

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

山内 愛里沙¹ 大西 俊輝² 武藤 佑太¹ 石井 亮³ 青野 裕司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。しかし、相手を上手く褒めるためには、具体的にどのような言語的・非言語的行動をとれば良いか明らかにされていない。このため、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がわからず、褒め方を上達させるトレーニングを行うことが困難であるという問題が存在する。そこで我々は、褒め方を上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムの構築を目指している。システムを構築するにあたり、表情・音声から褒め方の上手さを推定することができるか、褒め方の上手さを推定する上でどのような表情・音声情報が重要であるか明らかにする必要がある。本研究では、表情・音声から褒め方の上手さを推定することができるか検証し、どのような表情・音声情報が重要であるか明らかにする取り組みを行う。具体的には、実対話データから抽出された表情・音声に関する特徴量を用いて褒め方の上手さの評価値を推定する機械学習モデルを構築した。その結果、音声特徴量のみを用いたモデルよりも頭部・視線方向と表情特徴量のみを用いたモデルの性能が高く、さらに、頭部・視線方向と表情特徴量・音声特徴量を合わせたモデルが最も性能の高いモデルであることが明らかになった。次に、褒め方の上手さの評価値を推定する上で重要度の高い特徴量の分析を行った。その結果、頭部方向、視線方向、MFCC、顎の動きが重要であることが明らかになった。また、機械学習モデルを用いた褒め方の上手さの評価システムの構築により、ユーザに対して褒め方の上手さのスコアと、評価する上で重要であった特徴量を提示することが可能となった。

Study on a System for Rating Praising Behavior by Analysing Voice and Nonverbal Behavior

ARISA YAMAUCHI¹ TOSHIKI ONISHI² YUTA MUTO¹ RYO ISHII³ YUSHI AONO³
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。組織マネジメントの観点から、同僚や部下の行動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重要であるという認識が広がっている。褒める行為は、承認・称賛を表現すること、対象の良い側面を示唆する言語的・非言語的行動、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考えられている [1][2][3][4][5]。これらの既存知見より、対話相手を上手く褒めるためには、

言語的・非言語的行動を適切に用いる必要があると考えられる。しかし、対話相手を上手く褒めるために、言語的・非言語的行動をどのように用いれば良いか明らかにされていない。このため、相手を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がわからず、褒め方を上達させるトレーニングを行うことが困難であるという問題が存在する。このような背景より、自身の褒め方の上手さや、自身の褒め方においてどのような点が良かったかといったことを認識する必要がある。そこで我々は、褒め方を上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムの構築を目指している。これまでに我々は、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1つであり、頭部の縦軸回りの回転角度、

¹ 日本大学文理学部

² 日本大学大学院総合基礎科学研究科

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

a) miyata.akihiro@acm.org

視線の横方向の角度、顎を上げる動き、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を明らかにした [6][7]。また、表情・音声をを用いた褒め方の上手さを推定するモデルを構築し、褒め方の上手さの評価するシステムの基礎検討を行った [8]。しかし、褒め方の上手さの推定器の構築と褒め方の上手さを評価するシステムの基礎検討にとどまっており、推定器の詳細な評価や特徴量の分析は行われていない。そのため、褒め方の上手さを評価するシステムの構築にあたり、表情・音声から褒め方の上手さを推定することができるか、褒め方の上手さの評価を推定する上でどのような表情・音声情報が重要であるか明らかにする必要がある。そこで本研究では、表情・音声から褒め方の上手さを推定することができるか検証し、褒め方の上手さを推定する上で、どのような表情・音声情報が重要であるか明らかにする取り組みを行う。さらに、褒め方を上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムの構築を行う。本稿の貢献は、次のとおりである。

- 表情・音声から褒め方の上手さを推定することができるか、褒め方の上手さを推定する上で、どのような表情・音声情報が重要であるか明らかにしたこと。
- 褒め方を上達させるトレーニングを行うための、褒め方の上手さを評価するシステムの構築を行ったこと。

2. 関連研究

本研究は、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褒め方の上手さを評価するシステムの構築を目指すものであり、人間の言語的・非言語的情報を用いて対話中における行動・能力を分析する研究に関連している。

2.1 個人特性を推定する研究事例

Pianesi らは、BigFive と呼ばれる個人の性格特性（情緒不安定性・外向性・開放性・調和性・誠実性）を利用して、グループ対話における非言語的行動から個人の性格特性を推定する取り組みを行っている [9]。Batrincea らは、human-machine interaction (HMI) と human-human interaction (HHI) の 2 つのシナリオで収集された視聴覚データから、BigFive の出現度を分析する取り組みを行っている [10]。Biel らは、表情と BigFive を特徴量として用いて、動画での日常の出来事などを紹介する Vlogger の性格特性を推定する取り組みを行っている [11]。Aran らは、第三者による対象人物の性格に対する印象をアノテーションし、当該人物の性格特性を推定する取り組みを行っている [12]。Valente らは、対話における発話者の性格特性を推定する取り組みを行っている [13]。Lin らは、グループ活動における個人の音声をモデリングし、個人の性格特性を推定する取り組みを行っている [14]。

2.2 パフォーマンスを推定・評価する研究事例

Jayagopi らは、対面での発話や視線パターンから、グループ構成、対人知覚、グループのパフォーマンスといったグループ行動を分析する取り組みを行っている [15]。Nguyen らは、採用面接における応募者のコミュニケーション能力や採用に値する人材であるかを推定する取り組みを行っている [16]。Sanchez-Cortes らは、非言語的行動を用いて、グループディスカッションにおけるリーダーの出現を自動的に推定する取り組みを行っている [17]。Okada らは、対話における言語的・非言語的行動を用いてコミュニケーションスキルを分析する取り組みを行っている [18]。Ishii らは、話者交替/継続における視線行動と発話行為が、話者の共感スキルとどのような関係があるのか明らかにする取り組みを行っている [19]。Soleymani らは、対話における言語的・非言語的行動から自己開示レベルを推定する取り組みを行っている [20]。Wörtwein らは、非言語的行動を用いて、演説スキルの評価を推定する取り組みを行っている [21]。Ramanarayanan らは、言語的・非言語的行動を用いて、プレゼンテーションスキルの評価を推定する取り組みを行っている [22]。Tanaka らは、人間とコンピュータの相互作用によるソーシャルスキルトレーニングを行う対話システムを構築する取り組みを行っている [23]。Fukuda らは、熟練した高校教員によって行われたヒューマンアセスメントに基づき、教員のパフォーマンスを自動評価するためのモデルを構築する取り組みを行っている [24]。八木らは、プレゼンテーションデータから得られた言語的・非言語的行動を用いて、プレゼンテーションスキルの推定モデルを構築・評価する取り組みを行っている [25]。Chen らは、言語的・非言語的行動を用いて、プレゼンテーションスキルをスコアリングするシステムを構築する取り組みを行っている [26]。

3. 研究課題

人間の言語的・非言語的行動を用いて対話における行動や能力の推定、評価を行う研究事例は数多く存在するが、我々が行う褒め方の上手さに関する研究とは異なる点が存在する。[9], [10], [11], [12], [13], [14] は、対話全体を通じた人間の言語的・非言語的行動を用いて個々の性格特性を推定する取り組みを行っている。本研究においても人間の言語的・非言語的行動を用いて対話中の行動や能力の推定を行うが、対話全体を通してではなく個々の褒める行為に着目しているという点で異なる。[15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26] は、対話における言語的・非言語的行動を用いて、コミュニケーションスキルやパフォーマンスを推定・評価する取り組みや、スキルを評価するシステムを構築する取り組みを行っている。本研究においても対話における言語的・非言語的行動を用いた取り組みを行うが、推定・評価する対象が褒め方の上

手さであるという点で異なる。以上より、既存研究において、言語的・非言語的行動を用いて褒める行為のみに着目した分析、褒め方の上手さの推定、褒め方の上手さの評価システムを構築する取り組みは行われていない。

そこで我々は、上手く褒めるために人間がどのような言語的・非言語的行動を用いているのかを定量的な分析から明らかにし、褒め方を上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムの構築を目指している。これまでに我々は、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1つであること、さらに、対話における褒め方の上手さは、褒める人の頭部の縦軸回りの回転角度、視線の横方向の角度、顎を上げる動き、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を明らかにした [6][7]。しかし、褒め方の上手さに関係すると考えられる音声は扱われていない。また、表情・音声を用いて褒め方の上手さを推定する回帰モデルを構築し、推定モデルを用いた褒め方の上手さの評価システムの構築を行った [8]。しかし、褒め方の上手さの推定においてどの程度音声情報が有用であるか、具体的にどのような音声情報が重要であるか明らかにされていない。

本研究では、褒め方の上手さに関係すると考えられる表情・音声情報を用いて、表情・音声から褒め方の上手さを推定することができるか検証し、褒め方の上手さを推定する上で、どのような表情・音声情報が重要であるか明らかにすること、さらに、褒め方の上手さを上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムを構築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。

- **Step1 : 2 者対話の収録**

2 名 1 組の 2 者対話を 16 組分、計 160 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。

- **Step2 : アノテーション**

収録した映像データと音声データに対して、アノテーションを行う。

- **Step3 : 褒め方の上手さの評価**

話し手（褒められる人）であった参加者による、聞き手の褒め方の上手さの評価を行う。

- **Step4 : 特徴量抽出**

聞き手（褒める人）の映像データと音声データから、頭部・視線方向と表情特徴量、音声特徴量を抽出する。

- **Step5 : 褒め方の上手さに寄与する特徴量の分析**

褒め方の上手さを評価するにあたり、重要な特徴量を明らかにする。

- **Step6 : 褒め方の上手さの評価システムの構築**

褒め方の上手さの推定モデルを用いて、対話における褒め方の上手さの評価システムを構築する。

4. 対話コーパス

4.1 2 者対話の収録

2 者対話における、褒め方の上手さの評価値と各モダリティの記録を含む対話コーパスの作成を行った。2 者対話の参加者は、20 代の大学生 34 名（男性 28 名、女性 5 名）であり、2 名 1 組のペアを 17 組構成した。17 組のうち、初対面が 14 組、顔見知り 2 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、今まで頑張ってきたことについて 2 つ以上のエピソードを用意してもらった。参加者は、図 1 のように互いに向き合って着座する。参加者間の距離は 180cm とした。対話の収録は、各参加者の様子と 2 者対



図 1 2 者対話の様子

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を用いて行った。各組の参加者達（参加者 A、参加者 B）は、撮影者の合図に従い、次の (1) ~ (3) を行った。

(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 17 組分、計 170 分間の 2 者対話を収録する。対話において話し手となる参加者には、事前に用意した今まで頑張ってきたことについて話すように指示した。会話の自然さや話題の多様性を担保するために、事前に用意しなかった今まで頑張ってきたことについて話すことを許可した。対話において聞き手となる参加者には、話し手の発言内容について褒めるように指示した。しかし、一方的に褒めているだけのような不自然な対話にならないようにするために、自由に質問したり、リアクションしたりするこ

とを許可した．なお，(1) 自己紹介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

音声データや映像データに対して注釈付けを行うツールである ELAN[27] を用いて，収録した音声データと映像データに対して表 1 に示すアノテーション項目を手で付与した．発話シーンは，沈黙時間が 200 ミリ秒未満の連続

表 1 アノテーションの項目

項目	説明
発話シーン	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
Praise シーン	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

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．Praise シーンは，図 2 に示すように，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である．判断基準が揺るがないよう，1

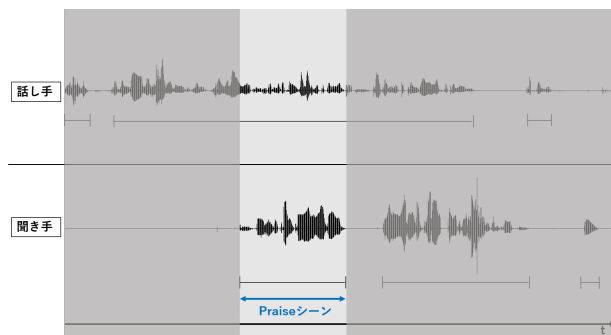


図 2 Praise シーンの抽出範囲

名のアノテータが全ての Praise シーンの抽出を行った．本研究では，聞き手の参加者の発話シーンから，相手のエピソードを認めていて，相手の良い側面を称賛しているシーンとした．これらのアノテーションで抽出した情報を表 2 に示す．参加者 34 名の話し手の発話シーンが 4,267 件，

表 2 アノテーションで抽出したデータの情報

項目	件数
話し手の発話シーン	4,267
聞き手の発話シーン	3,277
Praise シーン	327

聞き手の発話シーンが 3,277 件，Praise シーンが 327 件であった．

4.3 褒め方の上手さの評価

対話終了後に，4.2 節で抽出した Praise シーンについて，当該シーンで話し手であった参加者に，聞き手であった参加者の褒め方の上手さの評価を行ってもらった．具体的には，当該シーンで話し手であった参加者に聞き手であった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を参照してもらい，当該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，1（気分が良くなったと思わない）～7（気分が良くなったと思う）の 7 段階で評価してもらった．

5. モデル構築

本研究では，頭部・視線方向と表情特徴量，音声特徴量を用いた褒め方の上手さの評価値を推定するモデルを構築した．具体的には，4.1 節で収録したデータから頭部・視線方向と表情特徴量，音声特徴量の抽出を行い，これらの特徴量を用いて褒め方の上手さの評価値を推定する回帰モデルの構築を行った．さらに，褒め方の上手さの評価値を推定した結果から，褒め方の上手さの評価値を推定する上で重要であると考えられる特徴量について分析を行った．

5.1 特徴量抽出

5.1.1 頭部・視線方向と表情特徴量の抽出

顔画像処理ツールである OpenFace[28] を用いて，聞き手である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，視線，顔部における筋肉群の基本的な行動の単位を表す Action Units[29] に関する特徴量を抽出した．頭部に関する特徴量として，ビデオカメラ側から顔を見て左から右方向を x 軸，下から上方向を y 軸，手前から奥方向を z 軸とした場合，Praise シーンの前後 1 秒ずつを含む範囲における頭部の x 軸，y 軸，z 軸回りの回転角度 (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．視線に関する特徴量として，ビデオカメラ側から顔を見て左から右方向を x 軸，下から上方向を y 軸とした場合，Praise シーンの前後 1 秒ずつを含む範囲における視線の x 軸，y 軸方向の角度 (gaze_Ax , gaze_Ay) 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 に関する特徴量としては，表 3 に示す各 Action Units の強度の Praise シーンの前後 1 秒ずつを含む範囲における平均値を用いた．全 Praise シーンに対して上記の特徴量を抽出し，特徴量の取りうる値の範囲を揃える為に，各特徴量を平均値 0，分散 1 になるように正規化を行った．

5.1.2 音声特徴量の抽出

音声情報処理ツールである openSMILE[30] を用いて，聞き手である参加者に取り付けたマイクで録音した音声データから，音声の韻律の代表的な特徴量，発話に関する特徴

表 3 Action Units の内容

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	瞼を緊張させる	AU25	顎を下げて唇を開く
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		

量の抽出を行った。音声の韻律の代表的な特徴量として、具体的には、表 4 に示す各特徴量につき、表 5 に示す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（_de）の計 384 の特徴量を抽出した。なお、メル周

表 4 音声特徴量の内容

特徴量	内容
pcm_RMSenergy_sma	音のエネルギーの二乗平均平方根
pcm_fftMag_mfcc_sma[1]-[12]	1~12 次のメル周波数ケプストラム係数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数

表 5 抽出可能な統計量

統計量	内容	統計量	内容
_max	最大値	_linregc1	線形近似の勾配
_min	最小値	_linregc2	線形近似のオフセット
_range	最大値と最小値の差	_linregerrQ	線形近似の二乗誤差
_maxPos	最大値の絶対位置	_stddev	標準偏差
_minPos	最小値の絶対位置	_skewness	歪度
_amean	平均値	_kurtosis	尖度

波数ケプストラム係数とは、対数ケプストラム（20 次）の低次成分（1~12 次）に対して、人間の周波数知覚特性を考慮して重み付けをしたものであり、声道特性を表す特徴量である。また、ゼロ交差率とは、音の波形が音圧が 0 である軸を交差する割合を表す特徴量である。発話に関する特徴量としては、Praise シーンにおける聞き手の発話の長さ（秒）を用いた。5.1.1 項と同様に、全 Praise シーンに対して上記の特徴量を抽出し、各特徴量を平均値 0、分散 1 になるように正規化を行った。

5.2 モデル構築

本研究では、褒め方の上手さの評価値を推定するモデルとして、4.3 節で行った褒め方の上手さの評価値（以降、Praise スコアとする）を目的変数、5.1.1 項と 5.1.2 項で抽出した頭部・視線方向と表情特徴量、音声特徴量を説明変数とする回帰モデルを構築した。言語的・非言語的行動と褒め方の上手さの関連性を分析する手法として、説明変数の重要度を評価することができる Gradient boosting[31] を用いた。Gradient boosting は、弱学習器を 1 つずつ順番に構築していく手法であり、新しい弱学習器を構築する際に、

それまでに構築された全ての弱学習器の結果を利用することで高精度な推定モデルを構築可能である。学習率や木の深さなどのハイパーパラメータは、Hyperopt[32] でチューニングした。データセットを 90%の訓練データ、10%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 スコアを推定するタスクを 100 回行ったところ、各指標の平均値は表 6 のようになった。ベースラインは Praise スコアの平均値を推定するモデル、

表 6 各指標の平均値 (N=100)

指標	ベースライン	F	V	F+V
MSE	1.032	0.796	0.918	0.705
R ²	-0.033	0.198	0.102	0.216

F は頭部・視線方向と表情特徴量を用いたモデル、V は音声特徴量を用いたモデル、F+V は頭部・視線方向と表情特徴量、音声特徴量を用いたモデルを表している。また、MSE は平均二乗誤差（Mean Squared Error）、R² は決定係数を表している。モデル構築に用いた頭部・視線方向と表情特徴量、音声特徴量は、聞き手であった参加者から抽出したものである。モデル F+V における各特徴量の重要度（gain）の上位 10 個の結果を図 3 に示す。

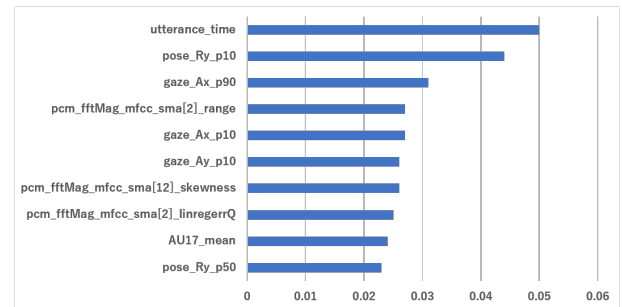


図 3 モデル F+V における特徴量の重要度の上位 10 個

5.3 考察

表 6 より、ベースラインの決定係数はほぼ 0 であり、推定モデルとしては機能していないと言える。一方で、本研究で構築したモデル F、モデル V、モデル F+V はベースラインよりも高い決定係数が算出された。ベースラインとこれら 3 つのモデルそれぞれに対して、対応のある t 検定を行ったところ、ベースラインとモデル F、ベースラインとモデル V、ベースラインとモデル F+V 全ての組み合わせにおいて 1%水準で有意差が認められた。このことから、モデル F、モデル V、モデル F+V は、ベースラインよりも性能を向上させることができたと考えられる。また、モデル F、モデル V 間には 1%水準で有意差が認められたことから、音声特徴量のみを用いたモデル V よりも頭部・視線方向と表情特徴量を用いたモデル F の方が性能が高いと考えられる。さらに、モデル F、モデル F+V 間には 5%水

準で有意差が認められたことから、頭部・視線方向と表情特徴量のみを用いたモデル F よりも、頭部・視線方向と表情特徴量、音声特徴量を用いたモデル F+V の方が性能が高いと考えられる。このことから、表情に関する情報に音声情報を加えることにより、褒め方の上手さの評価値を推定するモデルの性能を向上させることができ、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と言える。以上より、音声情報は褒め方の上手さの推定に有用な情報であり、頭部・視線方向と表情特徴量と合わせてモデルを構築することで、モデルの性能を向上させることができると考えられる。

次に、構築したモデルにおいて重要度の高かった特徴量について考える。図 3 より、utterance_time, pose_Ry, gaze_Ax, gaze_Ay, pcm_fftMag_mfcc_sma[2], pcm_fftMag_mfcc_sma[12], AU17 に関する特徴量が重要度の高い特徴量であった。

utterance_time は、Praise シーンにおける聞き手の発話の長さ（秒）である。今回の対話シーンでは、発話の長さが比較的長い 2 秒以上の発話に対して Praise スコアが高く評価されている傾向があった。聞き手であった参加者からは、相手を褒めるときに大切にしたこととして、“どういふ部分がすごいと思ったのかを伝えるようにした”、“ただ反射的に褒めるのではなく、具体性を持った褒め方を行うようにした”といった意見があった。相手に“すごい”などと思ったことを伝える際には、“すごい”という言葉だけでなく、なぜそう思ったのか理由とともに伝えることが重要であると考えられている [33]。このことから、相手を上手く褒めるためには、相槌を打つような一言の短い発話で褒めるよりも、具体的な内容を含めた比較的長い発話で褒めることが重要であると考えられる。

pose_Ry は、頭部の y 軸回り（左右方向）の回転角度である。また、gaze_Ax は視線の x 軸方向（左右方向）の角度、gaze_Ay は、視線の y 軸方向（上下方向）の回転角度である。これらの特徴量の値は、Praise シーンの序盤から中盤にかけて変化し、中盤から終盤にかけては値の変化が小さくなる傾向があった。このことから、褒める発話の開始直後に相手の顔・目を見るという動作を行い、相手を褒めている間は視線や顔の方向を変えることがなかったのではないかと考えられる。対話相手の方向を見ることや対話相手と目を合わせるということは、相手との対話に参加している・関心があることを表すと考えられており、上手く褒めるためには対話内容に関心があることを示すことが重要であると考えられている [5]。聞き手であった参加者からも、相手の話を聞くときに大切にしたこととして、“褒める時はなるべく相手の方向を見るようにした”、“相手の目を見ること”、“目を見て話を聞くこと”といった意見があった。

pcm_fftMag_mfcc_sma[2], pcm_fftMag_mfcc_sma[12] は、

メル周波数ケプストラム係数（MFCC）のうち、それぞれ 2 番目、12 番目の係数である。メル周波数ケプストラム係数は、人間の聴覚特性を考慮してスペクトルの概形を表現した音色に関する特徴量である。音声認識や話者識別に有用な特徴量であり、話者個人の音声の特徴を詳細に表すものであると考えられる。今回モデル構築に用いたデータセットは、34 名分の音声記録されており、各個人の音声の特徴が推定に影響していたことが考えられる。本研究で構築するシステムを用いて褒め方の上手さを評価するにあたり、どの程度ユーザ個人の音声の特徴が影響するか詳細な分析を行う必要があるが、褒め方の上手さを評価する上で重要な特徴量であると考えられる。

AU17 は、顎を上げる動きを表しており、人間が自信があるときに表れる動きであると考えられている [34]。しかし、顎を上げる程度によって、その動きに表されている感情が異なる可能性がある。大きく顎を上げる動きの場合、不満といった感情を表すと考えられるが、わずかに顎を上げる動きの場合は個人の癖であるという可能性も考えられる。本研究では、顎を大きく上げているのかわずかに上げているのかといった検証はできていないため、今後、実際のデータをもとに詳細な分析を行い、顎を上げる動きにどのような感情が表れているのか明らかにする必要がある。

6. システム構築

5.2 節で構築した褒め方の上手さの推定モデルを用いて、ユーザに対して褒め方の上手さのスコア、さらに、本システムで褒め方の上手さを評価するにあたり重要度の高かった特徴量を提示するシステムの構築を行う。本システム

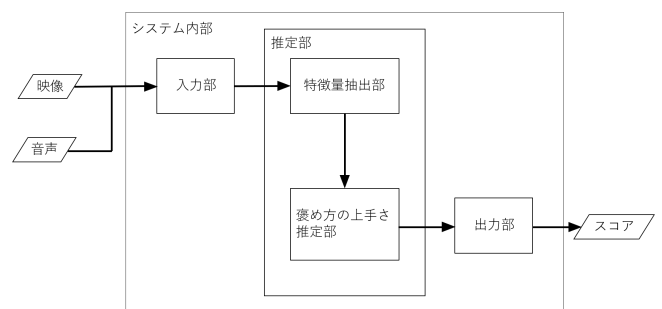


図 4 システムの構成図

は、図 4 に示す構成図のように、映像と音声の入力処理を行う入力部、褒め方の上手さを推定する推定部、推定結果の出力処理を行う出力部によって構成される。

6.1 入力部

入力部では、ユーザの顔部をビデオカメラで撮影した映像と、マイクから録音した音声に対して入力処理を行う。入力処理を行った後、映像データ、音声データを推定部へ渡す。

6.2 推定部

6.1 節の入力データに対して、特徴量抽出と褒め方の上手さの推定を行う。特徴量抽出では、6.1 節の映像データと音声データから、5.1.1 項と 5.1.2 項で示した頭部・視線方向と表情特徴量、音声特徴量を抽出し、抽出された頭部・視線方向と表情特徴量、音声特徴量のマージを行う。褒め方の上手さの推定では、マージされた特徴量と 5.2 節で構築した褒め方の上手さの推定モデルを用いて、褒め方の上手さの推定と重要度の高い特徴量の算出を行う。具体的には、5.2 節で用いたデータセットを訓練データ、マージされた特徴量を検証データとし、5.2 節で構築した表情・音声の特徴量としたモデルを用いて、Praise スコアの推定、重要度の高い特徴量の算出を行う。推定された Praise スコアと、褒め方の上手さを評価するにあたり重要度の高かった特徴量を出力部へ渡す。

6.3 出力部

褒め方の上手さの評価結果をディスプレイに出力する処理を行う。具体的には、6.2 節で算出された Praise スコアを、1~100 の数値にスケーリングし、100 点を満点としたスコアと重要度の高い特徴量上位 5 つをユーザに提示する。図 5 に、本システムの出力例を示す。

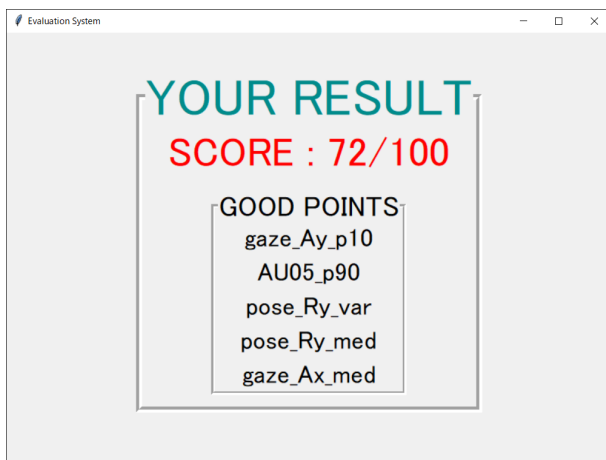


図 5 スコアの出力例

7. おわりに

本研究では、表情・音声から褒め方の上手さを推定することができるか検証し、褒め方の上手さを推定する上で、どのような表情・音声情報が重要であるか明らかにするために、対話コーパスの作成、褒め方の上手さの評価値を推定する機械学習モデルの構築を行った。対話コーパスの作成では、褒め方の上手さの評価値と各モダリティの記録を含む 2 者対話を収録し、アノテーションを行った。2 者対話の収録後、話し手であった参加者による、聞き手の褒め方の上手さの評価を行った。機械学習モデルの構築では、

収録した映像データと音声データから、それぞれ頭部・視線方向と表情特徴量、音声特徴量の抽出を行い、これらの特徴量を用いて褒め方の上手さの評価値を推定する回帰モデルを構築した。構築したモデルは、頭部・視線方向と表情特徴量のみを用いたモデル、音声特徴量のみを用いたモデル、頭部・視線方向と表情特徴量・音声特徴量を用いたモデルであり、これらの 3 つのモデルの比較を行った結果、頭部・視線方向と表情特徴量、音声特徴量を用いたモデルの性能が最も高く、褒め方の上手さの推定には音声情報が有用であることが明らかになった。褒め方の上手さの評価値を推定した結果から、褒める発話の長さ、頭部の左右方向の回転角度、視線の左右方向および上下方向の回転角度、MFCC、顎を上げる動きが褒め方の上手さを推定する上で重要であることが明らかになった。さらに、構築したモデルを用いて、褒め方の上手さを上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムを構築したことで、ユーザに対して褒め方の上手さのスコアと、評価する上で重要であった特徴量を提示することが可能となった。

本稿では、表情・音声からある程度妥当に褒め方の上手さの評価値を推定できること、褒め方の上手さを推定する上で重要である頭部・視線方向と表情特徴量、音声特徴量が明らかになったが、推定精度には向上の余地があり、対話相手を上手く褒めるためには、具体的にどのような表情・音声で褒めることが重要であるかは明らかにできていない。また、本稿においては頭部・視線方向と表情特徴量、音声特徴量のみを扱った分析を行ったが、相手を上手く褒めるためには発話内容やジェスチャーも重要であるため、これらのモダリティも含めて分析する必要があると考えられる。我々は、褒め方を上達させるトレーニングを行うことができる褒め方の上手さを評価するシステムの構築を目指しているが、今回構築した評価システムを利用して実際にユーザが褒め方のトレーニングを行うことで、褒め方のスキルを向上させることができるかは明らかになっていない。今後は、褒め方の上手さの推定に利用するモダリティを拡張し、推定精度の向上や、上手く褒めるためには具体的にどのような表情・音声で褒めれば良いかについての分析に取り組んでいきたい。さらに、ユーザが実際に本システムを利用して褒め方の上手さのトレーニングを行うことで、ユーザ自身の褒め方のスキルを向上させることが可能であるか検証を行っていきたい。

参考文献

- [1] Brophy, J.: Teacher praise: A functional analysis. Review of Educational Research, Vol.51, No.1, pp.5-32 (1981).
- [2] 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).
- [3] 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).
 - [4] 永崎一則：元氣と幸せを生み出す ほめかたの研究。早稲田教育出版 (2015)。
 - [5] 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)。
 - [6] 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 (DICOMO 2019), Vol.2019, pp.656-662 (2019)。
 - [7] 大西俊輝, 柴田万里那, 山内愛里沙, 呉健朗, 石井亮, 富田準二, 宮田章裕：褒め方の上手さの推定における頭部・顔部の効果。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 2019 論文集, Vol.2019, pp.1-6 (2019)。
 - [8] 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：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, pp.247-252 (2020)。
 - [9] Pianesi, F., Mana, N., Cappelletti, A., Lepri, B. and Zancanaro, M.: Multimodal recognition of personality traits in social interactions. *Proceedings of the 10th international conference on Multimodal Interfaces (ICMI '08)*, pp.53-60 (2008).
 - [10] Batrinca, L., Mana, N., Lepri, B., Sebe, N. and Pianesi, F.: Multimodal Personality Recognition in Collaborative Goal-Oriented Tasks. *IEEE Transactions on Multimedia*, Vol.14, No.4, pp.659-673 (2016).
 - [11] Biel, J.I., Teijeiro-Mosquera, L. and Gatica-Perez, D.: FaceTube: predicting personality from facial expressions of emotion in online conversational video. *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*, pp.53-56 (2012).
 - [12] Aran, O. and Gatica-Perez, D.: One of a kind: inferring personality impressions in meetings. *Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI '13)*, pp.11-18 (2013).
 - [13] Valente, F., Kim, S. and Motlicek, P.: Annotation and Recognition of Personality Traits in Spoken Conversations from the AMI Meetings Corpus. *Proceedings of INTERSPEECH*, pp.1183-1186 (2012).
 - [14] Lin, Y.S. and Lee, C.C.: Using Interlocutor-Modulated Attention BLSTM to Predict Personality Traits in Small Group Interaction. *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, pp.163-169 (2018).
 - [15] Jayagopi, D., Sanchez-Cortes, D., Otsuka, K., Yamato, J. and Gatica-Perez, D.: Linking speaking and looking behavior patterns with group composition, perception, and performance. *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*, pp.433-440 (2012).
 - [16] Nguyen, L.S., Frauendorfer, D., Mast, M.S. and Gatica-Perez, D.: Hire me: Computational Inference of Hirability in Employment Interviews Based on Nonverbal Behavior. *IEEE Transactions Multimedia*, Vol.16, No.4, pp.1018-1031 (2014).
 - [17] Sanchez-Cortes, D., Aran, O., Mast, M. S. and Gatica-Perez, D.: A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia*, Vol.14, No.3, pp.816-832 (2012).
 - [18] Okada, S., Nguyen, L. S., Aran, O. and Gatica-Perez, D.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol.15, No.1, pp.13:1-13:30 (2019) .
 - [19] 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. *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, pp.31-39 (2018).
 - [20] Soleymani, M., Stefanov, K., Kang, S. H., Ondras, J., Gratch, J.: Multimodal Analysis and Estimation of Intimate Self-Disclosure. *Proceedings of the 2019 ACM on International Conference on Multimodal Interaction (ICMI '19)*, pp.59-68 (2019).
 - [21] Wörtwein, T., Chollet, M., Schauerte, B., Morency, L. P., Stiefelwagen, R. and Scherer, S.: Multimodal Public Speaking Performance Assessment. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp.43-50 (2015).
 - [22] Ramanarayanan, V., Leong, C. W., Chen, L., Feng, G. and Suendermann-Oeft, D.: Evaluating Speech, Face, Emotion and Body Movement Time-series Features for Automated Multimodal Presentation Scoring. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp.23-30 (2015).
 - [23] Tanaka, H., Sakti, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H. and Nakamura, S.: Automated Social Skills Trainer. *Proceedings of the 20th International Conference on Intelligent User Interfaces (IUI '15)*, pp.17-27 (2015).
 - [24] Fukuda, M., Huang, H. H., Kuwabara, K., Nishida, T.: Multimodal Assessment on Teaching Skills in a Virtual Rehearsal Environment. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (IVA '19)*, pp.191-193 (2019).
 - [25] 八木 悠太朗, 岡田 将吾, 塩原 翔太, 杉村 聡太: ビジネスプレゼンテーションにおける言語・非言語的能力の自動推定. 人工知能学会全国大会論文集, pp.1-3 (2019).
 - [26] Chen, L., Feng, G., Joe, J., Leong, C.W., Kitchen, C. and Lee, C.M.: Towards Automated Assessment of Public Speaking Skills Using Multimodal Cues. *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp.200-203 (2014).
 - [27] Brugman, H. and Russel, A.: Annotating Multimedia Multi-modal resources with ELAN. *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC '04)*, pp.2065-2068 (2004).
 - [28] 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).
 - [29] Ekman, P. and Friesen, W. V.: *Manual for the Facial Action Coding System*. Palo Alto: Consulting Psychologists Press (1977).
 - [30] Eyben, F., Wöllmer, M. and Schuller, B.: openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor. *Proceedings of ACM Multimedia*, pp.1459-1462 (2010).
 - [31] Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* Vol.29, No.5, pp.1189-1232 (2001).
 - [32] Bergstra, J., Yamins, D. and Cox, D. D.: Hyperopt: A

Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. Proceedings of the 12th Python in Science Conferences (SciPy '13), pp.13–20 (2013).

- [33] 鈴木義幸: コーチングのプロが教える「ほめる」技術. 株式会社日本実業出版社 (2009).
- [34] 西松真子: 魅せる技術. インデックコミュニケーションズ (2004).