

褒め方の上手さの推定における頭部・顔部の効果

大西 俊輝¹ 柴田 万里那¹ 山内 愛里沙¹ 呉 健朗² 石井 亮³ 富田 準二³ 宮田 章裕^{1,a)}

概要: 対話において褒めることは重要である。例えば、初対面の相手と話す場合、相手の見た目やエピソードを積極的に褒めることで相手が心を開き、話が弾み、その後の人間関係が円滑になることが期待される。しかし、対話において相手を上手く褒めるために、発話内容、声色、表情、姿勢、ジェスチャといった多様なモダリティをどのように用いれば良いか明らかにはされていない。本研究では、これらの多様なモダリティを用いた上手い褒め方をモデリングする手法の確立を目指す。本稿では、対話における褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにする取り組みを行った。具体的には、2者対話コーパスの作成、特徴量抽出、褒め方の上手さに寄与する特徴量の分析を行った。2者対話コーパスの作成では、褒め方の上手さの評価と各モダリティの記録を含む2者対話を、10組分、計100分間記録した。また、記録したデータに対してアノテーションを付与した。特徴量抽出では、頭部、顔部、発話に関する特徴量の抽出を行った。褒め方の上手さに寄与する特徴量の分析では、頭部・顔部のどこが関係しているのかを明らかにするための分析を行った。その結果、頭部の縦軸回りの回転角度、視線の横方向の角度、顎をあげる動き、目を見開く（臉に力を入れる）動きが、褒め方の上手さに関係していることが明らかになった。

The Effect of Head and Face on Estimation of the Praise Ability

TOSHIKI ONISHI¹ MARINA SHIBATA¹ ARISA YAMAUCHI¹ KENRO GO² RYO ISHII³ JUNJI TOMITA³
AKIHIRO MIYATA^{1,a)}

1. はじめに

対話において褒めることは重要である。例えば、初対面の相手と話す場合、相手の見た目やエピソードを積極的に褒めることで相手が心を開き、話が弾み、その後の人間関係が円滑になることが期待される。あるいは、同僚や部下の行為を適切に褒めることで彼らのパフォーマンスを最大化することが、組織マネジメントの観点からも重要であるとの認識が広がっている。本研究で対象となる褒める行為は、海外において、承認・称賛を表現すること、対象の行動の良い側面を示唆する言語的または身体的な行動、対象の行動や性格に向けられた称賛を示す言語的・非言語的な

行為と定義されている [2][3][4]。また、国内においても、相手の良さ、優れている点、長所に気づき、価値を正しく認めて伝える行為と定義されている [5][6]。しかし、対話において相手を上手く褒めるために、発話内容、声色、表情、姿勢、ジェスチャといった多様なモダリティをどのように用いれば良いか明らかにはされていない。このため、日常生活の観点においては、相手を上手く褒めることが苦手な人は何を意識すれば褒め方が上達するのかわからないという問題がある。工学的な観点においては、人の褒め方の上手さを判定するシステムを構築することや、上手い褒め方を行う対話型エージェントを実現することが困難であるという問題がある。このような背景を踏まえ、本研究では、発話内容、声色、表情、姿勢、ジェスチャといった多様なモダリティを用いた上手い褒め方をモデリングする手法の確立を目指す。先行研究において、対話において相手を上手く褒めるためには、表情が欠かせない要素の一つであることを明らかにした [1]。しかし、対話において相手を上手く褒めるためには、どのような表情で褒めると良いか明

¹ 日本大学文理学部
College of Humanities and Sciences, Nihon University

² 日本大学大学院総合基礎科学研究科
Graduate School of Integrated Basic Sciences, Nihon University

³ 日本電信電話株式会社 NTT メディアインテリジェンス研究所
NTT Media Intelligence Laboratories, NTT Corporation

a) miyata.akihiro@acm.org

らかにはされていない。また、褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにはされていない。よって、本稿では、対話において相手を上手く褒めるためには、どのような表情で褒めると良いか、そして、褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにする取り組みを行う。本稿の貢献は、対話における褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにしたことである。

2. 関連研究

本研究は、対話において相手を上手く褒めるために、発話内容、声色、表情、姿勢、ジェスチャといった多様なモダリティを用いた上手い褒め方をモデリングする手法の確立を目指すものであり、対話中の行動・能力を分析する研究に関連している。対話中の行動・能力を分析する研究は数多く行われている。[7]は、現在の発話者の発話の終了と次の発話者の発話の開始との間の間隔を予測するために非言語的行動として参加者の頭部の動きを用いている。参加者の頭の動きは、多人数での会議の順番を変える際の次の発話のタイミングを予測することに有用であることが明らかになった。[8]は、複数人対話における共感スキルレベルを推定するために話者交替/継続の発話および対話行為の終わりに向けた注視行動、発話の意図を示す言語行動情報を調査している。[9]は、印象の性格特性、創発的リーダーシップスキル、コミュニケーション能力、そして採用決定を推論するための新しい特徴量抽出フレームワークを提案している。提案されたフレームワークは、各参加者の非言語的な活動を記述したマルチモーダル特徴量を抽出している。[10]は、会話参加者の表出するマルチモーダル情報からコミュニケーション能力を推定するモデルの構築・評価を行っている。[11]は、グループ会話における全メンバーの頭部方向、頭部動作、発話状態から、視線方向、視線状態とインタラクションの構造を推定する枠組みを提案している。[12]は、グループ会話から派生したマルチモーダル機能を用いて、タスクに対するグループパフォーマンスを自動的に予測している。[13]は、発話言語から得られる様々な種類の自然言語解析情報を用いて、人間と同様に適切なタイミングで発話を伴う全身運動を自動生成するモデルを提案している。頷き、頭部姿勢、表情、ハンドジェスチャ、上半身の姿勢を自動的に生成している。動きの自然さ、発話と動きの一致度、エージェントの人間性、そしてエージェントの好感度に関するエージェントの身体の動きや印象は、改善された。

3. 研究課題

対話の参加者の行動・能力を分析する研究は数多く行われているが、褒め方の上手さを推定することには向いていない[7][8][9][10][11][12][13]。例えば、[10]は、会話参加者

のマルチモーダル情報からコミュニケーション能力を推定しているが、褒め方の上手さを推定を行うことができない。そこで、本研究では、発話内容、声色、表情、姿勢、ジェスチャといった多様なモダリティを用いた上手い褒め方をモデリングする手法の確立を目指す。先行研究において、対話において相手を上手く褒めるためには、表情が欠かせない要素の一つであることを明らかにした[1]。しかし、対話において相手を上手く褒めるためには、どのような表情で褒めると良いか明らかにはされていない。また、褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにはされていない。よって本稿では、対話における褒め方の上手さは、頭部・顔部のどこが関係しているのか明らかにすることを研究課題とする。上記の研究課題に基づき、本研究を下記に示すステップに分けて推進する。

- Step 1: 2者対話の収録
2名1組の2者対話を10組分、計100分間のデータをビデオカメラ、マイク、深度センサを用いて収録する。
- Step 2: アノテーション
音声データと映像データに対してアノテーションを行う。
- Step 3: 褒め方の上手さの評価
話し手であった参加者による褒め方の上手さの評価を行う。
- Step 4: 特徴量抽出
頭部・顔部・発話に関する特徴量を抽出する。
- Step 5: 褒め方の上手さに寄与する特徴量の分析
褒め方の上手さには、頭部・顔部のどこが関係しているのか分析を行う。

4. 対話コーパス

4.1 2者対話の収録

本研究では、国内外の研究事例・書籍に基づいて褒める行為を下記のように定義する[2][3][4][5][6]。

褒める：言語的または非言語的表現を用いて相手の良い側面を称賛すること。

上記の定義に基づき、2者対話を対象とした対話における褒め方の上手さの評価と各モダリティの記録を含む対話コーパスの作成を行う。2者対話の参加者は、20代の大学生20名（男性19名、女性1名）であり、2名1組の組み合わせを10組構成する。10組のうち、初対面が8組、顔見知り1組、友人同士が1組であった。対話の収録を始めるにあたり、参加者に対話材料を準備させることを意図して、アンケート1（表1）に回答してもらう。参加者は、図1のように互いに向き合って着座する。対話の収録は、各参加者の様子と2者対話全体の様子を撮影するためのビデオカメラ、各参加者の声を録音するためのマイク、各参加者のハンドジェスチャなどの動きを計測するための深度センサを用いて行う。各組の参加者達（参加者A、参加者

表1 アンケート 1
Table1: Question 1

項目	質問事項	回答方法
Q1	年齢	単一選択
Q2	性別	単一選択
Q3	相手との関係	単一選択
Q4	今まで頑張ってきたこと (2 つ以上)	自由記述



図1 2者対話の様子

Fig.1: Condition of The Dialogue

Bとする)は、実験者の合図に従い、以下の(1)～(3)を行う。

- (1) 自己紹介 (5分間)
 - (2) 参加者 A が主に話し手となり今まで頑張ってきたことを話し、参加者 B が主に聞き手となり随時意見を述べる対話 (5分間)
 - (3) 参加者 B が主に話し手となり今まで頑張ってきたことを話し、参加者 A が主に聞き手となり随時意見を述べる対話 (5分間)
- (2), (3) を 10 組分、計 100 分間の 2 者対話を収録する。参加者には、それぞれの対話において、自由に質問をすることや、アンケート 1 で回答しなかった内容について発言をすることを許可した。なお、(1) は各組の多くが初対面同士であり、参加者の緊張をほぐす目的で行っているため、今回の分析では使用しない。

4.2 アノテーション

音声や映像データに対して注釈付けを行うツールである ELAN[14] を用いて、収録した音声データと映像データに対して表 2 に示すアノテーションの項目を手手で付与する。発話区間は、沈黙時間が 200 ミリ秒未満の連続した音声区間であり、さらに、発話が 1 つの文として成り立つ区間である。判断基準が揺るがないよう、1 名の実験者が全ての発話区間の付与を行う。発話内容は、参加者の発話内容を書き起こしたものである。4 名のアノテータが分担して発話内容の書き起こしを行う。Praise シーンは、聞き

表2 アノテーションの項目
Table2: Annotation List

項目	説明
発話区間	沈黙時間が 200 ミリ秒未満の連続した音声区間 さらに発話が 1 つの文として成り立つ区間
発話内容 Praise シーン	参加者の発話内容をテキストに書き起こしたもの 聞き手の参加者が相手の発話に対して褒める発話を行っている判断したシーン

手の参加者が相手の発話に対して褒める発話を行っていると判断したシーンである。判断基準が揺るがないよう、1 名の実験者が全ての Praise シーンの付与を行う。

4.3 褒め方の上手さの評価

Praise シーンで褒めていると判断した発話について、当該シーンで話し手であった参加者による褒め方の上手さの評価を行う。2 者対話の収録を行ってから 1 週間以上経過した後、当該シーンの映像データを参照した上で、参加者にアンケート 4 に回答してもらう。参加者にアンケート 4 として、このシーンで相手に褒められて気分が良くなったかについて 7 件法で回答してもらう。

5. 特徴量抽出

顔画像処理ツールである OpenFace[15] を用いて、参加者の正面に設置されたビデオカメラで撮影された映像データから特徴量を抽出を行った。抽出可能な頭部、視線、Action Units[16]、発話に関する特徴量を利用した。頭部に関する特徴量としては、カメラ側から顔を見て横方向を x 軸、縦方向を y 軸、奥行き方向を z 軸とした場合、頭部の x 軸、y 軸、z 軸方向の回転角度 (pose_Rx, pose_Ry, pose_Rz) の当該シーンの前後 1 秒ずつを含む区間における中央値 (_med), 分散 (_var), 10 パーセンタイル値 (_p10), 90 パーセンタイル値 (_p90) を用いた。視線に関する特徴量としては、視線の x 軸、y 軸方向の角度 (gaze_AX, gaze_Ay) の当該シーンの前後 1 秒ずつを含む区間における中央値、分散、10 パーセンタイル値、90 パーセンタイル値を用いた。Action Units に関する特徴量としては、各 Action Units の強度 (AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45) の当該シーンの前後 1 秒ずつを含む区間における中央値、分散、10 パーセンタイル値、90 パーセンタイル値を用いた。発話に関する特徴量としては、当該シーンにおける発話の長さ (秒) を用いた。全 Praise シーンに対して上記の特徴量を算出し、各特徴量を平均値 0、分散 1 になるよう正規化を行った。

6. 褒め方の上手さに寄与する特徴量の分析

褒め方の上手さには、頭部・顔部のどこが関係している

のかを明らかにするための分析を行う。分析方法として、Praise スコアを目的変数、前述の特徴量を説明変数とする回帰モデルを構築した上で、モデル中で重要度が高い特徴量を発見するアプローチをとる。Praise スコアは、各 Praise シーンにおいて、話し手である参加者が聞き手である参加者の褒め方の上手さについて 7 段階で評価をしたものである。モデル構築には Gradient boosting を用いる。学習率などのハイパーパラメータは Hyperopt でチューニングする。データセットを 80 % の訓練データ、20 % の検証データに無作為に分け、訓練データで構築したモデルを用いて、検証データの Praise スコアを予測するタスクを 100 回行ったところ、全タスクの各指標の平均値は表 3 のようになった。訓練データに対する決定係数が 0.881 の強

表 3 各指標の平均値 (N=100)

Table3: Average of the Index

指標	平均値
MSE (train)	0.214
MSE (test)	1.346
R^2 (train)	0.881
R^2 (test)	0.245

い相関、検証データに対する決定係数が 0.245 の弱い相関であることから、ある程度妥当に Praise スコアと特徴量の関係をモデル化できていると言える。このモデルにおける各特徴量の重要度 (gain) を図 2 に示す。この図を見ると、

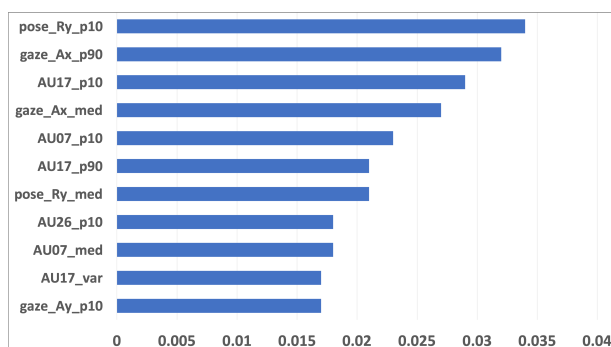


図 2 各特徴量における重要度の上位 10 個 (同率を含む)

Fig.2: Higher Rank Feature Importance

重要度が高い特徴量は、pose_Ry, gaze_Ax, AU17, AU07 に関するものが多いことが分かる。これらはそれぞれ、頭部の縦軸回りの回転角度、視線の横方向の角度、顎をあげる動き、目を見開く (瞼に力を入れる) 動きに該当する。

7. おわりに

2 者対話コーパスの作成、特徴量抽出、褒め方の上手さに寄与する特徴量の分析を行った。2 者対話コーパスの作成では、褒め方の上手さの評価と各モダリティの記録を含む 2 者対話を、10 組分、計 100 分間記録したまた、記録し

たデータに対してアノテーションを付与した。特徴量抽出では、頭部、顔部、発話に関する特徴量の抽出を行った。Praise スコアを目的変数、頭部、顔部、発話に関する特徴量を説明変数とする回帰モデルを構築した。データセットを 80 % の訓練データ、20 % の検証データに無作為に分け、訓練データで構築したモデルを用いて、検証データの Praise スコアを予測するタスクを 100 回行ったところ、全タスクの訓練データに対する決定係数が 0.881、検証データに対する決定係数が 0.245 となり、ある程度妥当に Praise スコアと特徴量の関係をモデル化できたと言える。褒め方の上手さに寄与する特徴量の分析では、頭部・顔部のどこが関係しているのかを明らかにするための分析を行った。その結果、頭部の縦軸回りの回転角度、視線の横方向の角度、顎をあげる動き、目を見開く (瞼に力を入れる) 動きが、褒め方の上手さに関係していることが明らかになった。今後は、頭部、視線、Action Units、発話に関する特徴量を増やしてモデル構築や、Praise スコアが高いシーンの分析などに取り組む予定である。

参考文献

- [1] 大西俊輝, 柴田万里那, 呉健朗, 石井亮, 富田準二, 宮田章裕: 対話における上手い褒め方のモデリングの基礎検討. 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 (DICOMO2019), Vol.2019, pp.656–662 (2019).
- [2] Brophy, J.: Teacher praise: A fullctional analysis: Review of Educational Research, Vol.51, No.1, pp.5–32 (1981).
- [3] Kalis, T. M., Vannest, K. J. and Parker, R.: Praise Counts: Using Self-Monitoring to Increase Effective Teaching Practices. Preventing School Failure: Alternative Education for Children and Youth, Vol.51, No.3, pp.20–27 (2007).
- [4] Jenkins, L. N., Floress, M. T. and Reinke, W.: Rates and Types of Teacher Praise: A Review and Future Directions: Psychology in the Schools, Vol.52, No.5, pp.463–476 (2015).
- [5] 永崎一則: 元気と幸せを生み出す ほめかたの研究. 早稲田教育出版 (2015).
- [6] 西村貴好: 結果を引き出す 大人のほめ言葉. 同文館出版株式会社 (2017).
- [7] Ishii, R., Kumano, S. and Otsuka, K.: Prediction of Next-Utterance Timing Using Head Movement in MultiParty Meetings. Proc. 5th International Conference on Human Agent Interaction (HAI'17), pp.181–187 (2017).
- [8] Ishii, R., Otsuka, K., Kumano, S., Higashinaka, R. and Tomita, J.: Analyzing Gaze Behavior and Dialogue Act during Turn-taking for Estimating Empathy Skill Level. Proc. 20th ACM International Conference on Multimodal Interaction (ICMI'18), pp.31–39 (2018).
- [9] Okada, S., Nguyen, L. S., Aran, O. and Gatica-Perez, D.: Modeling Dyadic and Group Impressions with Intermodal and Interperson Features. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol.15, No.1s, pp.13:1–13:30 (2019).
- [10] Okada, S., Ohtake, Y., Nakano, I. Y., Hayashi, Y., Huang, H., Takase, Y. and Nitta, K.: Estimating communication skills using dialogue acts and nonver-

- bal features in multiple discussion datasets. Proc. 18th ACM International Conference on Multimodal Interaction (ICMI'16), pp.169-176 (2016).
- [11] Otsuka, K., Sawada, H. and Yamato, J.: Automatic inference of cross-modal nonverbal interactions in multi-party conversations: “Who Responds to Whom, When, and How?” from gaze, head gestures, and utterances. Proc. 9th ACM International Conference on Multimodal Interaction (ICMI'07), pp. 255–262, (2007).
- [12] Murray, G. and Oertel, C.: Predicting Group Performance in Task-Based Interaction: Proc. 20th ACM International Conference on Multimodal Interaction (ICMI'18), pp.14–20 (2018).
- [13] Ishii, R., Katayama, T., Higashinaka, R. and Tomita, J.: Generating Body Motions using Spoken Language in Dialogue. Proc. 18th ACM International Conference on Intelligent Virtual Agents (IVA'18), pp.87–92 (2018).
- [14] Brugman, H. and Russel, A.: Annotating Multimedia / Multi-modal resources with ELAN. Proc. 4th International Conference on Language Resources and Language Evaluation (LREC'04), pp.2065–2068 (2004).
- [15] Baltrušaitis, T., Zadeh, A., Lim, Y. C. and Morency, L. P.: OpenFace 2.0: Facial Behavior Analysis Toolkit. 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG'18), pp.59–66 (2018).
- [16] Ekman, P. and Friesen, W. V.: Manual for the Facial Action Coding System. Palo Alto: Consulting Psychologists Press (1977).

付 録

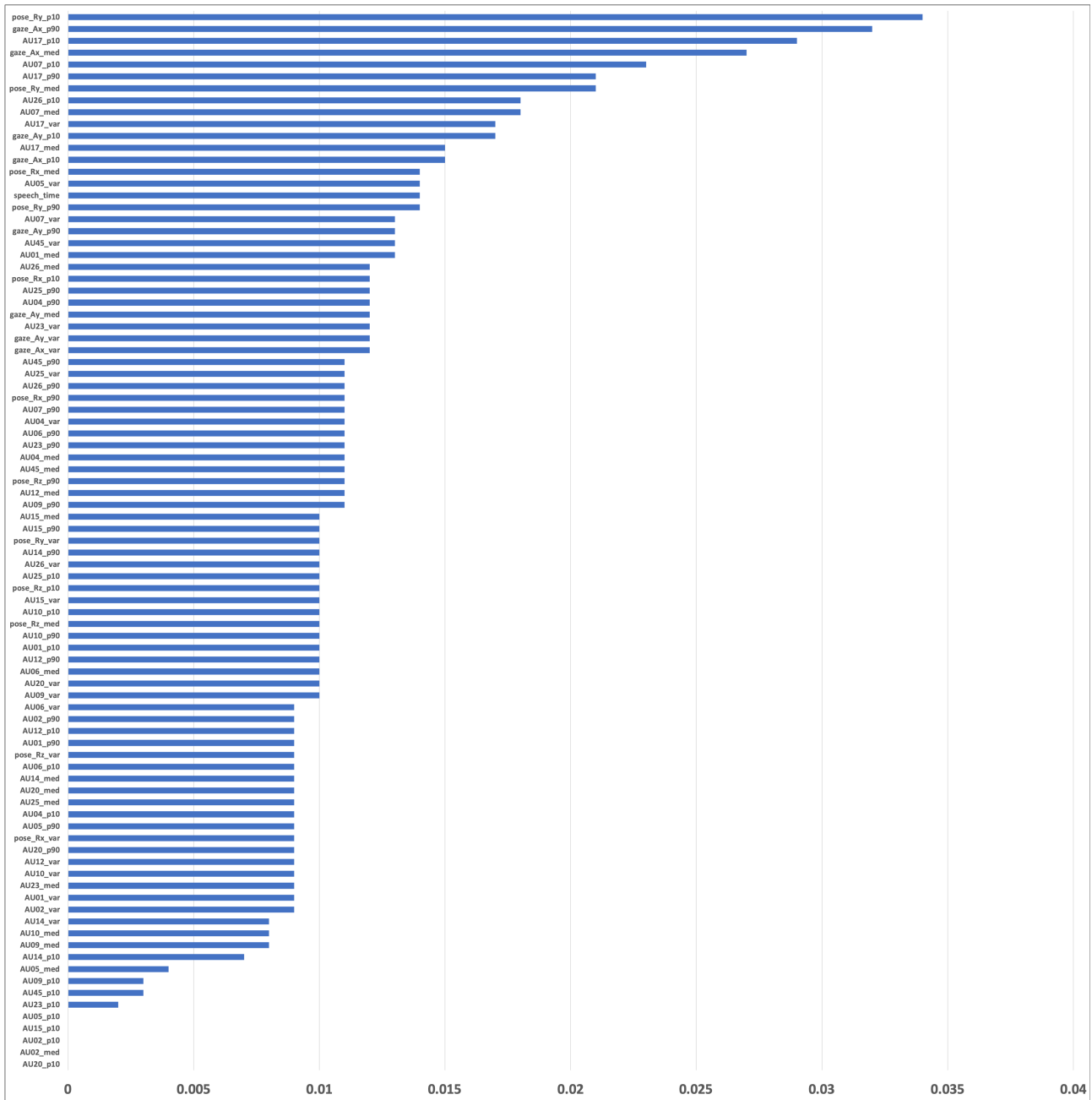


図 A.1 各特徴量における重要度
Fig.A.1: All Feature importance