Words Mark the Nerds: Computational Models of Personality Recognition through Language

François Mairesse (F.Mairesse@sheffield.ac.uk)
Department of Computer Science, University of Sheffield
211 Portobello Street, Sheffield S1 4DP, United Kingdom

Marilyn Walker (M.A.Walker@sheffield.ac.uk)
Department of Computer Science, University of Sheffield
211 Portobello Street, Sheffield S1 4DP, United Kingdom

Abstract

User models in human-computer interfaces have focused on various user characteristics, but there is little user modeling of the most fundamental dimension of variation between humans: personality traits. We explore here the possibility of automatically acquiring such models by simply observing the user's language. We automatically learn models for personality recognition from different corpora and sources of personality evaluation. The models are completely transparent, i.e. they can run in the background evaluating the user at every conversational turn, and provide input for the system to adapt to the user. Results show that recognition models based on observed personality perform significantly better than a baseline of the average personality score, as well as better than models using self-reports. An analysis of decision trees confirms previous findings linking language and personality, while revealing many new linguistic markers.

Keywords: Personality; Statistical model; Machine learning; Language; Conversation; EAR; Big Five; Pragmatics; Five factor model; Dialogue systems

Introduction

Many scientists have focused on how to optimize our interactions with machines, by finding the best way to select and communicate information. However, because users behave in different ways, computer systems need to adapt to their conversational partner, as humans do (Funder & Sneed, 1993; McLarney-Vesotski et al., in press); this is the purpose of user models in dialogue systems (Zukerman & Litman, 2001). One approach to user modeling is to elicit user preferences by asking the user (Linden, Hanks & Lesh, 1997), but other work develops such models through observation, i.e. by analyzing the user's language. Rich (1979) successfully maps user's keywords to content selection and language generation parameters, while Maloor & Chai (2000) estimate the user's expertise using dialogue features. Researchers have also learned user models from data, e.g. for sentence planning (Mairesse & Walker, 2005) and dialogue management (Litman et al., 2000).

Rather than modeling specific user preferences, we propose modeling personality traits. By definition, personality is the highest level variable characterizing individuals. Computer users can detect a machine's personality and prefer working with a computer exhibiting the same personality as theirs (Reeves & Nass, 1996). Personalities of real estate agents were used as a basis for the design of an e-commerce website (Fuchs, 2001).

In addition, many studies have shown that the success of interpersonal tasks depends on the personalities of the different parties involved. Personality traits influence many aspects of individual behaviour, such as the attitude toward machines (Sigurdsson, 1991), overall job performance (Furnham, Jackson & Miller, 1999), as well as academic motivation (Komarraju & Karau, 2005); this last finding suggests that training systems would be more efficient if they could adapt to the learner's personality.

Our goal is to develop models to recognize the user's personality, and use them to modify the output generation of a dialogue system. This paper presents non-linear statistical models of the five most essential personality traits, learned automatically via machine learning on different sources of data. Results show that extraversion, emotional stability and conscientiousness are easier to model, and recognition models based on observed personality perform significantly better than a baseline returning the average personality score, as well as better than models using self-reports. An analysis of decision trees confirms previous findings linking language and personality, while revealing many new linguistic markers.

Personality traits and language

Trait identification

Personality can be described as a set of attributes characterizing an individual. The number of those attributes seems to be extremely large. Indeed, when talking about a close friend, one can usually come up with many adjectives describing his or her behaviour. To be able to reason about personality, psychologists have tried to identify the most essential traits, referred to as the *Big Five* (Norman, 1963):

- Extraversion (sociability, activity, assertiveness)
- Emotional stability (as opposed to neuroticism)
- Agreeableness to other people
- Conscientiousness (discipline)
- Openness to experience (intellect)

Such traits were obtained by doing a factor analysis over the adjectives used in personality description questionnaires. They are therefore based on the assumption that the most relevant individual differences are encoded into the language, and the more important the difference, the more likely it is to be expressed as a single word. This is referred to as the *Lexical Hypothesis*.

While there are known limits to the Big Five model (Eysenck, 1991; Paunonen & Jackson, 2000), we shall use it in our models, as it provides a very general framework for reasoning about individual differences, and it has become a standard in psychology as it was replicated many times (Norman, 1963; Peabody & Goldberg, 1989).

Personality markers in language

Many studies have identified cues associated with personality at different linguistic levels, including acoustic parameters (Smith et al. 1975), lexical categories (Pennebaker & King, 1999) and more complex phrases (Gill & Oberlander, 2002). The extraversion/introversion dimension has received the most attention as it is the most important one for discriminating between people (Peabody & Goldberg, 1989).

A review by Furnham (1990) describes linguistic features linked to extraversion, emotional stability and other traits, and Dewaele & Furnham (1999) review studies focusing on the link between extraversion and both language learning and speech production. Findings include that there is a higher correlation between extraversion and oral language, especially when the study involves a complex task. Extraverts talk more and repetitively, with fewer pauses and hesitations, have higher speech rates, shorter silences, higher verbal output and a lower type/token ratio, while introverts use a broader vocabulary. Extraverts also use more positive emotion words and informal style, and show more agreements and compliments than introverts. Extravert students learning French as a second language produce more back-channels, and have a more implicit style and a lower lexical richness in formal situations. It seems that the more complex the task and the higher the level of anxiety, the easier it is to differentiate between introverts and extraverts.

Heylighen & Dewaele (2002) also noted that extraversion is significantly correlated with contextuality, as opposed to formality. Contextuality can be seen a high reliance on shared knowledge between conversational partners, leading to the use of many deictic expressions such as pronouns, verbs, adverbs and interjections, whereas formal language is less ambiguous and assumes less common knowledge.

Scherer (1979) showed that extraverts are perceived as talking louder and with a more nasal voice, and that American extraverts tend to make fewer pauses, while German extraverts produce more pauses than introverts. Thus personality markers are culture-dependent, even among western societies.

Pennebaker & King (1999) identify many linguistic features associated with each of the Big Five personality traits. They use their Linguistic Inquiry and Word Count (LIWC) tool to count word categories of essays written by students whose personality has been assessed using a questionnaire. The authors find small but significant correlations between their linguistic dimensions and personality traits. Relevant word categories for extraversion include social words, emotion words, first person pronouns, and present tense verbs.

Gill & Oberlander (2002) used content analysis tools and n-gram language models to identify markers in extravert and introvert emails. They replicated previous findings and identified new personality markers such as first person singular pronouns (e.g. *I don't*) and formal greetings (e.g. *Hello*) for introversion, while less formal phrases as *Take care* and *Hi* characterize extraverts.

Mehl, Golsing & Pennebaker (in press) analyzed personality in its natural habitat, by using an Electronically Activated Recorder (EAR) to collect conversation extracts from the participants' daily life over 2 days. They transcribed each participant's utterances and annotated them with subjective information about the type of interaction, location, activity, mood, language use (LIWC categories) as well as the participant's personality. They found that all Big Five traits were predominantly expressed in the participants' language use, although they observed important gender differences. Interestingly, the correlation between the LIWC variables and personality is usually much larger than with students' essays. Moreover, observers were found to significantly agree with self-reports for all personality dimensions, with the largest effect size for extraversion ($r = 0.41$, $p < 0.01$).

Experimental method

While previous work identified correlations between linguistic markers and personality ratings, none provide recognition results on unseen data. Here we conduct a set of experiments to examine whether automatically trained non-linear models provide better fits to the data, and whether these models can be used to *recognize* the personality of unseen subjects. Our approach can be summarized in five steps:

1. Collect individual corpora;
2. Collect associated personality ratings for each participant;
3. Extract relevant features from the texts;
4. Build statistical models of the personality ratings based on the features;
5. Test the learned models on the linguistic outputs of unseen individuals.

The following sections describe each of these steps in more detail.

Sources of language and personality

The first corpus contains 2,479 essays from psychology students (1.9 million words), who were told to write whatever comes through their mind for 20 minutes. The data was collected and analyzed by Pennebaker & King (1999); a sample is shown in Table 1. Personality was assessed by asking each student to fill in the Five Factor Inventory questionnaire (John, Donahue & Kentle, 1991), which asks participants to evaluate how well their personality matches a series of descriptions.

The second source of data consists of conversation extracts recorded using an Electronically Activated Recorder (EAR), collected by Mehl, Golsing & Pennebaker (in press). To preserve the participants' privacy, only random bits of conversation were recorded. This corpus is much smaller than the previous one (96 participants for a total of 97,468 words and 15,269 utterances). Moreover, only the *participants'* utterances were transcribed, making it impossible to reconstruct whole conversations. Nevertheless, the conversation extracts are less formal than the essays, and as personality should be best observed in the absence of behavioral constraints, our hypothesis is that they have a larger potential for exhibiting personality cues. Table 1 shows examples of conversations for two participants judged as introvert and extravert, respectively.

For personality ratings, the EAR corpus contains both self-reports and ratings from 18 independent observers. As people's self assessments might differ from what is observed through their behaviour, our second hypothesis is that observers' ratings produce better personality models than self-reports.

Table 1: Extracts from the essays and EAR corpus, for participants rated as extremely introvert and extravert. Only the participants' utterances are shown.

Introvert	Extravert
Stream of consciousness essays corpus: I've been waking up on time so far. What has it been, 5 days? Dear me, I'll never keep it up, being such not a morning person and all. But maybe I'll adjust, or not. I want internet access in my room, I don't' have it yet, but I will on Wed??? I think. But that ain't soon enough, cause I got calculus homework [...]	I have some really random thoughts. I want the best things out of life. But I fear that I want too much! What if I fall flat on my face and don't amount to anything. But I feel like I was born to do BIG things on this earth. But who knows... There is this Persian party today. My neck hurts.
EAR corpus: - Yeah you would do kilograms. Yeah I see what you're saying. - On Tuesday I have class. I don't know. - I don't know. A16. Yeah, that is kind of cool. - I don't know. I just can't wait to be with you and not have to do this every night, you know? - Yeah. You don't know. Is there a bed in there? Well ok just...	- That's my first yogurt experience here. Really watery. Why? - Damn. New game. - Oh. - That's so rude. That. - Yeah, but he, they like each other. He likes her. - They are going to end up breaking up and he's going to be like.

Observers were asked to make their judgments by rating descriptions of the Big Five Inventory (John & Srivastava, 1999) on a 7 point scale (from *strongly disagree* to *strongly agree*), without knowing the participants. Observers were divided into three groups, each rating one third of the participants, after listening to each participant's entire sound file. Mehl et al. (in press) report strong inter-observer reliabilities across all Big Five dimensions (intraclass correlations based on one-way random effect models: mean $r = .84$, $p < .01$). For each participant's transcribed text, the observers' ratings were averaged, to produce the final scores used in our experiments.

Feature selection

We extracted a set of linguistic features from each essay and conversation transcript, starting with 88 word categories from the LIWC utility (Pennebaker & Francis, 2001). These features include both syntactic (e.g. ratio of pronouns) and semantic information (e.g. positive emotion words). We also added 14 additional features from the MRC Psycholinguistic database (Coltheart, 1981), which contains statistics for over 150,000 words, such as estimates of the age of acquisition, frequency of use, and familiarity. To find the correct word in the database among a set of homonyms, we pick the entry with the same stem and Part-of-Speech tag as the target word. We computed the MRC feature values as the average value over all the words that match an entry in the database. Table 2 shows examples of LIWC word categories and MRC scales.

We also introduced features characterizing the types of speech act produced. We automatically tagged each utterance of the EAR corpus with speech act categories from Walker & Whittaker (1990), using heuristic rules based on each utterance's parse tree:

- Command: utterance using the imperative form, a command verb (e.g. *must*, *have to*) or a yes/no second person question with a modal auxiliary like *can*;
- Prompt: single word utterance used for back-channeling (e.g. *Yeah*, *OK*, *Huh*, etc.);
- Question: interrogative utterance which isn't a command;
- Assertion: any other utterance

We evaluated the automatic tagger by applying it to a set of 100 hand-labeled utterances randomly selected in the EAR corpus. We obtain 88% of correct labels, which are mostly assertions. Table 3 summarizes the partition and the evaluation results for each speech act type. The feature value is the ratio of the number of speech acts to the total number of utterances in each text.

Table 2: Examples of LIWC word categories and MRC psycholinguistic features (Pennebaker & Francis, 2001; Coltheart, 1981). [1] indicates an MRC feature, which associates each word to a numerical value.

Feature	Example
Anger words	hate, kill, pissed
Metaphysical issues	God, heaven, coffin
Physical state/function	ache, breast, sleep
Inclusive words	with, and, include
Social processes	talk, us, friend
Family members	mom, brother, cousin
Past tense verbs	walked, were, had
References to friends	pal, buddy, coworker
Imagery of words[1]	Low: future, peace - High: table, car
Syllables per word[1]	Low: a - High: uncompromisingly
Concreteness[1]	Low: patience, candor - High: ship
Frequency of use[1]	Low: duly, nudity - High: he, the

Table 3: Partition of the speech acts automatically extracted from the EAR corpus, and classification accuracies on a sample of 100 hand-labeled utterances.

Label	Fraction	Labeling accuracy
Assertion	73.0%	0.95
Command	4.3%	0.50
Prompt	7.0%	0.57
Question	15.7%	1.00
All	100%	0.88

As personality influences speech production (Dewaele & Furnham, 2000; Scherer, 1979), we added prosodic features based on the audio data of the EAR conversation extracts. As the EAR recorded the participants at anytime of the day, it was necessary to remove any non-voiced signal. We used Praat (Boersma, 2001) to compute features characterizing the voice's pitch and intensity (mean, extremas and standard deviation), and we added an estimate of the speech rate by dividing the number of words by the voiced time.

We included all the features mentioned in this section (117) in the models based on the EAR corpus. Models computed using the essays corpus contain only LIWC and MRC features (102), as speech acts are only meaningful in dialogues.

Statistical model

Depending on the adaptation capabilities of the target dialogue system, we will need two different types of personality models. First, for the case where the output generation of the dialogue system can be varied continuously along particular dimensions, we develop regression models of personality dimensions as continuous variables. Second, for the case where the output generation is limited to a few points at extremes of a personality scale, such as introvert vs. extravert language, or neurotic vs. emotionally stable, we develop classification models by splitting our subjects into two equal-size groups.

We use the Weka toolbox (Witten & Frank, 2005) for training and evaluating the different statistical models. For each trait, we compare a baseline model returning the mean personality score with a linear regression model, an M5' regression tree returning a linear model, regular M5' and REP-Tree decision trees, and a model based on Support Vector Machines (SMO). In order to evaluate models of personality classification, we compare six different learning algorithms against a baseline returning the majority class.

Results

For each Big Five trait, we trained regression models using the self-reports of the essays data, and we computed two additional models based on the self-reports and observer ratings of the EAR corpus. Regression results are summarized in Table 4. The baseline is a model returning the mean of all personality scores in the training set. We use the relative error for evaluation, which is the ratio between the model's prediction error and the error produced by the baseline. All results are averaged over a 10 fold cross-validation.

Table 4: Relative error for regression models, with observer ratings (Obs) and self-reports (Self) of the EAR corpus, and self-reports of the essays corpus (Essays). Models are linear regression (LR); M5' regression tree with linear models (M5R); M5' decision tree with regular leaves (M5D); REP-Tree decision tree (REPT)[2].

Dataset	Trait	LR	M5R	M5D	REPT
EAR-Obs	Extra	186.88	76.80•	83.49•	89.55•
EAR-Obs	Emot	331.30	96.20	96.08•	102.84
EAR-Obs	Agree	264.99	105.16	100.06	109.48
EAR-Obs	Consc	207.37	92.07•	85.25•	97.41
EAR-Obs	Open	318.69	117.44	106.66	104.48
EAR-Self	Extra	214.73	109.96	104.35	101.40
EAR-Self	Emot	357.86	116.76	105.05	104.27
EAR-Self	Agree	330.13	110.62	103.94	105.80
EAR-Self	Consc	181.63	118.14	103.95	103.75
EAR-Self	Open	372.30	121.47	103.21	106.68
Essays	Extra	99.31•	99.42•		
Essays	Emot	97.25•	97.14•		
Essays	Agree	99.07•	99.03•		
Essays	Consc	98.78•	98.72•		
Essays	Open	93.81•	93.83•		

• statistically significant improvement over the mean value baseline (two-tailed paired t-test, $p < 0.05$)

Paired t-tests show that all regression models based on the essays corpus significantly improve over the baseline (two-tailed, $p < 0.05$), but the improvements are smaller than for spoken language, confirming the results from Mehl et al. (in press). Interestingly, modeling openness to experience produces the best results (6.2% error decrease).

Concerning the EAR corpus, we observe that linear regression performs poorly for all traits, suggesting a highly non-linear relationship between language and personality recognition. Regression tree models produce the best improvement over the baseline: a paired t-test (two-tailed, $p < 0.05$) over the cross-validation folds shows that the error reduction is significant for extraversion (76.8% relative error, i.e. 23.2% improvement), emotional stability (3.92% improvement), and conscientiousness (14.75% improvement). Regression trees for extraversion and conscientiousness are in Figures 1 and 2. On the other hand, self-reports of the EAR corpus are clearly difficult to model: none of the models show significant improvement over the baseline.

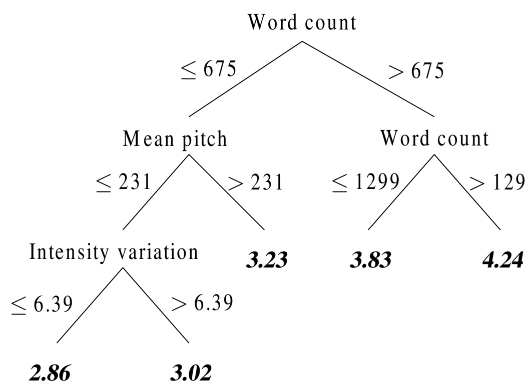

Figure 1: M5' regression tree for extraversion, computed using the EAR corpus. The target output ranges from 1 to 7, where 7 means strongly extravert. The mean pitch value is expressed in Hertz, and the intensity variation (standard deviation) in decibels.

Table 5: Classification accuracy with two equal-frequency bins on the EAR corpus, for observer ratings (Obs) and self-reports (Self). Models are J48 decision tree (J48); Nearest neighbour (NN); Naive Bayes (NB); JRip rules set (JRIP); AdaboostM1 (ADA); SMO support vector machine (SMO).

Data	Trait	J48	NN	NB	JRIP	ADA	SMO
Obs	Extra	67.26•	57.51	73.20•	64.86•	73.23•	65.48•
Obs	Emot	58.37	57.61	70.71•	59.00	58.47	62.79•
Obs	Agree	51.66	51.82	55.08	52.93	51.51	50.67
Obs	Consc	57.03	59.59•	65.68•	58.91•	60.13•	59.63•
Obs	Open	45.86	47.14	56.53	50.66	53.94	55.12
Self	Extra	46.87	47.34	57.48•	53.78	50.94	51.74
Self	Emot	47.72	47.90	50.28	50.46	48.51	45.82
Self	Agree	48.21	50.33	59.92	56.20	56.99	54.68
Self	Consc	46.39	44.72	42.16	46.68	47.08	53.47
Self	Open	52.84	44.07	64.54•	50.96	57.82	55.50

• statistically significant improvement over the majority class baseline (two-tailed paired t-test, $p < 0.05$)

[2] SVM regression models aren't included as they don't perform significantly better than the baseline. Only linear regression and M5' tree models were computed with the essays corpus, due to the large dataset size. The personality traits are extraversion (Extra), emotional stability (Emot), agreeableness (Agree), conscientiousness (Consc), and openness to experience (Open).

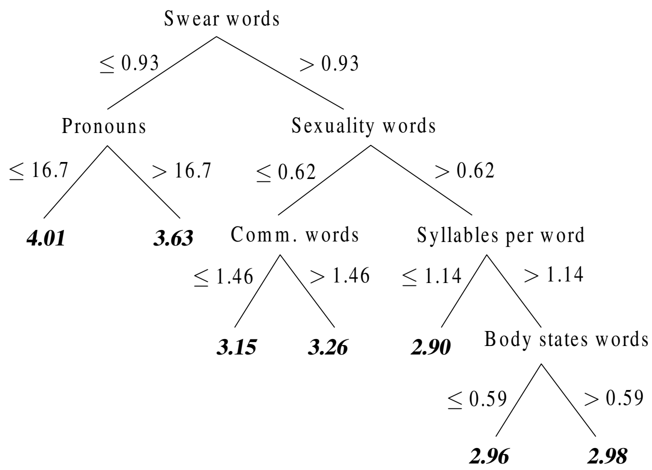

Figure 2: M5' regression tree for conscientiousness, computed using the EAR corpus. The target output ranges from 1 to 7, where 7 means strongly conscientious (*Comm. words* is the ratio of words related to communication).

Classification accuracies are in Table 5. The Naive Bayes algorithm produces the best result, significantly outperforming the baseline for extraversion (73.2% correct classifications), emotional stability (70.7%), and conscientiousness (65.7%). Extraversion is the easiest personality dimension to model through spoken language, as accuracies for all classifiers except nearest neighbour are higher for this trait. The J48 decision tree for extraversion is shown in Figure 3.

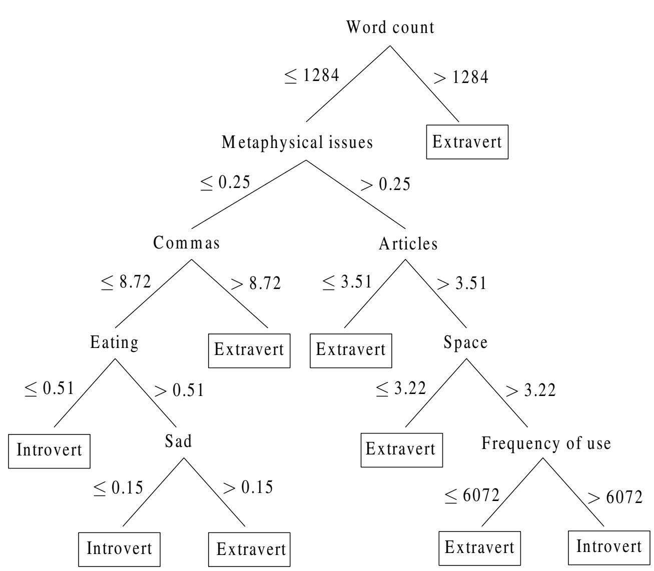

Figure 3: J48 decision tree for binary classification of extraversion, based on the EAR corpus.

Our models contain features characterizing many aspects of language production: speech acts, content and syntax (LIWC), psycholinguistic statistics (MRC), and prosody. In order to evaluate how each feature set contributes to the final result, we trained binary classifiers on the EAR corpus using the algorithm producing the best overall results (Naive Bayes) with each feature set. Table 6 shows that LIWC features perform well for extraversion and emotional stability, while MRC features are good indicators of extraversion and conscientiousness. Prosodic features are useful for modeling extraversion and especially openness to experience, and speech acts are the best features for modeling agreeableness.

Table 6: Classification accuracies over a 10 fold cross-validation using the Naive Bayes classifier, for different feature sets (the *Acts* column represents speech acts features).

Feature set	Acts	LIWC	MRC	Prosody
Set size	4	88	14	11
Extraversion	49.11	70.97•	67.96•	69.62•
Emotional stability	59.00	68.82•	61.04	61.88
Agreeableness	57.14	53.26	56.60	48.82
Conscientiousness	56.92	60.18•	65.76•	50.48
Openness	54.02	57.96	53.96	62.20•

• statistically significant improvement over the majority class baseline (two-tailed paired t-test, $p < 0.05$)

Decision tree models can be easily understood, and can therefore help to uncover new linguistic markers of personality. Our models replicate previous findings, such as the link between verbosity and extraversion (c.f. *Word count* node of Figure 3), but they also provide many new markers. Figure 1 shows that the voice's pitch and variation of intensity play an important role when modeling extraversion. Figure 2 shows that conscientious people use fewer swear words and content related to sexuality, while preferring longer words. Given particular ranges for features characterizing word count and the use of specific LIWC categories, the decision tree in Figure 3 classifies people using high-frequency words as introverts, contradicting previous hypotheses (Dewaele & Furnham, 1999; Furnham, 1990). Our models contain many additional personality cues which were not identified through a typical correlational analysis.

Conclusion

We showed that personality can be recognized by computers through language cues. To our knowledge, this is the first report of experiments testing automatically trained models on unseen subjects. The source of data is an essential factor: observed personality is easier to model than self-reports. This may be due to objective observers using similar cues as our models, while the perception of one's own personality is influenced by many other factors, such as the desirability of the trait. Moreover, spoken language is easier to model than written texts, probably because the speed of oral production prevents the cognitive system from doing the same amount of control as in a writing task. In future work, we would like to improve these models and examine how well they perform across dialogue domains. We need to test the models in our intended application (dialogue system adaptation) to assess whether the accuracies we achieve are high enough. Applications involving speech recognition will introduce noise in all features except for the prosodic features, probably reducing model accuracy, but since the EAR corpus is relatively small, we expect that more training data would improve performance.

Acknowledgements

We would like to thank James Pennebaker and Matthias Mehl for giving us access to their data and for their efficient and friendly collaboration.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A:497–505.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: the unloved variable in applied linguistic research. *Language Learning*, 49(3):509–544.

Dewaele, J.-M., & Furnham, A. (2000). Personality and speech production: a pilot study of second language learners. *Personality and Individual Differences*, 28:355–365.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3? Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12(8):773–790.

Fuchs, R. (2001). Personality traits and their impact on graphical user interface design: lessons learned from the design of a real estate website. In *Proceedings of the 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: an ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, 64(3):479–490.

Furnham, A. (1990). Language and Personality. In H. Giles & W. P. Robinson (Eds.), *Handbook of Language and Social Psychology*. Chichester: John Wiley & Sons.

Furnham, A., Jackson, C. J., & Miller, T. (1999). Personality, learning style and work performance. *Personality and Individual Differences*, 27:1113–1122.

Gill, A. J., & Oberlander, J. (2002). Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pp. 363–368.

Heylighen, F., & Dewaele, J.-M. (2002). Variation in the contextuality of language: an empirical measure. *Context in Context, Special issue of Foundations of Science*, 7(3):293–340.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The "Big Five" inventory: versions 4a and 5b. Technical report, Berkeley: University of California, Institute of Personality and Social Research.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: history, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality theory and research*. New York: Guilford Press.

Komarraju, M., & Karau, S. J. (2005). The relationship between the Big Five personality traits and academic motivation. *Personality and Individual Differences*, 39:557–567.

Linden, G., Hanks, S., & Lesh, N. (1997). Interactive assessment of user preference models: the automated travel assistant. In *Proceedings of the 6th International Conference on User Modeling*, pp. 67–78.

Litman, D., Kearns, M., Singh, S., & Walker, M. (2000). Automatic optimization of dialogue management. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pp. 502–508.

Mairesse, F. & Walker, M. (2005). Learning to personalize spoken generation for dialogue systems. In *Proceedings of Interspeech'2005 - Eurospeech: 9th European Conference on Speech Communication and Technology*, pp. 1881–1884.

Maloor, P., & Chai, J. (2000). Dynamic user level and utility measurement for adaptive dialog in a help-desk system. In *Proceedings of the 1st SIGdial Workshop on Discourse and Dialogue*, pp. 94–101.

McLarney-Vesotski, A. R., Bernieri, F., & Rempala, D. (in press). Personality perception: a developmental study. *Journal of Research in Personality*.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (in press). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality rating. *Journal of Abnormal and Social Psychology*, 66:574–583.

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality*, 68(5):821–836.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptor. *Journal of Personality and Social Psychology*, 57(3):552–567.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *LIWC: Linguistic Inquiry and Word Count*.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, 77:1296–1312.

Reeves, B., & Nass, C. (1996). *The Media Equation*. University of Chicago Press.

Rich, E. (1979). User modelling via stereotypes. *Cognitive Science*, 3:329–354.

Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech*. Cambridge University Press.

Sigurdsson, J. F. (1991). Computer experience, attitudes toward computers and personality characteristics in psychology undergraduates. *Personality and Individual Differences*, 12(6):617–624.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, 18:145–152.

Walker, M., & Whittaker, S. (1990). Mixed initiative in dialogue: an investigation into discourse segmentation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pp. 70–78.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Zukerman, I., & Litman, D. (2001). Natural language processing and user modeling: synergies and limitations. *User Modeling and User-Adapted Interaction*, 11(1-2):129–158.