対話における聞き手の理解度と
対話参加者の視覚的情報の関係の分析

令和６年度　卒業論文

日本大学　文理学部　情報科学科　宮田研究室

岡　哲平

概要

円滑な対話を行うために，聞き手が話し手の発言内容を理解していることを示すこと，話し手が聞き手の理解状況を把握することが重要である．限定されたシーンにおける個人の意図や心理状態と視覚的情報の関係性を分析する取り組みが行われているが，日常生活での対話における参加者の理解度と視覚的情報を分析する取り組みは行われていない．そのため，どのような視覚的情報が理解度と関係しているのか明らかになっていない．対話参加者の理解度と視覚的情報の関係性が明らかになると，話し手に理解を示すための行動や聞き手の理解状況を把握するための行動に役立つと考えられる．そこで我々は，対話における聞き手の理解度と対話参加者の視覚的情報の関係について分析を行う．具体的には，対話参加者の視覚的特徴を抽出し，聞き手の理解度との相関分析を行う．その結果，話し手と聞き手ともに笑顔に関連する表情の振舞いや頭部の前後の動き，視線の上下角度が聞き手の理解度と高い相関を示した．また，ネガティブな印象を与える振舞いは困惑や強調として機能し，結果的に理解を深める役割の一端を担っている可能性が示唆された．

目 次

第 1 章　序論 **1**
1.1　研究の背景 2
1.2　研究の目的 2
1.3　本論文の構成 2

第 2 章　関連研究 **4**
2.1　特定のタスクにおける理解度に関する研究事例 5
2.2　特定のタスクにおける視覚的情報を分析する研究事例 6

第 3 章　研究課題 **9**
3.1　問題の定義 10
3.2　研究課題の設定 10

第 4 章　対話コーパス **11**
4.1　2 者対話コーパス 12
4.2　理解度の評価 12
4.3　頷きの評価 12

第 5 章　理解度に応じた視覚的情報の分析 **14**
5.1　分析対象とする理解度について 15
5.2　分析対象とする頷きについて 15
5.3　視覚的特徴の抽出 15
5.4　分析 16
5.4.1　理解度と視覚的情報の相関分析 16
5.4.2　理解度の高群と低群における視覚的情報の比較 16

第 6 章　結果・考察 **20**
6.1　話し手の視覚的特徴量について 21
6.2　聞き手の視覚的特徴量について 22

第 7 章　結論 **23**

謝辞 **25**

参考文献 **27**

研究業績 **30**

図 目 次

4.1 各アノテータが評価した理解度の分布 13
4.2 アノテーションシステム 13

5.1 頷きの判定区間 16

表 目 次

5.1 Action Units の詳細 17
5.2 聞き手の理解度と視覚的特徴量の相関係数 18
5.3 聞き手の理解度と頷きの角度・回数の相関係数 18
5.4 話し手と聞き手それぞれで高い相関を示した特徴量と t 検定の結果 19

第1章　序論

1.1 研究の背景

円滑な対話を行うために，聞き手が話し手の発言内容を理解していると示すこと，話し手が聞き手の理解状況を把握することが重要である．話し手が聞き手は理解していると判断できる場面では，話し手は話を続けたり，次の話題に進めたりできる．一方で，話し手が聞き手は理解していないと判断できる場面では，話し手は同じ話題を繰り返したり，聞き手が理解できるように言い直したりする必要がある．しかし，コミュニケーションが苦手な人は，自身が聞き手となった時に相手に理解していると伝えられない可能性や，自身が話し手となった時に相手の理解状況を把握できていない可能性がある．そのため，対話が円滑に行われず，破綻してしまう問題が発生する．

1.2 研究の目的

円滑な対話を行うために，相手に理解していると伝えられるようにするための取り組みや理解状況を把握できるようにするための取り組みを行う必要がある．視覚的情報は個人の意図や心理状態を把握する際に手掛かりとなると考えられている [1][2][3][4][5]．現在，様々なシーンで個人の意図や心理状態と視覚的情報の関係を分析する取り組みが行われている．しかし，日常生活での対話における理解度と視覚的情報の関係を分析する取り組みは行われていない．これらの関係が明らかになると，聞き手の理解状況の把握に役立つと考えられる．また，多くの研究では，学習中の生徒や説明を受ける人など話を聞く側の視覚的情報に着目している．対話は話し手と聞き手の双方向コミュニケーションであるため，聞き手だけでなく話し手の視覚的情報にも着目する必要がある．そこで本研究では，対話における聞き手の理解度と対話参加者の視覚的情報の関係の分析を行う．具体的には，2 者対話データと第三者による聞き手の理解度の評価と聞き手の頷きの評価が記録されている対話コーパスから，対話参加者の視覚的特徴を抽出し，聞き手の理解度との相関分析や理解度が高いときと低いときでの各特徴量の値の比較を行う．本研究での視覚的情報とは，表情や視線，頭部の振舞いのことを指す．

1.3 本論文の構成

本論文の構成は次のとおりである．

2 章では，理解度に着目した研究事例と表情に着目した研究事例，頷きに着目した研究事例について述べる．
3 章では，本論文における問題の定義と研究課題について述べる．
4 章では，本論文で用いた対話コーパスについて述べる．
5 章では，聞き手の理解度と対話参加者の視覚的情報との関係の分析について述べる．

6 章では，聞き手の理解度と対話参加者の視覚的情報との関係の結果と考察について述べる．

最後に 7 章にて，本論文の結論を述べる．

第2章　関連研究

本研究では，聞き手の理解度の対話参加者の視覚的情報の関係の分析を行う．そこで本章では，理解度に着目した研究事例と表情に着目した研究事例，頷きに着目した研究事例について述べる．2.1 節では，特定のタスクにおける理解度に関する研究事例を紹介する．2.2 節では，特定のタスクにおける視覚的情報を分析する研究事例を紹介する..

2.1　特定のタスクにおける理解度に関する研究事例

本節では，特定のタスクにおける理解度に関する研究事例について述べる．

Mimura ら [6] は，被験者の問題を解く際の表情に基づき，三層ニューラルネットワークを利用して理解度を推定するシステムを提案している．実験では，20 名の被験者を対象に実施され，被験者はコンピュータ画面に提示された数学や地理の問題を解答するタスクを行った．被験者の表情はカメラで記録され，問題の「即座に理解」,「考え中」,「全く理解していない」という 3 つの異なる理解状態に対応する画像が収集された．顔の特徴点は，眉，目，口の主要領域に配置され，ニュートラルな表情画像とタスク中の画像を比較して表情情報が抽出された．システムは，三層ニューラルネットワークに特徴点データを入力し，バックプロパゲーションアルゴリズムによる学習を通じて理解度を 5 段階（全く理解していない〜完全に理解している）に分類した．実験結果として，平均 71.3 %の推定精度を達成しているが，被験者の表情の変化が小さい場合，推定精度が低下する問題があるということを示唆した．

Omata ら [7] は，e ラーニングにおいて学習者の視線追跡データと生理信号を比較し，理解度を推定するモデルの構築を行なっている．実験では，5 名の日本人大学生を対象に英語の読解課題を実施し，視線追跡データ（まばたき数，視線速度，注視時間など）と生理信号（脳波，血流，皮膚コンダクタンス，呼吸）を取得した．これらのデータを多層パーセプトロンによるニューラルネットワークで解析し，4 段階の理解度を推定した．その結果，視線速度や注視時間が理解度の主要な指標となる一方で，生理信号の低活動が理解状態と混乱状態の区別に寄与しないことが明らかとなった．

Clark ら [8] は，レゴモデルを組み立てるタスクでの，指示役と組み立て役の対話において，指示役が組み立て役の理解度をモニタリングし，必要に応じて発話内容を調整する仕組みを検討している.. 実験参加者は「指示役」と「組み立て役」に分かれ，指示役が組み立て役に指示を与えながら 10 種類のレゴモデルを組み立てるタスクを行った．顔や作業スペースが相互に見える場合，どちらかが見えない場合，録音された指示を用いる場合の 3 種類のやり取りを録画し，発話やジェスチャーの詳細な分析を行った．作業スペースが見える場合，ジェスチャーと視線を活用して相互理解が促進され，作業時間が短縮されることが確認された．作業スペースが見えない場合，発話が増加し，エラー率も増加した．録音された指示の場合では，エラー率がさらに増加した．視覚的なフィードバックはコミュニケーション効率を高める重要な要素であることを示した．

Olcay ら [9] は，ボードゲームのルールを説明するタスクでの説明者と被説明者の対話において，被説明者のマルチモーダル情報を利用し，理解している瞬間と理解していない瞬間を自動的に分類する機械学習モデルの構築・検証を行っている．ドイツ語で行われた

ボードゲームの説明を記録した MUNDEX コーパス [10] を使用し，理解・非理解の瞬間をビデオリコール手法でラベル付けし，4 秒間のウィンドウ内で頭部や手の動き，目の開閉度，音声の特性を解析した．その結果，非理解状態では頭部や手の動作が減少し，目の細めや頻繁なまばたきが確認され，これらの要素が理解状態を予測する重要な指標であることが示された．

2.2　特定のタスクにおける視覚的情報を分析する研究事例

本節では，特定のタスクにおける視覚的情報を分析する研究事例について述べる．

Ismail[11] は，教師と生徒の 1 対 1 の面談の際の教師の表情に基づき，教師の感情反応を解析する手法を提案している．研究手法として，1 対 1 のインタビューで収集したビデオデータを対象に，顔表情解析ソフトウェア (Sightcorp)[12] を用いて感情の変化を抽出した．顔の動作単位（Action Unit）を識別し，幸福，怒り，悲しみ，驚き，恐怖，嫌悪の 6 つの基本感情を定量的に測定した．さらに，データの信頼性を高めるため，手動のコーディングシートを用いて頻度，強度，持続時間，感情の肯定性または否定性を記録し，ソフトウェアの結果と統合した．その結果，教育現場における教員の表情変化が，学生との相互作用や教育手法の改善に寄与する可能性が示唆された．

Jiabei ら [13] は，表情認識技術を活用して，学生の集中度や理解度をリアルタイムで分析し，教師が個別化された指導を行えるシステムの構築を行なっている．既存の表情データベースを用いて表情認識手法の性能を比較し，ディープラーニング技術を用いたモデルが，単一画像や連続ビデオフレームの解析で高い精度を示すことを確認した．さらに，非理想的な環境（画像の遮蔽や低画質など）への対応能力を持つアルゴリズムが検証され，その利点と限界が議論された．その結果，学習感情分析モデルは，教師が学生の学習意欲や理解度に応じて教育方法を調整する際の有力なツールとなることが示された．

Whitehill ら [14] は，学生の顔表情を基にした自動的なエンゲージメント認識システムの構築を行なっている．実験では，34 人の大学生を対象に認知スキルトレーニングを実施した．実験中の表情を iPad で撮影し，表情データを収集後，人間観察者によるエンゲージメント評価データを用いて機械学習モデルを訓練した．認識システムは人間観察者と同等の精度を達成し，顔の静的特徴を用いた推定が有効であることが確認された．また，エンゲージメントとタスクパフォーマンスの相関が示され，学習セッション中のエンゲージメントが成績に影響することが示唆された．

Wang ら [15] は，e ラーニング環境で学習者がヒントを必要とするタイミングを顔表情から推定する手法を提案している．実験では，13 名の大学生を対象に，言語学オリンピアードの問題を使用した問題解決タスクを実施した．専用のウェブサイトを設計し，学習者がヒントボタンをクリックすることで，原理に基づいたヒントが表示される仕組みを導入した．実験中，学習者の顔表情をウェブカメラで記録し，OpenFace を使用して Action Units を抽出した．また，LightGBM と SVM の 2 つの機械学習モデルを構築し，学習者が「ヒントを求めている状態」と「問題解決中の状態」を分類するモデルを訓練した．その結果，LightGBM モデルは 85.0%の分類精度を達成し，重要な特徴として「眉を下げ

る（AU04）」,「唇を引き締める（AU23）」,「顎を開ける（AU26）」が特定された．また，LightGBM は SVM よりも訓練時間が短く，オンラインシステムへの実装に適していることが確認された．

Mioa ら [16] は，オンライン講義中の学生の表情を用いて，学生の集中度を推定する取り組みを行なっている．具体的には，学習者が講義に集中している際の注意水準を，タスク無関係な聴覚刺激への反応時間を指標に評価し，顔表情から反応時間を予測するモデルを構築した．．実験では，平均年齢 23.1 歳の 15 名の参加者を対象に，PHP の基礎を解説する YouTube 講義動画を視聴させた．参加者は動画視聴中に白色雑音の消失というタスク無関係な聴覚刺激に対してキーを押して応答し，その反応時間を記録した．顔画像データは，聴覚刺激の提示直前 3 秒間を解析対象として収集し，特徴量として Action Units の最小値，最大値，平均値，標準偏差などを抽出した．機械学習モデルを用いた反応時間の予測の結果，AU9（鼻をしわ寄せる），AU45（まばたき），AU15（唇の下げ），AU7（瞼の締まり）が反応時間予測に重要な特徴量であることが確認された．

Shujan ら [17] は，面接において顔表情解析ツールを用いて候補者の感情状態を数値化し，採用意思決定に影響を与える質問を特定する新たなアプローチを提案している．実験では，IT 業界のビジネスアナリストを対象にした模擬面接を実施し，12 名の業界専門家へのインタビューとオンライン調査を通じて求められる性格特性を特定し，36 件の模擬面接をビデオ録画してデータを収集した．顔表情解析により，候補者の 7 つの基本感情（喜び，怒り，軽蔑，嫌悪，恐怖，悲しみ，驚き）の変動を分析し，各質問に対する感情的関与を数値化した．その結果,「自分を説明する」「採用すべき理由を説明する」「他人との協調性を問う」といった感情的関与の高い質問が特定され，これらの質問が採用意思決定に大きく影響を与える可能性が示された．一方，感情変化が少ない質問では，採用意思決定に寄与する効果が低いことも明らかとなった．

Muralidhar ら [18] は，視線や表情と面接とホテル受付の 2 つの異なる職場シナリオにおける採用可能性や顧客対応能力との関係について分析している．338 本のビデオデータ（面接：169 本，ホテル受付：169 本）を収集し，視線データを Kinect v2 センサーで，表情データを Microsoft Azure Emotion API で取得した．これにより，視線の持続時間や方向，表情の中立性や感情表現の種類を数値化し，行動と評価の関連性を統計的に検証した．面接の結果では，視線の持続時間が採用可能性と正の相関を示し，一方で，表情が中立的であるほど採用可能性が低下する傾向が見られた．ホテル受付の結果では，顧客対応能力が視線の優位性や怒りの表情の頻度と関連しており，これらの要素が顧客との対話における評価基準となり得ることが示された．

Evelyn[19] は，スピーチコンテキストにおける頭部動作の言語学的機能を調査している．自然対話のビデオ記録を用い，頭部動作と音声の同期性，頻度，機能的役割を定量的および質的に分析した．特に，頷き，首振り，横方向の動き，頭部の固定の 4 つの主要な動作タイプに焦点を当て，それぞれが発話行動とどのように連携しているかを分析した．分析の結果，頭部動作はスピーチにおいて多様な言語的機能を果たしていることが明らかになった．具体的には，頭部動作は発話中の特定の単語やフレーズを強調するために使用され，特に頷きが頻出することが確認された．頭部の一定の動きは，発話の区切りやリ

ズムを明示し，構造化に寄与する役割を担っている．首振りや頷きは，相手の発言に対するフィードバックとして機能し，対話の継続や修正を示すシグナルとして重要であることが示された．頭部の大きな動きは，驚きや否定的な感情といった話者の感情状態を補完的に表現する傾向が見られ，非言語的な感情表出の手段としても機能することが明らかになった

Wakabayashi ら [20] は，仮想環境における販売員アバターの頷き模倣行動が，顧客との対話の楽しさや信頼関係に与える影響を検証している．実験では，仮想環境内に再現された家具販売店で販売員アバターが顧客にサービスを提供するシステムが用いられた．販売員アバターは，顧客の顔角度データを基に頷きを自動的に模倣するモデルと，ランダムに頷きを行うモデルの 2 つに設定し，14 名の参加者（20〜24 歳）に顧客役として参加してもらい，各アバターによる接客を 5 分間ずつ体験した．接客終了後，参加者は対話の楽しさ，個人的な繋がり，満足度，忠誠意図，口コミ意図に関するアンケートに回答した．アンケートは 7 段階のリッカート尺度で評価され，信頼性の高い顧客サービスの指標とした．その結果，模倣型アバターは対話の楽しさの項目で有意に高い評価を得た．また，女性参加者は模倣型アバターに対して男性参加者よりも高い満足度を示し，性別間の違いが頷き模倣の効果に影響することが示唆された．

Leone[21] は，教師と生徒の対話における生徒の頷きの多義的な役割について調査を行なっている．実験では，14〜15 歳の 27 人の中学生を対象に，教師と生徒，または生徒同士が 2 人 1 組で参加するシミュレーションゲームを実施した．このゲームでは，倫理的ジレンマに関するカードを読み上げ，意見を交換する場面が設定された．参加者のコミュニケーションはビデオ撮影され，頷き行動を含む非言語的信号が記録された．その後，頷きのタイミングや持続時間，体の向き，視線，表情などをマルチモーダル分析のフレームワークで評価した．その結果，観察された 45 回の対話のうち，頷きが見られたのは 23 回であった．そのうち 15 回は実際に理解を伴っていたが，5 回は完全に理解しておらず，3 回は部分的な理解に留まっていた．さらに，頷きが見られなかった 22 回の対話でも，似た傾向が確認された．特に，頷きが理解を示さない場合，身体の向きのずれや視線が相手から逸れている，表情に不安や困惑が見られることが分かった．

McClave[22] は，自然な会話や特定のシーンにおける頭部の動作を分析を行っている．実験では，参加者の自然な会話をビデオ録画し、その中での参加者の頭部の動作がどのように発話の強調やコミュニケーションの調整、対話の進行に影響を与えるかを測定した．その結果，話し手の頭部の動作は発話の強調や区切りを示す補助的な手段として機能し，聞き手の頭部の動作は相手の発話に対する理解や同意，反応として利用され，対話の円滑な進行に寄与していることを明らかにした．

第3章　研究課題

本章では，本研究における問題の定義と研究課題について述べる．

3.1　問題の定義

2.1 節に示したように，説明タスクにおける被説明者の理解度を推定する取り組みが行われている．これらの取り組みでは，視覚的情報が理解度を推定するのに役立つことが示されている．しかし，これらの取り組みは，レゴモデルの組み立てやボードゲームの説明における理解度に着目しているため，対話における理解度と視覚的情報の関係は明らかになっていない．一方，2.2 節に示したように，個人の意図や心理状態と視覚的情報との関係を分析する取り組みも行われている．これらの取り組みでは，表情，視線，頷きは，困惑，集中，理解などと関係していることが示されている．しかし，表情，視線が聞き手の理解度と関係しているのか，どのような頷きが聞き手の理解度に影響しているのか明らかになっていない．

3.2　研究課題の設定

3.1 節で述べたように，日常生活での対話における聞き手の理解度と対話参加者の視覚的情報の関係や表情明らかになっていないという問題がある．そこで本研究では，**聞き手の理解度と話し手と聞き手双方の視覚的情報の関係を明らかにすること**を研究課題として設定する．課題を達成するために，本研究はまず，2 者対話データを用いて頷きのアノテーションを行う．次に，2 者対話データから視覚的特徴量を抽出し，聞き手の理解度と視覚的情報の関係を分析する．

第4章　対話コーパス

本章では，本研究で用いた対話コーパスについて述べる．本研究は既存の 2 者対話コーパス [23] を利用する．この対話コーパスは，2 者対話データと聞き手の理解度の評価が記録されている．さらに，本研究は上記の 2 者対話コーパスに対して，後述のように聞き手の頷きの評価を追加した．

4.1　2 者対話コーパス

本節では，本研究で用いた対話コーパスに含まれる対面の 2 者対話データについて述べる．本研究で用いた 2 者対話データには，対話参加者の視覚的情報を記録するための映像データが記録されている．2 者対話の参加者は，初対面の 20 代から 50 代の日本人の男女合計 26 名である．対話データの収録には，ストーリテリングと呼ばれる手法が用いられている．具体的には，参加者は対話前にアニメーション「トムとジェリー」を鑑賞し，一方の参加者（話し手）が他方の参加者（聞き手）に鑑賞したアニメーション内容について説明を行っている．発話の単位は Inter-pausal units（IPU）を用いており，沈黙時間が 200ms 未満の連続した音声区間を 1 つとしている．この対話データでは，合計 7,805 件（話し手：4,940 件，聞き手：2,865 件）の IPU が記録されている．

4.2　理解度の評価

本節では，理解度の評価方法について述べる．本研究で用いる 2 者対話コーパスには，第三者のアノテータ 3 名による聞き手の理解度の評価が記録されている．アノテータは，対話全体を通して聞き手が理解できている度合いが変化したと感じたタイミングで次に示す評価を変更する方法で評価している．アノテータは，−2（全く理解していない），−1（あまり理解していない），0（中立の状態），+1（少し理解している），+2（とても理解している）の 5 段階で聞き手が理解できている度合いを評価している．本研究は，話し手の発話終了時点での各アノテータの評価を聞き手の理解度とする．各アノテータが評価した理解度の分布を図 4.1 に示す．話し手の発話終了時におけるアノテータ間の理解度の評価の一致率を評価するために，級内相関係数（ICC）[24] を求めた．話し手の発話終了時における 3 名のアノテータの評価の級内相関係数の平均は ICC(2, k)=0.484 であった．

4.3　頷きの評価

第三者のアノテータ 1 名によって，話し手の発話時における聞き手の頷きの判定を行った．アノテータは，本研究で構築したアノテーションシステムを用いて，聞き手の x 軸周りの回転角度の変化を表した波形と聞き手の映像を参照しながら聞き手が頷いているか判定した．その結果，対話全体で 5,451 件の頷きが確認された．

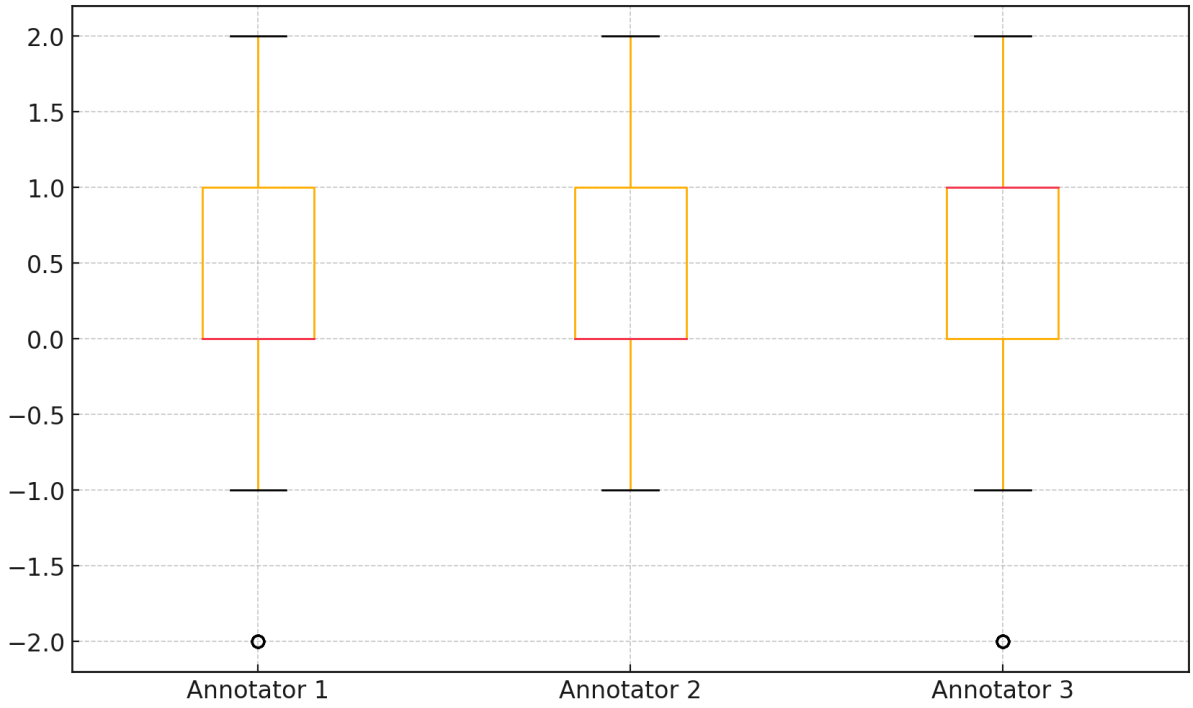

図 4.1: 各アノテータが評価した理解度の分布

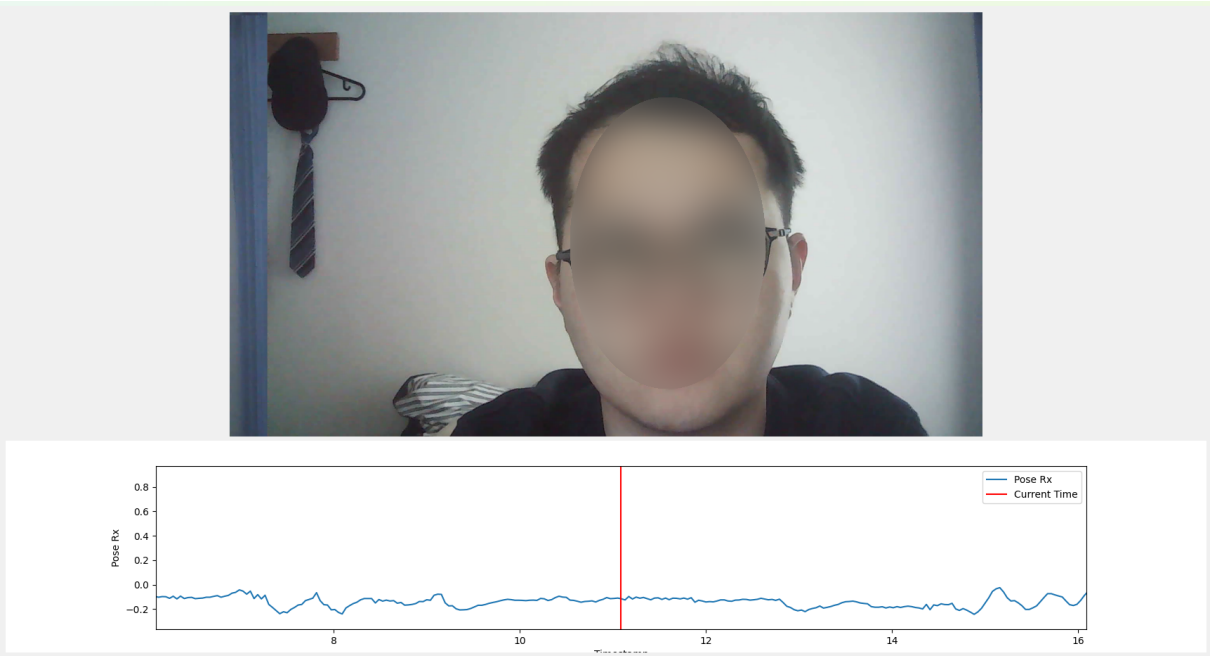

図 4.2: アノテーションシステム

第5章　理解度に応じた視覚的情報の分析

本章では，聞き手の理解度と話し手と聞き手双方の視覚的情報との関係の分析について述べる．

5.1　分析対象とする理解度について

4.2 節で述べたように，理解度の評価は対話全体を通して聞き手が理解できている度合いが変化したタイミングで評価されている．そのため，話し手が発話している最中にアノテータの評価が変更されている発話も存在する．本研究では，先行研究 [25] と同様に発話区間中の行動が発話終了時点での理解度に貢献すると考え，話し手の発話終了時点でのアノテータの理解度の評価を発話時の理解度として用いる．さらに，本研究では，相関分析の精度を高めるために，3 名のアノテータの理解度の評価が一致した発話 1,354 件を分析対象とする．

5.2　分析対象とする頷きについて

対話において，頷きは様々な役割を担っている．発話中の頷きには会話を弾ませる合いの手の役割があり，発話終了後の頷きには発話内容に対する反応を示す役割がある．そのため，発話中の頷きだけでなく発話終了後の頷きにも着目する必要がある．そこで本研究は，話し手の発話中に発生する聞き手の頷きだけでなく，話し手が発話を終了してから 1 秒後までに発生する頷きも当該発話における頷きとして扱う．分析対象となる頷きは，話し手の発話中および発話終了後に発生した 794 件である．頷きの分析を行うにあたり，本研究では当該発話における頷きの角度と回数を計測した．

5.3　視覚的特徴の抽出

聞き手の理解度と対話参加者の視覚的情報との関係を分析するために，4.1 節の対話データから顔画像処理ツールである OpenFace を用いて，話し手と聞き手の視覚的特徴量を抽出した．具体的には，対話参加者の正面に設置したビデオカメラで撮影した映像データから，頭部の向き，視線の角度，Action Units に関する特徴量を抽出した．頭部の向きは，カメラ側から対話参加者の顔を見て，左から右方向を x 軸，下から上方向を y 軸，手前から奥への方向を z 軸とし，話し手の発話時おける頭部の x 軸，y 軸，z 軸周りの回転角度（pose_Rx，pose_Ry，pose_Rz）の分散（_var），中央値（_med），10 パーセンタイル値（_P10），90 パーセンタイル値（_P90）を用いた．視線の角度は，カメラ側から対話参加者の顔を見て，左から右方向を x 軸，下から上方向を y 軸とし，話し手の発話時おける視線の x 軸，y 軸方向の角度（gaze_angle_x，gaze_angle_y）の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．Action Units は，話し手の発話時おける各 Action Units（表 5.1）の強度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

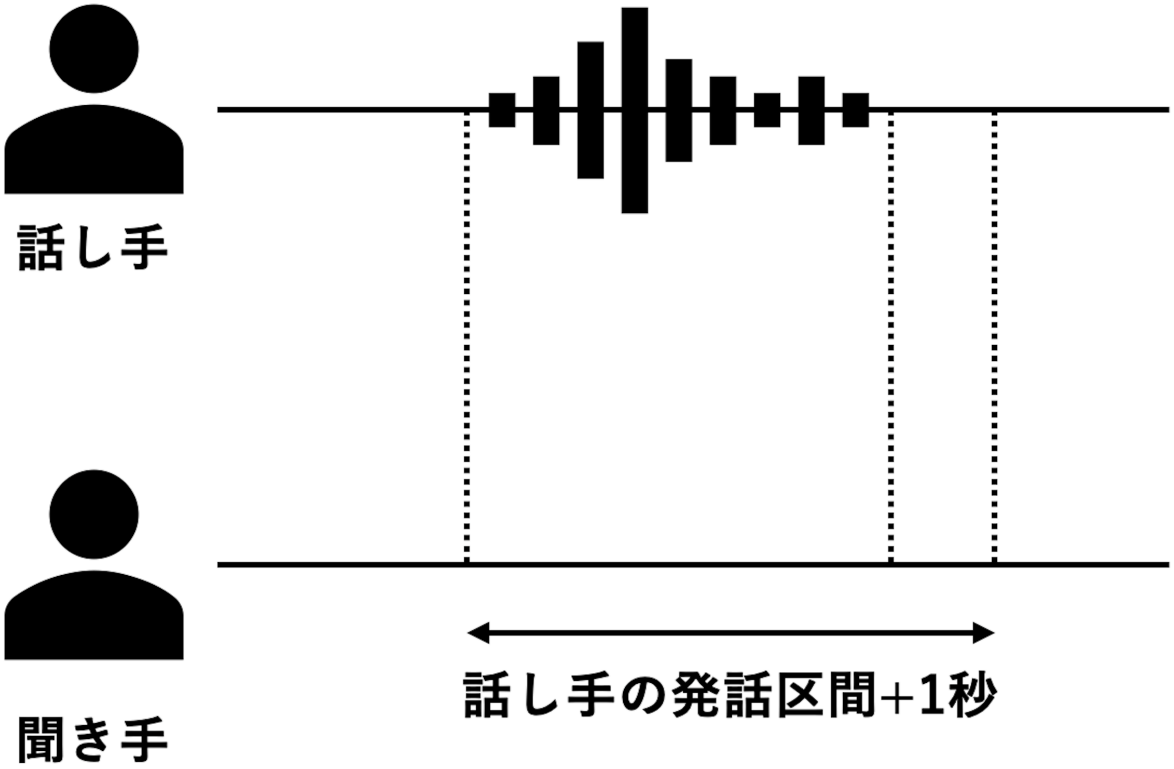

図 5.1: 頷きの判定区間

5.4　分析

本節では，聞き手の理解度と話し手と聞き手双方の視覚的情報との分析について述べる．

5.4.1　理解度と視覚的情報の相関分析

聞き手の理解度と視覚的情報の関係を明らかにするために，相関分析を行った．相関分析を行うにあたり，各特徴量のデータが正規分布に従っているか確認するために有意水準 5 %でシャピロ・ウィルク検定を行った．その結果，全ての特徴量において p 値が 0.05 未満となったため，各特徴量の分布は正規分布に従わないということが確認された．正規性が確認できなかったため，相関係数の算出にはスピアマンの順位相関係数を採用した．相関係数を算出した結果を表 5.2，表 5.3 に示す．

5.4.2　理解度の高群と低群における視覚的情報の比較

理解度の高さによって視覚的情報の傾向に違いがあるのか明らかにするために，発話を 2 群に分けて分析を行った．具体的には，3 名のアノテータの理解度の評価が一致している発話のうち，3 名とも +2（とても理解している）と評価であった発話を理解度高群（計

表 5.1: Action Units の詳細

項目	内容	項目	内容
AU01	眉の内側を上げる	AU14	笑窪を作る
AU02	眉の外側を上げる	AU15	唇の両端を下げる
AU04	眉を下げる	AU17	顎を上げる
AU05	上瞼を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を固く閉じる
AU07	瞳を緊張させる	AU25	顎を下げずに唇を開く
AU09	鼻に皺を寄せる	AU26	顎を下げて唇を開く
AU10	上唇を上げる	AU45	瞬きをする
AU12	唇の両端を引き上げる		

11 シーン）とし，3 名とも −1（あまり理解していない）であった発話を理解度低群*（計 36 シーン）とした．次に，5.3 節で抽出した視覚的特徴量を用いて，理解度高群と低群における各特徴量の値を対応のない t 検定を用いて比較を行った．表 5.2 において，話し手と聞き手それぞれで相関係数の高かった特徴量の値と，t 検定の結果を表 5.4 に示す．

*3 名のアノテータが −2（全く理解していない）と評価した発話が存在しなかったため，本稿では −1（あまり理解していない）であった発話を理解度低群とした．

表 5.2: 聞き手の理解度と視覚的特徴量の相関係数

	話し手				聞き手			
項目	分散	中央値	P10	P90	分散	中央値	P10	P90
AU01	0.0143	-0.1807	**-0.2329**	-0.0858	0.0514	-0.1423	**-0.2193**	-0.0611
AU02	0.0735	-0.1083	-0.1696	0.0195	0.0913	-0.0643	-0.0982	0.0248
AU04	**0.2574**	0.0684	0.0080	0.1200	**0.2539**	0.1252	0.0860	0.1652
AU05	0.1116	-0.0813	-0.1369	0.0622	0.0500	-0.0828	-0.0824	0.0028
AU06	**0.3841**	**0.2680**	0.1563	**0.3319**	**0.3435**	0.1308	0.0550	**0.2064**
AU07	**0.2051**	0.0782	0.0221	0.1433	**0.3131**	**0.2269**	0.1687	**0.2697**
AU09	**0.2904**	0.1120	-0.0554	**0.2450**	**0.3445**	0.1590	0.0537	**0.2952**
AU10	**0.3140**	0.1183	0.0194	**0.2059**	**0.3111**	0.1779	0.0941	**0.2531**
AU12	**0.3594**	**0.2322**	0.1151	**0.2943**	**0.3407**	0.1798	0.1009	**0.2648**
AU14	0.0501	-0.1101	-0.1744	-0.0277	0.1659	0.0534	0.0058	0.1078
AU15	0.1434	0.0025	-0.0886	0.1100	0.0438	-0.0860	-0.1265	-0.0260
AU17	**0.2071**	-0.0863	-0.1832	0.0491	0.1288	-0.1189	-0.1990	-0.0483
AU20	0.1723	-0.0305	-0.0958	0.1063	0.0859	-0.0475	-0.1033	0.0346
AU23	0.1365	-0.1150	-0.1596	0.0682	0.0676	-0.1020	-0.1456	-0.0126
AU25	**0.2915**	**0.2089**	0.0465	**0.2838**	**0.3435**	**0.2523**	0.1462	**0.3308**
AU26	**0.2120**	-0.0116	-0.1625	0.0977	**0.2890**	0.0859	-0.0455	0.1715
AU45	0.0736	-0.1253	-0.1731	0.0092	0.1808	0.0778	0.0074	0.1431
gaze_angle_x	**0.2538**	0.0465	-0.0041	0.1227	**0.2807**	-0.1303	-0.1466	-0.0994
gaze_angle_y	**0.3228**	-0.0610	-0.1754	0.0571	**0.4379**	0.0876	-0.0469	**0.2124**
pose_Rx	**0.3318**	-0.0780	-0.1898	0.0385	**0.4351**	0.0890	-0.0535	**0.2241**
pose_Ry	**0.2626**	-0.0559	-0.1382	-0.0034	**0.2884**	0.1247	0.0930	0.1449
pose_Rz	**0.2437**	0.0378	-0.0387	0.1023	**0.3116**	0.0362	-0.0157	0.0777

表 5.3: 聞き手の理解度と頷きの角度・回数の相関係数

項目	相関係数
頷きの角度	0.0846
頷きの回数	0.5396

表 5.4: 話し手と聞き手それぞれで高い相関を示した特徴量と t 検定の結果

特徴量	高群（N=11）	低群（N=36）	p 値
話し手_AU06_r_var	0.0656	0.0426	**0.0005**
話し手_AU10_r_var	0.1253	0.0552	**6.53E-12***
話し手_AU12_r_var	0.0711	0.0533	**0.016**
話し手_gaze_angle_y_var	0.0031	0.0018	**1.77E-05***
話し手_pose_Rx_var	0.0085	0.0095	0.8003
聞き手_AU06_r_var	0.0587	0.0364	**0.0051**
聞き手_AU09_r_var	0.0216	0.0051	**6.05E-15***
聞き手_AU25_r_var	0.1707	0.0603	**2.12E-17***
聞き手_gaze_angle_y_var	0.0020	0.0009	**5.54E-10***
聞き手_pose_Rx_var	0.0042	0.0016	**3.9E-10***

* $p < 0.05$

第6章　結果・考察

表 5.2，表 5.3，表 5.4 にあるように，話し手については，AU06（頬を持ち上げる），AU10（上唇を上げる），AU12（唇の両端を引き上げる）や gaze_angle_y（視線の上下角度），pose_Rx（頭部の x 軸周りの回転角度）などが聞き手の理解度と高い相関を示した．理解度高群と低群の比較の結果，AU06，AU10，AU12 や gaze_angle_y で統計的に差があることが確認された．一方で，聞き手については，AU06（頬を持ち上げる），AU09（鼻に皺を寄せる），AU25（顎を下げずに唇を開く）や gaze_angle_y（視線の上下角度），pose_Rx（頭部の x 軸周りの回転角度）などが聞き手の理解度と高い相関を示した．理解度高群と低群の比較の結果，AU06，AU09，AU25 や gaze_angle_y，pose_Rx で統計的に差があることが確認された．本章では，話し手と聞き手それぞれで高い相関を示した特徴量が聞き手の理解度とどのような関係があるのかについて考察を行う．

6.1　話し手の視覚的特徴量について

本節では，話し手に関する特徴量のうち，聞き手の理解度と高い相関を示したものについて議論する．AU06 は頬を上げる動作，AU12 は口角を上げる動作を表しており，どちらも笑顔に関連する Action Units である．笑顔はポジティブな感情の伝達だけでなく，相手に親しみやすさを感じさせる効果があるとされている [26]．そのため，話し手が笑顔を交えつつ発話することで，話し手の親しみやすさが聞き手に伝達され，結果的に理解度の向上に繋がったのではないかと考えられる．

AU10 は上唇を引き上げる動作を表しており，嫌悪や驚き，集中，強調の表現に関連する Action Units である．話し手が発話中にこの動作が生起するということは，強調表現を用いて聞き手の注意を引き付けていると考えられる．また，一般的に AU10 はネガティブな印象を与える動作だが，他の Action Unit との組み合わせで幅広い印象を与えるとされている [1]．AU10 と他の Action Unit との関係を分析することで，より聞き手の理解度との関係を明らかにすることができると考えられる．

gaze_angle_y は視線の上下角度を示している．話し手は発話中に発話内容の計画や情報の整理を行う際に，対話相手から視線を逸らすことがあるとされている．一方で，話し手が対話相手に視線を向けると理解の確認や相手の反応を引き出すための手段となる [27]．そのため，話し手は聞き手が理解しやすいような発話内容にする過程で視線を逸らし，聞き手の理解度を測るために視線を向けており，結果的に聞き手の理解度が高まったと考えられる．

pose_Rx は頭部の x 軸周りの回転角度であり，頷きなどの頭部を上下に動かす動作に関連している．話し手においてこの動作は発話の強調や区切りを示すとされている [22]．AU10 のみならず，pose_Rx からも発話中の強調表現が聞き手の理解度に影響することが明らかになった．

6.2 聞き手の視覚的特徴量について

聞き手に関する特徴量のうち，聞き手の理解度と高い相関を示したものについて議論する．AU06，gaze_angle_y，pose_Rx は話し手だけでなく，聞き手でも理解度との高い相関を示した．AU06 は，聞き手も話し手と同様に笑顔の相手に親しみやすさを伝達する効果が働いているものと考えられる．gaze_angle_y は，話し手では発話内容の計画や情報の整理を行う際に視線が変化したが，聞き手では頷きと連動して視線が変化したと考えられる．そして，pose_Rx も同様に頷きが影響したと考えられる．頷いているシーンに限定すると，理解度は，頷きの角度と相関が見られなかったものの，頷きの回数と高い相関が見られた．このことから，聞き手の頷きは理解度と深く関係しており，頷きの角度よりも頷いてる回数が聞き手の理解度を把握するためには重要であることが示された．

AU09 は鼻にしわを寄せる動作を表しており，嫌悪や困惑，考え込む過程で表れやすい Action Units である．話し手の発話に対し，内容を理解できていない状態から理解する状態までの思考過程で表出したと考えられる．

AU25 は唇を開く動作を表しており，相槌や驚き，納得などの反応として表れることが多い．聞き手は，話し手の発話中に相槌などの発話に対する反応を示すことが，自身の理解度を示すのに重要であるということが明らかになった．

第7章　結論

本研究では，日常生活での対話における聞き手の理解度と対話参加者の視覚的情報の関係を分析する取り組みを行った．具体的には，対話コーパスから対話参加者の視覚的特徴を抽出し，聞き手の理解度との相関分析を行った．その結果，AU06（頬の上昇），AU07（瞳の緊張），AU09（鼻に皺を寄せる），AU10（上唇を上げる），AU12（唇の両端を引き上げる），AU25（顎を下げずに唇を開く）などが理解度と正に相関することが明らかになった．これらは笑顔や肯定的表情，あるいは思案や集中を示す動作に関与するとされており，聞き手が深く理解・共感している際に自然に表出されやすい可能性がある．一方で，AU01（眉の内側を上げる）や AU02（眉の外側を上げる）などは驚きや疑問を表す動作として負の相関が見られ，AU15（唇の両端を下げる）や AU17（顎を上げる）なども混乱や不満との関連が指摘され，理解度とは負の相関を示した．これらは理解が進むにつれて減少する傾向が示唆される．頷きと理解度では，頷きの角度はほとんど相関が見られなかったが，頷きの回数は中程度以上の正の相関が得られた．これは，大きく頷くかどうかよりも，短い頷きを複数回行うほうが聞き手の積極的な関与や理解を示す可能性があることを示した．

今後の展望として，本研究で視線情報に着目していないため，視線移動の方向や注視などと聞き手の理解度の相関分析を行いたいと考えている．そしてこれらの知見を活かし，対話における聞き手の理解状況を把握するシステムを実装したいと考えている．

謝辞

本研究は，日本電信電話株式会社 NTT 人間情報研究所との共同研究の成果である．

参考文献

[1] Paul Ekman and Wallace V. Friesen. Facial action coding system: A technique for the measurement of facial movement. Technical report, Consulting Psychologists Press, Palo Alto, CA, 1978.

[2] P. Ekman. Facial expression and emotion. *American Psychologist*, Vol. 48, No. 4, pp. 384–392, 1993.

[3] C. Goodwin. Conversational organization: Interaction between speakers and hearers. In *Talk at Work*, pp. 1–30. Ablex, Norwood, NJ, 1981.

[4] C. L. Kleinke. Gaze and eye contact: A research review. *Psychological Bulletin*, Vol. 100, No. 1, pp. 78–100, 1986.

[5] A. Kendon. Some relationships between body motion and speech: An analysis of an example. In *Studies in Dyadic Communication*, pp. 177–210. Pergamon, Oxford, 1972.

[6] A. Mimura and M. Hagiwata. Understanding presumption system from facial images. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, Vol. 5, pp. 3270–3275 vol.5, 1999.

[7] Masaki Omata, Masaya Iuchi, and Megumi Sakiyama. Comparison of eye-tracking data with physiological signals for estimating level of understanding. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI '18, p. 563–567, New York, NY, USA, 2018. Association for Computing Machinery.

[8] Herbert H. Clark and Meredyth A. Krych. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, Vol. 50, No. 1, pp. 62–81, 2004.

[9] Olcay Türk, Stefan Lazarov, Yu Wang, Hendrik Buschmeier, Angela Grimminger, and Petra Wagner. Predictability of understanding in explanatory interactions based on multimodal cues. In *Proceedings of the 26th International Conference on Multimodal Interaction*, ICMI '24, p. 449–458, New York, NY, USA, 2024. Association for Computing Machinery.

[10] Olcay Türk, Petra Wagner, Hendrik Buschmeier, Angela Grimminger, Yu Wang, and Stefan Lazarov. Mundex: A multimodal corpus for the study of the understanding of explanations. In *Book of Abstracts of the 1st International Multimodal Communication Symposium*, 2023.

[11] Nashwa Ismail. Analysing qualitative data using facial expressions in an educational scenario. *International Journal of Quantitative and Qualitative Research Methods*, Vol. 5, pp. 37–50, 07 2017.

[12] Sightcorp. Sightcorp. (2017) [Online] Available at: www.Sightcorp.com. Online, 2017. Accessed: 05/04/2017.

[13] Jiabei He, Xiaoyu Wen, and Juxiang Zhou. Advances and application of facial expression and learning emotion recognition in classroom. In *Proceedings of the 2023 6th International Conference on Image and Graphics Processing*, ICIGP '23, p. 23–30, New York, NY, USA, 2023. Association for Computing Machinery.

[14] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing*, Vol. 5, No. 1, pp. 86–98, 2014.

[15] Guan-Yun Wang, Hikaru Nagata, Yasuhiro Hatori, Yoshiyuki Sato, Chia-Huei Tseng, and Satoshi Shioiri. Detecting learners' hint inquiry behaviors in e-learning environment by using facial expressions. In *Proceedings of the Tenth ACM Conference on Learning @ Scale*, L@S '23, p. 331–335, New York, NY, USA, 2023. Association for Computing Machinery.

[16] Renjun Miao, Haruka Kato, Yasuhiro Hatori, Yoshiyuki Sato, and Satoshi Shioiri. Analysis of facial expressions to estimate the level of engagement in online lectures. *IEEE Access*, Vol. 11, pp. 76551–76562, 2023.

[17] S. Shujan, A.T. Rupasinghe, N.L. Gunawardena, and D. Atukorale. Employee recruitment: Question formation for employment interviews based on facial cue analytics. *DecisionSciRN: Decision-Making & Performance Management (Topic)*, 2016.

[18] Skanda Muralidhar, Rémy Siegfried, Jean-Marc Odobez, and Daniel Gatica-Perez. Facing employers and customers: What do gaze and expressions tell about soft skills? In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM '18, p. 121–126, New York, NY, USA, 2018. Association for Computing Machinery.

[19] Evelyn Z. McClave. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, Vol. 32, No. 7, pp. 855–878, 2000.

[20] Takumi Wakabayashi, Yukihiko Okada, and Keiichi Zempo. Effect of salesperson avatar automatically mimicking customer's nodding on the enjoyment of conversation in virtual environments. In *Proceedings of the Augmented Humans International Conference 2023*, AHs '23, p. 334–337, New York, NY, USA, 2023. Association for Computing Machinery.

[21] Giovanna Leone. Nodding without understanding: An explorative study of how adolescents listen to their teachers. In *2012 International Conference on Social Informatics*, pp. 137–144, 2012.

[22] Evelyn Z. McClave. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, Vol. 32, No. 7, pp. 855–878, 2000.

[23] Ryo Ishii, Taichi Katayama, Ryuichiro Higashinaka, and Junji Tomita. Automatic generation of head nods using utterance texts. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 1143–1149, 2018.

[24] Patrick E. Shrout and Joseph L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, Vol. 86, No. 2, pp. 420–428, 1979.

[25] S. Kinoshita, T. Onishi, N. Azuma, R. Ishii, A. Fukayama, T. Nakamura, and A. Miyata. A study of prediction of listener's comprehension based on multimodal information. In *Proceedings of the the 23rd ACM International Conference on Intelligent Virtual Agents (IVA '23)*, No. 30, pp. 1–4, 2023.

[26] Robert E. Kraut and Robert E. Johnston. Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, Vol. 37, No. 9, pp. 1539–1553, 1979.

[27] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, Vol. 7. CUP Archive, 1990.

研究業績

研究会・シンポジウム

(1) 鹿摩大智, 岡哲平, 大西俊輝, 東直輝, 石井亮, 深山篤, 宮田章裕: 聞き手の相槌種類に応じた表情生成システムの基礎検討. 情報処理学会インタラクション 2024 論文集, pp.658–661 (2024).

(2) 鹿摩大智, 岡哲平, 大西俊輝, 東直輝, 石井亮, 宮田章裕: マルチモーダル情報に基づいて適切な相槌を提示するシステムの検討. 情報処理学会シンポジウム論文集, マルチメディア, 分散, 協調とモバイル (DICOMO ’24), pp.468–474 (2024).

(3) 岡哲平, 鹿摩大智, 大西俊輝, 東直輝, 石井亮, 宮田章裕: 対話における理解度と視覚的情報の関係の予備的分析. 情報処理学会研究報告コラボレーションとネットワークサービス（CN）, Vol.2025-CN-125, No.6, pp.1-6 (2025)