Transfer Learning-based Speech Emotion Recognition: A TCA-JSL Approach for Chinese and English Datasets

Sulian Sun,[1*] Libo Wu[1]
[1]School of Foreign Languages, Hubei College of Science and Technology, Xianning, Hubei 437100, China
E-mail: sslsunsl@hotmail.com
*Corresponding author

Keywords: speech emotion recognition, transfer learning, transfer component analysis, Chinese and English speech emotion library

Received: November 21, 2024

Speech emotion recognition (SER) has important application value in many scenarios and is a focus of current research. This paper developed a transfer component analysis-joint subspace learning (TCA-JSL) algorithm based on transfer learning for Chinese and English SER. It used TCA to reduce the dimension of speech emotion features, employed the JSL algorithm to generate categorical emotion features, and realized the recognition of different emotion types based on a support vector machine. An experimental analysis was carried out on Chinese speech emotion library CASIA, English speech emotion library eNTERFACE, and English speech emotion library SAVEE. The results showed that the p value of the TCA-JSL algorithm for CASIA→eNTERFACE was 0.4950 ± 0.0152, and the R value was 0.3542 ± 0.0163; the P value for eNTERFACE→CASIA was 0.4533 ± 0.0151, and the R value was 0.3511 ± 0.0161. Compared with the joint distribution adaptive regression (JDAR) method, $p < 0.05$. The P value of the TCA-JSL algorithm for CASIA→SAVEE was 0.4521 ± 0.0176, and the R value was 0.3544 ± 0.0161; the P value for SAVEE→CASIA was 0.4987 ± 0.0175, and the R value was 0.3511 ± 0.0158. Compared with the JDAR method, $p < 0.05$. The results verify that the TCA-JSL algorithm is effective in Chinese and English SER and can be applied to real tasks.

Povzetek: Predstavljen je pristop TCA-JSL za prepoznavanje čustev v govoru, ki uporablja analizo prenosnih komponent in učenje skupnega podprostora za prepoznavanje čustev v kitajskih in angleških naborih podatkov,

1 Introduction

With the continuous development of artificial intelligence, the application of machines in daily life and work is more and more extensive [1], and the research on human-computer interaction is deeper [2]. Speech is an important medium for communication between man and machine. Human speech contains a large amount of emotional information [3], and different emotions can convey different information. Machines can accept instructions and feedback through the recognition of speech. The understanding of speech also keeps advancing with the progress of technology. Speech emotion recognition (SER) is the process of recognizing the emotions in speech [4]. As a branch of speech recognition, SER has gradually become an important research content in human-computer interaction [5] and has wide application prospects in driving assistance [6], online education [7], psychological diagnosis [8], and other aspects. Nasersharif et al. [9] proposed a deep domain adaptive approach for general and variational autoencoders to achieve cross-corpus SER. They found that the approach can improve the accuracy of recognition. Kumar et al. [10] analyzed the performance of spectral contrast features in SER, conducted experiments on RAVDESS, SAVEE, and other datasets, and found that the spectral contrast feature had good performance on SER. Atmaja et al. [11] studied the effects of different loss functions on SER tasks and found that the loss function based on correlation and the loss function based on consistency correlation coefficient had better results. Xie et al. [12] designed a SER method using multi-head attention and found through experiments on the IEMOCAP dataset that this method can enhance the performance of SER. Different countries and nationalities do not speak the same language, but they all have the same emotions. SER of different languages can help learn the emotional commonality better, thus improving the recognition effect. Therefore, this paper used transfer learning to study Chinese and English SER and developed a transfer component analysis-joint subspace learning (TCA-JSL) algorithm, a method combining transfer learning with joint subspace learning, to further realize the processing of speech emotion features to improve the recognition performance. The effectiveness of this method in completing Chinese and English SER tasks was verified through experiments on datasets, providing a new and useful method for the current SER research.

2 Related works

Transfer learning can transfer the knowledge learned in one domain to another domain to obtain better learning effects, and relevant methods have also been increasingly applied in SER, such as TCA. Long et al. [13] proposed a

new transfer learning method called joint distribution adaptation (JDA). Zhang et al. [14] developed a joint distribution adaptative regression (JDAR) method to handle cross-corpus SER. This paper developed a TCA-JSL method by combining transfer learning with joint subspace learning. The P and R values of these methods in cross-corpus speech emotion recognition are shown in Tables 1 and 2. It can be found that the TCA-JSL method has more advantages compared with other current methods.

Table 1: Comparison of P values for different datasets.

	CASIA →eNTERFACE	eNTERFACE→CASIA	CASIA →SAVEE	SAVEE →CASIA
TCA	0.2864±0.0124	0.2512±0.0126	0.3321±0.0156	0.3551±0.0155
JDA	0.3132±0.0135	0.2884±0.0137	0.3676±0.0161	0.3972±0.0158
JDAR	0.4774±0.0141	0.4211±0.0142	0.4308±0.0172	0.4725±0.0173
TCA-JSL	0.4950±0.0152*	0.4533±0.0151*	0.4521±0.0176*	0.4987±0.0175*

Table 2: Comparison of R values for different datasets.

	CASIA →eNTERFACE	eNTERFACE→CASIA	CASIA →SAVEE	SAVEE→CASIA
TCA	0.2602±0.0133	0.2641±0.0131	0.2776±0.0137	0.2784±0.0138
JDA	0.2659±0.0137	0.2701±0.0135	0.2945±0.0142	0.2864±0.0139
JDAR	0.2843±0.0142	0.3032±0.0143	0.3318±0.0155	0.3346±0.0156
TCA-JSL	0.3542±0.0163*	0.3511±0.0161*	0.3544±0.0161*	0.3511±0.0158*

3 Transfer learning-based speech emotion recognition model

3.1 Transfer learning

With the increasing number of application scenarios, people's requirements for the speed and efficiency of machine models are becoming higher and higher. The establishment of traditional machine learning models requires a large amount of labeled data, but in actual scenarios, there are more unlabeled data. In this context, transfer learning appears [15]. Transfer learning refers to transferring knowledge from one domain (the source domain) to another domain (the target domain) in order to obtain better learning results. It is defined as follows.

Condition: Source domain D_s, task T_s on the source domain, target domain D, task T_t on the target domain are given, and $D_s \neq D_t$ or $T_s \neq T_t$ should be satisfied.

Objective: Its objective is to learn objective function $f(\cdot)$ using D_s and T_s and apply it in target domain.

3.2 Transfer component analysis

TCA is a classic algorithm in transfer learning [16]. It is assumed that the data distribution approximates to $P(\phi(D_s)) \approx P(\phi(D_t))$ after the feature mapping of D_s and D_t (ϕ). The core idea of TCA is to solve ϕ. The distribution distance is calculated based on the maximum mean difference (MMD). The objective function is:

$$dist(D_s, D_t) = \left\| \frac{1}{n_1}\sum_{i=1}^{n_1} \phi(d_{si}) - \frac{1}{n_2}\sum_{j=1}^{n_2} \phi(d_{tj}) \right\|_H,$$

where n_1 is the number of samples in D_s, $d_{si} \in D_s$, n_2 is the number of samples in D_t, $d_{tj} \in D_t$, and $\|\cdot\|_H$ is the reproducing kernel Hilbert space norm.

3.3 Speech emotion recognition model based on transfer learning

For Chinese and English SER, based on TCA, this paper uses JSL for the emotion features after dimensionality reduction to get the feature transform subspace to generate categorical emotion features and then uses a support vector machine (SVM) [17] as the classifier to get the final SER result. It is called the TCA-JSL algorithm.1

After extracting features from the source and target domains, TCA transformation is carried out to reduce the distribution distance between source domain sample feature X_s and target domain sample feature X_t. Then, the JSL method is used to find a feature transformation subspace (D_s) to minimize the conditional distribution distance between D_s and D_t:

$$\min_M(\|X_s^T M - f_s\|_F^2 + \|X_t^T M - f_t\|_F^2 + \|X_c^T M - f_c\|_F^2),$$

where f_s and f_t are the feature representation of the source domain label and target domain under M respectively, D_c is the set of samples identified as the same type in D_s and D_t, and f_c is the feature representation of the sample identified as the same type c in D_s and D_t under M. In order to avoid M overfitting, $L_{2,1}$ norm is introduced into the above equation for restriction. $L_{2,1}$ norm integrates the robustness advantages of L_2 and L_1 norms, which has appropriate sparsity and can reduce the impact of abnormal values. The $L_{2,1}$ norm of M is:

$$\|M\|_{2,1} = 2Tr(M^T V M),$$

$$V = [v_{ii}] \in R^{p \times p},$$

where V is the diagonal matrix. When ε tends towards 0, $v_{ii} = \frac{1}{2\sqrt{\|m^i\|^2 + \varepsilon}}$.

On this basis, distance metric matrix M is generated for the samples in D_s and D_t. The nearest K neighbors are found for each instance in the matrix, marked as 1, and the rest is marked as 0. Then,

$$dist1(M) = Tr(M^T X J X^T M),$$

$J = D - G$,

where J is the Laplacian matrix, D is the diagonal matrix, d_{ij} is the sum of each column of G: $d_{ij} = \sum_j g_{ij}$.

Then, the distance measure of D_s category label space (B) is regularized to minimize the distance of samples of the same type in D_s in this space:

$dist2(M) = Tr(M^T BJ'B^T M)$,

$J' = D' - G'$,

The resulting objective function is:

$dist(M) = Tr(M^T XJX^T M) + Tr(M^T BJ'B^T M)$.

In the TCA-JSL algorithm, the final solution formula can be written as:

$L(f_t, M) = \alpha(\|X_s^T M - f_s\|_F^2 + \|X_t^T M - f_t\|_F^2) + \beta\|X_c^T M - f_c\|_F^2 + \gamma[Tr(M^T VM)] + \delta[Tr(M^T XJX^T M)] + \varphi[Tr(M^T BJ'B^T M)]$,

where α, β, γ, δ, and φ are hyperparameters that are set based on the varying importance of different parts. The conditions are:

$$\begin{cases} \alpha > 0, \beta > 0 \\ \alpha + \beta = 1 \\ \gamma > 0 \\ \delta > 0 \\ \varphi > 0 \end{cases}.$$

The derivative of M is taken to obtain the updated formula of M, f_s, and f_t:

$M^* = (\alpha X_s X_s^T + \alpha X_t X_t^T + \beta X_c X_c^T - \gamma V - \delta XJX^T - \varphi BJ'B^T)^{-1} \cdot (\alpha X_s^T f_s + \alpha X_t^T f_t + \beta X_c^T f_c)$,

$f_s^* = X_s^T M^*$,

$f_t^* = X_t^T M^*$.

A SVM is used on the updated f_s and f_t for Chinese and English SER to obtain the final emotion recognition results.

4 Results and analysis

4.1 Experimental setup

The experiments were conducted in the Windows 10 and MATLAB environment, using the PyTorch framework and the Python language. In order to evaluate the performance of the TCA-JSL algorithm for Chinese and English SER, three speech emotion libraries were selected for experiments.

The first one was a Chinese speech emotion library (http://www.chineseldc.org/resource_info.php%20%20?rid=76): CASIA [18], including six categories: angry, fear, happy, sad, neutral, and surprise. Two males and two females were selected to read texts with different emotions as mentioned above.

The second library was an English speech emotion library (https://www.enterface.net/results/): eNTERFACE [19], including six categories: happiness, sadness, anger, surprise, disgust, and fear. Forty-two subjects from 14 different nationalities were told to listen to six consecutive short stories, each of which would trigger a specific emotion. Then, they responded to each situation and said specific words. The experimenters recorded the experimental process that met the requirements as videos as dataset samples.

The last one was an English speech emotion library (http://kahlan.eps.surrey.ac.uk/savee/Download.html): SAVEE [20], including seven categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. Four male speakers expressed seven emotion categories. There were 15 sentences for each emotion category.

The experiment adopted the ten-fold cross-validation method, and the average value was taken as the final result. The speech features used in the experiment were from the specified feature set of the INTERSPEECH2010 Emotion Challenge [21], and they were extracted by the openSMILE tool [22], with a total of 1,582 dimensions. The features are presented in Table 3.

Table 3: Model input characteristics.1

Feature	Dimension
F0 start time	1
Duration	1
F0 fundamental frequency	38
Local jitter	38
Local perturbation	38
Continuous jitter frame pairs	38
F0 envelope	42
Loudness	42
Dullness frequency distribution	42
Logarithmic Mel frequency band	336
Line spectrum pair frequency	336
Mel frequency cepstral coefficient	630

The following two indicators were used to evaluate the SER effect:

precision: $P = TP/(TP + FP)$,

recall rate: $R = TP/(TP + FN)$.

In the above equations, TP is the number of samples correctly identified in a certain category of emotion, FP is the number of samples that are wrongly identified as a certain category of emotion in other categories, and FN is the number of samples that are wrongly identified as another category of emotion in a certain category of emotion.

4.2 Analysis of results

After repeated experiments, the parameters of the TCA-JLS algorithm were determined as: $\alpha = 0.4$, $\beta = 0.6$, $\gamma = 0.2$, $\delta = 0.5$, and $\varphi = 0.1$. Under different combinations, the recognition results of the TCA-JSL algorithm were compared with other transfer learning-based algorithms, including TCA, joint distributed adaptation (JDA), and joint distribution adaptive regression (JDAR). Moreover, significance tests were conducted.

Table 4: Comparison of P values between different methods for CASIA and eNTERFACE.

	CASIA→eNTERFACE	**eNTERFACE→CASIA**
TCA	0.2864±0.0124	0.2512±0.0126
JDA	0.3132±0.0135	0.2884±0.0137
JDAR	0.4774±0.0141	0.4211±0.0142
TCA-JSL	0.4950±0.0152*	0.4533±0.0151*

Note: * indicates a significant difference compared to the JDAR method, $p < 0.05$.

Table 5: Comparison of R values between different methods for CASIA and eNTERFACE.23

	CASIA→eNTERFACE	**eNTERFACE→CASIA**
TCA	0.2602±0.0133	0.2641±0.0131
JDA	0.2659±0.0137	0.2701±0.0135
JDAR	0.2843±0.0142	0.3032±0.0143
TCA-JSL	0.3542±0.0163*	0.3511±0.0161*

Note: * indicates a significant difference compared to the JDAR method, $p < 0.05$.

From Tables 4 and 5, it can be found that the recognition effect of the TCA algorithm was poor for the SER of CASIA and eNTERFACE, and both p and R values were below 0.3. The performance of the JDA and JDAR algorithms was slightly better than that of the TCA method. Compared with these methods, the TCA-JSL algorithm showed a superior emotion recognition effect. The p and R values of the TCA-JSL algorithm for CASIA→eNTERFACE were 0.4950 ± 0.0152 and 0.3542 ± 0.0163, respectively, and the p and R values for eNTERFACE→CASIA were 0.4533 ± 0.0151 and 0.3511 ± 0.0161, respectively. Compared with the JDAR method, $p < 0.05$. These results demonstrate that the TCA-JSL algorithm had advantages in the SER of CASIA and eNTERFACE.

Table 6: Comparison of P values between different methods for CASIA and SAVEE.

	CASIA→SAVEE	**SAVEE→CASIA**
TCA	0.3321±0.0156	0.3551±0.0155
JDA	0.3676±0.0161	0.3972±0.0158
JDAR	0.4308±0.0172	0.4725±0.0173
TCA-JSL	0.4521±0.0176*	0.4987±0.0175*

Note: * indicates a significant difference compared to the JDAR method, $p < 0.05$.

Table 7: Comparison of R values between different methods for CASIA and SAVEE45

	CASIA→SAVEE	SAVEE→CASIA
TCA	0.2776±0.0137	0.2784±0.0138
JDA	0.2945±0.0142	0.2864±0.0139
JDAR	0.3318±0.0155	0.3346±0.0156
TCA-JSL	0.3544±0.0161*	0.3511±0.0158*

Note: * indicates a significant difference compared to the JDAR method, $p < 0.05$.

From Tables 6 and 7, it can be found that the TCA-JSL algorithm performed better in the Chinese and English SER of CASIA and SAVEE. The mean P and R values obtained by the traditional TCA method were low, while the mean P and R values obtained by the JDA algorithm were also below 0.4 and 0.3 respectively. The recognition effect of the JDAR algorithm were improved to a certain extent, but it was inferior to the TCA-JSL algorithm. The p and R values of the TCA-JSL algorithm for CASIA→SAVEE were 0.4521 ± 0.0176 and 0.3544 ± 0.0161, respectively. The p and R values of the TCA-JSL algorithm for SAVEE→CASIA were 0.4987 ± 0.0175 and 0.3511 ± 0.0158, respectively. Compared with the JDAR method, $p < 0.05$. The experiments on the two emotion databases revealed that the TCA-JSL algorithm had a better performance on Chinese and English SER and demonstrated statistically significant improvements over current algorithms in terms of Chinese and English cross-library SER.

5 Discussion

Chinese and English SER has very important research value with the wide use of Chinese and English worldwide. In this paper, a TCA-JSL algorithm was developed, and cross-database experiments were carried out on the Chinese speech emotion database CASIA and the English speech emotion databases eNTERFACE and SAVEE. The algorithm was compared with some existing transfer learning algorithms, with P value and R value as evaluation indicators.

From the results, it can be found that in cross-database emotion recognition, the p and R values of the TCA-JSL algorithm were both superior to those of the other compared algorithms. The P value for CASIA→eNTERFACE was 0.4950 ± 0.0152, and the R value was 0.3542 ± 0.0163. The P value for eNTERFACE→CASIA was 0.4533 ± 0.0151, and the R value was 0.3511 ± 0.0161. The P value for CASIA→SAVEE was 0.4521 ± 0.0176, and the R value was 0.3544 ± 0.0161. The P value for SAVEE→CASIA was 0.4987 ± 0.0175, and the R value was 0.3511 ± 0.0158. Analysis showed that algorithms such as TCA ignored the availability of the source-domain label features, while the TCA-JSL algorithm effectively balanced the overall features of the source and target domains and took into account the source-domain label features, enabling it to achieve higher performance in Chinese-English SER. The differences in the results of different cross-library emotion

recognition may be related to the number of emotion categories and the number of samples within the libraries, resulting in some differences in the p and R values. The results of the statistical significance test showed that compared with the JDAR algorithm, the differences of the TCA-JSL algorithm in different cross-library experiments were all significant ($p < 0.05$), verifying the superiority of the TCA-JSL algorithm in Chinese-English SER.

The TCA-JSL algorithm can achieve good results in Chinese and English speech emotion recognition. Therefore, it can be applied in the real world. For example, in human-computer interaction, this method can be used to effectively realize speech emotion recognition, provide support for the diagnosis of psychological emotions such as depression and anxiety, and also provide the customer emotion recognition function for telephone customer service to judge the needs of customers.

6 Conclusion

In this paper, a TCA-JSL algorithm was developed for Chinese and English pronunciation emotion recognition combined with transfer learning, and cross-library experiments were carried out on CASIA, eNTERFACE and SAVEE emotion libraries. The results showed that compared with some other transfer learning-based algorithms, the TCA-JSL algorithm could obtain better recognition results, which can be further promoted and applied in the actual SER.

References

[1] Oladipupo M A, Obuzor P C, Bamgbade B J, Adeniyi A E, Olagunju K M, Ajagbe S A (2023). An automated python script for data cleaning and labeling using machine learning technique. *Informatica*, 47, pp. 219-232.

[2] Gasteiger N, Lim J Y, Hellou M, MacDonald B A, Ahn H S (2024). A scoping review of the literature on prosodic elements related to emotional speech in human-robot interaction. *International Journal of Social Robotics*, 16(4), pp. 659-670. https://doi.org/10.1007/s12369-022-00913-x.

[3] Alam K, Nigar N, Erler H, Banerjee A (2023). Speech emotion recognition from audio files using feedforward neural network. *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2023, pp. 1-6. https://doi.org/10.1109/ECCE57851.2023.10101492.

[4] Aishwarya N, Kaur K, Seemakurthy K (2024). A computationally efficient speech emotion recognition system employing machine learning classifiers and ensemble learning. *International Journal of Speech Technology*, 27(1), pp. 239-254. https://doi.org/10.1007/s10772-024-10095-8.

[5] De Lope J, Grana M (2023). An ongoing review of speech emotion recognition. *Neurocomputing*, 528(APr.1), pp. 1-11. https://doi.org/10.1016/j.neucom.2023.01.002.

[6] Li W, Xue J, Tan R, Wang C, Deng Z, Li S, Guo G, Cao D (2023). Global-local-feature-fused driver speech emotion detection for intelligent cockpit in automated driving. *IEEE Transactions on Intelligent Vehicles*, 8, pp. 2684-2697. https://doi.org/10.1109/TIV.2023.3259988

[7] Vyakaranam A, Maul T, Ramayah B (2024). A review on speech emotion recognition for late deafened educators in online education. *International Journal of Speech Technology*, 27(1), pp. 29-52. https://doi.org/10.1007/s10772-023-10064-7.

[8] Wang X, Zhao S, Wang Y (2021). Bimodal emotion recognition for the patients with depression. *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, 2021, pp. 40-43. https://doi.org/10.1109/ICSIP52628.2021.9688837.

[9] Nasersharif B, Ebrahimpour M, Naderi N (2023). Multi-layer maximum mean discrepancy in auto-encoders for cross-corpus speech emotion recognition. *Journal of Supercomputing*, 79, pp. 13031-13049. https://doi.org/10.1007/s11227-023-05161-y.

[10] Kumar S, Thiruvenkadam S (2021). An analysis of the impact of spectral contrast feature in speech emotion recognition. *International Journal of Recent Contributions from Engineering Science & IT (iJES)*, 9, pp. 87-95. https://doi.org/10.3991/ijes.v9i2.22983.

[11] Atmaja B T, Akagi M (2021). Evaluation of error- and correlation-based loss functions for multitask learning dimensional speech emotion recognition. *Journal of Physics: Conference Series*, 1896(1), pp. 012004-. https://doi.org/10.1088/1742-6596/1896/1/012004.

[12] Xie Y, Liang R, Liang Z, Zhao X, Zeng W (2023). Speech emotion recognition using multihead attention in both time and feature dimensions. *IEICE Transactions on Information and Systems*, 106, pp. 1098-1101. https://doi.org/10.1587/transinf.2022edl8084.

[13] Long M, Wang J, Ding G, Sun J, Yu P S (2013). Transfer feature learning with joint distribution adaptation. *Proceedings of the 2013 IEEE International Conference on Computer Vision*, 2013, pp. 2200-2207. https://doi.org/10.1109/ICCV.2013.274.

[14] Zhang J, Jiang L, Zong Y, Zheng W, Zhao L (2021). Cross-corpus speech emotion recognition using joint distribution Adaptive Regression. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 3790-3794. https://doi.org/10.1109/ICASSP39728.2021.9414372.

[15] Priya Dharshini G, Sreenivasa Rao K (2024). Transfer accent identification learning for enhancing speech emotion recognition. *Circuits, Systems & Signal Processing*, 43(8), pp. 5090-5120. https://doi.org/10.1007/s00034-024-02687-1.

[16] Qu H, Dong H, Pang L (2023). Mental workload classification method based on transfer component

analysis with cross-session EEG data. *International Conference on Man-Machine-Environment System Engineering*, 941, pp. 17-23. https://doi.org/10.1007/978-981-19-4786-5_3.

[17] Gjoreski M, Gjoreski H, Kulakov A (2014). Machine learning approach for emotion recognition in speech. *Informatica: An International Journal of Computing and Informatics*, 38(4), pp. 377-384.

[18] Li H, Zhou Z, Sun X, Li C (2020). Multi-features integration for speech emotion recognition. *International Conference on Pattern Recognition and Artificial Intelligence*, 2020, pp. 191-202. https://doi.org/10.1007/978-3-030-59830-3_17.

[19] Martin O, Kotsia I, Macq B, Pitas I (2006). The eNTERFACE'05 audio-visual emotion database. *International Conference on Data Engineering Workshops*, 2006, pp. 8. https://doi.org/10.1109/ICDEW.2006.145.

[20] Yogesh C K, Hariharan M, Ngadiran R, Adom A H, Yaacob S, Berkai C, Polat K (2016). A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Systems with Applications*, 69, pp. 149-158. https://doi.org/10.1016/j.eswa.2016.10.035.

[21] Schuller B, Steidl S, Batliner A, Burkhardt F, Devillers L, Müller C A, Narayanan S S (2010). The INTERSPEECH 2010 paralinguistic challenge. *11th Annual Conference of the International Speech Communication Association*, 2010, pp. 2794-2797.

[22] Eyben F, Wllmer M, Schuller B (2010). Opensmile: the munich versatile and fast open-source audio feature extractor. *ACM International Conference on Multimedia*, 2010, pp. 1459-1462. https://doi.org/10.1145/1873951.1874246.