The discourse value of social signals at topic change moments

Francesca Bonin, Nick Campbell and Carl Vogel

School of Computer Science and Statistics
Trinity College Dublin, Ireland
{boninf,nick,vogel}@tcd.ie

10.21437/Interspeech.2015-433

Abstract

The dynamics of social interactions during human-human conversation have attracted the attention of many researchers aiming at improving the naturalness of automatic dialogue systems and creating socially aware machines. While many studies have investigate social dynamics in relation to the emotional sphere of a conversation, in this work we investigate whether these dynamics are correlated with the discourse structure, particularly topic changes. We conduct an observational study in two corpora of differing nature. In both corpora we find that at topic terminations, the interactional entropy, defined as the amount of social signals exchanged by the participants, decreases with the introduction of a new topic. We conclude that a relation exists between topic changes and amount of social interaction and, hence, that social signals have a discourse value in addition to a social function. This value, we suggest, is group acknowledgement that a new topic has gained the floor. Understanding these social dynamics within conversations can be exploited in the development of socially aware dialogue systems.
Index Terms: Social Signals, Topic Change, Paralinguistic signals, Dialogue Systems

1. Introduction

In recent years, much attention has been given to the study of paralinguistic events in conversations, both from an analytical and from a synthesis point of view. Domains like Affective Computing [1], Social Signal Processing (SSP) [2], Social Robotics [3], Intelligent Virtual Agents [4], Human Communication Dynamics [5] etc., have investigated the scope of paralinguistic signals in determining human-human and human-machine social interactions at several social and behavioral levels. This attention has led to work in automatic recognition of human personality in various contexts [6], artificial agents that sustain emotionally rich conversations with their users [7], affective interactions [8], and so on. Much effort has been devoted to increasing understanding of social signals, defined as the ensemble of non-linguistic vocalizations including turn taking, silences, overlaps and vocal outbursts, gazes etc. [9]. Such signals have been investigated in their relation with affective states, conflictual interactions, communication success, agreement and disagreement, among others factors [10, 11, 12, 13, 14].

While the social function of paralinguistic signals has been explored, few have attempted to investigate whether such signals underpin also a discourse function. However, a better understanding of the discourse function of social signals can represent a valuable resource in information extraction. Spontaneous conversations are usually a challenging scenario for traditional topic segmentation techniques (due to the noise in the transcriptions). Therefore, an understanding of whether and how those signals relate with topic changes can provide useful simple cues for the detection of topic boundaries in noisy, real world scenarios. In addition, a full understanding of a discourse function of social signals can be exploited in improving interactions with intelligent virtual agents, dialogue systems and speech interfaces in general, towards more socially aware machines.

Some [15] have investigated the relation among paralinguistic events and contextual factors, such as topic changes, gender, speaker role, mode of interaction. In previous work [16, 17, 18] we have analyzed the relation between laughter and topic changes, showing that the distribution of laughter is not random, nor uniform, and does not exclusively depend on the distribution of mirthful events, but rather correlates with discourse events. While such studies considering only an individual signal are useful, a wider perspective on an ensemble of social signals can provide an understanding of the variability in interaction among participants. Here we investigate the relation between such variability, which is expressed in the quantity and kinds of social signals expressed by the participants, and discourse events such as topic change. We call this variability in interaction *interactional entropy*, and we address the following research questions:

[RQ1] Is there a correlation between the interactional entropy of a conversation and discourse phenomena such as topic changes?

[RQ2] If there is such correlation, do topic changes influence interactional entropy or *vice versa*?

To answer these questions, we conduct an observational study over two distinct corpus types: a task-based dialogue (the AMI corpus) and spontaneous interaction (the TableTalk corpus), analyzing the social signals' distribution around topic changes and their interaction with the lexical volume. In what follows, we present a brief introduction of the five classes of social signals used in this study, together with a overview of related previous works, Sec. 2. We then describe the two datasets in Sec. 3, our methodology in Sec. 4 and finally experiments in Sec. 5. We provide a discussion and conclusions in Sec. 6.

2. Past studies of social signals

We define *interactional entropy* as the variability in interaction between participants, expressed by the amount of social signals exchanged by the participants. Here we analyze a subset of social signals, specifically: vocal outbursts such as laughs, and backchannels, overlaps and silences.

2.1. Laughter

While laughter has long being studied as the vocalization of a mirth, we start from the hypothesis the sometimes laughter can have a different function, not related to presence of a mirthful event either inside or accompanying the dialogue. Some have

Copyright © 2015 ISCA

studied the characteristics of controlled and uncontrolled laughter noticing a difference in the internal structure of the laughter: controlled laughter does not exhibit random structure but repetition; uncontrolled spontaneous laughter has been found to have random internal structure [19]. Laughter can be understood as a joint activity: one interlocutor may laugh alone, or a number may join the laughter. Previous studies have described laughter independently of its role in mirthful response to humor. In this context, laughter has been seen as a highly ordered phenomenon, internally and externally [20]. In this sense, it is also relevant to explore the timing of laughter with respect to other elements of interaction in dialog, such as topic changes. While the timing of mirthful laughter is effectively random, given the distribution of potential triggers, the timing of non-mirthful laughter might be related to conversation structure.[1] The hypothesis is that when laughter functions as a social signal, its timing is structured and conveys information about the underlying discourse structure. Previous works have explored other non-verbal features that can be predictive of discourse structure [22, 23]. Luz et al. [22] investigate the potential of non-verbal signals such as silences (among two speakers' vocalizations as well as within a single speaker's turn) and overlaps in predicting topic changes in meetings. Results show that pauses and overlaps on their own are good markers of the topic structure of meetings conversation, reaching performance comparable with lexical based methods. Vinciarelli et al. [15] also analyze laughter and topics, finding that laughs decreases when participant are concentrated on a topic rather than in social chitchats. In previous work, we have investigated the discourse function of laughter in isolation by looking at its distribution at a macro level throughout a topic [16]; here we investigate laughter and other signals at a micro level, providing a better insight into influence of the social signals on topic change.

2.2. Silences and Overlaps

Silences have long been studied in conversational analysis, and they play a fundamental role in vocal behavior [24]. Linguists have studied silence from different directions: from a functional perspective [25], from an acoustic perspective [26, 27], and from a discourse analysis perspective [28]. More recently silence has been studied as an important part of interaction and three kinds of silences have been distinguished: hesitation silence, psycholinguistic silence, and interactive silence [29]. Sacks et al. [28], and more recently [30], distinguished three different types of silences: pauses, gaps and lapses. A pause is defined as a silence within a turn. Gaps are defined as short pauses between turns and lapses as longer pauses between turns. Here we count as silence each soundless time interval bounded by human speech or audible sounds (see [15]). We start from the hypothesis that silence is more frequent in moments of high interaction characterized by shorter turns and a higher number of exchanges.

We also analyze overlaps, defining overlaps as time intervals during which at least two speakers talk at the same time, as in [15]. Many have looked at overlaps as indicators of conflicts, disputes or dominance displays [31]. Here we explore overlaps as a natural event occurring in spontaneous conversation, regardless of their social interpretation.

[1] While we see distinction between instances of mirthful laughter and structural laughter we do not here seek functional (or automatic) discrimination nor attempt to understand speakers' emotive state (others do attempt to infer speaker emotions [21]); rather, we treat all instances of laughter as members of the category 'social signals'.

We group silences and overlaps as complementary phenomena that jointly exhaust a single natural kind derived from the timing of individual linguistic interactional contributions.

2.3. Backchannels

In this work we consider backchannels short voiced utterances like "mm-hmm" or linguistic expressions like "yeap, yeah, yeas, right" that signal attention and encouragement. We do not consider the distinction among type of feedback that encourage the speaker to continue, from the type of feedback that anticipate a turn taking as in [32]. In this context, backchannels represent any vocal outburst which is direct to the speaker to reflect attention and engagement in the conversation.

3. Corpora

In this section we describe the two datasets considered: TableTalk and AMI corpus.

3.1. TableTalk

The TableTalk corpus contains multimodal recordings of free flowing natural conversations among five participants, recorded at the Advanced Telecommunication Research Labs in Japan [33]. To collect as spontaneous data as possible, neither topics of discussion nor activities were restricted in advance. Three sessions were recorded over three consecutive days in an informal setting over coffee, by three female (Australian, Finnish, and Japanese) and two male (Belgian and British) participants [34]. The conversations are fully transcribed and segmented for topic, and also annotated for affective state of participants and for gesture and postural communicative functions using MUMIN [35]. TableTalk has been analyzed in terms of engagement and laughter [36] and lexical accommodation [37]. Our analyses use transcripts of the entire corpus: about 3h 30, 31523 tokens and 5980 turns. The social signals considered were automatically extracted from the transcript.

3.2. AMI

The AMI (Augmented Multi-party Interaction) Meeting Corpus is a multimodal data set of 100 hours of meeting recordings [38]. The corpus contains role-play and real meetings. We base our analysis on the scenario based meetings, having topic annotation. Each meeting has four participants, and the same subjects meet over four different sessions to discuss a design project. The sessions correspond to four different project steps (Project kick-off meeting, Functional Design, Conceptual Design and Detailed Design). Each participant is given a role to play (project manager, marketing expert, industrial designer and user interface designer) and keeps this role until the end of the scenario. Conversations are all in English among 91 native speakers and 96 non-native speakers. Transcripts are topic annotated. All the social signals considered and the temporal elements were automatically extracted from the transcripts.

3.3. Topic annotation in TableTalk and AMI

Both TableTalk and AMI have topic annotations freely available. TableTalk topics were annotated manually by two labelers at a single level; AMI annotations include top-level or core topics whose content reflects the main meeting structure, and subtopics for small digressions inside the core topics. Here we use the core topic segmentation which is more in line with the TableTalk annotation.

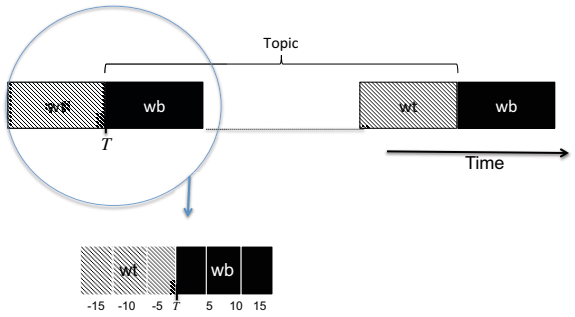


Figure 1: At the top: *wt* and *wb* segments - [Analysis 1]. At the bottom: fine grain segmentation of *wt* and *wb* - [Analysis 2].

4. Methodology

In order to investigate the relation between the interactional entropy, expressed by the variability of social signals, and topic changes, we conduct two analyses. With the first analysis we aim at addressing [RQ1] by understanding whether some sort of relation exists between interactional entropy and topic changes. With the second analysis we want to understand the directionality of such relation [RQ2].

4.1. Analysis 1: Topic termination vs topic beginnings

To answer [RQ1], we consider topic terminations and topic beginnings providing an operational model for such intervals.[2] We define a topic change, T, as the point in time in which a topic starts. Relying on [18], a threshold of fifteen seconds around topic changes has been chosen as a demarkation point for topic terminations as beginning, which are defined as follows:

wt: topic termination segments from T-15 seconds and T.

wb: topic beginning segments from T to T+15 seconds.

Gilmartin et al. [18] analyses the probability of laughter in 5 second bins at T minus multiples of 5 seconds (T-5, T-10, T-15, T-20) in order to study laughter trends near topic terminations. A meaningful threshold emerges (T-15 seconds) where a change in the laughter trend (number of laughs increasing significantly with respect to T-20) is visible. We use this to choose a threshold of T-15 seconds and T+15 seconds for defining *wt* and *wb*. Figure 1 depicts this operational segmentation. We analyze the distribution of social signals among these segments with non-parametric statistical tests with Bonferroni correction; a pattern indicates a non-random relationship between topic change and interactional entropy.

4.2. Analysis 2: Distributions around T

In a second analysis we address [RQ2]. We split the *wt* and *wb* segments into bins of 5 seconds which go from T to T-5, T-10, T-15 seconds and similarly from T to T+5, T+10, T+15 as shown in Fig 1 (bottom part). We then calculate the average amount of events per bin, over all the topics per AMI and TableTalk. Finally, we calculate the lexical volume per bin. We define lexical volume (LV_S) as the amount of lexical contributions (w) in a segment (S), excluding punctuation.[3]

[2]The analyses in this papers are conducted at topics' shifts. While it would be interesting to address the relation with turns' shifts, this is left to future works.

[3]No normalization over the length of the segment is necessary, as we always compare segments with the same duration.

5. Experiments and Results

5.1. Analysis 1: results

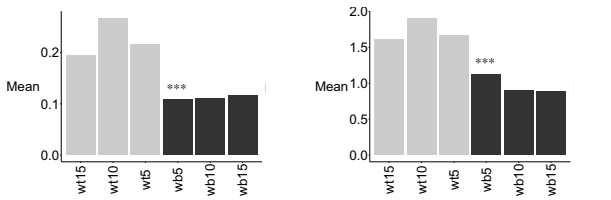
Being $\bar{x}$ the average amount of signals per segment, we analyze the distribution of $\bar{x}$ for wt and wb segments. Due to the non-normality of such distribution a non-parametric test is used (Wilcoxon Test). We define H_0: $\bar{x}_{wb}$=$\bar{x}_{wt}$ and H_1:$\bar{x}_{wb} < \bar{x}_{wt}$. The null hypothesis is rejected in favor of H_1 in the majority of the cases with $p < 0.001$. A general trend emerges in both TableTalk and AMI: topic terminations, compared to topic beginnings, show a significantly greater presence of social signals. In AMI, for each social signal—laughter, overlap, silence, backchannels—in *wt* and *wb* the non-parametric Wilcoxon test rejects the null hypothesis of $wb = wt$ and $wb \geq wt$, and validates the alternative hypothesis of $wb < wt$, $p < 0.001^{***}$. In TableTalk the same applies to laughter, overlaps, and lexical backchannels, but not silences. This pattern indicates a non-random relationship between topic change and interactional entropy [RQ1].

5.2. Analysis 2: results

From the preceding analysis a constant higher frequency of social signals in topic terminations rather than in topic beginnings emerges. However this result does not provide an answer to the causality effect, opening two possible scenarios: *1)* the decrease of interaction is the cause of the topic change; *2*) the topic change is the cause of the decrease in interaction. In order to answer this question, we examine more closely topic terminations and beginnings. As mentioned in Sec. 4, we divide *wt* and *wb* segments into bins of 5 seconds which go from T to T-5, T-10, T-15 seconds and similarly from T to T+5, T+10, T+15. We calculate the average number of events per bin, over all the topics per AMI and TableTalk. Fig 2a to Fig. 2d show the mean frequencies of signals from T-15 to T and T to T+15, in bin of 5 seconds for AMI. Fig 3a to Fig 3d refer to TableTalk. The bars *wt15, wt10, wt5* represent the bins before a topic change (light grey), the bars *wb15, wb10, wb5* represent the bins after a topic change (dark grey). The stars indicate significance difference in the distributions. With the exception of silences, a trend emerges: a drop in the frequency of events occurs exactly after the topic change. This is also confirmed by a series of statistical tests aimed at testing the differences in the distributions among consecutive windows. Once again the test applied is a non-parametric test (Wilcoxon, two tailed) due to the non-normal nature of the distribution. A significant drop in the frequency of signals is visible between wt_5 and wb_5 (signaled by the stars in the Figures), but not, generally, between the other bins. In the light of these results, it is reasonable to suppose that in the datasets examined, the decrease of social interaction, signaled by a decrease of social signals, is a consequence of topic change, as it happens immediately after the topic change [RQ2].

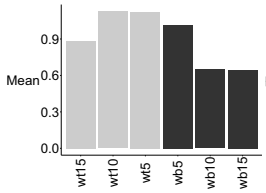
5.3. Lexical volume distribution around T

Finally, we investigate the lexical volume of between *wt* and *wb* events. If the beginning of a topic shows a lower social interaction, two hypotheses emerge: *wb* are characterized by greater speaker contribution of lexical content, or a greater number of silences (the latter could indicate a higher cognitive activity, additional to speech production). The second hypothesis can be rejected reflecting on the experiments in the previous section, as it is not found that *wb* segments have a significant higher number of silences than *wt* segments (the opposite is found in AMI). On the other hand, in order to explore the first hypoth-

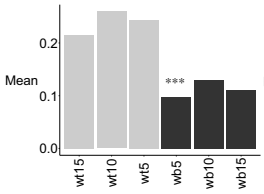


(a) AMI: Laughs distribution

(b) AMI: Overlaps distribution

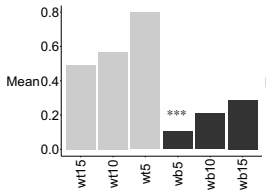


(c) AMI: Silences distribution

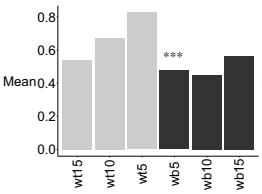


(d) AMI: Backchannels distribution

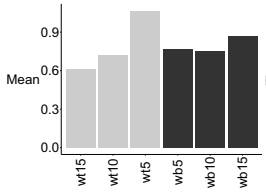
Figure 2: AMI distributions around topic changes.



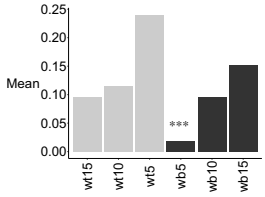
(a) TableTalk: Laughs distribution



(b) TableTalk: Overlaps distribution



(c) TableTalk: Silences distribution



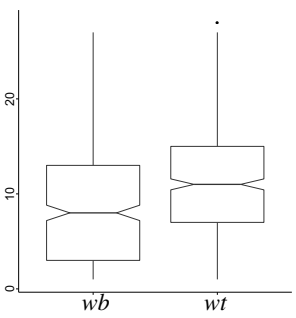
(d) TableTalk: Backchannels distribution

Figure 3: TableTalk distributions around topic changes.

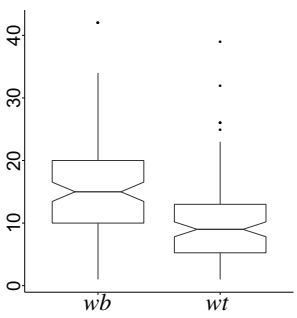
esis, we examine the distribution of lexical volume in wb_5 and wt_5 segments.[4] Fig. 4a and Fig. 4b show respectively the distribution of lexical volumes per AMI and TableTalk. In both cases, *wb* segments show significantly higher lexical volume than *wt* segments, as confirmed by T-student test, with H_0: $LV_{wb_5} = LV_{wt_5}$, and H_1: $LV_{wb_5} > LV_{wt5}$. The null hypothesis is rejected in favor of H_1, $p < 0.001$.

We investigate whether the higher lexical content in *wb* can be justified by a greater number of speakers. Interestingly, *wb* segments show a consistently lower number of speakers where compared with *wt* segments (T-test, $p < 0.001$). The average number of speakers in *wb* is 1.9 (sd=0.5) vs *wt* 2.25 (sd=0.7) in AMI. Similarly in TableTalk the average number of speakers in *wb* is 1.6 (sd=0.8) vs *wt* 2.3 (sd=0.9).

[4]Experiments conducted on wb_{15} and wt_{15} lead to the same result.



(a) LV in AMI



(b) LV in TableTalk

Figure 4: Lexical Volume in wb_5 and wt_5 - y axis distribution of LV

6. Discussion and Conclusions

We have analyzed the dynamics of five signals belonging to the social sphere around topic change moments in dialogue. Although not exhaustive, these social signals could represent the general level of social activity in the conversation, *interactional entropy*. We noticed that the beginnings of new topics show a lower presence of social activity, fewer speakers, but a greater amount of lexical content. In contrast, topic terminations show higher social activity and lower lexical volume. This is consistent with our view of this as a form of interactional entropy. We found that both in AMI and TableTalk there is a drop of interactional entropy when a new topic begins. One could interpret this as a social order introduced by the new topic: from a situation of high social interaction, with a higher number of overlaps, feedback, laughter, the new topic brings to a monological situation, in which one speaker takes the floor, reducing the interactivity among the participants. On this interpretation, one might understand the discourse function of diminished social signaling to be one of acknowledgment that a new topic has gained the floor, even if the topic is not "owned" by a single speaker in control of the floor, but a number of dialogue participants—this is the mult-party interactional equivalent of a hush coming over a group as a performance or oration commences. It remains to show that such an effect, diminishing interactional entropy at the start of a new topic, is discernible across various levels of topic-end interactional entropy, as such levels might differ in relation to communities of practice, familiarity among group participants, emotive qualities of topic, and so on.

Although limited to the datasets considered, a pattern has, therefore, been found between the fluctuation of social interaction (amount of social signals) and discourse phenomena such topic changes [RQ1]. In addition, we have noticed how topic changes are the cause of the drop of interactional entropy [RQ2]. Understanding such dynamics may contribute to social aware technologies able to engage in conversations following social rules of spontaneous conversations. Future work will be devoted to deeper explorations of this relationship on additional multimodal corpora, and exploiting this information on the social dynamics affordances of artificial conversational agents.

7. Acknowledgements

This work has been supported by the Innovation Bursary of Trinity College Dublin, the School of Computer Science and Statistics of TCD, the Speech Communication Lab at TCD, and by the SFI FastNet project 09/IN.1/1263.

8. References

[1] R. W. Picard, *Affective computing*. MIT press, 2000.

[2] A. Vinciarelli, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, Nov. 2009.

[3] C. L. Breazeal, *Designing sociable robots*. MIT press, 2004.

[4] J. Cassell, *Embodied conversational agents*. MIT press, 2000.

[5] L. Morency, "Modeling human communication dynamics," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 112–116, 2010.

[6] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Transaction on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[7] M. Schröder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. F. Valstar, and M. Wöllmer, "Building autonomous sensitive artificial listeners," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 165–183, 2012.

[8] S. Poria, E. Cambria, A. Hussain, , and G.-B. Huang, "Towards an intelligent framework for multimodal affective data analysis," *Neural Networks*, vol. 63, pp. 104–116, 2015.

[9] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland, "Social signals, their function and automatic analysis: A survey," in *Proceedings of ACM Int'l Conf. Multimodal Interfaces (ICMI'08)*, Chania, Greece, October 2008, pp. 61–68.

[10] M. Mehu and L. van der Maaten, "Multimodal integration of dynamic audio-visual cues in the communication of agreement and disagreement," *Journal of Nonverbal Behavior*, 2014.

[11] S. Kim, F. Valente, M. Filippone, and A. Vinciarelli, "Predicting continuous conflict perception with bayesian gaussian processes," *IEEE Transactions on Affective Computing*, 2014. [Online]. Available: http://www.dcs.gla.ac.uk/~vincia/papers/toacconflict.pdf

[12] D. Cheng, H. Salamin, P. Salvagnini, M. Cristani, A. Vinciarelli, and V. Murino, "Predicting online lecture ratings based on gesturing and vocal behavior," *Journal of Multimodal User Interafces*, 2014. [Online]. Available: http://www.dcs.gla.ac.uk/~vincia/papers/teaching.pdf

[13] A. Vinciarelli, A. Esposito, E. André, F. Bonin, M. Chetouani, J. F. Cohn, M. Cristani, F. Fuhrmann, E. Gilmartin, Z. Hammal *et al.*, "Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions," *Cognitive Computation*, pp. 1–17.

[14] F. Bonin, C. D. Looze, S. Ghosh, E. Gilmartin, C. Vogel, A. Polychroniou, H. Salamin, A. Vinciarelli, and N. Campbell, "Investigating fine temporal dynamics of prosodic and lexical accommodation," in *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*, 2013, pp. 539–543.

[15] A. Vinciarelli, P. Chatziioannou, and A. Esposito, "When the words are not everything: The use of laughter, fillers, backchannel, silence, and overlapping speech in phone calls," *Frontiers in ICT*, vol. 2, p. 4, 2015.

[16] F. Bonin, N. Campbell, and C. Vogel, "Time for laughter," *Knowledge-Based Systems*, vol. 71, no. 0, pp. 15 – 24, 2014.

[17] ——, "Laughter and topic changes: Temporal distribution and information flow," in *In Proceedings of CogInfoCom 2012 3rd IEEE International Conference on Cognitive Infocommunications*, 2012, pp. 53–58.

[18] E. Gilmartin, F. Bonin, C. Vogel, and N. Campbell, "Laughter and topic transition in multiparty conversation," in *Proceedings of SIGDIAL 2013 Conference,Metz, France, 22-24 August*, 2013, p. 304308.

[19] J. A. Bea and P. C. Marijuán, "The informational patterns of laughter," *Entropy*, vol. 5, no. 2, pp. 205–213, 2003.

[20] E. Holt, "On the nature of 'laughables' : laughter as a response to overdone figurative phrases," *Pragmatics*, vol. 21, no. 3, pp. 393–410, September 2011. [Online]. Available: http://eprints.hud.ac.uk/11553/

[21] M. Kurematsu, M. Ohashi, O. Kinosita, J. Hakura, and H. Fujita, "A study of how to implement a listener estimate emotion in speech," in *SoMeT*, 2009, pp. 531–540.

[22] S. Luz, "The nonverbal structure of patient case discussions in multidisciplinary medical team meetings," *ACM Trans. Inf. Syst.*, vol. 30, no. 3, pp. 17:1–17:24, sep 2012.

[23] S. Luz and J. Su, "The relevance of timing, pauses and overlaps in dialogues: detecting topic changes in scenario based meetings," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, 2010, pp. 1369–1372.

[24] B. Zellner, "Pauses and the temporal structure of speech," in *In*. John Wiley, 1994, pp. 41–62.

[25] V. J. Jensen, "Communicative functions of silence," *ETCA Review of General Semantics 30,*, pp. 249–257, 1973.

[26] E. D. Chapple, "Quantitative analysis of the interaction of individuals," *Proceedings of the National Academy of Sciences of the United States of America, 25(2),*, pp. 58–67, 1939.

[27] F. Goldman-Eisler, "The predictability of words in context and the length of pauses in speech," *Language and Speech*, vol. 1, no. 3, pp. 226–231, 1958.

[28] H. Sacks, E. A. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversationdynamic interpretation and dialogue theory," *Language*, pp. 696–735, 1974.

[29] V. P. Richmond, J. C. McCroskey, and M. Hickson, *Nonverbal behavior in interpersonal relations*. Prentice Hall Englewood Cliffs, NJ, 1991.

[30] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, no. 4, pp. 555–568, 2010.

[31] L. Smith-Lovin and C. Brody, "Interruptions in group discussions: The effects of gender and group composition," *American Sociological Review*, pp. 424–435, 1989.

[32] L. Cerrato, "Investigating communicative feedback phenomena across languages and modalities," Ph.D. dissertation, PhD Thesis, Speech Music and Hearing. KTH, Stockholm, Sweden., 2007.

[33] N. Campbell, "An audio-visual approach to measuring discourse synchrony in m ultimodal conversation data," in *Proceedings of Interspeech 2009*, 2009.

[34] K. Jokinen, "Gaze and gesture activity in communication," in *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, ser. Lecture Notes in Computer Science, C. Stephanidis, Ed. Springer Berlin / Heidelberg, 2009, vol. 5615, pp. 537–546.

[35] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena," *Language Resources and Evaluation*, vol. 41, no. 3-4, pp. 273–287, 2007.

[36] F. Bonin, R. Böck, and N. Campbell, "How do we react to context? annotation of individual and group engagement in a video corpus," in *SocialCom/PASSAT*, 2012, pp. 899–903.

[37] C. Vogel and L. Behan, "Measuring synchrony in dialog transcripts," in *Behavioural Cognitive Systems*, A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, and V. C. Müller, Eds. Springer, LNCS 7403, 2012, pp. 73–88.

[38] I. Mccowan, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner, "The ami meeting corpus," in *In: Proceedings Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research. L.P.J.J. Noldus, F. Grieco, L.W.S. Loijens and P.H. Zimmerman (Eds.), Wageningen: Noldus Information Technology*, 2005.