムを学習することで，標準的な応答タイミングで応答生成をするシステムの実現を期待できる．また，本実験データは，$N \in [4,5,6,7]$ の 4 種類存在するが，N の値が小さいほど正解の応答生成タイミングが多い．そのため，応答生成頻度に関して，$N = 4$ のデータで

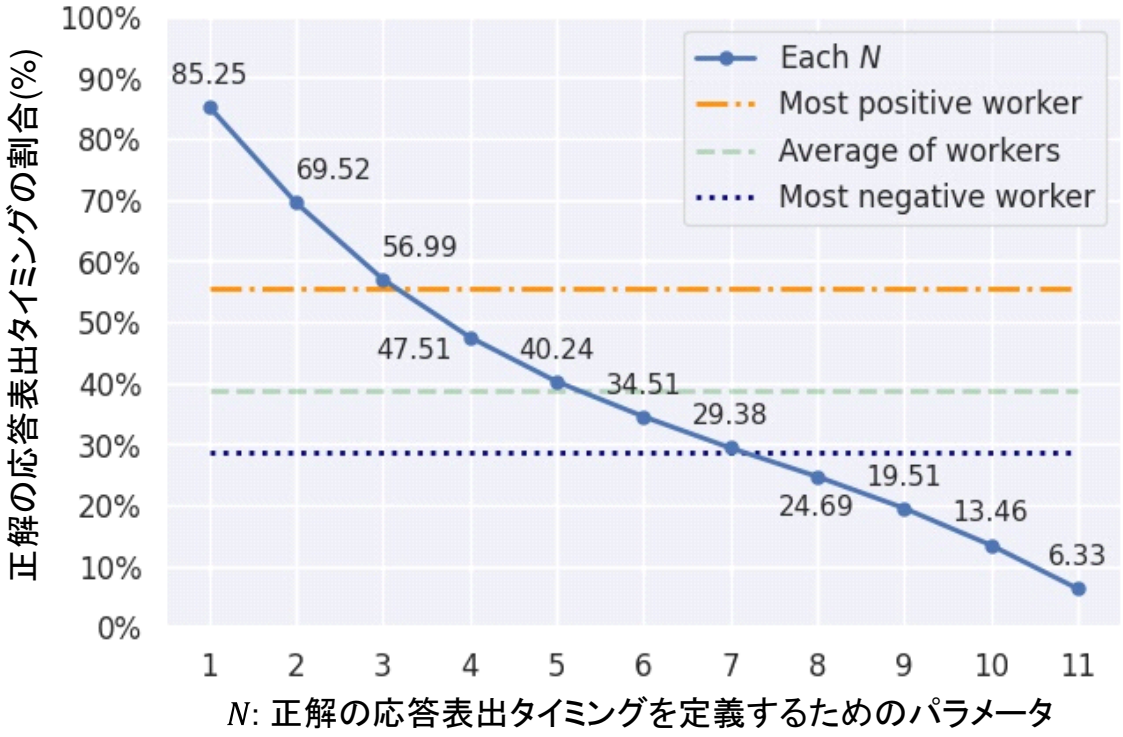


図 2.6: 表出タイミング候補に占める正解の応答生成タイミングの割合

学習されるシステムが最も積極的，$N = 7$ のデータで学習されるシステムが最も控えめとなるものと想定される．本実験データの作成は，本研究で同一の語りに対する複数の作業者の傾聴応答を収録しているため，実現可能となっている．

2.5.3 応答生成タイミングの検出

表出タイミング候補から応答生成タイミングを検出する手法の概略を説明する．表出タイミング候補直前の語りに含まれる文字列から，対象の表出タイミング候補が生成タイミングか否かの 2 クラス分類問題を解くことで，表出タイミング候補から生成タイミングを検出する．本実験では，事前学習済みの Transformer [72] ベースのモデルを fine-tuning する形で，この 2 クラス分類問題を解くためのモデルを構築する．事前学習済みの Transformer ベースのモデルとしては，huggingface[4]に公開されている

[4] https://huggingface.co/models

表 2.4: 各設定における各検出手法の検出性能

	$N=4$				$N=5$			
	(lr)	Pre.	Rec.	F1	(lr)	Pre.	Rec.	F1
random		0.460	0.498	0.478		0.395	0.431	0.412
BERT	(1e-6)	0.664	0.738	0.699	(1e-6)	0.657	0.738	0.695
RoBERTa	(5e-5)	0.679	0.707	0.692	(1e-6)	0.684	0.705	0.694
GPT2	(1e-6)	**0.688**	0.725	**0.706**	(1e-6)	**0.688**	0.707	**0.697**
DeBERTa-v2	(1e-5)	0.651	**0.757**	0.700	(1e-6)	0.650	**0.739**	0.691
参考値		0.792	0.792	0.792		0.825	0.800	0.812

	$N=6$				$N=7$			
	(lr)	Pre.	Rec.	F1	(lr)	Pre.	Rec.	F1
random		0.319	0.354	0.336		0.277	0.303	0.290
BERT	(1e-6)	0.639	**0.751**	0.690	(1e-6)	0.604	**0.740**	0.665
RoBERTa	(1e-6)	0.676	0.706	0.691	(1e-6)	0.639	0.701	**0.669**
GPT2	(1e-6)	**0.681**	0.716	**0.698**	(1e-6)	**0.649**	0.684	0.666
DeBERTa-v2	(1e-6)	0.639	0.739	0.685	(1e-6)	0.607	0.730	0.663
参考値		0.746	0.847	0.794		0.733	0.812	0.770

BERT [53] [5], RoBERTa [73] [6], GPT2 [62] [7], DeBERTa-v2 [74] [8]を採用した．いずれも，隠れ層のサイズが 768 次元，レイヤー数が 12 のモデルである．これらの事前学習済みモデルに，対象の表出タイミング候補が生成タイミングか否かという 2 クラス分類問題を解くための出力層を追加し，fine-tuning を実施した．

2.5.4 検出モデルの学習とテスト

生成タイミング検出モデルの学習について説明する．応答生成タイミングの検出手法の実装と評価のために，表出タイミング候補 29,846 個を，学習データ，開発データ，テストデータに分割した．学習データには 17,417 個，開発データには 6,270 個，テス

[5] cl-tohoku/bert-base-japanese-whole-word-masking
[6] nlp-waseda/roberta-base-japanese
[7] rinna/japanese-gpt2-small
[8] ku-nlp/deberta-v2-base-japanese

トデータには6,159個の表出タイミング候補が含まれている．

本実験では，表出タイミング候補直前の5個の文節に含まれる語りの文字列を入力とした．モデルの学習における損失関数はCross Entropy Lossとして，バッチサイズを32, エポック数を50とした．モデルの最適化手法には，weight decayを0.01としたAdamW [75] を用いた．学習率には，warmup ratioを10%とした，線形スケジューリングを採用した．すなわち，学習開始時から学習全体の10%が終了するまでの間は，事前に定めた学習率の最大値まで線形に学習率を増加させ，それ以降は線形に学習率を減少させた．本実験では，各Nにおける各モデルの学習率の最大値は，1e-6, 5e-6, 1e-5, 5e-5の中から，学習終了時のモデルの開発データにおけるF1値が最良となったものとした．ベースラインとして，学習データ内の表出タイミング候補における正解の生成タイミングの割合にしたがって，ランダムに生成タイミングを検出する手法を実装した．

本実験では，各表出タイミング候補について，応答生成タイミングか否かを判定する．評価には，テストデータにおける以下の式で計算される適合率と再現率，及び，その調和平均であるF値を用いた．

$$\text{適合率} = \frac{\text{検出された正解の生成タイミング数}}{\text{検出された生成タイミング数}} \tag{2.2}$$

$$\text{再現率} = \frac{\text{検出された正解の生成タイミング数}}{\text{正解の生成タイミング数}} \tag{2.3}$$

2.5.5 実験結果

表2.4に，各Nにおける各検出手法のテストデータに対する検出性能を示す．Pre.の列に適合率，Rec.の列に再現率，F1の列にF値を示す．また，randomを除く4つの手法については，その学習率の最大値を(lr)の列に示す．参考値として，11人の聴き手の中から順に1名を取り出し，その1名の聴き手の表出タイミングを検出タイミングとみなしたときの，各Nの正解の生成タイミングに対する適合率，再現率，F値を計算した．その結果，$N = 4$ではw_1を，$N = 5, 6$ではw_3を，$N = 7$ではw_2を取り出したときに，F値が最良となった．表2.4の最終行に，各NにおけるF値が最良と成った作業者の評価値を示す．

Transformerベースの4つの手法の適合率，再現率，F値は，いずれのNにおいても，0.60∼0.70程度であった．この値は，randomを大きく上回っており，その有効性を確認できた．これら4つの手法の評価値は同程度であるが，GPT2の値が比較的高い傾向にあった．また，いずれの手法も共通して，$N = 7$におけるF値が，他の3つ

の N よりも低かった．

本実験では，語りの文字列のみから，傾聴応答の生成タイミングの検出に取り組んだが，ピッチやパワーなどの語りの音響情報も有効であると考えられる．また，実際に傾聴応答を生成する際には，過去にいつ，どのような応答を生成したのかという履歴も考慮する必要がある．例えば，直前の傾聴応答の生成からの時間経過なども考慮されるべきである．本実験では音響情報や応答履歴を利用しなかったが，本研究で収集した応答データを用いれば，これらを踏まえた応答生成タイミングの検出に取り組むことが可能であり，今後の課題である．

2.6 おわりに

本章では，語りに対する傾聴応答の収集について述べた．あらかじめ収録された語りの音声に対し，表出するに相応しい傾聴応答の表現と表出タイミングを付与するという収集方式を採用した．本方式により自然で多様な応答を大量かつ網羅的に収集できることを確認した．最後に，収集した傾聴応答を利用した，応答生成タイミングの検出実験について述べた．本実験では，語りに含まれる文字列から応答タイミングを検出した．

第3章　語りに傾聴を示す応答の表出されやすさの推定

3.1　はじめに

社会の個人化が深刻化し，聴き手不在の生活シーンが増加している現代社会では，語りを傾聴する会話エージェントの実現は，人の語る機会を増やすための解決策の1つである．会話エージェントが人間の代わりに聴き手を担うには,「語りを傾聴していることを語り手に伝達する機能」を備える必要がある．傾聴していることを示す目的で語りに応答する発話である傾聴応答の表出は，傾聴態度を伝達するための有力な方法の1つである．傾聴応答を適切なタイミングで表出できれば，語り手の語る意欲を促進するなどの効果が期待できる．しかし，表出タイミングが不適切であれば逆効果となる．

本章では，適切なタイミングでの傾聴応答の自動生成に向けて，傾聴応答の表出タイミングとしての適切さの度合いを推定する手法を提案する．具体的には，この適切さを表す指標として，所与のタイミングにおいて傾聴応答を表出する聴き手の割合である**表出率**を導入し，その推定を行う．表出率は，語りを傾聴する会話エージェントが傾聴応答を表出するか否かを判断するための材料として利用できる．例えば，推定された表出率をもとに，語り手の嗜好や会話エージェントの個性などを加味して，表出するか否かを柔軟に決定することが考えられる．

本手法は，語りの音響情報と言語情報を用いて，傾聴応答の表出率を推定する．具体的には，これらの情報を Transformer ベースの手法 [72] でエンコードし，このエンコード結果を利用して表出率を計算する．

本手法の性能を評価するために，表出率の推定実験を行った．実験の結果，傾聴応答の表出率の推定において，本手法が高い推定性能を備えていることを確認した．また，語りの音響情報と言語情報を併用することの効果を確認した．

さらに本研究では，表出率の有用性を評価するために，表出率を入力に用いた傾聴応答の表出タイミングの検出実験を行った．実験の結果，表出率を利用することによ

り，検出性能が向上することを確認した．

本章の構成を以下に示す．3.2 節では傾聴応答の表出率について説明する．次に，3.3 節で傾聴応答の表出率の推定手法を提案し，3.4 節で提案手法を用いた表出率の推定実験について述べる．3.5 節では，傾聴応答の表出タイミングの検出を例に，表出率の有用性を検証する．最後に，3.6 節で本章のまとめを行う．

3.2 傾聴応答の表出率

本研究では，あるタイミングが傾聴応答の生成にどの程度適しているかを表す指標として，傾聴応答の表出率を定義する．あるタイミングでの傾聴応答の表出率は，そのタイミングで傾聴応答を表出する聴き手の割合であるとする．表出率が高いタイミングは，そのタイミングで傾聴応答を表出する聴き手が多いことを意味する．

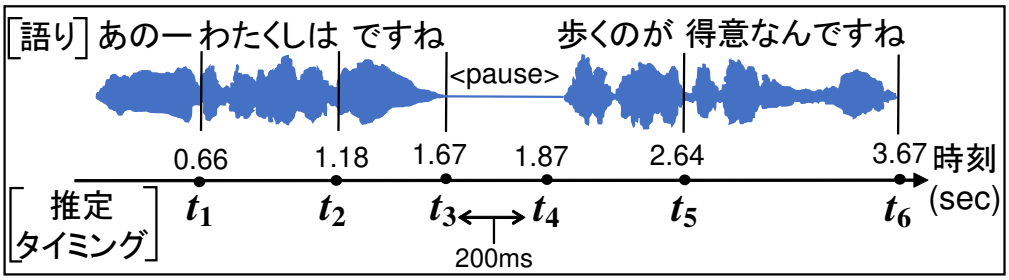
傾聴応答の表出タイミングの検出に関する従来研究 [30, 31, 32, 33, 34] では，語りから抽出されるピッチやパワーなどの音響情報が，表出タイミングの検出に有効であると報告している．また，語りに含まれる単語や文構造などの言語情報も，その有効性が示されている．これらの手法はいずれも，与えられたタイミングを傾聴応答の生成に適するか否かに分類するものであり，その結果に従って傾聴応答の生成を決定することを想定している．

一方，本章では，傾聴応答の表出率を推定する手法を提案する．表出率の推定は，傾聴応答の生成に関わる判断材料を提供することを目的とする．例えば，推定された表出率をもとに会話エージェントが自らの個性（積極的／控えめ，など）に応じた応答を生成することができる．また，表出率に加え，語り手の嗜好（密な応答を好む／好まない，など）も併せて考慮して，傾聴応答に関わる振る舞いを決定することも可能となる．それ以外にも，複数の会話エージェントが語りの聴き手を担う状況では，表出率に従って応答するエージェントを制御することが考えられる．

3.3 表出率の推定手法

3.3.1 推定タイミング

傾聴応答の代表である相槌が，文節境界，あるいは，ポーズで表出されやすいという知見 [40] に基づき，提案手法では，傾聴応答の表出率を推定するタイミング (以下，推定タイミング) は以下のとおりとする．



推定タイミング	語りのサブワードと推定タイミングを表すタグによって構成されるトークン系列
t_3	… あ の ー <T_a> わたくし は <T_a> です ね <E_a>
t_4	… あ の ー <T_a> わたくし は <T_a> です ね <T_a> <E_b>

図 3.1: 推定タイミング (上部) と言語情報 (下部) の例

(a) 語りの文節終端

(b) 語りの文節終端から 200ms ポーズが継続した点

ここで，文節終端とは，文節の最終形態素の発声の終了時を意味する．図 3.1 の上部に，推定タイミングの例を示す．$t_1 \sim t_6$ が推定タイミングであり，t_4 は上記の (b) に該当し，それ以外は (a) に該当する．

3.3.2 特徴量

傾聴応答の表出タイミングの検出に関する従来研究 [30, 31, 32, 33, 34] では，語りの音響情報と言語情報の有効性が示されており，これらは傾聴応答の表出率の推定にも効果を示す可能性がある．そこで，提案手法では，語りの音響情報と言語情報から，傾聴応答の表出率を推定する．音響情報としては，対象の推定タイミング直前のフレームの系列を用いる．フレーム単位の特徴量には，MFCC の下位 12 次元，ピッチ，パワー，及びそれらの Δ と $\Delta\Delta$ を用いる．

言語情報としては，対象の推定タイミング直前のトークンの系列を用いる．トークンは，語りのサブワードと推定タイミングを表現する 4 種のタグから構成される．表 3.1 に，4 種のタグを示す．トークン単位の特徴量には，トークンの埋め込み表現を用いる．言語情報の例を図 3.1 の下部に示す．

表 3.1: 推定タイミングを表現するタグ

$< \mathbf{E_a} >$	文節終端の対象の推定タイミング
$< \mathbf{E_b} >$	ポーズ継続地点の対象の推定タイミング
$< \mathbf{T_a} >$	文節終端の過去の推定タイミング
$< \mathbf{T_b} >$	ポーズ継続地点の過去の推定タイミング

3.3.3 表出率推定モデル

提案手法では，対象の推定タイミング直前のフレームとトークンの系列を入力として，表出率を推定する．このため，提案手法における表出率の推定モデル（以下，提案モデル）は，系列データの処理に適している必要がある．そこで提案モデルでは，機械翻訳 [72]，音声認識 [76]，マルチモーダル情報のエンコード [77] など，幅広いタスクでその有効性が示されている Transformer ベースのエンコード手法 [72] を採用する．

提案モデルの概略を図 3.2 に示す．ここで，$f_i \in [f_1, f_2, \cdots]$ はフレームを，$tok_j \in [tok_1, tok_2, \cdots]$ はトークンを表す．提案モデルは，対象の推定タイミング直前のフレーム系列とトークン系列を入力として，以下の手順で表出率を推定する．

(1) 音声分析ツールとトークン埋め込み層を用いて特徴量行列を抽出．

(2) positional encoding を加算．

(3) 線形変換と ReLU 関数を適用．

(4) Transformer encoder [72] で変換．

(5) 変換後の行列に対して要素ごとに平均値を求めて，音響情報と言語情報のベクトルを算出．

(6) 2 つのベクトルを連結．

(7) 線形変換と sigmoid 関数を適用．

以上により，0 以上 1 以下の 1 次元の値として，表出率が出力される．

3.4 実験

提案手法の推定性能を評価するために実験を行った．

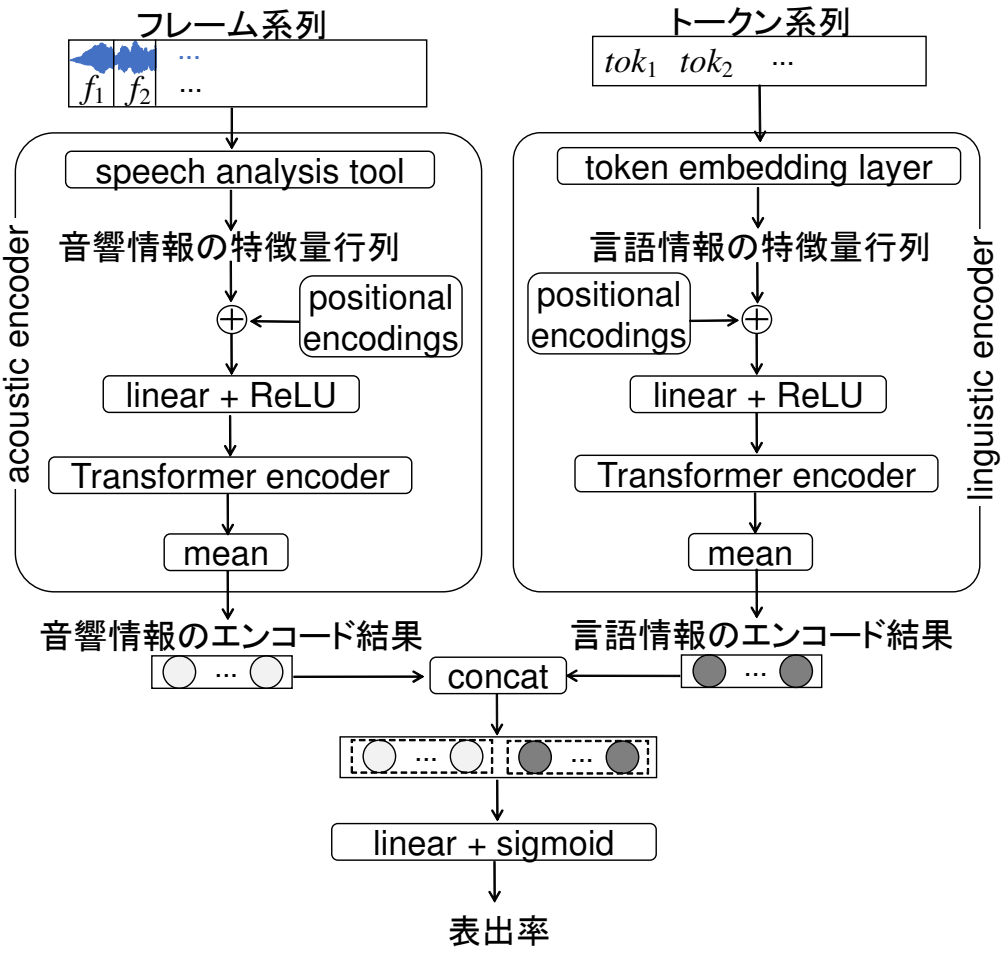


図 3.2: 提案モデルの概略

3.4.1 実験データ

本実験では，2 章で構築した傾聴応答コーパスの一部である 10 人の聴き手による傾聴応答 131,616 個を用いた[1]．各推定タイミングに対する傾聴応答の表出率の正解値は，この 10 人のうちの応答した聴き手の割合とした．まず，CaboCha [71] を用いて語りの文節境界を検出した後，推定タイミングを特定した．次に，応答の発話開始時刻を，語りにおける最も近い推定タイミングに対応付けた．最後に，対応付けの結果に基づき，傾聴応答の表出率の正解値を算出した．

表 3.2 に傾聴応答の表出率の正解値の算出例を示す．この例における語りと推定タ

[1]傾聴応答コーパスには 11 人の聴き手の傾聴応答が収録されているが，3.5 節で後述する表出率を用いた応答表出タイミングの検出実験と設定を合わせるために，本実験では 10 人の聴き手の傾聴応答を用いた．

表 3.2: 傾聴応答の表出率の正解値の算出例

推定タイミング	語り	聴き手の傾聴応答の有無										表出率
		L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	
t_1	あのー											**0.0**
t_2	わたくしは						✓			✓		**0.2**
t_3	ですね	✓								✓		**0.2**
t_4	<pause>							✓	✓		✓	**0.3**
t_5	歩くのが											**0.0**
t_6	得意なんですね	✓		✓		✓				✓	✓	**0.5**

イミングは，図 3.1 のものと同一である．$L_1 \sim L_{10}$ は 10 人の聴き手を表し，✓ は聴き手が応答したことを意味する．例えば，推定タイミング t_2 では，L_6 と L_9 の 2 人が応答しているので，表出率は $2/10 = 0.2$ となる．実験で用いるデータには，40,647 個の推定タイミングが存在した．図 3.3 に，語りに存在する 40,647 個の推定タイミングにおける表出率の正解値の出現割合を示す.

3.4.2 実験概要

実験データを学習データ，開発データ，テストデータに分割した．学習データには 23,846 個，開発データには 8,588 個，テストデータには 8,213 個の推定タイミングが含まれている．

音響の特徴量抽出には Praat [78] を使用した．フレームシフトを 10ms とし，対象の推定タイミングから遡って 25 フレーム（推定タイミングの直前 250ms）の特徴量を用いた．また，MFCC は，窓幅を 25ms とした．言語情報として，対象の推定タイミング直前の 3 文節の語りに含まれるトークン系列を使用した．語りのサブワードへの分割は，MeCab [68] (UniDic Ver. 2.1.2) で形態素解析したのち，subword-nmt [79][2] により実施した．サブワードの埋め込みの初期値には，日本語 Wikipedia コーパスを用いて GloVe [80] により事前学習した埋め込み表現を利用した．トークンに対する埋め込みの次元数は 300 とした．

表出率推定モデルの Transformer encoder については，層数を 1, multi-head atten-

[2] https://github.com/rsennrich/subword-nmt

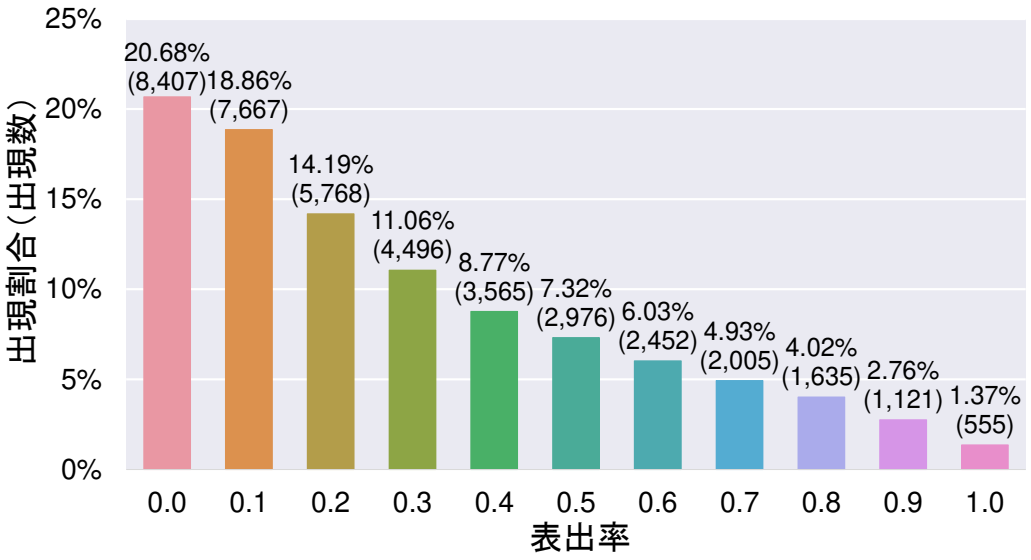


図 3.3: 傾聴応答の表出率の正解値の出現割合

tion のヘッド数を 8, 入出力の次元数 d_{model} の 4 倍を position-wise feedforward network の中間層の次元数とした．モデルの学習では，バッチサイズを 256，学習率を 2e-4 とし，損失関数には MSE (Mean Squared Error) Loss を，最適化手法には Adam [81] を用いた．

音響情報と言語情報のエンコーダにおける d_{model} については，それぞれ，64, 128, 256, 512, 1024 の 5 通りの中から，開発データを用いて定めた．5×5 の 25 通りのモデルをそれぞれ 100 エポック学習したところ，音響情報のエンコーダの d_{model} を 256, 言語情報のエンコーダの d_{model} を 512 として，18 エポック学習させた時点での開発データの損失が最小となったため，このモデルをテストデータに適用した[3]．

3.4.3 評価方法

本実験では，テストデータ中の全推定タイミングから算出される平均絶対誤差 MAE (Mean Absolute Error) を主要な評価指標とし，これを MAE (all) と表記する．さらに，より詳細な評価のために，表出率の正解値ごとの MAE も算出する．表出率の正解値 c に対する MAE を MAE (c) と表記する．具体的には，MAE (all) と MAE (c) を

[3]音響情報あるいは言語情報のエンコーダの d_{model} のうち，少なくとも一方が 1024 である大規模なモデルでは，100 エポックの学習の過程において，学習データと開発データ共に損失の減少を確認できなかったものも存在した．

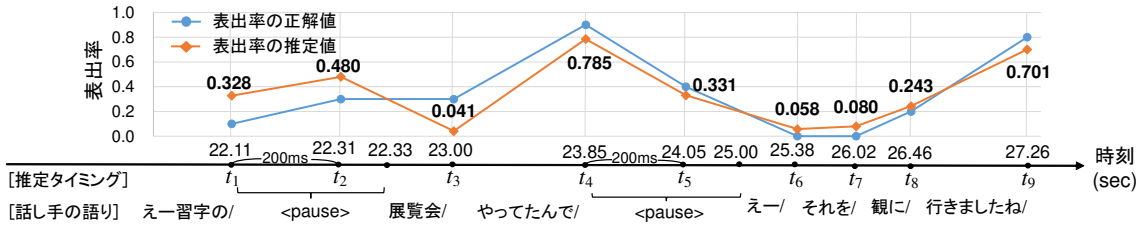


図 3.4: 提案手法による傾聴応答の表出率の推定例

以下の式に従って計算する．

$$\mathrm{MAE(all)} = \frac{1}{|S_{\mathrm{all}}|} \sum_{t \in S_{\mathrm{all}}} |p(t) - c(t)| \tag{3.1}$$

$$\mathrm{MAE}(c) = \frac{1}{|S_c|} \sum_{t \in S_c} |p(t) - c| \tag{3.2}$$

ただし，$p(t)$ と $c(t)$ はそれぞれ，推定タイミング t における表出率の推定値と正解値を，S_{all} はテストデータにおける全推定タイミングを，S_c はテストデータにおける表出率の正解値が c であるような推定タイミングの集合を表す．また，MAE (all) に加え，表出率の推定値と正解値におけるスピアマンの順位相関係数も主要な評価指標とする．

比較のため，提案手法に加え，以下の 3 つの手法を実装した．

- **random**: 学習データにおける表出率の出現分布に従ってランダムに表出率を推定する手法
- **proposed (A)**: 音響 (Acoustic) 情報のみを使用して表出率を推定する手法
- **proposed (L)**: 言語 (Linguistic) 情報のみを使用して表出率を推定する手法

ハイパーパラメータである Transformer encoder の入出力の次元数 d_{model} については，proposed (A) では 64，proposed (L) では 512 とした．これらの値は，提案モデルと同様に，開発データを用いて定めた．

3.4.4 実験結果

各推定手法に対する評価指標の値を表 3.3 に，提案手法による表出率の推定例を図 3.4 に示す．表 3.3 における MAE (c) の 1 段下の 0.0 ～ 1.0 の 0.1 刻みの数字は，MAE

表 3.3: 各推定手法における評価指標の値

	MAE(c)											MAE (all)	Spearman
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0		
random	0.297	0.237	0.215	0.223	0.242	0.296	0.341	0.418	0.529	0.598	0.689	0.291	0.017
proposed (A)	0.201	0.133	**0.096**	**0.098**	0.146	0.191	0.249	0.306	0.376	0.440	0.561	0.182	0.454
proposed (L)	0.175	0.123	0.109	0.112	**0.136**	**0.172**	0.219	0.254	0.294	0.356	0.442	0.165	0.582
proposed	**0.132**	**0.108**	0.115	0.129	0.157	0.173	**0.202**	**0.232**	**0.264**	**0.309**	**0.375**	**0.151**	**0.637**

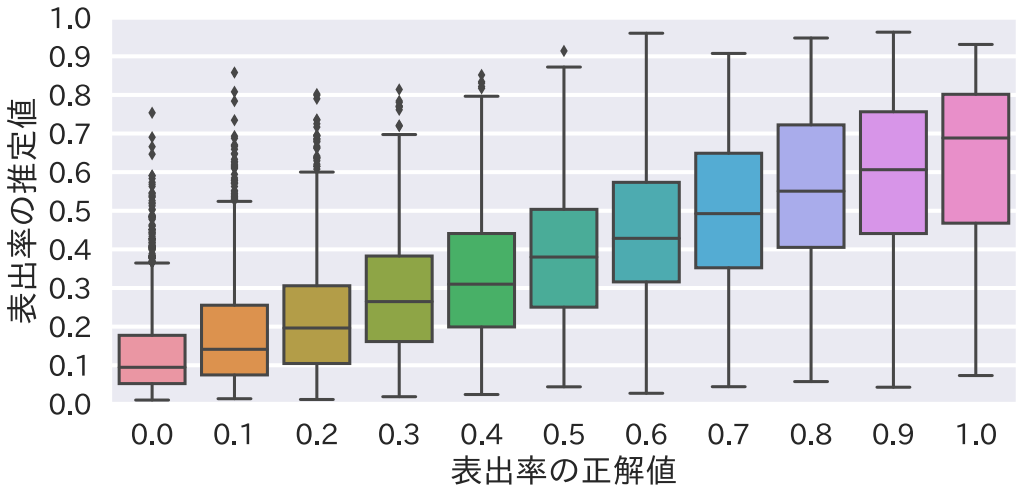


図 3.5: 提案手法による傾聴応答の表出率の推定傾向

(c) の c の値である．表出率の推定値と正解値との MAE について述べる．提案手法は，MAE (all) と 11 個の MAE (c) の計 12 個全てで，random を上回っており，提案手法による推定の有効性を確認した．proposed (A) と proposed (L) についても，12 個全ての MAE で random を上回っており，音響情報と言語情報がいずれも表出率の推定に寄与することが示された．一方，提案手法と proposed (A) 及び proposed (L) との比較では，提案手法が MAE (all) を含む 8 個の MAE で最良の結果を示しており，音響情報と言語情報の両方を使用して表出率を推定することの効果を確認した．また，提案手法とその他 3 つの各手法との間で，Wilcoxon の符号順位検定を用いて，表出率の推定値と正解値との絶対誤差の有意差を検定した．その結果，提案手法は，各手法を有意に上回っていることが認められた（$p < 0.05$）．

表出率の推定傾向について述べる．図 3.5 に提案手法の推定傾向を示す．図 3.5 は，正解の表出率が高いほど提案手法の推定値も高くなることを示している．また，proposed (A) 及び proposed (L) についても，同様の傾向が見られた．この推定傾向は，表 3.3 のスピアマンの順位相関係数の傾向とも合致しており，random を除いた 3 つの推定手法では，正の相関が示された．特に，提案手法が最も高い相関を示しており，音響情報と言語情報の両方を推定に用いることの有効性を確認した．また，提案手法とその他 3 つの各手法との間で，Williams 検定 [82, 83] を用いて，スピアマンの順位相関係数の有意差を検定した．その結果，提案手法は，各手法を有意に上回っていることが認められた（$p < 0.05$）．

3.5　傾聴応答の表出率の利用

表出率の推定は，傾聴応答の生成に関わる判断材料を提供することを目的とする．その一例として，推定された表出率をもとに，語りの聴き手としての会話エージェントが自らの特性に応じて，応答を生成するか否かを決定することが考えられる．表出率は，聴き手全体の応答傾向を表すとみなせるため，それを用いることで，傾聴応答の表出タイミングに関する普遍的な適切さを考慮することが可能となる．そこで本節では，表出率の有用性について考察するために，表出率を用いて傾聴応答の生成タイミングを検出する実験を行った．

3.5.1　実験設定

本実験の目的は，表出タイミングの検出における表出率の有用性を評価することである．本実験における表出タイミングとは，聴き手による応答が対応付けられた推定タイミングを指す．本実験では，聴き手ごとにその表出タイミングの検出を実施する．聴き手 L_1 の表出タイミングの検出を例に，本実験の概略を図 3.6 に示す．本実験では，図 3.6 に示されるような表出タイミングの検出を，聴き手 $L_1 \sim L_{11}$ の 11 人について実施する．以降の **(i)**∼**(iv)** で，本実験の詳細について説明する．

(i)　実験データ

本実験では，3.4 節の実験と同じく，語りに対して複数の聴き手による傾聴応答が付与されている傾聴応答コーパスを実験データとして用いた．3.4 節の実験では，傾聴応答コーパスの一部である，10 人の聴き手による傾聴応答が付与された語りのデータを用いており，語りには 40,647 個の推定タイミングが存在していた．本実験では，同一の語りに対する 10 人の聴き手とは別の 1 人の聴き手の傾聴応答を加えた，合計 11 人の聴き手による傾聴応答が付与された語りのデータを用いた．本実験で新たに使用する聴き手 1 人の傾聴応答についても，他の 10 人の聴き手の傾聴応答と同様に，3.4.1 節で述べた傾聴応答と推定タイミングの対応付けの処理を行った．11 人の聴き手の傾聴応答が付与された語りのデータである本節の実験データを，3.4 節の実験と同様の分割で，学習データ，開発データ，テストデータに分けた．すなわち，学習データには 23,846 個，開発データには 8,588 個，テストデータには 8,213 個の推定タイミングが含まれている．本実験では，11 人の聴き手 1 人 1 人について，その表出タイミングの検出を実施する．

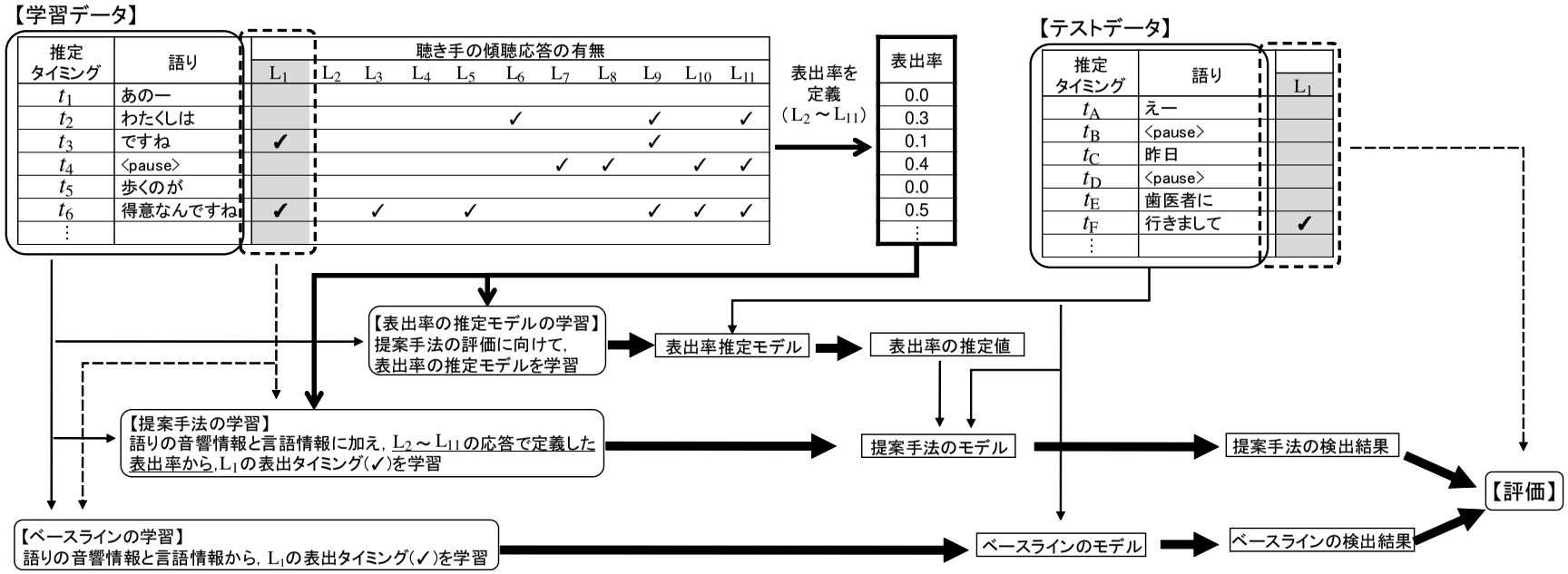


図 3.6: 聴き手 L_1 の表出タイミング検出実験の概要

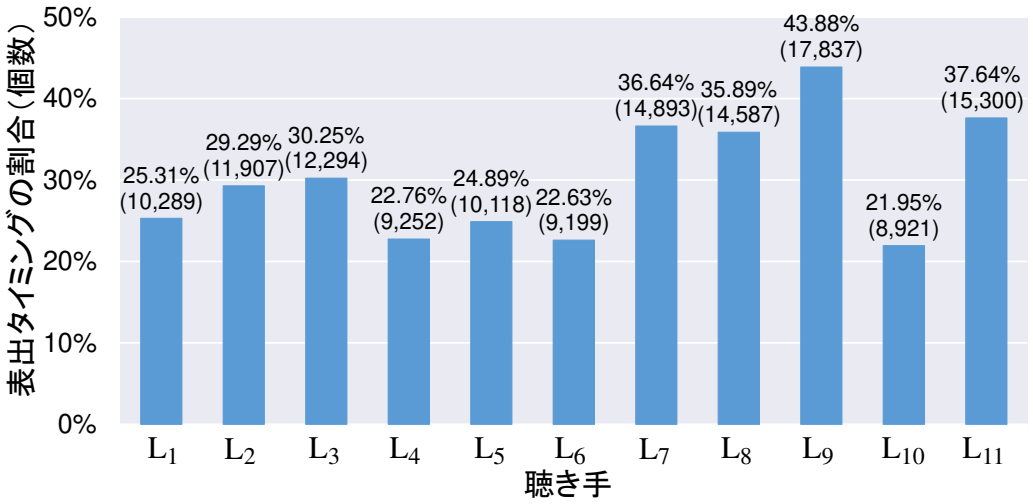


図 3.7: 推定タイミングにおける表出タイミングの割合

図 3.7 に，実験データに含まれる 40,647 個の推定タイミングにおける，11 人の聴き手それぞれの表出タイミングの割合を示す．$L_1 \sim L_{11}$ は，11 人の聴き手を表し，() 内の数字は，各聴き手の表出タイミング数を表す．推定タイミングに占める表出タイミングの割合は，11 人の聴き手の平均で，30.10%であった．

(ii)　表出タイミングの検出手法

図 3.6 に示す，語りから得られる音響情報と言語情報のみから，表出タイミングを検出する手法をベースラインとし，表出率も利用して表出タイミングを検出する手法を提案手法とする．図 3.8 に，これらの検出手法の概略を示す．ベースラインでは，3.3.2 節と同じ特徴量を入力とし，3.3.3 節と同じモデルで入力情報を処理する．ただし，出力は表出率ではなく，対象の推定タイミングが表出タイミングであるかに関する尤度である．提案手法では，表出率も入力として利用する．具体的には，音響情報と言語情報のエンコード結果の連結時に，表出率も併せて連結する．なお，提案手法で利用する表出率については，モデルの学習時には，学習データから算出された値を用いる．一方，テスト時には，3.3.3 節で述べた表出率の推定モデルの出力値を用いる．

(iii)　検出モデルの学習とテスト

ベースラインと提案手法のモデルの学習は，3.4.2 節と同様の設定で実施した．ただし，損失関数については，BCE (Binary Cross Entropy) Loss を用いた．いずれの

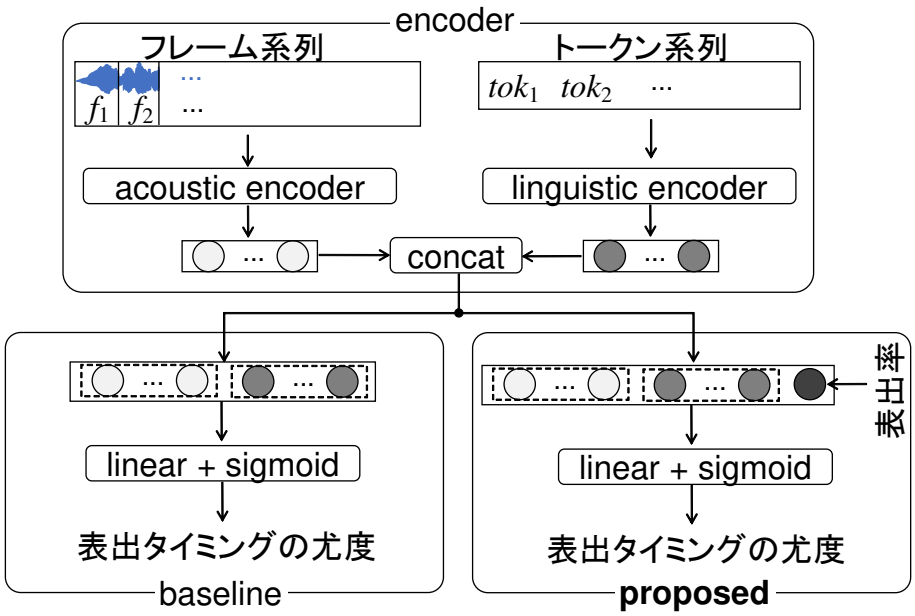


図 3.8: 表出タイミング検出モデルの概略

モデルの出力も，対象の推定タイミングが表出タイミングであるかに関する 0 以上 1 以下の尤度である．対象の推定タイミングにおける尤度が閾値以上であれば表出タイミングであると判定した．本実験では，開発データを用いて，本節の (iv) で述べる検出性能の評価指標が最良となるように，この閾値を決定した．

学習において推定タイミングの損失を計算する際の正解値は，評価対象の聴き手がそのタイミングで傾聴応答を表出しているか否かの 2 値ラベルである．ベースラインと提案手法のいずれのモデルも，学習において評価対象の聴き手のデータを利用する．さらに，図 3.6 に示す通り，提案手法のモデルの学習には，評価対象ではない 10 人の聴き手から定義される表出率の正解値も利用する．すなわち，提案手法の学習には，評価対象の聴き手 1 人と，それ以外の 10 人の聴き手，計 11 人の聴き手のデータを利用する．

なお，検出対象の聴き手以外の 10 人の聴き手の応答から定めた表出率の正解値を用いて，提案手法のモデルの学習と閾値の決定を行った．一方，テスト時には，表出率の推定値を用いて検出を行い，その性能を評価した．表出率の推定値は，3.4 節と同様の方法で学習された，3.3.3 節で述べた表出率の推定モデルの出力値である．以降では，この提案手法を **proposed (E)** と表記する．

表 3.4: 各検出手法における F 値

	baseline	proposed (E)	参考値
L_1	0.527	0.529 †	0.536 †
L_2	0.584	0.582	0.590 †
L_3	0.563	0.569 †	0.573 †
L_4	0.559	0.560 †	0.565 †
L_5	0.429	0.450 †	0.451 †
L_6	0.563	0.565 †	0.567 †
L_7	0.590	0.594 †	0.598 †
L_8	0.595	0.602 †	0.606 †
L_9	0.626	0.629 †	0.631 †
L_{10}	0.495	0.501 †	0.508 †
L_{11}	0.592	0.594 †	0.600 †

†: better than baseline

(iv) 評価方法

本実験では，傾聴応答の表出タイミングの検出性能を，正解の表出タイミングに対する適合率と再現率の調和平均である F 値で評価した．適合率と再現率は，それぞれ以下の式で計算される．

$$\text{適合率} = \frac{\text{正解の表出タイミングの検出数}}{\text{表出タイミングの検出数}} \tag{3.3}$$

$$\text{再現率} = \frac{\text{正解の表出タイミングの検出数}}{\text{正解の表出タイミング数}} \tag{3.4}$$

3.5.2 実験結果

表 3.4 に，表出タイミングの検出結果として，各検出手法ごとに，11 人の聴き手 $L_1 \sim L_{11}$ に対する F 値を示す．参考値として，テスト時に表出率の推定値ではなく，正解値を用いた際の F 値も示す．この参考値は，表出率の推定性能が 100%であるという理想的な状況における proposed (E) の検出性能である．proposed (E) の F 値は，聴き手 L_2 を除いたすべての聴き手において，ベースラインを上回った．なお，参考

表 3.5: ベースラインと proposed (E) による検出例

推定タイミング	語り	表出率の正解値	表出率の推定値	表出タイミング baseline	表出タイミング proposed (E)
t_{a}	3 時間ぐらい	0.2	0.353		
t_{b}	<pause>	0.6	0.474	✓	✓
t_{c}	えー	0.0	0.031		
t_{d}	<pause>	0.1	0.075		
t_{e}	何	0.1	0.016		
t_{f}	しゃべったんだか	0.2	0.314	✓	

値の F 値は，すべての聴き手において，ベースラインと proposed (E) の両方を上回った．proposed (E) とベースラインの 2 手法について，テストデータ中のすべての検出結果を対象にマクネマー検定を実施したところ，有意差は認められなかった．これらのことから，表出率の利用により表出タイミングの検出性能の向上の可能性が示唆されたものの，その効果には限界があるものと考えられる．

表 3.5 に，ベースラインと proposed (E) の検出例を示す．この例は，聴き手 L_1 が評価対象の検出例である．推定タイミング t_{b} は，評価対象の聴き手の応答表出タイミングであり，ベースラインでは検出できなかったが，proposed (E) では検出に成功した．t_{b} における表出率は比較的高く，このことを踏まえて proposed (E) は，t_{b} が評価対象の聴き手の応答表出タイミングであると判断できたものと考えられる．また，推定タイミング t_{f} について，ベースラインは評価対象の聴き手の応答表出タイミングであると誤検出していた．一方で，proposed (E) は，応答表出タイミングではないと正しく判定できていた．t_{f} における表出率は比較的低く，このことを踏まえて proposed (E) は，t_{f} が評価対象の聴き手の応答表出タイミングでないと判断できたものと考えられる．

3.6 おわりに

本章では，傾聴応答の表出率の推定手法を提案した．本手法は，Transformer ベースの手法でエンコードされた語りの音響情報と言語情報を用いて，表出率を算出する．本手法の推定性能を評価するために，表出率の推定実験を行った．実験には，10

人の聴き手の応答 131,616 個を含む傾聴応答データを利用した．実験の結果，本手法の推定性能が他の推定手法を上回った．また，語りの音響情報と言語情報を併用する効果が確認された．最後に，傾聴応答の表出タイミングの検出実験を実施した．その結果，限界はあるものの，表出率の利用により表出タイミングの検出性能が向上する可能性が示唆された．

第4章　語りの傾聴に不同意を示す応答の生成

4.1　はじめに

会話エージェントが語りの聴き手として認められるには，傾聴していることを示す目的で語りに応答する発話である傾聴応答の表出が有力である．これまでに，相槌をはじめとする傾聴応答の生成法が検討されている [30, 31, 32, 33, 34]．語りの傾聴では，語り手に理解を示し，共感を伝えることが重要とされており，反対に，傾聴において相手の話を否定することは，語り手の心を遠ざける原因になりかねない [84]．そのため，語り手の発話を受容することが，語りを傾聴する聴き手の基本的な応答方略となる．例えば，傾聴応答の代表例である相槌は,「語りを続けて」というシグナルや内容理解を示す機能を持っており [64]，これも，語りを受容していることを伝えるものといえる．一方で，語りでは，時として自虐や謙遜などの発話が行われることがある．この場合，その発話内容を否定することなくそのまま受容することは必ずしも適切ではなく，語り手の発話に同意しないことを示す応答，すなわち，**不同意応答**を積極的に表出することが求められる．このように，語りの傾聴を担う会話エージェントが不同意を示すべき発話を検出し応答できることは不可欠な機能であるものの，傾聴応答生成に関する従来研究において，不同意応答の生成に関する試みは行われていない．

そこで本章では，語りを傾聴する会話エージェントによる不同意応答生成の実現性を示す．語り手による自虐や謙遜などの発話をそのまま受容することは致命的であり，このような場面で適切に不同意応答を生成できれば，語りの傾聴を語り手に伝達する上で高い効果が見込まれる．

傾聴応答における不同意応答生成の適切なタイミングや表現は規則的に定まるものではなく，データに基づき決定することが現実的である．そのような不同意応答生成の実現性を示すために，

(1) 不同意応答の生成に利用できる応答コーパスを作成できること，並びに，

(2) 応答コーパスを用いて不同意応答を適切に生成できること

を示す必要がある．そのために本研究では，まず，語りデータに対して，不同意応答をタグ付けする基準を定め，不同意応答の生成に適したコーパスの作成可能性を実証する．続いて，事前学習済みの Transformer [72] ベースのモデルに基づく手法を実装し，不同意応答の表出に適したタイミング（以下，不同意応答タイミング）の検出実験，及び，表出する表現（以下，不同意応答表現）への分類実験を通して，不同意応答生成の実現性を実証する．

本章の構成を以下に示す．4.2 節では，傾聴応答における不同意応答について概説する．4.3 節では，不同意応答の生成に利用できるコーパスの設計について論じ，コーパスの具体的な作成法とその結果について述べる．4.4 節で不同意応答タイミングの検出実験について，4.5 節で不同意応答表現への分類実験について，それぞれ報告する．最後に，4.6 節で本章のまとめを行う．

4.2 傾聴応答における不同意応答

語りの傾聴では，語り手に理解を示し，共感を伝えることが重要とされており，反対に，傾聴において相手の話を否定することは，語り手の心を遠ざける原因になりかねない [84]．そのため，語りの傾聴では，語り手の発話を受容することが聴き手の基本的な応答方略となる．しかし，語りには，次に示すような発話が含まれることがある．

- 老い先短い私にとっては今回のイタリア旅行が人生最後の旅行でしょう

これは，話し手の謙遜が込められた発話であり，聴き手がこれをそのまま受容することは，話し手にとって必ずしも本意ではない．他にも，

- 私は子どもの頃から気が利かないというか空気が読めなくて

のような話し手の自虐が含まれる発話も同様である．このような発話に対して，会話エージェントがそれを受容するような応答をすれば，語りが理解されているかに関する疑念が生じるとともに，語り手が不快に感じる可能性がある．発話内容をそのまま受容することが相応しくない場面では，語り手の発話に同意しないことを示す応答，すなわち，**不同意応答**を確実に表出できることが重要となる．

これまでに，主に言語学の視点から，不同意という言語行為の定義とその分類が行われている．日本語の雑談における不同意を対象とした先行研究 [85] によると，不同

意は,「相手の発話が表す事実あるいは意見について，納得しない，または受け入れていないことを伝える発話（群）」と定義されている．さらに，不同意の先行発話の内容に基づき，不同意を以下の 2 つに分類している．

(1) 実質的不同意：不同意の対象となる先行発話（群）が，相手に対するプラス評価，または自己に対するマイナス評価ではない場合

(2) 儀礼的不同意：不同意の対象となる先行発話（群）が，相手に対するプラス評価，または自己に対するマイナス評価である場合

上述した，自虐や謙遜などの発話は，語り手が自己をマイナス評価するものであり，これらに対する不同意は儀礼的不同意に該当する．なお，儀礼的不同意のうち，相手をプラス評価する発話の例としては，相手を褒める発話が挙げられているが，本研究では，特定の聴き手の存在を前提としておらず，聴き手に対して言及する発話は含まれてない語りを対象とすることから，本研究の対象外とする．すなわち，本研究における不同意応答とは，語り手が自己をマイナス評価している発話に対して，聴き手がその評価を受容しないことを示す応答であるとする．

なお，応答を表出せずにあえて沈黙することも，不同意を示すための戦略の 1 つとして挙げられる．ただし，沈黙によって不同意を示すためには，単に応答を表出しなければよいというわけではなく，その代わりに，表情やジェスチャによって，不同意を伝える必要があると考えられる．本研究では，コミュニケーションロボットやスマートスピーカーなどの，主に音声によって応答する会話エージェントを対象としている．これらの会話エージェントにおいては，不同意を明示的に示す応答を生成できることが重要となる．

しかし，傾聴応答に関する従来研究では，相槌に代表されるように，語り手の発話内容をそのまま受容することを前提とした応答の生成が対象とされており，不同意応答の生成に関する検討は行われていない．そこで本研究では，不同意応答を生成することの実現性を明らかにするために，

(1) 不同意応答の生成に利用できる応答コーパスの作成，並びに，

(2) 応答コーパスを用いた不同意応答の適切な生成

に取り組む．

不同意応答が収録されたデータとして，2 章で述べた傾聴応答コーパスが挙げられる．傾聴応答コーパスでは，2.3 節で述べた方法で語りに対する傾聴応答を収録し，そ

の種類を人手で分類している．このコーパスには，16 種類の傾聴応答が合計 148,962 個収録されている．その 67.96%が相槌であり，相槌以外の傾聴応答の内訳は図 2.3 に示した通りである．感心や繰り返しが多数出現しているのに対し，不同意応答はわずか 0.87%にあたる 411 個の出現に留まっている．傾聴応答コーパスの収録では，作業者は，事前に収録された語りの音声の再生に同期して傾聴応答を表出する．このため，作業による応答の付与は時間制約下で行われる．このようなリアルタイム環境での応答付与では，作業者は必ずしも一様に振舞うことは難しく，不同意応答が期待される場面であっても，必ずしも適切に遂行されない可能性がある．また，対話における妥当な応答は唯一でないため，不同意応答が適する語りの発話であっても，そこで作業者が必ずしも不同意応答を表出するとは限らない．

不同意応答生成の実現性を明らかにするには，不同意応答タイミングの検出結果と不同意応答表現への分類結果を適切に評価できる必要がある．その評価に利用可能なコーパスには，不同意応答タイミングが網羅的に付与され，かつ，不同意応答表現が揺れが少なく，すなわち，安定的に付与されていることが求められる．上述の傾聴応答コーパスは，語りに対する自然な応答音声を収録したのちに，各応答をその種類で分類することにより構築されているため，この要求を満たしていない可能性がある．

4.3 不同意応答生成のための応答コーパス

語りに対して不同意応答を生成することの実現性を示すために，不同意応答の生成に利用可能なコーパスを作成できることを示す．

4.3.1 問題設定とコーパスの作成方針

語りの音声に同期して応答を表出するリアルタイム環境下での収録では，不同意応答タイミングが網羅的に，また，不同意応答表現が安定的に付与されたコーパスを作成することは容易ではない．この問題を解決するために，本研究では，語りに同期しない時間制約なしの環境で，不同意応答をタグ付けする方式を採用する．具体的には，語りのテキストに対して，その内容に基づき，不同意応答の生成に適したタイミングと表現を付与する．テキストに対して付与することで，網羅性と安定性を備えた結果を期待できる．

語りのテキスト上に付与するために，まず，不同意応答が生成されるタイミング上の候補を定める必要がある．一般に，同意や不同意の応答は，語り手の言明に対する

肯定または否定の提示であり，その対象は命題に相当するものであると考えられる．命題に対応する言語的な最小単位は,「節（述語を中心としたまとまり）」であることから，本研究では節を語りの構成単位とし，節の境界を不同意応答タイミングのタグ挿入位置の候補とする．すなわち，タグ付けでは，節の並びで表現された語りに対して，作業者は節ごとにその直後が不同意応答に適したタイミングであるか否かを判断する．

次に，付与可能な不同意応答表現の候補を定める必要がある．傾聴応答コーパスに含まれる 411 個の不同意応答について，その応答表現の出現分布を調査した．まず,「いえ」や「いえいえ」など「いえ」を含む応答表現の総数が全体の 58.15%にあたる 239 個と最も多かった．その他,「いや」や「いやいや」など「いや」を含む応答表現の総数は全体の 10.46%にあたる 43 個と少なかった．応答表現「いえ」と「いや」を比較した先行研究 [86] は,「いえ」の方がより丁寧であり,「いや」が非丁寧（ぞんざい，威圧的）であると位置づけている．傾聴における不同意応答表現として,「いえ」を含む応答表現の出現が最も多かったことは，その丁寧さによるものと考えられる．「いえ」を含む応答表現の中でも「いえいえ」が全体の 32.36%にあたる 133 個と最も多かったことから，本研究では「いえいえ」を不同意応答の代表表現と定めた．

不同意応答表現「いえいえ」は，ほとんどの場合，それが表出されれば不同意を示すことになる．しかし，不同意の対象や意図をより明確に語り手に提示するために,「いえいえ」のみではなく，それに適切な表現を付加することが効果的であると考えられる．上述の先行研究 [86] においても，応答表現「いえ」の後に何らかのコメント（提示された情報を否定する内容，根拠）が付加されやすく，単独の発話では若干の不足感を感じさせる，と指摘されている．そこで本研究では,「いえいえ」と付加表現を連接したものを不同意応答表現とし，付加表現のタイプごとに代表表現を定める．作業者は，付与した不同意応答タイミングに対し，そこで生成するのに適した付加表現を選択する．

4.3.2　語りデータへのタグ付け方針

タグ付けの対象の語りデータとして，傾聴応答コーパスの構築と同じく，高齢者のナラティブコーパス JELiCo [67] を用いた．この語りデータには，合計で 30 名の高齢者による 1 人約 20 分の語りの音声が収録されており，全高齢者共通の 10 個の質問に対し，その回答を語るという収録形式が採用されている．このうち本研究では，高齢

者 29 名が共通の質問 9 個に対して回答した語りをタグ付け対象とした[1].

語りの分割には，節境界解析ツール CBAP [87] を用いた．ツールによって挿入された節境界で挟まれた区間を「節」と定めた.

4.2 節で述べた通り，本研究における不同意応答とは，語り手が自己をマイナス評価している発話に対して，聴き手がその評価を受容しないことを示す応答のことをいう．そこで本研究では，不同意の対象となる先行発話（群）におけるマイナス評価のされ方に基づき,「いえいえ」に連接する付加表現のタイプを決定する．具体的には，傾聴応答コーパスにおいて不同意応答が付与された周辺の語りの発話における，マイナス評価のされ方の観察を通して，付加表現のタイプを定めた．その結果 6 種類に分類され，各分類に対して代表表現を定めた．定めた代表表現をそのタイプ，すなわち，どのようなマイナス評価に対して利用可能であるか，と共に以下に示す.

(1) そんなことないですよ
　以下の (2)∼(6) に該当しない

(2) いいと思いますよ
　プラス評価になりうることを否定的に捉える

(3) 十分ですよ
　プラス評価の水準に達していないことを示す

(4) これからですよ
　過去と比較し現在を否定的に捉える

(5) 大丈夫ですよ
　謝罪あるいは恐縮する

(6) 仕方ないですよ
　自責，後悔，無念，卑下を感じさせる否定的な事実（失敗談等）を自己開示する

上述の分類とそれに対する代表表現の割り当てが，過不足ないものとなっているかを確認するために，作業者が付加表現を選択する際には，上述の 6 つの代表表現に加えて，いずれにも属さない「その他」を選択することも可能とした.「その他」を選択し

[1] 10 個の質問のうち,「動物の名前を思いつくだけ言ってください」という質問に対して回答する形式の語りは，本研究が対象とする語りとは性質が異なるので，作業の対象外とした．また，高齢者 1 名分の語りについては，作業者への不同意応答のタグ付けの説明に利用したため，作業の対象外とした.

表 4.1: 作成した不同意応答コーパスの規模

節の数	10,662
不同意応答タイミング数	459
不同意応答タイミングの割合	4.31%

た際には，上述の代表表現以外の,「いえいえ」に連接する付加表現を付与するよう指示した．

タグ付けは，著者や所属研究室の関係者ではない，言語コーパスのアノテーション作業に精通した 1 名の作業者が実施し，実施にあたり参照可能な作業マニュアルを作成した．マニュアルでは，作業の目的，傾聴における不同意応答の説明，データの説明，タグ付け作業内容の説明，タグ付け作業の参考例を記すとともに，不同意応答の表出に適するタイミングを網羅的にタグ付けするよう指示した．

4.3.3 作成されたコーパスとその評価

(i) 不同意応答タイミングのタグ付け結果とその評価

作成した不同意応答コーパスの規模を表 4.1 に示す．459 個の不同意応答タイミングが付与された．節の直後の 4.31% (459 / 10,662) に不同意応答タイミングが出現している．比較には，2 章で述べた傾聴応答コーパスを用いる．傾聴応答コーパスは，語りに対する傾聴応答をリアルタイム環境下で収録することで構築されており，本研究の不同意応答コーパスと同じ語りデータが用いられている．両コーパスの不同意応答タイミングの比較のために，傾聴応答コーパスに含まれる不同意応答を節の直後に対応付ける処理を行った．具体的には，不同意応答を，その発声開始時刻が，語りデータ内の節の発声終了時刻のうち，最も近くにある節に対応付けた．この処理によって，対応付けられた不同意応答が存在する節の直後を，傾聴応答コーパスにおける不同意応答タイミングとみなした．その結果，傾聴応答コーパスにおける不同意応答タイミングは 197 個であり，節全体の 1.85% (197 / 10,662) であった．両コーパスにおける不同意応答タイミング数の差は著しく，時間制約のない環境で不同意応答のみに限定してタグ付けを実施することの効果が示された．

時間制約のない環境で作業することにより，網羅的にタグ付けが行われたかを評価するために，コーパス作成のタグ付け作業者 (X) とは別の作業者 (A) を設け，タグ付

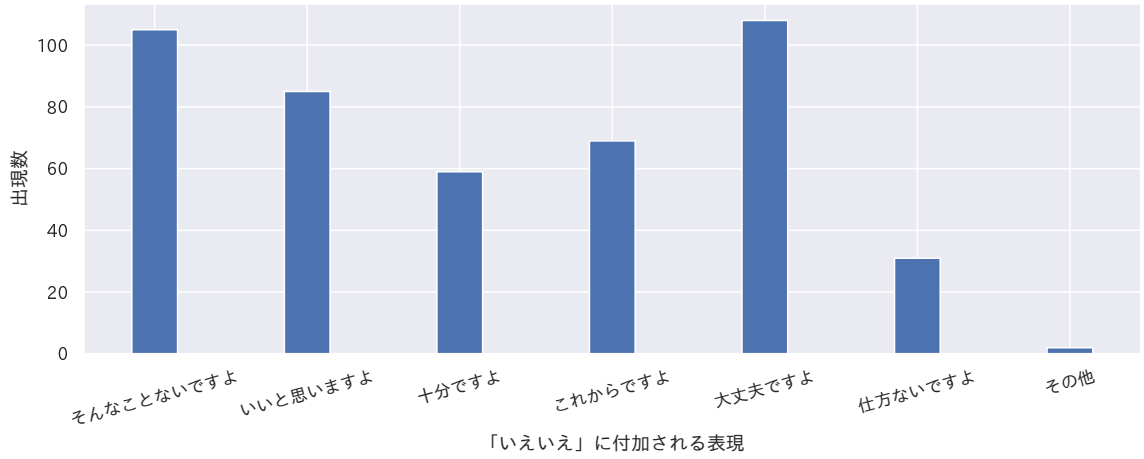


図 4.1: 「いえいえ」に付加される表現の出現数

けの一致率を測定した．一致率を測定するにあたって，作業者 (A) は，作業者 (X) と同じく，すべての節に対してタグ付け作業を行った．作業者間の一致度が高ければ，付与されたタグ付け位置の他にタグ付けの余地が少ないことを意味し，作成されたコーパスが高い網羅性を備えていることを示唆する．一致率の測定には，Cohen の kappa 値 [88] を用いた．この指標は，観測された一致率を $P(O)$, 期待される一致率を $P(E)$ とするとき，

$$\kappa = \frac{P(O) - P(E)}{1 - P(E)} \tag{4.1}$$

で算出する．$.80 < \kappa$ であれば good reliability, $.67 < \kappa < .80$ であれば usable quality の水準にあるとされる [89]．測定の結果，$\kappa = 0.771$ であり，実質的に一致している水準にあった．比較のために，傾聴応答コーパスに含まれる 11 名の作業者どうしで，不同意応答タイミングの一致率を調査したところ，kappa 値は最も高い作業者の組み合わせで 0.390 であった．以上の結果から，時間制約のない環境で不同意応答タイミングを付与することで，相対的に高い網羅性を備えた応答コーパスを作成できることが示された．

(ii)　不同意応答表現のタグ付け結果とその評価

図 4.1 に，作成されたコーパスにおいて，不同意応答タイミングに付与された不同意応答表現の出現数を示す．いずれのタイプの表現も多く出現していること，また，「その他」の出現はわずか 2 個に留まっていることから，4.3.2 節で述べた 6 つの分類

表 4.2: 不同意応答タイミングと不同意応答表現の例

語り	不同意応答表現
そこの間にもう一つのビー玉を入れるっていうのという遊びですねそれからお手玉ですねお手玉はなんか私不器用で	いえいえ，そんなことないですよ
あったんですがあたくしはもうどこに約束したならばいつも三十分前にはいあのーいつも早出をするものでもう性分って言いますか	いえいえ， いいと思いますよ
何とかあの健康で時々友達とも話がしたりできればいいかなまーそのくらいのことでしかありません	いえいえ，十分ですよ
若い頃でしたらしたいことはいっぱいありましたけれども今したいていうことはもうそういう力がなくなってしまったし	いえいえ，これからですよ
どちらから乗るか上り下りがあるからちょっとわかりませんね進行方向あそれはちょっとわたくしすみません	いえいえ，大丈夫ですよ
するようなそういう勉強をしなきゃいけないのかと思うとやっぱり医者にはなれませんでした	いえいえ，仕方ないですよ

が概ね過不足なく定められたものであることが示唆された[2]．表 4.2 に，表現ごとに，それがタグ付けられた不同意応答タイミングの例を示す．表では，不同意応答タイミングから遡って 5 つの節を記載している．

不同意応答表現のタグ付けが安定的に実施されているかを評価する．4.3.3 節の (i) で述べた作業者 (A) が，自身が付与した不同意応答タイミングに同様のタグ付けを実施した．作業者 (X) と (A) の両者でタグ付けが一致した不同意応答タイミング 349 個の節の直後を対象に，それぞれがタグ付けた表現に関する混同行列を図 4.2 に示す．

[2]代表表現として「その他」が選択された 2 つの不同意応答タイミングでは,「体が痛くても我慢する」といった内容の語りの発話に対して，作業者によって「無理しないでください」という表現が付与されていた．

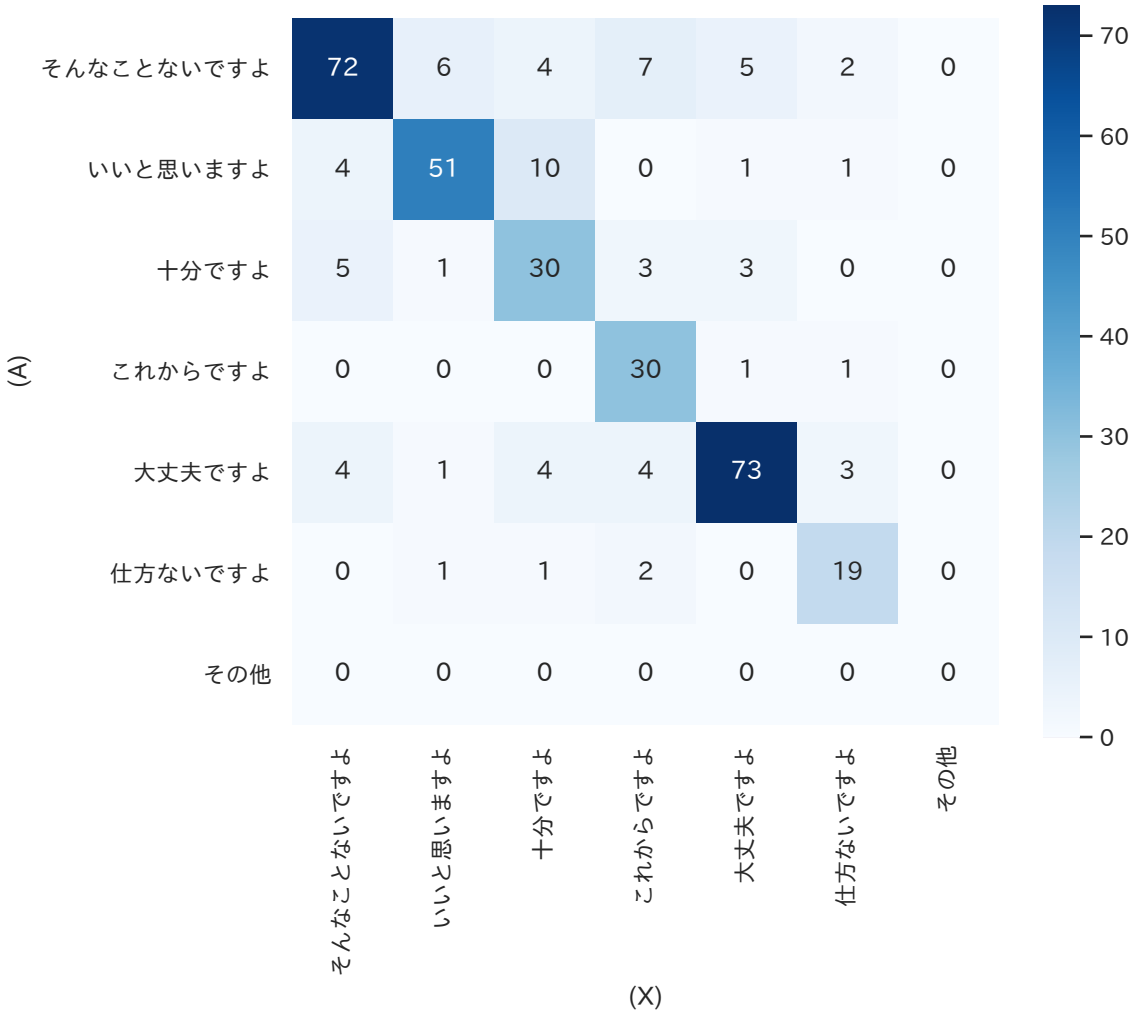


図 4.2: 「いえいえ」に付加される表現に関する作業者間の混同行列

78.80%（275/349）において表現が一致しており，安定的に不同意応答表現をタグ付けできているといえる．

4.4 不同意応答タイミングの検出実験

語りの傾聴において，不同意応答を生成することの実現性を示すために，本章の以下では，不同意応答コーパスが不同意応答の生成に効果的に機能することを示す．本節では，不同意応答タイミングの検出における応答コーパス利用の有効性を実験的に検証する．

表 4.3: 実験データにおける不同意応答タイミングの個数と割合

	不同意応答タイミング	
学習及びテスト	4.26%	(361/ 8,479)
開発	4.49%	(98/ 2,183)
Total	4.31%	(459/10,662)

4.4.1 実験概要

本実験では，語りの節ごとに，その直後が不同意応答タイミングであるか否かの判定を行う．実験は 5 分割交差検定で実施した．不同意応答コーパスを 8:2 に分割し，前者を学習及びテスト用データ，後者を開発データとした．学習及びテスト用データを 5 分割し，そのうちの 1 グループをテストデータ，残りの 4 グループを学習データとした実験を 5 回繰り返した．開発データは，交差検定の各試行における，ハイパーパラメータの調整に用いた．表 4.3 に，実験データにおける不同意応答タイミングの個数と節数に対する割合を示す．評価指標は，不同意応答タイミングに対する適合率，再現率，F 値とした．

4.4.2 不同意応答タイミングの検出手法とその実装

本節では，作成したコーパスを用いて不同意応答タイミングを検出する手法について述べる．本手法では，節の並び $c_1 \cdots c_n$ で表される語りに対して，語りにおける節が 1 つ入力されるごとに，その直後が不同意応答タイミングであるか否かを判定する．本研究の目的は，不同意応答タイミング検出の実現性を示すことであるため，不同意応答タイミングであるか否かの判定には，幅広いタスクで用いられている事前学習済みの Transformer [72] ベースのモデルを単純に用いた手法を採用した．具体的には，事前学習済みモデルを 2 クラス分類タスク用に fine-tuning することで，語りにおける節の直後が不同意応答タイミングであるか否かを判定するためのモデルを構築した．図 4.3 に本手法の概略を示す．語りにおける節 c_t の直後の判定の際には，節 c_t を含めて直前 m 個の節の並びを入力として用いる．すなわち，節の並び $c_{t-(m-1)}c_{t-(m-2)} \cdots c_t$ で表される語りの文字列を入力とする．入力の文字列を Transformer ベースの事前学習済みモデルでエンコードしたのち，全結合層と softmax 関数による分類層で処理することで，2 次元のベクトルを得る．最後に，argmax 関数を適用することによって，

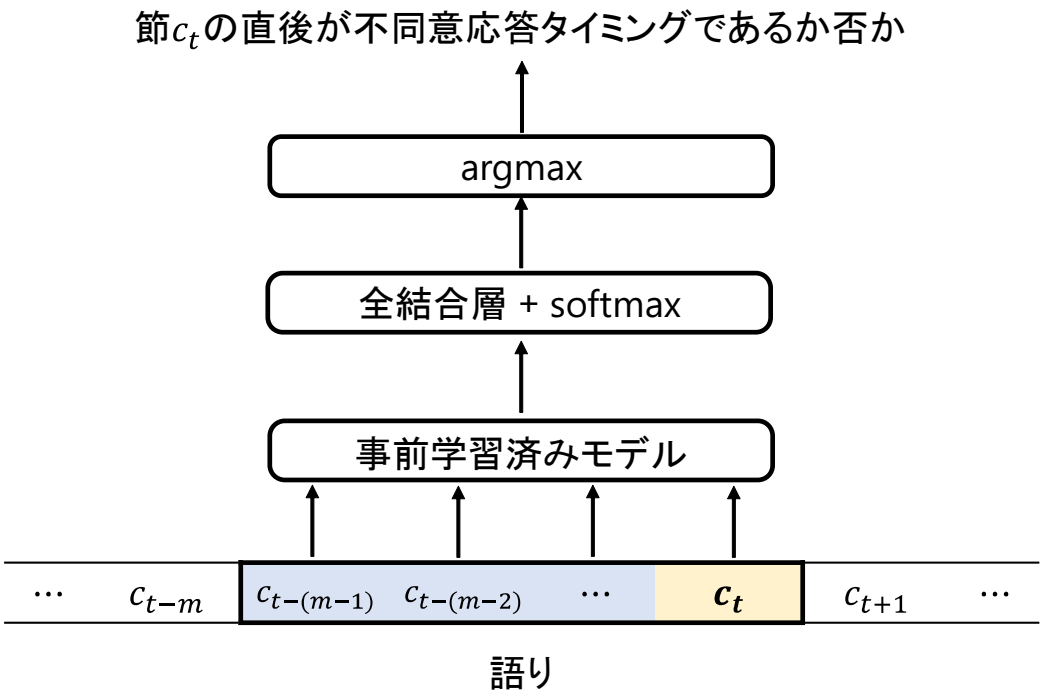


図 4.3: 不同意応答タイミングの検出手法の概略

節 c_t の直後が不同意応答タイミングであるか否かの判定結果が出力される．

不同意応答タイミングの検出のためには，語りの文脈を考慮することが望ましい．そこで本実験では，節 c_t を含めて直前 5 個の節の並びを入力として用いる．すなわち，図 4.3 における m を 5 に設定した．検出手法で用いる事前学習済みの Transformer ベースのモデルとしては，この実装を開始した時点における主流なものとして，BERT [53], RoBERTa [73], DeBERTa-v2 [74], GPT2 [62] を採用した．いずれも，huggingface[3]で公開されている，隠れ層のサイズが 768 次元，レイヤー数が 12 のモデル[4 5 6 7]である．モデルの学習における損失関数は Cross Entropy Loss として，バッチサイズを 32, エポック数を 50 とした．モデルの最適化手法には，weight decay を 0.01 とした AdamW [75] を用いた．学習率には，warmup ratio を 10%とした，線形スケジューリングを採用した．すなわち，学習開始時から学習全体の 10%が終了するまでの間は，事前に定めた学習率の最大値まで線形に学習率を増加させ，それ以降は線形に学習率を減少させた．本実験では，各モデルの学習率の最大値は，1e-6, 5e-6, 1e-5, 5e-5 のうち，学

[3] `https://huggingface.co/models`
[4] cl-tohoku/bert-base-japanesse-whole-word-masking
[5] nlp-waseda/roberta-base-japanese
[6] ku-nlp/deberta-v2-base-japanese
[7] rinna/japanese-gpt2-small

表 4.4: 不同意応答タイミング検出の実験結果

	適合率	再現率	F 値
random (even)	0.043	0.518	0.079
random (balanced)	0.035	0.039	0.037
BERT	0.473	0.247	0.324
RoBERTa	0.521	0.377	0.437
DeBERTa-v2	0.492	0.332	0.397
GPT2	0.516	0.349	0.417

習終了時のモデルの開発データにおける F 値が最良となったものとした．

上述の事前学習済みモデルによる手法の性能を評価するために，節の直後が不同意応答タイミングであるか否かを，不同意応答コーパスに基づくことなくランダムに分類する以下の 2 つの手法を実装し，その結果を参照することとした．

- **random (even)**: 50%の確率でランダムに分類する手法
- **random (balanced)**: 学習データにおける不同意応答タイミングの割合に従ってランダムに分類する手法

4.4.3 実験結果

表 4.4 に，各手法の適合率，再現率，F 値を示す．ランダム手法は，random (even) の F 値が 0.079, random (balanced) の F 値が 0.037 と極めて低い．一方で，事前学習済みモデルを用いた手法の F 値は，最低で 0.324, 最高で 0.437 を達成した．事前学習済みモデルを単純に fine-tuning した手法を用いており，その性能に向上の余地があるとはいえ，不同意応答タイミングを一定の水準で検出できること，並びに，不同意応答コーパスの利用の効果が示された．

表 4.5 に，事前学習済みモデルに基づく 4 つの手法のいずれかで検出に成功した不同意応答タイミングの例を，その直前の語りと共に示す．成功例 1 は末尾の「ごめんなさい」の直後（4 つの手法すべてで成功），成功例 2 は末尾の「悪いんですけれども」の直後（DeBERTa-v2 による手法と GPT2 による手法で成功）が，それぞれ正解の不同意応答タイミングである．これらの例では，謝罪の発話を含む語りを対象とした不同意応答のタイミング検出に成功している．

表 4.5: 検出に成功した不同意応答タイミングとその直前の語り

	不同意応答タイミング直前の語り
成功例 1	これもなんにもなかったのでお話しすることほんとないんでごめんなさい
成功例 2	必ずわたくしが同行してるんでもう自分のじ時間が大変少ないんですねでそれがやはりまー主人には悪いんですけれども

4.4.4 エラー分析

実験結果のエラー分析を行う．エラーの要因を明らかにするために，4.3.2 節の付加表現の分類ごとに，実験結果における不同意応答タイミング検出の再現率を算出した．表 4.6 にその結果を，事前学習済みモデルに基づく 4 つの検出手法ごとに示す．1 行目の（ ）内の数字は，学習及びテストデータ区分の各応答表現の個数である[8]．いずれも,「大丈夫ですよ」に対する不同意応答タイミング検出の再現率が相対的に高いという結果となった．この表現は，謝罪や恐縮などを含む発話に対して利用可能であり，そのような発話には,「申し訳ない」「すいません」「ごめんなさい」などのフレーズが含まれることが多い．したがって，モデルはその特徴を捉えやすかったものと考えられる．

一方で,「そんなことないですよ」「いいと思いますよ」については,「大丈夫ですよ」と同程度の出現頻度であるにも関わらず，その再現率は高くなかった．表 4.7 に，事前学習済みモデルに基づく 4 つの手法のいずれかが，検出すべきか否かの判断を誤った例を，その直前の語りと共に示す．失敗例 1 は末尾の「暇があるので」の直後，失敗例 2 は末尾の「私がその暇を持て余しているっていうことから」の直後が，正解の不同意応答タイミングである．それぞれ，付加表現としては,「いいと思いますよ」「そんなことないですよ」が選択されていた．これら 2 つの不同意応答タイミングは，4 つの手法すべてで検出できていなかった．いずれも，語り手自身に暇があることを自虐的に話す発話であるが，それぞれ語りの表現は異なる．一方，実験データには,「ちょっとできたんですけれどもね特別うーん遊びって小さい時はあんまり遊べない子でしたね今やっと暇ができて」という語りの発話も存在していた．これは，語り手が自身に暇があることを前向きに語っている発話であり，この発話の直後は不同意応

[8]実験データには「その他」が付与された不同意応答タイミングが 2 つ存在するが，いずれも開発データに割り当てられている．

表 4.6: 不同意応答タイミング検出の再現率（付加表現の分類別）

	そんなことないですよ (83)	いいと思いますよ (71)	十分ですよ (45)	これからですよ (56)	大丈夫ですよ (80)	仕方ないですよ (26)
BERT	0.205	0.113	0.311	0.232	0.400	0.192
RoBERTa	0.349	0.282	0.378	0.339	0.550	0.269
DeBERTa-v2	0.253	0.211	0.267	0.339	0.563	0.308
GPT2	0.265	0.282	0.244	0.411	0.538	0.269

表 4.7: 検出に失敗した不同意応答タイミングとその直前の語り

	不同意応答タイミング直前の語り
失敗例 1（未検出）	して一緒に聞きにいきますそれがやっぱり楽しみです年寄りは暇があるので
失敗例 2（未検出）	はい一つは私がその暇を持て余しているっていうことから
失敗例 3（誤検出）	形で私よりえーと十五歳ぐらい年下ですかね十歳ぐらい年下ですかねそんなことであの自分は誰も相談する人がないっていう

答タイミングではない．このように，語られる内容が似ていても，それが自虐や謙遜として語られるかは，語り手の立場や状況によって異なる．さらに，語りの発話に自虐や謙遜が含まれているか否かの判断には，語りの対象が何であるかにも依存する．例えば，表 4.7 の失敗例 3 の語りの末尾「人がないっていう」の直後は，不同意応答タイミングではない．しかし，RoBERTa による手法と DeBERTa-v2 による手法は，誤って不同意応答タイミングであると検出していた．失敗例 3 の語りの対象は，語り手ではなく，語り手よりも 10 歳から 15 歳年下の別の人物である．「自分は誰も相談する人がない」という語りの発話は，文脈によっては自虐や謙遜になりうる．しかし，失敗例 3 における「自分」は語り手ではないため，自虐や謙遜の発話とは異なる．このように，自虐・謙遜に対する不同意応答タイミングの検出には，語りをより深く理解することが必要であり，その検出は容易ではなかったと考えられる．

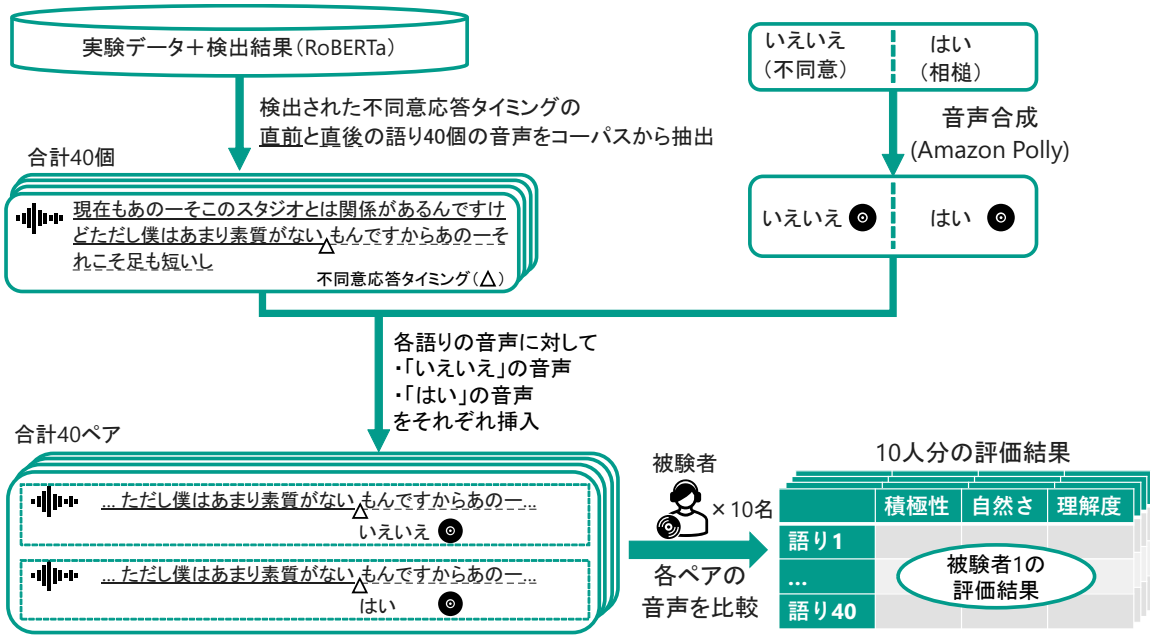


図 4.4: 不同意応答タイミングの検出結果の主観評価の手順の概略（ダミーのペアの比較の工程は除く）

4.4.5　検出結果の主観評価

4.4.3 節では，不同意応答タイミングを網羅的にタグ付けしたコーパスを正解データとし，それとどの程度一致するかという観点から，不同意応答タイミングを一定の水準で検出できること，並びに，不同意応答コーパスの利用の効果を確認した．しかしながら，会話エージェントに対する語り手の印象は様々な要因によって影響を受ける．そのため，語り手の印象に影響を与えるような水準の不同意応答を生成できることを確認するには，正解データとの一致に基づく評価だけでは十分ではない可能性がある．そこで本節では，主観評価に基づき，不同意応答タイミングの検出における不同意応答コーパスの利用の有効性を評価する．

(i)　評価の概要

主観評価の手順の概略を図 4.4 に示す．実験データにおいて，本手法が検出したタイミングで，実際に不同意応答を生成し，それを被験者が主観評価する[9]．被験者は，不同意応答「いえいえ」が挿入された語りの音声を聴取し，その印象を評価する．本

[9]一方で，正解のタイミングで検出されなかった場合についても，同様に主観評価することは有意義である．不同意応答が生成されないときの損失の程度を被験者実験により測定することが考えられ，その実施は今後の課題である．

表 4.8: 主観評価における評価項目

項目	説明
積極性	聴き手の応答は，話し手に対して語りを積極的に（集中して）聴いていることを伝えるものであったか．また，聴き手は，積極的に（集中して）語りを聴いていたと感じたか．
自然さ	聴き手の応答は，自然であったか．
理解度	聴き手の応答は，話し手に対して語りを理解していることを伝えるものであったか．また，聴き手は，語りを理解していると感じたか．

手法は，語りの節ごとに，その直後が不同意応答タイミングであるか否かを，直前の語りの文字列から判定することで検出を行う．不同意応答「いえいえ」の挿入位置は，検出された節の最終形態素の発声終了時刻の 200 ミリ秒後とした．検出には，表 4.4 の F 値が最も高かった RoBERTa による手法を用いた．

評価は，表 4.8 の積極性，自然さ，理解度の 3 つの項目で実施した．ただし，挿入された不同意応答を絶対評価することは必ずしも容易ではない．そこで本実験では，不同意応答のタイミングで，相槌が挿入された語りの音声を別途用意し，不同意応答が挿入された音声との比較評価を行った．すなわち被験者は，積極性，自然さ，理解度の 3 つの項目で，不同意応答が挿入された音声と相槌が挿入された音声のどちらが優れていたかを評価する．相槌の応答表現としては，傾聴応答コーパスで最も出現が多かった「はい」を用いた．不同意応答「いえいえ」と相槌「はい」の音声には，Amazon Polly [10]によって生成された合成音を用いた．いずれの応答についても，合成音ができるだけ自然となるように，アクセント，話速，発声の強度を調整して音声合成を行った．被験者が聴取する語りの音声の範囲は，検出されたタイミングの直前 5 つ目の節から直後 2 つ目の節までとし，コーパスから抽出した人間の語り手の音声を用いた．

実験データのうち，RoBERTa を用いた手法が検出したタイミングは 261 個存在していた．そのうち 40 個をランダムに抽出して主観評価に用いた．また，主観評価は，30 代から 50 代の被験者 10 名が各々実施した．すなわち，10 名の被験者が，同一の語り音声に対して不同意応答「いえいえ」が挿入された音声と相槌「はい」が挿入された音声の比較評価を 40 回行った．

[10]https://aws.amazon.com/jp/polly/

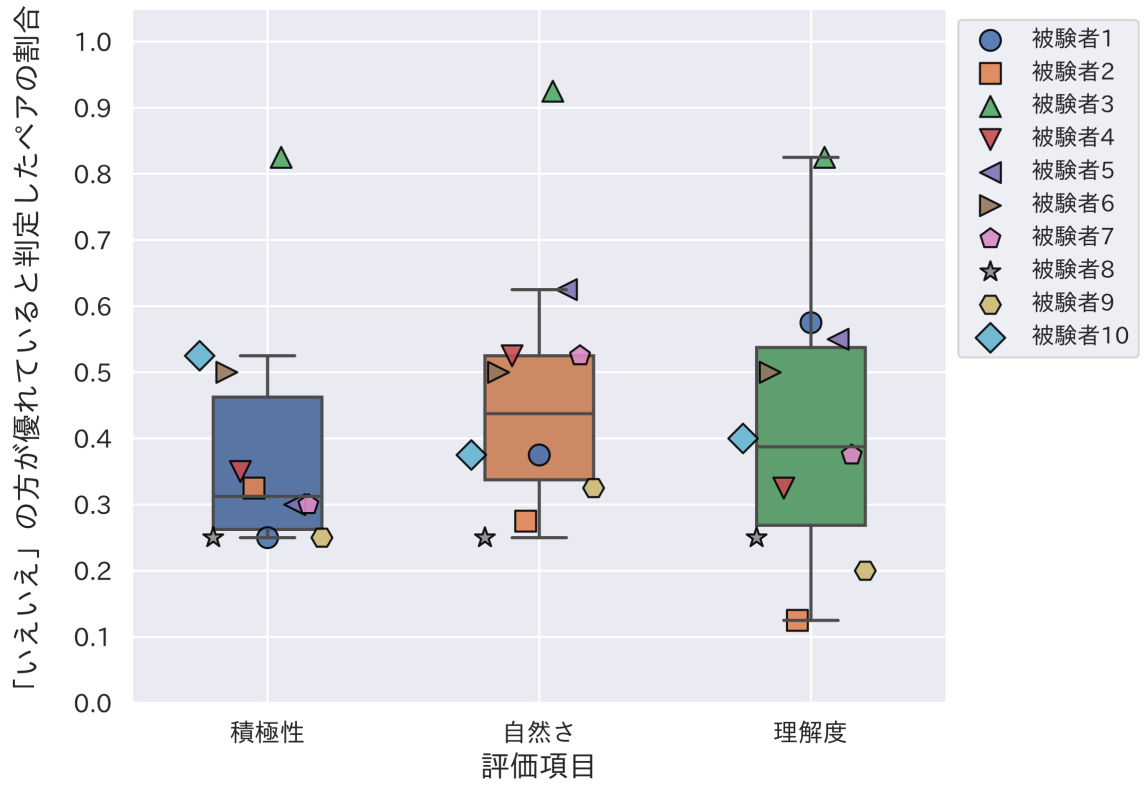


図 4.5: 被験者が不同意応答「いえいえ」の方が優れていると判定したペアの割合の分布

主観評価では，被験者が評価の目的や意図を推測して偏見を持っていると，純粋な評価が行われないリスクがある．これを軽減するために，上述の不同意応答「いえいえ」と相槌「はい」がそれぞれ挿入された音声のペア 40 個に加えて，ダミーのペアを 20 個用意した．すなわち，10 名の被験者が各々，不同意応答「いえいえ」と相槌「はい」がそれぞれ挿入された音声ペア 40 個に，ダミーのペア 20 個を加えた合計 60 個の音声ペアを比較評価した．ダミーのペアは，不同意応答「いえいえ」と相槌「はい」を挿入する際に，少なくとも一方を別の応答に置き換えることで作成した．置き換えに用いる応答は，傾聴応答コーパスに含まれる応答からランダムに選択し，Amazon Polly によって音声合成した．

(ii) 評価結果

図 4.5 に，被験者が不同意応答「いえいえ」の方が優れていると判定したペアの割合の分布を示す．被験者 10 人中，積極性では 3 人，自然さでは 5 人，理解度では 4 人

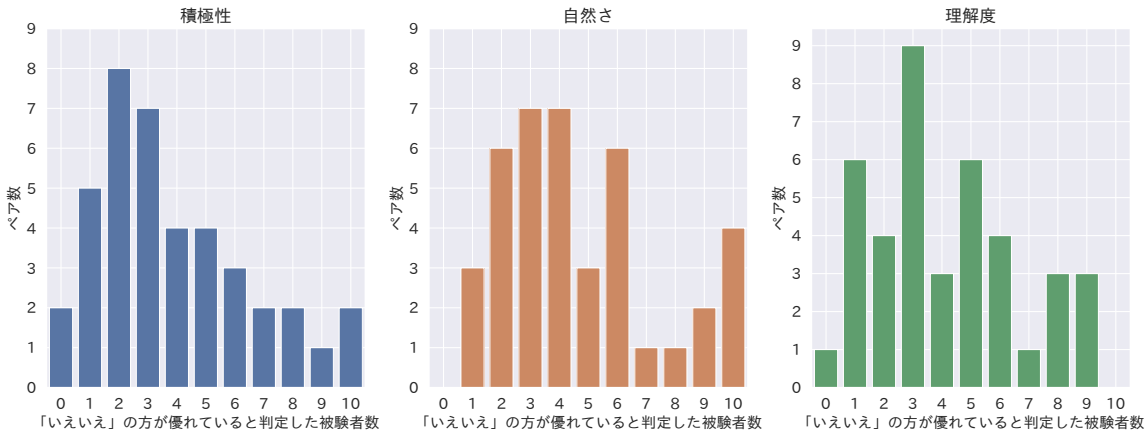


図 4.6: 不同意応答「いえいえ」の方が優れていると判定した被験者数ごとのペア数の分布

において，不同意応答「いえいえ」の方が優れていると判定した回数が，全体の半数以上となった．相槌「はい」を高く評価する傾向にある被験者の方が多かったものの，不同意応答「いえいえ」を高く評価する傾向にある被験者も存在することを確認できた[11]．語り手の個性や嗜好などによっても，不同意応答への印象は異なるものと考えられる．これらの情報を考慮した不同意応答の生成は今後の課題である．

図 4.6 に，不同意応答「いえいえ」の方が優れていると判定した被験者数ごとのペア数の分布を示す.「いえいえ」の方が優れていると判定した被験者数について，積極性では 2 名，自然さでは 3 名または 4 名，理解度では 3 名のペアが最も多かった．一方で,「いえいえ」の方が優れていると判定した被験者が多くを占めるペアも存在していた．このことから,「はい」の生成の方が好印象な場面が多かったものの，不同意応答コーパスに基づく本手法によって「いえいえ」を生成することで，好印象を獲得できた場面も存在することを確認できた．

表 4.9 に，被験者による主観評価の例を示す．不同意応答とその周辺の語りの列については，＜＞で囲んだ文字列が不同意応答であり，その他の文字列が語りである．積極性，自然さ，理解度の数値は，不同意応答「いえいえ」の方を優れていると判定した被験者の数である．例 1 と例 2 は，被験者からの評価が高かった例である．それぞれ，語り手の自虐的な発話と，謝罪の発話に対して，適切に不同意応答「いえいえ」

[11]被験者 3 は，特に「いえいえ」を好む傾向にあったため，その評価結果の信頼性を調査した．各評価項目において，被験者 3 が「いえいえ」の方が優れていると判定したペアと,「はい」の方が優れていると判定したペアのそれぞれにおける，他の 9 名の被験者の評価傾向を比較した．その結果，前者のペアの方が，他の 9 名の被験者も「いえいえ」が優れていると判定する傾向が高かった．このことから，被験者 3 の評価が不適当ではないことが示唆された．

表 4.9: 不同意応答タイミングの検出結果に対する主観評価の例

	不同意応答とその周辺の語り	積極性	自然さ	理解度
例 1	クレヨンだったかもしれませんクレヨンでこう色付けをしていきますけれどもそのわたくしはセンスがなくて ＜いえいえ＞ なんか野暮ったいー塗絵にあの色の塗り付けになるのですが小学校のクラスの女の子でとてもとてもその塗り絵を愛らしく綺麗な色で塗る	7	10	8
例 2	一番嬉しいですねそんなこと出来事といえばそのくらいですすいません ＜いえいえ＞ 終わります	8	9	9
例 3	おかげだって言ってくれたことが本当に嬉しかったですだいたい一般に人様のそうやってお世話焼きばっかりしているものですから ＜いえいえ＞ そのーなんていうんですかねその感謝の気持ちはよくまー言ってたんですけど	4	6	6
例 4	やはり現在のうー第一の趣味であるうーんゴルフあーあるいはあーまああーあとあーあ晴耕雨読ではないですが ＜いえいえ＞ まあー雨の日にはあー囲碁	1	1	1

を生成できたため，評価が高かったものと考えられる．例 3 は，被験者内で評価が割れた例である．語り手の「お世話焼きばっかりしている」という発話が，自虐や謙遜に当たるか否かの判断が被験者ごとに異なったため，評価が割れたものと考えられる．例 4 は，被験者による評価が低かった例である．「いえいえ」の直前には，語り手の「晴耕雨読ではないですが」という否定語と逆接を含む発話があるものの，自虐や謙遜，謝罪などには該当せず，評価が低かったものと考えられる．否定や逆接を表す表現の扱いは，不同意応答の生成における今後の課題であるといえる．

表 4.10: 不同意応答表現への分類実験で使用したデータの内訳

	学習及びテスト	開発
(1) そんなことないですよ	83	22
(2) いいと思いますよ	71	14
(3) 十分ですよ	45	14
(4) これからですよ	56	13
(5) 大丈夫ですよ	80	28
(6) 仕方ないですよ	26	5
Total	361	96

4.5 不同意応答表現への分類実験

不同意応答表現への分類における不同意応答コーパス利用の有効性を実験的に検証する．

4.5.1 実験設定

実験では，不同意応答タイミングを，4.3.2 節で設定した「その他」を除く 6 種類の不同意応答表現に分類する．実験は，不同意応答タイミングの検出実験と同様の交差検定で実施した．分類手法として，不同意応答タイミングの検出手法と同じく，Transformer ベースのモデルを単純に用いた手法を採用する．具体的には，事前学習済みモデルを 6 クラス分類タスク用に fine-tuning することで分類モデルを構築する．すなわち，不同意応答タイミングを，6 つの表現のいずれかに分類するためのモデルを構築する．モデルの入力には，不同意応答タイミングから遡って 5 つの節を用いる．

不同意応答表現への分類性能を評価するために，4.4.1 節と同じデータを用いた．ただし，表現として「その他」が付与された 2 個の不同意応答タイミングは，本実験データから取り除いた．表 4.10 に，本実験データでの不同意応答タイミングにおける,「いえいえ」に付加される表現の内訳を示す．評価指標には，正解率に加えて，マクロ適合率，マクロ再現率，マクロ F 値を用いた．

事前学習済みの Transformer ベースのモデルとしては，BERT, RoBERTa, DeBERTa-v2, GPT2 を採用した．モデルの学習における損失関数は Cross Entoropy Loss として，バッチサイズを 8, エポック数を 50 とした．最適化手法には weight decay を 0.01

表 4.11: 不同意応答表現への分類の実験結果

	正解率	適合率	再現率	F 値
random (even)	0.155	0.166	0.161	0.151
random (balanced)	0.205	0.172	0.173	0.168
BERT	0.410	0.355	0.359	0.353
RoBERTa	0.499	0.437	0.446	0.438
DeBERTa-v2	0.454	0.391	0.402	0.395
GPT2	0.402	0.363	0.360	0.361

とした AdamW を，学習率には warmup ratio を 10%とした線形スケジューリングを採用した．本実験では，各モデルの学習率の最大値は，1e-6, 5e-6, 1e-5, 5e-5 のうち，学習終了時のモデルの開発データにおける正解率が最良となったものとした．上述の事前学習済みモデルによる手法の他に，不同意応答タイミングにてランダムに分類する以下の 2 つの手法を実装した．これらの手法による結果は，その性能が分類の難易度を示すものとして参照できる．

- **random (even)**: 50%の確率でランダムに分類する手法
- **random (balanced)**: 学習データにおける表現の出現分布に従って，ランダムに分類する手法

4.5.2 実験結果

表 4.11 に，各手法の正解率，適合率，再現率，F 値を示す．事前学習済みモデルを用いた手法に関しては，正解率は 0.4 程度であった．一方で，ランダムな手法である random (even) の正解率は 0.155, random (balanced) の正解率は 0.205 であった．適合率，再現率，F 値についても，同程度の結果が得られた．これらの結果から，事前学習済みモデルを単純に用いた手法によって，一定の水準で不同意応答表現に分類できること，並びに，不同意応答コーパスの利用の効果が示された．

図 4.7 に，random(even), random(balanced), 事前学習済みモデルに基づく 4 つの手法の分類結果に対する混同行列を順に示す．混同行列内の括弧書きの数字は，表 4.10 における表現の番号と対応している．事前学習済みモデルに基づく手法の混同行列では，対角成分の値が比較的大きくなっている．このことからも，不同意応答表現への

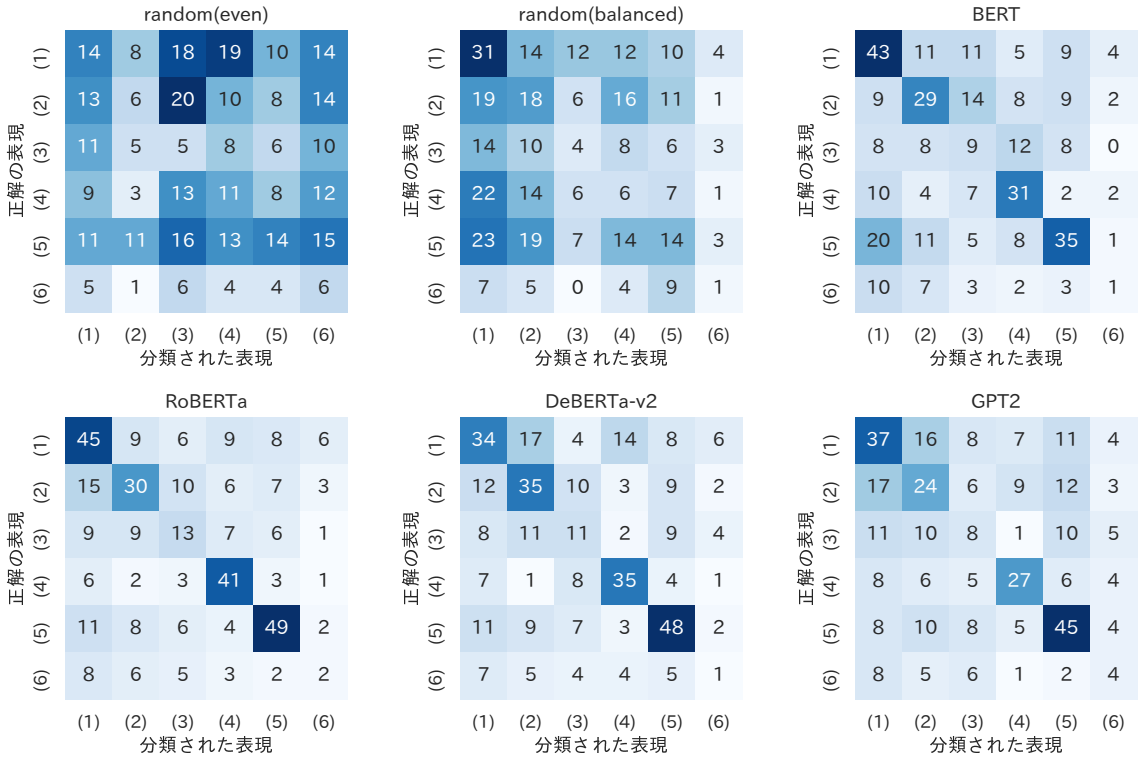


図 4.7: 不同意応答表現への分類に関する混同行列

分類可能性を確認した．表 4.12 の上段に，事前学習済みモデルに基づく手法の全てで分類に成功した例を，直前の語りと共に示す．図 4.7 から分かる通り，いずれの手法も,「いえいえ，大丈夫ですよ」については，適切に分類することができていた．この表現は，表 4.12 の成功例のように，謝罪の語りに対して表出されることが多く，モデルはその特徴を捉えていたものと考えられる．

4.5.3 エラー分析

表 4.13 に，各事前学習済みモデルに基づく 4 つの分類手法の F 値を，不同意応答表現ごとに示す．いずれの手法も,「いえいえ，仕方ないですよ」に対する F 値が極めて低く，ほとんど正しく分類できなかった．表 4.12 の下段に，事前学習済みモデルに基づく手法の全てが正しく分類できなかった例を，直前の語りと共に示す．この失敗例では,「いえいえ，そんなことないですよ」に誤って分類した手法が多かった.「いえいえ，仕方ないですよ」に正しく分類するには,「草書体は一般に難解であり，それを読める人は稀である」という知識を有し，それに基づき「仕方ない」事象であることを推論できる必要がある．本実験で実装した手法では，そのような知識や機能を備

表 4.12: 不同意応答表現への分類の成功例と失敗例

	不同意応答タイミング直前の語り	正解の不同意応答表現
成功例	特に特に出来事っていうのに遭遇してないですねごめんなさい	いえいえ，大丈夫ですよ
失敗例	帰ってきました写真に撮ってただ漢字とかかなとか草書で書いてあるんで読めないんで	いえいえ，仕方ないですよ

表 4.13: 各不同意応答表現に対する F 値（「いえいえ」は省略）

	そんなことないですよ	いいと思いますよ	十分ですよ	これからですよ	大丈夫ですよ	仕方ないですよ
BERT	0.470	0.411	0.191	0.508	0.479	0.056
RoBERTa	0.508	0.444	0.295	0.651	0.632	0.098
DeBERTa-v2	0.420	0.470	0.247	0.598	0.589	0.048
GPT2	0.430	0.338	0.186	0.509	0.542	0.160

えておらず，分類は困難であったと考えられる．

4.5.4 分類結果の主観評価

4.5.2 節では，不同意応答タイミングに不同意応答表現をタグ付けしたコーパスを正解データとし，それとどの程度一致するかという観点から，不同意応答表現への分類を一定の水準で実現できること，並びに，不同意応答コーパスの利用の効果を確認した．本節では，主観評価によって，不同意応答表現への分類における不同意応答コーパスの利用の有効性を評価する．

(i) 評価の概要

主観評価の手順の概略を図 4.8 と 4.9 に示す．評価には，不同意応答表現への分類の実験データを用いる．本手法が分類した不同意応答表現を実際に生成し，それを被

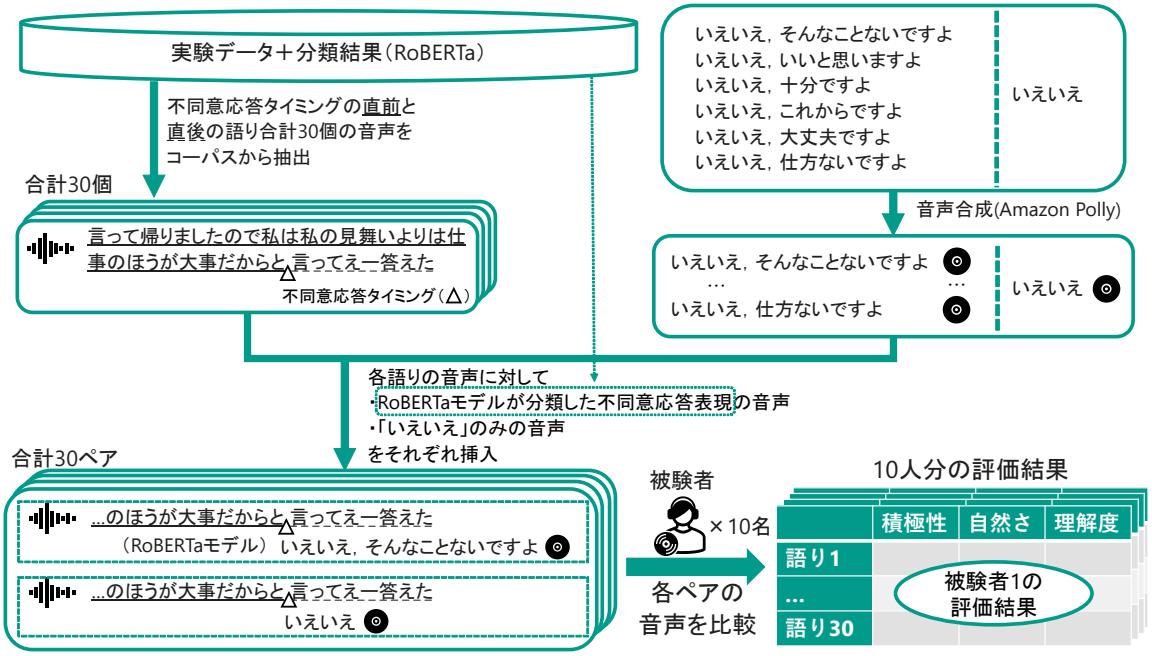


図 4.8: 不同意応答表現への分類結果の主観評価の手順の概略（「いえいえ」との比較）

験者が主観評価する[12]．被験者は，表 4.10 の 6 つの不同意応答表現のいずれかが挿入された語りの音声を聴取し，その印象を評価する．本手法は，語りの文字列を用いて，不同意応答タイミングを不同意応答表現に分類する．不同意応答表現の挿入位置は，不同意応答タイミング直前の最終形態素の発声終了時刻の 200 ミリ秒後とした．分類には，表 4.11 の正解率が最も高かった RoBERTa による手法を用いた．

評価は，表 4.8 の積極性，自然さ，理解度の 3 つの項目で実施した．本手法が分類した不同意応答表現とは別の応答が挿入された音声と比較する．比較音声として，下記の 2 つの音声を用意した．

(1) 不同意応答「いえいえ」を挿入した音声

(2) 表 4.10 の 6 つの不同意応答表現から，本手法が分類した表現を除いた 5 種類の表現のいずれかをランダムに挿入した音声

すなわち被験者は，本手法が分類した不同意応答表現が挿入された音声と，上記の 2 つの比較音声のいずれかとの比較をし，積極性，自然さ，理解度の 3 つの項目で，それぞれどちらが優れていたかを評価する．不同意応答「いえいえ」と表 4.10 の 6 つの

[12]一方で，正解の不同意応答タイミング以外で不同意応答表現を生成した場合についても，同様に主観評価することは有意義である．不同意応答表現が生成されないときの損失の程度を被験者実験により測定することが考えられ，その実施は今後の課題である．

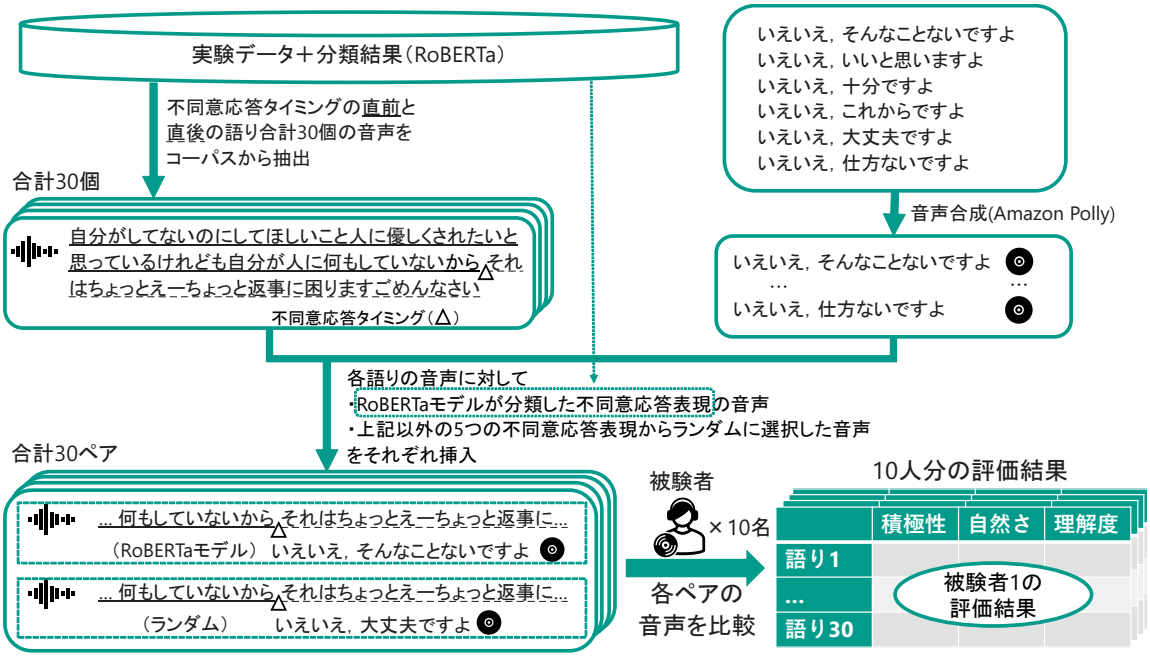


図 4.9: 不同意応答表現への分類結果の主観評価の手順の概略（ランダムとの比較）

不同意応答表現の音声合成には，Amazon Polly を用いた．音声合成の際には，応答の音声が自然となるよう，アクセント，話速，発声の強度を調整した．被験者が聴取する語りの音声の範囲は，不同意応答タイミングの直前 5 つ目の節から直後 2 つ目の節までとし，コーパスから抽出した人間の語り手の音声を用いた．

テストに用いた実験データから 60 個をランダムに抽出して主観評価に用いた．60 個のうち，30 個は不同意応答「いえいえ」を挿入した音声との比較，残りの 30 個は表 4.10 の 6 つの不同意応答表現のいずれかをランダムに挿入した音声との比較に用いた．また，主観評価は，30 代から 50 代の被験者 10 名が各々実施した．

(ii)　「いえいえ」との比較結果

本手法が分類した不同意応答表現が挿入された音声と，不同意応答「いえいえ」が挿入された音声との比較結果について述べる．図 4.10 に，被験者が本手法が分類した不同意応答表現の方が優れていると判定したペアの割合の分布を示す．被験者 10 人中，積極性では 10 人，自然さでは 5 人，理解度では 8 人が，本手法が分類した不同意応答表現の方が優れていると判定した回数が，全体の半数以上に達した．積極性と理解度において，本手法が分類した不同意応答表現の方を高く評価する傾向にある被験者の割合が高く，不同意応答コーパスに基づく本手法の有効性を確認できた．特に，

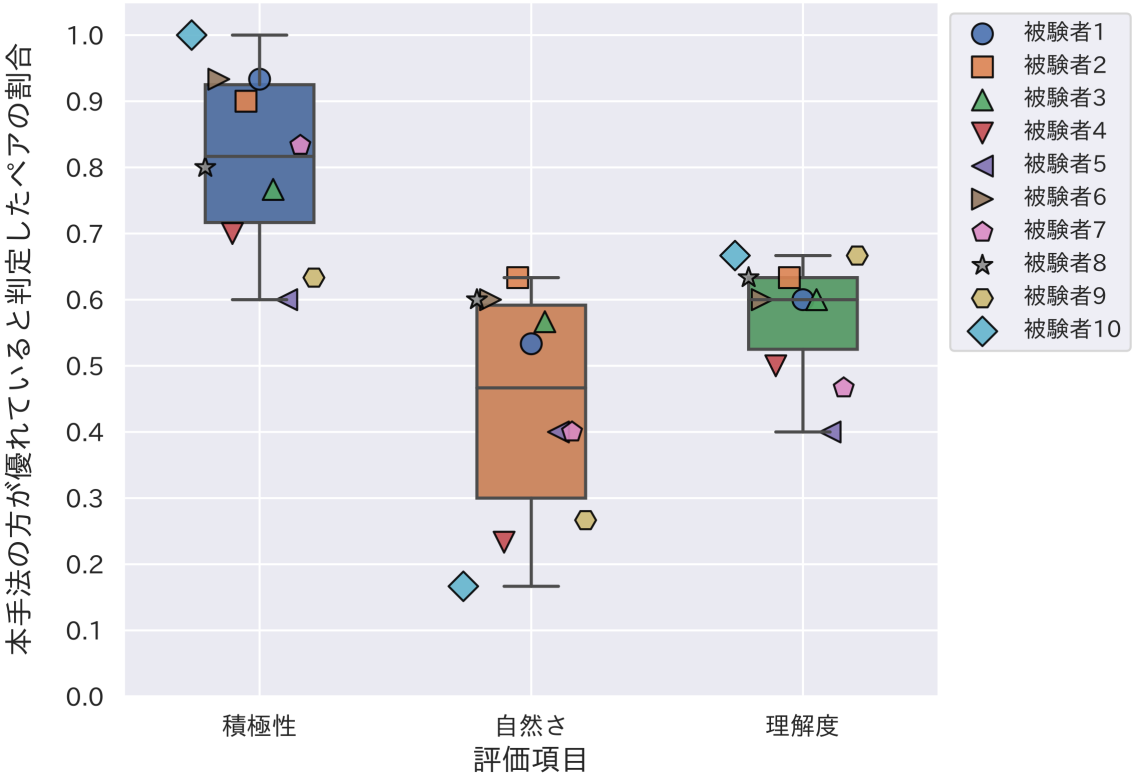


図 4.10: 「いえいえ」との比較において本手法が分類した不同意応答表現の方が優れていると被験者が判定したペアの割合の分布

積極性における割合は極めて高かったが，一方で，自然さにおける割合は相対的に低かった．この原因の 1 つとして，付加表現を伴う不同意応答は,「いえいえ」よりも長い応答となるため，合成音声の不自然さに意識が向きやすかったものと考えられる．これは，音声合成の性能向上によって改善が期待される．また，長く応答することに伴い,「いえいえ」よりも語りとの重複が発生しやすくなることも，自然さが低かった原因として考えられる．本コーパスと本手法では，語りの音響的な情報を考慮していない．例えば，語りの間の長さを考慮した不同意応答コーパスや不同意応答生成手法の実現によって，語りとの重複を解消することが期待できる．これらについても今後検討していきたい．

図 4.11 に，本手法が分類した不同意応答表現の方が優れていると判定した被験者数ごとのペア数の分布を示す．本手法が分類した不同意応答表現の方が優れていると判定した被験者の数について，積極性では 10 名，自然さでは 4 名，理解度では 3 名または 5 名のペアがもっと多かった．評価項目ごとに分布の形状は異なるものの，各評価項目において，本手法が分類した不同意応答表現の方が優れていると判定した被

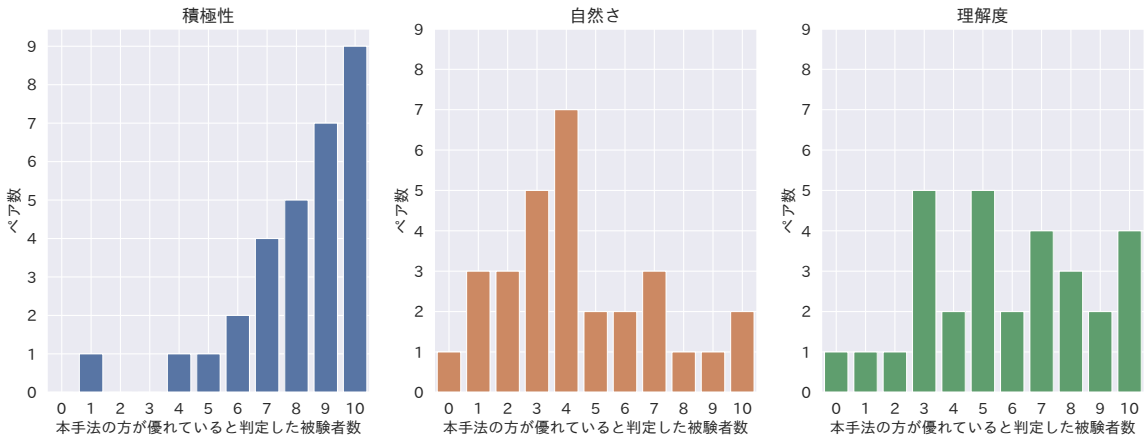


図 4.11: 「いえいえ」との比較において本手法が分類した不同意応答表現の方が優れていると判定した被験者数ごとのペア数の分布

験者が多いペアが存在していた．このことから，本手法に基づいて付加表現を伴う不同意応答を生成することで，より好印象を獲得できる場面の存在を確認できた．

表 4.14 に，被験者による主観評価の例を示す．不同意応答とその周辺の語りの列については，＜＞で囲んだ文字列が不同意応答であり，その他の文字列が語りである．積極性，自然さ，理解度の数値は，本手法が分類した不同意応答表現の方を優れていると判定した被験者数である．例 1 は，全ての評価項目で，被験者からの評価が相対的に高かった例である．語り手の申し訳なさを感じさせる発話に対して，本手法によって「いえいえ，大丈夫ですよ」と適切に応答を生成できたため，被験者からの評価が高かったものと考えられる．例 2 は，自然さと理解度の評価が相対的に低かった例である．この例では，語り手が自身の経験を自虐を交えて話しており，謝罪の発話や例 1 のような申し訳なさを感じさせる発話とは異なる．そのため，本手法によって生成された「いえいえ，大丈夫ですよ」は語りの文脈に合わず，評価が低かったものと考えられる．

(iii)　ランダムな分類結果との比較結果

本手法が分類した不同意応答表現が挿入された音声と，ランダムに分類した不同意応答表現が挿入された音声との比較結果について述べる．図 4.12 に，被験者が本手法が分類した不同意応答表現の方が優れていると判定したペアの割合の分布を示す．積極性，自然さ，理解度のすべての評価項目において，10 名の被験者全員が，本手法が分類した不同意応答表現の方が優れていると判定した回数の方が多かった．いずれの

表 4.14: 「いえいえ」との比較における不同意応答表現への分類結果に対する主観評価の例

	不同意応答とその周辺の語り	積極性	自然さ	理解度
例 1	地下鉄からえーと今度はあのーにしえええーと都庁前で乗り換えてこちらの駅までやって来ましたえーとその駅の名前をいうのはちょっと忘れちゃいましたですけども ＜いえいえ，大丈夫ですよ＞（本手法） ＜いえいえ，大丈夫ですよ＞（正解） それからま駅から歩いて来まして	10	6	9
例 2	風呂屋のペンキ画もう全く芸術性のない風呂屋のペンキ画はそれなりにいいんですけどもえー自分としてもなんじゃこれはと ＜いえいえ，大丈夫ですよ＞（本手法） ＜いえいえ，そんなことないですよ＞（正解） 小学校の頃のがよっぽど良かったなっていうんでまーそこで	9	1	3

評価項目においても，本手法が分類した不同意応答表現の方を高く評価する傾向にある被験者の割合が高く，不同意応答コーパスに基づく本手法の有効性を確認できた．

図 4.13 に，本手法が分類した不同意応答表現の方が優れていると判定した被験者数ごとのペア数の分布を示す．積極性，自然さ，理解度ともに，本手法が分類した不同意応答表現の方が優れていると判定した被験者が 10 名であるペアが，最も多かった．各評価項目において，本手法が分類した不同意応答表現の方が優れていると判定した被験者が多いペアが存在していた．このことから，本手法に基づいて付加表現を伴う不同意応答を生成することで，より好印象を獲得できた場面が存在することを確認した．

表 4.15 と 4.16 に，被験者による主観評価の例を示す．不同意応答とその周辺の語りの列については，＜＞で囲んだ文字列が不同意応答であり，その他の文字列が語りである．積極性，自然さ，理解度の数値は，本手法が分類した不同意応答表現の方を

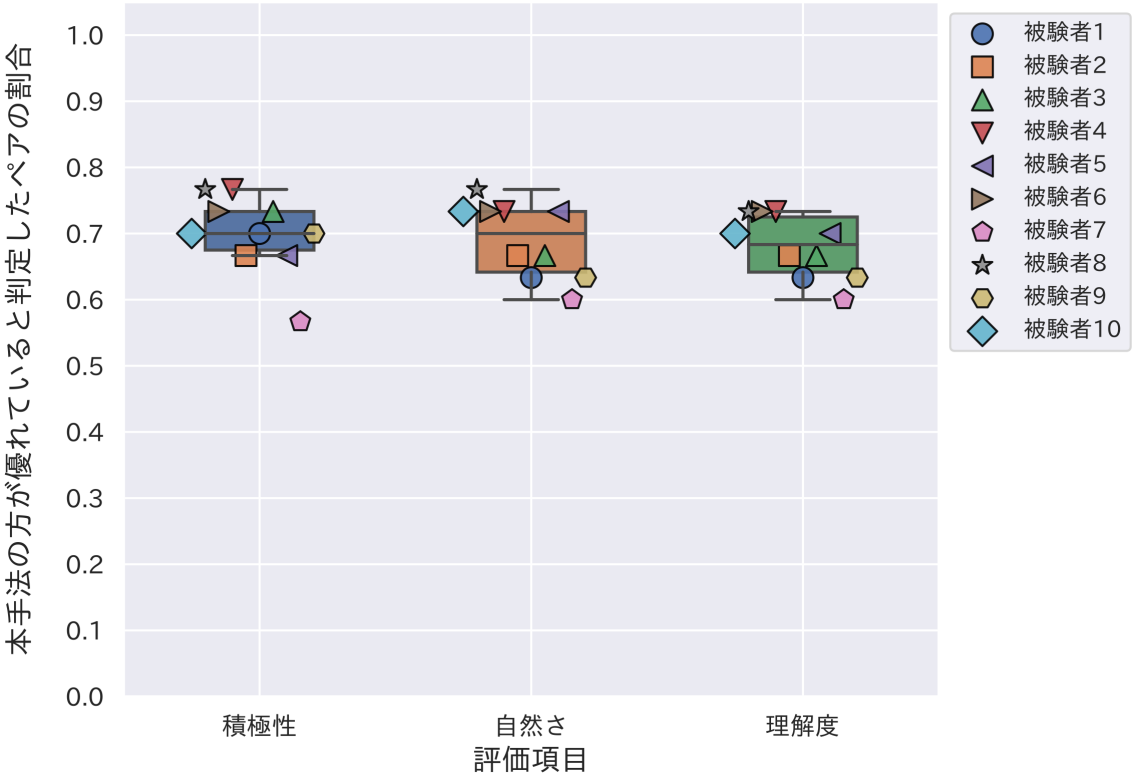


図 4.12: ランダムな分類結果との比較において本手法が分類した不同意応答表現の方が優れていると被験者が判定したペアの割合の分布

優れていると判定した被験者数である．例 1 は，全ての評価項目で，被験者からの評価が高かった例である．この例では，語り手が自身の趣味について部分的に自虐を交えて話しており，本手法では「いえいえ，いいと思いますよ」と適切に応答を生成できたため，被験者からの評価が高かったものと考えられる．例 2 では，語り手が過去の経験について謙遜を交えて話しており，自責，後悔，無念などを感じさせる語りの発話とはいえない．そのため，本手法によって生成された「いえいえ，仕方ないですよ」は語りの文脈に合わないといえる一方，ランダムな分類が正解と偶然一致しており，本手法が分類した不同意応答表現の方が優れていると判定した被験者数が 0 になったものと考えられる．

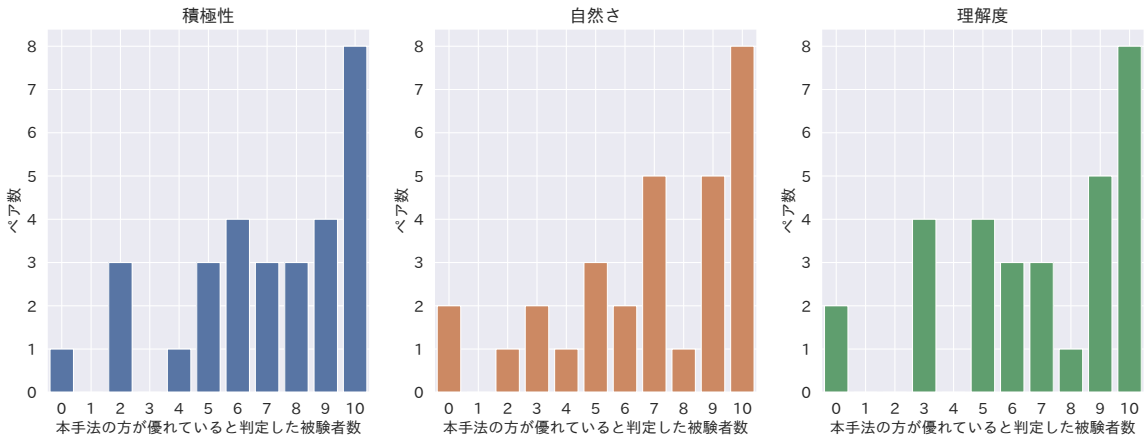


図 4.13: ランダムな分類結果との比較において本手法が分類した不同意応答表現の方が優れていると判定した被験者数ごとのペア数の分布

表 4.15: ランダムな分類結果との比較における不同意応答表現への分類結果に対する主観評価の例 1

	不同意応答とその周辺の語り	積極性	自然さ	理解度
例 1	歌っていうかねして歌っておりますあとはまー麻雀もねあのー頭の体操にもなるんでねえー仲間もね馬鹿話しながらね ＜いえいえ，いいと思いますよ＞（本手法） ＜いえいえ，これからですよ＞（ランダム） ＜いえいえ，いいと思いますよ＞（正解） あのー計算しながらやっておりますので	10	10	10

4.6 おわりに

本章では，会話エージェントなどによる語りの傾聴において，自虐や謙遜などの発話に対しては不同意応答を確実に表出できるべきである点に着目し，語りの傾聴における不同意応答生成の実現性を検証した．このために，本研究ではまず，高齢者の語りデータに対して，不同意応答タイミングと不同意応答表現をタグ付けし，応答コーパスを作成した．評価の結果，不同意応答タイミングが網羅的に，不同意応答表現が安定的に付与された応答コーパスを作成できることを確認した．次に，作成した応答

表 4.16: ランダムな分類結果との比較における不同意応答表現への分類結果に対する主観評価の例 2

	不同意応答とその周辺の語り	積極性	自然さ	理解度
例 2	そういう大学だったせいでおかげで別に中学高校からやってなくてもえー主将になれたんだと思うんですけども ＜いえいえ，仕方ないですよ＞（本手法） ＜いえいえ，そんなことないですよ＞（ランダム） ＜いえいえ，そんなことないですよ＞（正解） まーなんていうかにん人生って時々こうビューって自分の思っている以上に	0	0	0

コーパスを用いて，不同意応答タイミングの検出実験を実施した．検出手法は，事前学習済みの Transformer ベースのモデルを単純に fine-tuning することで実装した．実験の結果，不同意応答タイミングの検出が，一定の水準で可能であることが示された．同様にして，不同意応答表現への分類実験を行い，その分類の実現性を示した．

第5章　あとがき

5.1　本論文のまとめ

本論文では，語りを傾聴する会話エージェントの実現に向けた，傾聴応答コーパスの構築，傾聴応答の表出されやすさの推定，不同意応答の生成について述べた．

第 1 章では，語りに傾聴を示す発話である傾聴応答を生成するために解決すべき問題点として，傾聴応答タイミングの検出と傾聴応答表現の選定の 2 点を挙げたのち，傾聴応答の生成に関する研究動向を概観した．これまでの研究では，主に，典型的な傾聴応答である相槌を対象に，その生成タイミングの検出が行われており，ヒューリスティックなルールに基づくアプローチから，データに基づくアプローチまで，幅広く研究開発が進められてきたことを述べた．一方で，傾聴応答には相槌以外の応答も存在するものの，それらの生成タイミングの検出については，十分に解決されていないことを示した．また，傾聴応答を生成する際には，その応答表現のバリエーションが単調であることは好ましくないものの，傾聴応答表現の選定に関する試みは相対的に少ないことを指摘した．このような現状を踏まえ，本研究では，相槌以外も対象とした傾聴応答コーパスの構築，傾聴応答の表出されやすさの推定，傾聴応答の一種である不同意応答の生成について議論した．

第 2 章では，傾聴応答の生成手法の開発に利用可能な傾聴応答コーパスについて述べた．傾聴応答コーパスは，高齢者のナラティブコーパスに含まれる語りに対して，傾聴応答のタイミングと表現を付与することにより構築した．傾聴応答コーパスの特徴は，同一の語りに対して，11 名の聴き手役の作業者の傾聴応答が付与されていることである．合計して，148,962 個の傾聴応答が収録されている．また，収録されている傾聴応答には，その機能の観点から全 16 種類のうちのいずれかの種類が人手で付与されており，全体の 67.96%が相槌であり，残りの 32.04%が相槌以外の傾聴応答であった．相槌以外の傾聴応答は，感心，繰り返し，評価の順に多く出現していた．傾聴応答コーパスの評価と分析の結果，収録した傾聴応答の多頻度性，多様性，網羅性，自然さを確認した．最後に，傾聴応答コーパスを用いて，傾聴応答タイミングの検出実験を行い，その結果を報告した．

第 3 章では，傾聴応答のタイミング検出の要素技術である，傾聴応答の表出されやすさの推定手法を提案した．本研究ではまず，傾聴応答の表出されやすさとして，所与のタイミングにおいて傾聴応答を表出するであろう聴き手の割合である表出率を導入した．表出率が高いタイミングでは多数の聴き手が，表出率が低いタイミングでは少数の聴き手が，それぞれ傾聴応答を表出する．表出率は，語り手の嗜好などと併せて，傾聴応答の生成タイミングの検出に利用可能な要素である．次に，表出率の予測手法について説明した．本手法では，まず，表出率の予測タイミング直前の語りの音響情報と言語情報を，それぞれ Transformer エンコーダで処理する．音響情報としては，MFCC，ピッチ，パワーを，言語情報としてはサブワード化された語りのトークン系列を用いる．次に，エンコードされた音響情報のベクトルと言語情報のベクトルを連結したものを，出力層に入力することで，表出率を出力する．傾聴応答コーパスを用いた表出率の予測実験を行い，音響情報と言語情報の両方を用いる本手法の有効性を確認した．また，表出率を利用して傾聴応答タイミングを検出する実験を行い，表出率の予測の有用性を議論した．

第 4 章では，傾聴応答の 1 つである不同意応答の生成の実現性について述べた．本章ではまず，語りの傾聴における不同意について論じた．一般に，語りの傾聴においては，語り手の発話の内容をそのまま受容することが基本であるが，自虐や謙遜が含まれる発話などに対して，必ずしもその内容を受容することは適切ではない．そのような場合には，あえて語りに不同意を示すことが効果的である．次に，不同意応答の生成に利用可能なコーパスの作成について述べた．本研究では，高齢者のナラティブコーパスに含まれる語りに対して，不同意応答の生成に適したタイミングと，そのタイミングで生成すべき不同意応答表現をタグ付けた．不同意応答の対象となる語りの発話の性質に基づき，タグ付けする不同意応答表現の候補を 6 種類定めた．作成した不同意応答コーパスを評価した結果，不同意応答タイミングが網羅的に，不同意応答表現が安定的に付与されていることを確認した．最後に，不同意応答タイミングの検出実験と，不同意応答表現の分類実験について報告した．不同意応答コーパスを用いて，事前学習済みモデルを fine-tuning することで，不同意応答タイミングの検出モデルと，不同意応答表現の分類モデルを実装した．実装したモデルに対しては，不同意応答コーパスとの一致度に基づく評価を行うとともに，その性能に対する主観評価を実施した．

5.2 今後の課題と将来への展望

本論文が残した課題と将来への展望を述べる．
傾聴応答のコーパスの作成に関しては，以下の研究課題が挙げられる．

- 語りデータの多様化
  本研究では，高齢者のナラティブコーパスに対する傾聴応答を収集した．語り手の世代に応じて，語りの内容や使用される語句等の傾向は異なるため，有効な傾聴応答の生成方略も異なる可能性がある．日本語の独話が収録されているコーパスとしては，CSJ コーパス [90] が挙げられる．今後は，これらの利用も含めて，高齢者以外の語りに対しても，傾聴応答の収集と分析を進めていきたい．

- 視覚情報の考慮
  本研究の傾聴応答コーパスは，収録済みの語りの音声データの再生に同期して，聴き手役の作業者が傾聴応答を発声することで構築されている．傾聴応答を生成するか否かには，語り手の姿勢や目線，ジェスチャなど，視覚から得られる情報も有力であると考えられる．視覚情報を考慮した相槌生成に利用可能なコーパスとしては，千葉大学 3 人会話コーパス [91] が挙げられる．このコーパスは，人間どうしの会話を各自が装着したヘッドセットから収録するとともに，各自の正面と全員が映る外側の位置の 4 箇所から動画の撮影を行うことで構築されている．今後は，語りのビデオデータの再生に同期して，聴き手役の作業者が傾聴応答を発声するなどの方法で，本研究の傾聴応答コーパスに視覚情報を伴う拡張を施すことを検討したい．

表出されやすさの推定に関しては，以下の研究課題が挙げられる．

- 表出率の応用
  本論文では，傾聴応答タイミング検出タスクを例に，表出率の有用性を議論した．実験においては，傾聴応答タイミング検出モデルの素性として表出されやすさを利用した．近年，ターゲットタスクでモデルを学習する前に，中間タスクと呼ばれる別のタスクでモデルを学習するという機械学習の枠組みの有効性が報告されている [92, 93]．中間タスクとして表出されやすさの推定タスクを導入し，モデルに傾聴応答の表出されやすさの特徴を学習させたのち，個別の語り手用にチューニングすることで，後段の学習では語り手の個性や嗜好に着目した学習を効率的に行えるものと期待される．

不同意応答の生成に関しては，以下の研究課題が挙げられる．

- 不同意応答生成の性能改善
  本研究で実装した不同意応答の生成手法は，事前学習済みモデルを不同意応答コーパスで fine-tunign した単純なものであった．今後の課題として，不同意応答とその対象となる語りの特徴を踏まえた生成手法の提案が挙げられる．傾聴応答の生成タイミングの検出に関して，先行する語りの発話の極性（ポジティブ/ネガティブ）の予測をサブタスクとして導入してモデルを学習する試みが存在する [34]．不同意応答生成の性能改善に向けて，極性の利用を検討したい．

- 語りの音声を踏まえた不同意応答コーパスの作成と不同意応答生成
  本研究の不同意応答コーパスは，語りのテキストに対して，不同意応答の生成に適したタイミングと，生成するに相応しい応答表現をタグ付けることで作成されている．コーパスの作成にあたっては，タグ付けの作業者は，語りの音声の聴取を行っていない．しかし，ピッチやパワー，ポーズなど，語りの音響的な情報も不同意応答の生成に関わる有力な情報であるものと考えられる．今後は，語りの音声を踏まえた不同意応答コーパスの作成と不同意応答生成手法の開発に取り組みたい．

さらに，傾聴応答の生成に関しては，以下の研究課題が挙げられる．

- リアルタイム性の考慮
  本研究における表出されやすさの推定と不同意応答の生成では，語りの文字列を用いた．しかし，本研究が目指す会話エージェントを実世界で運用する際には，語りの文字列は音声認識を経て得られるものであり，その取得までにはタイムラグが生じる．本研究ではこのようなタイムラグの考慮をしていなかったが，所与のタイミングでの振る舞いの決定においては，利用可能な情報に制限を設ける必要がある．また，リアルタイム性に関しては，各種モデルの計算量の考慮も必要となる．本研究では，Transformer ベースのモデルを用いて実験に関する実装を行った．Transformer については，モデルの軽量化や計算量の削減に関する研究 [94, 95] が行われている．さらに近年では，相槌のタイミング検出やターンテイキングの予測を連続的に行うことができる VAP (Voice Activity Projection) モデルも提案されている [96]．VAP モデルは，対話参与者の音声信号を用いて，自己教師あり学習された Transformer ベースのモデルである．VAP モデルが韻律情報をどのように学習および利用しているのかの分析や [97]，CPU

環境における VAP モデルのリアルタイム処理を実験的に評価する試み [98] も進められている．リアルタイムなエージェントの振る舞いの決定に向けては，これらの技術の利用も検討したい．

- 傾聴応答のバリエーションの拡大
  傾聴応答の生成においては，その表現が多様であることが望まれる．相槌や繰り返し，評価については，比較的研究が盛んであるものの，その他の種類については，その生成に関する研究は多くない．このような状況を踏まえて，本研究では，計 16 種類の傾聴応答が収録されたコーパスを構築し，そのうち，不同意応答の生成について研究を行った．今後は，補完や言い換えなどの傾聴応答についても，その生成を検討していきたい．

- 多人数会話への対応
  本研究を含め，これまでの傾聴応答の生成に関する研究は，語り手と聴き手の 1 対 1 の会話を対象としたものがほとんどであった．しかし，近年は，会話への参与者が 3 名であるような場合での傾聴応答の生成に関する研究が進められている [99, 100]．多人数会話においては，会話エージェントは，話し手に引き続き語りを継続してもらうよう傾聴態度を示すだけでなく，聴き手に回っている会話参与者への発話を促すといった行為も求められる．多人数会話を対象とした傾聴応答の生成については，今後の研究の進展が待たれる状況にある．

- 応答系列と語りの展開の考慮
  本論文で述べた表出されやすさの推定手法や不同意応答の生成手法では，直前の語りから抽出される情報のみを利用している．相槌タイミングの検出に関する従来研究では，所与のタイミング以前にどのように応答したのかを表す応答履歴も用いられている [36, 40, 58]．また，今後の応答生成の見込みを踏まえて，現在の応答生成に関する振る舞いを決定する試みも存在する [101, 102]．今後は，これらも考慮して，表出率や不同意応答の生成に取り組みたい．意見や助言，質問に近いような傾聴応答の表出は，語りの発話を聴き終えてから表出することが好ましいとされており [16]，これらを適切に生成するには，語りの展開の予測が有効であると考えられる．日本語の独話文については，文節境界ごとに残存文長（文境界までの文節数）を推定する研究が行われており [103]，このような研究と組み合わせることでも，応答生成タイミングの検出性能の向上を期待できる．

- 感情の考慮
  傾聴を適切に遂行するには，感情に関する知識を持ち，人の感情を理解して，適切に活用できる力である，感情リテラシーが求められる [16]．そのための感情理解に関するモデルの 1 つに，プルチックの感情の輪 [104] が存在する．今後は，日本語におけるプルチックの 8 感情に関する注釈付きデータである WRIME [105] などを利用し，感情を考慮した傾聴応答の生成に取り組みたい．

将来の展望としては，人間の聴き手を代替可能な聴き役エージェントの開発が挙げられる．語ることは人間の基本的な欲求であり，人間は語るという行為を通して，自己肯定感を向上させ，ストレスを軽減し，自身の思考を整理することができる．このような聴き役エージェントが実現できれば，独居高齢者などの孤立している人であっても，語りたいときに自由に語ることが可能となる．また，語りたい内容によっては，たとえ人間の聴き手が身近にいたとしても，その人には語りづらく，聴き役エージェントを相手にした方が語りやすいこともあると考えられる．その実現のためには，本論文で述べたような傾聴応答の生成が不可欠である．さらに，音声会話というコミュニケーション形態に着目すると，音声認識技術や音声合成技術の向上も必要になる．本論文では，主に自然言語処理の観点から，傾聴応答の生成について論じたが，人間の聴き手を代替可能な聴き役エージェントの実現に向けては，言語，音声，視覚などの複数のモダリティを考慮することが必要であるといえる．

謝辞

本研究を進め，まとめるにあたり，日頃から様々な懇切丁寧な御指導と御鞭撻を賜りました，名古屋大学教授の松原茂樹先生に厚く感謝いたします．松原茂樹先生には，公私共に様々な御相談に乗っていただきました．心より感謝申し上げます．

本論文をまとめるにあたり，多大な御教示と御尽力をいただきました，名古屋大学教授の外山勝彦先生，東中竜一郎先生，並びに，名古屋工業大学教授の加藤昇平先生に深く感謝いたします．

本研究を進め，まとめるにあたり，多大な御示唆と御助言をいただきました，東京電機大学教授の大野誠寛先生に深く感謝いたします．大野誠寛先生には，本研究の初期の段階より有意義な御議論をいただきました．心より感謝の気持ちを申し上げます．

本研究の初期の段階より幅広い角度から御指導と御議論をいただきました，豊田工業高等専門学校准教授の村田匡輝先生に厚く御礼申し上げます．

名古屋大学松原研究室秘書の土井ひとみさんには，研究室での活動のサポートから，出張や事務手続き等，大変お世話になりました．心より感謝いたします．

研究に関する討論をはじめ，研究以外の面でもいろいろお世話になりました，松原研究室の皆様に深い敬意を表します．

高齢者のナラティブコーパスを提供いただいた，奈良先端科学技術大学院大学教授の荒牧英治先生，並びに，ソーシャル・コンピューティング研究室の皆様に，厚く御礼申し上げます．

本研究では，ナラティブコーパスに対する傾聴応答の収録，収録した傾聴応答の書き起こし，傾聴応答の分類等の作業を経て，傾聴応答コーパスを作成しました．また，収集した傾聴応答の自然さの主観評価を実施しました．本コーパスの作成及び評価に携わった方々に感謝いたします．

本研究では，ナラティブコーパスに対して，不同意応答のタイミングと応答表現を付与することで，不同意応答コーパスを作成しました．不同意応答コーパスの作成に携わった方々に謝意を表します．また，本研究では，不同意応答の生成結果に対する主観評価を実施しました．主観評価の被験者の方々に感謝いたします．

本研究を進めるにあたり，名古屋大学のスーパーコンピュータ「不老」を利用しま

した．管理及び運営に携わっている皆様に感謝の意を表します．

最後に，改めまして，本論文をまとめるにあたり御支援をいただいたすべての皆様に心より御礼申し上げます．

発表文献リスト

種別	論文名	関連する章
国際会議	Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara. Construction of Responsive Utterance Corpus for Attentive Listening Response Production, In *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC-2022)*, pp.7244-7252, 2022.	2 章
国際会議	Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara. Relation between Degree of Empathy for Narrative Speech and Type of Responsive Utterance in Attentive Listening, In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC-2020)*, pp.696–701, 2020.	2 章
論文誌	伊藤滉一朗，村田匡輝，大野誠寛，松原茂樹．語りに傾聴を示す応答の表出されやすさの推定，電子情報通信学会論文誌，Vol.J106-A, No.3, pp.136–145, 2023.	3 章
国際会議	Koichiro Ito, Masaki Murata, Tomohiro Ohno, Shigeki Matsubara. Estimating the Generation Timing of Responsive Utterances by Active Listeners of Spoken Narratives, In *Proceedings of the 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU-2021)*, pp.495–502, 2021.	3 章

種別	論文名	関連する章
論文誌	伊藤滉一朗，村田匡輝，大野誠寛，松原茂樹．語りの傾聴において不同意を示す応答の生成，自然言語処理，Vol.31, No.1, 2024.	4 章

参考文献

[1] Carl R. Rogers. *Client-Centered Therapy: Its Current Practice, Implications, and Theory*. Houghton Mifflin, 1951.

[2] Carl R. Rogers. *On Becoming a Person: A Therapist's View of Psychotherapy*. Houghton Mifflin, 1961.

[3] Michael White and David Epston. *Narrative Means to Therapeutic Ends*. W.W. Norton & Company, 1990.

[4] Labor Ministry of Health and Welfare. Summary report of comprehensive survey of living conditions 2019. `https://www.mhlw.go.jp/english/database/db-hss/dl/report_gaikyo_2019.pdf`, 2020.

[5] 内閣官房孤独・孤立対策担当室. 人々のつながりに関する基礎調査（令和 4 年）調査結果の概要. `https://www.cas.go.jp/jp/seisaku/kodoku_koritsu_taisaku/zittai_tyosa/r4_zenkoku_tyosa/tyosakekka_gaiyo.pdf`, 2023.

[6] Media Department for Culture and Sport (DCMS). Community life survey 2021/22: Wellbeing and loneliness. `https://www.gov.uk/government/statistics/community-life-survey-202122/community-life-survey-202122-wellbeing-and-loneliness`, 2023.

[7] William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria Luisa Gorno-Tempini, and Jennifer Ogar. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 27–37, 2014.

[8] Sweta Karlekar, Tong Niu, and Mohit Bansal. Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models. In *Proceedings*

of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 2, pp. 701–707, 2018.

[9] 柴田大作, 伊藤薫, 若宮翔子, 荒牧英治. 対照群付き高齢者コーパスの構築とそれを用いた認知症予備軍スクリーニング技術の開発. 人工知能学会論文誌, Vol. 34, No. 4, pp. B–J11_1–9, 2019.

[10] Jonathan Gratch, Ron Artstein, Gale Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, David Traum, Skip Rizzo, and Louis-Philippe Morency. The distress analysis interview corpus of human and computer interviews. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pp. 3123–3128, 2014.

[11] Alex Rinaldi, Jean Fox Tree, and Snigdha Chaturvedi. Predicting depression in screening interviews from latent categorization of interview prompts. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7–18, 2020.

[12] Chuyuan Li, Chloé Braud, and Maxime Amblard. Multi-task learning for depression detection in dialogs. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 68–75, 2022.

[13] 西尾実, 岩淵悦太郎, 水谷静夫, 柏野和佳子, 星野和子, 丸山直子. 岩波国語辞典第八版. 岩波書店, 2019.

[14] 山田忠雄, 倉持保男, 上野善道, 山田明雄, 井島正博, 笹原宏之. 新明解国語辞典第八版. 三省堂, 2020.

[15] Gerard Egan. *The Skilled Helper: A Problem-Management and Opportunity-Development Approach to Helping.* Brooks/Cole Cengage Learning, 2014.

[16] 大谷佳子. 対人援助の現場で使える 傾聴する・受けとめる技術 便利帳. 翔泳社, 2023.

[17] Senko K. Maynard. Conversation management in contrast: Listener response in Japanese and American English. *Journal of Pragmatics*, Vol. 14, No. 3, pp. 397–412, 1990.

[18] 細馬宏通, 富田彩加. うなずき運動とあいづちとの相互作用. 人工知能学会研究会資料 身体知研究会, Vol. SKL-09, pp. 13–18, 2011.

[19] Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, Vol. 57, pp. 233–243, 2014.

[20] Carlos Ishi Chaoran Liu and Hiroshi Ishiguro. Probabilistic nod generation model based on speech and estimated utterance categories. *Advanced Robotics*, Vol. 33, No. 15–16, pp. 731–741, 2019.

[21] 森大河, 伝康晴. 相槌の形態と頷きとの共起関係. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 90, pp. 140–145, 2020.

[22] 森大河, 伝康晴. 相槌の形態と頷きの物理的特徴の関係. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 91, pp. 62–67, 2021.

[23] Tomio Watanabe. A voice reaction system with a visualized response equivalent to nodding. *Advances in Human Factors/Ergonomics, A*, Vol. 12, pp. 396–403, 1989.

[24] Mutsuhiro Nakashige Tomio Watanabe, Masashi Okubo and Ryusei Danbara. Interactor: Speech-driven embodied interactive actor. *International Journal of Human–Computer Interaction*, Vol. 17, No. 1, pp. 43–60, 2004.

[25] Shinya Fujie, Yasushi Ejiri, Kei Nakajima, Yosuke Matsusaka, and Tetsunori Kobayashi. A conversation robot using head gesture recognition as paralinguistic information. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 159–164, 2004.

[26] 森大河, 伝康晴. 相槌の特徴に一致した頷き生成モデル. 人工知能学会論文誌, Vol. 37, No. 3, pp. IDS–H_1–12, 2022.

[27] 勝見久央, 井上昂治, 中村静, 高梨克也, 河原達也. 自律型アンドロイドによる対話における同調的笑いの生成. 情報処理学会研究報告, Vol. SLP-116, No. 4, pp. 1–6, 2017.

[28] Koji Inoue, Divesh Lala, and Tatsuya Kawahara. Can a robot laugh with you?: Shared laughter generation for empathetic spoken dialogue. *Frontiers in Robotics and AI*, Vol. Computational Intelligence in Robotics, pp. 1–11, 2022.

[29] 日本語記述文法研究会. 現代日本語文法 7. くろしお出版, 2009.

[30] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, Vol. 32, No. 8, pp. 1177–1207, 2000.

[31] Norihide Kitaoka, Masashi Takeuchi, Ryota Nishimura, and Seiji Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 20, No. 3, pp. 220–228, 2005.

[32] Takashi Yamaguchi, Koji Inoue, Koichiro Yoshino, Katsuya Takanashi, Nigel G. Ward, and Tatsuya Kawahara. Analysis and prediction of morphological patterns of backchannels for attentive listening agents. In *Proceedings of the 7th International Workshop on Spoken Dialogue Systems*, pp. 1–12, 2016.

[33] Kohei Hara, Koji Inoue, Katsuya Takanashi, and Tatsuya Kawahara. Prediction of turn-taking using multitask learning with prediction of backchannels and fillers. In *Proceedings of the Interspeech 2018*, pp. 991–995, 2018.

[34] Jin Yea Jang, San Kim, Minyoung Jung, Saim Shin, and Gahgene Gweon. BPM_MT: Enhanced backchannel prediction model using multi-task learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3447–3452, 2021.

[35] Ronald Poppe, Khiet P Truong, Dennis Reidsma, and Dirk Heylen. Backchannel strategies for artificial listeners. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, pp. 146–158, 2010.

[36] Nicola Cathcart, Jean Carletta, and Ewan Klein. A shallow model for backchannel continuers in spoken dialogue. In *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics*, pp. 51–58, 2003.

[37] Shinya Fujie, Kenta Fukushima, and Tetsunori Kobayashi. A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information. In *Proceedings of the 2nd International Conference on Autonomous Robots and Agents*, pp. 379–384, 2004.

[38] Hiroaki Noguchi and Yasuharu Den. Prosody-based detection of the context of backchannel responses. In *Proceedings of the 5th International Conference on Spoken Language Processing*, pp. 487–490, 1998.

[39] Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. A probabilistic multimodal approach for predicting listener backchannels. *Journal of Autonomous Agents and Multi-Agent Systems*, Vol. 20, No. 1, pp. 70–84, 2010.

[40] 大野誠寛, 神谷優貴, 松原茂樹. 対話コーパスを用いた相づち生成タイミングの検出. 電子情報通信学会論文誌, Vol. J100-A, No. 1, pp. 53–65, 2017.

[41] Robin Ruede, Markus Müller, Sebastian Stüker, and Alex Waibel. Enhancing backchannel prediction using word embeddings. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, pp. 879–883, 2017.

[42] Daniel Ortega, Chia-Yu Li, and Ngoc Thang Vu. Oh, jeez! or uh-huh? a listener-aware backchannel predictor on asr transcriptions. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8064–8068, 2020.

[43] Amalia Istiqlali Adiba, Takeshi Homma, and Toshinori Miyoshi. Towards immediate backchannel generation using attention-based early prediction model. In *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7408–7412, 2021.

[44] 室町俊貴, 狩野芳伸. 生成モデルによる傾聴応答タイミングの推定と動的 prompt-tune を用いた応答スタイルのパラメーター制御. 言語処理学会第 29 回年次大会発表論文集, pp. 3149–3154, 2023.

[45] 村田匡輝, 大野誠寛, 松原茂樹. 会話ロボットにおける繰り返し応答の生成. 第 14 回情報科学技術フォーラム講演論文集, No. 2, pp. 251–252, 2015.

[46] 伊藤滉一朗, 村田匡輝, 大野誠寛, 松原茂樹. 傾聴を示す応答で繰り返される語りの語句の検出. 言語処理学会第 25 回年次大会発表論文集, pp. 1316–1319, 2019.

[47] 石田真也, 井上昂治, 中村静, 高梨克也, 河原達也. 傾聴対話システムのための発話を促す聞き手応答の生成. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 77, pp. 1–6, 2016.

[48] 井上昂治, ラーラーディベッシュ, 山本賢太, 中村静, 高梨克也, 河原達也. アンドロイド ERICA の傾聴対話システム – 人間による傾聴との比較評価 –. 人工知能学会論文誌, Vol. 36, No. 5, pp. H–L51_1–12, 2021.

[49] Koichiro Yoshino and Tatsuya Kawahara. Conversational system for information navigation based on POMDP with user focus tracking. *Computer Speech & Language*, Vol. 34, No. 1, pp. 275–291, 2015.

[50] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–47, 2017.

[51] 田原俊一, 松本一則, 服部元. オウム返し応答を生成する対話システム. 第 20 回情報科学技術フォーラム講演論文集, No. 2, pp. 465–466, 2021.

[52] 吉井健治. カウンセリングの基本的技法: 相手のこころに近づく聴き方十二の技. 鳴門教育大学研究紀要, Vol. 30, pp. 41–51, 2015.

[53] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.

[54] 川本稔己, 長谷川駿, 上垣外英剛, 船越孝太郎, 奥村学. 傾聴の応答で繰り返される語句の検出性能の向上. 言語処理学会第 27 回年次大会発表論文集, pp. 1580–1584, 2021.

[55] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol. 12, No. 3, pp. 203–222, 2005.

[56] 国立国語研究所コーパス開発センター. 日本語アプレイザル評価表現辞書. https://www.gsk.or.jp/catalog/gsk2011-c/, 2012.

[57] 長岡技術科学大学自然言語処理研究室. SNOW D18: 日本語感情表現辞書. https://www.jnlp.org/GengoHouse/snow/d18, 2018.

[58] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Nigel G. Ward, 河原達也. 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成. 人工知能学会論文誌, Vol. 31, No. 4, pp. C–G31_1–10, 2016.

[59] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, pp. 5485–5551, 2020.

[60] Toshiki Kawamoto, Hidetaka Kamigaito, Kotaro Funakoshi, and Manabu Okumura. Generating repetitions with appropriate repeated words. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 852–859, 2022.

[61] 村田匡輝, 大野誠寛, 松原茂樹. 系列変換モデルに基づく傾聴的な応答表現の生成. 言語処理学会第 24 回年次大会発表論文集, pp. 821–824, 2018.

[62] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.

[63] Erik Ekstedt and Gabriel Skantze. TurnGPT: a transformer-based language model for predicting turn-taking in spoken dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2981–2990, 2020.

[64] 泉子・K・メイナード. 会話分析. くろしお出版, 1993.

[65] Patricia Clancy. Written and spoken style in Japanese narratives. *Spoken and Written Language: Exploring Orality and Literacy*, pp. 55–76, 1982.

[66] Senko K. Maynard. On back-channel behavior in Japanese and English casual conversation. *Linguistics*, Vol. 24, No. 6, pp. 1079–1108, 1986.

[67] Eiji Aramaki. Japanese elder's language index corpus v2. `https://figshare.com/articles/dataset/Japanese_Elder_s_Language_Index_Corpus_v2/2082706/1`, 2016.

[68] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, 2004.

[69] Akinobu Lee, Tatsuya Kawahara, and Kiyohiro Shikano. Julius — an open source real-time large vocabulary recognition engine. In *Proceedings of the 7th European Conference on Speech Communication and Technology*, pp. 1691–1694, 2001.

[70] Shoju Chiba Itsuko Fujimura and Mieko Ohso. Lexical and grammatical features of spoken and written Japanese in contrast: Exploring a lexical profiling approach to comparing spoken and written corpora. In *Proceedings of the 7th GSCP International Conference: Speech and Corpora*, pp. 393–398, 2012.

[71] Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. In *Proceedings of the 6th Conference on Natural Language Learning*, pp. 63–69, 2002.

[72] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, Vol. 30, pp. 5998–6008, 2017.

[73] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa:

A robustly optimized BERT pretraining approach, 2019. arXiv preprint arXiv:1907.11692.

[74] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced BERT with disentangled attention. In *Proceedings of the 9th International Conference Learning Representations*, pp. 1–21, 2021.

[75] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of the 7th International Conference Learning Representations*, pp. 1–8, 2019.

[76] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems*, Vol. 33, pp. 12449–12460, 2020.

[77] Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J. Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6558–6569, 2019.

[78] Paul Boersma and David Weenink. Praat: Doing phonetics by computer (version 6.0.37). `http://www.praat.org/`, 2018.

[79] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1715–1725, 2016.

[80] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 19th Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543, 2014.

[81] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[82] Evan James Williams. *Regression Analysis*, Vol. 14. Wiley, 1959.

[83] James H. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, Vol. 87, No. 2, pp. 245–251, 1980.

[84] 大谷佳子. 対人援助の現場で使える 聴く・伝える・共感する技術 便利帖. 翔泳社, 2017.

[85] 木山幸子. 日本語の雑談における不同意の相互作用 −「儀礼的不同意」に焦点を置いて−. 言語情報学研究報告，Vol. 9, pp. 251–265, 2006.

[86] 冨樫純一. 否定応答表現「いえ」「いいえ」「いや」. 矢澤真人・橋本修（編）, 現代日本語文法 現象と理論のインタラクション, pp. 23–46. ひつじ書房, 2006.

[87] 丸山岳彦, 柏岡秀紀, 熊野正, 田中英輝. 日本語節境界検出プログラム CBAP の開発と評価. 自然言語処理, Vol. 11, No. 3, pp. 39–68, 2004.

[88] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, 1960.

[89] Jean Carletta. Squibs and discussions assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, Vol. 22, No. 2, pp. 249–254, 1996.

[90] 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明. 日本語話し言葉コーパスの設計. 音声研究, Vol. 4, No. 2, pp. 51–61, 2000.

[91] Yasuharu Den and Mika Enomoto. A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. *Conversational Informatics: An Engineering Approach*, pp. 305–330, 2007.

[92] Jason Phang, Thibault Févry, and Samuel R. Bowman. Sentence encoders on stilts: Supplementary training on intermediate labeled-data tasks, 2019.

[93] Ting-Yun Chang and Chi-Jen Lu. Rethinking why intermediate-task fine-tuning works. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 706–713, 2021.

[94] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. In

Proceedings of the 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing, 2019.

[95] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. In *Proceedings of the 8th International Conference on Learning Representations*, 2020.

[96] Erik Ekstedt and Gabriel Skantze. Voice activity projection: Self-supervised learning of turn-taking events. In *Proceedings of the Interspeech 2022*, pp. 5190–5194, 2022.

[97] Erik Ekstedt and Gabriel Skantze. How much does prosody help turn-taking? Investigations using voice activity projection models. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 541–551, 2022.

[98] Koji Inoue, Bing'er Jiang, Erik Ekstedt, Tatsuya Kawahara, and Gabriel Skantze. Real-time and continuous turn-taking prediction using voice activity projection, 2024. arXiv preprint arXiv:2401.04868.

[99] Koji Inoue, Hiromi Sakamoto, Kenta Yamamoto, Divesh Lala, and Tatsuya Kawahara. A multi-party attentive listening robot which stimulates involvement from side participants. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 261–264, 2021.

[100] 森大河, 伝康晴, Kristiina Jokinen. 多人数会話におけるマルチモーダル聞き手反応予測. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 96, pp. 7–12, 2022.

[101] Matthew Roddy and Naomi Harte. Neural generation of dialogue response timings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2442–2452, 2020.

[102] Bing'er Jiang, Erik Ekstedt, and Gabriel Skantze. Response-conditioned turn-taking prediction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 12241–12248, 2023.

[103] 天暉河村, 誠寛大野, 茂樹松原. 漸進的な言語処理のための独話文に対する残存文長の推定. 情報処理学会第 82 回全国大会講演論文集, No. 2, pp. 447–448, 2020.

[104] Robert Plutchik. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, pp. 3–33. Academic Press, 1980.

[105] Haruya Suzuki, Yuto Miyauchi, Kazuki Akiyama, Tomoyuki Kajiwara, Takashi Ninomiya, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. A Japanese dataset for subjective and objective sentiment polarity classification in micro blog domain. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pp. 7022–7028, 2022.