Development of a Face-to-Face Meeting Capture and Indexing Process

Barry Piorkowski[1]
Richard Evans[1]*
James Gao[1]

[1] Centre for Innovative Product Development, University of Greenwich, UK
* Now working at Westminster Business School, University of Westminster

Development of a Face-to-Face Meeting Capture and Indexing Process

Barry Andrew Piorkowski, Richard David Evans and James Xiaoyu Gao
Centre for Innovative Product Development
University of Greenwich Chatham Maritime, United Kingdom R.D.Evans@greenwich.ac.uk and J.Gao@greenwich.ac.uk

***Abstract*— A face-to-face (F2F) meeting capture and indexing process could provide an internal feedback mechanism so that results can be monitored against change interventions decided at product lifecycle management (PLM) or performance development review (PDR) meetings. It could also assist people who do not currently have the knowledge or authority to act on what was discussed to share and reflect on the F2F meeting content after the event. In this paper a prototype is described in use with the continuous improvement Plan-Do-Check-Act (PDCA) cycle as a proof-of-concept in BAE Systems. More work should be completed to make the captioning of the F2F meeting content automatic or semi automatic. The presentation of the content in the groupware webpage may also be developed so that there is easier navigation. It would also be useful to have the variables that influence the profitability of the products to be displayed on the webpage so that changes can be tracked against business performance.**

Keywords- meeting; video; groupware; product development; BAE Systems; continous improvement

I. INTRODUCTION

Most humans can filter, decode and interpret whatever sensory data is available to a standard convention which can be learnt and remembered with neural pathways and brain activity. Common understanding in face-to-face (F2F) meetings may be facilitated through use of language. However, even with just one standard such as English, there are many subtle linguistic variations including syntax and semantics. This means that the signals from unstructured conversation in a F2F situation can be confusing to both humans and machines.

People are needed to investigate and make sense of complex situations but individuals are unique with their own story and life journey history [1, 2]. Each person therefore in the same F2F meeting can categorise and value the same audiovisual data in a different way to provide conflicting information biased to their own interest. Likewise the same person may interpret the same data differently depending on how they feel relative to their location and their function in time and space. In contrast, machines behave in the way in which they have been preset to behave against a set of rules or protocols chosen by a programmer. Artificial intelligence and neural networks also need to be prepared by people to train how the machine, such as an automatic speech recognition system, categorises an output. Similarly, leadership at the top of an organisation would like to believe that they have motivated people to perform according to a policy of set principles. This is not always the case and it is the unpredictability of human behaviour that drives both error and change.

The traditional way of publishing F2F meeting content used to be to employ a typist to write-up 'minutes' of what was said during the meeting. With the widespread acceptance of word processing, email and the web this work can now be delegated to a F2F meeting participant. The purpose of 'minutes' according to Whittaker et al. is to track group progress; to serve as a public record of past actions and decisions; to remind people about their commitments; and to resolve disputes about commitments [3]. Relying on only one person to write-up poses reliability issues because of bias. There are however groupware web applications that allow more than one user to contribute to content [4].

Many industrial managers and experts can spend a significant amount of time and money traveling to and attending F2F meetings so therefore it would make sense to try and get the best value out of that associated content. Formal standard processes in industry such as the Performance Development Review (PDR) for people to reflect on and set personal learning and achievement objectives and the Product Lifecycle Management (PLM) process are designed to link the actions and decisions people make to the success of product development and the profitability of the company. PLM is a systematic method whereby only products that have met the prearranged project criteria may progress with appropriate funding. This is discussed and decided at F2F review meetings. The PDR process is where the people in the company are motivated to achieve set goals, which is also normally the subject of a F2F meeting.

Automating the capture and indexing of F2F meetings for search and retrieval to make organisational decision making more transparent could reduce the associated costs. It may also provide an internal feedback mechanism to be used with a continuous improvement cycle so that performance results could be monitored against change interventions decided at those meetings.

This study is part of the three year project funded by BAE Systems and UK Engineering and Physical Science Research Council (EPSRC) in 'Knowledge Management for Integrated Product Development'.

II. STATE-OF-THE-ART

There has been a lot of research in the area of human computer interaction. A fusion of multiple wearable physical measurement sensors has featured in the literature whereby a person can log their daily activity into life-log media format [5, 6] which can then be retrieved and viewed through a computer browser or head mounted display [7]. Life-logging technology captures single person experience effectively. However, the thoughts and behaviour of many working towards a common goal may be more powerful. Logging the life of a F2F meeting room where many people share stories and make decisions could provide a greater volume of valuable knowledge than just following one person.

F2F meetings have had the attention of many past researchers over the years. The focus of published work has been on how to find the right people to attend the meeting [8], how to resolve issues raised in the meeting [9, 10] and the quantification of the audio signal contribution from each meeting participant displayed in graphical display on screen [11-14]. Isolating individual speakers in a group meeting in real-time has been achieved by either giving each participant an individual microphone to achieve a separate audio channel input [11, 13, 14] or forcing a participant to sit in a particular seat so that sounds can be selectively filtered coming from specific directions [12]. Contribution quantity does not always mean quality. Therefore an interest also needs to be paid to what people say and not just how often they exude sound waves.

Transcription or captions are useful for both human users to read quickly through text to find interesting parts and also for machine search engines. Captioning audiovisual data can be achieved by using phonetic keyboards and may provide a live as well as archived transcription of what is said when people speak in meetings. According to Wald citing his own previous work, trained stenographers can accurately cope with people talking at up to 240 words per minute. Whereas, the automatic verbatim captions created directly from the voice of a speaker using automatic machine technology has a much higher word error rate, sometimes in excess of 50%. Editing software therefore has been developed so that human intervention can edit the output caption errors and add in speaker name identification details [15]. Another approach to speech transcription has been presented. The semi-automatic tool finds a speech segment in the audio stream, segments the speech into small utterances, allows the user to listen to the speech segment and then type the transcription for the speech segment [16]. The researchers state that this method is up to six times faster than manual transcription or captioning tools. Utterances have also recently been made available online to leverage the lower man-hour rate expectations of some internet users. At this time it is an unsecured method but it does reduce the cost of professional transcription. Web-users are willing to be paid for as little as $0.01 per five second utterance transcribed [17].

Metadata tags also for search optimisation have been stored in an extensible markup language (XML) file to contain timestamps of when a participant types a document during a meeting [18, 19]. Attempts have been made to automatically identify where decisions are made during a meeting [20]. Features of the corpus such as the role of the speaker, prosody (pauses, energy, pitch, intonation) of speech, the lexical vocabulary being used and the distance in words from the end of a topic segment were suggested by the investigators to improve the automated results but at this time it still needs further development. Likewise automated argument diagramming has been attempted. The researchers found that prosodic information including a 200ms gap between a given utterance and the following and preceding utterance, which was tagged as a question, is very helpful in distinguishing from issues raised through questions as opposed to a backchannel (a short utterance from a listener that the speaker can continue to speak out of turn) [20]. These were similar results to an earlier study using a different tagging standard [20]. Further mention should be made of the CALO project, which “provides for distributed meeting capture, annotation, automatic transcription and semantic analysis of multi-party meetings”. The functionality of the CALO-MA architecture includes “real time and offline speech transcription, dialogue act segmentation and tagging” [21].

There has also been published work that has focused on the design and evaluation of user interface browsers and meeting mapping tools so that the meeting multimedia content can be viewed on a visual display unit. Buckingham Shum et al. have presented and replayed meeting content through a computer generated graphical map similar to a decision tree [22]. This type of approach has also been used for capturing design rationale in aerospace using visual symbols to represent items discussed like issue status, answers to questions and arguments [23, 24]. Another type of browser, developed by Jaimes et al. [25] was later evaluated in a study by Whittaker et al. [3]. The investigators found that despite good efforts, viewers using the browser misinterpreted the data unless they reviewed the meeting in its entirety to give context to the content. This particular browser failed user acceptance. Participants in the study wanted to focus only on the important parts of the meeting. They also wanted the browser to facilitate the easy navigation to summaries, agenda items, decisions and actions of the meeting.

III. ESTABLISHING A NEED

A Dynamic Knowledge Management framework has lead to a methodology being used in BAE Systems to uncover references to product development knowledge from analysis of F2F meetings [26]. This paper explains how the concepts used in that study can be further developed into a prototype F2F meeting capture and indexing process.

BAE Systems is a business that has a mission to deliver sustainable growth in shareholder value. There are thirteen phases that product development projects in BAE Systems may go through during the Product Lifecycle Management (PLM) process. Ongoing face-to-face (F2F) meetings required include bid status, contract, phase and design review meetings. There are

also specific technical bid reviews and requests for bid approval that can map onto the generic model. Although reviews look back to establish lessons learned and project maturity to provide a firm foundation for the review, their primary focus is to look ahead to the future targets for the project and to ensure that appropriate plans, processes and resources are in place to meet customer commitments.

Technology to support people learning and making sense of face-to-face (F2F) meetings would be of use to those who wish to enhance organisational hearing, sight and memory performance. It could assist people who do not currently have the knowledge or authority to act on what was discussed and so have to share and reflect on the F2F meeting content after the event - for example, checking with others before a decision is made and taking action.

The number of people that are effective in a F2F meeting depends on the complexity of the situation [2]. Generally in a complex product development environment, F2F meetings tend to be in smaller groups with less than a dozen people in attendance, with the meeting lasting an hour or longer. The meetings occur in comfortable indoor office rooms with people mostly in a seated position around a table with access to a computer, whiteboard and projector [27]. When more than twelve people are in a meeting then collaboration and interactions becomes more difficult and the meeting is more like a presentation with an audience rather than a group meeting. Therefore the need exists for a prototype F2F meeting capture and indexing process, which can easily capture the actual conversation and activity of up to twelve people having an unstructured (or semi-structured) conversation in a meeting room. Not all meetings take place in the same room so there is a need for the video and audio capture technology to be compact and portable.

Human decision makers can impact on society in a multitude of ways. Ontological concepts such as war, religion, science, technology and money are all constructs of human thought and decision making. These concepts have been formed and shaped with many F2F meetings over millennia. There is a need for an accurate record of what is said so that the outcome resulting from the actions checked in the F2F situation can be reflected on and reviewed. There is also therefore a need for a prototype to be able to track and monitor changes in performance following the actions of people after the meeting.

Since the interpretation of data from just one source can be biased or contain error there is a need to allow people to continuously improve the quality of the meeting content post publication.

IV. Proof-of-Concept

The prototype face-to-face (F2F) meeting capture and indexing process, which is part of the Dynamic Knowledge Management framework, is described in use with the continuous improvement Plan-Do-Check-Act (PDCA) cycle as a proof-of-concept (Figure 1).

After planning (P) to make a change based on the previous organisational results, the next step is to do (D) something to try and make an improvement. The outcome of these investigations would then be brought to a meeting pre-arranged through the organisations groupware technology. For example this could be Microsoft® SharePoint (Microsoft Corporation, Reading UK) which is popular with many large organisations. SharePoint uses the same Active Directory (AD) address book as email and instant messaging (IM) so it is possible to tag the content held in the infrastructure directly to the users that have uploaded (and downloaded). AD provides a proven method to apply security restriction to users or groups of users within the infrastructure. Different people in the F2F meeting with different viewpoints asked and answered questions; deliberated options of choice to make decisions; raised issues and reported on issue status.

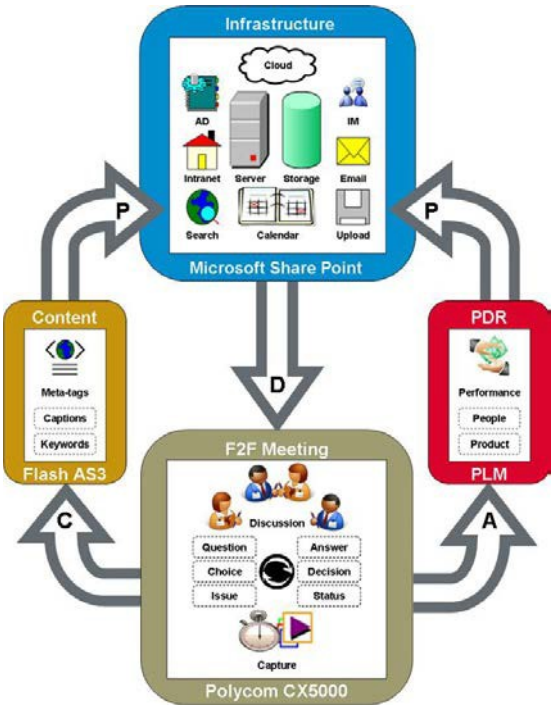

Figure 1. Example of a figure caption. (figure caption)

In the F2F meetings, video and audio data was acquired through one Polycom® CX5000 table mounted three-hundred and sixty degree web camera (Polycom, Slough UK) with PC and software. The video capture frequency was 15 Hz and audio microphone range 150 Hz - 3.4 kHz within the maximum room size 7.62 m x 4.57 m x 3.05 m. The Microsoft Live Meeting (Microsoft Corporation, Reading UK) software and Polycom® PVX™ (Polycom, Slough UK) software provided a Panoramic video resolution of 1056 x 144 pixels and audio. The output .avi file from the camera played with a YUV codec and was converted into smaller segments in Flash® video .f4v format.

Video segments were professionally transcribed by confidential (signed non-disclosure agreements) typists at a rate of £1.40 per minute with a 24 hour lag time from being uploaded through a secure server. The F2F meeting video and audio content in the .fv4 file was then linked to the captions with time codes compiled in XML format with Flash® AS3 code presented as a .swf movie (Figure 2).

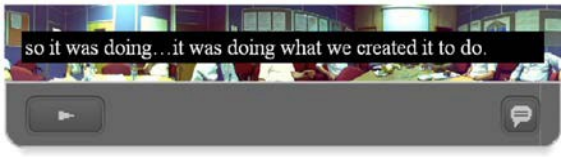

Figure 2. Flash Video with Captions

When published to a webpage in MS SharePoint there was an opportunity for the whole organisation (not just those that attended) to check (C) the original plan. Trade-offs could be evaluated or eliminated by experts who may not have been able to attend the meeting in person, but were able to review the corpus in a "Document Library" such as in the MS SharePoint screenshot below and make comments on the "Discussion Board" (Figure 3).

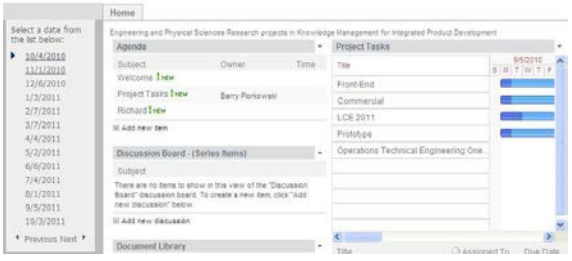

Figure 3. Example Groupware Webpage

If there is no need to check (C) and there has been authority granted at the F2F meeting to progress and act (A) on the proposed change shown above as a "Project Task" (Figure 3) then there should be some visible performance outcome improvements in the future. These would be either an improvement in the profitability of the products and/ or the people who had learnt something on reflection during the PDCA cycle. The profitability of the products is currently measured as part of the product lifecycle management (PLM) process and the employee learning as part of the performance development review (PDR) process. It is important that these are interoperable so that further continuous improvements can be planned on reflection and there is no electronic double handling and waste.

V. Further Work

More work should be done to make the captioning of the F2F meeting content automatic or semi automatic. Perhaps the editing of the XML document containing meta-data that informs the captions and keywords of the video and for search could be made available direct to web users. The presentation of the content in the groupware webpage may also be developed so that there is easier navigation to the page with the video segments. It would also be useful to have the variables that influence the profitability of the products to be displayed on the page so that changes can be tracked against business performance. This added functionality would require further interoperability with the current financial business tools used at BAE Systems, but would allow for changes to be effectively tracked against business performance.

Acknowledgment

This document is based on work funded by EPSRC through BAE Systems. The researchers are deeply grateful to Nick Martin, Clive Simmonds and all others for the support in making this study possible. Any opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect those of the EPSRC or BAE Systems.

References

[1] R. Parasuraman and C. D. Wickens, "Humans: Still Vital After All These Years of Automation," *Human Factors*, vol. 50, pp. 511-520, 2008.

[2] D. J. Snowden and M. E. Boone, "A leader's Framework for Decision Making," *Harvard Business Review* pp. 1-9, 2007.

[3] S. Whittaker, S. Tucker, K. Swampillai, and R. Laban, "Design and evaluation of systems to support interaction capture and retrieval," *Personal and Ubiquitous Computing*, vol. 12, pp. 197-221, 2008.

[4] M. Attaran, "Collaborative computing: a new management strategy for increasing productivity and building a better business," *Business Strategy Series*, vol. 8, pp. 387 - 393, 2007.

[5] I.-J. Kim, S. C. Ahn, H. Ko, and H.-G. Kim, "Automatic Lifelog media annotation based on heterogeneous sensor fusion," in *Multisensor Fusion and Integration for Intelligent Systems* Seoul: IEEE, 2008, pp. 703-708.

[6] D.-W. Ryoo and C. Bae, "Design of The Wearable Gadgets for Life-Log Services based on UTC," *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 1477-1482, 2007.

[7] B. H. Prananto, I.-J. Kim , and H.-G. Kim, "Multi-level Experience Retrieval for the Personal Lifelog Media System," in *Third International IEEE Conference on Signal-Image Technologies and Internet-Based System* Shanghai: IEEE, 2007, pp. 175-182.

[8] R. Crowder, G. Hughes, and W. Hall, "Approaches to Locating Expertise Using Corporate Knowledge," *International Journal of Intelligent Systems in Accounting, Finance and Management*, vol. 11, pp. 185-200, 2003.

[9] N. Khomenko, R. De Guio, and D. Cavallucci, "Enhancing ECN's abilities to address inventive strategies using OTSM-TRIZ," *International Journal of Collaborative Engineering*, vol. 1/2, 2009.

[10] M. Basadur, P. Pringle, G. Speranzini, and M. Bacot, "Collaborative Problem Solving Through Creativity in Problem Definition: Expanding the Pie," *Creativity and Innovation Management*, vol. 9, pp. 54-76, 2000.

[11] K. Karahalios and T. Bergstrom, "Social Mirrors as Social Signals: Transforming Audio into Graphics," *Computer Graphics and Applications, IEEE*, vol. 29, pp. 22-32, 2009.

[12] K. Bachour, F. Kaplan, and P. Dillenbourg, "Reflect : An Interactive Table for Regulating Face-to-Face Collaborative Learning," in *Times of Convergence. Technologies Across Learning Contexts*, 2008, pp. 39-48.

[13] T. Bergstrom and K. Karahalios, "Conversation Clock: Visualizing audio patterns in co-located groups," in *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*: IEEE Computer Society, 2007.

[14] J. M. DiMicco, K. J. Hollenbach, A. Pandolfo, and W. Bender, "The impact of increased awareness while face-to-face," *Human Computer Interaction*, vol. 22, pp. 47-96, 2007.

[15] M. Wald, "Captioning Multiple Speakers Using Speech Recognition to Assist Disabled People," in *Computers Helping People with Special Needs*, 2008, pp. 617-623.

[16] B. C. Roy and D. Roy, "Fast Transcription of Unstructered Audio Recordings," in *Interspeach 2009* Brighton, UK, 2009.

[17] S. Novotney and C. Callison-Burch, "Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription," in *Human Language Technologies: Association of Computational Linguistics 2010* Los Angles, California, USA, 2010, pp. 207-215.

[18] M.-M. Bouamrane and S. Luz, "An analytical evaluation of search by content and interaction patterns on multimodal meeting records," *Multimedia Systems*, vol. 13, pp. 89-102, 2007.

[19] M.-M. Bouamrane and S. Luz, "Uncovering non-verbal semantic aspects of collaborative meetings: iterative design and evaluation of the Meeting Miner," *Signal, Image and Video Processing*, vol. 2, pp. 337-353, 2008.

[20] P.-Y. Hsueh and J. Moore, "Automatic Decision Detection in Meeting Speech," in *Machine Learning for Multimodal Interaction*, 2008, pp. 168-179.

[21] G. Tur, A. Stolcke, L. Voss, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Hakkani-Tur, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, S. Peters, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, and F. Yang, "The CALO meeting speech recognition and understanding system," in *Spoken Language Technology Workshop, 2008. SLT 2008.* Goa, India: IEEE, 2008, pp. 69 - 72

[22] S. Buckingham Shum, R. Slack, M. Daw, B. Juby, A. Rowley, M. Bachler, C. Mancini, D. Michaelides, R. Procter, D. d. Roure, T. Chown, and T. Hewitt, "Memetic: An Infrastructure for Meeting Memory," in *7th International Conference on the Design of Cooperative Systems* Carry-le-Rouet, France: IOS Press, 2006.

[23] R. Bracewell, K. Wallace, M. Moss, and D. Knott, "Capturing Design Rationale," *Computer-Aided Design*, vol. 41, pp. 173-186, 2009.

[24] R. H. Bracewell, S. Ahmed, and K. M. Wallace, "DRed and Design Folders: a way of Capturing, Storing and Passing on - Knowledge Generated during Design Projects," in *2004 ASME International Design Engineering Technical Conferences (DETC'04)* Salt Lake City, Utah, USA, 2004.

[25] A. Jaimes, H. Bourlard, S. Renals, and J. Carletta, "Recording, Indexing, Summarizing, and Accessing Meeting Videos: An Overview of the AMI Project," in *Proceedings of the 14th International Conference of Image Analysis and Processing - Workshops* Modena, Italy IEEE Computer Society, 2007, pp. 59-64.

[26] B. A. Piorkowski and J. X. Gao, "A Case-study: Finding References to Product Development Knowledge from the Analysis of Face-to-Face Meetings," in *International Conference on Life Cycle Engineering* Universitat Braunschweig, Germany, to be presented 2011.

[27] G. Barczak, A. Griffin, and K. B. Kahn, "PERSPECTIVE: Trends and Drivers of Success in NPD Practices: Results of the 2003 PDMA Best Practices Study," *Journal of Product Innovation Management*, vol. 26, pp. 3–23, 2009.