

遠隔対話における上手い褒め方のモデリングの基礎検討

大串 旭¹ 大西 俊輝¹ 田原 陽平² 石井 亮³ 深山 篤³ 中村 高雄³ 宮田 章裕^{2,a)}

概要: 褒める行為は、コミュニケーションにおいて重要な要素である。しかし、相手を上手く褒めるためには、どのように振舞うとよいのか明らかにされていない。この問題に鑑み、これまで我々は対面対話において上手く褒めるために重要な振舞いを明らかにしてきた。本稿では、近年重要性が増す遠隔対話に着目し、上手く褒めるために重要な振舞いを分析する。具体的には、話者（褒める人・褒められる人）の言語・非言語行動から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、上手く褒めるための言語・非言語行動を分析する。これより、遠隔対話における褒め方の上手さを推定できることが確認でき、褒める人の視覚的、韻律的特徴量と褒められる人の韻律的特徴量が重要であることが示唆された。

1. はじめに

褒める行為は日常生活や社会活動において重要なコミュニケーションの1つであり、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、褒められる人の役割は褒める人の役割と同じくらい重要であることが考えられている [3]。だが、相手を上手く褒めることが苦手な人も少なくない。褒め方の上手さを向上させるためには、まず自分自身の褒め方の上手さを理解する必要があると考えられる。しかし、上手く褒めるために重要な行動が明らかでないため、自分の褒め方の上手さを理解することは困難である。この問題を解決するために、我々是对話において相手を上手く褒めるための重要な行動を明らかにする取り組みを行う。

多くの研究では、対話や特定のタスクにおける話者の行動を分析し、コミュニケーションスキル [4][5]、プレゼンテーションスキル [6][7]、共感 [8][9][10]、自己開示 [11]、説得力 [12] を明らかにしている。我々はこれらの研究事例と同様の方法で、対話における褒め方の上手さを分析する。先行研究では、対面対話中の頭の動きや表情、声の出し方などの非言語行動や話者の発話内容などの言語行動から上手く褒めるための振舞いを明らかにする取り組みを行っている [13][14][15]。これらの研究では、褒める人の視覚的・韻律的特徴量と褒められる人の視覚的特徴量が上手く褒め

るために重要であるということと、言語行動は非言語行動に比べて上手く褒めるために重要であるということを明らかにしている。さらに、対面対話中と遠隔対話中における褒める行為の視覚的特徴量に違いがあるのか分析を行っている [16]。その結果、上手く褒めるための行動は対面対話と遠隔対話で違いがあることを明らかにしている。しかし、これらの研究では遠隔対話中の言語・非言語行動を用いて褒め方の上手さを推定できるのか明らかにされていない。COVID-19の影響により、遠隔でコミュニケーションを行う機会が増えてきている。対面対話と遠隔対話では、言語・非言語情報の伝達のされ方に差異があることが明らかになっている [17]。しかし、先行研究 [16] では遠隔対話コーパスの作成と視覚的特徴量に関する振舞いの分析にとどまっており、相手を上手く褒めるためにはどのような振舞いが重要であるのか明らかにされていない。

そこで本稿では、遠隔対話中の話者の言語・非言語行動から褒め方の上手さを推定する機械学習モデルを構築することで、褒め方の上手さを推定するために重要な特徴量を明らかにする。本稿での研究課題は、遠隔対話中の言語・非言語行動を用いて、褒め方の上手さを推定する機械学習モデルを構築し、重要な特徴量を明らかにすることとする。

2. 対話コーパス

本研究では先行研究で構築した対話コーパス [16] を利用した。このコーパスには、遠隔による2者対話データと褒め方の上手さの評価値が記録されている。

2.1 2者対話データセットについて

オンラインコミュニケーションツールである Zoom [18]

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 2者対話の様子

表 1 発話区間と Praise シーンの情報

	Scenes	Mean (sec)	Max (sec)	Min (sec)
発話区間 (褒める人)	16351	1.233	11.180	0.450
発話区間 (褒められる人)	27834	1.959	13.430	0.450
Praise シーン	236	1.998	10.610	0.470

を用いて、2者遠隔対話中の映像・音声・発話情報を記録したデータである。話者は20代~50代の40名(男性20名,女性20名)であり、それぞれ異なる相手と3回ずつ対話が行われている。これより、合計60組の対話が記録されている。各組の参加者の年齢層・性別は同じであり、初対面*1であった。話者は図1に示すように、一般的な遠隔対話シーンを想定し机の上に置かれたPCの前に座り画面に映る対話相手を見ながら会話をしている。このデータセットは下記の対話1~対話2の対話60組分、計1200分間からなる。

対話1 話者Aが褒める人となり、話者Bが褒められる人となる対話(10分間)

対話2 話者Bが褒める人となり、話者Aが褒められる人となる対話(10分間)

各話者の映像は話者の前に設置されたPCの内蔵カメラで、各話者の音声はPCの内蔵マイクで記録された。上記で記録された映像データと音声データに対して、自動音声認識システム*2を利用して、自動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、沈黙時間が400ミリ秒未満の連続した音声区間とした(表1)。

2.2 アノテーション

2者対話の収録に参加していない第三者のアノテータ5名が褒め方の上手さを評価した。各アノテータは文系・理系の教育を受けている20代の大学生であり、日常生活に支障のないコミュニケーション能力を有している。アノテータは収録した映像データと音声データを参照し、褒める人

*1 人間関係は多様で複雑なため、本研究では初期検討として初対面とした。

*2 本研究では日本電信電話株式会社によって開発された音声認識システムを利用した。

の発話区間ごとに下記の判定・評価を行った。

- 対話相手を褒めているシーンであるか、そうでないかの判定
- 褒めているシーンであると判定した場合、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段階での褒め方の上手さの評価

本研究では、先行研究と同様に各発話区間において褒めていると判定したアノテータが3名以上であるシーンをPraiseシーン(表1)とする。さらに、Praiseシーンにおいて褒めているシーンであると判定したアノテータの評価の平均値をPraiseスコア(平均値と標準偏差: 4.280 ± 0.931)とする。アノテータ間のPraiseスコアの一致率を級内相関係数(ICC)[19]で評価した。具体的には、3~5人のアノテータのそれぞれの組み合わせごとに級内相関係数を算出した。次に、サンプル数を考慮して組み合わせごとの級内相関係数の重み付き平均を算出した。その結果、Praiseスコアの級内相関係数の平均は $ICC(2, k) = 0.701$ であった。この結果から、Praiseスコアはアノテータ間で高い一致率があり、信頼性のあるデータであることが示唆された。

3. モデル構築

3.1 特徴量抽出

本稿では、先行研究[13][14][15]で用いている視覚的、韻律的、言語的特徴量を抽出する。

3.1.1 視覚的特徴量

顔画像処理ツールであるOpenFace[20]を用いて、話者の正面に設置したPCの内蔵カメラで撮影した映像データから、各話者の頭部・顔部の振舞いに関連する特徴量とAction Units*3[21]に関する特徴量の抽出を行った。頭部の向きに関しては、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。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。視線の角度に関しては、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。Praiseシーンの前後1秒ずつを含む範囲における視線の向きのx軸、y軸回りの回転角度(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に関しては、Praiseシーンの前後1秒ずつを含む範囲における各Action Units(表2)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。

3.1.2 韻律的特徴量

音声情報処理ツールであるopenSMILE[22]を用いて、話者の正面に設置したPCの内蔵マイクで録音された音声

*3 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。

表 2 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	上脣を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を硬く閉じる
AU07	脣を緊張させる	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
AU010	上唇を上げる	AU45	瞬きをする
AU012	唇の両端を引き上げる		

表 3 韻律的特徴量の内容

特徴量	内容
pcm_intensity_sma	正規化された強度の値
pcm_loudness_sma	正規化された強度に 0.3 乗した値
mfcc_sma[1]-[12]	1~12 次のメル周波数ケプストラム係数
lspFreq_sma[0]-[7]	8 つの LPC 係数から計算される周波数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数
FEnv_sma	基本周波数のエンベロープ

表 4 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	._linregerrA	線形近似と実際の値の誤差の差
_minPos	最小値の絶対位置	._skewness	歪度
_amean	平均値	._kurtosis	尖度
_stddev	標準偏差	._iqr1-2	四分位範囲:quartile2-quartile1
._quartile1	25 パーセンタイル	._iqr2-3	四分位範囲:quartile3-quartile2
._quartile2	50 パーセンタイル	._iqr1-3	四分位範囲:quartile3-quartile1
._quartile3	75 パーセンタイル		

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。表 3 に示した音声の韻律の代表的な特徴量につき、表 4 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 988 の特徴量を抽出した [23][24]。メル周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。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。

3.1.3 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の BERT モデル [25] を利用し、褒める人の発話を 768 次元のベクトルに変換し、言語的特徴量とした。

3.2 モデル構築

褒め方の上手さを推定するために分類モデルの構築を行った。分類モデルを構築するにあたり、2.2 項で判定した Praise シーンを Praise スコア低群、中群、高群の 3 クラスに分割した。各群に属する Praise シーン数（全 236 シー

ン）ができるだけ均等になるように、下記のように Praise スコア低~高群を定義した*4。

Praise スコア低群 Praise スコアが 3.8 点以下の Praise シーン（計 86 シーン）

Praise スコア中群 Praise スコアが 3.8 点より大きく 4.6 点未満の Praise シーン（計 62 シーン）

Praise スコア高群 Praise スコアが 4.6 点以上の Praise シーン（計 88 シーン）

本稿ではモデルを構築するアルゴリズムとして Random Forest[26] を利用した。具体的には、目的変数を Praise スコア低群、中群、高群の 3 クラスとし、説明変数を 3.1 節で抽出した特徴量とする分類モデルを構築した。決定木の木の深さや木の深さといったハイパーパラメータは Hyperopt[27] を用いて最適化した。加えて、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。本研究では、褒め方の上手さを推定するために寄与するモダリティを明らかにするために、各モダリティの組み合わせを変えて機械学習モデルの構築を行った。各モダリティを組み合わせさせた機械学習モデルについて、次の (1) ~ (2) を 100 回ずつ行った。

- (1) 特徴量が含まれているデータセットを訓練データとテストデータを 9:1 に無作為に分ける。
- (2) 訓練データで構築したモデルを用いてテストデータにおける目的変数を推定する。

3.3 推定性能

3.2 項で構築した機械学習モデルの推定性能（F 値）を表 5 に示す。ベースラインとして、推定結果が Praise スコア低群、中群、高群となる確率がそれぞれ 37%, 26%, 37%（データセットにおける各群の割合）となるモデルを用いた。各モデルの F 値の比較を行うにあたり、正規性の確認のために M1~M32 の F 値に対して Shapiro-Wilk 検定を行った。その結果いくつかのモデルは正規分布に従わなかったため ($p > 0.05$)、ノンパラメトリック検定による比較を行うこととした。具体的には、各モデルの F 値の組に対して有意水準 5% のマン・ホイットニーの U 検定を行い、多重検定の問題を回避するために p 値をホルム法で補正した（p 値の詳細は表 A.1 を参照）。表 5 より、最も性能の良いモデルは、褒める人の視覚的・韻律的特徴量と褒められる人の韻律的特徴量を抽出して用いたモデル（M18）であった。これより、褒め方の上手さの推定をするためには、褒める人の視覚的・韻律的特徴量と褒められる人の韻律的特徴量が重要であると考えられる。M18 と比べて有意に低い F 値を示したモデルは M1, M3, M4, M7, M8, M13, M19 であった。これらのモデルは、褒める人の視覚

*4 各群に属するシーン数は等しいことが理想ではあるが、スコアが同じシーンが多数存在したため、各群のシーン数を等しくすることができなかった。

表 5 機械学習モデルの性能

	褒める人			褒められる人		F 値		褒める人			褒められる人		F 値
	非言語		言語	非言語				非言語		言語	非言語		
	視覚	韻律	言語	視覚	韻律			視覚	韻律	言語	視覚	韻律	
Baseline						0.335	M16	✓	✓	✓			0.556
M1	✓					0.437*	M17	✓	✓		✓		0.532
M2		✓				0.540	M18	✓	✓			✓	0.561
M3			✓			0.431*	M19	✓		✓	✓		0.448*
M4				✓		0.370*	M20	✓		✓		✓	0.555
M5					✓	0.556	M21	✓			✓	✓	0.550
M6	✓	✓				0.559	M22		✓	✓	✓		0.551
M7	✓		✓			0.440*	M23		✓	✓		✓	0.556
M8	✓			✓		0.441*	M24		✓		✓	✓	0.555
M9	✓				✓	0.546	M25			✓	✓	✓	0.553
M10		✓	✓			0.531	M26	✓	✓	✓	✓		0.552
M11			✓			0.558	M27	✓	✓	✓		✓	0.560
M12		✓			✓	0.542	M28	✓	✓		✓	✓	0.541
M13			✓		✓	0.450*	M29	✓		✓	✓	✓	0.551
M14			✓		✓	0.540	M30		✓	✓	✓	✓	0.560
M15				✓	✓	0.548	M31	✓	✓	✓	✓	✓	0.544

F 値に下線が引いてあるモデルは M18 との間に有意差 (p < .05) が認められたモデルである。

的・言語的特徴量、褒められる人の視覚的特徴量を抽出してユニモーダル (M1, M3, M4) あるいはマルチモーダル (M7, M8, M13, M19) で構築している。これより、褒め方の上手さの推定をするためには、褒める人の言語的と褒められる人の視覚的に関する特徴量は重要度が低いと考えられる。

3.4 考察

3.4.1 機械学習モデルの性能について

3.3 節より、ベースラインと M1~M3, M5~31 との間に有意差が認められたことから、構築したモデルが褒め方の上手さを推定可能であると考えられる。最も性能の良いモデル (M18) は褒める人の視覚的・韻律的特徴量と褒められる人の韻律的特徴量を組み合わせて構築したモデルである。M18 と M1 との間に有意差が認められたことより、褒める人の視覚的特徴量はユニモーダルではなく、他の特徴量と組み合わせて用いることが重要であると考えられる。褒め方の上手さの推定は、頭の動きや表情、声の出し方といった非言語行動をユニモーダルではなく、マルチモーダルで用いることが重要であると考えられる。

3.4.2 重要な特徴量について

M18 で重要度の高かった特徴量上位 20 件を図 2 に示す。図 2 より、上位 20 件は韻律的特徴量である praiser_mfcc_sma と praiser_lspFreq_sma であった。これは、遠隔対話において褒め方の上手さを推定するためには褒める人の韻律的特徴量が重要であることを示唆している。praiser_mfcc_sma は人間の聴覚を考慮した声道の音響特性を示している。praiser_lspFreq_sma は声道の周波数特性である LPC 係数から導かれる周波数を示している。これらの特徴は、声色を表すと考えられている。そのため、上手に褒めるには、褒める人の声の強さや高さなどの質が重要

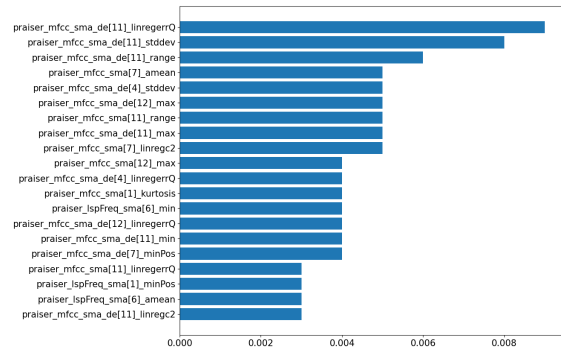


図 2 M18 における重要度の高い特徴量上位 20 件

であると考えられる。今後、これらの特徴量を可視化することで、上手に褒めるためにどのような声色であればいいのか検討したい。

4. おわりに

本稿では、遠隔対話において褒め方の上手さを推定することができるのか明らかにする取り組みを行った。具体的には、褒める人の視覚的、韻律的、言語的特徴量と褒められる人の視覚的、韻律的特徴量を用いて褒め方の上手さを推定する機械学習モデルを構築し、褒め方の上手さの推定に重要な特徴量を明らかにした。その結果、遠隔対話中の褒め方の上手さを推定することが可能であることが明らかになった。褒め方の上手さを推定するためには、褒める人の視覚的・韻律的と褒められる人の韻律的に関する特徴量が重要であると示した。

今後は対面対話と遠隔対話での上手い褒め方の振る舞いを比較していく予定である。さらに、話者の人間関係を考慮した場合の褒め方の上手さの分析も行っていきたい。

* M18 との間で p < .05

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., 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] Okada, S., Ohtake, Y., Nakano, Y., Hayashi, Y., Huang, H., Takase, Y. and Nitta, K.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets, *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 169–176 (2016).
- [5] Rasipuram, S., Rao, P. and Jayagopi, D.: Asynchronous video interviews vs. face-to-face interviews for communication skill measurement: A systematic study, *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 370–377 (2016).
- [6] 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).
- [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*, pp. 1–16 (2021).
- [8] Park, S., Shim, H., Chatterjee, M., Sagae, K. and Morency, L.: Computational Analysis of Persuasiveness in Social Multimedia: A Novel Dataset and Multimodal Prediction Approach, *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp. 50–57 (2014).
- [9] 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).
- [10] Tan, Z., Goel, A., Nguyen, T. and Ong, D.: A Multimodal LSTM for Predicting Listener Empathic Responses Over Time, *14th IEEE International Conference on Automatic Face and Gesture Recognition (FG '19)*, pp. 1–4 (2019).
- [11] Soleymani, M., Stefanov, K., Kang, H., Ondras, J. and Gratch, J.: Multimodal Analysis and Estimation of Intimate Self-Disclosure, *Proceedings of the 21st ACM International Conference on Multimodal Interaction (ICMI '19)*, pp. 59–68 (2019).
- [12] Naim, I., Tanveer, M., Gildea, D. and Hoque, M. E.: Automated prediction and analysis of job interview performance: The role of what you say and how you say it, *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG '15)*, pp. 1–6 (2015).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 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).
- [17] Doherty, S.G. and Anderson, A. and O'malley, C. and Langton, S. and Garrod, S. and Bruce, V.: Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance, *Journal of experimental psychology*, Vol. 3, No. 2, pp. 105–125 (1997).
- [18] Zoom Video Communications Inc: Zoom Cloud Meetings. <https://zoom.us/>.
- [19] Shrout, P. E. and Fleiss, J. L.: Intraclass correlations: uses in assessing rater reliability, *Psychological bulletin*, Vol. 86, No. 2, pp. 420–428 (1979).
- [20] Baltrušaitis, 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).
- [21] Ekman, P. and Friesen, W.: Manual for the facial action coding system (1977).
- [22] 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).
- [23] Schuller, B. and Steidl, S. and Batliner, A.: The interspeech 2009 emotion challenge, *Proceedings of Interspeech '09* (2009).
- [24] 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).
- [25] 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).
- [26] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5–32 (2001).
- [27] Bergstra, J. and Yamins, D. and Cox, 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).

