Cross lingual speech emotion recognition using canonical correlation analysis on principal component subspace

Hesam Sagha, Jun Deng, Maryna Gavryukova, Jing Han, Björn Schuller

Angaben zur Veröffentlichung / Publication details:

Sagha, Hesam, Jun Deng, Maryna Gavryukova, Jing Han, and Björn Schuller. 2016. "Cross lingual speech emotion recognition using canonical correlation analysis on principal component subspace." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 20-25 March 2016, Shanghai, China*, 5800–5804. Piscataway, NJ: IEEE. https://doi.org/10.1109/icassp.2016.7472789.

CROSS LINGUAL SPEECH EMOTION RECOGNITION USING CANONICAL CORRELATION ANALYSIS ON PRINCIPAL COMPONENT SUBSPACE

Hesam Sagha[†], *Jun Deng*[‡†], *Maryna Gavryukova*[†], *Jing Han*[†], *Björn Schuller*[‡†♭]

[†]Chair of Complex & Intelligent Systems, University of Passau, Passau, Germany
[♭]Department of Computing, Imperial College London, London, UK
[‡]Machine Intelligence & Signal Processing group, Technische Universität München, Munich, Germany

ABSTRACT

This paper proposes an analytical approach based on Kernel Canonical Correlation Analysis (KCCA) for domain adaptation. To generate paired instances for KCCA, we mapped source and target data onto both source and target principal components. We performed pair-wise domain adaptation between four emotional speech corpora with different languages (English, German, Italian, and Polish) to validate the approach. We compared our approach with the Shared-Hidden-Layer Auto-Encoder (SHLA) and kernel based principal components. On average, the proposed approach yields higher classification performance.

Index Terms— Transfer learning, domain adaptation, cross lingual, emotion recognition, canonical correlation analysis

1. INTRODUCTION

The fast pace of progress in the ubiquitous Internet facilitates collection of more data in less time. This is beneficial for the machine learning domain where more data can represent better the feature distribution. However, the side effect is that the labels of the collected data may not be available and they need to be annotated by human effort. This could be costly, tedious, cumbersome and time consuming, even by using crowed-sourcing platforms [1]. Semi-supervised approaches such as active learning try to reduce this effort by automatic labeling of the data which have a high probability in a class and get the label from a human when the certainty is not adequate [2]. Nevertheless, for big databases this approach will be also less practical. Instead, transfer learning (TL) approaches try to use the knowledge which is already gained from other databases and use this knowledge for a new database [3]. Therefore, no more human effort would be needed for the annotation. In addition, in TL, it is not necessary to hold the assumption of having the same distribution for training and test data. This is beneficial when the training and test data are not obtained in the same way (e. g., studio vs. real environment audio recording) or their types are different (e. g., image vs. speech). TL approaches are categorized based on the domain and the task of the source and the target corpora. Generally, if both domains and both tasks are the same and the target corpus is not annotated, the problem is called *domain adaptation* [4].

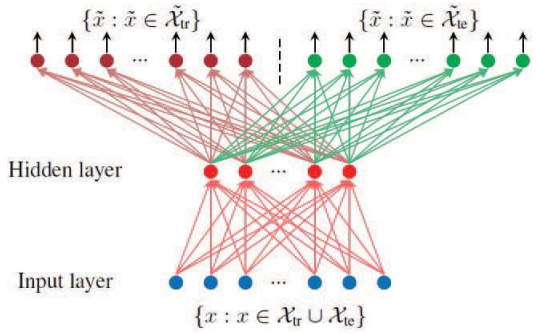

Fig. 1. Domain adaptation using Shared-Hidden-Layer Auto-Encoder.

Kernel Mean Matching (KMM) was proposed to deal with domain adaptation problems by directly estimating the resampling weights through matching training and test distribution means in a reproducing kernel Hilbert space [5]. Recently, it was applied to reduce the acoustic and speaker difference across training and test data for speech emotion recognition [6]. Furthermore, Deng et al. proposed the use of Shared-Hidden-Layer Auto-Encoders (SHLA) to obtain shared views of source and target emotional speech corpora [7]. In this approach, having two data corpora $\chi_{tr} \in \mathbb{R}^{n \times Q}$ and $\chi_{te} \in \mathbb{R}^{m \times Q}$ with different sample size (m and n) and Q features, an artificial neural network with Q neurons in the input layer, $H < Q$ neurons in the hidden layer, and $2Q$ neurons in the output layer is created (cf. Fig. 1). A gradient descent approach is performed to tune the weights. Finally, the outputs of the hidden layer are used for the training and classification. They compared this method with KMM, showing an improvement in cross-lingual emotion classification from speech.

In this paper, motivated by the success of SHLA, we

propose a domain adaption method, which applies Canonical Correlation Analysis (CCA) to find the views with the highest correlations between source and target corpora. CCA has been applied for speaker adaptation in the domain of speech recognition [8, 9], and audiovisual synchronization [10] among others. CCA is a statistical method to find linear bases so that the correlations between the projections of the variables onto these bases are mutually maximized [11]. Kernel CCA (KCCA) is a variant of CCA where it uses the kernel trick to capture non-linear correlation hidden in data. KCCA has been widely used for multimodal dimensionality reduction, such as for fMRI analysis [12] and speaker identification [13]. Further, it was found that CCA based feature reduction can overcome the problem of over-fitting and provide a compact set of high quality features for computational paralinguistics applications [14]. Additionally, it has also been used as multi-view transfer learning for cross-language information retrieval where a parallel corpus is generated by text translation [15, 16].

Rather than using CCA as a feature reduction, we extend CCA to alleviate the mismatch between different languages for emotion recognition from speech. We generate two views of each training and test corpora to construct two paired corpora. Finally, we use Kernel CCA to find the views of the two paired corpora where their mappings onto those views have the highest correlation.

The remainder of this paper is organized as follows. The next section provides the basis for CCA and Kernel CCA and our approach to apply it on domain adaptation. In Section 3, we describe the databases, and in Section 4, we provide the results and compare it with SHLA. Finally, Section 5 draws conclusions and points out future directions.

2. METHOD

Similar to Auto-Encoder transfer learning, the idea is to seek a shared representation of features for the source and target databases. Then, a model is built on these features from the source database, and is used to label the target database. In the following, first we introduce general CCA and Kernel CCA, and then we describe the proposed approach on how to deploy Kernel CCA for transfer learning.

2.1. CCA

Consider two databases $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ and $\boldsymbol{Y} \in \mathbb{R}^{n \times p}$ having n paired multivariate random vectors $(\boldsymbol{x_i}, \boldsymbol{y_i})$ with dimensions d and p, respectively. CCA finds mappings (views) for X and Y so that the mapped data are highly correlated. In other words, CCA maximizes:

$$\rho = \max_{\boldsymbol{w},\boldsymbol{v}} corr(\boldsymbol{X}\boldsymbol{w}, \boldsymbol{Y}\boldsymbol{v}) = \max_{\boldsymbol{w},\boldsymbol{v}} \frac{\boldsymbol{w}^T \boldsymbol{C}_{xy} \boldsymbol{v}}{\sqrt{\boldsymbol{w}^T \boldsymbol{C}_{xx} \boldsymbol{w} \boldsymbol{v}^T \boldsymbol{C}_{yy} \boldsymbol{v}}}. \tag{1}$$

where $\boldsymbol{C}_{xy}$ is the cross covariance matrix between $\boldsymbol{X}$ and $\boldsymbol{Y}$. $\boldsymbol{w}$ can be found through Lagrangian approach and is the eigenvectors in the form of:

$$\boldsymbol{C}_{xy} \boldsymbol{C}_{yy}^{-1} \boldsymbol{C}_{yx} \boldsymbol{w} = \lambda^2 \boldsymbol{C}_{xx} \boldsymbol{w}. \tag{2}$$

Then, we select the N vectors corresponding to the N largest eigenvalues. $\boldsymbol{v}$ is equal to

$$\boldsymbol{v} = \frac{\boldsymbol{C}_{yy}^{-1} \boldsymbol{C}_{yx} \boldsymbol{w}}{\lambda}. \tag{3}$$

$\boldsymbol{X}\boldsymbol{w}$ and $\boldsymbol{Y}\boldsymbol{v}$ have the highest correlation on the vector corresponding to the largest eigenvalue, and the second highest on the second vector corresponding to the second largest eigenvalue and so on. The upper limit of N is the minimum of rank of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

2.2. Kernel CCA (KCCA)

The Kernel CCA defines a Kernel on data and similar to CCA it seeks to maximize the correlation between mappings of these kernels:

$$\rho = \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{\boldsymbol{\alpha}^T \boldsymbol{K}_x \boldsymbol{K}_y \boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^T \boldsymbol{K}_x^2 \boldsymbol{\alpha} \boldsymbol{\beta}^T \boldsymbol{K}_y^2 \boldsymbol{\beta}}}, \tag{4}$$

where $\boldsymbol{K}_x$ and $\boldsymbol{K}_y$ are the kernel matrices corresponding to the two representations. As linear kernel they are $\boldsymbol{K}_x = \boldsymbol{X}\boldsymbol{X}^T$ and $\boldsymbol{K}_y = \boldsymbol{Y}\boldsymbol{Y}^T$, and as RBF kernels they are defined as $\boldsymbol{K}_x(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}}$, where σ is a free parameter. The solution to (4) is in the form of an eigenproblem:

$$(\boldsymbol{K}_x + \kappa \boldsymbol{I})^{-1} \boldsymbol{K}_y (\boldsymbol{K}_y + \kappa \boldsymbol{I})^{-1} \boldsymbol{K}_x \boldsymbol{\alpha} = \lambda^2 \boldsymbol{\alpha}, \tag{5}$$

where κ is the regularization parameter. Moreover,

$$\boldsymbol{\beta} = \frac{(\boldsymbol{K}_y + \kappa \boldsymbol{I})^{-1} \boldsymbol{K}_x \boldsymbol{\alpha}}{\lambda}. \tag{6}$$

Finally, the two mapped vectors are $\boldsymbol{K}_x \boldsymbol{\alpha}$ and $\boldsymbol{K}_y \boldsymbol{\beta}$ [17].

2.3. KCCA-based domain adaptation

CCA and KCCA are useful when $\boldsymbol{x}_i$s and $\boldsymbol{y}_i$s are paired together. For example, Kaya et al. used CCA and KCCA for feature reduction where $\boldsymbol{x}_i$ are the features and $\boldsymbol{y}_i$ are the binarized class labels [14]. Different from the use of KCCA for feature reduction, this paper makes use of KCCA in conjunction with Principal Component Analysis (PCA) for domain adaptation. Note that, PCA is used to create two representations (views) of each corpora on two sets of orthogonal vectors, as principal components, to preserve information on lower dimensions. The schematic of the approach is shown in Fig. 2. First, source data $\boldsymbol{X}$ is mapped once on its principal components, $\boldsymbol{X}^{p_x}$, and another time on target data

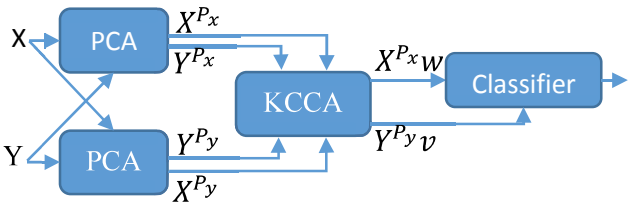

Fig. 2. proposed approach. Superscript P_x and P_y denote the data mapping to $\boldsymbol{X}$'s and $\boldsymbol{Y}$'s principal components, respectively.

principal components, $\boldsymbol{X}^{p_y}$ to create two views of $\boldsymbol{X}$. Similarly, target data is mapped on its principal components, $\boldsymbol{Y}^{p_y}$, and on source data principal components, $\boldsymbol{Y}^{p_x}$ to create two views of $\boldsymbol{Y}$. Then, we reduce dimensions for each mapping to keep 99% of variation on principal components (This will also help to avoid singularity during CCA process). Then, we find the shared view between the paired mapped data on the source principal components, $[\boldsymbol{x}^{p_x}; \boldsymbol{y}^{p_x}]$ and the target principal components $[\boldsymbol{x}^{p_y}; \boldsymbol{y}^{p_y}]$ using canonical correlation analysis. Afterwards, we pick top N dimensions of the mapped data with largest correlations. Finally, we train a classifier on the mapped training data and test it on the mapped test data.

In general, there is no need to have the two views with the same number of transferred features. However, for the following analyses we kept them the same, equal to the maximum of the two reduced dimensions.

3. EXPERIMENT

We compare the results of using no domain adaptation with five models of transfer learning on four emotional speech databases. Two of the domain adaptation methods are the proposed KCCA approach with RBF and linear kernels. We set $N = 30, 40, 50$ to create three models and we used Bayesian fusion to combine the decisions of each model. We set the κ value as 10. The two other approaches are based on Kernel PCA with linear and RBF kernels. In this case, we map data on three subspaces; $\boldsymbol{X}$'s principal components, $\boldsymbol{Y}$'s principal components, and $[\boldsymbol{X}; \boldsymbol{Y}]$'s principal components. We perform the classification on these three views and combine the decisions. Additionally, we compare the methods with the SHLA with the same number of hidden layers ($N = 30, 40, 50$) and decision fusion approach. We ran this process ten times and provided the average of the results.

3.1. Databases

Four emotional speech databases with different languages have been investigated. Some information about these databases are provided in Table 1. Utterances of all corpora are generated by actors/actresses in studio environment. EMODB is a German emotional speech corpus where 10 sentences with emotionally neutral content is uttered in different emotions. In the SAVEE corpus, each actor uttered 15 sentences in different emotions and they are validated by 10 subjects. The Italian corpus (EMOVO) contains utterances of 14 sentences simulating six emotional states plus neutral state. In the Polish Emotional Speech Dataset, each speaker uttered five different sentences with six types of emotional load. To uniform labels, we have mapped emotions onto two classes: positive and negative valence. This mapping is provided in Table 1.

3.2. Feature extraction

We extracted 384 features as in the Interspeech 2009 Emotion Challenge using openSMILE [22]. It comprises 12 functionals of 2×16 acoustic Low-Level Descriptors (LLDs) including their first delta regression. The LLDs are zero-crossing-rate, root mean square of frame energy, pitch frequency, harmonics-to-noise ratio by autocorrelation function and Mel-frequency cepstral coefficients 1-12. The 12 functionals are minimum, maximum, mean, standard deviation, kurtosis, skewness, relative position, ranges, and two linear regression coefficients with their mean square error. Additionally, we removed the features which are highly (positive or negative) correlated with each other ($|\rho| > 0.95$) or if they have small variance ($< 10^{-10}$). This feature pruning keeps 311 features. Moreover, we removed the outlier data where a feature value is larger than 10 times of the standard deviation. Finally, we performed subject based normalization followed by corpus-based normalization (SC-normalization) which is shown to boost cross-language emotion recognition [23]. The SC-normalized data is fed to the transfer learning method.

4. RESULT

Table 2 shows the performance of classification using Simple Logistic classifier. The choice of this non-parametric classifier was to avoid parameter tuning and have a fair comparison between databases. Unweighted Average Recall (UAR) is used as performance measure. Classifications without transfer learning are denoted as *'Direct C'* for corpus normalized data and *'Direct SC'* for SC-normalized data. Only in one case (out of 12) the performance has not been improved by transfer learning. In 7 cases (out of 11) KCCA, in 3 cases SHLA, and in 1 case PCA provide the highest accuracy. As it can be seen, there is no clear winner between the methods. However, on average, *Direct SC* yields 2.5% improvement with respect to *Direct C*. Furthermore, KCCA (Linear), KCCA (RBF), KPCA (Linear), KPCA (RBF) and SHLA yield 2.19%, 2.81%, 2.71%, 2.61%, and 2.0% average improvement over *Direct SC*, respectively.

The advantage of the KCCA and KPCA over SHLA is

Table 1. Corpora information and the mapping of class labels onto Negative/Positive valence. (#m): number of male speaker, (#f): number of female speakers, (Rate): Sampling rate.

Corpus	Language	#m	#f	Rate	Negative Valence (#)		Positive Valence (#)	
EMODB [18]	German	5	5	16	Anger, Sadness, Fear, Disgust, Boredom	(385)	Neutral, Happiness	(150)
SAVEE [19]	English	4	0	44	Anger, Sadness, Fear, Disgust	(240)	Neutral, Happiness, Surprise	(240)
EMOVO [20]	Italian	3	3	44	Anger, Sadness, Fear, Disgust	(336)	Neutral, Joy, Surprise	(252)
Polish [21]	Polish	4	4	44	Anger, Sadness, Fear, Boredom	(160)	Neutral, Joy	(80)

Table 2. UAR of transfer learning methods.

		EMODB	SAVEE	EMOVO	Polish
EMODB	Direct C		55.8	56.4	71.9
	Direct SC		59.2	**58.7**	69.0
	KCCA (Lin.)		64.6	57.9	**72.8**
	KCCA (RBF)		**65.2**	57.0	72.5
	KPCA (Lin.)		64.4	57.6	72.2
	KPCA (RBF)		63.5	57.2	70.0
	SHLA		63.7	56.8	59.7
SAVEE	Direct C	62.5		54.6	65.0
	Direct SC	63.1		57.6	62.8
	KCCA (Lin.)	70.6		59.3	69.4
	KCCA (RBF)	**71.9**		**59.5**	67.5
	KPCA (Lin.)	70.1		58.7	72.8
	KPCA (RBF)	67.9		58.3	**74.1**
	SHLA	67.7		59.0	70.2
EMOVO	Direct C	58.0	51.2		55.3
	Direct SC	59.1	55.4		71.0
	KCCA (Lin.)	62.9	**58.5**		65.9
	KCCA (RBF)	60.2	56.0		**71.9**
	KPCA (Lin.)	65.4	56.0		62.8
	KPCA (RBF)	66.7	56.0		67.5
	SHLA	**67.3**	58.2		64.3
Polish	Direct C	65.1	55.4	57.3	
	Direct SC	65.9	56.9	54.3	
	KCCA (Lin.)	70.9	58.7	56.1	
	KCCA (RBF)	68.4	**61.9**	54.1	
	KPCA (Lin.)	69.3	58.7	57.8	
	KPCA (RBF)	67.9	57.1	57.6	
	SHLA	**71.1**	60.6	**58.3**	

the analytical solution instead of gradient descent. Therefore, there is no risk of falling into a local minima and the learning speed is faster. Moreover, using KCCA prevents the necessity of having the same number and type of features. On the other hand, autoencoders with large number of layers and nodes with non-linear activation function can represent better non-linearity in the data distribution. However, to achieve this non-linear mapping, there is a need for much more data samples.

5. CONCLUSION

This paper proposed an approach to use Kernel Canonical Correlation Analysis (CCA) on principal component subspaces for domain adaptation. Cross corpora transfer learning for an emotion recognition task from four emotional speech corpora with different languages have been chosen to investigate the validity of our approach. The Kernel CCA has been compared with the Kernel PCA as well as an autoencoder approach named as Shared Hidden Layer Autoencoder (SHLA). On average, the proposed approach performs better than the others.

Our future study will focus on the use of non-linear mapping instead of linear PCA to generate subspaces. Additionally, discriminant mappings (such as Discriminant Locality Preserving CCA [24]) could be applied on the training corpora to increase the level of discrimination on the corresponding subspace. Finally, having bigger datasets, we will also investigate Deep Canonical Correlation Analysis [25] on domain adaptation.

Acknowledgments

The research leading to these results has received funding from the the European Union's Horizon 2020 Programme research and innovation programme under grant agreements Nos. 644632 (MixedEmotions) and 645094 (SEWA) and from the German national BMBF IKT2020-Grant under grant agreement No. 16SV7213 (EmotAsS).

6. REFERENCES

[1] S. Hantke, F. Eyben, T. Appel, and B. Schuller, "iHEARu-PLAY: Introducing a game for crowed-sourced data collection for affective computing," in *Proc. of the 6th International Conference on Affective Computing and Intelligent Interaction*. IEEE, 2015, pp. 891–897.

[2] Z. Zhang, F. Ringeval, B. Dong, E. Marchi E. Coutinho, and B. Schuller, "Enhanced Semi-Supervised Learning for Multimodal Emotion Recognition," in *Proc. of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, To appear.

[3] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[4] S. Sun, H. Shi, and Y. Wu, "A survey of multi-source domain adaptation," *Information Fusion*, vol. 24, pp. 84–92, 2015.

[5] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt, and B. Schölkopf, "Covariate shift by kernel mean matching," *Dataset shift in machine learning*, vol. 3, no. 4, pp. 5, 2009.

[6] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: compensating for covariate shift," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1458–1468, 2013.

[7] J. Deng, R. Xia, Z. Zhang, Y. Liu, and B. Schuller, "Introducing Shared-Hidden-Layer autoencoders for transfer learning and their application in acoustic emotion recognition," in *Proc. of the 39th IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014, vol. 338164, pp. 4818–4822.

[8] Y. Grenier, "Speaker adaptation through canonical correlation analysis," in *Proc. of the 5th IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr 1980, pp. 888–891.

[9] K. Choukri and G. Chollet, "Adaptation of automatic speech recognizers to new speakers using canonical correlation analysis techniques," *Computer Speech & Language*, vol. 1, no. 2, pp. 95–107, 1986.

[10] M. E. Sargin, Y. Yemez, E. Erzin, and A. M. Tekalp, "Audiovisual synchronization and fusion using canonical correlation analysis," *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1396–1403, 2007.

[11] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[12] D. R. Hardoon, J. Mourao-Miranda, M. Brammer, and J. Shawe-Taylor, "Unsupervised analysis of fMRI data using kernel canonical correlation," *NeuroImage*, vol. 37, no. 4, pp. 1250–1259, 2007.

[13] K. Livescu and M. Stoehr, "Multi-view learning of acoustic features for speaker recognition," in *Proc. of the 11th IEEE Workshop on Automatic Speech Recognition & Understanding*. IEEE, 2009, pp. 82–86.

[14] H. Kaya, F. Eyben, A. A. Salah, and B. Schuller, "CCA based feature selection with application to continuous depression recognition from acoustic speech features," in *Proc. of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 3729–3733.

[15] Y. Li and J. Shawe-Taylor, "Using KCCA for Japanese–English cross-language information retrieval and document classification," *Journal of intelligent information systems*, vol. 27, no. 2, pp. 117–133, 2006.

[16] B. Fortuna and J. Shawe-Taylor, "The use of machine translation tools for cross-lingual text mining," in *Proc. of the ICML Workshop on Learning with Multiple Views, Germany*, 2005, pp. 28–33.

[17] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[18] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech," in *Proc. of Interspeech*, 2005, vol. 5, pp. 1517–1520.

[19] S. Haq, P. JB. Jackson, and J. Edge, "Audio-visual feature selection and reduction for emotion classification," in *Proc. of the International Conference on Auditory-Visual Speech Processing, Tangalooma, Australia*, 2008, pp. 185–190.

[20] G. Costantini, I. Iaderola, A. Paoloni, and M. Todisco, "Emovo corpus: an italian emotional speech database," in *Proc. of the 9th International Conference on Language Resources and Evaluation, Iceland*, 2014, pp. 3501–3504.

[21] P. Staroniewicz and W. Majewski, "Polish emotional speech database — recording and preliminary validation," in *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*, vol. 5641 of *Lecture Notes in Computer Science*, pp. 42–49. Springer Berlin Heidelberg, 2009.

[22] F. Eyben and B. Schuller, "openSMILE:): The Munich open-source large-scale multimedia feature extractor," *SIGMultimedia Records*, vol. 6, no. 4, pp. 4–13, 2015.

[23] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: variances and strategies," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 119–131, 2010.

[24] X. Zhang, N. Guan, Z. Luo, and L. Lan, "Discriminative locality preserving canonical correlation analysis," in *Pattern Recognition*, pp. 341–349. Springer, 2012.

[25] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep Canonical Correlation Analysis," in *Proc. of the 30th International Conference on Machine Learning*, 2013, pp. 1247–1255.