Modelling Participant Affect in Meetings with Turn-Taking Features

Catherine Lai, Jean Carletta, Steve Renals

Centre for Speech Technology Research,
School of Informatics, University of Edinburgh, United Kingdom
clai@inf.ed.ac.uk, j.carletta@ed.ac.uk, s.renals@ed.ac.uk

Abstract

This paper explores the relationship between turn-taking and meeting affect. To investigate this, we model post-meeting ratings of satisfaction, cohesion and leadership from participants of AMI corpus meetings using group and individual turn-taking features. The results indicate that participants gave higher satisfaction and cohesiveness ratings to meetings with greater group turn-taking freedom and individual very short utterance rates, while lower ratings were associated with more silence and speaker overlap. Besides broad applicability to satisfaction ratings, turn-taking freedom was found to be a better predictor than equality of speaking time when considering whether participants felt that everyone they had a chance to contribute. If we include dialogue act information, we see that substantive feedback type turns like assessments are more predictive of meeting affect than information giving acts or backchannels. This work highlights the importance of feedback turns and modelling group level activity in multiparty dialogue for understanding the social aspects of speech.

Index Terms: Turn-taking, dialogue, affect, groups, meetings, dialogue acts, satisfaction, cohesion, leadership.

1. Introduction

Humans are social animals. The success of meetings is not just a matter of the correct information flow, it also requires that participants to build consensus and rapport. Understanding the relationship between what happens in a meeting and affective outcomes like satisfaction, cohesiveness and perception of leadership is useful not just for meeting analysts but also for getting the design right in a range of technologies. Being able to characterize what makes for an affectually good meeting has implications for the turn-taking strategies of conversational agents attempting to convey information while building group cohesion. Similarly, understanding participant affect has implications for how technology can be used to facilitate meetings. For example, we would like to know if turn-taking patterns change information flow in a way that influences user satisfaction or productivity in computer mediated environments like video-conferencing.

Modelling meeting affect adds more broadly to our knowledge of how speech is used in social signalling. Most previous work on this has focused on involvement (similarly, interest and engagement) [1, 2, 3] or dominance [4, 5]. While understanding these social components is important in its own right, the implications of these studies for higher level meeting interpretation is unclear. For example, does dominant behaviour lead to perception of good leadership? Does active participation (i.e. involvement) from more individuals lead to greater group cohesiveness? In general, progress in this area has been hampered by the lack of direct automatic measures of meeting behaviour at the group level and the cost of employing methods that require these properties to be hand-annotated. To make some headway on these issues, we use a corpus of role-played meetings to show that automatic measures of interaction can explain differences in affectual outcome.

2. Background

Previous work on social signals in speech and group dynamics suggests that direct measures of how much participants speak or don't speak, i.e. the turn-taking structure, should play a major role in objective and affectual meeting outcomes. For example, brainstorming groups with fewer inter-speaker pauses produce fewer ideas but are more liked by participants [6]. Greater amounts of silence also correlated with higher disagreeability ratings of others in [7]. Substantive contributions were found to improve ratings in [8]. However, face-to-face meetings, characterized by shorter turns and more overlapping speech, had higher satisfaction ratings than video-conference meetings in [9], even when no difference in the quality of solutions the groups came up with was found. This suggests that more talk and less strict turn-taking in general leads to a more positive outlook, even if it doesn't lead to better task solutions.

Feedback utterances and their timing have also been identified as potentially important factors for understanding affect in spoken dialogue. However, results on the role of backchannels, for example, have not been consistent across different studies. Participants who interrupted more and produced more backchannels were less liked by external raters in task-oriented dialogues [10]. Similarly, male speed date participants perceived as friendly used less backchannels but more turn overlap in [11]. In [12], however, participants' production of short utterances correlated with how much participants said they were seeking to avoid disagreements, while backchannel counts were correlated with self-ratings of agreeableness, as well as conscientiousness and openness in ratings by other non-dialogue participants in [13]. Backchannels were not correlated with group likeability measures in [7], although the difference in the top two leadership scores per group was correlated with skew in backchannel distribution across participants.

To develop realistic conversational dialogue strategies we need to know when and how much to talk. Previous work incorporating automatic measures of turn-taking have focused on detection of involvement or dominance rather than evaluative aspects of meeting affect like satisfaction. So, we would like to know if equal turn-taking and less rigid turn-taking structure lead to more positive participant affect, and if so, what measures we can use to characterize this automatically. We also need to consider what sort of participation is appropriate. In that vein, we would like to know whether characteristics of conversational speech, like short feedback and overlap, are beneficial in this re-

	Satisfaction
1	I am satisfied with the group's discussion.
7	All in all, I am very satisfied.
9	I am satisfied with the process by which the group made its decision.
	Cohesiveness
2	I find the members of the group helpful.
8	I find the members of the group trustworthy.
11	I find the members of the group pleasant to be with.
13	I find the members of the group irritating.
14	I find the members of the group kind.
	Leadership
3	The meeting was directed in a good manner
4	Decisions were made in a democratic way
12	Every team member had sufficient opportunity to make his or her contribution.
16	All team members received sufficient attention.
	Information processing
5	All available information is being used.
6	I trust that we will find a good solution for the design problem.
10	I think my contributions affected the group discussion.
15	There was too much information.

Table 1: *Questionnaire Items.*

spect. To investigate this, we model post-meeting ratings about satisfaction, cohesiveness, leadership and information processing using individual and group level measures of turn-taking based on speaker activity and dialogue act usage. Experiments based on the AMI corpus meetings [14], are described in the following sections.

3. Experimental Setup

3.1. Meeting Data

The experiments described in the following were carried out on 'scenario' data from the AMI meeting corpus [14].[1] In the scenario, participants played out a series of four meetings focusing on different stages of remote control design for a fictitious company given various informational and budget constraints. Each group comprised of four participants, each assigned a specific role (project manager, user interface designer, marketing expert and industrial designer). This study uses the subset of the corpus where groups rated meetings for affective properties (120 meetings, 30 groups).

3.2. Participant Ratings

The ratings used in this study come from the post-meeting questionnaire described in [15] . This questionnaire was designed to help determine what makes a meeting successful with the goal of facilitating development of meeting browsers. Three main aspects of meeting success are queried over 16 items (see Table 1): process satisfaction, cohesiveness, leadership, and information processing.[2] Items were rated on a 1 to 7 scale (i.e. 'not applicable at all' to very much applicable'). In the following, ratings are centered from -3 to 3 and (reversed if necessary) so that positive scores reflect a positive affect. We model averages for individual participants ratings for satisfaction, cohesiveness, and leadership to get a general notion of these dynamics. We leave discussion of information processing (with the other dynamics) to individual items since it is not clear that aggregation over these items is meaningful [15].

[1] `http://corpus.amiproject.org/`

[2] NB: only a subset of the original question set described in [15] was administered at all sites.

3.3. Turn-Taking Features

The turn-taking measures used in the following are calculated using *spurts*: segments separated by at least 500ms silence [16], where we use word alignments to mark silence. In the following all continuous group and individual level features are converted to z-scores so that estimated effects are easier to compare. At the group level we look at measures of participation equality and predictability of turn-taking structure. Participation equality P_{eq} is defined as [17]:

$$P_{\text{eq}} = 1 - \frac{\sum_i^N (T_i - T)^2 / T}{E}, \quad (1)$$

where N is the number of participants, T_i total spurt time for participant i, $T = (\sum_i^N T_i)/N$ (i.e. equal participation). E represents the maximum possible value of the term under the sum: the average distance from equal participation (so E represents the case when only one participant speaks for the entire meeting). Values closer to 1 indicate greater equality. Similarly, let $H(Y|X)$ be the conditional entropy of speaker Y being the next participant to speak after X begins their spurt, with $H_{\text{max}}(Y|X)$ representing the maximal possible value for groups of a given size. Turn-taking freedom F_{cond} is defined as

$$F_{\text{cond}} = 1 - \frac{H_{\text{max}}(Y|X) - H(Y|X)}{H_{\text{max}}(Y|X)}. \quad (2)$$

So, F_{cond} is 0 when turn-taking follows a strict order (i.e. only speaker y follows x) and is 1 when every speaker follows everyone else in equal proportion. To examine the role of overlaps and possible interruptions, we measure barge-in rate (*bargein.rate*[3]) as the number of times any spurt is overlapped by a later starting spurt. We also measure the proportion of the interval that is silence (*sil.prop*).

At the individual level we look at participant speaking time (*ispk.prop.abs*) and number of Very Short Utterances (VSUs, *ivsu.rate*), both divided by the meeting duration. Here VSUs are defined as spurts that have duration less than 500 ms. These are likely to represent backchannels or other forms of short feedback [19].

3.4. Dialogue Acts

To get an idea of the type of contribution the participants make we include measures of dialogue act usage based on manual annotations of the AMI corpus [20]. This annotation scheme includes 15 dialogue acts involving the giving and eliciting of information, assessments, suggestions, offers, and comments about understanding. The label set also includes backchannels, stalls, and fragments, 'be positive' and 'be negative' categories. The latter two labels apply to utterances that do not fit into any of the previously mentioned categories and are deemed to have a purely social impact. The remaining utterances are assigned the category 'other'. To normalize across different meeting lengths we look at the number of instances of each dialogue act category proffered by the individual divided by the meeting duration (i.e. DA rates). We also look at the combined rates for participants other than the current rater.

3.5. Linear Model

We model the ratings averaged over group dynamics and for individual items for each participant in terms of a multilevel linear regression [21]. Indicators for team membership and role,

[3] We take the term from the dialogue systems literature, e.g. [18]

Feature	sat	lead	coh
sil.prop	**-0.15**	-0.07	-0.05
bargein.rate	-0.02	0.04	-0.05
P_{eq}	-0.00	0.08	-0.03
F_{cond}	**0.27**	**0.18**	**0.13**
ispk.prop.abs	0.06	-0.03	0.02
ivsu.rate	**0.14**	0.09	**0.12**

Table 2: *Turn-taking feature fixed effect estimates for aggregate satisfaction, leadership and cohesiveness for the speaker activity model. Items in bold are significantly different from zero (95% confidence interval).*

as well as group based turn-taking features described above are included as group level predictors. Individual participant features are included as either speaker activity features (the speaker activity model) *or* in terms of DA rates (the DA model). Coefficients were estimated using the R package `lme4`. Adding in the DA rates of other participants as individual level predictors did not improve the model fit. Similarly, including indicators for the role of the person who spoke the most did not explain any of the variance in the data, so for brevity we omit them in the models discussed in the following.

4. Results

4.1. Average Ratings

Estimates for the speaker activity model for aggregate satisfaction, leadership and cohesiveness are shown in Table 2. We consider effects to be significant where the 95% confidence interval of the estimate excludes zero. The results show a positive effect for F_{cond} in each of the rating categories, while silence proportion has a negative effect on satisfaction. There are no clear effects for participation equality or individual speaking time. That is, participants have a more positive attitude towards these meetings when floor-taking is less predictable, but not necessarily when everyone talks an equal amount. Similarly, people are more positive when there is less silence, but the fact that an individual talks more in general does not indicate that they will be more satisfied. Moreover, we don't see any clear effects for participant roles for any of the group dynamics. The positive effects for individual VSU observed suggests that short feedback utterances are important for modelling meeting affect and that more is better.

Incorporating DA rates, we again see that participants are more positive in meetings with higher F_{cond}, while meetings with less barge-ins and silence proportion also have greater satisfaction and cohesiveness (Table 3). Adding this turn-type information, it appears that the type of short utterances we are interested in are substantive and non-overlapping: participants had higher satisfaction and cohesiveness ratings when they made more assessments, not when they used more backchannels, where assessments include most short utterances that do more than just invite continuation, e.g. affirmative cue words. People who made more inform acts didn't appear to be more positive, so again the substance of an utterance is important. Suggestions appear to have a negative relationship with satisfaction and leadership, implying that people didn't like it when they had to suggest courses of action to others. Interestingly, we do not see any clear effects for the socially oriented DAs (be positive/negative), although this may be due to the low frequency of this category since they are defined via exclusion of other DA types.

Feature	sat	lead	coh
sil.prop	**-0.24**	-0.11	**-0.10**
bargein.rate	**-0.17**	-0.05	**-0.12**
P_{eq}	0.05	0.10	-0.00
F_{cond}	**0.24**	**0.16**	**0.12**
Assess	**0.23**	0.10	**0.12**
Backchannel	-0.03	0.03	0.06
BeNegative	0.01	0.05	-0.00
BePositive	0.07	0.07	0.03
CmtUnderstanding	0.07	-0.02	**0.09**
ElicitAssessment	0.02	0.04	-0.01
ElicitCmtUnd	0.07	0.09	0.03
ElicitInform	-0.08	-0.07	**-0.09**
ElicitOffOrSug	0.05	0.03	0.03
Inform	-0.03	0.04	-0.06
Offer	0.07	-0.04	0.06
Fragment	**0.15**	0.09	0.07
Other	-0.03	-0.03	0.04
Stall	-0.00	-0.06	-0.07
Suggest	**-0.18**	**-0.16**	-0.06

Table 3: *Coefficient estimates for the DA model.*

4.2. Individual Items

Results for the individual items paint a similar picture to that above for satisfaction and cohesion, although F_{cond} features less prominently for cohesion than satisfaction. Looking at the leadership and information processing questions, however, gives us a better indication of where sources of overall satisfaction come from (Table 4). On the one hand, silence proportion, elicit inform moves and suggestions had a negative impact on perception of how well the meeting was directed (Q3). On the other hand, turn-taking freedom had a positive effect for Q4 (*Decisions were made in a democratic way*) and Q12 (*Every team member had sufficient opportunity to make his or her contribution*). Similarly, participants in teams with higher turn-taking freedom also felt that information usage was better (Q5), while those who produced more elicit inform moves were more negative. In this vein, we see more information elicitations in groups that felt that there was too much information overall (Q15), while fewer backchannels and more inform moves were associated with higher ratings of irritation (Q13).

Overall, it appears that having a less predictable turn-taking structure leads to feelings of sufficient participation, even if speaking time isn't actually equal. Providing assessments similarly increased the feeling of contributing (Q10) although, interestingly, inform acts did not. Even so, feelings of sufficient participation may not result in the impression of good leadership as much as having less silence and not having to explicitly ask for information or make suggestions.

5. Discussion

Other studies have used turn-taking features to predict dominance with the motivation that this should help understand leadership behaviour [22]. In this study, dominance features (i.e. speaking time/inequality) do not appear to bear on how well directed meetings were perceived to be. However, looking at more specific process questions, we saw greater turn-taking freedom in meetings with a higher ratings of how democratic the decision making process was and how well information was flowing. This suggests that it's not so important how much people talk so much as the fact that individuals are able to take the floor and that turn-taking is generally less regimented. We might also

	Satisfaction			*Cohesiveness*					*Leadership*				*Information Processing*			
	1	7	9	2	8	11	13	14	3	4	12	16	5	6	10	15
sil.prop	-0.27	-0.21	-0.25	-0.25				-0.15	-0.25				-0.22		-0.18	
bargein.rate		-0.26			-0.19			-0.16								
P_{eq}																
$F^{'}_{\text{cond}}$	0.30	0.18	0.23					0.14		0.25	0.20		0.41			
Assess	0.25	0.25	0.19	0.13	0.15	0.14		0.15	0.22						0.21	
Backchannel							-0.16									
BePositive													0.16			
CAUnderstanding				0.12			-0.14							0.18		
EAssessment																0.19
ECUnderstanding															0.11	
EInform							0.21		-0.17				-0.18			0.21
Fragment	0.15	0.15														
Inform																
Offer							-0.12				-0.12					
Other								0.10								
Stall								-0.14								
Suggest	-0.22	-0.17	-0.15	-0.15					-0.33	-0.19			-0.20			

Table 4: *Coefficient estimates for individual items. Only effects where the 95% confidence interval for the estimate excludes zero. Questions are listed in Table 1.*

have expected higher features of group involvement to lead to more positive affect in this regard. However, we did not find that higher participation equality led to higher satisfaction or cohesiveness, nor did we find effects for overall individual speaking time. Some aspects of turn-taking usually associated with face-to-face conversational dialogue were associated with positive affect, i.e. very short utterances, while others were not, i.e. overlapping turns.

These results highlight the importance of identifying different types of turns when modelling the social aspect of meetings. While we can approximate feedback rates with VSUs, what we are really interested in are turns that form actual dialogue contributions (e.g. assessments) rather than continuation signals (e.g. backchannels). Many past studies have excluded backchannels and overlapped utterances altogether seemingly due to difficulties in transcript alignment (see references in [17]). However, recent studies indicate that the distinction between backchannels and contentful cue words is not so easy to delineate manually or automatically [23, 24]. As such, it appears that a greater focus on modelling feedback utterances is necessary for dialogue modelling, especially with respect to floor control [25, 26]. Given the cost and low agreement rates associated with manual annotation of social signals [27], development of automatic measures is crucial for carrying out large scale studies in these areas and for checking whether insights from qualitative studies generalize to other types of dialogue.

The current study makes some steps into understanding the relationship between turn-taking structure and meeting affect. However, we should note that the AMI meetings were also designed so that everyone had to participate to complete the task. In other conversational modes, people might be quite happy if just one person talks a lot. So, to generalize the results we would need to look at dialogues with more participants and where participants can easily opt out of speaking. Moreover, this sort of study is somewhat encumbered by the low stakes nature of the task. In more high stakes cases different leadership styles may result in equally good outcomes, while leaving the groups with different levels of satisfaction. Similarly, differences in the task structure may affect the reception of different acts. For example, in instructor/follower style dialogues, the need for frequent feedback from a follower may actually be a sign of task difficulty or problems with alignment between participants. This may explain the negative correlations with backchannels/affirmatives and (external) likeability reported in [10]. In contrast, the AMI scenarios involve more open ended discussion and opinion sharing, so short feedback is more likely to be an expression of involvement in the decision making process. In this vein, it seems unlikely that the negative impact of actively directing task structure or information flow in our study would be the case for instructor/follower dialogues. Of course, this doesn't mean that those sorts of acts should be avoided in AMI style meetings - they may be necessary to achieve goals. However, we need further evaluation of the quality of meeting task outcomes to be able to investigate this further.

6. Conclusion and Future Work

This study examined the relationship between turn-taking measures and participants affectual evaluations of meetings. Overall, the most broadly applicable predictor was found to turn-taking freedom. In general, participants were more satisfied and cohesive when group turn-taking freedom and individual VSU rates were higher. In fact, turn-taking freedom turned out to be a better indicator than actual speaking time equality of whether individuals felt they had a chance to contribute to the discussion. From a facilitation point of view, it appears that encouraging floor grabbing to be more unpredictable is beneficial for participant satisfaction. However, this needs to be conditionalized on the meeting task.

In general, feedback turns like assessments were more predictive of meeting affect than information giving acts. Given the positive effects for VSUs and assessments, it seems likely that video-conferencing systems that inhibit the production of short feedback, may reduce the satisfaction of participants. Similarly, conversational dialogue systems should include strategies that invite assessments from other participants. Even though we did not find effects for equality of participation, it has been linked to more diverse idea generation [6, 28, 17]. To investigate this further we need to leverage lexical features. In particular, it makes sense to look at the relationship between social signals, participant affect and coherence based features. Although adding combined DA rates for other group members did not improve model fit, a more sophisticated notion of participant influence (e.g. [12]) may also help increase our understanding of how turn-taking relates to meeting affect.

7. Acknowledgements

This work was supported by the European Union under the FP7 project inEvent (grant agreement 287872).

8. References

[1] B. Wrede and E. Shriberg, "Spotting "hot spots" in meetings: Human judgments and prosodic cues," in *Eighth European Conference on Speech Communication and Technology*, 2003.

[2] B. Schuller, R. Müller, F. Eyben, J. Gast, B. Hörnler, M. Wöllmer, G. Rigoll, A. Höthker, and H. Konosu, "Being bored? Recognising natural interest by extensive audiovisual integration for real-life application," *Image and Vision Computing*, vol. 27, no. 12, pp. 1760–1774, Nov. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0262885609000316

[3] C. Oertel, S. Scherer, and N. Campbell, "On the use of multimodal cues for the prediction of involvement in spontaneous conversation," in *Interspeech 2011*, 2011, pp. 1541–1544. [Online]. Available: http://www.isca-speech.org/archive/interspeech_2011/i11_1541.html

[4] R. Rienks and D. Heylen, "Dominance detection in meetings using easily obtainable features," *Machine Learning for Multimodal Interaction*, pp. 76–86, 2006.

[5] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez, "Modeling Dominance in Group Conversations Using Nonverbal Activity Cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 3, pp. 501–513, Mar. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4782034

[6] R. B. Ruback, J. M. Dabbs, and C. H. Hopper, "The process of brainstorming: An analysis with individual and group vocal parameters." *Journal of Personality and Social Psychology*, vol. 47, no. 3, pp. 558–567, 1984. [Online]. Available: http://content.apa.org/journals/psp/47/3/558

[7] D. Jayagopi, D. Sanchez-Cortes, K. Otsuka, J. Yamato, and D. Gatica-Perez, "Linking speaking and looking behavior patterns with group composition, perception, and performance," *Proceedings of the 14th ACM international conference on Multimodal interaction - ICMI '12*, p. 433, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2388676.2388772

[8] J. A. Bonito, "The Effect of Contributing Substantively on Perceptions of Participation," *Small Group Research*, vol. 31, no. 5, pp. 528–553, Oct. 2000. [Online]. Available: http://sgr.sagepub.com/cgi/doi/10.1177/104649640003100502

[9] R. van der Kleij, J. Maarten Schraagen, P. Werkhoven, and C. K. W. De Dreu, "How Conversations Change Over Time in Face-to-Face and Video-Mediated Communication," *Small Group Research*, vol. 40, no. 4, pp. 355–381, Apr. 2009. [Online]. Available: http://sgr.sagepub.com/cgi/doi/10.1177/1046496409333724

[10] A. Gravano, R. Levitan, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova, "Acoustic and prosodic correlates of social behavior," in *Interspeech 2011*, 2011, pp. 97–100. [Online]. Available: http://www.cs.columbia.edu/~rlevitan/papers/p95485.pdf

[11] D. Jurafsky, R. Ranganath, and D. McFarland, "Extracting social meaning: Identifying interactional style in spoken conversation," in *NAACL '09 Proceedings of Human Language Technologies*, no. June, 2009, pp. 638–646. [Online]. Available: http://dl.acm.org/citation.cfm?id=1620847

[12] A. Pentland, "Social dynamics: Signals and behavior," in *International Conference on Developmental Learning*, vol. 5, 2004. [Online]. Available: http://web.media.mit.edu/~sandy/TR-579.pdf

[13] F. Mairesse, M. Walker, M. Mehl, and R. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of Artificial Intelligence Research*, vol. 30, pp. 457–500, 2007. [Online]. Available: http://www.aaai.org/Papers/JAIR/Vol30/JAIR-3012.pdf

[14] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus," *Language Resources and Evaluation*, vol. 41, no. 2, pp. 181–190, 2007.

[15] W. M. Post, M. Huis in t Veld, and S. van den Boogaard, "Evaluating meeting support tools," *Personal and Ubiquitous Computing*, vol. 12, no. 3, pp. 223–235, Mar. 2007. [Online]. Available: http://link.springer.com/10.1007/s00779-007-0148-1

[16] D. Hillard, M. Ostendorf, and E. Shriberg, "Detection of agreement vs. disagreement in meetings: Training with unlabeled data," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. Association for Computational Linguistics, 2003, pp. 34–36.

[17] J. Carletta, S. Garrod, and H. Fraser-Krauss, "Placement of authority and communication patterns in workplace groups the consequences for innovation," *Small Group Research*, vol. 29, no. 5, pp. 531–559, 1998.

[18] K. Komatani and A. I. Rudnicky, "Predicting barge-in utterance errors by using implicitly supervised asr accuracy and barge-in rate per user," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ser. ACLShort '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 89–92. [Online]. Available: http://dl.acm.org/citation.cfm?id=1667583.1667612

[19] J. Edlund, M. Heldner, S. Al Moubayed, A. Gravano, and J. Hirschberg, "Very short utterances in conversation," in *Proceedings of fonetik*, 2010, pp. 11–16.

[20] S. Renals, T. Hain, and H. Bourlard, "Recognition and understanding of meetings The AMI and AMIDA projects," in *Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on*. IEEE, 2007, pp. 238–247.

[21] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press Cambridge, 2007.

[22] D. Jayagopi and D. Gatica-Perez, "Mining Group Nonverbal Conversational Patterns Using Probabilistic Topic Models," *IEEE Transactions on Multimedia*, vol. 12, no. 8, pp. 790–802, Dec. 2010. [Online]. Available: http://publications.idiap.ch/downloads/papers/2010/Jayagopi_IEEETRANS.%ONMULTIMEDIA_2010.pdfhttp://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5545443

[23] A. Gravano, S. Benus, J. Hirschberg, S. Mitchell, and I. Vovsha, "Classification of discourse functions of affirmative words in spoken dialogue," in *Interspeech 2007*, 2007, pp. 1613–1616.

[24] C. Lai, "Prosodic Cues for Backchannels and Short Questions: Really?" in *Proceedings of the Fourth Conference on Speech Prosody*, no. 1989, 2008.

[25] L. Chen, M. Harper, A. Franklin, T. R. Rose, I. Kimbara, Z. Huang, and F. Quek, "A Multimodal Analysis of Floor Control in Meetings," *Machine Learning for Multimodal Interaction*, pp. 36–49, 2006. [Online]. Available: http://www.cs.umd.edu/~zqhuang/index_files/MLMI06.pdf

[26] L. Chen and M. P. Harper, "Multimodal floor control shift detection," in *Proceedings of ICMI-MLMI '09*. New York, New York, USA: ACM Press, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1647314.1647320

[27] S. Afzal and P. Robinson, "Natural affect datacollection & annotation in a learning context," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–7.

[28] S. Silver, B. Cohen, and J. Crutchfield, "Status differentiation and information exchange in face-to-face and computer-mediated idea generation," *Social Psychology Quarterly*, pp. 108–123, 1994.