Linking Speaking and Looking Behavior Patterns with Group Composition, Perception, and Performance

Dinesh Babu Jayagopi[1] and Dairazalia Sanchez-Cortes[1,2] and Kazuhiro Otsuka[3] and Junji Yamato[3] and Daniel Gatica-Perez[1,2]
[1] Idiap Research Institute, Martigny, Switzerland
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[3] NTT Communication Science Laboratories, Japan.
(djaya,dscortes,gatica)@idiap.ch, otsuka@eye.brl.ntt.co.jp, yamato@brl.ntt.co.jp

ABSTRACT

This paper addresses the task of mining typical behavioral patterns from small group face-to-face interactions and linking them to social-psychological group variables. Towards this goal, we define group speaking and looking cues by aggregating automatically extracted cues at the individual and dyadic levels. Then, we define a bag of nonverbal patterns (Bag-of-NVPs) to discretize the group cues. The topics learnt using the Latent Dirichlet Allocation (LDA) topic model are then interpreted by studying the correlations with group variables such as group composition, group interpersonal perception, and group performance. Our results show that both group behavior cues and topics have significant correlations with (and predictive information for) all the above variables. For our study, we use interactions with unacquainted members i.e. newly formed groups.

Keywords

Small groups; Nonverbal behavior; Group Mining

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Collaborative computing

1. INTRODUCTION

The automatic analysis of face-to-face group interaction integrates knowledge from signal processing, machine learning, and social psychology. The possibility of recording audio-visual data and extracting multimodal nonverbal cues helps to relate nonverbal behavior with group constructs such as composition, interpersonal perception, and performance [4, 19]. With the global workforce becoming increasingly team-based, understanding group processes, performance, and satisfaction measures has become relevant.

While building supervised learning models to infer interpersonal perception or performance from individual and group behavior has been studied in a few previous works [14, 7, 11, 19], characterizing group behavior in unsupervised frameworks to mine typical behaviors of groups is a relatively unexplored problem [8]. Such an approach can also yield intermediate representations that could help understand some group constructs in a holistic manner, because they describe groups based on co-occurence of observations, rather than directly choosing a measure like performance or satisfaction as it is done in supervised frameworks. The automatically extracted behavioral cues and intermediate representations could serve as complementary information to those measures extracted from self-reports, e.g. personality or interpersonal perception [14].

In this work, we address two research questions: First how to characterize and extract typical speaking and looking patterns; second, to understand how group behavior patterns relate to how group members perform and perceive themselves and other members of their group. In order to address these two questions, we define a data-centric framework as shown in Fig. 1. Specifically, we use LDA, a probabilistic topic model, to extract recurrent nonverbal patterns on five-minute segments of group interactions. We show that the extracted topics are socially meaningful by correlating them with group composition, interpersonal perception, and performance variables obtained from questionnaires. These variables relate to the group input-process-output model in social psychology literature [13]). For this study, we use the Emergent Leadership (ELEA) corpus [17]. This corpus contains, apart from the audio-visual recordings, comprehensive data about self-reported personality (Big-Five traits), interpersonal perception (of dominance, leadership, competence, and liking), and group performance.

Our paper has two contributions. First, mining speaking and gaze behavior patterns in groups is a relatively unexplored problem. While mining approaches have been recently used for audio cues [8], to our knowledge the group-level mining of both speaking and looking patterns has not been addressed. We define turn-taking and gaze-based group behavioral cues. In particular, the proposed gaze-based group cues are novel. The mining framework is powerful as it potentially allows to work with large datasets, make statistical inferences, and validate the results on datasets for which ground-truth is available. Second, relating the cues and the topics with multiple group-level constructs such as group composition, perception, and performance on a single dataset is also novel. Our findings show several interesting correlations between the personality of the group members, how they perceive each other, or how they perform, and the

ICMI'12, October 22–26, 2012, Santa Monica, California, USA.

automatically extracted group's speaking and looking patterns. The ELEA corpus allows us to study this aspect.

Section 2 presents related work. Section 3 introduces the data and the definition of questionnaire variables. Section 4 introduces the group nonverbal cues. Section 5 gives the background for the topic model. In Section 6, we present the correlations between the cues and the questionnaire data. Section 7 presents the topic-based analysis. Section 8 provides conclusions.

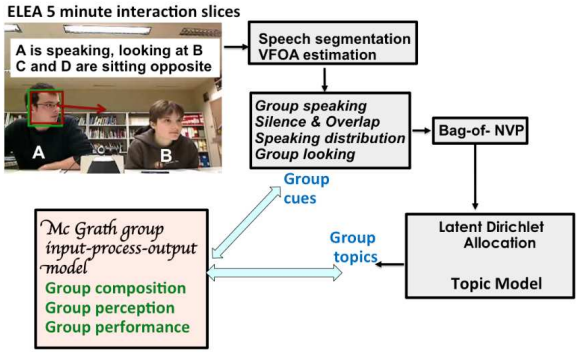

Figure 1: Our approach: Mining and validating group speaking and gaze patterns by defining a bag of nonverbal patterns and employing a topic model.

2. RELATED WORK

The works related to our study belong to either social psychology or social computing.

In social psychology literature, how group composition affects group processes or group performance has been extensively investigated across a variety of group tasks both in lab and field settings. In order to characterize and understand group processes and its effects, self-reports have been used predominantly, as manual annotation of behavior for large interaction datasets is laborious.

A classic framework to study groups was proposed by Mc Grath in 1964, termed the Input-Process-Output model [13]. Group input refers to the characteristics of the group members, including their personality, age, sex, or a formal role. Group process addresses the interpersonal relationship aspects between group members, both perceived as well as observed. This could include both the vertical ('getting ahead') and horizontal ('getting along') facets of relationships. Finally, the outputs could capture performance or satisfaction measured quantitatively or qualitatively. Recent works, such as [1], have validated and reinforced this approach. This study shows that the framework could be used to study teams accomplishing relatively modern tasks such as software development.

Depending on the task characteristics, the relationship between group composition, specifically group personality composition (GPC), and group performance has varied. For example, teams exhibit a significant positive correlation between the personality factor extraversion and software product quality [1]. GPC has been operationalized by considering mean, maximum, or minimum over individual personality scores [2]. The method that computes maximum or minimum score assumes that one person can have significant effect on the group's performance or other group outcomes.

Apart from the individual attributes, the interpersonal perceptions have also been used to predict group outcomes such as performance or satisfaction. Group cohesion and conflict (task or social) have been shown to affect group performance [2]. Also, how the personality of individuals affect group process has also been documented [2]. The relationships among perceived variables such as dominance and leadership or dominance and likeability [18] have also been studied.

In social computing literature concerning small groups, the fact that nonverbal cues are a rich source for automatic social inference about people's traits, or interpersonal perception or performance has been exploited. The existing studies include both individual and group-level inferences. Personality of individuals in groups was estimated in [14]. Formal roles were automatically estimated in [16]. Dominant behavior was inferred in [7] and emergent leadership was studied in [17]. Group level constructs such as group interest and group conversational context (for e.g. cooperative vs competitive behavior and brain-storming vs decision-making behavior) has been studied [7]. Leadership styles were discovered in [8], using a mining framework similar to this work, albeit using only audio cues. The topics were validated with perception annotation using external annotators (unlike our work here that uses self-reported questionnaires from the participants themselves).

The inference of performance in groups using nonverbal behavior has also received attention in recent years [19, 11]. The most comprehensive work been by Woolley et al. [19], in which the effect of collective intelligence (a novel way to characterize group composition) on group performance was studied on a range of cognitive tasks. It was also found that collective intelligence correlated with average social sensitivity of group members and the equality in distribution of conversational turn-taking, which relate to the group composition and processes respectively.

As compared to the literature in social psychology, our work adds the definition of automatic extraction group behavioral cues to the group variables studied. We also propose to mine the group behavior to extract recurrent patterns. Further, as compared to the existing works in social computing, our work explores jointly group composition, perception, and performance. We use both turn-taking and gaze cues. Only few works (such as [6, 12]) have explored gaze cues as compared to the turn-taking cues. Finally, unsupervised approaches to characterize group behavior is still a relatively unexplored direction in group inference [8].

3. INTERACTION DATASET

3.1 Data

We use 18 group interactions from the ELEA corpus [17] to extract and mine group nonverbal behavior and relate it to group-level social psychological constructs (see Fig. 1 for an overview of our approach). Each interaction has 4 unacquainted participants solving the winter survival task [10]. Kickul et al. claim that "the winter survival task allows team members to observe the social or relational skills of others while interacting in this problem solving task." The participants filled a questionnaire about themselves (i.e. their personality using the NEO-FFI inventory [9]) before the interaction started. They were instructed to come up with a ranked list individually before discussing with others and then come up with a joint list after a fifteen-minute discussion. The ranked list would contain twelve items important for surviving a airplane crash in harsh winter conditions.

Later, each of them filled a questionnaire about the behavior of their group members.

The room set-up had a rectangular table, with two people on either side. A Microcone, a commercial array microphone, was used to record the audio and obtain the speech segmentations of participants. The audio sample rate was 16kHz. For video recordings, a portable setup with two webcameras (Logitech R Webcam Pro 9000, 30 fps, 640 by 480) was used. The average duration of an interaction was 14 minutes 8 seconds. The dataset has 4.4 hours of interaction data in total.

3.2 Questionnaire variables

In our work, we operationalize group composition as measured by the personality of individual members. We study group interpersonal perception using dominance, leadership, competence, and liking. Group performance is objectively measured by comparing the joint ranking proposed by the group with that of the experts. Next, we describe the questionnaire data and then define the variables of interest at the level of groups.

Composition (Personality): Personality attempts to capture individual differences i.e. the dimensions along which people differ from each other, and at the same time describe individual persons as unique, integrated wholes [9]. Big Five factors of personality - Agreeableness, conscientiousness, Extraversion, Neuroticism, and Openness to experience have been shown to be a parsimonious and cross-culturally valid approach to personality [9]. We used the NEO-FFI questionnaire commonly used in personality research which has 60 statements measuring these five factors.

Interpersonal perception: After the survival test, participants were asked to answer 16 statements that capture how they perceived each participant. This questionnaire was designed adapting existing questionnaires in leadership with the help of a social psychologist. 16 of the statements were evaluated on a five-point scale. The variables included in these statements are: Perceived Dominance (denoted PDom: dominates, is in a position of power, asserts him- or herself), Perceived Leadership (denoted PLead: directs the group, imposes his or her opinion, is involved), Perceived Competence (denoted PCom: is competent, is intelligent, has a lot of experience), and Perceived Liking (denoted PLike: is kind, friendly, not unpleasant).

Performance: The performance of the group (denoted $GPerf$) was measured by the negative distance between an 'ideal' list created by experts and the list proposed by a group. Therefore higher the performance score, better performing the group is. Zero being the highest score.

Defining group-level variables

Performance is defined at the level of groups. We need to define personality and perception about interaction partners at the group level as well. We compute the *mean*, *maximum*, and *range* (maximum - minimum) over the personality score of the individuals in a group to characterize 'group composition'. This approach is quite standard in the small group literature [5]. Perceived constructs - PDom, PLead, PComp, and PLike are defined at the dyadic level. For example $PDom(i, j)$ denotes what i perceives about j (Note: $PDom(i, i)$ is not defined). Our first four variations are defined by averaging over i, the row variable of PDom.

$$ARDom(j) = \frac{\Sigma_{i=1,i\neq j}^{P} PDom(i,j)}{P-1} \quad (1)$$

$ARDom(j)$ is what the group (on an average) says about j. Next, we compute *mean*, *maximum*, *toprange* (maximum - second maximum), *range* over ARDom(j) to characterize the perception about the group. *maximum* is abbreviated as $max \rightarrow$.

The final variation is computed by averaging over j.

$$ACDom(i) = \frac{\Sigma_{j=1,j\neq i}^{P} PDom(i,j)}{P-1} \quad (2)$$

$ACDom(j)$ is what j says about the group (on an average). Next, we compute maximum over ACDom(j), denoted $max \leftarrow$. While $max \rightarrow$ is related to what group perceives about an individual, $max \leftarrow$ is related to what an individual perceives about the group. Also, *mean* is what the group perceives about the group. Finally, *toprange* and *range* is what the group perceives about two individuals, the difference between them to be exact. Similar to PDom, we compute these variations for PLead, PComp, and PLike as well. In total, we have defined 36 group-level variables i.e. group composition(15), group perception (20), and group performance (1) from the questionnaire data.

4. GROUP NONVERBAL CUE EXTRACTION

Nonverbal cues, particularly turn-taking and gaze patterns, are known to reveal social-psychological constructs such as traits [14], interpersonal perceptions [7] and performance [11, 19]. Below, we describe and define the speaking and looking cues that capture a group's turn-taking and gaze behavior in multiple ways.

4.1 Group speaking cues

Individual cues: The Microcone outputs the speaking status - a binary variable indicating speaking (1) and non-speaking (0) - of each of the participant. The speaking status was downsampled to a rate of 5 frames per second (AFps). This rate is sufficient to analyze conversational behavior at the level of turns. Short conversational events, for example backchannels are of the order of 1 or 2 second duration.

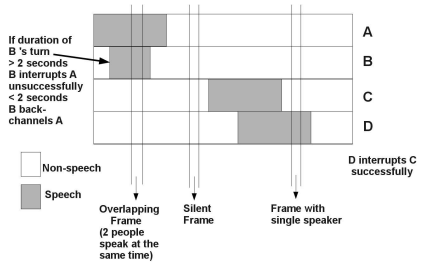

Figure 2: Turn-taking based nonverbal cues.

From the speech segmentation, we compute individual speaking length, speaking turns, successfully interrupting, being unsuccessfully interrupted, and being backchanneled defined below:

Speaking Length (sl): Cumulates the total time that a person speaks according to their binary speaking status.

Speaking Turns (st): Cumulates speaking turns, where a speaking turn is an interval in time for which a person's speaking status is active.

Successfully Interrupting (si) : The feature is defined by the cumulative number of times that speaker $i \in \{1, 2, 3, 4\}$

starts talking while another speaker $j \in \{l : l \neq i\}$ speaks, and speaker j finishes his turn before i does,
Being Unsuccessfully interrupted (ui) : The feature is defined by the cumulative number of times that speaker $i \in \{1, 2, 3, 4\}$ unsuccessfully interrupts (by speaking atleast for 2 seconds) another speaker $j \in \{l : l \neq i\}$
Being Backchanneled (bc) : The feature is defined by the cumulative number of times that speaker $i \in \{1, 2, 3, 4\}$ speaks less than 2 seconds while another speaker $j \in \{l : l \neq i\}$ is talking.

Fig. 2 illustrates the conversational cues at the level of individuals. From these individual cues, three types of group conversational cues are extracted.

Group participation cues: A first set of cues characterize the participation rates of the group by accumulating it over the participants. Let D denote the duration of the meeting in seconds. We compute the following five cues - Speaking Length (SL), Speaking Turns (ST), Successful Interruptions (SI), Unsuccessful Interruptions (UI), Backchannels (BC) - from the individual cues. For e.g. $SL = \Sigma_{i=1}^{P} sl(i)/D$

Silence and Overlap cues: A second set of cues attempts to capture the overlap and silence patterns of a group as a whole. Let $AF = D * AFps$ be the total number of frames in a meeting, S be the number of frames when no participant speaks, M be the number of frames when only one participant is speaking, $O2$ and $O3$ be the number of frames when more than two or three participant talk at the same time. $AFps$ being frames-per-second, the rate at which speaking status is available. Then we define the following four cues - Fraction of Silence(FS), Fraction of Non-overlapped Speech(FN), Fraction of two-people and three-people Overlapped Speech(FO2 and FO3) - defined as follows: $FS = \frac{S}{AF}$, $FN = \frac{M}{AF}$, $FO2 = \frac{O2}{AF}$, $FO3 = \frac{O3}{AF}$.

Speaking distribution cues: A third set of cues characterizes which meeting is more 'egalitarian' with respect to the use of the speaking floor i.e. everyone gets equal opportunities. Let **sl** denote the vector composed of P elements, whose elements are $sl(i)/\Sigma_i sl(i)$ for the ith participant. Employing an analogous notation for **st**, **si**, **ui** and **bc**, these vectors are first ranked (**p**) and then compared with the uniform (i.e. "egalitarian") distribution i.e. a vector of the same dimension with values equal to $\frac{1}{P}$ (**q**). The comparison is done using the Hellinger distance, a measure useful to compare probability distributions and bounded between 0 and 1. The Hellinger distance is defined in terms of the Bhattacharya coefficient as follows: $HD(\mathbf{p}, \mathbf{q}) = \sqrt{1 - BCoeff(\mathbf{p}, \mathbf{q})}$ and $BCoeff(\mathbf{p}, \mathbf{q}) = \Sigma_i \sqrt{p(i) * q(i)}$. Hellinger distance of 0 would correspond to a egalitarian meeting. This results in five cues: Speaking Length Skew (SLS), Speaking Turns Skew (STS), Successful Interruption Skew (SIS), Unsuccessful Interruptions Skew (UIS), Backchannels Skew (BCS).

4.2 Group looking cues

The visual focus of attention (i.e. 'what a participant is looking at') is denoted VFOA and was estimated using the head pose angle as the resolution to track the eye gaze was not enough. The head pose is estimated following the technique described in [15]. Tracking and pose recognition are treated as two coupled problems in a dynamic, probabilistic framework. The method uses a formulation using particle filters, with the state space accounting for the location and scale of the head as well as discretized head pose. The observation model uses both texture [based on Histograms of Oriented Gradients (HOG)] and color features. The left image in Fig. 3 shows the tracker output location which is computed as the mean (in green color) and median (in red color) of all particle filter outputs. The top right part of Fig. 3 shows the estimated pan and tilt head pose angles. The video sampling rate was 30 frames per second (VFps).

Figure 3: Tracking, head-pose estimation, and VFOA estimation for a meeting participant.

Individual cues: Using the head-pose (considering pan and tilt only and not the roll), the VFOA is later estimated using the Maximum Aposteriori (MAP) rule. The MAP rule assumes a Gaussian distribution with mean and standard deviation prespecified manually (in the pan and tilt space), for each of the 5 targets T1 to T5. The bottom right part of Fig. 3 shows the estimated VFOA target. T1, T2 are the participants sitting opposite to the participants shown. T3 is the participant sitting next to the tracked participant. T4 and T5 represent the table area close to tracked participant and participant T3. UN stands for unfocussed.

In order to assess the accuracy, we carried out the annotations for VFOA. For every 15 second, the VFOA of every participant was manually annotated using one annotator for one randomly chosen meeting. This resulted in $61 * 4 = 244$ ground truth samples in total for all four participants. The automatic method had an accuracy of 42% when compared to the manual annotation. The estimation errors were mainly due to tracking failures (which were typically due to background color effects or illumination issues) or inaccuracies in head-pose estimation. Typical VFOA accuracies in group interactions are in this order.

Group Looking cues: From the individual gaze cue i.e. 'what is the visual target of each of the participant', the following group cues are defined, in order to characterize the gaze behavior of the group. Fig. 4 illustrates the visual target and three group cues for an interaction with four participants.

Fraction of People Gaze (FPG): Let $VF = D * VFps$ be the total number of video frames and P1, P2, P3, P4 be the number of frames in an interaction where a participant $i \in \{1, 2, 3, 4\}$ is being looked at. Then $FVTP = \frac{P1+P2+P3+P4}{4VF}$. This feature captures the intution that some groups look at people a lot while some others look at table a lot.

Fraction of Convergent Gaze (FCG): Let C be the number of frames when a participant $i \in \{1, 2, 3, 4\}$ is being looked at by all the other participants $j \in \{l : l \neq i\}$, $FCG = \frac{C}{VF}$.

Fraction of Mutual gaze (FMG): Let MG be the number of frames when two participants look at each other i.e. for a participant $i \in \{1, 2, 3, 4\}$ the visual target is $j \in \{l : l \neq i\}$ and for the participant j the visual target is i, $FMG = \frac{MG}{2VF}$.

Fraction of Shared gaze (FSG): Let SG be the number of frames when two participants look at the same participant.

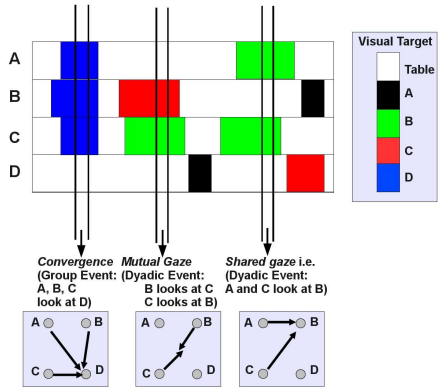

Figure 4: Visual targets and group looking cues

i.e. for a participant $i \in \{1, 2, 3, 4\}$ and $j \in \{l : l \neq i\}$ the visual target is k, $k \in \{l : l \neq i, j\}$, $FMG = \frac{SG}{2VF}$ (see Fig. 4 for an illustration of FCG, FMG, and FSG).

Gaze Skew (GS): This feature follows the definition of distribution measures in audio. First the vector, Total Attention Length (TAL) composed of [P1, P2, P3, P4] is formed, normalized to one, and then compared with the egalitarian vector to obtain a group measure (Definition of this cue follows the speaking distribution cues).

5. TOPIC MODELING OF NONVERBAL GROUP CUES

In order to describe groups using a discrete probabilistic framework, we define the bag-of-NVPs by quantizing the group cues and then employ LDA based topic modeling. Working in the discrete domain helps to describe group behavior in intuitive categorical terms. The categories themselves were constructed using a data-centric approach described in the subsequent subsection. Finally, clustering the group cues in the discrete domain using LDA allows for a probabilistic interpretation of typical group behavior.

5.1 Bag-of-NVP definition

We define our documents as five-minute meeting slices, 163 of them obtained by slicing the 18 group interactions with an overlap of 80%. After computing the group cues in continuous domain, we discretize them to produce a bag-of-NVPs. This quantization was done using K-means procedure for every group cue. We set $K = 5$, rank each of the clusters and assign one of the 5 words corresponding to (HIGHEST, HIGH, AVERAGE, LOW, LOWEST) and color coded with (BLACK, BLUE, GREEN, MAGENTA, RED) for displaying in Figures 5-8.

5.2 LDA based topic extraction

We use LDA, a topic model to cluster our group cues in a discrete space. Topic models are co-occurrence based probabilistic generative models that were originally used in text modeling. In LDA [3], a text document is modeled as a distribution over topics, and a topic as a multinomial distribution over words.

Let there be D documents in a corpus and let a document contain N_d words. Let V denote the total number of unique words in the corpus. The probability of a given word w_i assuming T topics is $p(w_i) = \sum_{t=1}^{T} p(w_i|z_i = t)P(z_i = t)$, where z_i is a latent variable indicating the topic from which the i^{th} word was drawn. Each topic is characterized by a word distribution $p(w|z = t) = \phi_w^{(t)}$ over the vocabulary of words V. Each document is generated by choosing a distribution over topics $p(z = t|d) = \theta_t^{(d)}$. When multiple slices of interaction are available for a particular chosen group g, $d \in D_g$, $p(z|g)$ can be computed by averaging over the multiple slices. This distribution can then be used to characterize and compare groups.

6. CORRELATION ANALYSIS

In this section, we study the correlation among the group behavior cues and the questionnaire variables (group composition, perception, and performance). All variables are 18 dimensional vectors, corresponding to the number of groups. Table 1 gives the significant correlations. For all the results reported below, we compute the Pearson correlation coefficient, hence reporting the r and p values. p-values for Pearson's correlation are computed using a Student's t distribution for a transformation of the correlation. Correlations with $p < 0.05$ are denoted by $*$ and $p < 0.01$ by $**$ in the superscript.

The group participation cues capture one aspect of group composition (extraversion *mean*) and one aspect of group perception (the difference in perceived dominance between top two individuals). Extraversion *mean* had a correlation of 0.56, 0.58, and 0.49 with SL, SI, UI. This means extraverted groups talk more and interrupt more. PDom *toprange* had a correlation of -0.69, -0.66, and -0.55 with ST, BC, and SI. This implies groups with a top-two dominance hierarchy have lesser turns, interruptions, and backchannels.

The silence and overlap cues captured three aspects of group composition (extraversion *mean*, extraversion *max*, and agreeableness *max*) and one aspect of group perception (PDom *toprange*). Extraversion *mean* has a correlation of 0.53 with FO2 and PDom *toprange* has a correlation of -0.64 with FO3. Extraverted groups or groups with lesser hierarchy among top two dominant people have higher overlapping speech. Groups with a disagreeable or an introverted person have a higher fraction of silence, FS (Agreeableness *max* and Extraversion *max* have a negative correlation of -0.57 and -0.5 with FS)

The speaking distribution cues charaterizing the skew in conversational opportunities, expectedly, capture the *range*, *toprange*, and *max* aspects of the perceived constructs. No significant correlation with group composition was found. One of the cue related to group performance. For example, PDom *toprange* had a significant correlation of 0.54, 0.51, 0.55 with SLS, STS, and BCS i.e. groups with top-two dominance hierarchy have unequal speaking time, turns, and backchannels distribution among group members. PLead *toprange* was found to correlate with BCS (r = 0.57). This indicates that groups with a skew in the distribution of backchannels have an emergent leader among the top two scorers on leadership. Group performance had a significant correlation with unsuccessful interruption skew.

The looking cues showed some complementary correlations w.r.t the speaking cues. Several correlations with group perception were found and only one cue captured group composition (Extraversion *max*) and the group performance. While audio cues showed correlation with PDom *max* $\rightarrow$, many of the visual cues were correlated with PDom *max* $\leftarrow$. Particularly, FPG, FCG, GS had correlations of 0.59, 0.52, and 0.61. This implies the likelihood of an individual perceiving the group as dominating is high when there was more

people-gaze, convergent gaze, and gaze skew (i.e. everyone are not being looked at equally). The fraction of convergence indicates groups with a competent person and performing groups (FCG had a correlation of 0.63 and 0.57 with PComp *max* $\rightarrow$ and GPerf). Groups with an extraverted person have higher mutual gaze (FMG had a correlation of 0.63 with Extraversion *max*). Group with gaze skew report higher mean dominance score for their team members i.e. in dominant groups everyone is not being looked at equally (PDom *mean* had a correlation of 0.49 with GS).

Group cue	Group variable	r
Group participation cues		
Speaking Length (SL)	Extrav. *mean*	$+0.56^{*}$
Speaking Turns (ST)	PDom *toprange*	-0.69^{**}
Successful Interruptions (SI)	Extrav. *mean*	$+0.58^{*}$
	PDom *toprange*	-0.55^{*}
Unsucc. Interrup. (UI)	Extrav. *mean*	$+0.49^{*}$
Backchannels (BC)	PDom *toprange*	-0.66^{**}
Silence and overlap cues		
Fr. of Silence (FS)	Agreeab. *max*	-0.57^{*}
	Extrav. *max*	-0.50^{*}
Fr. of Non-overlapped Speech (FN)	PLike *toprange*	$+0.53^{*}$
Fr. 2-people Ovp. (FO2)	Extrav. *mean*	$+0.53^{*}$
Fr. 3-people Ovp. (FO3)	PDom *toprange*	-0.64^{**}
Speaking distribution cues		
Speaking Length Skew (SLS)	PDom *max* $\rightarrow$	$+0.51^{*}$
	PDom *toprange*	$+0.54^{*}$
	PLead *max* $\rightarrow$	$+0.59^{*}$
	PLead *range*	$+0.58^{*}$
Speaking Turns Skew (STS)	PDom *toprange*	$+0.51^{*}$
	PLead *max* $\rightarrow$	$+0.54^{*}$
Successful Interruptions Skew (SIS)	PLead *max* $\rightarrow$	$+0.59^{*}$
	PLead *range*	$+0.51^{*}$
Unsucc. Interruptions Skew (UIS)	PDom *range*	$+0.53^{*}$
	PComp *max* $\rightarrow$	$+0.59^{*}$
	PComp *range*	$+0.58^{*}$
	GPerf	$+0.52^{*}$
Backchannels Skew (BCS)	PDom *toprange*	$+0.55^{*}$
	PLead *toprange*	$+0.57^{*}$
Group looking cues		
Fr. of People Gaze (FPG)	PDom *max* $\leftarrow$	$+0.59^{*}$
	PLike *toprange*	$+0.51^{*}$
Fr. of Convergent Gaze (FCG)	PDom *max* $\leftarrow$	$+0.52^{*}$
	PComp *max* $\rightarrow$	$+0.63^{**}$
	PLike *range*	$+0.54^{*}$
	GPerf	$+0.57^{*}$
Fr. Mutual Gaze (FMG)	Extrav. *max*	$+0.50^{*}$
	PLead *max* $\leftarrow$	$+0.58^{*}$
Fr. Shared Gaze (FSG)	PLead *max* $\leftarrow$	$+0.50^{*}$
Gaze Skew (GS)	PDom *mean*	$+0.49^{*}$
	PDom *max* $\leftarrow$	$+0.61^{**}$
	PComp *range*	$+0.53^{*}$
	PLike *max* $\leftarrow$	-0.58^{*}

Table 1: Correlation between group cues and questionnaire variables ($**$: $p < 0.01$, $*$: $p < 0.05$)

7. TOPIC-BASED ANALYSIS

For the 163 meeting slices from 18 groups, we systematically study the effect of the data representation on the extracted topics i.e. the Bag-of-NVPs with each set of group cues. The vocabulary size V is five times (corresponding to the five categories from 'highest' to 'lowest') the number of group cues employed to build the bags. We used a symmetric Dirichlet distribution with $\alpha = 1$ and $\beta = 0.01$. We experimented with a small number of topics T, due to limited number of meeting slices or documents. For space reasons, we report the results with $T = 4$ only.

Figures 5-8 shows the top document for each of the four topics as illustration. Table 5-8 documents the correlation between $p(z|g)$ and the questionaire variables (each being a 18 dimensional vector) for various bag representations. Link to couple of demos here: `www.idiap.ch/~djaya/ICMI2012/`

7.1 Topics with group participation cues

The extracted topics with group participation cues are illustrated in Fig. 5 by their top document and in Table 2 by the questionnaire group variables with whom they have significant correlations. Topic 2 represents a group that is extroverted and the group members like each other (correlation with extraversion *mean* is 0.50^{*} and PLike *mean* is 0.47^{*}). There is no one perceived to be dominant (correlation with PDom *max* $\rightarrow$ is -0.53^{*}). Looking at Fig. 5, we see that this group takes more turns, interrupts, and backchannels a lot (i.e. high ST, SI, BC). It is interesting to contrast Topic 2 with Topic 4 in which groups talk even more. Although this topic did not have significant correlations with any variables. Topic 1 captures a group that has a clear hierarchy between the top two dominant people (correlation with PDom *toprange* is 0.68^{**}) and top two liked people (correlation with PLike *toprange* is 0.48^{*}). The group does not talk or interrupt much.

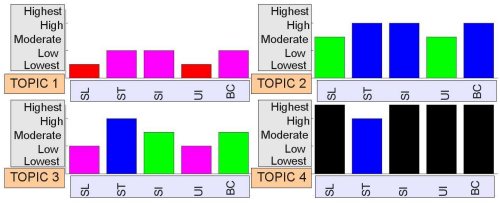

Figure 5: Top documents of four topics using group participation cues

Group cue set (SL, ST, SI, UI, BC)			
Construct	r	Construct	r
TOPIC 1		TOPIC 2	
PDom *toprange*	$+0.68^{**}$	Extrav. *mean*	$+0.50^{*}$
PLike *toprange*	$+0.48^{*}$	PDom *max* $\rightarrow$	-0.53^{*}
		PDom *toprange*	-0.60^{**}
		PDom *range*	-0.47^{*}
		PLead *max* $\leftarrow$	$+0.47^{*}$
		PLike *mean*	$+0.47^{*}$
TOPIC 3		TOPIC 4	
Extrav. *max* $\rightarrow$	-0.52^{*}		
PLead *max* $\leftarrow$	$+0.63^{**}$		

Table 2: Correlation between the group participation topics and perceived questionnaire variables

7.2 Topics with silence and overlap cues

Fig. 6 shows the top documents and Table 3 lists the significant correlations. Topic 1 represents an agreeable group

in which the group members like each other. Correlation with agreeableness *mean* is 0.52^{*} and PLike *mean* is 0.63^{**}. It is to be noted that while group agreeableness is a self-reported quantity, group liking is perception about others. In these groups, the number of silent frames is lowest, with moderate non-overlapping and overlapping speech (see Fig. 6). Topic 4 represents a group that is extroverted and dominating (unlike the Topic 2 with group participation cues in Section 9.1 which is extroverted and group members like each other). Correlation with PDom *mean* is 0.55^{*} and Extraversion *mean* is 0.52^{*}. Topic 3 captures a group with a hierarchy between the top two members (correlation with PDom *toprange* is 0.54). Interestingly, the group members like each other less (correlation with PLike *mean*: -0.50^{*}). Nonoverlapping speech (FN) is highest, which means a lot of monologues (see Fig. 6).

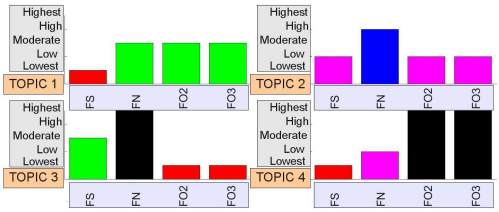

Figure 6: Top documents - silence and overlap cues

Group cue set (FS, FN, FO2, FO3)			
Construct	r	Construct	r
TOPIC 1		TOPIC 2	
Agreeab. *mean*	$+0.52^{*}$		
PDom *max* $\rightarrow$	-0.53^{*}		
PDom *range*	-0.54^{*}		
PDom *max* $\leftarrow$	-0.48^{*}		
PLike *mean*	$+0.63^{**}$		
PLike *toprange*	-0.49^{*}		
PLike *range*	-0.53^{*}		
TOPIC 3		TOPIC 4	
PDom *toprange*	$+0.54^{*}$	Extrav. *mean*	$+0.52^{*}$
PLike *mean*	-0.50^{*}	PDom *mean*	$+0.55^{*}$
PLike *toprange*	$+0.69^{**}$	PDom *max* $\leftarrow$	$+0.50^{*}$

Table 3: Correlation between the silence and overlap topics and perceived questionnaire variables

7.3 Topics with speaking distribution cues

Fig. 7 shows the top documents and Table 4 lists the significant correlations. Topic 1 corresponds to a group with a disagreeable person (correlation with agreeableness *max* $\rightarrow$ is -0.46^{*}), but no emergent leader (correlation with PLead *max* $\rightarrow$ is -0.53). Everyone in the group contributes to the conversations as the distribution measures indicate equality i.e. skew is lowest (see Fig. 7). Topic 3 is where the distribution measures indicate inequality (high skew) and in these groups there is a hierarchy. The correlation with PDom *toprange* is 0.55 and correlation with PLead *range* is 0.6.

7.4 Topics with group looking cues

The results are shown in Fig. 8 and Table 5. The group corresponding to Topic 1 is introverted. Correlation with extraversion *mean* is -0.55^{*} and *max* $\rightarrow$ is -0.47^{*}. Very

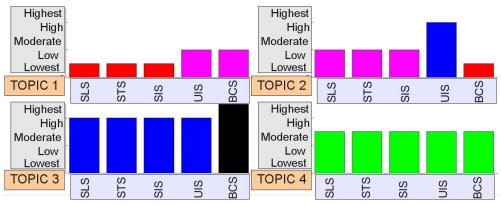

Figure 7: Top documents - speaking distribution cues

Group cue set (SLS, STS, SIS, UIS, BCS)			
Construct	r	Construct	r
TOPIC 1		TOPIC 2	
Agreeab. *max* $\rightarrow$	-0.46^{*}	PDom *toprange*	-0.47^{*}
PLead *max* $\rightarrow$	-0.53^{*}	PLike *max* $\leftarrow$	-0.48^{*}
TOPIC 3		TOPIC 4	
PDom *toprange*	$+0.55^{*}$		
PDom *range*	$+0.48^{*}$		
PLead *max* $\rightarrow$	$+0.48^{*}$		
PLead *toprange*	$+0.60^{**}$		

Table 4: Correlation between the speaking distribution topics and perceived questionnaire variables

interestingly, the table and not people, is the main visual target, with lowest convergence and mutual gaze (see FCG and FMG in Fig. 8 top left). But everyone is being looked at equally (Gaze skew, GS is lowest). Furthermore, the likelihood of someone perceiving the group members as dominant or taking the lead is less (correlation with PDom *max* $\leftarrow$ is -0.54^{*} and PLead *max* $\leftarrow$ is -0.48^{*}). The groups corresponding to Topic 2 are in direct contrast. People look at each other a lot, with high mutual and shared gaze (see FMG and FSG in Fig. 8 top left). But the likelihood of someone perceiving the group members as dominant is high (correlation with PDom *max* $\leftarrow$ is 0.62^{**}). Furthermore, the group members do not like each other. Topic 3 represents a case where group looking is moderate, the correlation with performance is -0.45 ($p = 0.06$). So this is one signature of a good performing group.

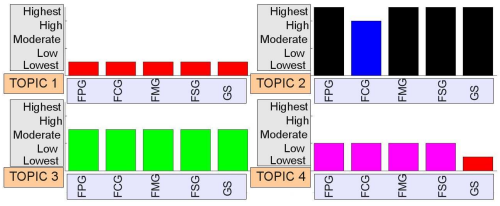

Figure 8: Top documents - group looking cues

7.5 Regression experiments

We used a step-wise linear regression procedure to measure the power of the group cues to predict the questionnaire variables, by combining group cues two sets at a time. We also experimented with group topics. We report the best results here. Regarding group composition, the combination of Group participation cues and Silence-Overlap cues could predict 77% ($F = 11.2$; $p = 0.0004$) of the variance in Agreeableness *max*. Regarding perception, the combination of Speaking distribution cues and Silence-Overlap cues could predict 67% ($F = 15.3$; $p = 0.0009$) of the variance in PDom

Group cue set (FPG, FCG, FMG, FSG, GS)			
Construct	r	Construct	r
TOPIC 1		TOPIC 2	
Extrav. *mean*	-0.55^{*}	PDom *max* $\leftarrow$	$+0.62^{**}$
Extrav. *max* $\rightarrow$	-0.47^{*}	PLead *max* $\leftarrow$	$+0.47^{*}$
PDom *max* $\leftarrow$	-0.54^{*}	PComp *range*	$+0.50^{*}$
PLead *max* $\leftarrow$	-0.48^{*}	PLike *mean*	-0.50^{*}
		PLike *toprange*	$+0.52^{*}$
TOPIC 3		TOPIC 4	
PLead *max* $\leftarrow$	$+0.63^{**}$	Op. Exp. *mean*	-0.49^{*}
GPerf	+0.45		

Table 5: Correlation between the group looking topics and perceived questionnaire variables

toprange. Using the topics, a new group composition variable (apart from agreeableness and extraversion) could be modeled. The topics with group looking cues could explain 42% (F = 5.5; p = 0.0089) of the variance in Openness to experience *mean*. Regarding group perception, the topics with Silence-Overlap cues could explain 50% (F = 8.0; p = 0.0054) of the variance in PDom *mean* and 39% (F = 10.3; p = 0.0054) of the variance in PLike *mean*. Group topics show complementary properties compared to the group cues. for certain constructs.

8. CONCLUSION

In this work, we presented a framework to define and extract group behavioral cues characterizing the speaking and looking patterns in face-to-face interactions. We mined these patterns using different bag representations and LDA. First, the relationship between the group constructs and the group cues was documented. Later the correlation with group topics was presented. Group interactions with unacquainted participants, having no prior hierarchy, were used for the study. The group variables were defined using the self-reported questionnaires.

Our study shows multiple significant connections between nonverbal features and variables characterizing the group composition, interpersonal perception, and performance. To summarize, group composition variables such as the average and maximum extraversion of group members, and maximum agreeableness found in the group were significantly related to group participation and the silence-overlap cues. Group perception, like the dominance hierarchy among the top members, and the perceived leadership were captured by many of the group cues. Group performance was captured by two group cues (unsuccessful interruptions skew and convergent gaze). Apart from these trends, it was interesting to observe that higher convergent gaze indicated the presence of a competent group member; that groups with a highly extroverted person have higher mutual gaze; and that gaze skew signalled groups with higher average dominance.

The discovered group topics show certain complementarity as compared to the raw group cues. For instance, while no group cue had significant correlation with mean group liking, a topic with silence-overlap cues correlated well with this construct. It is interesting to note that the topics captured dominance, leadership, and liking aspects more than competence and performance. Also, regarding group composition, the group topics captured the extraversion and agreeableness aspects.

Finally, the regression experiments showed that the group cues and group topics could predict certain constructs, for example the hierarchy between the top two dominant people, and the presence of a highly disagreeable person. For the future, we need to increase the corpus size, and perform classification experiments and model selection to choose the number of topics automatically. Fine-grained gaze cues and speaking cues that capture the dynamics of the interaction over time could also be explored.

Acknowledgments: This research was funded by the NISHA collaborative project between Idiap and NTT, and was also partly supported by the SNSF SONVB project and the EU HUMAVIPS project. We also thank Jean-Marc Odobez (Idiap) for sharing code for VFOA estimation.

9. REFERENCES

[1] S.T. Acuña et al. How do personality, team processes and task characteristics relate to job satisfaction and software quality? *Information and Software Technology*, 51(3):627–639, 2009.

[2] B. Barry and G.L. Stewart. Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology*, 82(1):62, 1997.

[3] D. M. Blei et al. Latent Dirichlet Allocation. *J. Machine Learning Research*, 3:993–1022, January 2003.

[4] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: a review. *IVC*, 2009.

[5] T. Halfhill et al. Group personality composition and group effectiveness. *Small Group Research*, 36(1):83–105, 2005.

[6] H. Hung et al. Investigating automatic dominance estimation in groups from visual attention and speaking activity. In *Proc. ICMI*, pages 233–236. ACM, 2008.

[7] D. Jayagopi. *Computational modeling of face-to-face social interaction using nonverbal behavioral cues*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2011.

[8] D. Jayagopi and D. Gatica-Perez. Mining group nonverbal conversational patterns using probabilistic topic models. *IEEE Trans. Multimedia*, 12(8):790 – 802, 2010.

[9] O. P. John et al. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and research*. Guilford, 1999.

[10] J. Kickul et al. Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and ksas. *JBP*, 2000.

[11] B. Lepri et al. Automatic prediction of individual performance from thin slices of social behavior. In *Proc. ACM MM*, 2009.

[12] B. Lepri et al. Employing social gaze and speaking activity for automatic determination of the extraversion trait. In *Proc. ICMI-MLMI*, page 7. ACM, 2010.

[13] J. E. McGrath. *Groups: Interaction and performance*. Prentice-Hall Englewood Cliffs, NJ, 1984.

[14] F. Pianesi et al. Multimodal recognition of personality traits in social interactions. In *Proc. ICMI*, Greece, 2008.

[15] E. Ricci and J.M. Odobez. Learning large margin likelihoods for realtime head pose tracking. In *Proc. ICIP, Cairo*, 2009.

[16] H. Salamin et al. Automatic role recognition in multiparty recordings: using social affiliation networks for feature extraction. *IEEE Trans. Multimedia*, 11(7), 2009.

[17] D. Sanchez-Cortes et al. An audio visual corpus for emergent leader analysis. In *Proc. ICMI Workshop*, 2011.

[18] J.S. Wiggins. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 37(3):395, 1979.

[19] A.W. Woolley et al. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 2010.