The Annotation and Analysis of Importance in Meetings

Robert Eklund
Rebecca Bates
Chad Kuyper
Elizabeth Willingham
Elizabeth Shriberg

TR-06003

June 2006

Abstract

Meetings typically contain important regions that are likely to be the focus of summarization and recall requests. We present a new approach for labeling speech corpora with categories of importance at the level of utterance groups; these labels may help to identify focus regions for browsing, summarization, or question-answering. We ask whether importance can be consistently labeled by humans with the idea that these regions might be used to improve speech understanding and automatic summarization of speech and text. We present information about related annotation schemes for high-level speech labeling, including the relationship between this labeling scheme and pre-existing labels at the levels of utterances and groups of utterances. We provide a summary of the annotation system and labeling procedure, as well as preliminary inter-annotator reliability statistics on the ICSI Meeting Recorder Corpus.

1

Annotation and Analysis of Importance in Meetings

Robert Eklund
Elizabeth Shriberg

International Computer Science Institute
University of California, Berkeley
Berkeley, CA USA 94074

eklund@icsi.berkeley.edu
ees@icsi.berkeley.edu

Rebecca Bates
Chad Kuyper
Elizabeth Willingham

Computer & Information Sciences
Minnesota State University, Mankato
Mankato, MN USA 56003

bates@mnsu.edu, chad.kuyper@mnsu.edu
elizabeth.willingham@mnsu.edu

1 Introduction

Automatic analysis of meetings has gained increased interest in recent years, as exemplified by such large-scale collaborative projects as *Augmented Multi-Party Interaction* (AMI),[1] the *ICSI Meeting Recorder Project* (Janin et al., 2003), *Computers in the Human Interaction Loop* (CHIL),[2] and the *Cognitive Assistant that Learns and Organizes* (CALO).[3] Explicit goals of these and similar projects include the creation of meeting browsers (Gruenstein, Niekrasz & Purver, 2005; Waibel et al., 1998), text summarization (Murray, Renals & Carletta, 2005; Murray et al., 2005), and hot-spot recognition (Wrede & Shriberg, 2003a, 2003b, 2005; Wrede et al., 2005).

In this study, we attempt an alternative approach to analysis of meetings which aims to develop and test a taxonomy for "importance", i.e. stretches of speech in meetings that labelers consider important without being given any detailed instruction. For the purpose of this work, we explore three types of importance in meetings: content (topic related), social interaction, and meeting flow; we ask whether labelers consistently annotate regions as belonging to both the same type of importance and the same level of importance. We also compare the results of this importance labeling with group interaction labeling (Meeting Acts) (Bates et al., 2005a, 2005b).

This paper is organized as follows. We will present information about related work, followed by a description of the corpus. After a description of the labeling approach and guidelines, analysis of the current labels and their interaction with Meeting Acts will be presented.

2 Related Work

Content analysis can be traced back to 18th century Sweden (Dovring, 1954; Wilson & Rayson, 1993) where religious controversy over a hymn book created a new field of research and generated ideas that are still part and parcel of the field (Krippendorff, 2004). The idea of labeling meetings for content has been carried out at different sites, using different schemes. This section will describe four of these.

Speech Acts

The notion that human communication can be described in terms of "actions", or speech/dialog acts, has been around since Austin (1962/1975) and Searle (1969) provided the first in-depth analyses of human communication as an interactive activity with the main objective of meeting goals common to speakers and listeners. More recently, Allwood (1997a, 1997b, 1995, 1994, 1977, 1976), Clark (1996), and others have provided alternative analyses and theories on human communication as an interactive, or collaborative, phenomenon. There are several different taxonomies to choose from, and an excellent overview of different schemas is provided in Traum (2000).

Of particular interest to this work is the Dialog Act (DA) which can be defined from an *is* perspective: i.e., a specific DA *is* a `statement` or *is* a `question`. Another way to define DAs is from an intentional perspective,

i.e., what a specific DA is intended to achieve. DA annotation sets have been defined for use in different conversational contexts: a task-oriented scheme, *DAMSL* (Allen & Core, 1997); a two-party conversation-oriented scheme, *SWBD-DAMSL* (Jurafsky et al., 1997); and multi-party meeting schemes, *MRDA* (Shriberg et al., 2004) and *MALTUS* (Popescu-Belis, 2004).

Meeting Acts (Minnesota State)

At Minnesota State University, labeling of *Meeting Acts* (MAs) was done with the goal of capturing meeting styles by categorizing speaker interaction states (Bates et al., 2005a, 2005b). MAs describe a higher level of abstraction than DAs and could be as short as a single linguistic segment or as long as an entire meeting. Allowable MA boundaries were the linguistic segment boundaries used with DAs (rather than acoustic-based boundaries, such as pauses or set-length time chunks). The primary labels (in their five groups) are shown in **Table 1**.

Table 1: Meeting Acts (from Bates et al., 2005a).

Group	Label	Tag Title
1	G	Agenda Work
	P	Planning of Future Work
2	N	Negotiation
	E	Voting
3	R	Reporting
	Z	Brainstorming
	T	Other Discussion
4	H	Humor
5	C	Commentary
	BE	Technical/Equipment Break
	BO	Breaks: Interrupts, filler talk, off-topic self-talk, etc.

Action Items (Stanford University)

Within the CALO project, an office assistant is being developed at Stanford University, employing ontology-based discourse understanding (Niekrasz et al., 2005; Niekrasz & Purver, 2005; Gruenstein, Niekrasz & Purver, 2005). Included in the labeling scheme are *Action Items* (AIs), defined as a task which is discussed and assigned to a participant to complete at some point after the meeting (Gruenstein, Niekrasz & Purver, 2005, section 2.2). AIs can span sets of utterances and include `proposals, acceptances, rejections,` and others (see also Pallotta, Niekrasz & Purver,

2005; Purver, Niekrasz & Peters, 2005). One of the outcomes of this work is a meeting annotation tool (NOMOS) that presents video as well as text and audio to labelers (Niekrasz & Gruenstein, 2006).

Meeting Actions (IDIAP)

At IDIAP, a scheme for annotation of joint participant behavior has been developed (McCowan et al., 2003). The issues addressed by this scheme allow for identification of the physical actions speakers are taking as well as what level of interaction the speaker may have with other participants or equipment (such as a projector or whiteboard). Labels include `monologue`, `presentation` (i.e., including a projector), `whiteboard`, `discussion`, `disagreement`, `consensus,` and `note-taking`.

3 Importance Labeling

What the previous schemes have in common is that they distinguish different kinds of DAs or other conversational moves from a content perspective: i.e., they describe the linguistic content without any kind of importance ranking. Our approach, which starts with importance, is described below.

3.1 Label Set

Although we hold that the importance labeling should be done with as little instruction as possible, it was still obvious to us that discourse items could be important in a different dimension at a meta-level. Consequently, the label set we decided to use in this pilot consists of the following:

- *Socially Important (S)*
Refers to socially important utterances, e.g., utterances which build or maintain relationships or that express emotive states.

- *Topic Important (T)*
Refers to utterances that are important as to the content of the meeting, i.e., the task or topic under discussion.

- *Meeting Flow Important (M)*
Refers to utterances that are important from an organizational or administrative point of view.

The labelers read transcripts of the meeting while simultaneously listening to audio files of the same meetings. They were instructed to use linguistic segment boundaries as importance boundaries. An utterance could be labeled as either *1* (important) or *0* (not important), resulting in a total of six labels: *S0/S1, T0/T1,* and *M0/M1*. Since there was also the option to not use a label (Null), utterance regions labeled with S0, T0, or M0 are still more pertinent to the meeting than those areas not labeled. As was pointed out above, a central and crucial feature in our study was to stay away from detailed instruction. Therefore, the basic instruction to the labelers was "If you were reading these meetings as a textbook, what would you highlight?".

3.2 Data

The ICSI Meeting Recorder Corpus is a collection of over 90 recorded and transcribed technical meetings (Janin et al. 2003). The recorded meetings consist of regularly scheduled meetings of various groups meeting for different purposes. The conversations are natural in that the meetings would have happened even if they were not being used for data collection. Seventy-five of the meetings are hand-annotated with DA labels at the sentence level (Shriberg et al., 2004) and with MA labels describing group interaction above the sentence level (Bates et al., 2005a, 2005b). Three meeting groups were used

in the development of our labeling system. Two of these (BED and BMR) had similar styles in that they consisted of groups of people with generally similar status discussing projects. The third (BRO) had a distinctly different style in that there was typically a clear leader; in these meetings, the group interaction tended to be "reporting" rather than "brainstorming" or "discussion". We chose five meetings each from BED (out of 15) and BMR (out of 27) for this study as well as a few other meetings (out of BRO and BMR) for development purposes.

4 Experiments

The meetings were labeled for importance using text, linguistic segment boundaries, dialog act labels, and audio. Although meeting act labels were not used during labeling, the labelers were familiar with the MA labels. Three conversations (one each from BRO, BED, and BMR) were used for initial discussions about approaches before labelers independently labeled the ten meetings in the initial set. For our developmental assessment purposes, we used two measures: raw time counts of label types and counts of labels that have matching overlap regions of importance. Additionally, mappings between importance regions and available MA labels were used to see if there is any correlation between group interaction and importance.

Table 2. Percentage of labels by raw time.

TYPE	Labeler	Count	T1	T	S1	S	M1	M	Null
BED	1	5	14.2	73.8	0.3	5.0	1.0	1.5	19.7
	2	5	47.2	77.8	5.2	11.4	0.9	3.0	7.8
	Both	10	30.7	75.8	2.8	8.2	0.9	2.3	13.7
BMR	1	6	14.7	65.6	1.4	2.4	1.2	1.9	30.3
	2	5	49.0	88.9	2.3	6.7	1.4	2.3	2.2
	Both	11	30.3	76.2	1.8	4.3	1.3	2.1	17.5
BRO	2	1	41.8	88.7	1.1	3.3	2.9	3.3	4.9

Table 3. Percentage of label matches using one labeler as truth and comparing to a second labeler.

Type	Truth	Label Count	Exact Match (%)	Type Match (%)	Any Label (%)
BED	Labeler 1	91	51.6	82.4	97.8
	Labeler 2	186	30.1	62.4	80.6
	Combined	277	37.2	69.0	86.3
BMR	Labeler 1	75	56.0	94.7	100.0
	Labeler 2	190	35.8	68.4	78.9
	Combined	265	41.5	75.8	84.9
Both	Total	542	39.3	72.3	85.6

5 Results

5.1 Label Assessment

The first method of assessment uses raw time measurements to look at two aspects: 1) labeler tendencies and 2) meeting differences. Results are shown in **Table 2**. In all cases, the overall topic label (label T) was the majority label, on average being used for about 75% of both BMR and BED meetings. Not surprisingly, the meeting and social labels were used significantly less than the topic label. While the BMR and BED meetings had similar amounts of time spent controlling the agenda and meeting flow (label M), the BED meetings had more time spent on aspects that could be called social (label S). The lone BRO meeting had more time spent on the agenda that was considered "important", possibly because this meeting type was typically guided by a single person rather than by a collaborative group. Labeler 2 was more likely to label a topic segment as actually being important than labeler 1. Labeler 1 was more likely to not categorize segments (Null) than labeler 2. For both labelers, there was a negligible amount of overlapping labels (less than 0.5%).

Labeling results were assessed using both labelers as "truth" and counting label matches at the level of exact match (e.g., T0 and T0), type match (e.g., S1 and S0), and any importance label used (e.g., T1 and M0). Matches were counted when a label was used by the second labeler within the boundaries of the first label. Results are presented in **Table 3** for the two types of meetings labeled by both labelers. One of the reasons for lower agreement is that labeler 1, who tended to allow more Null areas, categorized discussion related to the Digit Task (recording strings of digits that happened while meeting participants wore microphones) as Null, while labeler 2 categorized these tasks (which occurred at least once in every meeting) as T0. Otherwise, there was strong agreement in category labeling, suggesting that further instruction and discussion about the importance levels may improve labeler consistency.

5.2 Labeling Observations

As was shown in the time-based measurements as well as in label counts, the meetings were not uniquely divided by the three importance types (M/S/T). Instead, there were occasional sections of Null and overlapping labels. Similarly, the importance regions were not uniformly distributed among the three label types. Rather, there was a much higher occurrence of T labels than of M or S labels, regardless of importance level (0 or 1).

The different meeting groups (BED, BMR, BRO) displayed the importance levels with varying frequency. BRO had short snippets of importance, while BMR often had large sections labeled T, where the importance level shifted frequently between 0 and 1. BED meetings showed large sections of discussion or negotiation that were relevant but of little importance and thus took the label T0.

Audio was highly beneficial to the labelers in establishing importance boundaries, especially in determining which backchannels were contextually relevant. Hearing the meeting made clearer to the labelers instances of important social interactions and of dialog acts that influenced meeting flow. Audio also determined the order in which labelers asked themselves two questions "Is this important?" and "Is this social, topic, or meeting flow related?" With audio, labelers tended to notice importance first.

Labelers initially wanted a "lukewarm" label (M/S/T2); however, lacking this importance level label forced labelers to make stronger decisions about dialog act importance. This may have affected labeler agreement, as one labeler tended to grade these lukewarm segments up an importance level, while the other usually marked these as less important. This inconsistency will be addressed in further labeling.

Labelers found that labels tended to overlap, if only because natural speech overlaps; e.g., one participant trying to capture the floor will overlap the end of another participant's speech in order to establish his/her own right to speak. Such instances of floor-grabbing were often labeled M0, as they affected the flow of the meeting but made no reference to an established agenda.

5.3 Interaction with Meeting Acts

Though labelers did not have access to MA labels, they sometimes noticed that they were using similar criteria to determine importance as

they would have used to mark some meeting acts. For example, utterances that might have been labeled *G* using MAs received a label of M1 in this study. Also, there was some correlation between the MA *R* and the MI *T*; instances of reporting and discussing were almost always given a topic importance label (label T0 or T1).

These observations are supported by an assessment of the MA labels contained within MI regions. For each importance label, counts of affiliated MA labels were done. The distributions are shown in **Table 4**. Negotiation was rarely marked as important in any region type. Although the group 3 MAs (reporting, brainstorming and discussion) are the most common, they are more likely to appear in importance regions related to the topic. Planning and social interaction tend to co-occur more than planning and topic interaction. We also see that Group 1 MA labels are more associated with the meeting importance labels.

Table 4. Percentage of MA labels by importance label. Outside topic MAs were excluded.

	G1	G2	G3	Group 3			G5	Count
	GP	N	RZT	R	Z	T	CB	Count
M0	40	0	57	24	3	30	0	30
M1	32	5	51	27	5	27	2	41
S0	16	2	70	23	9	39	11	44
S1	18	0	71	37	0	34	5	38
T0	12	3	72	31	6	35	5	152
T1	14	7	78	31	9	38	0	121

6   Summary and Conclusions

Several issues arose in the labeling process. One that arose for the labelers was that of familiarity with the meeting topics. Our labelers hold that their importance levels would probably change based on whether or not they were knowledgeable about the topics under discussion. Because these labelers knew very little about the meeting content, they based their importance labels on their perceptions of the meeting participants' reactions to utterances. These interpretations are subjective but are consistent for each labeler. Since this is similar to a naïve summarization system, this type of labeling system, as well as any relationships with MAs or DAs, may be valuable for assessment or for guidance in developing

meeting browsing, summarization and/or question-answering systems.

Another issue is the ephemeral nature of importance. Those involved in this project have a strong notion that "importance" is subject to half-life and suspect that ideas might well fall from very important (level 1) to not at all important (level 0 or Null) either gradually or overnight. This, however, does not preclude that importance labels can be of value, not only while still "hot" but also after their immediate value has diminished. To address this and labeler familiarity with topics, the meetings should be reviewed by the labelers at a different point in time. Labelers found that their importance levels changed slightly when reviewing one meeting's labels after labeling other meetings involving the same topics.

It may prove useful to use somewhat more specified categories tuned to domain, such as "administratively important" or "scientifically important", but this proposal runs counter to the basic idea of the study, which was to identify non-content-related importance levels. More specific labels would also require labelers trained in specific areas, whereas general importance can be labeled by those who are not specialists in a field.

The basic research issue in this paper was to make a first attempt to answer the question of importance: can it be evaluated objectively, or is it completely in the eye of the beholder? While this is not yet completely answered, we have shown that labels of importance can be generated quickly and that regions are consistently labeled as important, even though consistency on importance level (0 vs. 1) should be improved by further clarification for the labelers. We plan to explore automatic detection of importance regions and labels. We expect that meeting summarization and automatic segmentation will both benefit from this approach.

7   Future Work

There are many open avenues to explore in the realm of importance. In the short term, we expect that studies including more labelers (using labelers who are knowledgeable about the topic content as well as naïve labelers),

evaluating more types of meetings, and using more rigorous assessment methods will be done.

There is definitely a difference when labelers work with text alone and with both text and audio, since the reactions of listeners were useful in importance annotation. Whether this should be assessed as a difference between expectations for automatic systems that use text-only vs. text and acoustic information or whether labelers should always have audio available is an open question.

A study of "lukewarm" importance labels would prove interesting. This label might apply to areas that the labeler thinks will change based on later work or to regions that might be important for a week, but will later prove moot, e.g., short-range plans and questions that need to be addressed to or by people who are not present at the meeting.

A future study with video data would also make it easier to define subsets of meeting participants and would help labelers to decide whether or not un-microphoned meeting participants' utterances were important or not. It may also facilitate "counting" the participants for whom utterances are important, since utterances that are important to a subset of meeting participants are not always relevant to the meeting as a whole. The NOMOS tool would make this type of annotation feasible (Niekrasz & Gruenstein, 2006).

Acknowledgments

References

Allen, James & Mark Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. Available at <http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html>.

Allwood, Jens. 1997a. Notes on Dialog and Cooperation. In: Kristiina Jokinen, David Sadek & David Traum (eds.), *Proceedings of the IJCAI '97 workshop "Collaboration, Cooperation and Conflict in Dialogue Systems"*, 25 August 1997, Nagoya, Japan, pp. 9–21.

Allwood, Jens. 1997b. Dialog as Collective Thinking. In: Paavo Pylkkänen & Pauli Pylkkö (eds.), *New Directions in Cognitive Science. Publications of the Finnish Artificial Intelligence Society. International Conferences*, no. 2, Helsinki, pp. 222–226. Also in: Paavo Pylkkänen, Pauli Pylkkö & Anti Hautamäki (eds.), *Brain, Mind & Physics*, Amsterdam: IOS Press, pp. 205–210.

Allwood, Jens. 1995. Reasons for Management in Spoken Dialogue. In: Robbert-Jan Beun, M. Baker & M. Reiner (eds.), *Dialogue and Instruction*, Springer-Verlag, pp. 241–250.

Allwood, Jens. 1994. Obligations and Options in Dialogue. In: *Think*, vol. 3, ITK Tilbury University, pp. 9–18.

Allwood, Jens. 1977. A Critical Look at Speech Act Theory. In: Östen Dahl (ed.), *Logic, Pragmatics and Grammar*. Lund: Studentlitteratur, pp. 53–69.

Allwood, Jens. 1976. *Linguistic Communication as Action and Cooperation. A study in pragmatics*. PhD thesis, Göteborg University, Sweden.

Austin, John. 1962/1975. *How to do things with words*. Oxford: Clarendon Press.

Bates, Rebecca, Patrick Menning, Elizabeth Willingham & Chad Kuyper. 2005a. Meeting Acts: A Labeling System for Group Interaction in Meetings. *Proceedings of InterSpeech 2005*, 4−8 September 2005, Lisbon, Portugal, pp. 1589−1592.

Bates, Rebecca, Patrick Menning, Elizabeth Willingham & Chad Kuyper. 2005b. *Mapping Meetings Project: Group Interaction Labeling Guide*. Internal document, Aug. 24, 2005. Available at <http://bates.cs.mnsu.edu/MRMA.pdf>.

Clark, Herbert H. *Using language*. 1996. Cambridge, Cambridge University Press.

Dhillon, Rajdip, Sonali Bhagat, Hannah Carvey & Elizabeth Shriberg. 2004. Meeting Recorder Project: Dialog Act Labeling Guide. ICSI Technical Report TR-04-002, February 9, 2004.

Dovring, Karin. 1954. Quantitative Semantics in 18th Century Sweden. *Public Opinion Quarterly*, vol. 18, pp. 389−394.

Gruenstein, Alexander, John Niekrasz & Matthew Purver. 2005. Meeting Structure Annotation: Data and Tools. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, 2−3 September 2005, Lisbon, Portugal, pp. 117−127.

Janin, Adam, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke & Chuck Wooters. 2003. The ICSI Meeting Corpus. *Proceedings of IEEE/ICASSP 2003*, 6−10 April 2003, Hong Kong, vol. 1, pp. 364−367.

Jurafsky, Daniel, Elizabeth Shriberg, & Debra Biasca. 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13*. University of Colorado, Boulder. Institute of Cognitive Science Technical Report 97-02

Krippendorff, Klaus. 2004. *Content analysis: an introduction to its methodology*. Thousand Oaks: Sage Publications.

McCowan, Iain, Samy Bengio, Daniel Gatica-Perez, Guillaume Lathoud, Florent Monay, Darren Moore, Pierre Wellner & Hervé Bourlard. 2003. Modeling Human Interaction in Meetings. *Proceedings of ICASSP 2003,* 6-10 April 2003. Hong Kong, vol. 4, pp. 748-751.

Murray, Gabriel, Steve Renals & Jean Carletta. 2005. Extractive Summarization of Meeting Recordings. *Proceedings of InterSpeech 2005*, 4−8 September 2005, Lisbon, Portugal.

Murray, Gabriel, Steve Renals, Jean Carletta & Johanna Moore. 2005. Evaluating Automatic Summaries of Meeting Recordings. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Measures for Machine Translation and/or Summarization*, June 2005, Ann Arbor, Michigan, pp. 33−40.

Niekrasz, John & Alexander Gruenstein. 2006. NOMOS: A Semantic Web Software Framework for Annotation of Multimodal Corpora. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006),* 24-26 May 2006, Genoa, Italy.

Niekrasz, John & Matthew Purver. 2005. A Multimodal Discourse Ontology for Meeting Understanding. *Proceedings of the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, 11−13 July 2005, Edinburgh, Scotland.

Niekrasz, John, Matthew Purver, John Dowding & Stanley Peters. 2005. Ontology-Based Understanding for a Persistent Meeting Assistant. *Proceedings of the AAAI Spring Symposium on Persistent Assistants: Living and Working with AI*, 21−23 March 2005, Stanford, CA.

Pallotta, Vincenzo, John Niekrasz & Matthew Purver. 2005. Collaborative and Argumentative Models of Meeting Discussions. *Proceedings of the 5th Workshop on Computational Models of Natural Argument*, 30 July 2005, Edinburgh, Scotland.

Popescu-Belis, Andrei. 2004. Abstracting a Dialogue Act Tagset for Meeting Processing. *Proceedings of the 4th Annual Conference on Language Resources and Evaluation*, 26-28 May 2004, pp. 1415−1418.

Purver, Matthew, John Niekrasz & Stanley Peters. 2005. Ontology-Based Multi-Party Meeting Understand-ing. *Proceedings of the CHI'05 Workshop on the Virtuality Continuum Revisited*, 3 April 2005, Portland, Oregon.

Searle, John. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Shriberg, Elizabeth, Raj Dhillon, Sonali Bhagat, Jeremy Ang & Hannah Carvey. 2004. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In: M. Strube & C. Sidner (eds.), *Proceedings 5th SIGdial Workshop on Discourse and Dialogue*, 30 April − 1 May 2004, Cambridge, MA, pp. 97−100.

Traum David. 2000. 20 Questions for Dialog Act Taxonomies. *Journal of Semantics*, 17(1), pp. 7−30.

Purver, Matthew, John Niekrasz & Stanley Peters. 2005. Ontology-Based Multi-Party Understanding. *Proceedings of the CHI'05 Workshop on the Virtuality Continuum Revisited*, 3 April 2005, Portland, Oregon.

Waibel, Alex, M. Bett, Michael Finke & Rainer Stiefelhagen. 1998. Meeting Browser: Tracking and Summarizing Meetings. In: D.E.M. Penrose (ed.), *Proceedings of the Broadcast News Transcription and Understanding Workshop*, Lansdowne: Morgan Kaufmann, pp. 281−286.

Wilson, Andrew & Paul Rayson. 1993. The Automatic Content Analysis of Spoken Discourse. A Report on Work in Progress. In: Clive Souter & Erik Atwell (eds.), *Corpus-Based Computational Linguistics. Amsterdam*: Rodopi, pp. 215−226.

Wrede, Britta, Sonali Bhagat, Raj Dhillon & Elizabeth Shriberg. 2005. *Meeting Recorder Project: Hot Spot Labeling Guide*. ICSI Technical Report TR-05-004, May 12, 2005.

Wrede, Britta & Elizabeth Shriberg. 2005. *Reliability for Hot Spot Annotations in the MRDA Corpus*. ICSI internal document, April 11, 2005.

Wrede, Britta & Elizabeth Shriberg. 2003a. Spotting "Hot Spots" in Meetings: Human Judgments and Prosodic Cues. *Proceedings of Eurospeech 2003*, 1−4 September 2003, Geneva, Switzerland, pp. 2805-2808.

Wrede, Britta & Elizabeth Shriberg. 2003b. The Relationship Between Dialogue Acts and Hot Spots in Meetings. *Proceedings of the IEEE Speech Recognition and Understanding Workshop*, 1−4 December 2003, St. Thomas, US Virgin Islands.