Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and Challenges

Bridget Kane, Saturnino Luz, Jing Su

Department of Computer Science
Trinity College
kaneb@tcd.ie, luzs@scss.tcd.ie, sujing@tcd.ie

Abstract

This paper highlights the issues involved in gathering a corpus of data on the multimodal interaction that occurs at a team meeting of medical specialists. Difficulties in capturing the data are described, and the ethical issues are emphasised. Methods to investigate the internal structure of meetings, at the level of discussion topic (patient case discussion) are summarised and the potential benefit that such meeting records promise are reviewed. The hospital setting where the corpora are proposed experience issues in common with any business venture, but in addition demonstrate additional sensitivities because of health service complexities and patient privacy issues.

1. Introduction

This paper discusses the issues involved in the recording, annotation and analysis of conversation among a group of hospital specialists at multidisciplinary medical team meetings (MDTMs). We report on progress so far in the automatic annotation and analysis of these multimodal corpora and the potential usefulness of our approach to information capture in MDTM settings.

A teaching hospital setting has all the issues that one encounters in any workplace setting such as policies, procedures, culture, company business concerns and privacy, with additional requirements because of the fact that the collection is in healthcare.

We see that corpus gathering in this situation is useful and important for three reasons. Firstly, as a research tool to improve our understanding of the role interactions, information sharing and collaboration in this type of work. Challenges encountered in a teaching hospital with regard to the multitude and complexity of specialist roles is normally greater than in most organisations. Through using a hospital forum to study collaborative work, the findings are likely to find application in a variety of other work settings. Secondly, we believe a meeting record in this setting would provide a useful artefact or co-ordinating mechanism for the co-operative work involved in patient care. By demonstrating its utility in this respect we expect that electronic meeting records will be eventually adopted. Finally, a meeting record would facilitate organizational level tasks such as audit and quality review for organizational development and learning. We expect that by integrating meeting records into existing data gathering methods in the hospital will prove more efficient than current practices of independent specialised data collection for a single audit question. Changing the methods for information monitoring and audit review has the potential to reshape organisations and enhance quality improvement initiatives in the long term.

We summarise the main difficulties in achieving meeting corpora in business meeting scenarios, particularly in the health sector, and we also consider issues in conducting this type of research, i.e. different role perspectives, privacy constraints and ethics.

The main goal of this paper is to identify the key issues that workplace researchers need to address, together with the issues specific to the healthcare service. We also discuss how these concerns can be addressed.

2. Background

Routine multidisciplinary medical team meetings are becoming an important event in modern hospital life, particularly in cancer centres, as more and more professional organizations and regulatory bodies are making recommendations for the adoption of MDTMs into patient care pathways (Calman and Hine, 1995). The rationale for their adoption is twofold: i) they provide a useful forum for triple assessment of the patient's clinical findings, thus improving the quality of diagnosis, and ii) as the management of disease becomes more and more complex, multidisciplinary discussion is a useful co-ordinating mechanism for treatment planning and the management of individual patients. Many related work activities outside of the meeting require careful planning and co-ordination for the system of multidisciplinary team patient care to work efficiently (Kane and Luz, 2009; Kane and Luz, 2006a). For example, a cancer patient might require treatment through a few modalities, namely surgery, medical oncology and radiation oncology. The sequence and timing of these interventions might be sequential, concurrent, or in a combination, depending on the tumour type, size and anatomic location. The co-ordination and timing of such treatment strategies can be crucial to a successful outcome for a patient, and requires high levels of co-ordination and interaction among the associated specialities involved.

MDTMs provide a valuable resource for information gathering to inform patient management tasks subsequent to the meeting, and they have potential to be used as an information resource for audit and planning purposes. With the potential value of meeting recordings in mind, we investigated human and technological issues involved in building advanced computing support for collaboration, production

Kane, B., Luz, S., and Jing, S. Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and challenges. In *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (Malta, 2010), LREC, pp. 140–145.

and access of electronic medical records in the context of MDTMs. It is apparent that although recent technological and organisational developments have made digital recording of entire meetings a distinct possibility, the usefulness of this kind of audiovisual database is dependent on how effectively its contents can be accessed, among other factors. We believe that the internal structure within the MDTM can be harnessed so that elements of the discussion, or particular information, can be retrieved from recordings more effectively than linear methods alone will allow.

In conducting our research on meetings in a busy hospital setting, we experience many of the difficulties in common with any complex work setting, such as limited resources, interruptions and rescheduling of tasks due to national and personal holidays, staff illness, etc. Additional issues more prominent in healthcare are encountered, such as medical emergencies, patient concerns for privacy and confidentiality and new developments in technology and treatments. The main issues that we identify here and discuss in this paper relate to respecting the patient's privacy and that of the health professionals collaborating at the meeting.

3. Methods

The work reported here is based on several years of ethnographic observation, supplemented with audiovisual recording, together with questionnaires and interviews conducted with the multidisciplinary team members. Specific exercises that targeted particular research questions are reported elsewhere, such as in (Kane and Luz, 2006b). This paper reports in a more general way on the overall issues concerning the multimodal corpus collection. The meetingroom where the corpus was gathered is shown in Figure 1.

Figure 1 shows the team engaged in discussion. Radiological images from either disk, radiological film or the PACS[1] system and pathology (tissue) samples from a microscope are shown on the main screen display. Images may also be used, from time to time, that were taken at patient procedures that pre-date the meeting, such as video clips taken at surgery or at endoscopy.

3.1. Audio Capture Requirements

The human voice frequency band range is approximately from 80Hz to 1100Hz, and the frequency response range of a selected microphone should span this voice band range, as a minimum. One type of widely used conference microphone offers the frequency response range from 30Hz to 20,000Hz. Given our interest in indexing the recorded meeting data, we adopted the sampling rate of 16kHz, which is commonly used in speech recognition (Lee et al., 1989), in order to convert the microphone's analog output signal into a digital format.

MDTMs are held in a closed meeting room with 10 to 20 participants. Clinical specialists sit beside each other in rows, and face the main monitor. We record voice from clinical specialists for further analysis. The recorded audio needs to offer satisfying audio volume, have a high signal to noise ratio and a moderate frequency response range for human voice.

We aim to evaluate how a speaker influences meeting structure through topic change, so the optimal recording strategy is to separate each speaker's voice through the recording devices. Throat microphones are superior to traditional microphones for this task because they capture the sound wave more directly from the vocal chords thus reducing outside noise interference. Audio signals from one throat microphone can be recorded in a single channel, on which the target speaker's voice is prominent over peripheral speakers and noise. In post-capture processing, the audio files can be filtered through speaker diarisation so as to generate files containing the voice of a single speaker. As we note below however, despite the fact that throat microphones would have been an optimal choice from an audio processing perspective, they proved unacceptable for this particular data gathering project.

In the current MDTM setup (see Figure 1), two cardioid condenser boundary microphones are mounted on the front wall and side wall of the meeting room. The distance between microphone and the main speakers is about 3 meters. Cardioid microphones pick up sounds from all directions, so that voices from all speakers are recorded in a single channel. In comparison with the throat microphones, cardioid microphones record lower quality audio, but they do not interfere with the meeting participants. In order to locate vocalization boundaries of each speaker from the cardioid microphones recordings, speaker diarisation and segmentation algorithms can also be executed, though the results will be far less satisfactory. These proceedures are required if one aims to perform high-level segmentation, categorisation, and other forms of indexing on the meeting data (Bouamrane and Luz, 2007). Chen (Chen and Gopalakrishnan, 1998) suggests Bayesian Information Criterion (BIC) (Schwarz, 1978) as a standard to evaluate the coherence of continuous speech, for speaker segmentation. Speaker identification techniques can be used to label all vocalizations from the same speaker. Gaussian Mixture Models (GMM) can be used for this task (Reynolds and Rose, 1995). The use of Gaussian Mixture Models for modeling speaker identity is motivated by the interpretation that the Gaussian components represent some general speaker-dependent spectral shapes. In other words, these are acoustic classes, which are useful for modeling speaker identity.

3.2. Data gathering in practice

Due to the fact that we aimed to record real MDTMs, subjected to the stringent constraints of a medical setting, the actual practice of data gathering deviated significantly from the requirements scenario outlined above. The multimodal corpus was gathered from two media sources. The first was an S-VHS recording facility in the Telesynergy® system (Martino et al., 2003) of the audio in videoconference together with the screen display being broadcast to the meeting. A proportion of the meetings were held in videoconference, and for those meetings the recording captured the incoming video stream together with the audio discussion in videoconference. Outgoing video data were captured through the picture-in-a-picture view that was displayed on a TV monitor during the conferences. Given that the research involved a second institution, approval was also

[1] Picture Archiving and Communication System

Kane, B., Luz, S., and Jing, S. Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and challenges. In *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (Malta, 2010), LREC, pp. 140–145.

Figure 1: Multidisciplinary medical team meeting

sought from the staff at the second hospital site.

The second recording source was a video camera placed at the back of the room, in the same location from which the still image in Figure 1 was taken. This recording captured gestures and movements among the participants as they pointed, turned to gaze in particular directions or engaged in personal note-taking. These observations were valuable in helping identify information gathering needs of individual roles within the group.

Participants at the meeting did not wear any microphone devices, because of the principle agreed at the outset not to interfere in the work of the staff in any way (Section 3.3.). This resulted in less than perfect recordings, because of background noise from people handling papers, coughing, sneezing, moving in seats, etc. Furthermore, while the staff were agreeable to the recording of the meeting in order to gather annotations and verify observations, there were some reservations expressed concerning potential breach in confidentiality. Once assurances were given and trust was established, then cooperation was achieved.

The wall-mounted cardioid condenser boundary microphones were used to capture the speech from the participants in discussion. As mentioned above, we would have preferred if the participants had worn throat microphones, to improve the signal to noise ratio of the recorded audio. However because throat microphones exert pressure on the throat and may distract clinical specialists, and the danger of this having a detrimental indirect effect on patient care, it was decided to use the wall mounted cardioid condenser boundary microphones only. While this method proved satisfactory for our purposes, we noted that the audio recordings from videoconference was superior to that of co-located discussion. This may be due to the way the Telesynergy system (McAleer et al., 2001) used in the recordings was configured.

The S-VHS recording was transferred to digital tape and both recordings were converted to MPEG-4 format. These media files were imported into the Elan annotation tool (MPI, 2005), synchronised and annotated. Annotations were prepared manually in the first instance at the level of individual patient discussion. Following the full meeting being segmented into individual patient discussions, each discussion was then sub-divided into its natural sections, defined as D-Stages, and fully described in (Kane and Luz, 2009). The identifiable discussion tasks were further subdivided into four sub-sections. At a deeper level of detail, vocalisation events by participants were annotated and labelled with the individual's identifier and professional role. Analysis was subsequently conducted at individual and specialist role levels (reported elsewhere), since there was more than one individual for any particular role. For example there were two consultant radiologists in attendance and three respiratory physicians at most meetings.

Automatic annotation has also been performed at different

Kane, B., Luz, S., and Jing, S. Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and challenges. In *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (Malta, 2010), LREC, pp. 140–145.

levels based on the recorded speech. These included speech segmentation and speaker diarisation (Su et al., 2008), segmentation of meetings into patient case discussions (Luz, 2009), categorisation of such discussions (Luz and Kane, 2009), and speaker role identification (Su et al., 2010). Although speaker diarisation in noise settings remains a difficult problem, higher level segmentation tasks can be performed accurately enough to facilitate the process of manual annotation by researchers (e.g. as an add-on to tools such as ELAN) or to support browsing of meeting records by users.

Following collection of the detailed annotations the original recordings containing confidential patient data and actual medical discussions were destroyed, as agreed at the outset.

3.3. Privacy and Ethics

In the first instance ethical approval was required by the hospital Board before commencing any research with the multidisciplinary team. Two areas of ethical concerns were required to be addressed: i) The protection of the business interests of the hospital as would be required in any company, and ii) concerns for patient confidentiality.

In gaining confidence of the hospital staff, the ethical committee required that one of the senior medical consultants (staff member) vouched for and supported the research proposal. This individual agreed to mentor the study and also accepted responsibility for maintaining the highest ethical standards on behalf of the hospital. All staff were informed of the nature of the study by the lead researcher at the outset. As part of maintaining on-going cooperation and trust, regular progress reports and result data are provided to the staff.

Because our primary interest was in the hospital work systems and the role and effectiveness of the multidisciplinary team meeting in those processes, our research was given approval. Had this project focussed on any individual patient data, further ethical process would have been required as part of the research approval procedures of the hospital.

The researchers undertook not to interfere in any way in patient management, as well as giving undertakings that any patient information that was incidentally learned in the course of the research would be respected and maintained in the strictest confidence. The issue of how to preserve anonymity and the sensitive content in recording medical and other types of meetings while maintaining enough of the original data (speech and video signals) to allow researchers to investigate automatic meeting indexing methods has received attention from the research community in recent years. Promising directions include the gathering of "sociometric" signals though unobtrusive wearable devices (Olguin et al., 2009), and the use of digtal signal processing techniques for anonymising speech data (Parthasarathi et al., 2009). Achieving an acceptable way of recording and storing multimodal data is crucial to corpus gathering and data indexing research in medical settings.

4. Structure of the meeting

In an analysis of the collaboration and interaction exhibited at MDTMs, an apparent structure was identified, described in (Kane and Luz, 2009). MDTM are composed of several patient case discussions (PCDs) and internal structures have been identified. These structures reflect the highly structured tasks undertaken in the discussion and are a testament to the medical tradition of conducting a patient assessment in a predictable way, i.e. information is methodologically reviewed and the underlying cause of the presenting problem is assessed in the first instance, before treatment is prescribed. During PCDs the narrative follows with tradition in the conduct of the two main tasks, namely, the patient diagnosis and the next step in the patient's management. For each of these tasks specialities interact, collaborate, exchange and share information. Images are used by some of these specialities, for example radiologists use patient radiological imaging, pathologists show microscopic images and surgeons may use video to demonstrate their findings or procedures. Participants have been observed to point at images, and use their hands to describe the complexities of size and shape of tumours. Drawings have also been used by participants to explain the orientation of a tumour, or finding, at MDTMs. Representational gestures have been found to play an important role in medical meetings (Becvar et al., 2008). Therefore, multimodal meeting corpora from MDTMs would include the artefacts used, gestures and annotations, but in the corpora described here, we confine ourselves to audio recordings of the speech interactions.

As well as identifying internal structures of MDTMs and PCDs through our ethnographic observations, we are investigating automatic methods, as outlined in Section 3.. PCDs segments in MDTMs are analogous to *topics segments* in more general meeting corpora such as AMI (Carletta, 2007). We found that features of the vocal interactions such as the speaker ID, role, length of vocalisation, pauses and overlaps are useful in helping segment the meeting data into individual topics, or individual patient case discussions (PCDs). These features may also be useful in segmentation of the internal PCD sub-section boundaries which we define as D-Stages.

5. Utility of MDTM Records

The fact that an internal structure is identifiable through automatic means suggests to us that these techniques could be applied for the retrieval of individual discussions that match particular criteria. Such a development would potentially facilitate a number of important hospital functions listed in Table 1, particularly review of meeting proceedings without necessitating a full meeting review and the development of a corpus of patient cases that would inform future decisions, including the development of clinical practice guidelines. These potential uses are discussed below.

Individual Contributors would be able to review or check their input to a discussion. Sometimes a patient's results might be given an emphasis in discussion that is not reflected in the formal written report in the patient's file and this can lead to later confusion. A PCD record would allow for the contributor to check that their contribution was not misleading, or in contradiction to any formal written reports.

Kane, B., Luz, S., and Jing, S. Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and challenges. In *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (Malta, 2010), LREC, pp. 140–145.

Perspective	Utility
	Current
Individual contributor to PCD	Record of contribution made and context of any comments Evidence of image data provided that informed discussion
Individual listener	Ability to expert advice given in PCD Facility to review any task assigned
Specialist in training	Bank of cases for educational purposes
Hospital	Record for individual patient's record Audit Development of improved practice guidelines
	Potential
Hospital	Automatic data collection for National Statistics Health Insurers Department of Health
MDTM	Facilitate real time review of similar case to current discussion

Table 1: Potential Utility for MDT meeting records

Individual Listeners could review a meeting record and the need for individual note taking, which might be a distraction at a meeting, would be obviated.

Specialists in training would be able to review a corpus of cases of a particular type in order to educated themselves in a particular type of problem.

Hospitals would reduce the risk of errors by being assured that decisions taken at MDTM were fully documented. Furthermore, the availability of MDTM records would improve current audit practices and provide useful data for the development of clinical practice guidelines. The MDTM is also a potentially a very useful forum for data gathering for National Registries and required by the Department of Health and other agencies.

MDTMs could potentially access prior similar cases to the case under discussion which would help in making the decision about the current case. It may be, for example, that a similar case had an unexpected outcome during treatment, which may moderate the treatment decision on the current patient. Having a corpus of PCDs together with follow-up data on the outcome of their treatment undertaken would provide evidence (data) for the development of clinical practice guidelines that would influence future decisions.

6. Discussion and Conclusion

The opportunities that a multimodal meeting record would provide to a multidisciplinary medical team are well recognised. However, difficulties are experienced in capturing such a record from technical and behavioural perspectives. The technical difficulties in making the recordings could be overcome through the use of a dedicated meeting room with suitable microphones and recording devices to capture the images used to inform the discussion. Difficulties however in maintaining data security, including respecting patient privacy and confidentiality, while making electronic records available in a hospital network poses a great challenge.

The skepticism demonstrated by medical specialists in the adoption of technology into the healthcare workflow has been documented. Staff experience of failed IT projects and lack of knowledge of the potential contribution that IT might bring to MDTMs are both significant and are not to be underestimated (Heeks, 2006; Southon et al., 1999). Establishing and maintaining trust between the researchers, developers and hospital staff is a key factor in the success of any study. Involving individuals from the group, inviting research mentors and providing frequent progress reports to the group was critical to undertaking this study. We believe that if the technical issues can be satisfactorily addressed and potential benefits demonstrated, then the development of multimodal meeting records will be found to directly improve patient care and make the health services more efficient in the long term.

Acknowledgements

We wish to thank the multidisciplinary teams at St. James's hospital Dublin for their co-operation in this on-going study. We especially thank the members of the lung MDT for facilitating this research, and our mentors Dr. F. O'Connell, Prof. K. O'Byrne, Prof. D. Hollywood and Mr.

M. Buckley. This work is funded under the IRCSET Enterprise Partnership scheme with St. James's hospital.

7. References

Amaya Becvar, James Hollan, and Edwin Hutchins. 2008. Representational gestures as cognitive artifacts for developing theories in a scientific laboratory. In *Resources, Co-Evolution and Artifacts: Theory in CSCW*, pages 117–143. Springer-Verlag, London.

Matt-M. Bouamrane and Saturnino Luz. 2007. Meeting browsing. *Multimedia Systems*, 12(4–5):439–457.

Kenneth Calman and Deirdre Hine. 1995. *A Policy Framework for Commissioning Cancer Services*. Department of Health, Welsh Office.

Jean Carletta. 2007. Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus. *Language Resources and Evaluation*, 41(2):181–190.

S. Chen and P. Gopalakrishnan. 1998. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In *Proc. of DARPA Broadcast News Transcription and Understanding Workshop*.

Richard Heeks. 2006. Health information systems: Failure, success and improvisation. *International Journal of Medical Informatics*, 75(2):125 – 137.

Bridget Kane and Saturnino Luz. 2006a. Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing. *Computer Supported Co-operative Work (CSCW)*, 15(5-6):501 – 535, December.

Bridget Kane and Saturnino Luz. 2006b. Probing the use and value of video for multi-disciplinary medical teams in teleconference. In *Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems*, pages 518–523. IEEE Computer Society, July.

Bridget Kane and Saturnino Luz. 2009. Achieving diagnosis by consensus. *Computer Supported Cooperative Work (CSCW)*, 18(4):357 – 392, April. DOI:10.1007/s10606-009-9094-y.

Kai-Fu Lee, Hsiao-Wuen Hon, and Mei-Yuh Hwang. 1989. Recent progress in the sphinx speech recognition system. In *HLT '89: Proceedings of the workshop on Speech and Natural Language*, pages 125–130, Morristown, NJ, USA. Association for Computational Linguistics.

Saturnino Luz and Bridget Kane. 2009. Classification of patient case discussions through analysis of vocalisation graphs. In *Proceedings of the 11th International Conference on Multimodal Interfaces and Machine Learning for Multimodal Interaction (ICMI-MLMI'09)*, pages 107–114, New York, NY, USA. Association for Computing Machinery, ACM.

Saturnino Luz. 2009. Locating case discussion segments in recorded medical team meetings. In *SSCS '09: Proceedings of the ACM Multimedia Workshop on Searching Spontaneous Conversational Speech*, pages 21–30, Beijing, China, October. ACM Press.

R L Martino, K M Kempner, F S McGovern, D Chow, M E Steele, J E Elson, and C N Coleman. 2003. A collaborative telemedicine environment for the ireland - northern ireland - national cancer institute international partnership in cancer care. In *25th Annual International Conference of the IEEE EMBS*. IEEE Computer Society, Sept 17-21.

J McAleer, D O'Loan, and D Hollywood. 2001. Broadcast quality teleconferencing for oncology. *Oncologist*, 6(5):459–462.

MPI. 2005. ELAN: Eucido Linguistic Annotator. Max Planck Institute for Psycholinguistics, March. http://www.lat-mpi.eu/tools/elan/.

D.O. Olguin, B.N. Waber, Taemie Kim, A. Mohan, K. Ara, and A. Pentland. 2009. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics*, 39(1):43–55, February.

Sree Hari Krishnan Parthasarathi, Mathew Magimai.-Doss, Daniel Gatica-Perez, and Hervé Bourlard. 2009. Speaker change detection with privacy-preserving audio cues. In *ICMI-MLMI '09: Proceedings of the 2009 international conference on Multimodal interfaces*, pages 343–346, New York, NY, USA. ACM.

D.A. Reynolds and R.C. Rose. 1995. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83.

Gideon Schwarz. 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Gray Southon, Chris Sauer, and Kit Dampney. 1999. Lessons from a failed information systems initiative: issues for complex organisations. *International Journal of Medical Informatics*, 55(1):33 – 46.

Jing Su, Bridget Kane, and Saturnino Luz. 2008. Automatic content segmentation of audio recordings at multidisciplinary medical team meetings. In *International Conference on Information Technology*, Gdansk, Poland.

J. Su, B. Kane, and S. Luz. 2010. Automatic meeting participant role detection by dialogue patterns. In *Proceedings of COST 2102 Int. School - Development of Multimodal Interfaces: Active Listening and Synchrony*, volume 5967.

Kane, B., Luz, S., and Jing, S. Capturing multimodal interaction at medical meetings in a hospital setting: Opportunities and challenges. In *Proceedings of Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (Malta, 2010), LREC, pp. 140–145.