Multimodal Recognition of Personality Traits in Social Interactions

Fabio Pianesi
FBK-irst
via Sommarive 18 Povo, Italy
+39-0461-314570
pianesi@fbk.eu

Nadia Mana
FBK-irst
via Sommarive 18 Povo, Italy
+39-0461-314570
mana@fbk.eu

Alessandro Cappelletti
FBK-irst
via Sommarive 18 Povo, Italy
+39-0461-314570
cappelle@fbk.eu

Bruno Lepri
FBK-irst
via Sommarive 18 Povo, Italy
+39-0461-314570
lepri@fbk.eu

Massimo Zancanaro
FBK-irst
via Sommarive 18 Povo, Italy
+39-0461-314570
zancana@fbk.eu

ABSTRACT
This paper targets the automatic detection of personality traits in a meeting environment by means of audio and visual features; information about the relational context is captured by means of acoustic features designed to that purpose. Two personality traits are considered: Extraversion (from the Big Five) and the Locus of Control. The classification task is applied to thin slices of behaviour, in the form of 1-minute sequences. SVM were used to test the performances of several training and testing instance setups, including a restricted set of audio features obtained through feature selection. The outcomes improve considerably over existing results, provide evidence about the feasibility of the multimodal analysis of personality, the role of social context, and pave the way to further studies addressing different features setups and/or targeting different personality traits.

Categories and Subject Descriptors
H.5.3 [INFORMATION INTERFACES AND PRESENTATION]: Group and Organization Interfaces: *Computer-supported cooperative work - Synchronous interaction.*

General Terms
Algorithms, Measurement, Experimentation, Human Factors.

Keywords
Personality Modeling, Group Interaction, Support Vector Machines, Intelligent Environments.

1. INTRODUCTION
Humans have the tendency to understand and explain other humans' behavior in terms of stable properties that are variously assorted on the basis of the observation of everyday behavior. In this sense, the attribution of a personality and its usage to infer about the others is a fundamental property of our naïve psychology.

Scientific psychology has maintained the importance of personality as a high-level abstraction encompassing sets of stable dispositions towards action and towards belief and attitude formation. The concept of personality is commonly used to explain human behavior in several domains: clinical and social psychology, educational studies and so on.

In every-day intuition, the personality of a person is assessed along several dimensions: we are used to talk about an individual as being (non-)open-minded, (dis-)organized, too much/little focused on herself, etc. Several existing theories have formalized this intuition in the form of multifactorial models, whereby an individual's personality is described in terms of a number of more fundamental dimensions known as traits, derived through factorial studies. A well known example of a multifactorial model is the Big Five [18] which owes its name to the five traits it takes as constitutive of people's personality:

- Extraversion vs. Introversion (sociable, assertive, playful vs. aloof, reserved, shy)
- Emotional stability vs. Neuroticism (calm, unemotional vs. insecure, anxious)
- Agreeableness vs. Disagreeable (friendly, cooperative vs. antagonistic, faultfinding)
- Conscientiousness vs. Un-conscientiousness (self-disciplined, organised vs. inefficient, careless)
- Openness to experience (intellectual, insightful vs. shallow, unimaginative)

Besides models that, as the Big Five, attempt to provide a comprehensive assessment of people personality, others have privileged specific dimensions, possibly useful to characterize people attitudes and behaviour in specific domains. An interesting example is the so-called Locus of Control (LoC) [31], which

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

measures whether causal attribution [16] for one's behavior or beliefs is made to oneself or to external events or circumstances. Hence, it consists of a stable set of belief about whether the outcomes of one's actions are dependent upon what the subject does (internal orientation) or on events outside of her control (external orientation) [31]. LoC has been used as an empirical tool in several domains; for instance, it was shown that people who feel they are the source or cause of their own attitudes and behaviors (internal LoC), tend to see the computer as a tool that they can control and use to extend their capabilities [19]. On the other hand, those who attribute their own behavior or attitudes to external factors (external LoC) are much proner to regard computers as an autonomous, social entity with which they are forced to interact.

The field of human computer interaction has shown a recurring, interest in the notion of personality. For instance, the latter has found a place in the repertoire of features a lifelike character should possess in order to improve its belivability; the underlying assumption is that a virtual agent would appear more realistic, understandable, and, ultimately, human-like, if, as a human, it exhibited a personality through consistent behaviours that the interacting humans could use understand its goals, form expectations about future behaviours, etc. [1], [7]. In the user modeling literature, information about the personality has been used to help inferring people's goals from their behavior, as in the work of Zhou and Conati [36] in the context of a tutoring system.

At a more general level, theoretical frameworks have been studied to provide principled links between people personality and a number of technology-related variables, such as attitudes towards and acceptability of technology. A notable example is the CASA—Computer as Social Agents—framework [19] positing that, in certain conditions, the relationship between humans and technology may be modeled in terms of social relations. One might therefore expect that personality plays a role in the way people use and experience technology, an intuition that Goren-Bar et al. [14] demonstrated to be true for adaptive systems: strong external orientation correlates with a preference for non-adaptive systems over adaptive ones: people who are highly sensitive to the social facets of technology because of their external LoC are not comfortable with adaptivity, or other forms of control delegation, in technology.

Despite those and other demonstrations of the importance of the notion of personality in shaping the human-machine relationship, and the declared interest in using it at the interface, progress have been slow, mainly for a lack of computational models of personality. In particular, at present only few works have attempted at providing theoretical and empirical frameworks for the automatic analysis of personality based on the observation of behavioural manifestations [22] [23]. This work intends to contribute to the specific task of the automatic analysis of people's personality.

To that end, a first and important issue to address is the characterization of the set of behavioural signs to be used in the analysis task. Social interaction is an ideal setting, in many respects, wherein people manifest their personality through verbal and non-verbal behaviour, exploit behavioural information from the partners to build intuitive models of their personality and use those insights to infer about their goals and intentions, anticipate actions, and regulate and tune own behaviour accordingly. For these reasons, in this work we consider acoustic and visual features, measured in the course of small working groups meetings.

The task we are considering is a classification one: on the basis of 1-minute-long behavioral sequences, the system must assign the subjects to the right class on two personality traits, Extraversion (one of the dimensions of Big Five) and Locus of Control (LoC); to this end, the continuous distributions of Extraversions and LoC will be turned into discrete ones (Low, Medium and High).

In relevant respects, the task is similar to the one we, as humans, are routinely involved in when judging about strangers' personality from very short behavioural sequences. Those *intuitions*, based on so-called *thin slices* of behaviour, and the process they come by have been the subject of extensive investigation by social psychologists in the last years [20].

The term 'thin slices' was coined by Ambady and Rosenthal [2] to refer to the short amount of information that we, humans, rely on to produce impressively precise judgments on an individual's or group's properties (personality, teaching capabilities, negotiation outcomes, etc.). In the mentioned paper, female college students evaluated 30-sec. silent videos of instructors teaching a class; their judgments showed very high correlations (r=.76) with end-of-semesters ratings of those same instructors by their students. Importantly, this result was replicated with high school teachers using thinner slices of visual data (6 sec. for each instructor). After this, and the other studies, the term 'thin slices' has come to be used to refer to an essential component of our social cognitive capabilities.

Another impressive related result is found in the marital research conducted by the psychologist John Gottman [13]. For example, Carrere and Gottman [6] were able to predict marital outcomes over a six year period only by observing the human micro-coding of positive and negative affect over the first three minutes of a marital conflict. Finally, Cuhran and Pentland [9] showed how the conversational dynamics occurring within the first 5 minutes of a negotiation can predict its outcomes.

Given that we take the golden standard to consist of scores on standard personality tests (Big Five's Extraversion, and the LoC scale), the present study bears close similarity to expert (e.g., psychologist) judgments provided quickly and on the basis of small amounts of information s.

2. PREVIOUS AND RELATED WORK

Psychologists have shown the existence of a correlation between extraversion and verbal behavior, and in particular with prosodic features. Extraversion is associated with higher pitch and higher variation of the fundamental frequency [34], with fewer and shorter silent and filled pauses, and with a higher voice quality and intensity [24]. Moreover, studies on the differences between the communication styles of introverts and extroverts suggest that the latter speak more and more rapidly, with fewer pauses and hesitations [11]. We were not able to find studies addressing similar issues for the Locus of Control.

Perhaps, the first work addressing the automatic recognition of personality was [3], who used the relative frequency of function words and of word categories based on Systemic Functional Grammar, to train Support Vector Machines with linear kernel for the recognition of Extraversion and Emotional Stability. The data concerning the two personality traits were based on self-reports.

Oberlander and Nowson [26] trained Naive Bayes and Support Vector Machines with linear kernel for four of the Big Five traits on a corpus of personal weblogs, using n-gram features extracted from the dataset. Their personality data were obtained through self-reports. A major finding of theirs is that the model for Agreeableness was the only one to outperform the baseline.

Mairesse et al. [22] [23] applied classification, regression and ranking models to the recognition of the Big Five personality traits. They also systematically examined the usefulness of different sets of (acoustic and textual) features suggested by the psycholinguistic and psychosocial literature. As to the personality data, they compared self-reports with observed data. Mairesse et al. could show that Extraversion is the easiest personality trait to model from spoken language, that prosodic features play a major role, and that their results were closer to those based on observed personality than on self-reports.

We are not aware, at present, of any study attempting to automatically evaluate LoC.

3. THE MISSION SURVIVAL 2 CORPUS

For our study, we used the MS-2 (Mission Survival 2) Corpus [25], an annotated multimodal corpus of multi-party meetings around a table, based on audio and video recordings[1].

In order to have an active discussion in a laboratory setting, we employed the "Mission Survival Task". This task is often used by experimental and social psychologists to elicit decision-making processes in small groups. Originally designed by the National Aeronautics and Space Administration (NASA) to train astronauts, the task proved to be a good indicator of the group decision making processes [15]. The exercise consists in promoting group discussion by asking the participants to reach a consensus on a list of appropriate items that can allow survival to disaster scenario. Each participant is asked to express her own opinions, while the group is encouraged to discuss each individual proposal through the weighing and evaluation of decision quality, and finally rank the proposed items according to their importance for survival.

The corpus consist of 12 meetings of 4 participants each, for a total length of over 6 hours.

The meetings were video-recorded with four fire-wire cameras placed in the corners of the lab and one directly above the round table; for audio recording, four wireless closed-talk microphones worn by the participants and one omni-directional microphone placed on the tabletop were used.

3.1 Description of the Corpus

As said in the introduction, our personality finds a major manifestation in the way we interact with other humans, by means of a number of diverse verbal and non-verbal behaviours. Moreover, we submit that, once situational aspects are controlled out, those manifestations are remarkably stable. Hence, a controlled setting such as the one described above provides an interesting testing field for studying personality.

3.1.1 Speech features

The audio corpus was annotated with a number speech features computed using 1-minute-length windows. Earlier works [27], [28], [35] suggest, in fact, that this window size is large enough to compute the features in a reliable way while being small enough to capture the transient nature of social behavior. For the relevant analyses, we employed the speech feature extraction toolbox developed by the Human Dynamics group at Media Lab[2].

The set of acoustic features (labeled F1-F22) extracted from the audio corpus is reported in Table 1.

Table 1. Extracted acoustic features (Mean and Standard Deviation calculated on 1 minute)

LABELS	ACOUSTIC FEATURES
F1	Mean of Formant Frequency (Hz)
F2	Mean of Confidence in formant frequency
F3	Mean of Spectral Entropy
F4	Mean of Largest Autocorrelation Peak
F5	Mean of Location of Largest Autocorrelation Peak
F6	Mean of Number of Autocorrelation Peaks
F7	Mean of Energy in Frame
F8	Mean of Time Derivative of Energy in Frame
F9	SD of Formant Frequency (Hz)
F10	SD of Confidence in formant frequency
F11	SD of Spectral Entropy
F12	SD of Value of Largest Autocorrelation Peak
F13	SD of Location of Largest Autocorrelation Peak
F14	SD of Number of Autocorrelation Peaks
F15	SD of Energy in Frame
F16	SD of Time Derivative of Energy in Frame
F17	Average length of voiced segment (seconds)
F18	Average length of speaking segment (seconds)
F19	Fraction of time speaking
F20	Voicing rate
F21	Fraction speaking over
F22	Average number of short speaking segments

We focused on four classes of features: 'Activity', 'Emphasis', 'Mimicry', and 'Influence', [28], [29], measuring vocal signals in social interactions. In Pentland's view, these four classes of features are honest signals, "behaviors that are sufficiently expensive to fake that they can form the basis for a reliable channel of communication" [29], and they can be used to predict and explain the human behavior in social interactions.

Activity, meant as conversational activity level, usually indicates interest and excitement. Such level is measured by the z-scored percentage of speaking time (features F7, F17, F18, F19 and F20). For this purpose, the speech stream of each participant is first segmented into voiced and non-voiced segments, and then the voiced ones are split into speaking and non-speaking.

Emphasis is often considered a signal of how strong is the speaker's motivation. Moreover, the consistency of emphasis (the lower the variations, the higher the consistency) could be a signal

[1] The audio and video recordings and the annotations are available for the research community at the url http://tcc.itc.it/research/i3p/ms-2/

[2] http://groupmedia.media.mit.edu/data.php

of mental focus, while variability may signal an openness to influence from other people. Emphasis is measured by the variation in prosody, i.e. pitch and amplitude. For each voiced segment, the mean energy, frequency of the fundamental format and the spectral entropy are extracted (features F1, F2, F3, F4, F5, F6 and F8). The mean-scaled standard deviation of these extracted values is then estimated by averaging over longer time periods (features F9, F10, F11, F12, F13, F14 and F16).

Mimicry, meant as the un-reflected copying of one person by another during a conversation (i.e. gestures and prosody of one are "mirrored" by the other), is expressed by short interjections (e.g. "uh-huh", "yup") or back-and-forth exchanges consisting of short words (e.g. "OK?", "done!"). Usually, more empathetic people are more likely to mimic their conversational partners: for this reason, mimicry is often used as an unconscious signal of empathy. Mimicry is a complex behavior and therefore difficult to computationally measure. A proxy of its measure is given by the z-scored frequency of these short utterances (< 1 second) exchanges (features F22).

Finally, *Influence*, the amount of influence each person has on another in a social interaction, was measured by calculating the overlapping speech segments (feature F21). Influence is a signal of dominance. Moreover, its strength in a conversation can serve as an indicator of attention. It is difficult, in fact, for a person maintain the rhythm of the conversational turn-taking without paying attention to it.

3.1.2 Visual features

We focused on few features related to the fidgeting, that is the amount of energy associated with body gestures, under the assumption that it correlates well with traits such as Extraversion and LoC.

In the MS-2 corpus, given the video sequences captured by the cameras in the meeting lab, the fidgeting cues have been automatically annotated by employing the MHI (Motion History Images) techniques [8]. These techniques use skin region features and temporal motions to detect repetitive motions in the images and associate such motions to an energy value in such a way that the higher the value, the more pronounced the motion.

In the corpus each annotation consists of an absolute timestamp, followed by three parameters relating to the fidgeting energy of head, hands and body.

3.1.3 Personality traits

Extraversion was measured by means of the extraversion sub-scale of the Italian version of the Big Five Marker Scale [30], while LoC through the Italian version of Craig's Locus of Control of Behavior scale [10]. The scales were administered to participants before they engaged in the interaction task.

To provide for a classification schema, the two traits' scores were clustered into three classes (Low, Medium and High) each, with Medium comprising scores ranging ±1SD around the mean, and the Low and High classes including the scores below and above that interval, respectively. Hence, in both cases the Medium class accounted for approximately 66.67% of the instances, while the other two classes were of similar size (approx. 16.5% instances each).

3.2 Feature selection

Feature selection was performed on the acoustic features by comparing their means through ANOVA: each feature was treated as a dependent variable in two between subject analysis of variance, with factor Extraversion (3 levels: L, M, H) and LoC (3 levels: L, M, H)); significance level was $p<.05$. No adjustment for multiple comparisons was performed, in order to have a more liberal test. Only the features for which the analysis of variance gave significant results were retained, for the given factor: F1, F2, F6, F14, a subset of the Emphasis class, and F21, the Influence feature, for Extraversion, and F1, F6, F14, the same subset of the Emphasis class apart for the mean energy, and F22, the Mimicry feature, for LoC.

4. THE EXPERIMENT

As said in Section 1, in our classification task classifiers must predict personality traits by considering the behavior of a subject in a 1-minute temporal window (similarly to a psychologist asked to recognize personality traits of a person by observing 1-minute behavioral sequences).

As classifier, we used Support Vector Machines (SVMs): these classifiers try to find a hyper-plane that not only discriminates the classes but also maximizes the margin between these classes [5]. SVMs were originally designed for binary classification but several methods have been proposed to construct multi-class classifier [17]. The "one-against-one" method [21] was used whereby each training vector is compared against two different classes by minimizing the error between the separating hyper-plane margins. Classification is then accomplished through a voting strategy whereby the class that most frequently won is selected.

The bound-constrained SVM classification algorithm with a RBF kernel was used. The cost parameter C and the kernel parameter γ were estimated through the grid technique by cross-fold validation using a factor of 10^3. Furthermore, the cost parameter C was weighted for each class with a factor inversely proportional to the class size.

4.1 Experimental design

The assumption underlying this study is that personality shows up in social behaviour, and we expect that the acoustic and visual features described in section 3.1.1 and 3.1.2 are appropriate to provide the 'thin slices' an automatic system can exploit to classify personality traits.

That task can be pursued in (at least) two different manners, each corresponding to a different hypothesis about the way personality, as manifested in social interaction, can be assessed. According to the first, the sole consideration of the target subject' behaviour (her thin slices) is enough: the way she moves, the tone and energy of her voice, etc., are sufficiently informative to get at her personality. The second view maintains that, the appreciation of personality requires information not only about the target's behaviour, but also about the social context: the same behaviour might have a different import for personality assessment if

[3] We used the BSVM tool available at http://www.csie.ntu.edu.tw/~cjlin/bsvm/.

produces in a given social environment than in another. Hence, thin slices of the other group members are needed as well.

A second aspect to test is the effectiveness of the feature selection procedure.

To test these two dimensions, and focusing on the acoustic features, we designed a between-subject experiment with factors ‘target’ and ‘others’, each relating to different arrangements of the target subject’s (target) and of the other group members’ (others) features.

- Target has two levels: all acoustic features+ visual features (ALL) vs. selected features + visual features (SEL).
- Others has three levels: no acoustic features + visual features (No-Feat); all acoustic features+ visual features (ALL); selected features + visual features (SEL).

A given combination in the experiment—e.g., (ALL, No-Feat)—corresponds to a specific arrangement of the feature vectors used to train and test the classifiers—in the example, all the acoustic plus the visual features of the subject, and, for each of the other group members, only the visual features—and to a specific combination of the hypothesis dimensions discussed above—in the example, that it is enough to consider thin slices of the sole target subject, and that the whole set of acoustic features are needed. The result is a 2×3 design that gives the possibility of fully testing the hypotheses combinations.

For each experimental condition, the training instances included the average values of the relevant acoustic and visual feature, computed over a 1-minute window; this way, the total number of generated instances corresponded to the total meetings’ duration in minutes (i.e. 366 minutes).

The analysis was conducted by means of 15-fold stratified cross-validation, with the same 15 training/test sets pairs being used in all the design 6 conditions. Stratification was conducted in order to closely reproduce in the training and test sets the distribution of Extraversion and LoC in the whole corpus.

4.2 Results

Tables 2-3 report the results in terms of accuracy, while Tables 4-5 report the average macro-F figures. In this paper, we will limit our discussion to accuracy, comparing our results with those of the trivial classifier that always assigns the most frequent class to each instance (Accuracy=0.6667).

Both for Extraversion and for LoC, the global average values of accuracy are well above the performance of the trivial classifier (0.8914 and 0.8718, respectively).

Two analysis of variance, one for Extraversion and one for LoC, showed that all the main effects are significant (p<.0001); interaction effects were not significant (p>.05) for Extraversion, and were significant for LoC (p<.05). With reference to the marginal means, both for Extraversion and LoC the usage of all the features for the target subjects yields much better results in terms of accuracy, the advantage being even more marked for LoC (0,9116 vs. 0.8713 for Extraversion and 0.9197 vs. 0.8238, for LoC).

Table 2. Means and SDs of accuracy for Extraversion

	Others			
Target	No-Feat	ALL	SEL	
ALL	0.8889 (.029)	0.9021 (.028)	**0.9438** (.021)	0.9116 (.035)
SEL	0.8493 (.024)	0.8611 (.036)	0.9035 (.026)	0.8713 (.037)
Total	0.8691 (.033)	0.8816 (.038)	0.9237 (.031)	0.8914 (.041)

Table 3. Means and SDs of accuracy for LoC

	Others			
Target	No-Feat	ALL	SEL	
ALL	0.9014 (.026)	0.9090 (.021)	**0.9486** (.016)	0.9197 (.030)
SEL	0.7847 (.040)	0.8396 (.042)	0.8472 (.039)	0.8238 (.048)
Total	0.8431 (.068)	0.8743 (.048)	0.8979 (.059)	0.8718 (.062)

Table 4. Means and SDs of mean macro-F for Extraversion

	Others			
Target	No-Feat	ALL	SEL	
ALL	0.8399 (.048)	0.8529 (.055)	**0.9198** (.037)	0.8708 (.058)
SEL	0.7774 (.039)	0.7837 (.038)	0.8630 (.039)	0.8081 (.054)
Total	0.8087 (.053)	0.8183 (.058)	0.8914 (.047)	0.8395 (.064)

Table 5. Means and SDs of mean macro-F for LoC

	Others			
Target	No-Feat	ALL	SEL	
ALL	0.9404 (.016)	0.9488 (.012)	**0.9722** (.012)	0.9538 (.019)
SEL	0.8740 (.023)	0.7628 (.063)	0.9059 (.019)	0.8476 (.073)
Total	0.9072 (.039)	0.8558 (.104)	0.9390 (.03)	0.9007 (.075)

Concerning the effect of the context, as captured through the factor ‘others’, contrast analysis shows that the usage of acoustic features yields better results for both Extraversion (contrast value=0.067, p<.0001) and LoC (contrast value=0.086, p<.0001). Moreover, the best results are obtained when the social context is capture by means of the selected features (condition SEL), both for Extraversion (contrast value=0.097, p<.0001) and LoC (contrast value=0.078, p<.0001). Finally, we can neglect the interaction effect for LoC, since it is due to the rather low accuracy value in condition (SEL, No-Feat) and absence of a significant difference between (SEL, ALL) and (SEL, SEL), see Fig. 2 and Table 3.

4.3 Discussion

Contrary to our expectations, the features selected according to the procedure described in section 3.2 are not effective: when applied to the target subject they constantly yield worse results, as the summary curves in Fig. 1 and Fig. 2 show. Clearly, the feature selection procedure was not that effective, as far as the target subject is concerned. A more detailed analysis (not reported here for reasons of space) shows that, for Extraversion the precision and recall figures at classes Low and Medium do not differ according to whether ALL or SEL conditions are used for the target subject; in the High class, on the other hand, SEL is significantly worse on both measures. Concerning LoC, the same analysis highlights a degradation of SEL with respect to ALL in all three classes (Low, Medium and High), which is, however, stronger at classes Low and High. A possible explanation for the ineffectiveness of the feature selection procedure is that it picked up features that appropriately describe only some of the classes we divided the personality traits into.

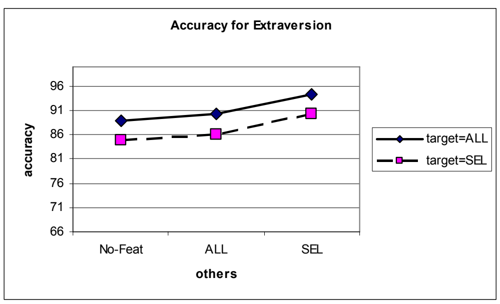

Figure 1. Accuracy for Extraversion

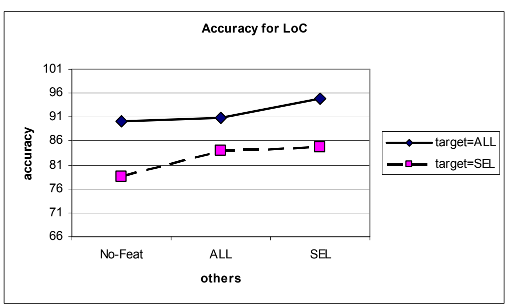

Figure 2. Accuracy for LoC

Concerning the other hypothesis, it is confirmed that the encoding of the social context (what the other members of the group do) improves personality classification. It is important to emphasise, however, that even in the absence of any attempt to (acoustically) capture the social context, the performance obtained are all much higher than the baseline provided by the trivial classifier: 0.8691 for Extraversion and 0.8431 for LoC. Considering that the baseline is 0.6667, the relative improvement is 0.607 and 0.529, respectively. Hence, thin slices of the sole target subject's behaviour are enough to obtain quite a good automatic classification of the two personality traits we are considering.

Our results show that the way the social context is encoded matters a lot: the best performances are obtained when the selected features are used. A more detailed analysis on the figures reported in Tables 2 and 3 shows that when all the acoustic features are used for the target subject, no advantage is obtained by using the same features for the social context with respect to not using any acoustic features at all; the comparisons between (ALL, No-Feat) and (ALL, ALL) are not significant at p=.05, both for Extraversion and LoC. An improvement is obtained only when the social context is captured through the selected features (comparisons between (ALL, SEL), and (ALL, No-Feat) and (ALL, ALL) both significant at p<.0001, for both personality traits. The role of the selected features in capturing the social context is striking and unexpected, given that a) the choice procedure aimed to improve the recognition of personality on the target subject, and b) the same features are inefficacious to that purpose.

5. CONCLUSION AND FUTURE WORK

The aim of this paper was to contribute to advance the state of the art in the automatic analysis of people personality. With respect to other similar works, we based our approach on the assumption that a) personality shows up in a clearer form in the course of social interaction and b) that thin slices of social behaviour are enough to allow personality traits classification. The first assumption was realized by exploiting classes of acoustic features encoding specific aspects of social interaction (Activity, Emphasis, Mimicry, and Influence) and three visual features (head, body, and hands fidgeting). As to the second, we considered 1-minute long behavioral sequences. The resulting task for the classifier is similar to that of an expert (e.g., a psychologist) that must provide a personality assessment of strangers based only on short sequences of theirs.

The classification study was based on two main hypotheses: a) that a simple feature selection procedure could provide a smaller, but still effective, subset of the acoustic features, and b) that the encoding of the social contexts (in the form of the other group members' acoustic features) could contribute to the accuracy of classification.

The results discussed in the previous sections show that the first hypothesis cannot be kept: the feature selection procedure was not effective, when applied to the target subject. The second hypothesis was confirmed, though in an unexpected form: capturing the social context by means of the selected features greatly improves classification accuracy, whereas, at least when the classifier has access to all the acoustic features of the target subject, the availability of the same features for all the other members does not bring about any improvement.

At a more general level, we believe that our results largely support the idea that social interaction is an ideal context to conduct automatic personality assessment in, and prove the feasibility of the thin slice approach to personality classification: our accuracy figures are all much higher than the baseline, and higher than those reported in the few studies on the topic published so far (e.g., [23]).

Given these initial encouraging results, several research directions open for future research.

- Move towards a task that is closer to that of psychologists, by considering regression techniques, or techniques for ordinal scales [23].
- Work out more efficacious feature selection procedures, while better understanding the role and balance between the features used for the target subject and those for the social context.
- Explore the role of visual features (a task we have not pursued in this work), even beyond fidgeting, to include, e.g., amount of social attention (focus of attention) received.
- Provide for more comprehensive personality assessment that can be actually used in realistic setting—e.g., by considering the full set of Big Five's scales. Conceivably, this move might require enlarging the scope of the context explored beyond the social ones. It is well known, in fact, that traits such as Extraversions are more deeply involved in social behaviour than others, such as conscientiousness. Another direction for a move towards practical impact is towards addressing traits that, much as the Locus of Control considered here, have been argued to be important for the relationship and the interaction between humans and machines (e.g., Computer Anxiety [12]).

6. ACKNOWLEDGMENTS

This work has been partially supported by the EU-funded CHIL (FP6) project.

7. REFERENCES

[1] André, E., Klense, M., Gebhard, P., Allen, S., Rist, T. Integrating models of personality and emotions into lifelike characters. In *Proceedings of the Workshop on Affect in Interaction - Towards a New Generation of Interfaces*, pp 136-149, 1999.

[2] Ambady, N., and Rosenthal, R. Thin slices of expressive behaviors as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, pp. 256-274

[3] Argamon, S., Dhawle, S., Koppel, M., Pennbaker, J. Lexical predictors of personality type. In *Proceedings of Interface and the Classification Society of North America*. 2005.

[4] Blanck, P.D., Rosenthal, R., Vannicelli, M., and Lee, T.D. Therapists' tone of voice: descriptive, psychometric, interactional and competence analyses. *Journal of Social and Clinical Psychology*, 4, pp. 154-178. 1986

[5] Burges, C.J.C., A Tutorial on Support Vector Machines for Pattern Recognition. In *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 121-167, 1998.

[6] Carrere, S., and Gottman, J.M. Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process*, 38, pp. 293-301

[7] Castelfranchi, F. D. Rosis. How can personality factors contribute to make agents more 'believable', *Proceedings of the Workshop on Behavior Planning for Lifelike Characters and Avatars*, Sitges, Spain, 25-35, 1999.

[8] Chippendale, P. Towards Automatic Body Language Annotation. *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition - FG2006 (IEEE)*, pp. 487-492, Southampton, UK, 2006.

[9] Curhan, J.R., and Pentland, A. Thin slices of Negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, Vol. 92, No. 3, pp. 802-811, 2007

[10] Farma, T., and Cortivonis, I. Un Questionario sul "Locus of Control": Suo Utilizzo nel Contesto Italiano (A Questionnaire on the Locus of Control: Its Use in the Italian Context). *Ricerca in Psicoterapia*. Vol. 2, 2000.

[11] Furnham, D. Language and Personality. In Giles, H., & Robinson, W. (Eds.), *Handbook of Language and Social Psychology*. Winley, 1990.

[12] Giannoutsos, V. 'Stress, Attitudes and Personality in Computing'. In Proceedings of ETHICOMP 2004, Syros, Gr. 2004

[13] Gottman, J.M, and Notarius, C.I. Decade review: observing marital interaction. *Journal of Marriage and the Family*, 62, pp. 927-947, 2000

[14] Goren-Bar, D., Graziola, I., Pianesi, F., Zancanaro, M. Influence of Personality Factors on Visitors' Attitudes towards Adaptivity Dimensions for Mobile Museum Guides'. In *User Modeling and User Adapted Interaction: The Journal of Personalization Research*, 16 (1), 31-62. 2006

[15] Hall, J. W., and Watson, W. H. The Effects of a normative intervention on group decision-making performance. In *Human Relations*, 23(4), 299-317, 1970.

[16] Heider, F. *The psychology of interpersonal relations*. Wiley. New York. 1957

[17] Hsu, C.W. A. Hsu C.-W., Lin C.-J. A Comparison of Methods for Multi-Class Support Vector Machines. *IEEE Transactions on Neural Networks*, 13, pp. 415-425, 2002.

[18] John, O. P., Srivastava, S. The Big five trait taxonomy: History, measurement and theoretical perspectives. In Pervian, L. A. & John, O. P., (Eds.) *Handbook of personality theory and research*. Guilford Press. New York, 1999.

[19] Johnson, R. D., Marakas, G., Plamer, J. W. Individual Perceptions Regarding the Capabilities and Roles of Computing Technology: Development of The Computing Technology Continuum of Perspective. Ms. 2002.

[20] Kenny, D. A. *Interpersonal perception: A social relations analysis*. New York: Guilford Press. 1994.

[21] Kressel U. Pairwise Classification and Support Vector Machines. In B. Scholkopf, C.J.C. Burges, and A.J. Smola (eds.) *Advances in Kernel Methods – Support Vector Learning*. MIT Press, Cambridge, MA, 1999.

[22] Mairesse, F., Walker, M. Words marks the nerds: computational models of personality recognition through language. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pp 543-548, 2006.

[23] Mairesse F., Walker M.A., Mehl M.R., and Moore R.K. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In *Journal of Artificial Intelligence Research* 30, pp.457-500. 2007

[24] Mallory P., and Miller V. A possible basis for the association of voice characteristics and personality traits. Speech Monograph, 25, pp. 255-260, 1958.

[25] Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P, and Zancanaro, M. Multimodal Corpus of Multi-Party Meetings for Automatic Social Behavior Analysis and Personality Traits Detection. In *Proceedings of Workshop on Tagging, Mining and Retrieval of Human-Related Activity Information, at ICMI07, International Conference on Multimodal Interfaces*, Nagoya, Japan, 2007.

[26] Oberlander, J. and Nowson, S. 2006. Whose thumb is it anyway? Classifying author personality from weblog text. In Proceedings of the Annual Meeting of the ACL. Association for Computational Linguistics, Morristown, NJ, 627-634, 2006.

[27] Pentland, A. Socially Aware Computation and Communication. IEEE Computer, May 2005.

[28] Pentland, A. A Computational Model of Social Signaling. *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*. Volume 1, Issue 2006 , Page(s):1080 – 1083, 2006.

[29] Pentland, A. *Honest Signals: how they shape our world*. (In press). MIT Press, September 2008.

[30] Perugini, M. and Di Blas L. Analyzing Personality-Related Adjectives from an Eticemic Perspective: the Big Five Marker Scale (BFMS) and the Italian AB5C Taxonomy. In De Raad, B., & Perugini, M. (Eds.), *Big Five Assessment*, Hogrefe und Huber Publishers. Göttingen, 281-304, 2002.

[31] Rotter J.B. Generalized Expectancies for Internal versus External Control of Reinforcment. In *Psychological Monographs*, 80 (1, Whole N. 609), 1965

[32] Siegman A., and Pope B. Personality variable associated with productivity and verbal fluency in the initial interview. *Proceedings of the 73rd Annual Conference of the American Psychological Association*, 1965.

[33] Scherer K.R. Personality Inference from Voice Quality: the Loud Voice of Extraversion. *European Journal of Social Psychology*, vol. 8, pp. 467-487, 1978.

[34] Scherer K.R. Personality markers in speech. In Scherer K.R. and Giles H. (eds.) *Social Markers in Speech*, pp. 147-209 Cambridge University Press, 1979.

[35] Stoltzman, W. Toward a Social Signaling Framework: Activity and Emphasis in Speech. MEng Thesis, MIT, 2006

[36] Zhou, X. and Conati, C. Inferring user goals from personality and behavior in a causal model of user affect. In *Proceedings of the 8th international Conference on intelligent User interfaces* (Miami, Florida, USA, January 12 - 15, 2003). IUI '03. ACM, New York, NY, 211-218. DOI= http://doi.acm.org/10.1145/604045.604078, 2003.