Employing Social Gaze and Speaking Activity for Automatic Determination of the *Extraversion* Trait

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

Ramanathan Subramanian
DISI, University of Trento
via Sommarive 14 Povo, Italy
subramanian@disi.unitn.it

Kyriaki Kalimeri
CIMEC, Univ. of Trento and FBK-irst
Corso Bettini, 38068 Rovereto, Italy
kalimeri@fbk.eu

Jacopo Staiano
DISI, University of Trento
via Sommarive 14 Povo, Italy
jacopostaiano@gmail.com

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

Nicu Sebe
DISI, University of Trento
via Sommarive 14 Povo, Italy
sebe@disi.unitn

ABSTRACT
In order to predict the *Extraversion* personality trait, we exploit medium-grained behaviors enacted in group meetings, namely, speaking time and social attention (social gaze). The latter will be further distinguished into *attention given* to the group members and *attention received* from them. The results of our work confirm many of our hypotheses: a) speaking time and (some forms of) social gaze are effective in automatically predicting Extraversion; b) classification accuracy is affected by the size of the time slices used for analysis, and c) to a large extent, the consideration of the social context does not add much to accuracy prediction, with an important exception concerning social gaze.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human Information Processing; I.5.4 [**Pattern Recognition Applications**]: Computer Vision

General Terms
Algorithms, Measurement, Experimentation, Human Factors.

Keywords
Personality Prediction, Visual Social Gaze, Speaking Activity, Support Vector Machines

1. INTRODUCTION
It is customary for all of us to describe people as being more or less *talkative*, *bold* or *sociable*. We all exploit these descriptors in our everyday life to explain and/or predict people's behavior, attaching such labels to well-known as well as to new acquaintances. *Extraversion*, the trait dimension they refer to, is so familiar that we continuously exploit it inconspicuously.

ICMI-MLMI'10, November 8-10, 2010, Beijing, China.

Social and personality psychology have been long concerned with Extraversion, which has emerged as one fundamental dimension of personality [8] because it is: a) capable of explaining a wide range of behaviors; b) able to predict functioning across a number of domains, ranging from cognitive performance [24] and social endeavors [11] to socio-economic status [31], and c) useful to assess the risk of different types of psychopathology [39].

The importance of personality for technology and human-computer interaction in particular has also been acknowledged. Studies have shown that personality traits determine people's attitudes toward machines in general [34], towards basic dimensions of adaptivity [13] as well as towards conversational agents [30; 7]. It has been argued that social networking websites could increase the chances of a successful relationship by first analyzing text messages, and then matching personalities [10]; Tutoring systems, in turn, would be more effective if they could adapt to the learner's personality [17; 43]. Moreover, given its relevance in social settings, information on people's personality, and in particular on their Extraversion level, can be useful for providing personalized support to group dynamics [27]. A brief discussion on the characterization of meeting and personality types from meeting videos is presented in [39].

Consequently, a number of works have started exploring automated personality analysis [2; 19; 20; 25; 21; 28; 18], often targeting the so called Big Five model of personality [8], of which Extraversion is a major dimension. The general approach aims at isolating promising behavioral correlates of the targeted traits, to use them for classification or regression purposes. For instance, Pianesi *et al.* [28] and Lepri *et al.* [18] have exploited the well known correlating ion between Extraversion and prosodic features - higher pitch and higher variation of the fundamental frequency [33], higher voice quality and intensity [22] - while Mairesse *et al.* [21] have considered both verbal and non-verbal (acoustic) cues.

In this work, we will exploit medium-grained behaviors enacted in interactive settings (small group meetings): amount of speech activity and social attention (social gaze). The latter will be further distinguished into attention (gaze) given to the other members of the group and attention (gaze) received.

There is a general consensus about the speech behavior of extraverts- they tend to speak more, more rapidly and with shorter pauses [3]. Assuming that the gazing behavior of the peers is affected by the speaking pattern of the target subject, one might expect that the amount of attention the latter receives depend on his/her Extraversion level.

It seems also reasonable to expect Extraversion to affect the social gazing of the target subject him/herself. The evidence, however, is mixed in this respect: some studies [15; 3] have concluded that extraverts gaze more frequently and with longer glances when talking; other research works [32], while confirming that extraverts look more frequently than introverts, could not find evidence that this translates into a longer time spent looking at the others.

Humans are capable of forming accurate impressions of people personality on the basis of very short sequences of expressive behaviors – the so-called *thin slices* [1]. In this work, we consider thin slices of speaking activity and social gaze, and investigate the effect of slice length on accuracy. The relationship between exposure time and human classification accuracy has not been unambiguously established. While meta-analytic studies [1] did not find significant effects of exposure time on judgment accuracy for many target constructs, Blackman and Funder [4] found a significant linear increase in agreement between observers' and targets' assessments as a function of exposure time, with a correlation coefficient increasing from r=.22 for 5-10 minutes slices to r=.26 for 25-30 minutes ones. More to our point, Carney *et al.* [6] found a significant linear increase in the assessment accuracy of Extraversion for slices ranging between 5 and 300 seconds. In this work, we are going to consider 2, 5 and 6 minutes long slices.

Finally, most thin-slice-based studies on human assessment of Extraversion let judges have access to the full social context in the form of, *e.g.*, videos where both the target subject and his/her meeting companions are simultaneously present while interacting. One might wonder whether the availability of the social context is useful to Extraversion prediction. To this end, we will systematically compare conditions in which only information about the target subject is fed into the system against others where information about the behavior of the other participants is also made available.

The results of our work confirm many of our hypotheses: a) speaking time and (some forms of) social gaze are indeed effective in automatically predicting Extraversion; b) classification accuracy is affected by the size of the slice used; c) to a large extent, consideration of the social context does not add much to accuracy prediction, with an important exception concerning social gaze.

The organization of this paper is as follows: Section 2 describes the 'Mission Survival' corpus that we used for our experiments, the methods employed to automatically extract medium-grained behavioral features, as well as observational data concerning human-annotation (ground-truth). Extraversion assessment though Support vector Machine (SVM)-based classification on automatically extracted behavioral features is described in Section 3, while experimental results are discussed in Section 4. We end by outlining key conclusions in Section 5.

2. THE MISSION SURVIVAL CORPUS

The Mission Survival corpus is a multimodal corpus of meetings where groups of four people, seated around a table, are involved in the Mission Survival Task (MST). The MST involves reaching a consensus through discussion about the appropriate ranking of 12 items for survival purposes after a plane crash. First, each participant expresses his/her own personal opinion and then the group discusses the merit of each proposal, to finally rank the items according to their importance. See [14] for more details on the MST.

The corpus consists of audio and video recordings of 12 meetings for a total of over 6 hours. Audio is recorded through close-talk microphones worn by each participant, and through one omni-directional microphone placed in the middle of the table. The visual recording equipment consists of five Firewire cameras, four placed in the corners of the room and one directly above the table, and four web cameras installed on the meeting table. For each participant, Extraversion is measured by means of the Big Marker Five Scales [26]. Readers are referred to [23] for a detailed description of the Mission Survival corpus.

2.1 Feature Extraction

Four meetings out of the twelve of the Mission Survival corpus were used to extract three sets of acoustic and visual cues. The meetings were selected to comply with some basic tracking criteria, such as sufficient facial visibility throughout the entire meeting duration. Starting from one minute-long intervals, the definitions of the three indices extracted for each participant p are the following:

- Speaking time (ST_p): the percentage of time p is speaking; It is given by the summation of the overall speaking activity described in the following section for a specific time fraction.
- Attention received (AR_p): the percentage of time at which at least one member of the group is looking at p.
- Attention given (AG_p): the percentage of time at which p is looking to other participants.

For automated analysis, the one minute-long intervals were aggregated into temporally overlapping windows (thin slices) of length 2/5/6 minutes and shifted by 1/2.5/3 minutes respectively, so as to cover the entire meeting duration. Then, the mean and standard deviation for each of the above features, computed over the basic intervals contained in the relevant window, were used for analysis.

In the following subsections we describe how these indices are automatically computed.

2.1.1 Speaking Activity

Speaking activity cues are extracted from close-talk microphone speech signals. The long-term spectral divergence algorithm [29] is used to discriminate between speech and non-speech signals. In order to detect vocal activity, we assume that the most significant information is contained in the time-varying spectral magnitude of the signal. After segmenting the initial signal, the long-term spectral envelope (LTSE) and the long term spectral divergence (LTSD) are estimated, in order to formulate the decision rule for voice activity detection.

Let $x(n)$ be the initial signal, segmented into overlapping frames and $X(k,l)$ it is amplitude spectrum for the k-th band in frame l. The N-order LTSE is defined as:

$$LTSE_N(k,l) = \max_{j=-N}^{j=+N}\{X(k,1+j)\}$$

The N-order long-term spectral divergence between speech and noise is defined as the deviation of LTSE with respect to the average noise spectrum magnitude $N(k)$ for the k-th band, with k = 0,1,...,NFFT-1, where NFFT is the length of the Fast Fourier Transform.

$$LTSD_N(l) = 10\log_{10}\left(\frac{1}{NFFT}\sum_{k=0}^{NFFT-1}\frac{LTSE^2(k,l)}{N^2(k)}\right)$$

The decision rule for voice activity detection is based on the LTDS between speech and noise, while the threshold distinguishing speech from non speech-regions is adjusted to maximize accuracy with respect to manually annotated ground-truth data. All meetings are annotated with the same binary schema, where speaking frames are labeled as `1' and non-speaking frames as `0'. The automated speech detection algorithm is 92% accurate with respect to the ground-truth, which is sufficient for our analysis.

The Speaking Time (ST) for the subject p is then calculated as

$$ST_p = \frac{Speaking\ Frames_p}{Total\ Number\ Frames} * 100\%$$

2.1.2 Gaze

The gaze direction of a subject, a reliable indicator of his/her focus-of-attention (FOA), is computed by fusing head pose and eye location information by means of the eye-localization scheme proposed in [41], which is able to accurately extract head-pose and eye-center locations from a monocular video sequence (see Figure 1).

Figure 1. (a) View-normalized eye-center estimation using CHM and eye-center locator. (b),(c) are examples where direction of social visual attention is different from head-pose direction. Blue triangular-normal denotes head-pose while green circles represent eye locations.

The method combines a robust cylindrical head model (CHM) pose tracker [42] and an isophote-based eye center locator [40], to obviate limitations of both methods when considered independently. The system integrates the eye locator with the CHM by interleaving the transformation matrices obtained by both systems. This way, eye-center locations are estimated given the pose, and conversely, pose is adjusted given the eye-center locations.

The 2D eye-center locations detected on a frontal face are used as reference points to initialize the CHM and used for eye-center estimation in subsequent video frames. The estimated eye-centers are projected on to a view-normalized model and therefore, displacement between the reference and current eye-center locations are independent of head pose. We approximate the gaze direction as horizontal and vertical shifts of the eye-centers from their resting positions on the head surface. The displacement vectors for the two eyes are averaged to compute the gaze-based FOA from the head pose. The process is outlined in Figure 2.

Table 1. Means and standard deviations (in parenthesis) of relevant indices.

ST	AR	AG	Extra
24.07 (9.56)	40.23 (12.42)	15.23 (9.52)	40.63 (10.07)

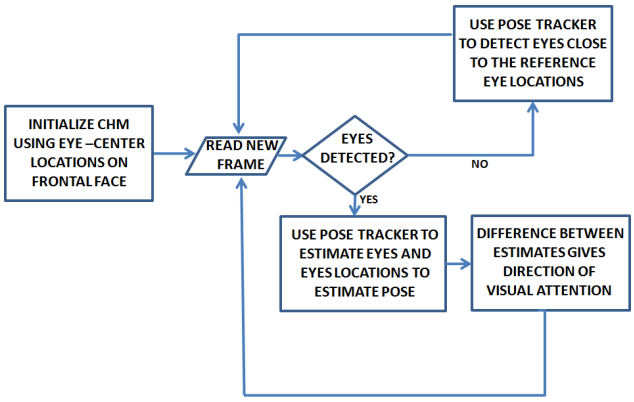

Figure 2. Overview of gaze-direction estimation.

Upon computing the gaze direction, we map it to classes 'L' (Left), 'O' (Opposite) 'R' (Right), and 'S' (Self-attention) using a Bayesian approach [37; 38]. Classes L, O and R include gazes directed at the person sitting on the left, opposite, and on the right hand side of the target subject, respectively. Class S includes gazes directed anywhere else (in most cases, this corresponds to participants looking at the list of items provided to them). Once the FOA classes are determined for each meeting participant over the entire meeting duration, the *Attention Received* and *Attention Given* features are computed for each participant according to the definitions given above.

The accuracy for the estimated attention given and attention received were assessed against manually labeled ground-truth data. For Attention Given, the mean Euclidean error between the estimates obtained when using only the Head Pose and the ground truth is 7.62, while it is 5.33 when the eye gaze-based social attention is employed together with the head pose. This corresponds to an error reduction of 30%. Similarly, for the *Attention Received* the social attention error obtained with head-pose only is 6.28, while the use of eye-gaze information in conjunction with head-pose reduces the error to 4.59, yielding an error reduction of 26.7%.

2.2 Feature Analysis

For each of the four meetings, speaking/non-speaking frames as well as the FOA targets were manually annotated to compute the speaking time (ST), attention received (AR) and attention given (AG) for each subject in 1-minute intervals. Means and standard deviations of the three indices and of the Extraversion score computed across the 16 subjects are reported in Table 1. Table 2 reports the correlation matrix.

Table 2. Correlation matrix. (* =>p<.05, **=> p<.01)

	ST	AR	AG	Extra
ST		0.584**	0.1	0.664**
AR			0.095	0.680**
AG				0.088
Extra				

As can be seen, AG never correlates in a significant manner with any of the other variables. To further investigate the relationships among those variables, a linear regression analysis of Extraversion on ST and AR was performed. The model yields R^2=0.571 and partial correlation coefficients $r_{partial}$(Extra, ST)=0.449; $r_{partial}$(Extra, AR)=0.482 (all significant at p<.05). Hence, despite their significant mutual correlation, ST and AR still bear a high partial correlation to Extraversion, together explaining 57.1% of its variance. It can be concluded that there exists a positive linear relationship between Extraversion, on one hand, and ST and AR, on the other, confirming that *people higher in Extraversion speak more and receive more attention.* There does not seem to be any significant linear relationship involving Extraversion and social gaze of targets (AG).

If instead of considering the social gaze as a global variable (amount of visual attention given to the rest of the group), we articulate it into its components (L, O and R gazes), then a correlation of -0.558 (p<.05) is found between the attention given to the person sitting directly opposite and the subject's Extraversion score. Nothing significant is found for the other components.

The data concerning ST and AR are expected, confirming evidence in the literature. The absence of a significant correlation between the Extraversion score and global social gaze, in turn, seem to support the findings in Rutter *et al.* [32], Iizuka [15] and Argyle [3]. *The negative correlation between Extraversion and the attention given to the person sitting right opposite* is, to our knowledge, new and somewhat surprising: extraverts gazed less at the person sitting in front of them than introverts. One possible explanation appeals to the task-related nature of our meetings: in task-oriented contexts introverts exhibit a greater capability of attending to stimuli and a lower level of distractibility [35]. Now, suppose that in our task-oriented meetings looking at the person sitting right opposite is not merely a sign of social attention (as it would be in a free-wheeling conversation) but is somehow task-related (*e.g.*, talking to the person sitting right opposite as a way to talk to the whole group); then the datum would follow: introverts increase or maintain their attention to task-related objects, including the person sitting in front of them, whereas extraverts keep distributing their attention around, driven by social concerns. More work is needed to confirm this hypothesis.

3. MODELING EXTRAVERSION USING SUPPORT VECTOR CLASSIFICATION

Following up on the results obtained from the analysis of ground-truth data, we attempted the automatic assessment of Extraversion from thin-slices, using features extracted according to the methodologies discussed in sections 2.1.1 and 2.1.2. In doing so, we also investigated the effects of slice length and of information about the social context on the classification accuracy. The resulting experimental design was a factorial one, with three factors:

- 'Time' with 3 different thin-slice durations, namely, T2, T5 and T6, corresponding to 2, 5 and 6 minute-long slices respectively.
- 'Target' with 4 levels: (i) Speaking time features (ST); (ii) Attention Given (AG); (iii) Attention Received (AR) and (iv) all features (ALL). Correspondingly, the feature vectors contains the average and standard deviation of ST, AG, AR or all of them simultaneously (Target=ALL) for the target subject. The case when Target=ALL allows for a simple multimodal analysis strategy.
- 'Others' with 5 levels, the same 4 as 'Target' plus 'No_Feat', corresponding to the absence of any features for the other participants. When Others='No_Feat', the feature vectors contain only information about the target; hence, the factor 'Others' is the means we use to address the contextual hypothesis discussed above. When `Others' value is different from 'No_Feat', *e.g.*, Others=AR, then the feature vectors contain, in addition to the feature for the target, the relevant ones for all the other participants (*e.g.*, average and standard deviation of AR for all the other participants). The order of other participants in the feature vector is standardized and based on the target: first the features of the subject to the left, then those of the person in front and finally the ones of the person on the right.

To elaborate further, a condition such as (T2, ST, ALL) indicates, the case where two minute-long slices are considered, the information concerning the target is limited to (the average and SD of) ST, and that about the others consists of (the averages and SDs) of all behavioral cues considered in this paper. Notice, also, that besides addressing a simple form of multimodal analysis by means of the conditions where Target=ALL or Others=ALL, our factorial design allows for cross-subject multimodal cases, as, in (*, ST, AR).

The task is modeled as a classification one. To this end, Extraversion scores were dichotomized along the median value, yielding two classes: 7 people with low extraversion (LOW) and 9 people with high Extraversion (HIGH). Those labels were added to the feature vectors described above for each time slice considered.

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[1].

For testing, cross validation was performed by using a leave-one-meeting out procedure: at each fold, all the slices relative to three meeting were used for training, and those of the left-out meeting for testing.

4. Results

The first step in the analysis of the results consisted in finding out which conditions yielded an accuracy significantly higher than that of the baseline classifier - namely, the classifier that assigns slices to LOW or HIGH according to their prior probabilities, 0.44 and 0.56, respectively. The procedure was the following:

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

- We first computed the expected accuracy of the baseline classifier (0.5072).
- Regarding accuracy as the probability of success in a Bernoulli process, we formed the Null Hypothesis according to which the distribution of hits and errors in the confusion matrix of each experimental condition was generated by a Bernoulli process with success probability of 0.5072.
- To test the Null Hypothesis, we used the binomial test. Given that this test is sensitive to the number of trials, and the latter varies according to slice size, three series of tests were performed, one for each level of the Time factor. For each run, the significance level was fixed at α=0.05, with Bonferroni correction for multiple comparisons.
- The tests were one-tailed and conducted only for conditions whose measured accuracy values appeared to be equal or higher than the expected one. The identified threshold where: $accuracy_{T2}$>0.57; $accuracy_{T5}$>0.61; $accuracy_{T6}$>0.62.

Table 3 reports the average accuracy values for each condition. Stricken-through items indicate conditions whose accuracy is not significantly higher than that of the baseline classifier (according to the procedure just discussed).

All conditions with Target=AG never yield accuracies higher than the baseline criteria. Similar considerations hold for cases when all the available features are used to characterize the context, with the exception of (*, ALL, ALL). Hence, we will disregard (*,AG, *) and (*, *, ALL) altogether in the following. Finally, any conditions with Others=ST has performance that, even when higher than the relevant criteria, is always very close to it. Not much is lost, therefore, if (*, *, ST) is not considered for analysis.

Table 3. Average accuracy for Extraversion

		Others				
	Target	No_Feat	ST	AG	AR	ALL
T2	ST	**.74**	.63	.59	.62	~~.52~~
	AG	~~.32~~	~~.44~~	~~.47~~	~~.47~~	~~.45~~
	AR	.58	.63	~~.54~~	~~.50~~	~~.48~~
	ALL	.71	.58	.58	.59	.62
T5	ST	**.82**	.61	.65	**.78**	~~.48~~
	AG	~~.31~~	~~.40~~	~~.49~~	~~.57~~	~~.51~~
	AR	.68	.62	**.74**	~~.56~~	~~.48~~
	ALL	**.76**	.64	.61	.64	.63
T6	ST	**.83**	~~.57~~	~~.51~~	.69	~~.47~~
	AG	~~.31~~	~~.34~~	~~.51~~	~~.41~~	~~.56~~
	AR	.65	~~.39~~	**.76**	~~.52~~	~~.53~~
	ALL	**.76**	~~.53~~	.67	.64	.64

We are now left with three 'interesting' levels for Target - ST, AR and ALL - and three 'interesting' levels for Others - No_Feat, AG, AR. Hence, our original design reduces to a 3×3×3 one and we can now turn to considering the accuracy values produced at each fold of our cross-validation procedure and submit them to a 3×3×3 repeated measures ANOVA. Significant effects are summarized in Table 4.

Table 4. Significant effects for Extraversions. (*=>p<.10, ** =>p<.05, * =>p<.01).**

Effect	F (df)
Time	9.033*** (1.585, 14)
Others	3.209* (1.717, 14)
Target*Others	3.813** (1.903, 12)

Starting from the main effect of Time, its marginal means are μ_{T2}=0.607, μ_{T5}=0.694 and μ_{T6}=0.671; pair-wise comparisons between them (with Bonferroni adjustment) have it that μ_{T2} is significantly lower than both μ_{T5} and μ_{T6}. Polynomial contrast analysis reveals both a significant linear (F=9.724, p<0.01) and quadratic effect (F=7.357, p<0.05). All this evidence converges towards showing that in general, classification performance tends to increase from 2 to 5 minute long slices, and then decreases.

Though weak (only p<.10), the main effect of Others is worth discussing. The marginal means are μ_{No_Feat}=0.726, μ_{AG}=0.629 and μ_{AR}=0.617. According to pairwise comparisons, μ_{No_Feat} is significantly higher than μ_{AR} (p<.10). As said, though only marginally significant, these data suggest that there is at least a tendency for better performance to come about when no contextual information is considered.

Finally, we turn to the 'Target*Others' effect. The relevant data are reported in Fig. 3.

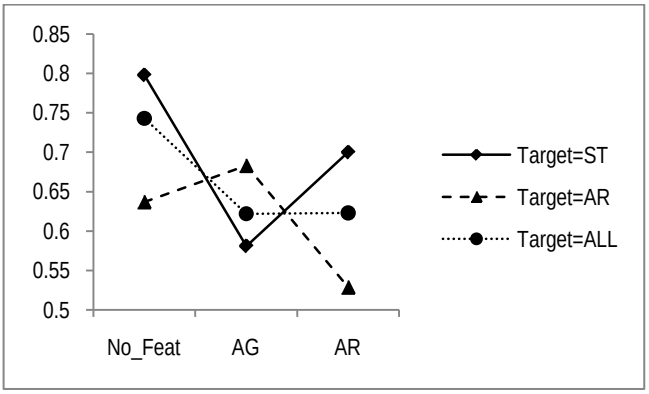

Figure 3. Marginal means for Target*Others. Categories for Others are on the x-axis.

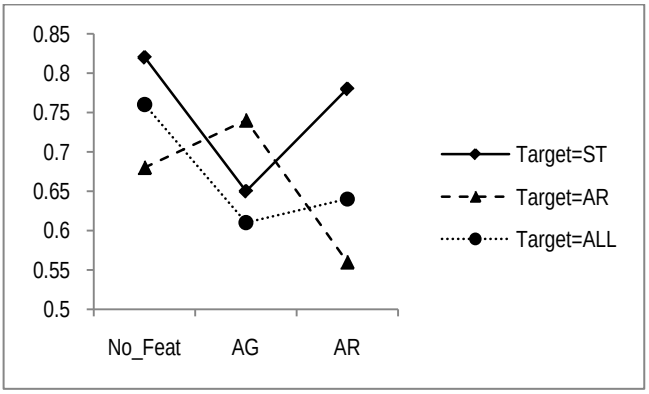

Figure 4. Accuracy at T5. Categories for Others are on the x-axis.

The most interesting results are: a) the supremacy of (*, ST, No_Feat) and (*, ALL, No_Feat); b) the fact that, though low everywhere else, Target=AR has a peak when Others=AG. These two data are best seen in Fig. 3 which focuses on results at Time=5 minutes (T5).

Taking stock of the above discussion, a number of conditions that were subjected to experimental evaluation turned out either not to yield accuracy values significantly higher than the baseline criterion or to do so only marginally. For (*, AG, *), it can be concluded that knowledge concerning the attention behavior of the target (how much attention he/she is devoting to the rest of the group) is not relevant to the task of getting at his/her Extraversion level. Similarly, encoding the social context by means of the speech activity of the other parties or by using all the available cues never results in interesting performance.

The repeated-measure ANOVA we have run on the restricted design have shown that: a) the size of the time slice used is important, with the best results obtained with 5 minutes long slices. b) No Target feature level among ST, AR and ALL is better than the others (no Target main effect). c) the usage of contextual information does not bring in any advantage. d) The best results are obtained with (*, ST, No_Feat) and (*, ALL, No_Feat). e) Although attention related features do not seem capable of improving performance, the combinations (T5, AR, AG) and (T6, AR, AG) give quite respectable accuracy values (.74 and .76 respectively). Noticeably, (*, AR, AG) describes in a complete way the attention-related behavior of the group with respect to the target.

Of some interest is also condition (*, ST, AR) which reaches 0.78 accuracy with 5-minute long slices. This value does not statistically differ from the accuracy value of 0.82 obtained by (T5, ST, No_Feat), and both are clearly superior to any other combination.

4.1 Error Analysis

We conclude this section by briefly discussing the way errors and hits distribute in the three best conditions considered above (T5, ST, No_Feat), (T5, AR, AG) and (T5, ST, AR). We do so in terms of Pearson residuals, computed for each of the three relevant observed confusion matrices with respect to the expected one for the baseline classifier (see discussion above). Pearson residuals are useful because being $N(0, 1)$ they allow for straightforward comparisons. They can be interpreted as follows: on hit categories, their absolute magnitude measures how much the relevant classifier does better (positive sign) or worse (negative sign) than the baseline; on error categories, the reverse is true. In general, values close to zero on a given category signal that the performance is similar to that of the baseline classifier. The results are reported in Table 5.

Table 5. Pearson residual for (T5, ST, No_Feat), (T5, AR, AG) and (T5, ST, AR) at Time=5'. Baseline is along columns and the classifiers' yields are on rows.

	ST, No_Feat		AR, AG		ST, AR	
	L	H	L	H	L	H
L	4.10	-4.61	5.13	-1.73	3.24	-4.46
H	-3.70	4.15	-4.61	1.59	-2.94	4.01

Two things are worth reporting: (T5, ST, No_Feat) fares better than (T5, AR, AG) and slightly better than (T5, ST, AR) with slices in the H category; (T5, AR, AG), in turn, does better on L, while exhibiting very poor performance on H. In other words, attention-related cues in (T5, AR, AG) seem to be biased towards low values of Extraversion, while speaking time produces a (lower) bias towards higher Extraversion values. In future work, it might be worth investigating whether a late fusion strategy combining the output of (T5, ST, No_Feat) and (T5, AR, AG) classifiers is able to improve accuracy above the values measured here.

We close this section on data analysis by reporting a few more data concerning the usage of the attention given to the person sitting directly opposite to the target subject. In section 3.2, we noticed that it has a significant negative correlation with the subject's Extraversion score. If we now use this information for classification (attention given to the opposite person=AGO) then a condition such as (T5, AGO, No_Feat) yields an accuracy of 0.80. Although we cannot increase the size of our design to accommodate for an additional level of Target and/or Others and perform tests of statistical significance (there would not be enough degrees of freedom), we think that this result is worth pointing to.

5. CONCLUSIONS

This work is devoted to the systematic analysis of the importance of a number of middle-level behavioral cues (speaking time, attention received by the rest of the group, attention given to the rest of the group and attention given to the person sitting right opposite) for the automatic classification of Extraversion. In doing so, we have adopted a thin slice perspective [1], investigating the effects of: a) slice dimension; b) the explicit encoding of the social context in terms of behavioral cues for the rest of the group. The results are very promising: speaking activity is a powerful behavioral cue; context representation is useless when the target behavior is represented in terms of the target's speaking activity or by means of the early multimodal fusion of the considered features. We also obtained some results that challenge naïve assumptions as well as data in the literature: though exhibiting a high correlation with Extraversion, the attention (social gaze) a target subject receives from the rest of the group is not powerful enough to provide good classification performance when used alone. However, when we turn from the gaze received by the target (*, AR, No_Feat) to account for the full social gaze behavior of the rest of the group, (*, AR, AG), then much better performances are obtained: in other words, *extraverts/introverts do not differ simply because of the amount of social gaze they receive, but because of the general gaze behavior they induce in the rest of the group*. Moreover, while the target subject's global gaze is ineffective as a cue, the amount of attention paid to the person sitting opposite yields a very good performance, a datum that could be explained with the mixed task and social oriented nature of attention in the context we have exploited. Finally, considering of all our features together, condition (*, ALL, *), did not yield results superior to those of the other conditions; hence, the simple multimodal strategy exploited here was not superior to the mono-modal ones. We can extend this conclusion to the representation of the social context: only in one case, (T5, ST, AR), the usage of features from different modalities for the target and his/her parties yielded an interesting performance. In all the other cases, either the social context is detrimental, as when Target=ST, or feature of the same modality are better, as with (T5, AR, AG) and (T6, AR, AG).

The difference between speaking behavior (as captured by speaking time) and attention-related behavior is worth some more discussion: the coincidence of our findings with those in the literature concerning the former can be interpreted as a sign that the extravert's/introvert's speaking behavior is relatively insensitive to the particular social context in which it is measured so that knowing how you speak is enough to know how much introvert/extravert you are. The predictive value of attention-related behaviors, in turn, seems to be more sensitive to the specific social context; it might be expected, therefore, that by changing the latter different results be obtained for gaze.

Before concluding, let us observe that the limited amount of subjects used in our study (16 subjects) might have limited the statistical power of our analysis, meaning that some effect might well have gone undetected. Those we could single out, however, are robust. Needless to say, the findings of this work need to be submitted to further verification, with larger and different data sets. Still, we believe that they shed important light on the prospects for the automatic analysis of a personality trait that occupies a central position in our social life.

6. ACKNOWLEDGMENTS

Bruno Lepri's research was funded by Marie Curie – COFUND – 7th Framework fellowship. The research of Ramanathan Subramanian and Nicu Sebe was partially supported by the S-PATTERNS FIRB project.

7. REFERENCES

[1] Ambady, N. and R. Rosenthal. 'Thin slices' of expressive behaviors as predictors of interpersonal consequences. A meta analysis. *Psychological Bulletin*. 111, 156-274, 1992.

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

[3] Argyle, M.1992. The social psychology of everyday life. Routledge, London. 1992.

[4] Blackman, M. C., and D. C. Funder. The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, 34, pp. 164–181, 1998.

[5] Blumenthal, T. D. Extraversion, attention and startle response reactivity. *Personality and Individual Differences*, 30, pp. 495-503, 2001.

[6] Carney, D. R., C. R. Colvin and J. A. Hall. A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41, pp. 1054-1072, 2007.

[7] Cassell, J., and Bickmore, T. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13, pp. 89–132, 2003.

[8] Costa, P. T. and R. R. McCrae. Four ways why five factors are basic. *Personality and Individual Differences*, 13, pp. 653-665, 1992.

[9] Costa, P. T., and McCrae, R. R. NEO PI-R Professional Manual. *Psychological Assessment Resources*, Odessa, FL., 1992.

[10] Donnellan, M., B., Conger, R. D., and Bryant, C. M. The Big Five and enduring marriages. *Journal of Research in Personality*, 38, pp. 481–504, 2004.

[11] Eaton, L. G.and D. C. Funder. The creation and consequences of the social world: An international analysis of extraversion. *European Journal of Personality*, 17, pp. 375-395, 2003.

[12] 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.

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

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

[15] Iizuka, Y. Extraversion, introversion and visual interaction. *Perceptual and Motor Skills*. 74, pp. 43-50, 1992.

[16] Ickes, W. J. Strangers in a strange lab: how personality shapes our initial encounters. *Oxford University Press*, New York, 2007.

[17] Komarraju, M., and Karau, S. J. The relationship between the Big Five personality traits and academic motivation. *Personality and Individual Differences*, 39, pp. 557–567, 2005.

[18] Lepri, B., Mana,N., Cappelletti,A., Pianesi, F., Zancanaro, M. Modeling Personality of Participants during Group Interaction. *In Proceedings UMAP – User Modeling, Adaptation and Personalization*. Trento, Italy, 2009.

[19] Mairesse, F., and Walker, M. Automatic recognition of personality in conversation. *In Proceedings of HLT-NAACL*, 2006.

[20] Mairesse, F., and Walker, M. Words mark 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.

[21] 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.

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

[23] 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*, International Conference on Multimodal Interfaces, Nagoya, Japan, 2007.

[24] Matthews, G. Extraversion. In A. P. Smith and D. M. Jones (eds.) *Handbook of Human Performance - Vol. 3: State and Trait*, pp. 95-126. Academic Press, 1992.

[25] Oberlander, J. and Nowson, S. Whose thumb is it anyway? Classifying author personality from weblog text. In *Proceedings of the Annual Meeting of the ACL*. Association for Computational Linguistics, 627-634, 1992.

[26] 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., and Perugini, M. (Eds.), Big Five Assessment, Hogrefe und Huber Publishers*, pp. 281-304, 2002.

[27] Pianesi F., Zancanaro M., Not E., Leonardi C., Falcon V., Lepri B. 2008.Multimodal Support to Group Dynamics. In Personal and Ubiquitous Computing, 12(2), 2008.

[28] Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., and Zancanaro. Multimodal Recognition of Personality Traits in Social Interactions. *In Proceedings of International Conference on Multimodal Interfaces*, Chania, Crete, Greece, 2008.

[29] Ramirez, J., Segura, J.C., Benitez, A., de la Torre, A., and Rubio, A. Efficient voice activity detection algorithms using long-term speech information. *Speech Communication*, 42(3-4): 271 – 287, 2004.

[30] Reeves, B., and Nass, C. The Media Equation. *University of Chicago Press*, 1996.

[31] Roberts, B W., N. R. Kuncel, R. Shiner, A. Caspi and L. R. Goldberg. The power of personality: The comparative validity of personality traits, socioeconomic status and cognitive ability for predicting important life outcome. *Perspectives on Psychological Science*, 2, pp. 313-345, 2007.

[32] Rutter, D. R., Morley, I. E., Graham, J. C.Visual interaction in a group of introverts and extraverts. *European Journal of Social Psychology*. 2, pp. 371-384, 1972.

[33] 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.

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

[35] Stelmak, R. M. The psychophysics and psychophysiology of extraversion and arousal. In *H. Nyborg (ed.) The scientific study of human nature: tribute to Hans J. Eysenck at eighty*, 388-403, 1991.

[36] Stenberg, G., I. Rosen and J. Riseberg. Attention and personality in augmenting/reducing of visual evoked potentials. Personality and Individual Differences, 11, pp. 1243-1254, 1990.

[37] Stiefelhagen, R., Yang, J., and Waibel, A. Modeling focus of attention for meeting indexing. *In ACM MM*, pp. 3–10, 1999.

[38] Stiefelhagen, R., Yang, J., and Waibel, A. Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks*, 13, pp. 928–938, 2002.

[39] Subramanian, R., Staiano, J., Kalimeri, K., Sebe, N., Pianesi, F. Putting the Pieces Together: Multimodal Analysis of Social Attention in Meetings, *ACM Multimedia*, 2010.

[40] Trull, T. J and K. J. Sher. Relationship between the five-factor model of personality and Axis I disorders in a nonclinical sample. *Journal of Abnormal Psychology*, 103, pp. 350-360, 1994.

[41] Valenti, R., and Gevers, T. Accurate eye center location and tracking using isophote curvature. *In IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[42] Valenti, R., Yucel, Z., and Gevers, T. Robustifying eye center localization by head pose cues. *In IEEE Conference on Computer Vision and Pattern Recognition*, 612–618, 2009.

[43] Xiao, J., Kanade, T., and Cohn, J. Robust full motion recovery of head by dynamic templates and re-registration techniques. *In IEEE Face and Gesture Recognition*, 2002.

[44] 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*, 2003.