Comparative Study of Speaker Personality Traits Recognition in Conversational and Broadcast News Speech

Firoj Alam, Giuseppe Riccardi

Department of Information Engineering and Computer Science, University of Trento, Italy

{alam, riccardi}@disi.unitn.it

Abstract

Natural human-computer interaction requires, in addition to understand *what* the speaker is saying, recognition of behavioral descriptors, such as speaker's personality traits (SPTs). The complexity of this problem depends on the high variability and dimensionality of the acoustic, lexical and situational context manifestations of the SPTs. In this paper, we present a comparative study of automatic speaker personality trait recognition from speech corpora that differ in the source speaking style (broadcast news vs. conversational) and experimental context. We evaluated different feature selection algorithms such as information gain, relief and ensemble classification methods to address the high dimensionality issues. We trained and evaluated ensemble methods to leverage base learners, using three different algorithms such as SMO (Sequential Minimal Optimization for Support Vector Machine), RF (Random Forest) and Adaboost. After that, we combined them using majority voting and stacking methods. Our study shows that, performance of the system greatly benefits from feature selection and ensemble methods across corpora.

Index Terms: Speaker Personality trait Recognition, Ensemble methods, Information gain, Relief

1. Introduction

In our daily communication, we interact with unknown individuals, even with machines that exhibit human-like features and behaviors, including robots, embodied virtual agents, animated characters etc. [1]. To make these automated systems more like human, we need to understand non-verbal characteristics of human speech such as speaker personality traits.

It has been a long-term goal for psychologists to understand human personality and its impact on human behavior. Behavior involves an interaction between a person's underlying personality trains and situational variables. The situation, that a person finds himself or herself plays a major role in how the person reacts. However, in most of the cases, people often respond based on their underlying personality traits. With time, this area has attracted researchers from different fields, especially for the researchers in the human-machine interaction and behavioral analytics.

It is suggested in [1, 2] that naturalness and efficiency of interaction to a user increases by matching user's personality. Studies have been done on how style of communications like emails, blog entries [11] etc. depends on the author's personality, and the choice of particular parts-of-speech [12].

This paper follows previous research [3,9,21] on designing algorithms to extract features from speech that best predict SPTs as well as machine learning algorithms that tackle the high-dimensionality and variability of the classification problem. In particular, this paper comparatively evaluates SPTs automatic recognition algorithms on two speech corpora drawn from different speaking styles and data collection conditions. We evaluated the SPTs specific feature selection algorithms as well as their impact on the base and the ensemble classification systems.

The paper is organized as follows: Section 2 describes the related work; Section 3 describes the corpora, which was used in the experiment; Section 4, defines the experimental method. Details of the classification results and discussion are given in Section 5. Finally, conclusion appears in Section 6.

2. Related Work

Personality is defined as the coherent patterning of affect, behavior, cognition and desire over time and space, which are used to characterize unique individuals. There are several theories for personality traits in the literature; however most widely used personality traits model is the Big-5, five broad personality dimensions [17]. It describes the human personality as a vector of five values corresponding to bipolar traits. This is a popular model among the language and computer science researchers, as it has been used as a framework for both personality traits identification and simulations. The Big-Five personality traits are:

O (**Openness**): Artistic, curious, imaginative, etc.

C (**Conscientiousness**): Efficient, organized, etc.

E (**Extraversion**): Energetic, active, assertive, etc.

A (**Agreeableness)**: Compassionate, cooperative etc.

N (**Neuroticism**): Anxious, tense, self-pitying, etc.

There has been active research since the first study done by Sapir [18] to understand the effect of speech on personality traits. However, automatic recognition of personality is relatively very recent. Major contributions have been done in the Interspeech 2012 speaker traits challenge [5, 21-19], where one of the sub-challenges was the recognition of the speaker personality traits. The contributions [21-29] in the evaluation campaign include studying different feature selection and classification techniques along with combining acoustic and linguistic features. From this evaluation campaign, it can be concluded that there is a great challenge of understanding, which features are important and which classification approach provides a better hypothesis. This also suggests that more investigation is needed in order to understand personality traits from speech. Among these studies, Kartik et al. [21] evaluated their system by using SPC development set (SPC dev) as a test set, which is comparable with our study on the SPC corpus.

One of the major problems in SPTs recognition is the ecological data collection and annotation. Interspeech-2012 challenge has triggered interests on the recognition of SPTs in the speech community and in this paper we used the corpus published in [5] (broadcast news). Persia corpus has been studied in [9] and to the best of our knowledge is the first corpus collected from human-human interactions designed to elicit SPTs.

3. Corpora

In this study, we experimented with two different corpora: (i) Speaker Personality Corpus (SPC), (ii) Personable and Intelligent virtual Agents (PerSIA) corpus.

3.1. SPC

SPC was obtained from the organizers of the Interspeech 2012 Speaker Trait Challenge [5]. The data set consists of training, development and test set, where each set is labeled as OCEAN tag, and each trait is mapped into two classes, positive and negative. This corpus consists of 640 audio files, that were randomly collected from the French news bulletins, broadcasted in February 2005, with the quality of 16 bit, 8kHz sample rate. Out of those clips professional speakers were produced 307 clips and 333 clips were from 210 non-professional speakers. Only one speaker was used for each audio clip and there were altogether 322 individual speakers. The corpus was assessed by 11 judges by listening to all the clips and individually evaluated the clips using BFI-10 [14]. The judges did not understand French, so the personality assessment could only be motivated by the nonverbal behavior. The dataset also consists of extracted acoustic features from those speech files, feature extraction *configuration file* (IS2012.conf) and a tool – *openSMILE* [6]. The annotation of the SPC test set was not available to us, as the organizers had not released it. For this reason, we used SPC dev set as a test set. The SPC train set consists of 256 instances and the dev set consists of 183 instances. Moreover, the distribution of the corpus is quite balanced.

3.2. Persia

The Personable and Intelligent virtual Agents (PerSIA[1]) [9] corpus is an Italian human-human spoken dialog corpus, recorded in a simulated tourist call center. Speakers played randomly the "customer" and the "agent" role over a telephone conversation. Each customer was given a tourism task to perform and the agent provided relevant answers. The task scenarios' difficulty ranged from easy to no-solution [9]. Out of the 24 speakers 12 were users and 12 were agents. Personality label was assigned based on the self-assessment questionnaire during the data collection. At the end, out of 144 (each user × agent) calls, 119 calls of Agent sub-corpus were used in the experiment. A distribution of the corpus consists of openness (Y-0.62%, N-0.38%), conscientiousness (Y-0.84%, N-0.16%), extraversion (Y-0.50%, N-0.50%), agreeableness (Y-0.66%, N-0.34%) and neuroticism (Y-0.50%, N-0.50%), where Y and N represent positive and negative.

4. Experimental Method

We conducted several experiments for this comparative study and to examine the performance of different feature selection and classification algorithms. For the experiment, we first extracted acoustic features and then used feature selection algorithms to select subset of features. After that, we applied ensemble methods as opposed to say 'classifier combination methods' for the final classification, which is explained in Section 4.3. It is evident that ensemble methods have also been studied for emotion and personality traits recognition from speech [21, 30]. A conceptual design of the system is given in Figure 1.

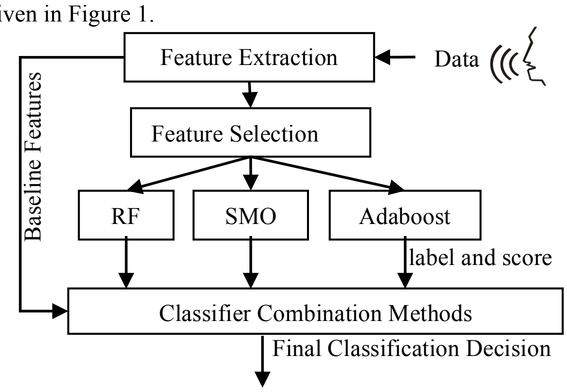

Figure 1: Conceptual flow of the system

4.1. Features

We extracted acoustic features using *openSMILE* [6] with the predefined *configuration file* provided in the Interspeech-2012 Speaker trait evaluation campaign. The low-level acoustic features extracted with approximately 100 frames per second with 10-30ms per frame. These low-level descriptors (LLDs) were then projected on single scalar values by descriptive statistical functionals [10]. More details of the acoustic features can be found in [5]. In this paper, we denote these acoustic features as our baseline features.

4.2. Feature Selection

We have high-dimensional problems p>>N, the number of features p, (6125) is much larger than the number of instances N. Therefore, to avoid high variance and overfitting we worked on two different feature selection techniques such as Information Gain (IG) [15] and Relief [20] along with equal frequency discretization method. Feature values were discretized into 2 equal frequency bins before applying feature selection algorithms. All acoustic features were continuous valued and converted into discrete value. This is because some feature selection algorithms like IG is not able to handle continuous value. Additionally, we applied discretization for relief feature selection as we were getting better results after applying discretization.

IG was proposed as a measure of estimating the features quality [16]. It tells us how well a given feature separates the training examples according to their target classification and calculated using the equation 1.

$$\text{Gain} = -\sum_{y \in Y} p(y) log_2 p(y) - \left(-\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log_2 p(y|x)\right) \quad (1)$$

where x and y are the feature value and the class label. $p(x)$ and $p(y)$ are the probabilities of x and y. $p(y|x)$ is the probability of y given x.

The idea of relief is to weight features according to how well their values discriminate with respect to a class label (in our case binary). Relief estimates weight of feature A ($W[A]$), by the following equation:

$$W[A] = P(x \mid nearest\ miss) - P(x \mid nearest\ hit) \quad (2)$$

[1] This corpus could be provided for research purpose upon request to sisl-data@disi.unitn.it.

where x is a different value of A, *nearest miss* and *nearest hit* are the nearest instance from a different and same class, respectively.

To identify the top ranked most informative features using these feature selection algorithms we generated feature learning curves by incrementally adding top ranked features. These learning curves were generated using our chosen classification algorithms – RF, SMO and Adaboost. From the feature learning curves we were able to figure out what range of feature we should select for different categories of personality traits. Figure 2 shows an example of feature learning curve for SPC using IG, relief and random feature selection with SMO classifier, where random feature selection was considered as a baseline study. In each learning point we also computed standard deviation from the cross validation results to see the statistical variation. Each of the feature selection algorithms behaves differently for each personality trait with different classification algorithms. Therefore, for different personality traits and for different feature selection algorithms we selected different number of features.

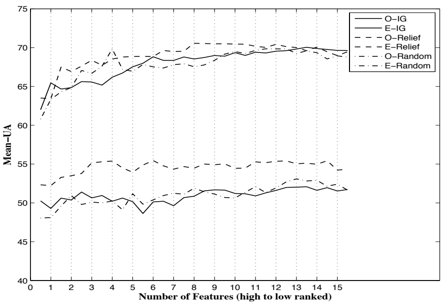

Figure 2: IG, relief and random feature selection learning curves with SMO classifier for the O and C categories, which shows different patterns. In x-axis, each point represents multiple of 400 top ranked features from left to right, whereas y-axis represents mean-UA of the LSGO cross validation.

4.3. Ensemble of Classifiers

For the classification of personality traits we conducted experiments with ensemble (classifier combination) methods where to design base learners we used RF, SMO and Adaboost (Ada). Ensemble methods were chosen due to their higher generalization ability [13] than just a single base learner. We choose three different classification algorithms in ensemble methods because of their different characteristics in classification. SMO [7] is an optimization technique for solving quadratic optimization problem, which arises during the training of SVM and it has better generalization capability. RF [8] is a combination of tree predictors and it builds a series of classification trees and each tree on its own makes a prediction. These predictions vote to make the RF prediction. RF reduces variances in classification by randomizing features and training instances. Adaptive Boosting (Adaboost) [4] is a meta-learner that uses greedy search for a linear combination of classifiers by overweighting the examples that are misclassified by each classifier. Similar to RF, Adaboost also reduces variances by randomizing the training instances. We used weka [19] for feature selection and classification.

As combiners in the ensemble methods, we conducted experiment using majority voting and stacking. Voting is the most popular and fundamental combination method for nominal outputs and the majority vote [13] is computed with the following equation 3.

$$H(x) = c_{\hat{j}};\ where\ \hat{j} = \mathrm{argmax_j} \textstyle\sum_{i=1}^{T} w_i h_i^j(x) \quad (3)$$

where $H(x)$ is the combined output of instance x; $h_i^j(x)$ is the output of the classifier h_i for the class label c_j; i=1…T is the number of classifiers; j=1…C is the number of classes; w_i is the weight assigned to classifier h_i. Here, we considered weight as 1.

Stacking [13] is a general procedure where a learner is trained to combine the base learners and the combiner is called second level learner or meta-learner. To train the meta-learner we used LSGO (leave speaker group out) cross validation. In LSGO, speakers were drawn randomly to make groups and the instances of the speaker groups were selected for the train and test set by leaving speaker-group-out approach. Base level classifier's decision and class probability were used as features in the meta-learner and we designed meta-learner using multi-response linear regression (MLR) [19].

4.4. Evaluation Methods

The performance of the system was measured in terms of weighted average (WA) and un-weighted average (UA) that have recently been used in the paralinguistic tasks [5]. For the sake of simplicity we are only showing UA in this paper.

For the SPC corpus we tuned parameters and selected features using LSGO cross validation on the SPC train set with macro averaging. In macro-average, UA and WA were calculated for each cross validation folds and then took the average. To measure the performance of the system we used the SPC dev set as a test set.

Moreover, for the Persia corpus we used LOSO (leave one speaker out) cross validation with micro averaging to tune parameters, feature selection and to measure the performance of the system. Micro-averaged values were calculated by constructing a global confusion matrix from all cross validation folds and then calculated UA_{micro} and WA_{micro} as given in equation 4 and 5. The reason to choose micro averaging is the imbalance class distribution in LOSO cross validation of the Persia corpus.

$$UA_{micro} = \frac{1}{2}\left(\frac{\sum_{i=1}^{F} TP_i}{\sum_{i=1}^{F} TP_i + FN_i} + \frac{\sum_{i=1}^{F} TN_i}{\sum_{i=1}^{F} TN_i + FP_i}\right) \quad (4)$$

$$WA_{micro} = \frac{\sum_{i=1}^{F} TP_i + TN_i}{\sum_{i=1}^{F} TP_i + FP_i + TN_i + FN_i} \quad (5)$$

where i=1…F is the number of folds. TP-true positive, TN-True negative, FP-False positive, FN-False negative.

5. Classification Results and Discussion

We evaluated BIG-5 personality traits binary classification models on both the SPC and the Persia corpora.

5.1. Baseline Results

Baseline results were measured using all the acoustic features (baseline features) for both the SPC and the Persia corpora. The SPC corpus was evaluated using the SPC dev set and we obtained baseline results using baseline features with RF, SMO and Ada as shown in Table 1. We estimated the performance of the SPC dev set by using LSGO cross validation on the SPC train set.

For the evaluation of the Persia corpus we used micro-averaged LOSO cross validation. Table 2 shows the results using baseline features with RF, SMO, Adaboost (Ada) and Chance [9]. Chance (%) is the performance computed by randomly drawing labels using the prior distribution, more details can be found in [9].

Class	UA-RF	UA-SMO	UA-Ada
O	58.5	60.4	60.5
C	71.6	71.6	72.2
E	81.9	82.0	78.7
A	65.8	66.3	59.0
N	68.7	68.7	62.7
Mean	**69.3**	**69.8**	**66.6**

Table 1: Baseline results on the SPC dev set using baseline features with RF, SMO and Ada.

Class	UA-RF	UA-SMO	UA-Ada	Chance %
O	44.5	45.5	26.6	53.0
C	54.5	52.1	73.2	73.2
E	56.4	58.9	58.9	50.0
A	53.7	63.3	56.2	54.8
N	48.1	45.4	44.6	50.0
Mean	**51.4**	**53.0**	**51.9**	**56.2**

Table 2: Micro-averaged baseline results on the LOSO cross validation using baseline features of the Persia corpus. Chance (%) is the performance of the randomly drawing labels.

5.2. Feature Selection Results

After applying feature selection methods IG and relief on the SPC corpus we obtained improved results using relief feature selection with SMO. Table 3 shows the results on the SPC dev set using relief feature selection where we obtained better results with SMO. However, performance had been dropped in the agreeableness category.

Similarly, for the Persia corpus, we obtained improved results using relief feature selection with SMO as shown in Table 4. Though, after feature selection, performance had been reduced in conscientiousness category using RF and Adaboost, and in neuroticism category using Adaboost.

Class	Feat-RF	UA-RF	Feat-SMO	UA-SMO	Feat-Ada	UA-Ada
O	1200	61.2	1200	63.4	600	56.3
C	2200	73.2	2200	75.5	1000	66.0
E	1000	79.2	3200	84.2	1200	74.9
A	400	63.4	3800	65.4	800	56.4
N	400	65.6	1800	69.8	400	61.5
Mean		**68.5**		**71.6**		**63.0**

Table 3: Results on the SPC dev set with relief feature selection. Feat-* represents number of features selected for RF, SMO and Adaboost.

Class	Feat-RF	UA-RF	Feat-SMO	UA-SMO	Feat-Ada	UA-Ada
O	2200	47.2	2400	47.0	600	46.7
C	4800	50.6	800	74.6	1200	47.6
E	3600	64.8	200	64.7	200	58.8
A	1400	56.8	2400	71.8	200	69.7
N	3200	51.3	3000	54.6	1600	42.1
Mean		**54.1**		62.6		53.0

Table 4: Micro-averaged results on the LOSO cross validation using relief feature selection on the Persia corpus

5.3. Ensemble Methods

Table 5 shows the results of the SPC corpus with the ensemble of majority vote where classifier ensemble is formed by the best models of three classification algorithms: baseline features for RF, relief feature selection for SMO and baseline features for Adaboost. We used same models for stacking and obtained mean-UA: 69.0 for Big-5 traits.

	Our results		**Results [22]**
Class	**UA:** SPC train	**UA:** SPC-dev	**UA:** SPC-dev
O	52.5	65.2	67.0
C	67.2	75.3	73.2
E	70.8	83.0	80.9
A	57.8	66.0	69.0
N	67.7	69.2	71.0
Mean	63.2	71.7	72.2

Table 5: Results on the SPC dev set using the ensemble of the majority vote, which is comparable with [22]. With **UA:** SPC train is the mean results across the LSGO cross validation runs and for all traits we obtained 63.2±3.7 (mean±std).

With the Persia corpus, the ensembles (majority vote and stacking) of the three best models (relief feature selection with three different classifiers, Table 4) we obtained mean-UA: 56.4 with majority vote, and mean-UA: 49.1 with stacking. However, we obtained improved results with ensemble methods on extraversion (majority voting: UA-67.3) and agreeableness (stacking: UA-80.2) categories. The reason of poor performance is the higher correlation between lower performing classifiers (e.g. RF and Ada). Applying weighted majority voting could probably alleviate this problem, where proper weight needs to assign to the individual classifier.

The results of SPC on dev set are comparable with the results in [22], where our system performs better in conscientiousness and extraversion categories. However, overall, in five categories of OCEAN traits our results are close to their results in [22]. From the cross validation on the SPC-train set it is observed that our results are within statistical variation 63.2±3.7 (mean±std) in all traits. Another difference is that, in [22], they obtained their best results by considering the best models, and using the majority voting of all of their models they did not obtain better results compare to us. For the Persia corpus, the results in [9] showed the performance in terms of WA_{micro} where they obtained overall 57.5 and we obtained 64.4 with our best system (SMO with relief feature selection).

6. Conclusions

In this paper, we investigated automatic recognition of SPTs from speech using two different corpora – conversation and broadcast news. We studied different feature selection techniques such as IG and relief with different classification algorithms. It is observed that relief with SMO performs better than other models on both corpora and also relief feature selection performs well than IG. We obtained better results using majority voting ensemble method on the SPC corpus. Moreover, the stacking ensemble method did not perform well in any corpus with all personality traits categories. Future directions of this study include integrating linguistic information, understanding feature overlap in different feature selection algorithms and studying the contextual information.

7. References

[1] C. Nass and S. Brave, “Wired for Speech: How Voices Activates and Advances the Human-Computer Relationship”, MIT Press 2005

[2] T.W. Bickmore and R.W. Picard., “Establishing and maintaining long-term human-computer relationships”, ACM Trans. Comput.-Hum. Interact., 12:293–327, June 2005.

[3] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, “Using linguistic cues for the automatic recognition of personality in conversation and text”, J. Art. Intelligence Res., 30:457–500, 2007.

[4] Freund, Y. and Schapire, R. “A decision-theoretic generalization of online learning and an application to boosting”, Journal of Computer and System Sciences, 55: 119–139. 1997.

[5] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, B. Weiss, “The Interspeech 2012 Speaker Trait Challenge”, Proc. Interspeech 2012, ISCA, Portland, OR, USA, 2012.

[6] F. Eyben, M. Wöllmer, and B. Schuller, “openSMILE - The Munich Versa- tile and Fast Open-Source Audio Feature Extractor”, In Proc. ACM Multi- media. Florence, Italy, ACM, pp. 1459–1462, 2010.

[7] John C. Platt, “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines”, Microsoft Research, Technical Report, April 21, 1998.

[8] Leo Breiman, “Random Forests”, Statistics Department, University of California Berkeley, CA 94720, January 2001.

[9] A. V. Ivanov, G. Riccardi, A. J. Spork and J. Franc. 2012. Recognition of Personality Traits from Human Spoken Conversations. In: Proceedings of the 12th Annual Conference of ISCA. p. 1549-1552, 2011.

[10] Björn Schuller, “Voice and Speech Analysis in Search of States and Traits”. In Computer Analysis of Human Behavior, Albert Ali Salah, Theo Gevers (eds.), Advances in Pattern Recognition, Springer, pp. 227-253, 2011.

[11] A. J. Gill and R. M. French. “Level of representation and semantic distance: Rating author personality from texts”, In Proc. of the 2nd European Cognitive Science Conference (EuroCogsci07), 2007.

[12] J. Oberlander and A. J. Gill., “Individual differences and implicit language: Personality, parts-of-speech and pervasiveness”, In Proc. of the 26th Annual Conference of the Cognitive Science Society, Chicago, IL, USA, 2004.

[13] Zhi-Hua Zhou, “Ensemble Methods – Foundation and algorithms”, CRC Press, 2012.

[14] Rammstedt, B. & John, O. P., “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. 41, pp. 203-212, 2007.

[15] Yiming Yang, Jan O. Pedersen, “A Comparative Study on Feature Selection in Text Categorization”, In Proceeding ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning, pp. 412-420, 1997.

[16] Hunt E., Martin J & Stone P. “Experiment in Induction”, New york, Academic Press, 1966.

[17] G. Matthews, I. Deary, and M. Whiteman, “Personality traits”, Cambridge University Press, 2003.

[18] E. Sapir, “Speech as a Personality Traits”, American Journal of Sociology, vol. 32, no. 6, pp. 892-905, 1927. Online: http://www.brocku.ca/MeadProject/Sapir/Sapir_1927_b.html, accessed on 5 March 2008.

[19] Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, January 2011.

[20] Igor Kononenko, “Estimating Attributes: Analysis and Extensions of RELIEF”, In European Conference on Machine Learning, 171-182, 1994.

[21] Kartik Audhkhasi, Angeliki Metallinou, Ming Li and Shrikanth S. Narayanan, “Speaker Personality Classification Using Systems Based on Acoustic-Lexical Cues and an Optimal Tree-Structured Bayesian Network”, In Proceedings of Interspeech, 2012.

[22] Clément Chastagnol, Laurence Devillers, “Personality traits detection using a parallelized modified SFFS algorithm”, In Proceedings of Interspeech, 2012.

[23] Jouni Pohjalainen, Serdar Kadioglu, Okko Räsänen, “Feature Selection for Speaker Traits”, In Proceedings of Interspeech, 2012.

[24] Johannes Wagner, Florian Lingenfelser, Elisabeth André, A., “Frame Pruning Approach for Paralinguistic Recognition Tasks” In Proceedings of Interspeech, 2012.

[25] Dongrui Wu. “Genetic Algorithm Based Feature Selection for Speaker Trait Classification”, In Proceedings of Interspeech, 2012.

[26] Michelle Hewlett Sanchez, Aaron Lawson, Dimitra Vergyri, Harry Bratt, “Multi-System Fusion of Extended Context Prosodic and Cepstral Features for Paralinguistic Speaker Trait Classification”, In Proceedings of Interspeech, 2012.

[27] Harm Buisman and Eric Postma, “The log-Gabor method: speech classification using spectrogram image analysis”, In Proceedings of Interspeech, 2012.

[28] Yazid Attabi and Pierre Dumouchel, “Anchor Models and WCCN Normalization For Speaker Trait Classification’, In Proceedings of Interspeech, 2012.

[29] Claude Montacié and Marie-José Caraty, “Pitch and Intonation Contribution to Speakers’ Traits Classification”, In Proceedings of Interspeech, 2012.

[30] Björn Schuller, Ronald Müller, Manfred Lang, Gerhard Rigoll, “Speaker Independent Emotion Recognition by Early Fusion of Acoustic and Linguistic Features within Ensembles”, Proc. Interspeech 2005, Special Session: Emotional Speech Analysis and Synthesis: Towards a Multimodal Approach, Lisbon, Portugal, pp. 805-809, 2005.