Perception of Speaker Personality Traits Using Speech Signals

Leilani H Gilpin
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
lgilpin@mit.edu

Danielle M Olson
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
dolson@mit.edu

Tarfah Alrashed
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
tarfah@mit.edu

CHI'18 Extended Abstracts, *April 21–26, 2018, Montreal, QC, Canada*

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188557

Abstract
As conversational agents continue to replace humans in consumer contexts, voice interfaces must reflect the complexity of real-world human interaction to foster long-term customer relationships. Perceiving the personality traits of others based on the way they look or sound is a key aspect of how humans unconsciously adapt their communication with others. In an effort to model this complex human process for eventual application to conversational agents, this paper presents the results of (1) building SVM and HMM classifiers for perceived personality prediction using speech signals using a data corpus of 640 speech signals based on 11 Big Five personality assessments, (2) determining correlations between feature and speaker subgroups, and (3) assessing the SVM classifier performance on new speech signals collected and assessed through a user study. This work is a small step towards the greater goal of designing more emotionally intelligent conversational interfaces.

Author Keywords
Personality traits; social computing; voice input.

ACM Classification Keywords
H.5.2 User Interfaces: Voice I/O

	MFCCs	Pitch	Female	Male
1	X		X	X
2		X	X	X
3	X	X	X	
4	X		X	
5		X	X	
6	X	X		X
7	X			X
8		X		X
9	X		X	X
10		X	X	X
11	X		X	X
12		X	X	X

Table 1. Some of the data and feature subsets used to train/test SVMs (continued onto Table 2)

Introduction

Given the widespread application of speech recognition technology in consumer contexts (e.g.: call center automated attendants and customer care agents), conversational interfaces must be designed in a natural and user-centric way to ensure customer satisfaction. Speech recognition advances have typically been applied to conversational agents to improve dictation, translation, and directory access. This paper presents an approach to apply these advances to enable personalization of voice-based conversational interfaces and agents using the perceived personality traits of its users, as a small step towards the larger goal better personalizing and humanizing conversational agents.

In 2012, Gelareh Mohammadi and Alessandro Vinciarelli [5] proposed a computational approach to model the complex social phenomenon of how humans perceive a speaker's personality traits (grounded in Big 5 Personality theory, which focuses on: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) when listening to their voice for the first time. This work is a result of 3 major goals which build upon their work: (1) building SVM and HMM classifiers for perceived personality prediction using speech signals, (2) determining correlations between feature and speaker subgroups, and (3) assessing the SVM classifier performance on new speech signals.

In this paper, we present an approach that can accurately predict perceived personalities using speech signals with a small set of features, and a relatively small data corpus. We motivate this idea using an SVM and HMM classifier, rather than a neural network, which needs orders of magnitude more data to produce a confident prediction.

Background and Related Work

Face, body, and speech in judgment of personality

In an early study, Ekman et al. [1] considered speech-related cues as well as other forms of nonverbal behavior (e.g., the amount of energy associated with body gestures) and face expressions to judge personality. They showed that the claim in the literature that the face is most important or that the nonverbal visual cues are more important than verbal cues have not been supported, and that it varies depending on the characteristics that one is trying to judge and the situation.

Automatic personality perception using speech signals

Some of the earliest approaches were proposed by Mairesse et al. [3,4], in which they considered both personality perception and personality recognition and use written data as well as speech samples for their experiments. Both psycholinguistic, like Linguistic Inquiry and Word Count (LIWC) or MRC, and prosodic features (average, minimum, maximum, and standard deviation of pitch, intensity, voiced time, speech rate) have been used, separately and in combination. The recognition is performed using different statistical approaches. The results show that it is possible to predict whether a person is perceived to be below or above average along the Big Five dimensions with an accuracy between 60 and 75 percent, depending on the trait and on the features used.

In a similar study that maps nonverbal vocal behavior into trait attributions, Mohammadi et al. [6], used statistical functions of the main prosodic features (pitch, energy, first two formants, length of voiced, and unvoiced segments) to predict whether a speaker is perceived as above or below average along each of the

	Journalists	Non - Journalists
1	X	X
2	X	X
3	X	X
4	X	X
5	X	X
6	X	X
7	X	X
8	X	X
9	X	
10	X	
11		X
12		X

Table 2. A continuation of the data and feature subsets used to train/test SVMs.

Big Five dimensions. The prediction was performed with Support Vector Machines (SVM) and the accuracies range between 60 and 75 percent depending on the trait. Polzahl et al. [7] conducted a personality assessment paradigm to speech input, and compared human and automatic performance on this task. They applied a total of 1450 features based on statistics of intensity, pitch, loudness, formants, spectral energy, and Mel Frequency Cepstral Coefficients. These are first submitted to a feature selection approach and then fed to Support Vector Machines (SVM) to recognize 10 different personality types acted by the same speaker, and the recognition rate was 60 percent. Golbeck et al. [2] conducted a similar small data study, and developed a method to predict user's personality through the publicly available information on their Facebook profile. Although, Golbeck uses a similarly small set of samples and features, they use regression as their method while we use classifiers like HMM and SVM.

Methods

Build SVM and HMM classifiers for perceived personality prediction using speech signals

Building the SVM classifier includes 3 mains steps: extracting low-level features from the speech clips using the Kaldi speech recognition toolkit [8], processing the extracted data into an appropriate format using various scripts and toolkits, and finally, training and testing each of the SVM classifiers for each of the Big 5 personality traits using MATLAB.

Building the HMM classifier includes 4 main steps: extracting low-level features from the speech clips using the Kaldi[1] speech recognition toolkit [8], creating the HMM network topology using Python, processing the extracted data into an appropriate format using various scripts, and finally training and testing each of the HMM classifiers for each of the Big 5 personality traits using Python.

Determine correlations between features, speaker subgroups, and personality accuracy

Determining the correlations between features, speaker subgroups, and personality prediction accuracy involved subdividing the data into the appropriate speaker subgroups, and then repeating the steps for building the SVM classifier and modifying which features are extracted as appropriate. The following feature and speaker subgroups were used to train and test 12 SVMs for each of the Big 5 personality traits (resulting in 60 total SVMs), which are shown in Table 1 and Table 2.

Survey

Finally, to assess the accuracy of the SVM classifiers on a small data set (which was trained and tested on all 640 speech clips), we assessed the performance on new speech clips includes 4 main steps: recording 15 new speech clips (3 unique speakers; 5 speech clips each), recruiting 12 assessors to listen to the speech clips and fill out an online BFI-10 [9] about each of the 3 speakers' personalities, calculating the perceived Big 5 personality traits from the collected data, and running the SVM classifier on the new speech clips to compare the output personality classification to that of the assessors.

[1] http://kaldi-asr.org

SVM Classification Accuracy		HMM Classification Accuracy	
MFCCs & Energy	Pitch	MFCCs & Energy	Pitch
64.53%	78.83%	64.07%	63.29%
90.78%	90.78%	92.19%	93.75%
70.16%	70%	65.62%	67.19%
66.72%	65.16%	74.22%	66.41%
77.03%	77.65%	82.81%	79.69%

Table 3. Accuracies observed for the SVM and HMM classifiers. Rows correspond to (in order) openness, conscientiousness, extraversion, agreeableness and neuroticism.

MFCCs and Energy Only	Pitch Only
64.53%	78.83%
90.78%	90.78%
70.15%	70%
66.72%	65.15%
77.03%	77.66%

Table 4: The accuracies observed for each of the feature subgroups. Rows correspond to (in order) openness, conscientiousness, extraversion, agreeableness and neuroticism.

Experimental dataset

Dataset for Experiments

The SSPNet Speaker Personality Corpus [5] was used for training and testing the SVMs and HMMs. The Social Signal Processing Network (SSPNet) Speaker Personality Corpus contains 640 speech clips (from 322 unique speakers) mapped to the Big 5 personality traits determined by 11 assessors. The corpus also includes the raw personality questionnaires and the overall personality scores and metadata associated with each clip: speaker gender, speaker status (journalist or non-journalist) and speaker ID. The label of each speaker's journalist or non-journalist status is a feature of the data corpus used.

Dataset for User Study

Three unique speakers (1 male, 2 females; all non-journalists) were recruited and asked to record 5 speech clips reading aloud 5 distinct news articles. 12 assessors were recruited to listen to the 15 speech clips for and assess personalities of the 3 unique speakers using the BFI-10 [9], the results for which were mapped to the Big 5 personality scores.

Statistical Significance

The SVM showed a significant effect with a p-value of 0.031, and similarly, the HMM had a similar effect with a p-value of 0.027. Therefore, the classifiers reject the null hypothesis by providing more accurate classification than achieved by randomized selection.

Experiments

Build SVM and HMM classifiers for perceived personality prediction using speech signals

To build the SVM classifiers, Kaldi was used to convert each speech clip from the SSPNet Speaker Personality Corpus into the following 2 sequences of frame feature vectors (sampled at a rate of 8 kHz) including the Mel-frequency cepstral coefficients (MFCCs), energy, and pitch. For each of the feature sequences, 5 SVM classifiers were trained and tested in MATLAB for each of the Big 5 personality traits, resulting in 10 new classifiers. Cross-validation was performed on each of the SVM models and the out-of-sample misclassification rate was estimated to determine the class loss for each trait.

To build the HMM classifiers, Kaldi was used to convert each speech clip from the SSPNet Speaker Personality Corpus into a sequence of frame feature vectors (sampled at a rate of 8 kHz) containing the Mel-frequency cepstral coefficients (MFCCs), energy, and pitch for each frame. Various scripts were used to process the data into an appropriate format. An HMM instance was built for each of the Big 5 personality traits with a 12-node, strongly-connected network topology.

Determine correlations between features, speaker subgroups, and personality prediction

Four new .scp files were created corresponding to each of the 4 speaker subgroups (*i.e.: females only, males only, journalists only, non-journalists only*) and used to extracted from the metadata provided in the SSPNet Speaker Personality Corpus. Kaldi was used with the new .scp files to convert each speech clip within each of the 4 speaker subgroups into different sequences of frame feature vectors (sampled at a rate of 8 kHz) including the Mel-frequency cepstral coefficients (MFCCs), energy, and pitch. For each of these features and speaker subgroups, 5 new SVM classifiers were trained and tested for each of the 5 personality traits,

All Data	Females Only	Males Only
78.83%	77.37%	61.43%
90.78%	96.35%	89.26%
70.15%	72.26%	71.17%
66.72%	78.83%	65.61%
77.66%	69.34%	79.32%

Table 5: The accuracies observed for each of the gender subgroups. Rows correspond to (in order) openness, conscientiousness, extraversion, agreeableness and neuroticism.

All Data	Journalists Only	Non-Journalists Only
78.83%	75.57%	56.76%
90.78%	98.69%	83.48%
70.15%	91.53%	62.16%
66.72%	63.84%	70.87%
77.66%	76.87%	78.67%

Table 6: The accuracies observed for each of the professional subgroups. Rows correspond to (in order) openness, conscientiousness, extraversion, agreeableness and neuroticism.

resulting in 50 new classifiers. Cross-validation was performed on each of the SVM models and the out-of-sample misclassification rate was estimated to determine the class loss for each trait within each feature and speaker subgroup.

Assess classifier performance with a short user study
Kaldi was used to convert each of the 15 new speech clips into two sequences of frame feature vectors (sampled at a rate of 8 kHz), including the Mel-frequency cepstral coefficients (MFCCs), energy, and pitch.

These feature vectors were loaded into MATLAB and truncated to match the dimensions of the training matrices. The first sequence of features was provided as input to be classified [18] by the 5 SVM classifiers trained on only the MFCCs and energy features from the entire SSPnet Speaker Personality Corpus. The second sequence of features was provided as input to be classified by the 5 SVM classifiers trained on only the pitch features from the entire corpus. The classifier results of "*high*" or "*low*" for each of the Big 5 personality traits on all of the new speech clips were compared to the perceived personality trait results obtained from the user study to calculate overall accuracy.

Results and Analysis

Build SVM and HMM classifiers for perceived personality prediction using speech signals
For the SVM and HMM classifiers corresponding to each of the Big 5 personality traits, the following accuracies were observed in Tables 3-6.

Within the feature subgroups, it appears that for *openness* and *neuroticism*, extracting pitch features resulted in higher accuracies, which may suggest that voice pitch is most salient to listeners when assessing these specific traits. Within the gender subgroups, a gain in accuracy was observed across all Big 5 personality traits except for *openness*. 3 out of 4 of these gains in accuracy were gender-specific:

- prediction accuracy for *conscientiousness* and *agreeableness* improved only for the ***female*** subgroup
- prediction accuracy for *neuroticism* improved only for the ***male*** subgroup

Predicting *extraversion* improved for both the ***female*** and ***male*** subgroups. Finally, within the professional subgroups, the highest gain in prediction accuracy was observed for *extraversion*, which improved by about 21% for the ***journalist*** subgroup.

Survey
Both classifiers performed well on 3 out of 5 personality traits for both Speakers 1 and 2, with the most accurately predicted traits being *conscientiousness* and *agreeableness*. However, the classifiers did not perform well for Speaker 3. We attribute this discrepancy to the presence of noise on the speech clips associated with Speaker 3 (*i.e.: "clicking" noises, background noise, etc*). Overall, the SVM classifiers trained on MFCC and energy features produced slightly more accurate results across all traits and speakers compared to the SVM classifiers trained on pitch features.

Conclusion

This paper has presented methods and experiments demonstrating that it is possible to predict with high

accuracy whether a person is perceived to have high/low levels of each of the Big 5 personality traits using a relatively small amount of data and features. Of all 5 traits, predicting a speaker's perceived *conscientiousness*, *openness*, and *neuroticism* using SVM classifiers resulted in the highest accuracies. The results in this paper also demonstrate that training SVM classifiers on certain feature and speaker subgroups may result in higher accuracies for specific traits, even if the data set is relatively small (i.e.: predicting *extraversion* using classifiers trained only on the ***male***, ***female***, or ***journalist*** subgroups resulted in higher accuracies than all of these groups combined). Given the small size of the dataset used, future work should investigate larger datasets to determine the statistical significance of these differences and understand the interrelationship between speech and identity. Finally, the user study verified that these small data samples and features can accurately predict perceived personality traits. Our methods demonstrated that the SVM classifiers assessed 3 out of 5 personality traits from new speech clips with high accuracy, with the most accurately predicted traits being *conscientiousness* and *agreeableness.*

References

1. Paul Ekman, Wallace V. Friesen, Maureen O'Sullivan, and Klaus Scherer. 1980. "Relative Importance of Face, Body, and Speech in Judgments of Personality and Affect," *Journal of Personality and Social Psychology*, vol. 38, no. 2, pp. 270-277, 1980.
2. Jennifer Gobeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI '11 extended abstracts on human factors in computing systems*. ACM, 253-262
3. François Mairesse, Marilyn Walker, and others. 2006. Words mark the nerds: Computational models of personality recognition through language. In *Proceedings of the Cognitive Science Society*, Vol. 28
4. François Mairesse, Mariln A Walker, Matthias R Mehl and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research* 30 (2007) 457-500.
5. Gelareh Mohammadi and Alessandro Vinciarelli. 2012. Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing* 3, 3 (2012), 273-284.
6. Gelareh Mohammadi, Alessandro Vinciarelli, and Marcello Mortillaro. 2010. The voice of personality: Mapping nonverbal vocal behavior into trait attributions. In *Proceedings of the 2nd international workshop on Social signal processing.* ACM, 17-20,
7. Tim Polzeho, Sebastian Moller, and Florian Metze. 2010. Automatically assessing personality from speech. In *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on.* IEEE, 134-140.
8. Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glebek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, and others. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding.* IEEE Signal Processing Society.
9. Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute of less: A 10-item short version f the Big Five Inventory in English and German. *Journal of research in Personality* 41, 1 (2007), 203-212.