A Semi-Supervised Learning Approach for Acoustic-Prosodic Personality Perception in Under-Resourced Domains

Rubén Solera-Ureña[1], *Helena Moniz*[1,2,6], *Fernando Batista*[1,3], *Vera Cabarrão*[1,2], *Anna Pompili*[1,4], *Ramón Fernández-Astudillo*[1,6], *Joana Campos*[4,5], *Ana Paiva*[4,5], *Isabel Trancoso*[1,4]

[1]Spoken Language Systems Laboratory, INESC-ID Lisboa, Lisboa, Portugal
[2]FLUL/CLUL, Universidade de Lisboa, Lisboa, Portugal
[3]Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
[4]Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[5]Intelligent Agents and Synthetic Characters Group, INESC-ID Lisboa, Lisboa, Portugal
[6]Unbabel Lda, Portugal

{rsolera,helenam,fmmb,anna,ramon.astudillo,imt}@l2f.inesc-id.pt, veracabarrao@gmail.com

Abstract

Automatic personality analysis has gained attention in the last years as a fundamental dimension in human-to-human and human-to-machine interaction. However, it still suffers from limited number and size of speech corpora for specific domains, such as the assessment of children's personality. This paper investigates a semi-supervised training approach to tackle this scenario. We devise an experimental setup with age and language mismatch and two training sets: a small labeled training set from the Interspeech 2012 Personality Sub-challenge, containing French adult speech labeled with personality OCEAN traits, and a large unlabeled training set of Portuguese children's speech. As test set, a corpus of Portuguese children's speech labeled with OCEAN traits is used. Based on this setting, we investigate a weak supervision approach that iteratively refines an initial model trained with the labeled data-set using the unlabeled data-set. We also investigate knowledge-based features, which leverage expert knowledge in acoustic-prosodic cues and thus need no extra data. Results show that, despite the large mismatch imposed by language and age differences, it is possible to attain improvements with these techniques, pointing both to the benefits of using a weak supervision and expert-based acoustic-prosodic features across age and language.

Index Terms: computational paralinguistics, automatic personality perception, OCEAN, cross-language, cross-age, semi-supervised learning

1. Introduction

The analysis of personality traits has a plethora of applications, such as discriminating natural from disordered behaviors or automatically assessing personality traits, either in human-human communications or in human-computer interactions. Much of the literature on automatic processing of personality traits is still mostly focused on assessing and detecting the traits based on several sets of distinct features. Artificial intelligence applications are, however, taking steps towards endowing robots and virtual agents with certain traits to better interact with humans, making the communication more idiosyncratic and tuned to the paralinguistic fingerprints of an interlocutor. The *Big-Five* (**OCEAN**) personality traits is a widely used psychological model that aims at describing human personality in terms of five broad dimensions: **O**penness (artistic, imaginative, original), **C**onscientiousness (organized, efficient, thorough), **E**xtroversion (energetic, outgoing, talkative), **A**greeableness (kind, generous, sympathetic), and **N**euroticism (anxious, self-pitying, worrying).

The automatic perception/classification of personality traits is still a very challenging task, either due to the individual spectrum of a speaker, or to the spectrum of the trait itself: whenever the richness of a person is defined by the *Big-Five* model in five personality dimensions, it may not cover all the subspecifications or the boundaries between such classes, as psychological studies have pointed out [1]. It is clear in the literature that some traits can be more easily recognized by means of automatic procedures than others, but this fact may vary according to the data and the methodologies applied (see [2] for a survey). Moreover, it has been timidly pointed out that different personality traits are revealed in spontaneous speech by means of different sets of representative acoustic/prosodic features [2, 3, 4, 5, 6], but exhaustive categorizations of such features and studies across ages, cultures, etc. are still very scarce.

In addition, computational perception of personality is a recent field of research and speech datasets annotated in terms of personality traits are still scarce and small, which hinders the development of robust and accurate personality models. Psychological studies have shown a strong debate between change and continuity of personality traits from childhood to adult age, or even elderly in longitudinal studies [7, 8, 9, 10]. The studies in [7] show that children's personality traits are linked to the ones displayed in adult age. In this same line, we hypothesized the existence of a consistent set of acoustic/prosodic features for Extroversion and Agreeableness in both adult and children speech, pointing out to reasonable performance rates for the perception of personality traits across different languages and ages [11]. This opens the door to the use of heterogeneous data sets in personality perception tasks as a way to circumvent the scarcity of labeled data in under-resourced domains.

Building upon our previous work, we devise here a more solid experimental setup by adding new personality annotations of the test set, and by incorporating a large, unlabeled database with Portuguese children's speech that is used in a semi-supervised learning approach to learn more solid personality models. For this purpose we apply the concept of transfer learning, following other successful applications of this approach to emotion recognition in speech [12, 13].

The paper is organized as follows: Section 2 presents the speech databases employed in this work. Section 3 describes the

acoustic/prosodic features used as the basis for the automatic personality perception models described in Section 4. Experimental results are presented in Section 5 and the paper ends with conclusions in Section 6.

2. Cross-age and cross-language datasets

This work pursues performing automatic perception of children's personality by taking advantage of heterogeneous corpora in a cross-language and cross-age setup. The well-known Speaker Personality Corpus (SPC) [14, 15, 16] database has been used here to train statistical models (binary classifiers) for each personality trait in the *Big-Five* model (**OCEAN**). The more populated, unlabeled CNG Corpus of European Portuguese Children's Speech (CNG) [17] was then used in a self-learning (semi-supervised) approach to iteratively refine the initial models. Finally, the Game-of-Nines (GoN) corpus [18] has been used as a test set to study how personality models built up from French adults' speech can be used to assess the *Big-Five* dimensions of personality of Portuguese children, and to evaluate the performance of the semi-supervised learning procedure adopted in this work.

2.1. Speaker Personality Corpus

The Speaker Personality Corpus consists of 640 speech files from 322 different Swiss-French speaking adult individuals. Each file contains 10 seconds of speech from just one speaker (around 1 hour and 40 minutes in total). All the files were independently assessed by 11 judges using the BFI-10 personality questionnaire [19]. For each file, a high or a low level is assigned for every personality trait (denoted as O/NO, C/NC, E/NE, A/NA, N/NN, respectively) using a majority vote procedure. We refer to [15] for a detailed description of the division of the SPC corpus into train, development and test subsets.

2.2. Game-of-Nines Corpus

The Game-of-Nines corpus was originally designed to study how conflict unfolds in social interactions by looking at behavioral cues (e.g. gaze) in a mixed-motive social interaction (i.e. a scenario with competitive and cooperative incentives) with children. It comprises synchronized video- and audio-recordings of 11 dyadic sessions with 22 Portuguese children aged 10 to 12 years-old, playing a bargaining card game (a modified version of the *Game of Nines* [20]). The duration of the recordings varies between 9 and 18.6 minutes, with an average duration of 12.8 minutes and a total of 2 hours and 20 minutes.

A preliminary pre-processing of the original GoN database was performed in order to adapt it for our purposes [11]. As a result, three different speech subsets were generated, which allow for a comparison of the effect of long/medium/short acoustic cues on personality perception systems:

1. *GoN-complete*: all the speech segments for a given child during the game session were concatenated together in one single speech file. As a results, the *GoN-complete* subset consists of 22 files ranging from 49 seconds to 8.1 minutes of speech (average duration of 4.2 minutes).
2. *GoN-20seconds*: for each child, 4 different files with around 20 seconds of speech were generated by concatenating their longer speech segments in the session. Very short segments (below 2 seconds) were discarded in order to avoid an excessive variability in the speech characteristics, resulting in just 2 files for one of the participants. As a result, the *GoN-20seconds* subset consists of 86 files with an approximate duration of 20 seconds.
3. *GoN-10seconds*: this subset was constructed by splitting each file in the *GoN-20seconds* subset into approximately 2 halves, resulting in a subset of 172 files with an approximate duration of 10 seconds each.

The original video recordings in the GoN database have been independently annotated in terms of the *Big-Five* personality dimensions by three experienced assessors (1 psychologist and 2 professional speech practitioners) using the BFI-10 personality questionnaire. These annotations have been used as the ground-truth labels in this work. The inter-annotator agreement values in terms of the Fleiss' Kappa coefficient are 0.673 for Openness, 0.151 for Conscientiousness, 0.292 for Extroversion, 0.07 for Agreeableness, and 0.209 for Neuroticism (mean value of 0.279). Although it is not straightforward to make comparisons across different experimental setups, these values are in line with those reported in the literature, e.g., in [21]. We refer to [11] for a more detailed description of the GoN subsets, including the number of examples in each class.

2.3. CNG Corpus of EP Children's Speech

The CNG Corpus of European Portuguese (EP) Children's Speech [17] comprises around 20 hours of speech from 484 speakers. The corpus contains four different types of utterances spoken by children aged 3 to 10 years old: phonetically rich sentences, musical notes, isolated cardinal numbers, and sequences of cardinal numbers. Data is organized in two different subsets with children aged 3 to 6 years old and children aged 7 to 10 years old, respectively. Depending on their age and reading skills, the children either read the prompts, or repeated them after a supervisor.

In this study, we have just used the subset of phonetically rich sentences uttered by children aged 7 to 10 years-old (6 hours of speech). The reasons are: i) that is the more similar subset to the target domain (Portuguese children aged 10 to 12 years-old), and ii) phonetically rich sentences are more prone to display personality traits, rather than the other types of utterances in the CNG database.

3. Feature extraction

The experiments performed in this work use two sets of features extracted with openSMILE [22], and a set of knowledge-based features known in the literature to have impact on the classification of personality-related tasks, henceforth referred to as KB-features.

3.1. Baseline features

The first set (IS2012) consists of 6125 features and was created in the scope of the Interspeech 2012 Speaker Trait Challenge-Personality Sub-challenge. We have also used the eGeMAPS feature set [23], an extended version of GeMAPS - Geneva Minimalistic set of Acoustic Parameters for Voice Research and Affective Computing, that consists of 88 features well-known for their usefulness in a wide range of paralinguistic tasks.

3.2. Knowledge-based features

Our knowledge-based features (KB-features) are based on phone tokenizations of the speech files using the neural network-based acoustic models of the AUDIMUS speech recognizer [24]. The phonetic tokenizations provide phone align-

ments for each speech file, which can be used to extract duration-related features and to generate more advanced features. In this way, for instance, it is possible to extract the silence ratio, speech duration ratio, and speech rate features in terms of phones per second. The phone tokenizations also make it possible to characterize each speech segment using n-grams of phones. Based on these tokenizations, we then derive *Inter Pausal Units* (IPUs), that consist of sequences of phones delimited by silences. Our experiments use both French and Portuguese phone models.

The experiments presented in this work use a set of 41 knowledge-based features, including duration of speech with and without internal silences, and tempo measurements such as speech and articulation rates (number of phones or syllables divided by the duration of speech with and without internal silences, respectively) and phonation ratio (duration of speech without internal silences divided by the duration of speech including internal silences). Other features involve pitch (f0), energy, jitter and shimmer, including pitch and energy average, median, standard deviation, dynamics, range, and slopes, both within and between IPUs [25]. Pitch related features were calculated based on semitones rather than frequency. On top of such features, we extracted elaborated prosodic features for the whole sentence involving the sequences of derived IPUs, that were expressed in terms of standard deviation, slope and concavity. The Snack Sound Toolkit[1] was used to extract the pitch and energy from the speech signal. Jitter and shimmer were extracted from openSMILE low-level descriptors. For the time being KB-features are still not extensive and have been used in combination with eGeMAPS features in order to achieve improved performances, amounting to a total of 129 features.

4. System for automatic personality perception

In this work, the same experimental setup as that employed in the Interspeech 2012 Speaker Trait Challenge-Personality Subchallenge [15, 16] has been adopted. We use support vector machines (SVM) with logistic functions fitted to the SVM soft outputs as statistical models (binary classifiers) for automatic personality perception. Special attention has been paid to feature normalization given the heterogeneous characteristics of the different corpora used in this work. Two different normalization techniques have been used ($[0, 1]$ range -denoted as NORM.-, and zero-mean and unit-variance -denoted as STAND.-).

4.1. Supervised learning

Five different models were trained in this work corresponding to the *Big-Five* dimensions of personality (**OCEAN**). Each model is trained to assign a high/low level on that trait (denoted as O/NO, C/NC, E/NE, A/NA, N/NN) to every speech file. A grid-search approach using the train and development subsets of the SPC corpus was applied to find the optimal value for the complexity parameter C of the SVMs. The value for C providing the higher unweighted average recall (UAR) on the development subset was then selected. Then, the training and development subsets were merged together and the definitive SVM models were trained on this data set, using the selected values for C. Finally, the UAR on four different test sets (SPC test set, *GoN-complete*, *GoN-20seconds* and *GoN-10seconds* sets) was calculated to evaluate the models on both same- and cross-language conditions.

[1] http://www.speech.kth.se/snack/

4.2. Semisupervised learning

A iterative self-learning (semi-supervised) approach has been applied starting from the initial models described above. At each iteration, the current model is used to classify the remaining samples in the unlabeled CNG dataset, and the 100 samples with the maximum output probabilities for each class (high/low on a given trait) are then extracted from the CNG dataset and joined together to the current training set, forming a new training set to be used in the next iteration. The labels assigned by the current model are used as ground truth labels to train the model in the next iteration. This process is iterated 8 times.

5. Experimental results

The results are presented in terms of unweighted average recall (UAR) and accuracy (Acc). When the data is substantially unbalanced, which is the case of the data sets employed here, UAR should be used as the most relevant measure.

Tables 1, 2, 3 and 4 present the results for the initial models learnt by means of the completely supervised approach. Table 1 reveals that Conscientiousness and Extroversion can be easily perceived in the SPC corpus. Overall, openSMILE features achieve the best performance, except for Neuroticism. When using $[0, 1]$ range normalization, the eGeMAPS features achieve considerably lower values for Openness and for Conscientiousness, but the difference is much smaller for Openness when using zero-mean and unit-variance normalization. The combination of KB-features and eGeMAPS is almost consistently better than the eGeMAPS features alone.

Tables 2, 3 and 4 present the results achieved for the *GoN-complete*, *GoN-20seconds* and *GoN-10seconds* subsets, respectively. In comparison with our previous results in [11], we observe that the Openness trait can be reasonably perceived now in the GoN subsets. This fact may be related with the recent additional personality annotations of this dataset, which may have lead to more consistent labels. Overall, the results show reasonable performance rates for the perception of Openness, Extroversion and Agreeableness traits across languages and ages.

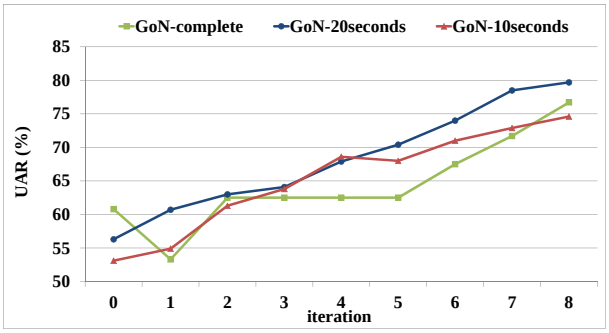

Figure 1: *Oppeness: results achieved on the GoN subsets -self-learning.*

Figures 1 and 2 show the results (UAR) achieved by the models trained at each iteration of the semi-supervised approach for Oppenness and Extroversion, respectively. The zero-mean and unit-variance normalization and the eGeMAPS feature set was used in these experiments. Although still in a preliminary stage, these results point out the potentials of applying self-learning approaches as a method to overcome the lack of

Table 1: *Results achieved on the SPC data set -initial models-.*

TRAIT	NORM.	openSMILE					eGeMAPS					eGeMAPS+KB-feats.				
		C	DEVELOP		TEST		C	DEVELOP		TEST		C	DEVELOP		TEST	
			UAR	Acc	UAR	Acc		UAR	Acc	UAR	Acc		UAR	Acc	UAR	Acc
O	NORM.	3E-1	61.4	65.0	59.1	58.2	1E-1	50.0	61.7	52.3	43.3	1E-5	51.8	62.3	54.8	55.2
	STAND.	1E-7	63.8	66.7	58.9	61.7	3E-4	64.2	67.2	56.4	58.7	1E-4	64.0	66.7	53.7	56.2
C	NORM.	3E-5	73.1	73.2	75.7	75.6	3E-1	70.2	68.3	69.4	69.2	1E-5	73.6	73.2	70.8	70.6
	STAND.	3E-4	73.9	74.3	78.6	78.6	3E-2	71.3	71.6	75.1	75.1	3E-5	73.3	73.2	75.0	75.1
E	NORM.	3E-3	82.0	82.0	74.9	75.6	1E-1	82.0	82.0	67.7	69.7	1E-4	81.4	81.4	68.5	70.1
	STAND.	1E-4	82.0	82.0	75.0	75.1	1E-2	79.8	79.8	72.5	72.6	3E-3	82.0	82.0	74.1	74.1
A	NORM.	3E-4	67.3	66.1	54.3	53.7	1E-5	66.2	65.6	56.6	56.2	3E-4	67.0	66.7	57.0	56.2
	STAND.	3E-8	68.4	66.7	60.3	60.2	1E-5	65.8	63.9	58.1	58.2	3E-4	66.3	64.5	58.2	58.2
N	NORM.	3E-5	64.9	64.5	62.3	61.7	3E-6	69.1	68.9	62.7	62.2	1E0	69.7	69.9	66.2	68.2
	STAND.	3E-4	65.7	65.6	65.1	64.7	1E-7	69.6	69.4	63.2	63.2	1E-7	69.1	68.9	64.8	64.7

Table 2: *Results achieved on the GoN-complete data set -initial models-.*

TRAIT	NORM.	openSMILE			eGeMAPS			eGeMAPS+KB-feats.		
		C	TEST		C	TEST		C	TEST	
			UAR	Acc		UAR	Acc		UAR	Acc
O	NORM.	3E-1	61.7	63.6	1E-1	65.8	68.2	1E-5	55.0	59.1
	STAND.	1E-7	70.0	72.7	3E-4	60.8	63.6	1E-4	70.8	72.7
C	NORM.	3E-5	38.3	36.4	3E-1	43.3	40.9	1E-5	50.0	45.5
	STAND.	3E-4	42.5	40.9	3E-2	37.5	36.4	3E-5	33.3	31.8
E	NORM.	3E-3	67.9	59.1	1E-1	71.4	63.6	1E-4	71.4	63.6
	STAND.	1E-4	71.4	63.6	1E-2	65.2	59.1	3E-3	68.8	63.6
A	NORM.	3E-4	64.3	68.2	1E-5	59.8	59.1	3E-4	59.8	59.1
	STAND.	3E-8	64.3	68.2	1E-5	59.8	59.1	3E-4	57.1	59.1
N	NORM.	3E-5	27.5	27.3	3E-6	27.5	27.3	1E0	26.7	27.3
	STAND.	3E-4	31.7	31.8	1E-7	27.5	27.3	1E-7	27.5	27.3

Table 3: *Results achieved on the GoN-20seconds data set -initial models-.*

TRAIT	NORM.	openSMILE			eGeMAPS			eGeMAPS+KB-feats.		
		C	TEST		C	TEST		C	TEST	
			UAR	Acc		UAR	Acc		UAR	Acc
O	NORM.	3E-1	57.4	59.3	1E-1	62.1	64.0	1E-5	57.7	60.5
	STAND.	1E-7	65.5	67.4	3E-4	56.3	58.1	1E-4	58.8	60.5
C	NORM.	3E-5	39.4	36.0	3E-1	43.5	43.0	1E-5	46.5	41.9
	STAND.	3E-4	40.1	38.4	3E-2	46.4	45.3	3E-5	45.6	44.2
E	NORM.	3E-3	64.3	53.5	1E-1	68.9	60.5	1E-4	64.5	55.8
	STAND.	1E-4	59.8	54.7	1E-2	71.1	67.4	3E-3	71.9	67.4
A	NORM.	3E-4	64.4	69.8	1E-5	63.0	64.0	3E-4	63.0	64.0
	STAND.	3E-8	63.5	68.6	1E-5	61.3	61.6	3E-4	61.3	61.6
N	NORM.	3E-5	31.7	31.4	3E-6	28.2	29.1	1E0	39.4	43.0
	STAND.	3E-4	36.7	36.0	1E-7	31.1	31.4	1E-7	32.3	32.6

Table 4: *Results achieved on the GoN-10seconds data set -initial models-.*

TRAIT	NORM.	openSMILE			eGeMAPS			eGeMAPS+KB-feats.		
		C	TEST		C	TEST		C	TEST	
			UAR	Acc		UAR	Acc		UAR	Acc
O	NORM.	3E-1	57.6	59.3	1E-1	55.1	58.1	1E-5	52.5	55.8
	STAND.	1E-7	60.9	62.8	3E-4	53.1	54.7	1E-4	54.2	55.8
C	NORM.	3E-5	42.2	38.4	3E-1	48.8	50.0	1E-5	49.5	44.2
	STAND.	3E-4	37.8	36.0	3E-2	45.4	44.8	3E-5	47.7	45.9
E	NORM.	3E-3	61.8	52.3	1E-1	65.9	57.6	1E-4	61.8	51.7
	STAND.	1E-4	58.5	54.1	1E-2	67.2	63.4	3E-3	65.0	60.5
A	NORM.	3E-4	60.0	64.5	1E-5	61.5	64.5	3E-4	61.9	64.5
	STAND.	3E-8	59.0	62.2	1E-5	59.1	59.3	3E-4	62.5	62.2
N	NORM.	3E-5	32.6	32.0	3E-6	28.7	30.8	1E0	45.8	51.2
	STAND.	3E-4	34.6	34.3	1E-7	33.2	33.7	1E-7	33.9	34.3

labeled data in severely under-resourced domains such as the task of children's personality perception addressed in this work.

6. Conclusions

This paper investigates a semi-supervised training approach to tackle personality perception tasks in severely under-resourced domains, where heterogeneous and partially non-labeled data sets can be used in order to circumvent the scarcity of labeled data for the target domain (Portuguese children in this work). Thus, we devise here an experimental setup with age and language mismatch and two training sets: a small labeled training set from the Interspeech 2012 Personality Sub-challenge, containing French adult speech labeled with personality OCEAN traits (SPC), and a large unlabeled training set of Portuguese children's speech (CNG). As test set, a corpus of Portuguese children's speech labeled with OCEAN traits is used (GoN). Based on this setting, we investigate a weak supervision approach that iteratively refines our initial models by using the unlabeled data-set. We also investigate knowledge-based features, which leverage expert knowledge in acoustic-prosodic cues and thus need no extra data. Results show that, despite the large mismatch imposed by language and age differences in the training and test data sets, it is possible to attain improvements with these techniques, pointing both to the benefits of using a weak supervision approach and expert-based acoustic-prosodic features across age and language.

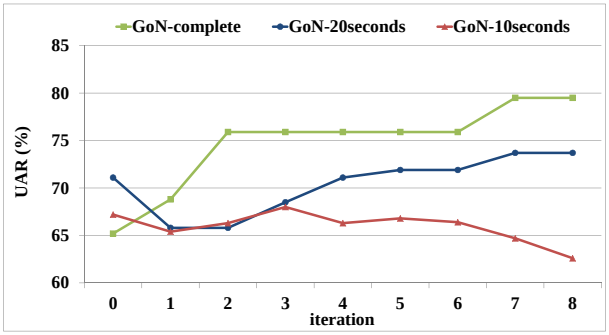

Figure 2: *Extroversion: results achieved on the GoN subsets -self-learning.*

7. Acknowledgements

This work has been supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, under Post-doc grant SFRH/PBD/95849/2013 and grant SFRH/BD/97187/2013, by project H2020-EU.3.7 contract 653587 (LAW-TRAIN), project CMUP-ERI/HCI/0051/2013 (INSIDE), and project CMUP-ERI/TIC/0033/2014 (BioVisualSpeech).

8. References

[1] S. Cloninger, "Conceptual issues in personality theory," in *The Cambridge Handbook of Personality Psychology*, P. Corr and G. Matthews, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2009, vol. 4, ch. 8, pp. 3–26.

[2] A. Vinciarelli and G. Mohammadi, "A Survey of Personality Computing," *IEEE Transaction on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[3] F. Mairesse, M. Walker, M. Mehl, and R. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research*, vol. 30, no. 1, pp. 457–500, 2007.

[4] T. Polzehl, S. Moeller, and F. Metze, "Automatically assessing acoustic manifestations of personality in speech," in *Spoken Language Technology Workshop*, Berkeley, CA, USA, Dec. 2010, pp. 7–12.

[5] ——, "Automatically Assessing Personality from Speech," in *International Conference on Semantic Computing*, Pittsburgh, PA, USA, Sep. 2010, pp. 134–140.

[6] ——, "Modeling Speaker Personality Using Voice," in *Proc. of Interspeech 2011*, Florence, Italy, Aug. 2011, pp. 2369–2372.

[7] A. Caspi, H. Harrington, B. Milne, J. Amell, R. Theodore, and T. Moffitt, "Children's Behavioral Styles at Age 3 Are Linked to Their Adult Personality Traits at Age 26," *Journal of Personality*, vol. 71, no. 4, pp. 495–514, 2003.

[8] J. Kagan, "Three pleasing ideas," *American Psychologist*, vol. 51, pp. 901–908, 1996.

[9] M. Lewis, *Altering fate: Why the past does not predict the future*. New York: Guilford Press, 1997.

[10] B. Roberts, K. Walton, and W. Viechtbauer, "Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies," *Psychological Bulletin*, vol. 132, no. 1, pp. 1–25, 2006.

[11] R. Solera-Ureña, H. Moniz, F. Batista, R. Fernández-Astudillo, J. Campos, A. Paiva, and I. Trancoso, "Acoustic-prosodic Automatic Personality Trait Assessment for Adults and Children," in *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference, IberSPEECH 2016*, Lisbon, Portugal, Nov. 2016.

[12] E. Coutinho, J. Deng, and B. Schuller, "Transfer learning emotion manifestation across music and speech," in *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, Jul. 2014, pp. 3592–3598.

[13] J. Deng, Z. Zhang, and B. Schuller, "Linked source and target domain subspace feature transfer learning–exemplified by speech emotion recognition," in *22nd International Conference on Pattern Recognition (ICPR 2014)*, Stockholm, Sweeden, Aug. 2014, pp. 761–766.

[14] G. Mohammadi and A. Vinciarelli, "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 273–284, 2012.

[15] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, and F. Burkhardt, "The INTERSPEECH 2012 Speaker Trait Challenge," in *Proc. of Interspeech 2012*, Portland, OR, USA, Sep. 2012.

[16] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "A Survey on perceived speaker traits: Personality, likability, pathology, and the first challenge," *Computer Speech & Language*, vol. 29, no. 1, pp. 100–131, Jan. 2015.

[17] A. Hämäläinen, F. M. Pinto, S. Rodrigues, A. Júdice, S. M. Silva, A. Calado, and M. S. Dias, "A Multimodal Educational Game for 3-10-Year-Old Children: Collecting and Automatically Recognising European Portuguese Children's Speech," in *Workshop on Speech and Language Technology in Education*, Grenoble, France, Aug. 2013.

[18] J. Campos, P. Oliveira, and A.Paiva, "Looking for Conflict: Gaze Dynamics in a Dyadic Mixed-Motive Game," *Autonomous Agents and Multi-Agent Systems*, vol. 30, no. 1, pp. 112–135, 2016.

[19] B. Rammstedt and O. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, vol. 41, pp. 203–212, 2007.

[20] H. Kelley, L. Beckman, and C. Fischer, "Negotiating the division of a reward under incomplete information," *Journal of Experimental Social Psychology*, vol. 3, no. 4, pp. 361–398, 1967.

[21] O. P. John and R. W. Robins, "Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self," *Journal of personality*, vol. 61, no. 4, pp. 521–551, 1993.

[22] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent Developments in openSMILE, the Munich Open-source Multimedia Feature Extractor," in *Proc. of the 21st ACM International Conference on Multimedia*, New York, NY, USA, Oct. 2013, pp. 835–838.

[23] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "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, 2016.

[24] H. Meinedo, M. Viveiros, and J. Neto, "Evaluation of a Live Broadcast News Subtitling System for Portuguese," in *Proc. of Interspeech 2008*, Brisbane, Australia, Sep. 2008, pp. 508–511.

[25] F. Batista, H. Moniz, I. Trancoso, and N. J. Mamede, "Bilingual Experiments on Automatic Recovery of Capitalization and Punctuation of Automatic Speech Transcripts," *IEEE Transactions on Audio, Speech and Language Processing, Special Issue on New Frontiers in Rich Transcription*, vol. 20, no. 2, pp. 474–485, Feb. 2012.