表情特徴量に着目した 褒め方の上手さフィードバックシステムの基礎検討

大串 旭[1]　大西 俊輝[1]　石井 亮[3]　深山 篤[3]　宮田 章裕[2,a)]

概要：褒める行為はコミュニケーション中で多く行われている．しかし，自身の褒め方がどの程度上手いのかを把握できている人は少ない．その理由として，自身で自分の褒め方を把握することや，褒め方の上手さを向上するための改善方法を適切に把握することが困難であることが挙げられる．これらを解決するために，本稿では話者の褒め方の上手さの評価を提示しながら，褒め方の上手さの改善案を提示するフィードバックシステムの実装を行った．褒め方の上手さの改善案を作成する手法として表情に関する特徴量である Action Units を使用した．上手く褒めるために重要である Action Units をもとに対話中の話者の Action Units を変化させた顔画像を生成することで表情の改善案を提示した．

1. はじめに

コミュニケーションを行う際に褒める行為は多く行われている．褒める行為とは，対象の行動や性格に向けられた称賛を表現する言語・非言語行動と考えられている [1][2]．加えて，褒める行為は，褒める人から褒められる人への一方的な意思の伝達ではなく複雑な社会的コミュニケーションであることが考えられている [3]．一方で，褒める行為が苦手な人や褒める行為がより上手くなりたいと考えている人が存在する．しかし，自分で自身の褒め方を認識し，褒め方の上手さを向上することは困難である．この問題を解決するために，我々は話者の褒め方の上手さの評価を提示し，褒め方の上手さを向上するための改善案をフィードバックするシステムを構築する取り組みを行う．

対話や特定のタスクに関するスキルを自動で評価しそのフィードバックを提示するシステムの研究としてソーシャルスキル [4][5][6]，プレゼンテーションスキル [7][8] に関する研究がある．これまで我々は，対話中の言語・非言語行動から褒める行為を明らかにする取り組みを行っている [9][10][11][12]．さらに，1 発話中の非言語行動から褒め方の上手さを自動で評価した結果を提示する取り組みも行っている [13][14]．しかし，これらの研究では褒め方の上手さの推論や褒めているか否かの検出，単に褒め方の評価を提示するシステムにとどまっている．褒め方の評価を提示するのみでは，自身の褒め方の上手さを向上させることは困難であると考える．褒め方の上手さを向上させるためには，現在の自身の振る舞いと上手い褒め方の振る舞いを比較し，具体的にどのように振る舞いを変化させればよいのか把握する必要がある．

そこで本稿では，褒め方の上手さの評価を提示しながら，褒め方の上手さを向上させる改善案を提示するフィードバックシステムを実装する．本稿での研究課題は対話している映像を入力することで，褒め方の上手さを把握し向上させる手法を明らかにすることである．

2. 提案手法

1 章にて設定した研究課題を解決するために，対話中の行動をもとにコミュニケーションに関するスキルを訓練する研究 [4][5][6][7][8] を参考に褒め方の上手さのフィードバックシステムの構築を行う．褒め方の上手さを向上させるためのフィードバック手法として，表情特徴量に着目した．表情に関するフィードバックは視覚的なフィードバックの生成が可能であり，自身の振舞いと改善案の比較が容易である．話者が褒めている際の表情を，上手く褒められている際の表情に変化させた顔画像を提示することで，褒め方の上手さを向上できると考えた．そこで褒め方の上手さを向上させるフィードバック手法として，対話中の本人の顔画像とその画像の表情を上手く褒められている際の表情に変化させた顔画像を表示する手法を提案する．上手く褒められている際の表情は，褒めるシーンを収録した対話コーパスから，上手く褒められているシーンから抽出した表情特徴量を使用した．

1 日本大学大学院総合基礎科学研究科
2 日本大学文理学部
3 日本電信電話株式会社 NTT 人間情報研究所
a) miyata.akihiro@acm.org

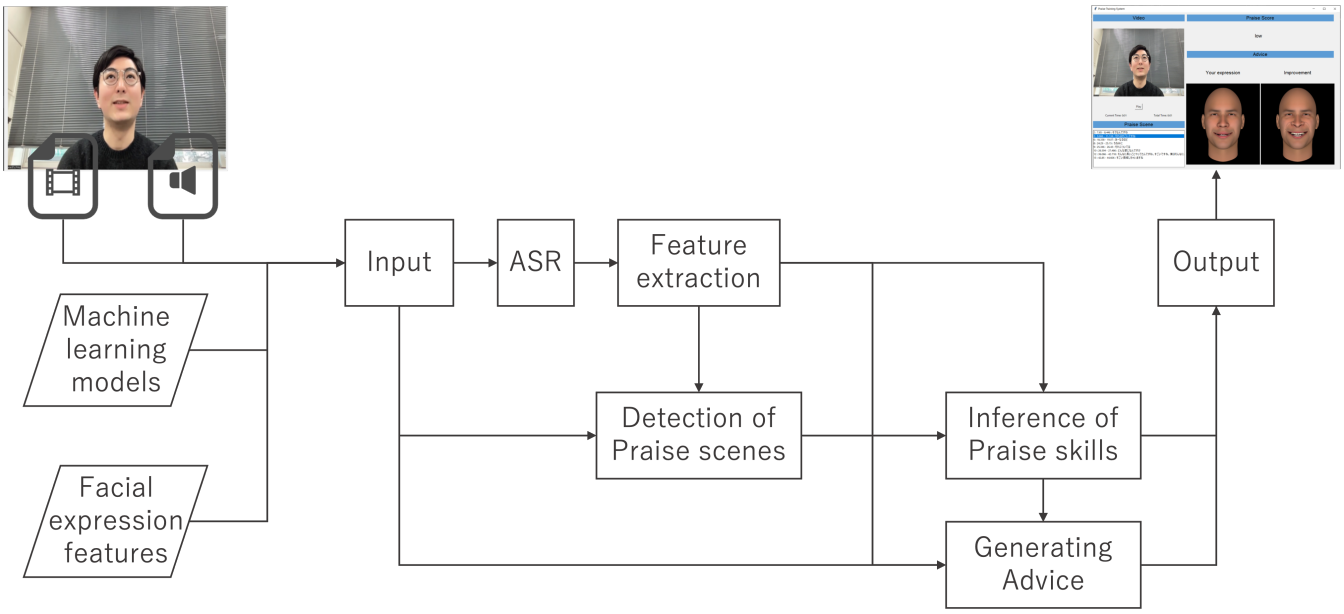

図 1 提案システムのフローチャート

3. 対話コーパス

本稿では先行研究にて構築された対話コーパスを使用した [11]．この対話コーパスには、2 者の対面対話コーパスと褒め方の上手さの評価データが含まれている．

2 者対話コーパスは，日本語による 2 者間対話を収録し、言語・非言語行動を記録したものである．2 者対話の参加者は 20 代の大学生 34 名（男性 28 名、女性 6 名）で，17 組に分けた．17 組のうち 14 組が初対面，2 組が知人，1 組が友人であった．対話は話し手と聞き手の役割を設けて実施した．話し手には過去頑張った話をしてもらい聞き手には話し手の話を聞き褒めるように教示した．収録された映像と音声をもとに人手で発話区間の付与と発話内容の書き起こしを行った．発話区間は Inter–Pausal Unit（IPU）[15] を使用し，無音区間が 400ms 未満の連続した音声区間とした．

褒め方の上手さの評価データは，5 名のアノテータが聞き手の各発話に対する褒め方の評価が記録されている．アノテータは，聞き手の各発話に対して褒めているか/褒めていないかの判定を行い，褒めていると判定した場合は 1（上手く褒められていない）〜7（上手く褒められている）の 7 段階で当該発話の褒め方の上手さを評価した．この時，アノテータ 3 名以上が褒めていると判定した発話は Praise シーンとし，当該シーンを褒めいていると判定したアノテータの評価の平均値を Praise スコアとした．Praise シーンは合計で 228 シーンであった．Praise シーンを下記のように Praise スコア低群，中群，高群の 3 クラスに分割した．

Praise スコア低群: Praise スコアが 3.8 点以下の Praise シーン （計 82 シーン）

Praise スコア中群: Praise スコアが 3.8 点より大きく 4.4 点未満の Praise シーン （計 65 シーン）

Praise スコア高群: Praise スコアが 4.4 点以上の Praise シーン （計 81 シーン）

本稿では Praise スコア高群を，上手く褒められているシーンとした．

4. 実装

2 章にて提案したシステムの具体的な実装について述べる．提案システムのフローチャートを**図 1** に示す．提案システムは，入力部，音声認識部，特徴量抽出部，褒める行為の検出部，褒め方の上手さ推論部，フィードバック生成部，出力部からなる．

4.1 入力部

入力部では，話者が対話している映像と音声データ，事前に構築された褒める行為の検出と褒め方の上手さを推論する機械学習モデル，フィードバック生成のための表情特徴量を受け取る．入力される映像データは，話者がカメラの正面を向いて対話しているシーンを撮影したデータである．音声データに関しては，話者本人の音声のみが収録されたデータである．機械学習モデルとフィードバック生成のための表情特徴量の構築については，以降の項で説明する．

4.2 音声認識部

音声認識部では，入力された音声から，発話されている区間の検出とその区間の書き起こしを行った．入力された音声データから音声区間検出モデルである Silero VAD[16]

を用いて音声区間を検出した．発話区間は先行研究において IPU 400ms 未満の連続した音声区間と定義されている．本稿でも先行研究を参考に検出された音声区間から発話区間を決定した．次に，音声認識モデルである whisper[17] を用いて検出した発話区間の発話内容の書き起こしを行った．

4.3 特徴量抽出部

発話区間から言語・非言語行動に関する特徴量を抽出した．抽出する特徴量は先行研究 [18][9][10][11] と同様とした．抽出した特徴量について順番に説明を行う．

4.3.1 視覚的特徴量

顔画像処理ツールである OpenFace[19] を用いて，入力された映像データから各話者の頭部・顔部の振舞いに関連する特徴量と Action Units [*1][20] に関する特徴量の抽出を行った．発話区間における頭部の回転角度，視線の回転角度，Action Units の強度について，それぞれ分散，中央値，10 パーセンタイル値，90 パーセンタイル値を算出し計 88 の特徴量を抽出した．

4.3.2 韻律的特徴量

音声情報処理ツールである openSMILE[21] を用いて，入力された音声データから，音声の韻律の代表的な特徴量，発話に関する特徴量の抽出を行った．音声の韻律の代表的な特徴量につき，統計量を算出した特徴量と，これらの各特徴量を一次微分したものの計 988 の特徴量を抽出した [22][23]．

4.3.3 言語的特徴量

言語的特徴量の抽出方法として，発話内容をベクトル化する方法を用いた．具体的には，日本語事前学習済み[*2]の BERT モデル [24] を利用し，褒める人の発話を 768 次元のベクトルに変換し，特徴量を抽出した．

4.4 褒める行為の検出部

褒める行為の検出部では，4.3 節にて抽出した特徴量を用いて発話区間における褒めているシーンの検出を行った．先行研究 [12] で構築された褒める行為の検出する機械学習モデルを参考に，3 章で説明した対話コーパスから 4.3 節と同様の特徴量を抽出し機械学習モデルの構築を行った．4.3 節で抽出した言語・非言語行動に関する特徴量を入力として，当該シーンが褒めているシーンであるかそうでないかを出力する機械学習モデルである．

4.5 褒め方の上手さの推論部

褒め方の上手さの推論部では，4.3 節にて抽出した特徴量を用いて褒めているシーンの褒め方の上手さの推論を行った．先行研究 [9][10] で構築した褒め方の上手さを推論する機械学習モデルをを参考に，3 章で説明した対話コーパスから 4.3 節と同様の特徴量を抽出し機械学習モデルの構築を行った．4.3 節で抽出した言語・非言語行動に関する特徴量を入力として，当該シーンの褒め方の上手さを 3 クラスで出力する機械学習モデルである．出力されるクラスは 3 章にて説明した 3 クラスと同様である．

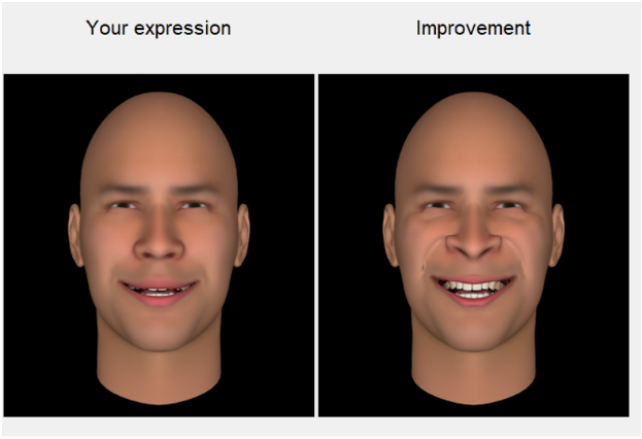

図 2 表情に関するフィードバックの生成例

4.6 フィードバックの生成部

褒め方の上手さを改善できるようにするために表情に関するフィードバックを生成した．発話中に表出した表情を使用した顔画像と，上手く褒められている際の表情を参考に，話者の表情を変化させた顔画像の生成を行った．上手く褒められている際の表情として，3 章の対話データにおける各 Action Units の強度を使用した．

上手く褒められている顔画像を生成するために，話者の表情から変化させる Action Units の対象と値を決定した．変化させる Action Units の対象は，Praise スコア高群と低群において各 Action Units の中央値の間で 5%水準で t 検定を行い，有意差のあった Action Units とした．変化せる Action Units の値は，Praise スコア高群の各 Action Units の値の平均値を使用した．顔画像の生成には，Action Units の強度を使用して表情を変化させた顔 3D CG モデルを作成できる FaceGen[25] を使用した．実際の出力例を**図 2** に示す．左側が当該発話区間の Action Units を使用して作成した顔画像，右側が有意差のあった Action Units を Praise スコア高群の Action Units の平均値に変化させて作成した顔画像となっている．

4.7 出力部

出力は，実際の対話中の映像，褒めている発話区間のリスト，当該シーンの褒め方の上手さの評価とフィードバックとした．実際の出力例を**図 3** に示す．左上には実際の対話中の映像，左下には褒めている発話区間のリスト，右上には当該シーンの褒め方の上手さの評価，右下には褒め

[*1] Action Units とは，頭部の向きや視線の角度と筋肉群の基本的な行動の単位を表す．
[*2] https://huggingface.co/cl-tohoku

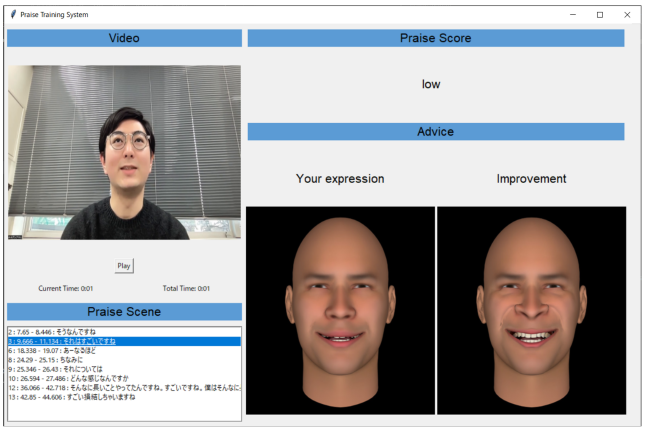

図 3 提案システムの出力例

方の上手さを向上させるための表情に関するフィードバックを表示した．まず，ユーザは左下に示されているリストから参照したい発話を選択する．発話を選択することで当該発話区間に対応する，映像，褒め方の上手さの評価とフィードバックが表示される．そして，ユーザは自身が褒めていたシーンと褒め方を確認することで自身の褒める行為を振り返ることが可能である．さらに，右下の表情に関するフィードバックを確認することで自身の表情をどのように変化させればよいか判断可能である．このように，自身が褒めているシーンを振り返り，もし上手く褒めれれていない場合は改善案の具体例を視覚的にユーザに提示することで，褒め方の上手さを向上することが可能であると考えている．

5. おわりに

本稿では対話している映像を入力することで，褒め方の上手さを向上させるための表情に関するフィードバックを行う手法を明らかにするために，表情特徴量に着目した褒め方の上手さのフィードバックの生成を行った．表情に関するフィードバックとして，上手く褒めるために重要である考えた Action Units の値を使用し，話者の Action Units を変化させた顔画像を生成し提示した．

本システムには制約がある．まず，変化させる対象の表情を決める際の Action Units の値が中央値であることである．表情は発話中に様々に変化している．そのため，瞬間的に変化がある Action Units については対応できていない可能性がある．今後は，表情の変化を考慮した上で，変化させる Action Units の指標を検討する必要がある．さらに，提示する顔画像についても発話中の表情の変化に対応させた映像として提示することが必要である．次に，本システムでは表情に関するフィードバックまでにとどまっていることである．褒める行為は表情以外の非言語行動や言語行動も伴っている．これらの行動に関するフィードバックも含めることでより褒め方の上手さの改善に貢献できると考えている．具体的な行動として，話し方や話す内容に関するフィードバックを生成することを予定している．

今後の展望として，実際に提案システムを使用して，褒め方の上手さにどのような影響があるのか実験を行う予定である．

参考文献

[1] Brophy, J.: Teacher praise: Afullctional analysis, *Review of Educational Research*, Vol. 51, No. 1, pp. 5–32 (1981).

[2] Kalis, T., Vannest, K. and Parker, R.: Praise Counts: Using Self-Monitoring to Increase Effective Teaching Practices, *Preventing School Failure*, Vol. 51, No. 3, pp. 20–27 (2007).

[3] Jenkins, L., Floress, M. 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).

[4] Saga, T., Tanaka, H., Iwasaka, H. and Nakamura, S.: Objective Prediction of Social Skills Level for Automated Social Skills Training Using Audio and Text Information, pp. 467–471 (2020).

[5] Saga, T., Tanaka, H., Matsuda, Y., Morimoto, T., Uratani, M., Okazaki, K., Fujimoto, Y. and Nakamura, S.: Automatic evaluation-feedback system for automated social skills training, *Scientific Reports*, Vol. 13 (2023).

[6] Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B. and Picard, R. W.: MACH: My Automated Conversation Coach, *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*, Association for Computing Machinery, pp. 697–706 (2013).

[7] Yagi, Y., Okada, S., Shiobara, S. and Sugimura, S.: Predicting multimodal presentation skills based on instance weighting domain adaptation, *Journal on Multimodal User Interfaces*, Vol. 16, pp. 1–16 (2021).

[8] Schneider, J., Börner, D., van Rosmalen, P. and Specht, M.: Presentation Trainer, Your Public Speaking Multimodal Coach, *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, Association for Computing Machinery, pp. 539–546 (2015).

[9] Onishi, T., Yamauchi, A., Ogushi, A., Ishii, R., Fukayama, A., Nakamura, T. and Miyata, A.: Modeling Japanese Praising Behavior by Analyzing Audio and Visual Behaviors, *Frontiers in Computer Science*, Vol. 4 (2022).

[10] Ogushi, A., Onishi, T., Tahara, Y., Ishii, R., Fukayama, A., Nakamura, T. and Miyata, A.: Analysis of praising skills focusing on utterance contents, *Proceedings of Interspeech '22*, pp. 2743–2747 (2022).

[11] Onishi, T., Ogushi, A., Tahara, Y., Ishii, R., Fukayama, A., Nakamura, T. and Miyata, A.: A Comparison of Praising Skills in Face-to-Face and Remote Dialogues, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Marseille, France, European Language Resources Association, pp. 5805–5812 (2022).

[12] 田原陽平，大西俊輝，大串 旭，石井 亮，深山 篤，中村高雄，宮田章裕：対面・遠隔対話からの称賛行為検出の基礎検討，情報処理学会グループウェアとネットワークサービスワークショップ 2022 論文集，Vol. 2022, pp. 36–43 (2022).

[13] 山内愛里沙，大西俊輝，武藤佑太，石井 亮，青野裕司，宮田章裕：表情・音声を用いた褒め方の上手さを評価するシステムの基礎検討，情報処理学会インタラクション

2020 論文集，Vol. 2020, pp. 247–252 (2020).
[14] 山内愛里沙，大西俊輝，呉　健朗，武藤佑太，石井　亮，青野裕司，宮田章裕：音声および視線・表情・頭部運動に基づく上手い褒め方の評価システムの検討，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 ’20)，Vol. 2020, pp. 98–106 (2020).
[15] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. and Den, Y.: An Analysis of Turn-Taking and Backchannels Based on Prosodic and Syntactic Features in Japanese Map Task Dialogs, *Language and Speech*, Vol. 41, No. 3-4, pp. 295–321 (1998).
[16] Team, S.: Silero VAD: pre-trained enterprise-grade Voice Activity Detector (VAD), Number Detector and Language Classifier, `https://github.com/snakers4/silero-vad` (2021).
[17] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C. and Sutskever, I.: Robust speech recognition via large-scale weak supervision, *International Conference on Machine Learning*, PMLR, pp. 28492–28518 (2023).
[18] Onishi, T., Yamauchi, A., Ishii, R., Aono, Y. and Miyata, A.: Analyzing Nonverbal Behaviors along with Praising, *Proceedings of the 22nd ACM International Conference on Multimodal Interaction (ICMI ’20)*, pp. 609–613 (2020).
[19] Baltrusaitis, T. and Robinson, P. and Morency, L.P.: OpenFace: An open source facial behavior analysis toolkit, *IEEE Winter Conference on Applications of Computer Vision (WACV ’16)*, pp. 1–10 (2016).
[20] Ekman, P. and Friesen, W.: Manual for the facial action coding system (1977).
[21] Eyben, F. and Wöllmer, M. and Schuller,B.: openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor, *Proceedings of ACM Multimedia*, pp. 1459–1462 (2010).
[22] Schuller, B. and Steidl, S. and Batliner, A.: The INTERSPEECH 2009 emotion challenge, *Proceedings of Interspeech ’09* (2009).
[23] Schuller, B. and Steidl, S. and Batliner, A. and Burkhardt, F. and Devillers, L. and Müller, C. and Narayanan, S.: The INTERSPEECH 2010 paralinguistic challenge, *Proceedings of Interspeech ’10*, pp. 2794–2797 (2010).
[24] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT ’19)*, pp. 4171–4186 (2019).
[25] Inversions, S.: FaceGen Modeller, `https://facegen.com/modeller.htm`.