INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Predicting Collaborative Task Performance using Graph Interlocutor Acoustic Network in Small Group Interaction

Shun-Chang Zhong[1,3], *Bo-Hao Su*[1,3], *Wei Huang*[4], *Yi-Ching Liu*[2], *Chi-Chun Lee*[1,3]

[1]Department of Electrical Engineering, National Tsing Hua University
[2]College of Management, National Taiwan University
[3]MOST Joint Research Center for AI Technology and All Vista Healthcare
[4]Gamania Digital Entertainment Co., Ltd. (HQ)

flank03200@gmail.com, cclee@ee.nthu.edu.tw

Abstract

Recent works have demonstrated that the integration of group-level personality and vocal behaviors can provide enhanced prediction power on task performance for small group interactions. In this work, we propose that the impact of member personality for task performance prediction in groups should be explicitly modeled from both *intra* and *inter*-group perspectives. Specifically, we propose a Graph Interlocutor Acoustic Network (G-IAN) architecture that jointly learns the relationship between vocal behaviors and personality attributes with intra-group attention and inter-group graph convolutional layer. We evaluate our proposed G-IAN on two group interaction databases and achieve 78.4% and 72.2% group performance classification accuracy, which outperforms the baseline model that models vocal behavior only by 14% absolute. Further, our analysis shows that Agreeableness and Conscientiousness demonstrate a clear positive impact in our model that leverages the inter-group personality structure for enhanced task performance prediction.
Index Terms: group interaction, personality, attention mechanism, graph convolutional network

1. Introduction

Small group interaction is a communication unit consists of three to six members exchanging verbal and non-verbal messages in an attempt to influence one another during the decision-making process [1]; this particular type of interaction mechanism provides advantages for human to complete intellectually challenging tasks which often require teamwork to complete. Each group's ability to complete a given cooperative task varies not only with their intellectual knowledge but also with their group-level interactive relationship. Studies of group dynamics suggested a group-level influence between personality and performance may be associated with the match of personality characteristics with group member roles. For example, groups engaged in a cooperative task perform best when they are composed of one relatively dominant member and two or three average- or relatively low-dominance members [2]. During such an interaction, a good performance outcome is commonly considered as the result of the *right* group composition.

Personality plays an important role in affecting the dynamics of the group interaction. It is well-known when assessing each group member's contribution to task performance at an individual level, each member's own personality shows a significant impact. However, it is important to also acknowledge that the role of these traits within the group when considering it as a whole may differ, e.g., a conscientious and extroverted team would be composed of not only conscientious and extroverted members. In fact, the configuration of personality attributes have already been conceptually associated with group processes since the early days of group dynamics research [3, 4]. Combinations of group member personality attributes often form different behavioral dynamics, and it will affect group process and the quality of group performance with either their collaborative talk or opinion conflict. Aside from the well-established literature that intra-group personality composition affects their task performance, understanding how between-group structures are similar for a given task further help in analyzing and understanding the seemingly heterogeneous group interaction behaviors [5].

Recently, computational research has progressed in developing methods that automatically predict group-level task performance from verbal/non-verbal behaviors during small group interactions [6, 7], and some research has started to investigate joint modeling approach in considering the intertwining effect between member's vocal behaviors and *intra-group* personality compositions [8, 9]. Where these past research has laid the solid foundation in predicting group performances using vocal behaviors by jointly modeling the effect of *intra-group* personality composition, these works do not leverage the *inter-group* personality structures into consideration. It should be intuitive that groups with similar group personality compositions should have more correlated performance outcome; the ability to explicitly exploit this inter-group personality structural dependency could lead to a better prediction performances. Thus, in this work, we propose a Graph Acoustic Interlocutor Network (G-IAN) which models not only the intertwining effect between acoustic behaviors and personality for each group, and further represent the inter-group relationship using a group-based personality graph structure that is imposed on the acoustic representations in predicting task performances.

Specifically, our proposed G-IAN architecture predicts group-level performance on two datasets, the NTULP and the Gamania Group Interactive Database (GGID), consists of face-to-face collaborative small group interactions using acoustic features as inputs. The inter-group personality structure is encoded with the inspiration from the successful use of graph convolutional network (GCN) [10] in applications such as social network [10], traffic problem[11], or disease prediction[12]. Our proposed G-IAN considers both intra-group and inter-group effects of personality on vocal behavior jointly: the intra-group personality effect on behavior is modeled by applying personality control attention mechanism and the inter-group personality effect is represented using GCN with the adjacency matrix obtained from group-level personality characterization. Our result shows that G-IAN achieves promising accuracy of

http://dx.doi.org/10.21437/Interspeech.2020-1698

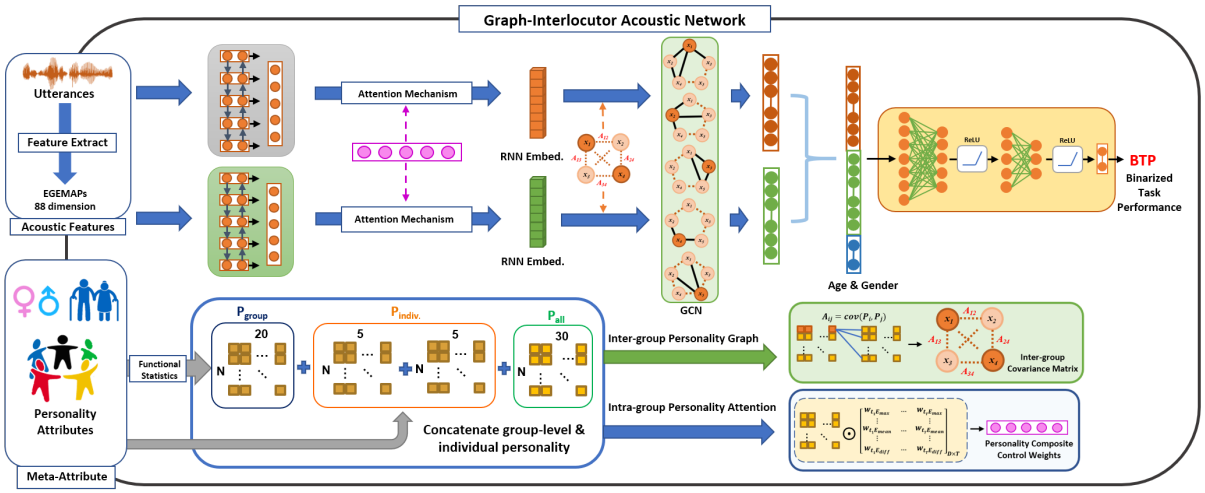

Figure 1: *A complete schematic of our Graph Interlocutor Acoustic Network (G-IAN). It applies modified attention mechanism controlled by group-level personality, and models the inter-group relationship of personality with a graph convolutional layer for the recognition task.*

78.4% and 72.2% unweighted recall rate (UAR) in classifying group performances on the NTULP and the GGID, respectively. Further, our analysis reveals that Agreeableness and Concentiousness constitutes the two *key* factors in linking between group's vocal representations for improved performance prediction accuracy.

2. Research Methodology

2.1. Dasaset

2.1.1. The NTULP Audio-Video Database

The NTULP dataset includes 97 interaction sessions of people engaged in solving a collaborative school policy task [13]. There are three participants in the interaction, and each takes on a different roles randomly: vice president of university, vice president of business school, and a member of the business school teachers committee. The goal of the task is to come up with a solution for a pre-designed issue, and the participants should communicate and discuss the information on their own collaboratively. In addition to audio and video recording of each session, the NTULP contains meta data: age, gender, individual Big5 personality trait and group performance score.

Personality. Participants were administered the five 10-item scales that measure the Big-Five personalities[14]. They were asked to assess how accurately each statement described them on a 5-point scale, with anchors of 1 = very inaccurate and 5 = very accurate.

Group performance. The task performance of each group was evaluated by two trained research assistants according to the scoring manual for the task developed by Wheeler and Menneck (1992) [13]. It consists of over 300 possible solutions with two scores: a problem-solving score for how well the solution solved the problem, and a feasibility score to how feasible a solution is to the problem. Two assistants both coded all the 97 groups privately by finding the best match between the groups solutions and the potential solutions listed in the manual. When scoring disagreements, they would discuss with the third coder, one of the researchers, and the reconciled score were be used in the analysis.

2.1.2. The Gamania Group Interactive Database

The Gamania Group Interactive Database (GGID) is a novel and innovative group interaction corpus proprietary collected by Gamania Digital Entertainment Company in Taiwan. Each session includes a four persons interaction, where the participants jointly engage in a collaborative board game. Participants were asked engage in a puzzle game to rearrange the facilities in four routes in order to reconnect the corresponded entrances to exits. Each group will receive game rewards at the group-level depending on how well they solve this four-route puzzle in time, or receive a game punishment if they fail to solve it.

The GGID includes 31 sessions with 124 subjects totally (age ranges from 21 to 55 years old, 50 males and 74 females). The audios and videos recorded in the database by using one panoramic camera and four separate wireless directional microphones. The database also contains the following meta data: age and gender, individual personality traits, group scores. Table 1 shows the classes distribution of the binraized group performance scores (BGP) provided by the Gamania Group.

Personality. Each participants personality attribute, i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism and Imagination, is measured by IPIP-15 scale with 15 items [15, 16]. The scale is translated and modified from the original IPIP scale with 50 items[16].

Group Performance. The binarized task performance of each team is determined by the task completion distribution of all groups in this board game. The groups completing over 60% of the four-route puzzle game are defined as the high performing group, while those completing below 60% are defined as the low performing group.

2.2. Graph Interlocutor Acoustic Network (G-IAN)

Figure 1 shows our proposed Graph Interlocutor Acoustic Network (G-IAN) framework. We model two participants (since the third participant is also a pre-set examiner) in NTULP database and four participants in the GGID within each session. Specifically, for each of the i-th group, our model uses Bi-GRU with attention mechanism trained on acoustic inputs to form the acoustic embedding, x_m, of the m-th interlocutor in the group,

Table 1: *It shows the classes distribution of the binarized group performance (BGP).*

Dataset	Low(0)	High(1)	Total
NTULP	36	61	97
GGID	18	13	31

where m = 1,..., M (with the number of group members of M).

2.2.1. Speaker Acoustic Features

Firstly, the audio wav files in the NTULP and the GGID are segmented into speaker utterances using an automated voice activity detector (VAD). We extract the sentence-level extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) on each of the target speaker as acoustic features [17] using the openSMILE tookit [18]. It computes 88 dimensional features including statistical properties of mel-frequency cepstral coefficients (MFCCs), associated delta, and prosodic information.

2.2.2. Speaker Personality Features

An intuitive method to model the group-level personality is to compute personality statistics within each group. The group-level personality features are obtained by computing statistics: maximum, minimum, mean and standard deviation value (difference value for NTULP dataset) within the group. Each member has 5 personality scores, and we derive a 20-dimensional features, P_{group}, as group-level personality inputs. Additionally, we retain each member's raw personality attributes as the individual-level personality inputs, P_{indiv}. Then the composite personality, P_{all}, is derived by concatenating P_{group} and P_{indiv}.

2.2.3. Personality Graph

The adjacency matrix, A, is built to represent the inter-group personality relationship between groups. A, a symmetric matrix of size $N \times N$, is defined as following:

$$A_{ij} = \begin{cases} 1, & \text{if } i = j \\ cov(P_i, P_j), & \text{otherwise.} \end{cases}$$

where P_i is the personality inputs (P_{indiv}, P_{group} or P_{all}); $cov(P_i, P_j)$ indicates the covariance value computed between group i and group j on the personality attributes. Intuitively, this personality graph models the inter-group personality relationship, i.e., how similar the personality of the group is to one another. It connects two groups with a larger edge weight when their personality composition are similar.

2.2.4. Graph Interlocutor Acoustic Network

For each session, we first rank and label participants according to their speak times from the most to the least, e.g., we assign the interlocutors as either *talkative* or *talk-less* subject in the NTULP database. We train a Bi-GRU for each subject with personality re-weighted attention mechanism as in our previous work[9], defined as:

$$\alpha_t = \frac{exp[(u^T y_t)]}{\sum_t exp(u^T y_t)} \tag{1}$$

$$\alpha'_t = \alpha_t + ctrl_t \tag{2}$$

$$ctrl_{i \times t} = P_{i \times D} \times W_{D \times t} \tag{3}$$

where α'_t is the personality re-weighted attention weight, α_t is the self attention weight, y_t is the hidden layer of time step t and $ctrl_t$ indicates the personality control weight controlled by the group composite personality inputs mentioned in section 2.2.2. D is the dimension of the personality attributes and W is a trainable matrix. z is the reweighed hidden layer. We consider a 1-layer GCN with the following layer-wise propagation rule:

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)}) \tag{4}$$

Here, A is the adjacency matrix calculated in Section 2.2.3; $W^{(l)}$ is a layer-specific trainable weight matrix; σ is the activation function (we use ReLU here); $H^{(l)}$ is the matrix of activations in the lth layer; Then, we can pass z into the GCN layer.

$$z' = f(z, A) = (AReLU(AzW^{(0)})W^{(1)}) \tag{5}$$

After obtaining z', we then concatenate the meta attributes (age & gender) to z' and feed it into the prediction layer including five fully-connected layers to perform binary classification. All of the parameters of our G-IAN are updated batch-wised by using cross entropy loss function with L2 regularization term to prevent overfitting.

3. Experiment Setup and Results

3.1. Experiment Setup

We evaluate G-IAN in both the NTULP and the GGID using 5-folds cross validation using the metric of unweighted average recall (UAR). In this section we compare different methods, model parameters, and our evaluation scheme.

3.1.1. Model Comparison

Bi-GRU+ATT-Vocal Behavior Only
Training a typical Bi-GRU for each subject with attention to perform recognition directly.
Personality Network (PN)-Vocal Personality Only
Using the PN model in our previous work[9], which uses 5-layer DNN on personality composite features to perform recognition.
Bi-GRU+ATT+CTRL-Vocal Behavior+Intra-Group Effect
Integrating the personality control mechanism to the Bi-GRU attention weight to perform recognition.
G-IAN without CTRL-Vocal Behavior+Inter-Group Effect
Using our proposed architecture without personality control mechanism to perform recognition.
Graph Interlocutor Acoustic Network (G-IAN)
Using our proposed architecture to perform recognition.

3.1.2. Model Parameters

Our proposed G-IAN is trained with same parameters on the NTULP and the GGID. For the NTULP, the number of the hidden nodes in the Bi-GRU is 10, the number of the hidden nodes in the GCN layer is 20. The only difference is the node size of the prediction layer which is composed of 5 fully-connected layers: [44,50,50,32,16,2] for the NTULP and [82,50,50,32,16,2] for the GGID. We use ReLU activation function and batch normalization for the fourth layer. The model is optimized using ADAM optimizer with learning rate equals to 0.0005, batchsize equals to 16 and the λ regularization term equals to 0.007.

3.2. Result and Analysis

3.2.1. Analysis on Model Performance

Table 2 shows the complete prediction results. Our proposed G-IAN obtains the best overall performance at UAR (78.4% on the GGID and 72.2% on the NTULP) group performance classification task. Our method also outperforms the baseline model by 14.3% on GGID and 14.1% UAR on the NTULP absolutely. The baseline model is a standard Bi-GRU architecture with attention mechanism, which only consider subjects' vocal behavior. The PN model uses a 5-layer DNN with group-level personality as input only to perform task performance recognition. The accuracy obtained with these two baseline models (vocal behavior only, personality attributes only) are around 64%.

Table 2: *It shows a comparison of model performance using the metric of unweighted average recall (UAR). The overall result show that the G-IAN outperforms all other methods in group performance classification task achieving 72.2% UAR in NTULP and 78.4% UAR in Gamania.*

		NTULP			Gamania		
Type	Adj	Low UAR	High UAR	UAR	Low UAR	High UAR	UAR
Bi-GRU+ATT (Baseline)		55.6	60.7	58.1	66.7	61.5	64.1
PN	indiv	72.2	50.8	61.5	55.6	61.5	58.5
	group	58.3	68.5	63.6	61.1	61.5	65.3
	all	72.2	55.7	64.0	66.7	61.5	64.1
Bi-GRU+ATT+Control	indiv	55.6	62.3	58.9	66.7	61.5	64.1
	group	61.1	59.0	60.1	72.2	61.5	66.9
	all	63.9	67.2	65.6	66.7	69.2	67.9
G-IAN without Control	indiv	63.9	68.9	66.4	66.7	61.5	64.1
	group	63.9	65.6	64.7	66.7	69.2	67.9
	all	63.9	70.4	67.2	72.2	69.2	70.7
G-IAN	indiv	63.9	68.9	66.4	72.2	61.5	64.1
	group	69.4	62.3	65.9	72.2	76.9	74.6
	all	**72.2**	**72.1**	**72.2**	**72.2**	**84.6**	**78.4**

We find that neither baseline model nor PN has sufficient predictive power of the group performance, and by modeling the intra-group personality effect on vocal behavior using personality control attention mechanism, i.e., Bi-GRU+ATT+Control, it increases slightly to around 66% and 68% on the NTULP and the GGID respectively. Additionally, we compare different types of personality representations (P_{indiv}, P_{group} and P_{all} as mentioned in section 2.2.2 and 2.2.3), the model with the personality feature P_{all} which incorporates P_{indiv} and P_{group} obtains the best prediction effect among these three personality representations.

Generally, the prediction results of the G-IAN that takes into account of the inter-group personality effect improves 2-3% over Bi-GRU+ATT+Control without the graph convolutional structure. By jointly modeling the intra-group and inter-group personality effect on vocal behavior, i.e., our proposed G-IAN with personality control attention mechanism, it achieves the best performing model of 72.2% on the NTULP and 78.8% on the GGID. Our G-IAN also outperforms the results obtained in our previous work [9] (evaluated only on the NTULP) which only models the intra-group personality effect and acoustic behavior. In summary, the results indicate the need of taking into account of both inter-group and intra-group effects of personality, which are shown to be beneficial in this group performance predicting task.

3.2.2. Analysis of Personality Graph

Our experiments demonstrate that modeling intra-group and inter-group personality effects help improve the overall prediction accuracy. We would like to further analyze the differences in the impact of different personality composition on this graph structure. Specifically, we quantify the graph structure using *clustering analysis* and *connectivity analysis*. We would like to investigate which personality trait has the greatest impact on the graph structure; we will perform the following graph analysis by removing each of the five personality attribute to examine the changes in the structure.

Table 3: *Clustering analysis and connectivity analysis of the personality graph with different personality composition. The nodes includes the sessions in Fold 3 in NTULP (78 nodes) and Fold 2 in the GGID (24 nodes).*

	GGID			NTULP		
Removing Attribute	$\overline{C}$	C_N	C_E	$\overline{C}$	C_N	C_E
None	0.139	17	18	0.465	78	78
Extraversion	0.124	**16**	18	0.461	78	78
Agreeableness	**0.143**	**16**	**17**	**0.473**	**77**	**77**
Conscientiousness	**0.147**	**15**	**17**	**0.482**	**77**	**76**
Neuroticism	**0.144**	17	**17**	0.447	78	78
Openness/Imagination	0.136	17	19	0.391	78	78

We evaluate the clustering level of the graphic structure by calculating the *average clustering coefficient*, $\overline{C}$, defined as follows:

$$C_i = \frac{2|e_{jk}|}{k_i(k_i-1)} \tag{6}$$

$$\overline{C} = \frac{1}{N_{nodes}} \sum_v C_v \tag{7}$$

where k_i is the number of neighbours of a vertex, $v_j, v_k \in G, e_{jk} \in E$. C_i is the fraction of pairs of nodes, that are neighbors of a given node v, that are connected to each other by edges, and $\overline{C}$ is the average of all local coefficients. The higher the clustering coefficient, the denser the graph. For connectivity analysis, we calculate the node connectivity (C_N) and edge connectivity (C_E), which is the minimum number of nodes/edges need to be removed in order to split the network.

Table 3 shows the changes in our personality graph of A_{all} in the training set of the GGID and the NTULP as it removes each personality attribute. We observe that Agreeableness and Conscientiousness are the two attributes that make the graph more clustered and reduce the connectivity. In contrast, the distance between groups will become further apart if these two attributes are missing which causes the graph to be more divergent and makes it more difficult for our model leverage similar patterns between groups.

4. Conclusions

In small group collaborative task, the ability to automatically predict task performance from vocal cues is not only related to intra-group effect of personality composition but also the inter-group relationship of personality composition. In this work, we proposed a Graph Interlocutor Acoustic Network that not only integrates the intra-group effect of personality attributes into acoustic behaviors as attention mechanism, but also models the inter-group personality relationship as a graphical structure with GCN. We obtain a competitive prediction accuracy of group performance on the NTULP (72.2%) and the GGID (78.4%) datasets. In summary, the combination of time-series model and graph-based deep learning network provides a novel approach in studying the personality effect and the speech dynamics within group. We will continue to investigate the effect of the personality traits and advance our technical framework so that it can adapt to the more complex conversational environment.

5. References

[1] S. Tubbs, *A systems approach to small group interaction*. McGraw-Hill, 1995. [Online]. Available: https://books.google.com.tw/books?id=PCjTVXIIPI0C

[2] E. E. Ghiselli and T. M. Lodahl, "Patterns of managerial traits and group effectiveness." *The Journal of Abnormal and Social Psychology*, vol. 57, no. 1, p. 61, 1958.

[3] W. Haythorn, "The influence of individual members on the characteristics of small groups." *The Journal of Abnormal and Social Psychology*, vol. 48, no. 2, p. 276, 1953.

[4] B. Barry and G. L. Stewart, "Composition, process, and performance in self-managed groups: The role of personality." *Journal of Applied psychology*, vol. 82, no. 1, p. 62, 1997.

[5] M. Kompan and M. Bieliková, "Social structure and personality enhanced group recommendation." in *UMAP Workshops*, 2014.

[6] G. Murray and C. Oertel, "Predicting group performance in task-based interaction," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 14–20.

[7] U. Kubasova, G. Murray, and M. Braley, "Analyzing verbal and nonverbal features for predicting group performance," *arXiv preprint arXiv:1907.01369*, 2019.

[8] Y.-S. Lin and C.-C. Lee, "Using interlocutor-modulated attention blstm to predict personality traits in small group interaction," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018, pp. 163–169.

[9] S.-C. Zhong, Y.-S. Lin, C.-M. Chang, Y.-C. Liu, and C.-C. Lee, "Predicting group performances using a personality composite-network architecture during collaborative task," *Proc. Interspeech 2019*, pp. 1676–1680, 2019.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[11] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu, "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3656–3663.

[12] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimers disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.

[13] B. C. Wheeler and B. E. Mennecke, "The school of business policy task manual: Working paper# 92-524c," 1992.

[14] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.

[15] L. Zheng, L. R. Goldberg, Y. Zheng, Y. Zhao, Y. Tang, and L. Liu, "Reliability and concurrent validation of the ipip big-five factor markers in china: Consistencies in factor structure between internet-obtained heterosexual and homosexual samples," *Personality and individual differences*, vol. 45, no. 7, pp. 649–654, 2008.

[16] R.-H. Li and Y.-C. Chen, "The development of a shortened version of ipip big five personality scale and the testing of its measurement invariance between middle-aged and older people," *Journal of Educational Research and Development*, vol. 12, no. 4, pp. 87–119, 2016.

[17] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[18] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.