INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Attentive to Individual: A Multimodal Emotion Recognition Network with Personalized Attention Profile

Jeng-Lin Li[1,2], *Chi-Chun Lee*[1,2]

[1]Department of Electrical Engineering, National Tsing Hua University, Taiwan
[2]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

cllee@gapp.nthu.edu.tw, cclee@ee.nthu.edu.tw

Abstract

A growing number of human-centered applications benefit from continuous advancements in the emotion recognition technology. Many emotion recognition algorithms have been designed to model multimodal behavior cues to achieve high performances. However, most of them do not consider the modulating factors of an individual's personal attributes in his/her expressive behaviors. In this work, we propose a Personalized Attributes-Aware Attention Network (PAaAN) with a novel personalized attention mechanism to perform emotion recognition using speech and language cues. The attention profile is learned from embeddings of an individual's profile, acoustic, and lexical behavior data. The profile embedding is derived using linguistics inquiry word count computed between the target speaker and a large set of movie scripts. Our method achieves the state-of-the-art 70.3% unweighted accuracy in a four class emotion recognition task on the IEMOCAP. Further analysis reveals that affect-related semantic categories are emphasized differently for each speaker in the corpus showing the effectiveness of our attention mechanism for personalization.

Index Terms: personal attribute, multimodal emotion recognition, attention, psycholinguistic norm

1. Introduction

The surge of artificial intelligence has brought major advancements for a wide range of human-centered applications, e.g., human robotics, personal assistance, and recommendation system [1, 2, 3]. These technologies rely strongly on the understanding of human with supportive feedback in order to provide higher quality personalized experiences. Technology involves affect sensing, hence, has become crucial as emotion plays a critical role in human's decision-making, cognitive process, and social communication [4, 5]. Numerous algorithms have already been proposed that learn to recognize emotion from multimodal humans expressive behavior data. Several recent works have further shown a significant improvement in utilizing deep learning network architecture for modeling speech and language modalities simultaneously for the task of emotion recognition.

Specifically, Poria et al. propose to use multiple kernel learning with deep convolution neural networks (CNN) to model speech acoustic and text data for emotion recognition [6]. Beard et al. present a recursive attention fusion network to model the complementary effect between acoustic and lexical data [7]. Most recently, Cho et al. propose to use a multi-resolution CNN for lexical transcript and long short term memory network (LSTM) for acoustic features and finally combine the both using a deep forward neural network [8]. While many of these algorithms achieve promising recognition accuracies, none of these multimodal emotion recognition algorithms have considered the *personalized* differences at an *individual*-level.

Individual differences in the emotional expressions result from a wide of personalized factors. These factors include categorization attributes and also individualized attributes. Categorization attributes, such as culture, age, and gender, are well known meta-factors in affecting the emotion expressions [9, 10, 11]. Individualized attributes, such as personality, motives, and beliefs, are formed through a long term self-understanding and socialization process that affects the emotional experiences and expressions at an *individual*-level [12, 13, 14]. Other meta-factors like social status and social role have also been indicated to shape the emotion regulation strategy for an individual when experiencing similar circumstances [15, 16].

While *personalized* emotion recognition system is critical, only a handful of research has integrated modeling of personal factors into the recognition algorithms. For example, Stefanie et al. and Igor et al. develop attribute-specific emotion recognition by simply training on hard-segmented subgroups of data [17, 18], Sicheng et al. use hypergraph structure to formulate the relationship of physiological signal and personality for emotion recognition [19], and more recently, Sagha et al. incorporate personality trait, age, and gender as a model selection strategy in task of valence recognition [20]. Developing advanced algorithms that integrate complex modulation of multimodal affective expressions by an individual's personalized factors is a crucial next-step to enhance the model capacity.

In this work, we propose a Personalized Attribute-Aware Attention Network (PAaAN) to perform emotion classification using speech and text modality. The key idea centers on the fact that each subject would possess different personalized attributes that could be integrated jointly to the recognition network. Specifically, the PAaAN integrates a re-weighting mechanism of *personalized* attention profile that is jointly optimized with multimodal attention bi-directional LSTM (BLSTM). The *personalized* attention profile is learned jointly from embeddings of target subject's personal profile with each utterance's acoustics and lexical embeddings. This unique personal profile embedding is computed as a dot product between the psycholinguistic norm vector of each target speaker and a large speaker set of movie script corpus; this can be conceptualized as a high-dimensional representation of a speaker's personal attribute space. This approach further mitigates the common requirements of obtaining a pre-defined set of personal attribute classes such as gender, age, personality trait, etc.

We evaluate our framework on a benchmark emotional corpus, the IEMOCAP database [21]. Our PAaAN achieves the state-of-the-art 70.3% unweighted average recall (UAR) in a four class emotion recognition task, which is a 5.55% relative improvement over multimodal BLSTM network without personalized attention profile. Furthermore, our analysis on PAaAN reveals an *individualized* reweighting effect on affect related word acoustically and textually.

http://dx.doi.org/10.21437/Interspeech.2019-2044

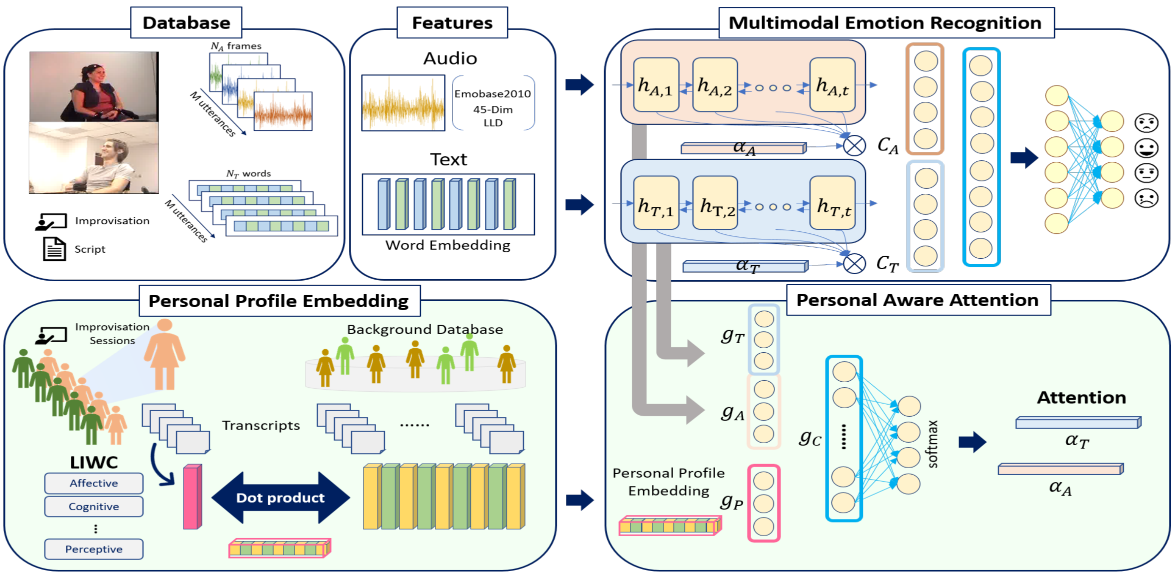

Figure 1: *This is the overall PAaAN framework. We compute dot product of each target speaker's LIWC features with a large speaker set of movie scripts to project the target speaker into a personal profile space. Then, we introduce the personalized attention profile to jointly model inter-modality relation and personalized attributes in a multimodal framework for emotion recognition.*

2. Methodology

2.1. Databases

2.1.1. The IEMOCAP Emotion Database

In this work, we use a multimodal dyadic interaction corpus, the IEMOCAP [21], as our main emotion recognition evaluation database. It comprises of 12 hours of audio-video recordings with word alignment and manual transcripts available. There are a total of 10 different speakers paired in dyads performing either scripted or improvised hypothetical scenarios. There are a total of 10039 utterances with each rated by at least three annotators. We use a subset of the database to be comparable to recent multimodal emotion recognition works [8]. It includes a total of 5531 samples from four emotion classes (angry: 1103, happiness: 1636, sad: 1084, and neutral: 1708).

2.1.2. The Background Movie Script Databases

Our PAaAN integrates personalized profile into the recognition framework. The IEMOCAP database only includes 10 speakers, in order to robustly extract representation of personal profile for each speaker, our framework leverages other large speaker set corpora with text transcripts. In this work, we gather two additional background movie script databases each with a large speaker set. One of them is the Cornell movie dialogs corpus [22], which includes movie scripts from 5531 distinct acting roles. The other one is the EmotionLines corpus [23] including scripts from 656 speakers derived from the TV shows Friends and private Facebook messenger texts. Both corpora contain dialogue contents that are similar to the IEMOCAP.

2.2. Acoustic and Textual Representations

2.2.1. Acoustic Features

We extract 45 dimensional low level descriptors (LLDs) including 12 dimensional Mel-Frequency Cepstral Coefficients (MFCCs), fundamental frequency (F0), loudness, voice probability, zero cross rate along with the first and the second derivatives of MFCCs and loudness using the openSMILE toolbox [24]. These LLDs are extracted using 60ms frame size and 10ms step size; speaker-wise zscore normalization is applied. Each time step in our BLSTM corresponds to 4 frames (40ms), and the input to each time step is the average of these LLDs.

2.2.2. Word Embeddings

Texts in the transcripts used in this work are encoded using the GloVe word2vec pretrained model that is originally trained on 42 billion tokens and 1.9 million vocabularies [25]. An utterance of N words is represented as a set of word embedding vectors, $U = \{w_1, w_2, ..., w_N\}$, where w is the word embedding and $U \in \mathbb{R}^{N\times 300}$. Each word is encoded as a 300-dimensional vector at every time step for the BLSTM model.

2.3. Personalized Attribute-Aware Attention Network

Our complete PAaAN architecture is illustrated in Figure 1. It includes a dual modality BLSTM with Deep Neural Network (DNN) as emotion recognition network using both acoustic and textual features. Each of the BLSTM is learned together with a novel personalized profile attention mechanism. We will first describe the derivation of personal profile embedding that is needed for computing personalized attention for BLSTM recognition networks.

2.3.1. Personal Profile Embedding

An individual speaker's personal factors would shape the particular speaker's speech and language behaviors. In this work, instead of coming up with pre-defined personal attributes, such as personality, age, gender, social status, etc, we decide to encode personal attributes to a high-dimensional space using the semantic lexicons (the psycholinguistic norm). Specifically, we use the 'Linguistic Inquiry Word Count' (LIWC-2007) tool [26], which includes 64 semantic word categories such as standard linguistic words, words related to psychological processes, personal concerns, spoken patterns, etc. The LIWC have been known to be related to a variety of one's personal attributes such as age, gender, and personality traits [27, 28].

Table 1: *It summarized the 4-class recognition results in the emotion recognition task*

	Audio	Text	A+T	Prev1	Prev2	I_{A+T}	P-D	P-T	$PAaAN_{WE}$	$PAaAN_{LIWC}$	$PAaAN_{LC}$
ang	0.570	0.663	0.657	0.724	0.625	0.666	0.721	0.666	0.697	0.727	**0.769**
hap	0.527	0.647	0.666	0.675	0.652	0.685	0.718	0.757	0.701	0.721	0.719
neu	0.581	0.584	0.628	0.574	0.696	0.607	0.601	0.559	0.611	0.588	0.591
sad	0.608	0.551	0.666	0.665	0.633	0.705	0.666	0.675	0.714	0.690	**0.732**
UAR	0.571	0.611	0.654	0.659	0.651	**0.666**	0.676	0.665	0.681	0.682	**0.703**

To compute a personal profile embedding vector, g_p, we first perform LIWC counts on the improvised portion of each speaker in the IEMOCAP database to derive a 64-dimensional psycholinguistic norm vector per speaker. Each speaker's personal factor representation is then derived by computing dot product between this target speaker's psycholinguistic norm vector to the 5987 background speaker's psycholinguistic norm vector. This 5987 dimensional vector is termed as the personal profile embedding, which can be conceptualized as a point in a high-dimensional personal attribute space characterized by the psycholinguistic word usage.

2.3.2. Multimodal Emotion Recognition Network

Our recognition network is a dual modality BLSTM networks with personalized profile attention mechanism feeding into DNN classification layers. Specifically, for a BLSTM at time step, t, of modality m with hidden state $h_{m,t} = [\overleftarrow{h_{m,t}} \bigoplus \overrightarrow{h_{m,t}}]$ where $\overleftarrow{h_{m,t}}$ and $\overrightarrow{h_{m,t}}$ denotes the forward and backward hidden states, we first summarize the $h_{m,t}$ using a fully connected layer with hyperbolic tangent activation:

$$g_{m,t} = tanh(w_m^T h_{m,t} + b_m) \tag{1}$$

We can therefore obtain the time-normalized personalized profile attention weight at time t, α_t, by applying softmax function on the following concatenated latent vector $g_{c,t}$:

$$g_{c,t} = [g_{T,t}, g_{A,t}, g_{P,t}] \tag{2}$$

$$\alpha_t = \frac{\exp(g_{c,t})}{\sum_t^T \exp(g_{c,t})} \tag{3}$$

where $m = \{T, A\}$ denotes text and audio modality. The audio and text attention weights are learned separately using each modality-specific BLSTM resulting in two attention weights α_A and α_T; however, noted that in deriving $g_{c,t}$ attention weights for each modality, the inter-modality information is utilized (equation 2). The two modality-specific attention weights provide additional flexibility in our emotion modeling. After multiplying these learned personalized attentions back to the BLSTMs, we can obtain the re-weighted context vector of C_A and C_T. The concatenation of C_A and C_T serves as the representation to the DNN layers for emotion classification.

In summary, our proposed PAaAN learns two attention-based BLSTMs feeding into DNN layers for emotion classification task. The time-varying and modality-specific attention weights are learned by jointly considering the interaction of inter- behavior modalities (speech and language) with personal profile embedding mentioned in section 2.3.1.

3. Experimental Setup and Results

3.1. Experimental Setup

In this work, we evaluate the performances on a 4-class emotion recognition task, and the comparison models are listed:

- **Audio**: Using single-modal of audio features only
- **Text**: Using single-modal of word embeddings only
- **A+T**: Concatenating C_A and C_T without inter-modality nor personal attention as a multimodal baseline
- **Prev1**: A recent multimodal audio and text framework proposed by Cho et al. [8]
- **Prev2**: Another multimodal audio and text framework proposed by Poria et al. [6]
- $\mathbf{I_{A+T}}$: Using inter-modality attention mechanism without personal attention profile (i.e., $g_{c,t}$ in equation 2 is derived without $g_{P,t}$)
- **P-D**: Concatenating personal profile embedding directly with inter-modality attention based BLSTMs context vector, C_A and C_T
- **P-T**: Concatenating personal profile embedding with audio and word embedding in each BLSTM time step with inter-modality attention
- $\mathbf{PAaAN_X}$: Using X-personal profile embedding generating approach within the PAaAN network

For I_{A+T}, we use the same architecture as PAaAN without personal profile embedding as a multimodal baseline to examine the effectiveness of personalize attention profile. *P-D* and *P-T* are different strategies to integrate personal profile embedding in the network. $PAaAN_X$ is our proposed framework where X indicates other potential personal profile generation approaches, i.e., *WE* is simply the average word embedding computed for each target speaker's improvisation transcripts, *LIWC* indicates the use of LIWC word count vector of the target speaker directly, and *LC* indicates our proposed profile embedding detailed in section 2.3.1.

The BLSTM for audio and text has the same structure, which includes 128 nodes, and $g_{m,t}$ is of dimension 64. Each modality-specific BLSTM's context vector is first fed into a 128 dimensional fully-connected layer then to a 256-node layer after concatenating the outputs. The final four class recognition is done using a softmax layer. We use tanh as activation function for $g_{m,t}$ whereas relu for all other layers. The model is trained with 50 epochs using 32 as batch size and 0.0001 as learning rate. The experiments are carried out using leave dyad out cross validation and the unweighted average recall (UAR) is used as the evaluation metric.

3.2. Experimental Results

Table 1 summarizes the emotion classification results. Our proposed $PAaAN_{LC}$ achieves the best 70.3% UAR, which improves 5.55% relative compared to the best performing multimodal framework without personal profile embedding (I_{A+T}).

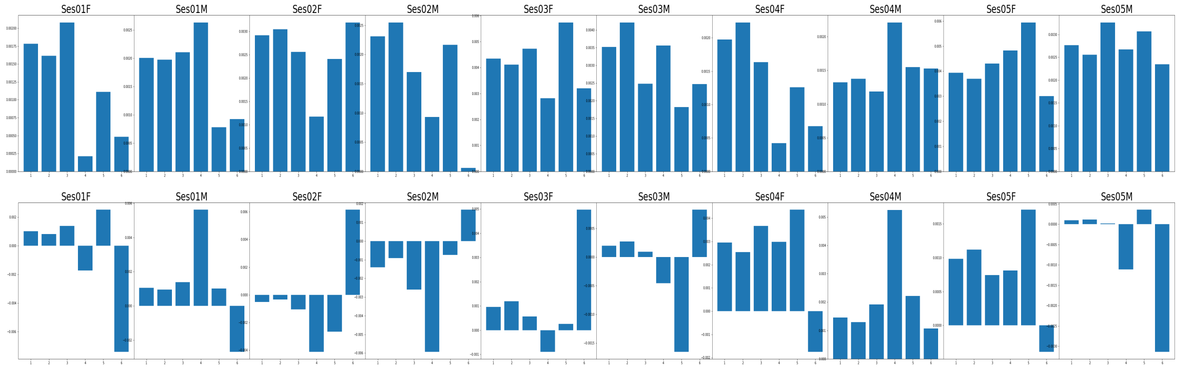

Figure 2: *It shows the individual difference between accumulated attention values from PAaAN$_{LC}$ and A+T on the six affect-related word class usage for each speaker in the IEMOCAP. The upper side distributions are from α_A and the lower ones are from α_T.*

This suggests the importance in modeling personal attributes. In our multimodal scheme, I_{A+T} achieves an 66.6% UAR outperforming the other recent multimodal approaches on the same database [8, 6], which indicates the better modeling power in the use of our inter-modality attention mechanism. Also, I_{A+T} demonstrates an improvement over single modality baseline, *Audio* and *Text*, with 16.64% and 9% relative improvement.

We additionally compare different methods in integrating the personal profile embedding into the recognition networks. We report UARs obtained by using the personal vector detailed in section 2.3.1 with different integration strategies. The results of *P-D* and *P-T* are 67.6% and 66.5%, which are both 3.99% and 5.71% relative worse than our proposed *PAaAN$_{LC}$*. The method of *P-D* treats personal attributes as an additional static features to be integrated, whereas the method of *P-T* directly replicates the personal profile vector at every time step of BLSTM. *P-T* performs the worst likely due to the fact that it introduces too much redundancy into the network. *PAaAN$_{LC}$* outperforms *P-D* not only demonstrates the superior performances of attention-based modeling, but also corroborates with past knowledge that personal factors do interact with the temporal dynamics of an individual speaker's expressive behaviors.

Within our proposed PAaAN framework, we examine different approaches in generating personal profile embedding, g_p. This can be thought as a task in seeking a proper feature space to represent the underlying personal attributes. We observe that *PAaAN$_{WE}$* and *PAaAN$_{LIWC}$* are comparable in overall accuracy but behave slightly different between the 4 emotion classes. The use of LIWC would be more effective for non-neutral emotion, which is likely due to the composition of the existing LIWC categories covering mostly affective, cognitive, or perceptive word semantic class. Furthermore, we observe a significant improvement in angry and sad classes when comparing our *PAaAN$_{LC}$* versus *PAaAN$_{LIWC}$*. Our proposed used of projected embedding *PAaAN$_{LC}$* represents personal attributes in a high dimensional space by computing the similarity to a large set of diverse movie acting roles. The space constructed by these movie scripted roles includes a larger personal factor variability and is likely to provide a more robust space representation with more emotional expressivity.

3.3. Analyses

In this section, we examine the change of modality-specific attention weights between *PAaAN$_{LC}$* and I_{A+T} on the six affect-related word class usage of each speaker (Table 2). The category-wise accumulated attention difference values, i.e., the values from *PAaAN$_{LC}$* minus values from *A+T* are demonstrated

Table 2: *The six affect-related word class displayed in Figure 2*

1	2	3
Affective Processes	Positive Emotions	Negative Emotions
4	5	6
Anxiety	Anger	Sadness

in Figure 2. We align the acoustic frames with the words in an utterance and sum up the frame-level attention values in α_A for each spoken word as accumulated acoustic attention. Similarly, accumulated textual attention on word-level can be derived. For each selected affect-related categories, we then aggregate the derived accumulated attention values to compute the difference before and after we integrate the personal profile embedding.

Figure 2 shows the modality-specific attention difference in these word categories for each subject in the IEMOCAP database. Generally, these affect-related word categories in α_A increase for *PAaAN$_{LC}$* while some of them from α_T descends after considering personal attributes. We observe a generally distinct pattern for each speaker indicating that there may be a different 'personalized speech and language baseline' for each subject individually; for example, some people might just generally like to use emotional words in their daily conversation, the attention weight after considering personal profile embedding should counter-emphasize those instances. A more detailed analysis at the individual-level will still be required.

4. Conclusions

The individual differences between people are known to affect our multimodal expressive behaviors; however, there is a lack of appropriate framework that models this intertwining effect for emotion classification. In this work, we propose *PAaAN$_{LC}$* that learns a personalized attention by integrating a speaker-level attribute space computed using psycholinguistic norm. Our *PAaAN$_{LC}$* outperforms the state-of-the-art multimodal, i.e., speech and language, emotion recognition on the benchmark emotion corpus. The introduction of our personal profile vector can also easily be deployed to many databases where there is no explicit annotations on speaker-wise personal attributes.

We will continue to evaluate the robustness of our *PAaAN* framework on other corpora and also investigate these data-driven personal profile vector's relation to known personal attributes, e.g., personality. We hope to further understand the contributing factors of personal attributes in modulating our emotionally expressive behaviors to develop personalized emotion sensing technology across human-centered applications.

5. References

[1] M. Tkalcic, B. D. Carolis, M. de Gemmis, A. Odic, and A. Kosir, Eds., *Emotions and Personality in Personalized Services - Models, Evaluation and Applications*, ser. Human-Computer Interaction Series. Springer, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-31413-6

[2] P. Beckerle, G. Salvietti, R. Unal, D. Prattichizzo, S. Rossi, C. Castellini, S. Hirche, S. Endo, H. B. Amor, M. Ciocarlie *et al.*, "A human–robot interaction perspective on assistive and rehabilitation robotics," *Frontiers in neurorobotics*, vol. 11, p. 24, 2017.

[3] T. Schnabel, P. N. Bennett, and T. Joachims, "Improving recommender systems beyond the algorithm," *CoRR*, vol. abs/1802.07578, 2018.

[4] R. W. Picard, *Affective computing*. MIT press, 2000.

[5] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, 2013.

[6] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional mkl based multimodal emotion recognition and sentiment analysis," in *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 2016, pp. 439–448.

[7] R. Beard, R. Das, R. W. Ng, P. K. Gopalakrishnan, L. Eerens, P. Swietojanski, and O. Miksik, "Multi-modal sequence fusion via recursive attention for emotion recognition," in *Proceedings of the 22nd Conference on Computational Natural Language Learning*, 2018, pp. 251–259.

[8] J. Cho, R. Pappagari, P. Kulkarni, J. Villalba, Y. Carmiel, and N. Dehak, "Deep neural networks for emotion recognition combining audio and transcripts," *Proc. Interspeech 2018*, pp. 247–251, 2018.

[9] A. M. Kring and A. H. Gordon, "Sex differences in emotion: expression, experience, and physiology." *Journal of personality and social psychology*, vol. 74, no. 3, p. 686, 1998.

[10] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[11] D. Matsumoto, "Are cultural differences in emotion regulation mediated by personality traits?" *Journal of Cross-Cultural Psychology*, vol. 37, no. 4, pp. 421–437, 2006.

[12] L. Parks-Leduc and R. Guay, "Pesonality, values, and motivation," *Personality and Individual Differences*, vol. 47, pp. 675–684, 11 2009.

[13] M. Emotion, "The development of the person: Social understanding, relationships, self, conscience," 2006.

[14] S. E. Hampson, "Personality processes: Mechanisms by which personality traits get outside the skin," *Annual review of psychology*, vol. 63, pp. 315–339, 2012.

[15] U. Hess, R. Adams Jr, and R. Kleck, "Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger," *Cognition & Emotion*, vol. 19, no. 4, pp. 515–536, 2005.

[16] A. S. Morris, J. S. Silk, L. Steinberg, S. S. Myers, and L. R. Robinson, "The role of the family context in the development of emotion regulation," *Social development*, vol. 16, no. 2, pp. 361–388, 2007.

[17] S. Rukavina, S. Gruss, H. Hoffmann, J.-W. Tan, S. Walter, and H. C. Traue, "Affective computing and the impact of gender and age," *PloS one*, vol. 11, no. 3, p. e0150584, 2016.

[18] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese, and A. Sciarrone, "Gender-driven emotion recognition through speech signals for ambient intelligence applications," *IEEE transactions on Emerging topics in computing*, vol. 1, no. 2, pp. 244–257, 2013.

[19] S. Zhao, G. Ding, J. Han, and Y. Gao, "Personality-aware personalized emotion recognition from physiological signals." in *IJCAI*, 2018, pp. 1660–1667.

[20] H. Sagha, J. Deng, and B. Schuller, "The effect of personality trait, age, and gender on the performance of automatic speech valence recognition," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 86–91.

[21] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, no. 4, p. 335, 2008.

[22] C. Danescu-Niculescu-Mizil and L. Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs," in *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, 2011, pp. 76–87.

[23] C.-C. Hsu, S.-Y. Chen, C.-C. Kuo, T.-H. Huang, and L.-W. Ku, "Emotionlines: An emotion corpus of multi-party conversations," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. European Language Resource Association, 2018. [Online]. Available: http://aclweb.org/anthology/L18-1252

[24] 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*. ACM, 2010, pp. 1459–1462.

[25] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162

[26] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," Tech. Rep., 2015.

[27] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.

[28] V. Zeigler-Hill and T. Shackelford, *The SAGE Handbook of Personality and Individual Differences: Volume I: The Science of Personality and Individual Differences*, 01 2018.