This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2016 conference.

Speaker Trait Prediction for Automatic Personality Perception using Frequency Domain Linear Prediction features

S. Jothilakshmi and R. Brindha
Department of Computer Science and Engineering
Annamalai University, Annamalainagar
Tamilnadu, India.
Email: jothi.sekar@gmail.com and brinzram@gmail.com

***Abstract*—The aim of automatic personality perception is to predict the personality of the speaker perceived by the listener from nonverbal behavior. Extroversion, Conscientiousness, Agreeableness, Neuroticism and Openness are the speaker traits used for personality evaluation. In this work, a speaker trait prediction approach for automatic personality assessment is proposed to model the relationship between speech signal and personality traits using frequency domain linear prediction (FDLP) technique. Among several feature extraction techniques, FDLP features render increased performance. SSPNet Speaker Personality Corpus is used for experiments and evaluation. The proposed method predicts the speaker traits with 90-99% classification accuracy.**

***Index Terms*—Personality perception, Nonverbal behavior, Speaker traits, Frequency domain linear prediction.**

I. Introduction

Individuals are uniquely characterized by their personality. Trait theory states that personality of an individual are often outlined as a collection of habitual patterns of comportment, temperament, etc., which are stated using measurable traits. Several pioneering researches aimed towards integrating personality psychology in Human-Computer Interaction [1]. Several techniques measured the traits that has a role in technology and showed that personality has a strong bond in computing [2], [3], [4], [5], [6], [7].

Most of the personality computing approaches address solely three elementary issues: Automatic Personality Recognition (APR), Automatic Personality Perception (APP) and Automatic Personality Synthesis (APS). The aim of APR is to acknowledge actual personality of an individual whereas APP predicts the personality perceived by the observer. APS is to come up with artificial personalities. With the help of written scripts, para-language, and information from mobile or wearable devices and online games, APR approaches recognize personality. APP approaches focus principally on para-language particularly in speech and social media. Speech conveys an excellent deal of knowledge regarding the speaker additionally to their linguistic content.

Personality is usually assessed on five dimensions referred to as the Big Five [8]:

- Extroversion: Perspective towards outside like active, assertive, energetic, outgoing, talkative, etc.
- Agreeableness: Ability of social reflection and trust like benevolent, tolerant, innocent, compassionate, liberal, grateful, etc.
- Conscientiousness: Quality of diligence like competent, systematized, trustworthy, liable, etc.
- Neuroticism: Emotional stability like miserable, nervous, stressed, sensitive, unbalanced, distressing, etc.
- Openness: Attitude towards new experiences in everyday life like creative, interested, visionary, shrewd, etc.

In the last decade, solely few computing researches were dedicated to personality. The Speaker trait Challenge [9] broadens the scope of speaker traits within the procedural analysis of personality test. During this challenge, with the help of out sized range of options and machine learning techniques, completely different approaches are projected and also the comparisons are created with same experiments and corpus.

In existing approaches to predict the personality numerous machine learning approaches are used like support vector machine, Gaussian mixture model, artificial neural networks, decision tree, nearest neighbor, Naive Bayes and Adaboost. From the literature, it is determined that several enhancements ought to be created at numerous levels like information creation, machine learning techniques, feature set and methodology. Therefore to reinforce machine learning techniques, many tries are required to relate the speech signal and personality traits additionally to incorporate computing in human sciences.

In this work, a speaker trait prediction approach for automatic personality assessment is developed to model the speech signal and personality traits using frequency domain linear prediction features. These features provide improved performance than different feature extraction techniques.

Further sections are structured as follows: II Speaker personality Corpus, III Frequency Domain Linear Prediction methodology, IV Experiments and classification results, V Conclusion.

978-1-4673-9338-6/16/$31.00 ©2016 IEEE

II. Speaker Trait Prediction Database

In INTERSPEECH 2012 Speaker Trait Challenge, SSPNet Speaker Personality Corpus [10] (SPC) was used for examination and evaluation. It is used in this work also. The corpus contains audio clips randomly extracted from French news bulletins broadcasted by Radio Suissee Romande during February 2005. Only one speaker is involved per clip in order to elude conversational effects in the analysis. The personality trait assessment may be influenced by words that can be easily recognized which help the hearers who do not speak French. Hence such words are avoided in the speech clips. Professional and nonprofessional categories of speakers are involved in the audio clips. In this event, journalists are the professionals who work for the radio and talk regularly, and common people are the nonprofessionals. The details of the database are given in Table I.

TABLE I
Details of the Database

Total number of clips	640
Length of clips	593 clips of 10 seconds 47 clips of less than 10 seconds
Number of speakers	322 61% of the speakers talked in only one clip. 20.2% of the speakers talked in two. The remaining speakers talked in more than two.
Nature of speakers	307 speakers are professional 333 speakers are nonprofessional

The Big Five (BF) speaker traits, Extroversion, Conscientiousness, Agreeableness, Neuroticism and Openness, are considered in the database for personality assessment. In psychology, BFs are usually measured by using Big Five Inventory 10 (BFI-10). It is a set of ten questionnaires to assess the personality traits by providing scores. Ten dimensional questionnaire used in the database is shown in Table II.

TABLE II
BFI-10 Questionnaire

Q_1. This person is reserved
Q_2. This person is generally trusting
Q_3. This person tends to be lazy
Q_4. This person is relaxed, handles stress well
Q_5. This person has few artistic interests
Q_6. This person is outgoing, sociable
Q_7. This person tends to find fault with others
Q_8. This person does a thorough job
Q_9. This person gets nervous easily
Q_{10}. This person has an active imagination

BFI questionnaire was filled by 11 judges using online system. The score for each question is from 1 to 5 corresponding to the answers ranging from 'strongly disagree' to 'strongly agree'. The constraints followed to fill up the questionnaire are:

- The judges were not known to each other and they could not understand French.

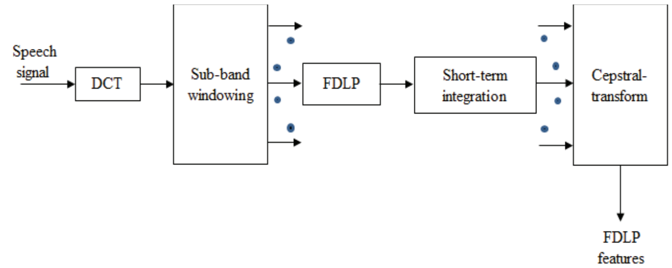

Fig. 1. Frequency Domain Linear Prediction short-term (FDLP-S) feature extraction schema.

- The assessment was performed in different places and at different moments.
- Each judge filled the score immediately after listening a clip before moving on to the next clip.
- It was not possible to edit scores once it was completed.
- For each judge the audio clips were presented in a random order.
- The judges were not permitted to work more than 60 minutes per day with two 30 minutes continuous sessions.

BFI-10 scores given by the judges are used to obtain the personality scores for each trait as given in Table III. For each clip and for all the judges, the procedure is repeated and finally, 11 set of scores are acquired.

TABLE III
Personality Scores

Extroversion = $S_6 - S_1$
Agreeableness = $S_2 - S_7$
Conscientiousness=$S_8 - S_3$
Neuroticism=$S_9 - S_4$
Openness=$S_{10} - S_5$
S_1 to S_{10} are the scores corresponding to Q_1 to Q_{10}

III. Frequency domain linear prediction (FDLP)

In speech analysis the spectral structures convey important linguistic information. Usually the short time spectral analysis is performed over windows of 10 to 30 ms. Vital cues in the temporal structure of these segments is for the perception of natural sounds and also for the understanding of stop bursts in speech. At the opposite extreme, the gross temporal distribution of acoustic energy in windows of upto one sec. has verified to be a thriving domain for the recognition of the whole phonemes and also for the description of their dynamics. Frequency domain linear prediction (FDLP) is a parametric description of the temporal dynamics of speech.

FDLP is the frequency domain dual of the standard time domain linear prediction (TDLP).

FDLP is computed in a simple way which has two parts:
1. On long time frames, DCT is applied.
2. Linear prediction is applied on the transformed DCT output.

For the purpose of feature extraction usually subband FDLP is used. In this case FDLP is applied on logarithmically-split octave bands, namely 0-0.5, 0.5-1, 1-2, and 2-4 kHz. The input signal energy in two dimensional representation is formed by the whole set of sub-band temporal envelopes. Thus short term cepstral features are obtained by converting these energies. For speaker trait prediction, the FDLP contours of various speaker traits are analyzed by selecting the audio clips with high score in that particular category. A one sec. long speech signal, whose sampling rate is 8 khz, is selected and DCT of the whole sample is taken. A single FDLP polynomial has been used to extract the temporal envelope of the signal. The FDLP contour of the trait Extroversion is displayed in Figures 2, 3.

FDLP is calculated as follows:
1. The entire signal is transformed using DCT.
2. The obtained DCT coefficients are divided into sub-band DCT using sub-band windowing.
3. FDLP envelope is formed by applying linear prediction on the spectral auto-correlations which is gained by the application of DFT on the hilbert envelope.
4. The sub-band FDLP envelopes are integrated in short-term windows.
5. The short-term energies are converted into cepstral co-efficients using cepstral transform which are called as FDLP short-term features.

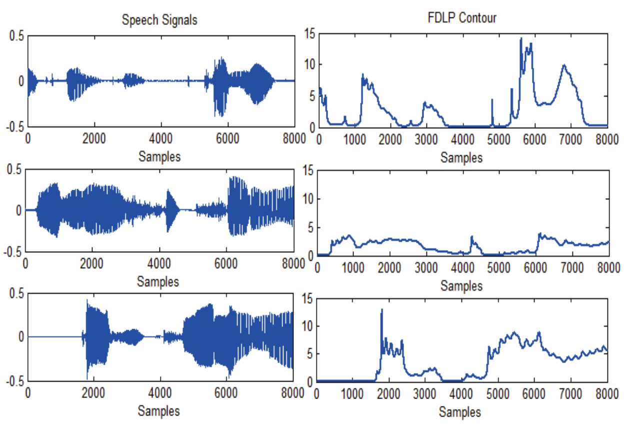

Fig. 2. FDLP contour - Extroversion high.

IV. Experiments and Results

This section describes the experiments that are done to automatically predict whether an audio clip corresponds to High or Low category for each speaker trait and the results are obtained. From the Speaker Personality Corpus personality scores of 640 audio clips given by 11 judges for each trait are

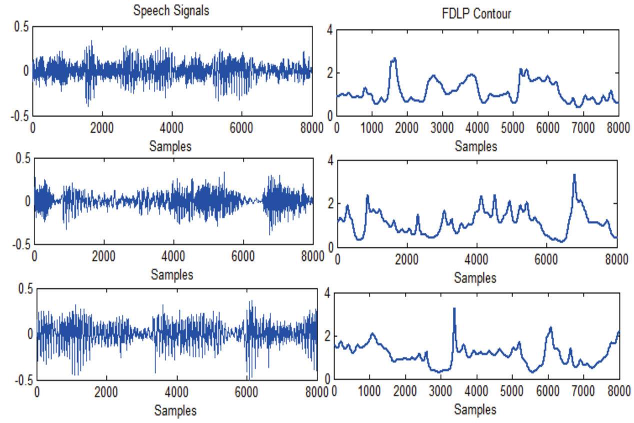

Fig. 3. FDLP contour - Extroversion low.

obtained. These scores are used to label each clip as High or Low for each trait. For each trait, the average score of six judges should be high in order to label a clip to be high; otherwise it is labeled as Low. Based on this procedure each clip is labeled as High or Low for each trait. Finally the corpus is segregated into two categories namely high and low for each trait.

FDLP features are extracted as described in Section III. Pattern classification techniques like support vector machine (SVM) with polynomial kernel, multilayer perceptron (MLP) and k-nearest neighbor (k-NN) with different values of k are applied on the extracted features. These techniques are chosen based on their diverse properties and are implemented with default and additional parameters using WEKA (Waikato Environment for Knowledge based Analysis) tool [11].

For experimental purpose, training and testing groups are created by partitioning the data with help of k-fold cross validation method. For all ten categories audio clips are classified to be k equal subgroups. For the purpose of training, k-1 subgroups are used and for testing, the remaining one is used such that every audio clip in the corpus is tested at least once and also to differentiate the training and testing groups. In this work 10 folds are used. Classification accuracy is used for computing the performance measure.

A. Classification results

Performance of the chosen classification techniques on speaker trait prediction are calculated. In k-NN technique, various values for k are applied and performances are listed in Table IV. k value is chosen to be odd to avoid the tie problem in choosing the nearest neighbor. In Table V, results of SVM, MLP and high accuracy rate of k-NN (from Table IV) are listed. The classification accuracy for both SVM and MLP are in the range 63-75 percent whereas k-NN provides better performance which ranges between 90 and 99 percent.

V. Conclusion

In this work, the experiments are conducted for inferring the personality traits. The database, SSPNet speaker personality

TABLE IV
PERFORMANCE OF k-NN (IN %) FOR BF TRAITS (VARIOUS NEAREST NEIGHBORS)

Name of the trait	$k(1)$	$k(3)$	$k(5)$	$k(7)$
Extroversion	98.06	93.76	86.42	82.95
Agreeableness	97.83	93.31	85.84	81.75
Conscientiousness	95.42	91.46	83.66	80.55
Neuroticism	98.37	93.89	85.99	81.77
Openness	97.41	92.6	84.51	79.85

TABLE V
PERFORMANCE OF SVM, MLP, k-NN

Name of the trait	SVM (%)	MLP (%)	k-NN (%)
Extroversion	73.05	71.81	98.06
Agreeableness	70.61	68.77	97.83
Conscientiousness	72.13	63.45	95.22
Neuroticism	63.14	68.25	98.37
Openness	68.33	63.74	92.66

corpus, is used for experiments since it has huge number of speech clips and speakers. Several pattern classification techniques are applied and performances are measured using classification accuracy.

Performances of the three classifiers are compared and it is noticed that k-NN achieves better accuracy than SVM and MLP. Also, as a replacement for the standard parameters (mel-frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients (LPCC) and perceptual linear prediction cepstral coefficients (PLP)), frequency domain linear prediction (FDLP) features can be used because it captures the amplitude fluctuations of a speech signal.

Each trait has its own characteristics. So it is not good to use same features to distinguish various speaker traits. Thus, the future research will be in the direction to analyze various features to identify a better feature for each speaker trait. Hence, the experiments will be conducted with multiple features.

REFERENCES

[1] C. Nass and K. M. Lee, "Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction and consistency attraction," *Experimental psycology applied*, vol. 7, no. 3, pp. 171–181, 2001.

[2] S. Love and J. Kewley, "Does personality affect peoples attitude towards mobile phone use in public places? in mobile communications," *ser. Computer Supported Cooperative Work, Springer*, vol. 51, pp. 273–284, 2005.

[3] S. Butt and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior*, vol. 24, no. 2, pp. 346–360, 2008.

[4] R. Guadagno, B. Okdie, and C. Eno, "Who blogs? personality predictors of blogging," *Computers in Human Behavior*, vol. 24, no. 5, pp. 1993–2004, 2008.

[5] D. Johnson and J. Gardner, "Personality, motivation and video games," in *Proc. Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 2010, pp. 276–279.

[6] L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on twitter," *Journal of Research in Personality*, vol. 46, no. 6, pp. 710–718, 2012.

[7] D. Quercia, D. L. Casas, J. Pesce, D. Stillwell, M. Kosinski, V. Almeida, and J. Crowcroft, "Facebook and privacy: The balancing act of personality, gender, and relationship currency," in *Proc. AAAI International Conference on Weblogs and Social Media*, 2012, pp. 306–313.

[8] G. Matthews, I. Deary, and M. Whiteman, *Personality traits*. Cambridge Univ. Presss, 2003.

[9] B. Schuller, "(2012) the interspeech 2012 speaker trait challenge," [Online]. Available: http://emotion-research.net/sigs/speechsig/is12-speaker-trait-challenge.

[10] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic fetures," *IEEE Trans. Affective Computing*, vol. 3, no. 3, pp. 273–284, 2012.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/