SMART MEETING MINUTES APPLICATION SPECIFICATIONS

Version 2.4, August 2002

Stephane.Sire@epfl.ch, Denis.Lalanne@unifr.ch

This document presents the global specifications of the Smart Meeting Minutes Application (SMM), the core project of IM2. It describes the objectives, requirements, functional parts and the various annotations of the meeting data. More generally, we expect this document to generate feedbacks and to improve IPs coordination, especially concerning the annotations of the recorded meeting data.

Certain parts of the document are highlighted (in light gray) in order to extract important definitions and requirements for all the IPs.

SMM APPLICATION SPECIFICATIONS

Version 2.4, August 2002

Stephane.Sire@epfl.ch, Denis.Lalanne@unifr.ch

1 Project goal and overview 2
2 Meeting form and style 2
2.1 Meeting room layout 2
2.2 Participants' roles 3
2.3 Meeting structure 3
2.4 Meeting documents 3
3 System requirements 4
3.1 Hardware configuration and room setup 4
3.2 Audio and video quality 4
3.3 Audio and video storage 5
3.4 Audio and video synchronization 6
4 SMM Functional specifications 7
4.1 Functional diagram 7
4.2 Meeting Organizer system 8
4.3 Meeting capture and recording system 9
4.4 Data integration and storage system 9
4.5 Retrieval system 11
5 Smart meeting minutes and annotations 12
5.1 Inventory of annotations 12
5.2 From annotations to minutes 17
5.3 Summary of IP's contributions 18
5.4 Note on annotations and XML 19
6. Related works 19
Conclusion 20
References 20
Annex 1: Glossary 21
Annex 2: Global meeting descriptors DTD 24
Annex 3: Sample meeting recording 26
Annex 4: Sample queries 34

1 Project goal and overview

The goal of this project is to develop the Smart Meeting Minutes application (later referred to as the SMM application), a multimedia workspace that enables people to organize, record, archive and retrieve meetings extracts and meetings minutes.

The SMM application is constituted by a set of hardware devices and software tools. It aims at different users:

- Meeting participants will use the system to prepare and to record meetings on different media (audio and video) and to archive meeting documents.
- Multimodal interaction researchers will use the recorded data to stress and to improve multimodal analysis algorithms. These algorithms will produce sets of annotations of the captured media (audio, video and documents) according to different modalities (including but not limited to speech, gestures, facial expressions). The annotations will be fed into a multimedia meeting database where they will provide convenient indexes to the captured media.
- Meeting participants and other authorized external people will be allowed to interact with the multimedia meeting database to retrieve diverse forms of meeting minutes on different kind of devices (including but not limited to personal computers and PDAs).

This document describes the overall specifications of the system. Its purpose is to clarify the use of the system from the previous users' perspectives and the expected contributions of each IP. First, it gives the context of the application by describing the type of meetings that will be recorded. Second, it lists the main system requirements. Third, it presents the four logical components of the system and the way they relate to each other. Finally, the last section is dedicated to the annotations to collect in order to generate meeting minutes† from the database.

As this document is targeted at an audience of experts in different domains (speech, dialog, video analysis and document analysis, human-computer interaction), it may use some terms inaccurately. Some of the terms have been defined in a glossary in Annex 1. They are marked with an † after their first occurrence in the text.

2 Meeting form and style

The meetings of IM2 researchers will constitute a test-bed for the SMM application. These are **scientific coordination meetings** characterized by decision-making, planification and analysis tasks. As described in this section, this type of meeting provides a simple yet powerful framework. In particular, it provides a natural yet controlled situation, where lot of multimodal† interactions should occur between participants. While reading this section, you may refer to the annex 3 to view a sample meeting agenda.

2.1 Meeting room layout

The participants are seated around a table in a room tailored for up to 12 participants. Each participant is filmed by a fixed camera and is recorded by a personal microphone. The personal cameras are placed so that each meeting member can see all the other ones without moving and without being disturbed. For that purpose, the cameras are settled in the middle of the table in front of each participant. One or two extra cameras are used to film an overview of the meeting room.

The addition of a whiteboard and/or of a projection screen to the meeting room is considered as an extension to the standard SMM application. In that case, the participants will be placed so that they can observe the screen(s) without changing place. It is also possible to have an interactive electronic whiteboard (such as a SmartBoard) providing both the benefits of a projection screen and of a whiteboard for handwriting.

2.2 Participants' roles

The meeting is organized on the behalf of one of the participants called the organizer. Before the meeting, s/he is charged to write an agenda, to gather documents related to the meeting and to send them to the participants. During the meeting, the organizer or another volunteer participant plays the role of a president and/or moderator. The president's main task is to follow the meeting agenda and to allocate speech turns to the participants. The SMM application purpose of automatically producing meeting minutes removes the need for a scribe (or a secretary) taking notes during meetings.

In the standard SMM application, the participants use only paper documents distributed before or during the meeting. Thus they should not need to move away from their regular place. In extended versions with a whiteboard and/or a projection screen, some participants may become presenters and make outward and arrival journeys towards the screen and/or the whiteboard.

2.3 Meeting structure

The meeting president allocates speaking turns in order to minimize speech overlapping. This imposes a relatively rigid structure on meetings, with the use of explicit gestures for requesting speech floor. We recommend a raised hand gesture, as it should be easy to track on video recordings. The existence of an agenda is also an important point as it simplifies the description of the meeting structure. When the agenda topics are discussed sequentially, they provide a natural index to audio and video segments.

The classical structure of a meeting is a sequence of **thematic episodes**. Each thematic episode involves one or more speakers engaged in a dialog about an agenda topic, or about a new unplanned topic. Each episode is composed of **sections** themselves composed of **turns** where each turn is defined as a contiguous part of speech from one speaker. Finally, each turn can be decomposed into **utterances** with an utterance being a small coherent part of one speaker's speech (more or less corresponding to a sentence in a written document).

The official language for the meetings is English. However, for demonstration purposes where speech recognition will be replaced by a manual annotation process, we may use other languages.

2.4 Meeting documents

The documents used or mentioned during the meeting are integrated into the multimedia† meeting database. Documents contribute to the definition of the meeting focus just as speech and others modalities. This is particularly true for the agenda. The SMM application will provide some tools to gather these documents.

By documents, we mean any piece of information, which can be printed on paper or displayed on a screen. To simplify document analysis at first, we exclude all the documents containing animations or temporal data, such as multimedia clips†, from the document analysis process.

Documents available only in paper form will be scanned and converted to PDF after optical recognition. All the other documents will also be converted to PDF for simplifying their delivery and/or their analysis. The output of a document analysis is a description of the document structure†. For instance, structural elements such as headings, paragraphs, tables and lists are tagged as part of the document. The structure gives important clues about the semantic of a document. It can be used to extract useful information that can be linked across the other media (audio and video). For instance when participants are talking about a budget table, it is up to the document analysis modules to extract that table from the set of available documents.

Usually we distinguish between a physical and a logical view of a document structure. In the former case, a layout structure† describes the general presentation in term of blocs with typographical attributes. It represents the document as it is printed or displayed on a screen. In the latter case, logical structure† is considered; it describes the abstract organization of the document content in term of chapters, sections, lists, captions, and so on. For the SMM application, both types of structure have to be recovered by the document analysis tool in order to provide the necessary information for a suitable indexing.

The document analysis in the standard SMM application includes the analysis of figures (such as tables, or diagrams). However, the analysis of photographs or non-vectorial graphics inside documents is left as a future extension for any interested IP. Former, handwriting recognition could be applied to the notes taken by participants, either collaboratively on a shared whiteboard, or individually on a notebook.

3 System requirements

The SMM application intends to minimize the perturbations brought to a meeting by the audio and video recording. In that sense, it is different from augmented conference rooms or from Group Decision Support System (GDSS) that support meeting participants in real time at the cost of users' interactions with the room equipments. Our purpose is to obtain high quality meeting recordings† and to postpone users' interactions with the system for after the meeting, when users interact with the recorded materials. Thus our requirements are different from augmented conference rooms and GDS systems. This section describes our requirements and some of the targeted performances.

3.1 Hardware configuration and room setup

The meeting room hardware setup must be **scalable** to fit meetings from 2 participants up to 12. Scalability should have little impact on the hardware volume and on the noise level to keep the room as a real meeting room and not as a computer room. In section 4.3 we present a scalable hardware configuration based on a replicable basic capture box, which is a screenless PC connected to a few microphones and cameras.

The area covered by each participant' camera will be materialized with small marks around the table. Thus, it will be possible to record meetings without disturbing participants with a video feedback channel to align themselves with their close-up camera. Privacy of the audio and video data collected will be guaranted afterwards by adequate access control methods. If necessary, the application will use microphones with an on/off button to individually control audio recording. The effect will be to insert silence in the audio track. A similar feature for controlling video cameras will be considered as an extension to the standard SMM application.

3.2 Audio and video quality

The audio and video media have two purposes in the SMM application. The first purpose is to be analyzed for producing relevant annotations. The second purpose is to be inserted in multimedia clips when retrieving meeting minutes from the database or to replay meeting extracts. Consequently, audio and video need to be stored and delivered in two different quality levels: a **data analysis quality level** for data analysis and a **broadcasting quality level**. The broadcasting quality level may itself be divided into sub-levels taking into account different bandwidths and screen sizes, as it is common for multimedia broadcasting. The tables below summarize our choices of audio and video formats.

Audio quality level	Format
data analysis	uncompressed 16 KHz / 16 bits bandwidth: 31.25 KB/s

broadcasting	MP3 but legal problem of fees
	any other available codec ex: Qualcomm Purevoice

Table 1: audio formats

As indicated in the table above, they are legal issues about the choice of an MP3 codec for audio broadcasting. Any input on that subject is welcome.

The SMM application is based on a fully digital video recording solution. This is a convenient solution permitted by the availability of small Firewire† and USB† webcams on the market at attractive prices. The goal is to get an image quality at least as good as the analog VHS format. After a few preliminary tests, we propose to capture raw video in two different resolutions: 640x480 for the overview cameras and 320x240 for participant's close-up cameras. This should optimize the storage requirements, as video is very demanding in disk space.

The targeted raw video frame rate of the standard SMM application is of 15 frames per second. This is a convenient choice to allow the connection of multiple cameras on a single capture box. If necessary, and in conformance with other IPs needs, some extensions of the standard application may allow a better frame rate for a few specialized cameras. We propose to use a lower frame rate of 12 fps for broadcasting. The broadcasting video codecs are usually parameterized to compress for a given bandwidth.

Video quality level	Format
data analysis: meeting room overviews	640 * 480 15 fps non interlaced uncompressed YUV 422 (16 bits by pixel) bandwidth: 8.8 MB/s
data analysis: participants' close-up views	320 * 240 15 fps non interlaced uncompressed l420 (12 bits by pixel) bandwidth: 1,65 MB/s
broadcasting	MPEG-4 or H-263 320*240 12 fps bandwidth: variable

Table 2: video formats

It is important to notice that at this stage of the project specifications, we have retained some uncompressed Y'CbCr video formats, which only perform some reductions of the color space. These formats correspond to the signal delivered by Firewire webcams. The next section on storage requirements shows that these formats are still very demanding in disk space.

3.3 Audio and video storage

The estimation of the required storage capacity depends on the following variables:

- number of participants (with one camera / microphone for each participant);
- number of extra cameras for meeting room overviews;
- meeting duration;
- number of meetings to archive in the long term.

The table below gives an estimation of the storage space required. It corresponds to data analysis quality audio and video. Volumes have been calculated according to the bandwidths given in the previous section.

Meeting duration	1 audio	1 video close-up 320x240	1 video overview 640x480	Total 6 participants 1 overview camera
1 minute	1,83 MB	99 MB	528 MB	629 MB
10 minutes	18,31 MB	990 MB	5,16 GB	11 GB
1 hour	110 MB	5,8 GB	31 GB	66,4 GB

Table 3: storage requirements (top estimation, data analysis quality)

According to that table, it is realistic to store data analysis quality audio on CD-ROM and/or DVD (one DVD contains up to 4,7 GB). It is also realistic to store a few hours of data analysis quality video on a disk server, but we have to find other solutions if we want to keep more than a few hours of meeting. Some of these solutions may be:

- to use a more compact format than uncompressed Y'CbCr;
- to store video at data analysis quality level only during video analysis and to keep just the broadcasting quality level when it has been processed;
- to find other mass-storage media.

The standard SMM application will compress video in order to store it at an acceptable image quality level for data analysis with lower space storage needs than raw Y'CrCb. The choice of a video codec is an opened issue that will be solved in collaboration with all the IM2 IPs. For instance, preliminary tests with atemporal compression formats such as Photo-JPEG show a reduction of up to 10 times of the storage requirements.

The standard SMM application will keep only the most recent video recordings at the data analysis quality level on file servers. The former recordings will be moved to another less expensive storage support. The broadcasting quality video data requires much less storage space. Consequently it will be stored on broadcasting servers that can store hours of meeting recordings. It is also possible that future improvements in compression algorithms will make broadcasting quality sufficient for video analysis too.

3.4 Audio and video synchronization

The goal of synchronization is to guaranty a regular recording of each audio and video track (intra-stream synchronization) and to get audio and video tracks aligned together (inter-stream synchronization). We consider synchronization during playback as another issue that is handled by multimedia streaming servers and by player applications.

The standard SMM application will record audio at least at 16 KHz / 16 bits. It will record video at 15 frames per second without dropped frames. Video jittering should not exceed one frame every two minutes. Video jittering can be estimated with the chronograph test (cf. Figure 1). This test consists in recording a chronograph with a 1/60 sec clock hand. If there is not jittering, the clock hand should appear at exactly the same position every 15 frames. If we admit one frame of jittering every two minutes, the clock hand position is allowed to vary by 24 degrees (360 / 15) every 120 turns. The chronograph test can also be used to check that there are no dropped frames. The intra-stream recording constraints are summarized below:

- be able to retrieve the time code of each frame in a video file;
- be able to retrieve a sample sound starting at a given time code in an audio file;
- to check that no frame has been lost;
- to check that video jittering is less than 1 frame every two minutes.

The inter-stream synchronization will be achieved at two levels. First, each pair of a participant's microphone and of a participant's close-up camera will be recorded together. Within each pair, the shift between audio and video will not exceed a duration equivalent to 3 frames (0,2 seconds when recording at 15 fps). Second, all the synchronized audio and video pairs will be synchronized together. That step implies a post-synchronization process for which we will admit a maximum shift between two different pairs of audio and video of a duration equivalent to 5 frames (0,3 seconds when recording at 15 fps). The same constraints apply to the synchronization of the video overviews with the different pairs of participant's audio and video.

Figure 1: The chronograph test

4 SMM Functional specifications

This section describes the overall SMM application architecture. It presents a functional diagram of the main parts involved. Then it describes each part.

4.1 Functional diagram

The SMM application is made of four distinct logical systems. The main purpose of each system is to collect and/or to process different kinds of input data. Each system also produces new data for the other systems. The systems may work in parallel or sequentially depending on the type of data they manage. Figure 2 summarizes the roles and relations between the different systems. Each system is detailed in the following sections. The integration and storage system is of most concern to the other IPs, as it will coordinate the process of producing annotations from all the available media (audio and video and documents).

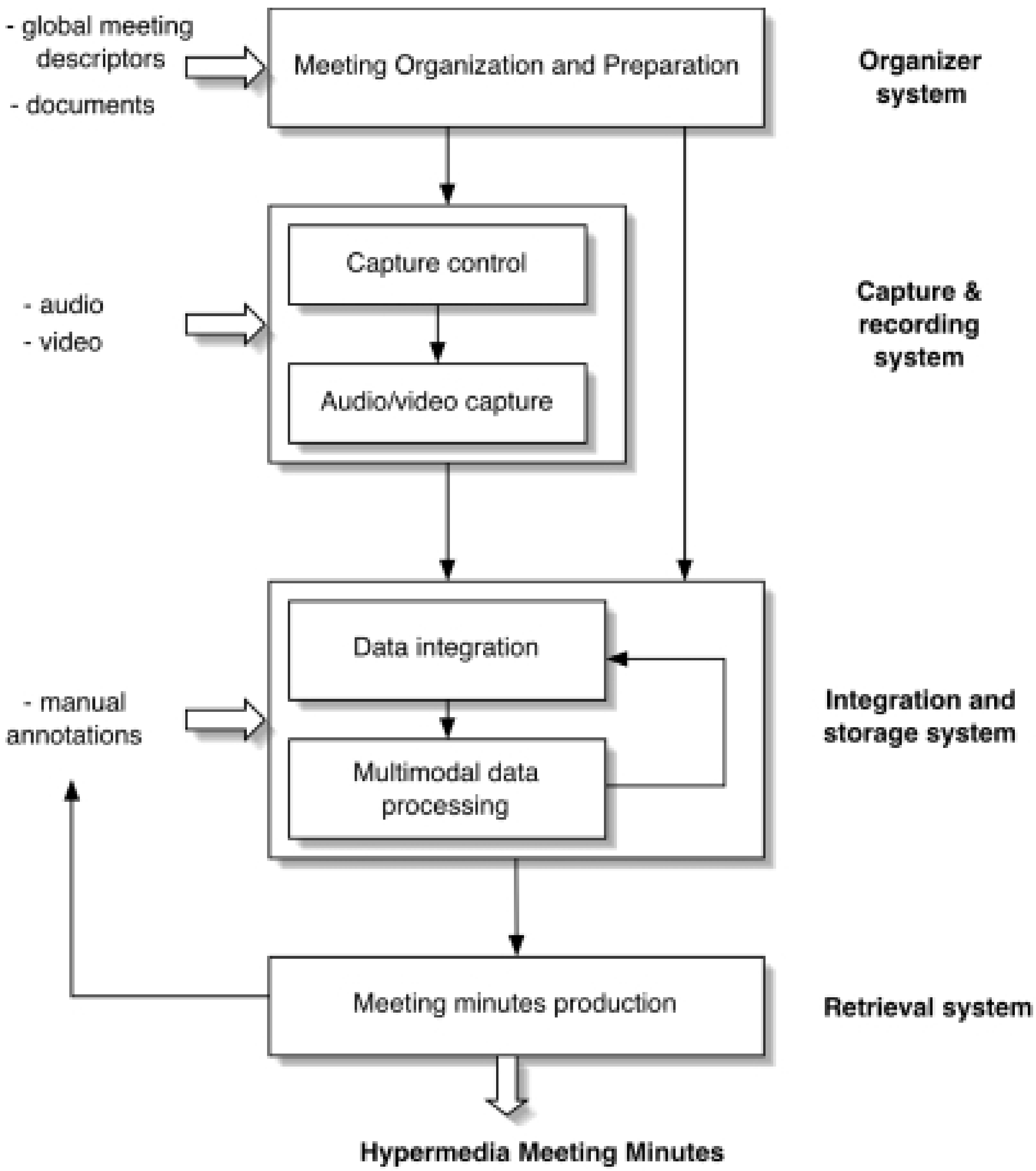


Figure 2: SMM application functional diagram

4.2 Meeting Organizer system

The organizer system manages the collection of **global meeting descriptors** and of the meeting documents. It also prepares some configuration parameters for the capture and recording system.

The global meeting descriptors include:

- meeting contextual information (place, time, duration, subject, etc.);
- participants' related information (identity, sex, photograph sample, voice sample, etc.);
- meeting documents descriptors (authors, file names, etc.);
- participants' seat around the meeting table.

The global meeting descriptors are gathered before and after the meeting. They are stored in the multimedia meeting database along with the meeting documents. The Annex 2 gives a DTD for the global meeting descriptors. In the standard SMM application, these descriptors are edited by an assistant in a customizable editor, such as XML Spy (using XSLT). We plan to improve this process in the future with a collaborative web application. This application will also support some meeting organization functionalities such as automatically sending a meeting invitation by email to the participants.

Once the global meeting descriptors are collected, they are transferred to the integration and storage system for inclusion in the multimedia meeting database. Then, they become accessible to other IPs during the data analysis phase. In that respect, each IP is welcome to define any global descriptor that it may need during analysis. For instance, we have suggested some photographs of participants or some voice samples for automatic identification.

The second purpose of the organizer system is to compute some parameters used internally by IM2.AP for configuring the capture and recording system. These parameters are used to configure the meeting room (number of cameras, microphones, etc.).

4.3 Meeting capture and recording system

The capture and recording system is responsible for recording audio and video, for synchronizing the different tracks, and for transcoding them to the required quality levels. For each meeting, its input is a configuration file generated by the organizer system. Its outputs are a set of audio and video files and a description of the mapping of these files with the cameras and microphones.

The capture and recording system is based on a modular setup made of a combination of **capture boxes**. Each capture box is a small PC that records a few microphone and camera pairs. First tests have shown that we can put at least two USB and/or Firewire cameras and two USB microphones on a standard PC. Thus for a 6 participants meeting, we plan to use 3 PCs. An extra PC may be required to add overview cameras. The capture boxes are also connected to a central **control capture box** from where an assistant can start and stop recordings. The capture boxes and the control capture box are networked together with a low latency network to get a reliable network time. We propose to use a fast Ethernet switch, which is disconnected from external networks during meeting recordings.

During a meeting, the capture boxes physically record and segment video files to keep manageable file sizes. A limit of 2 GB per file seems to be a good setting. The files are stored locally on each capture box hard drives. An **audio and video file map** keeps a track of the file segmentation and of the file camera and/or microphone association.

After a meeting has been recorded, the capture boxes are used to transcode audio and video to the data analysis and broadcasting format. Then, the files are moved to the multimedia meeting database for annotation.

4.4 Data integration and storage system

The first purpose of the integration and storage system is to gather all the raw meeting data (audio, video, documents, global descriptors) and all the annotations produced within each IP. The second purpose is to make this data available to every IP, in predefined formats, and through predefined Application Programming Interfaces (API). Finally, the last purpose is to control and to coordinate the annotation process.

The data gathered by the integration and storage system is stored in a **multimedia meeting database**. The exact database schemas and the database architecture are not in the scope of this document. They are defined in a joint effort between IM2.DS and IM2.AP and any other IP with special requirements regarding audio and video data access and/or annotations.

The database system is a storage and query server. It serves as a front-end to a network file server for delivering audio and video at the data analysis quality level, and as a front-end to a streaming server for delivering audio and video at the broadcasting quality level. The network file server will be accessible through standard file sharing protocols (NFS, FTP, webDAV, etc.).

Data taxonomy

The database contains at least the following objects:

- One **Meeting** object that represents the meeting as a whole, with attributes such as location, date, time, etc.
- One **Participant** object is associated with each participant. It contains some biographical information.
- One **Agenda** object made of **Topic** objects that represent the linear sequence of topics which have been discussed during a meeting.
- Multiple **Stream** objects that refer to every audio or video tracks. Stream objects can be annotated through a referencing mechanism that defines anchor points or segments such as "the audio stream *x* data at time *t*" or "the video stream *y* between *start-time* and *end-time*".
- Multiple **Document** objects that refers to every document file. Document objects can be annotated through a referencing mechanism that defines anchor points or segments such as "the table named *budget* in the document *x*".
- Multiple **Annotation** objects. An annotation describes some properties of one or more objects (including other annotations) stored in the database. For instance, the speech transcript annotation is the annotation of an audio stream, which contains a textual transcript of what is said in some fixed audio segments.
- One **Minutes** object made of **Minutes item** objects. The minutes are similar to the agenda but they describe what has really happened during a meeting. The section 5 gives an overview of the minutes purpose and production.

Some of the annotations may be "hard-wired" into the database schema for efficiency. For instance, the discourse structure into thematic episodes, sections, turns and utterances may be part of the initial database schema. Thus there will be extra objects such as episode, section, turn and utterance. We must notice here that with carefully designed database objects, it is possible to consider some annotations as attributes of these objects. For instance, if there is an object utterance, the dialog act tagging (an annotation like "greetings", "question", "commitment", etc.) of every utterance becomes an attribute of that object. Whatever annotations are hard-wired into the database, the system should be flexible enough to allow the definition of new annotations, as they are required.

Data integration API

The initial database contains the meeting object, one participant object for each participant, and one stream/document object for each audio/video/document file. The physical segmentation of audio and video files into multiple "small" files is transparent at that level. Then IM2.AP adds the agenda and topic objects (probably after a manual indexing process). Next, the IPs concerned with speech add speech related annotations (speech segments, speech transcript, etc.). Next, IM2.MDM adds the annotations that enable the creation of the discourse objects. In parallel, all the IPs add their custom annotations such as annotations of document objects, annotations of video streams and/or of video images extracted from the video streams. Once the database has been filled, the scribe can access it to create the minutes related objects.

The database provides a specific data integration API to retrieve audio, video and documents during the annotation process and to store new annotations. This API is also used to retrieve annotations as they are becoming available, for producing new higher level annotations. A part of this API may be shared with the retrieval system described in section 4.5.

The data integration API can be divided into the following sub-categories:

- **Generic meeting access:** to store and to retrieve descriptions of the meeting and of all the other objects stored in the database, including the "hard-wired" annotations, but excluding the other annotations. This part of the API support queries such as "get all the meeting document objects" or "get the audio stream for Denis".

- **Stream and document access:** to store and to retrieve parts of audio and video tracks, and document parts for data analysis. For audio and video, in a first stage, the data returned by this API can be a list of file names with time boundaries for the segment to extract from each of these files. The files will be available on a network file server through standard file sharing protocols. In a second step, further developments are needed to offer a more complete API that reconstructs on the fly an exact audio or video file and send it as a response stream. This part of the API support queries such as "get video from the video stream *x* from *start-time* to *end-time*" or "get the 2nd table in the document object *z*".
- **Annotation access:** to query the database about the annotations available in the database (introspection purpose), to check the completion of the annotation process for each type of annotation (useful to build workflow management tools for the annotation process), to store and to retrieve annotations in an easy to parse format. This part of the API support queries such as "get the name of the current speaker at time *t*" or "get the speech segmentation of the audio stream *x*".
- **Annotation definition:** to define and to enter new annotation formats in the database and to include them in the annotation process.

The data integration API may require other specific APIs. The exact specification of data integration API will be the subject of a report written by IM2.AP and IM2.DS as the database API becomes available to other IPs.

Opened question

We do not know yet if it necessary to create an additional audio stream obtained by merging all the other audio streams. This global meeting audio stream could be useful to playback all the speakers at once. In case it exists, this audio stream could also become a reference stream for some of the audio annotations.

4.5 Retrieval system

The retrieval system produces usable views of the multimedia meeting database. By "views" we mean dynamical documents, which are generated on-the-fly as an answer to one or more queries to the database, and more advanced user interfaces, which allow more intimate interaction with the database content or with extracted parts of the database. These views will be accessed by at least three kind of users: 1) the meeting participants, 2) some scribes that participate into the production of meeting minutes and, 3) people external to the meeting authorized to access the meeting database. The following short scenarios give some examples:

- Pierre has participated to a meeting. Back in his office and before calling Bridget who also attended to the meeting, he wants to quickly review what she has said.
- Bruce has been appointed as a scribe to add minutes to the meeting database. These minutes are a short textual summary for every item discussed during the meeting and an account on each vote and its result. Bruce uses the retrieval system to visualize parts of the meeting.
- Sarah didn't attend to a meeting, but her boss has asked her to make a written summary of the meeting. She uses a web browser to retrieve and to display an Hypermedia Meeting Minutes† document generated from the minutes added to the database by Bruce.
- Judith is on the road to a computer shop where she is supposed to buy some new webcams for a new Smart Meeting Room. She doesn't know yet which cameras to buy, but a meeting has just been held by her colleagues to decide that point. Using her

mobile phone directly from the computer shop, and interacting with voice, she asks to the Retrieval System what cameras have been selected for the new room.

These scenarios illustrate three types of functionalities that the retrieval system should support. First, it should support a **query engine** to access database content. This query engine may share some queries with the data integration and storage system. Second, the retrieval system should support a **multimedia generation service**. This is a service which converts the results of the queries, which can include texts, images, documents, audio and video into a multimedia format. Finally, the retrieval service must support a **dialog manager service** for natural language user interfaces. A dialog manager infers users' goals and generate speech acts from internal users' representations and internal discourse models.

The query engine of the SMM application will allow a scribe to extract significant audio and video extracts of a meeting. These significant extracts will be based on the available annotations. Once the scribe has viewed these extracts, s/he will write and add related minutes into the database.

The standard SMM application query engine will be based on IM2.IIR work on the multimedia query language MRML. Thus it should be reusable and easily extensible. The extension of MRML for querying multimedia meeting data requires collaboration between all the interested IPs.

The multimedia generation service of the standard SMM application will be based on the XSLT transformation engine. In that context, data returned by the query engine is an XML† document, which is processed with one or more style sheets to generate one or more target documents that may include multimedia presentations defined in SMIL†. The audio and video extracts of the multimedia presentations are received in real time from a streaming server. This schema is well suited for web based interfaces. For more advanced interfaces, we will define custom multimedia presentation formats and build our own players.

The standard SMM application is an opportunity for collecting a corpus of annotated meeting recordings. As these corpuses become available, any IP with knowledge in natural language dialog systems, such as IM2.MDM, are welcome to use the SMM application as a test-bed for advanced dialog manager services. These will be integrated into extensions of the SMM application. These dialog management services will also benefit from personal identification technologies developped by IM2.ACP.

5 Smart meeting minutes and annotations

The smart meeting minutes are a natural extension of traditional meeting "minutes" to the domain of multimedia meeting database. They are stored as special minutes objects that summarize in textual form the significant meeting moments and link these moments with the available media (audio, video and documents).

This section presents the set of annotations collected by the standard SMM application. It also describes briefly how to use these annotations to support the task of a scribe adding smart meeting minutes to the database. Finally, it summarizes each IP's contributions to the annotation process. This section does not pretend to be exhaustive in the choice of interesting annotations. The notion of "interesting annotation" depends itself highly on the targeted usage for the database. That presentation is guided by an intuitive analysis of what could be useful to a scribe writing minutes.

5.1 Inventory of annotations

The annotations are grouped into the following categories:

- global meeting descriptors (description of meeting settings and agenda);
- audio annotations (annotations of audio files);
- discourse annotations (description of the structure of a meeting in terms of dialogs)
- document annotations (annotations of meeting documents);

- images / video annotations (annotations of video files).

For each category of annotation we define one or more annotation classes. Each annotation class is instanciated into multiple annotation objects in a multimedia meeting database. It is convenient to define **the value of an annotation** as the possible value(s) taken by the annotation and **the domain of an annotation** as its referred object(s). It is also convenient to define an abstract domain that we call "the meeting". Depending on the context of use, this domain corresponds to the meeting object in the database and/or to all of the audio/video streams.

Global meeting descriptors annotations

AG1. Meeting settings

Information about a meeting context. Its values are a set of attributes such as: its location, title, date, starting time, and duration. Its values also describe the meeting room configuration (seats, close-up cameras, microphones, and overview cameras). It depicts as well a list of the meeting participants with their characteristics (role, expertise, a photo, seat in the room, close-up camera and personal microphone, etc). It finally contains a list of the documents brought to the meeting. Its domain is the meeting. The DTD given in Annex 2 includes the meeting settings.

AG2. Agenda (or topics)

List of the topics that were planed to be discussed during a meeting. The value of each topic annotation is the topic name and a list of documents attached to that topic. The domain of each topic is its corresponding physical bloc in the meeting agenda document and the associated meeting segment. The first part of the domain will be obtained with document analysis methods, the meeting segment will be identified manually. The DTD given in Annex 2 includes the agenda. The Annex 3 contains a sample agenda document.

Audio annotations

AS1. Speech segmentation

Segmentation of audio streams into speaking versus silent parts. Each speaking part is a "speech segment" defined by time boundaries. Example: "the audio-x contains speech from *start-time* to *end-time*". Its values are "speech" or "silence" and its domain is an audio stream segment.

AS2. Speech transcription

Textual transcription of what is said on a speech segment of an audio stream.

AS3. Speaker

Annotation of the meeting that gives the name of the speaker (or of the speakers in case of speach overlapping) at any time during the meeting.

AS4. Non-verbal acoustical clues

Annotation of audio streams that identifies non-verbal acoustical clues (such as "applause", or "laugh") and their time boundaries. Example "the audio stream *x* contains applause from *start-time* to *end-time*". Its values are depending on a lexicon of non-verbal clues which is still to be defined.

Discourse annotations[1]

Discourse (linguistic) phenomena related to multiparty dialogs are annotated after the audio annotations. These phenomena are related to the structure of the dialogs, in terms of episodes, turns and utterances, but also in terms of the entities referred to and the role of each utterance in the evolution of the whole dialog. In principle, these annotations are made on the speech transcriptions of each participant's audio channel. They rely at least on the transcribed words and on temporal markers inserted between them. The temporal markers allow discourse annotations to

[1]this section has been writen in collaboration with Andrei Popescu-Belis from IM2.MDM

be mapped to audio or video streams. A first lexicon for dialog annotations will be defined by IM2.MDM and may evolve during the project.

A meeting (or its transcription) is divided in several **episodes** (thematic segmentation). Each speaker produces 0 or more **turns** in an episode, each turn being composed of 1 or more **utterances**. An utterance has an associated **dialog act** (such as "question", "order", "acceptation"), and two utterances may constitute an **adjacency pair** (such as "question/answer", "greeting/greeting", etc.). An utterance can mention 0 or more **named entities** (such as a speaker, Albert Einstein, Mt. Everest, etc.), and some named entities are **coreferent**, i.e. they point to the same person or object.

ADI1. Utterances + Dialog Acts

Utterances decompose the (transcribed) speech segments of audio streams into atomical units of discourse, each one corresponding to a dialog act. Utterances more or less correspond to simple sentences in written documents. There are two aspects in ADI1 : 1) segmentation of a speech transcription into utterances, 2) association of each utterance with the correct dialog act (from a predefined list).

ADI2. Discourse relations between utterances : Adjacency Pairs

Utterances can be grouped into higher level discourse structures. On one level of grouping, which must be tractable enough for an application, the adjacency pairs correspond to sequentially connected pairs of utterances, for instance a question and an answer, a request and a refusal, etc. Adjacency pairs are not necessarily sequentially connected, and it is not always possible to identify adjacency pairs all along the dialog (cf. breakpoints in a discussion).

ADI3. Turns

A turn is a sequence of speech or transcription between two pauses of the same speaker. A turn has one or more utterances. Turns correspond to the "dash + paragraph" notation of direct dialogs in literature; a dash marks a new turn.

ADI4. Thematic Episodes + Sections

Thematic episodes are a higher level discourse structure which gather turns into units corresponding to a coherent theme. For instance, the meeting can be divided into several episodes corresponding to the meeting agenda topics. In that respect, some of the thematic episodes defined in ADI4 will also be linked with the topics defined in AG2. It is also possible that some thematic episodes will correspond to new unplanned topics introduced during the meeting. The domain of ADI4 is list of turns. As necessary, IM2.MDM may also introduce some intermediate section levels between thematic episodes and turns.

ADI5. Named Entities (NE) + Coreference

Named entities are expressions in the (transcribed) speech that refer to persons, objects, ideas in the real world (e.g., proper names, acronyms, definite descriptions, etc.). It is essential, for dialog retrieval, to identify named entities in utterances--for instance to assign a topic to a dialog fragment. NEs that refer to the same entity are called coreferent, and these links are also useful, for instance to identify which are the entities most referred to in a dialog (the "main entities").

Documents annotations

ADO1. Document structure (layout + logical)

Description of the structure of a document (layout and logical†). The value of a layout document annotation is a set of typographical attributes (i.e. a font, a font size, etc.) and/or any other relevant attributes. Its domain is a set of blocs in the physical document. The value of a logical document annotation is an element type (like "heading", "image", "paragraph", etc.) and/or any other semantic attribute (such as a list of named entities). Its domain is a set of blocs in the physical document. The topics annotation (see AG2) is a particular type of document structure annotation that identifies agenda topics in an agenda document.

ADO2. Document alignment

The document alignment annotation indicates which document physical blocs, if any, are in the focus of the meeting at any time. It is similar to the speaker annotation (AS3). Its value is a set of physical document blocs (see ADO1). Its domain is a meeting segment during which the blocs are in the focus (because they are either explicitly referenced in speech or displayed on a projection screen).

Video/image annotations

Image annotations refer to an image from a video stream. That image, or key frame, can be defined as a time code reference to a video stream or it can also be extracted and stored in the database for a quicker access. The video annotations can apply to a video extract (with start time and end time), but they can also be applied only to one image.

AV1. Gestures

Gestures annotations describe when and where a particular gesture is visible on a video stream. Participant's turn taking and voting actions will be annotated in the standard SMM application. The corresponding gesture is a raised hand. A more complex gesture lexicon will be defined for future extensions in collaboration with IM2.SA, IM2.IIR and IM2.ACP.

AV2. Facial expressions

Facial expression annotations describe the apparition of some stereotyped participant's face appearance over a video stream. Facial expressions can be described as basic emotions, but other lexicons can be defined. In the standard SMM application we plan to extract only low level information about facial expressions such as "talking face", "smiling face" and "listening head". Facial expression characterization will be defined with the experience of IM2.VE.

AV3. Eye gaze

Eye gaze direction consists in a description of the evolution of the gaze direction of each participant on a video stream. On a higher level, eye gaze direction can be translated into a "looking at" annotation that describes who is looking at who during the meeting.

Summary

The table below summarizes the previous annotations, their possible values, their domain (objects to which they refer) and proposes on or more IP(s) for their production.

SMM standard annotations	**Values** **Domain**	**IP**
AG1. Meeting settings	strings meeting	IM2.AP
AG2. Agenda topics	string (topic name) physical document bloc(s) and meeting segment	IM2.AP
AS1. Speech segmentation	"silence" or "voice" audio stream segment	IM2.SP
AS2. Speech transcription	string audio stream segment	IM2.SP

SMM standard annotations	Values Domain	IP
AS3. Speaker	participant's name meeting segment	IM2.SP
AS4. Non-verbal acoustical clues (lexicon to be define)	clue type ("applause", "laugher", etc.) audio stream segment	IM2.SP
ADI1. Utterances (dialog act lexicon to define)	dialog act speech transcription extract	IM2.MDM
ADI2. Adjacency Pairs	none two utterances	IM2.MDM
ADI3. Turns	none list of utterances	IM2.MDM
ADI4. Thematic Episodes (partially linked with AG2)	string (theme definition) list of turns	IM2.MDM
ADI5. Named Entities (NE) + Coreference	entity definition speech transcription extract	IM2.MDM IM2.SP
ADO1. Document structure (layout + logical)	list of characteristics document blocs	IM2.AP
ADO2. Document alignment	document bloc(s) meeting segment	IM2.AP
AV1. Gesture (lexicon to be defined)	gesture type ("raised hand") video stream segment or image	IM2.SA, IM2.IIR
AV4. Facial expression (lexicon to be defined)	face type ("smile", "talk", "listen") video stream segment or extracted image	IM2.SA
AV3. Eye gaze (directions or "looking at" function)	eye gaze direction video stream segment or extracted image	IM2.SA, IM2.VE

Table 4: Summary of the annotations for the SMM application

Other annotations

The use of a whiteboard with a camera focused on it can lead to interesting annotations such as:

- regular snapshots of the whiteboard video and the result of handwriting analysis applied to its content;
- whiteboard user identification (the name of the participant in front of the whiteboard).

The use of a projection screen should also be completed with a current projected document annotation giving the name of the document visible. This annotation can be obtained directly from the projection software, but it could also be a good exercise to recognize the projected document only with image analysis of a video stream of the screen captured by an external camera.

It is also possible to extend the system for applying image analysis algorithms to some images extracted from the video streams at regular intervals. This could also bring useful information into the database without the cost of analysing full video streams. For instance, when there is an overview camera of the full meeting room, a snapshot taken during the meeting can be used to automatically extract the participant's position around the table.

5.2 From annotations to minutes

The annotations listed in the previous section represent a subset of all the possible annotations. However, we believe that this subset is reasonable in regard of the actual state-of-the-art in speech, video and document analysis. Together, these annotations represent multiple modalities (such as voice, gesture, a visual media for documents, etc.) that can be combined to generate higher level, cross-modalities annotations as needed.

The SMM application must demonstrate interesting and innovative usages for the annotations stored in a multimedia meeting database. Through the retrieval system presented in section 4.5, it is obvious that meeting participants and other authorized people will benefit from a query system based on the annotations. Some of the potential queries are listed in annex 4. In particular, a special kind of user, the scribe, will benefit from the annotations: s/he will use the retrieval system to extract significant meeting moments and to write corresponding minutes that will be added to the database. The database enriched with these "smart meeting minutes" will then offer further functionalities for generating hypermedia meeting minutes on demand and for different media (multimedia workstations, PDAs, mobile phones, etc.). Some tools that automatically extract, generate and display multimedia sequences and interactive views from the database will be useful to support the scribe's task, for instance:

- a **meeting movie**, i.e. a multimedia clip showing only the current speaker (and additional speakers in case of speech overlapping);
- a **votes movie**, i.e. a multimedia clip that for each vote in a meeting summarizes the vote topic, the vote options, the vote count for each option, and the final decision;
- a **meeting playwriter transcript**, i.e. a textual transcription of a meeting structured like a theatre play with acts corresponding to thematic episodes, scenes corresponding to sections, and replicas to utterances;
- a **visual abstraction** representing the **dialog structure**, i.e. an overall view of the full meeting dialogs with keyframes extracted from the meeting linked to their audio and video counterparts, for example it is possible to generate a video manga view of the meeting (i.e. "a comic-book style video summarization system" [1]);
- a **visual interaction graph,** i.e. a symbolic view of the meeting table with a time slider that shows at any moment who is looking at who; this graph could be linked with the meeting movie and to each participant's video stream for a quick access to interesting moments;
- a **document focus timeline**, i.e. a document window with a time slider that display for any point during the meeting the documents parts that are being discussed;
- a **linked agenda**, i.e. an hypermedia list of the agenda topics (including the new topics that showed up during the meeting) linked with: a) the corresponding meeting movie

extract, b) the corresponding extract in each participant's movie, c) related parts of the meeting documents.

Some of the previous sequences and interactive views will be developed by IM2.AP in collaboration with any IP interested in data visualization methods and multimedia generation algorithms. The definition of the smart meeting minutes schema will be done in collaboration between IM2.AP and IM2.DS. IM2.AP will also design and prototype some hypermedia meeting minutes documents for different devices.

For instance, a minimal browser version of hypermedia meeting minutes will present at least: the meeting date, place and title; a map of the meeting room with meeting participants and links to personal information (email address, home page, institute), a list of the thematic episodes with a a textual summary (including the goal of the dialog, participant's opinions, decisions taken, etc.). Each episode will be display with a key frame pointing to the corresponding extract of the meeting movie and of each participant's movies. Finally, the hypermedia minutes will also show a list of all the meeting documents linked to their PDF or HTML version and to their corresponding referencing thematic episodes in the meeting movie.

5.3 Summary of IP's contributions

The production of the annotations listed above concern most of IM2 IPs. Annotations will be integrated into the database with the data integration system introduced in section 5.3. The list below summarizes the previous sections. Feedbacks are very welcomed in order to make this document a useful reference.

- IM2.AP Smart Meeting Minutes

  Global meeting descriptors (see annex 3), agenda and current document annotations and, for each document, document composition annotation.

- IM2.ACP Multimedia Information Access and Content Protection

  Speaker identification through face recognition on images extracted from video streams (this could automatize the filling of the "participant *x* was sited at place *p*" descriptors) and through voice recognition on audio streams.

- IM2.DS Multimodal Information: Deployment, Storage and Access

  Constitution of the multimedia meeting database.

- IM2.IIR Information Indexing and Retrieval

  Definition of an image and/or video lexicon related to video annotations (in collaboration with IM2.SA and IM2.VE). This lexicon should include and go beyond the simple "raised hand" gesture annotation and the facial expression and eye gaze annotations. Definition of a query language for the retrieval system (based on previous experiments with MRML). Keyframe extraction to select interesting images from video streams to be used as poster images in multimedia clips.

- IM2.MDM Multimodal Dialogue Management

  Discourse annotations.

- IM2.MI Multimodal Input and Modality Integration

  To be defined.

- IM2.SA Scene Analysis

  See IM2.IIR. More particularly "raised hand" gesture, facial expression and eye gaze direction recognition, handwriting recognition of whiteboard snapshots (when available), identification of documents projected on the projection screen (when projection screen available), recognition of the presenter in front of the projection screen and/or whiteboard.

- IM2.SP Speech Processing

  Speech segmentation, speech transcription, current speaker and non verbal acoustical clues annotations.

- IM2.VE Virtual Entities

  Facial expressions analysis (definition of a lexicon, see IM2.IIR). Virtual reality based minutes generation (for instance a 3D view of the meeting table with avatars reproducing participant's eye gaze and raised hand).

5.4 Note on annotations and XML

There are several reasons for choosing XML as the data representation language for uploading and retrieving annotations to and from the multimedia meeting database. These reasons are:

- it is device independent;
- it allows data interchange between applications and archival without software dependencies;
- it lower the need for middleware;
- it is readable both by humans and machines.

The representation of annotations in XML supposes that all the IPs that produce or consume annotations agree together for defining the corresponding data models with a corresponding DTD or XML schema. These data models can be inspired from the existing practices, such as the ICSI Meeting Transcriptions [2], or emerging standards such as ATLAS [3]. It is also possible that some annotations may be better expressed with the RDF formalism. We welcome any input on that subject to eventually include it in the SMM application specifications.

6. Related works

A few research projects around the world share some common goals with the SMM application. These following projects are also about recording meetings, their web sites contain lot of useful information:

- NIST Automatic Meeting Transcription Project, www.nist.gov/speech/test_beds/mr_proj/
- NIST Smart Space Laboratory, www.nist.gov/smartspace/
- The Meeting Recorder Project at ICSI, www.icsi.berkeley.edu/Speech/mr/index.html

The eClass project is focused on the recording of teaching lessons in the classroom. It is not focused on automatic speech transcription but instead it has investigated the capture of handwriten students' notes and its synchronization with audio, video and teachers' whiteboard presentations:

- eClass media enabled classroom, www.cc.gatech.edu/fce/eclass/index.html

The following projects are more generally speaking some "augmented meeting room" projects or relate to "reactive environments". However, recording meetings is a major application for these rooms and environments:

- SmartSpaces, Xerox FxPAL media enabled meeting room, www.fxpal.com/smartspaces/
- AMBIENTE, augmented meeting room, www.darmstadt.gmd.de/ambiente/
- Jeremy Cooperstock's home page (reactive environments), www.cim.mcgill.ca/~jer/

- CMU Project Aura, www-2.cs.cmu.edu/~aura/

Finally, the W3C has launched a working group on multimodal interaction. This group will certainly play a role in the definition of future standards:

- W3C working group on Multimodal Interaction Activity: www.w3.org/2002/mmi

Conclusion

The SMM application described in this document offers a rich framework for integrating results and technologies from different domains. It is a complex application that requires a coordination effort between all the IM2 IPs. In the next months, some efforts should be particularly put on the definition of common data formats for the annotations and on the data integration process and middleware.

The benefits of the full project are multiple. Firstly, the database setup by IM2.DS and IM2.AP will provide all the members with a rich corpus of multimedia data annotated according to multiple human-to-human communication modalities. Secondly, this database will be available for experimenting advanced interaction techniques with rich multimedia content (including visualization and natural language dialog techniques). Thirdly, the analysis of the annotated meeting recordings will be a first-class resource for understanding multimodality in human-to-human communication. Finally, the knowledge and experience acquired previously will allow the development of a new generation of multimodal human-computer interfaces. It seems obvious then that interaction with the meeting database will also be a natural test bed for multimodal interfaces.

References

[1] J. Boreczky, A. Girgensohn, G. Golovchinsky, and S. Uchihashi. An Interactive Comic Book Presentation for Exploring Video. In CHI 2000 Conference Proceedings, ACM Press, pp. 185-192, 2000 [Abstract, PDF].

[2] C. Barras, E. Geoffrois, Z. Wu, and M. Liberman, "Transcriber: development and use of a tool for assisting speech corpora production." Speech Communication special issue on Speech Annotation and Corpus Tools, 2000.

[3] Muhr, Thomas (2000, December). Increasing the Reusability of Qualitative Data with XML [64 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], 1(3). Available at: http://qualitative-research.net/fqs/fqseng.htm

ANNEX 1: GLOSSARY

A

Audio annotations

Annotations of the audio recordings, such as "audio *x* from *start-time* to *end-time* contains some speech" .

D

Discourse annotations

Annotations describing the structure of the speech. The structure is a hierarchy of thematic episodes made of sections, made of turns, made of utterances. Utterances are associated with a dialog act (for instance inform, agree, reject, request, justify, etc.). Two utterances can be grouped into an adjacency pair (for instance invitation <-> accept/decline, question <-> answer, etc.).

Utterances can be seen as annotations of speech transcripts (which are annotations of speech). In that way an utterance is of the form "the speech transcript *t* from *start-pos* to *end-pos* is an utterance".

Document layout structure

A layout structure describes the physical document presentation in term of blocs with typographical attributes. It describes the document as it is printed or displayed on a screen. For instance "the bloc in the box *(x1,y1,x2,y2)* of the document *x* is in Helvetica 18 bold".

Document logical structure

A logical structure describes the abstract organization of the document content in term of structural elements such as chapters, sections, lists, captions, etc, and/or any other logical description of the content. For instance "the bloc *b* in the document *d* is a figure entitled *budget 2002*" or "the 3rd bloc of the document *d* is a chapter title".

F

Firewire (IEEE 1394)

A very fast external bus standard that supports data transfer rates of up to 400 Mbps (400 million bits per second). Products supporting the 1394 standard go under different names, depending on the company. Apple, which originally developed the technology, uses the trademarked name *FireWire*. Other companies use other names, such as *i.link* and *Lynx,* to describe their 1394 products.

A single 1394 port can be used to connect up 63 external devices. In addition to its high speed, 1394 also supports *isochronous data* -- delivering data at a guaranteed rate. This makes it ideal for devices that need to transfer high levels of data in real-time, such as video devices.

Like USB, 1394 supports both Plug-and-Play and hot plugging, and also provides power to peripheral devices.

G

Global meeting descriptors

Description of the meeting settings (date, place, participants, seats, documents, cameras, microphones, etc.) and of the meeting agenda.

H

Hypermedia meeting minutes

Hypermedia document generated from the meeting database based on the agenda (the list of topics discussed), augmented with multimedia clips and textual summaries (decisions, comments, participant's list, etc.) and links to meeting documents.

M

Meeting recordings

The records collected during a meeting include static documents (distributed to the participants or projected during the meeting), several sound tracks as well as several video tracks. Those records come with structural and contextual descriptors.

Meeting movie

The movie of the meeting. It is a single video track showing only the current speaker (and additional speakers in case of speech overlapping). It has the same duration as the full meeting.

Minutes

“Minutes” is the term designating an official record of the proceedings of a meeting. It is a summary covering points to be remembered: a memorandum.

Multimedia clip

In editing, a video clip refers to a short section removed from a picture shot. Often called a cut or trim. We define a multimedia clip being a composition of different media clips (video, graphics, animation, etc).

Multimedia database

A database that gives access to different raw media data (video, audio, text, document) through different indexes (or annotations).

Multimedia

The term multimedia corresponds to the use of computers to present text, graphics, video, animation, and sound in an integrated way. Long touted as the future revolution in computing, multimedia applications were, until the mid-90s, uncommon due to the expensive hardware required. With increases in performance and decreases in price, however, multimedia is now commonplace. Nearly all PCs are capable of displaying video, though the resolution available depends on the power of the computer's video adapter and CPU.

Multimodal

In the general sense, a multimodal system supports communication with the user through different modalities such as voice, gesture, and typing. Literally, “multi” refers to “more than one” and the term “modal” may cover the notion of “modality” as well as that of “mode”.

- Modality refers to the type of *communication channel* used to convey or acquire information. It also covers the way an idea is expressed or perceived, or the manner an action is performed.
- Mode refers to a state that determines the way information is interpreted to extract or convey meaning.

In a communication act, whether it is between humans or between a computer system and a user, both the modality and the mode come into play. The modality defines the type of data exchanged whereas the mode determines the context in which the data is interpreted. Thus, if we take a system-centered view, multimodality is the capacity of the system to communicate with a user along different types of communication channels and to extract and convey meaning automatically.

We observe that both *multimedia* and *multimodal* systems use multiple communication channels. But in addition, a multimodal system is able to automatically model the content of the information at a high level of abstraction. A multimodal system strives for meaning.

S

SMIL

Short for ***S**ynchronized **M**ultimedia **I**ntegration **L**anguage,* a markup language being developed by the World Wide Web Consortium (W3C) that enables Web developers to divide multimedia content into separate files and streams (audio, video, text, and images), send them to a user's computer individually, and then have them displayed together as if they were a single multimedia stream. The ability to separate out the static text and images should make the multimedia content much smaller so that it doesn't take as long to travel over the Internet.

SMIL is based on the eXtensible Markup Language (XML). Rather than defining the actual formats used to represent multimedia data, it defines the commands that specify whether the various multimedia components should be played together or in sequence.

U

USB

Short for ***U**niversal **S**erial **B**us,* an external bus standard that supports data transfer rates of 12 Mbps. A single USB port can be used to connect up to 127 peripheral devices, such as mice, modems, and keyboards. USB also supports *Plug-and-Play* installation and *hot plugging*.

Starting in 1996, a few computer manufacturers started including USB support in their new machines. It wasn't until the release of the best-selling iMac in 1998 that USB became widespread. It is expected to completely replace serial and parallel ports.

V

Video annotations

Annotations of video recordings, such as "participant *x* has a raised hand on video *y* from *start-time* to *end-time*" or "the image *i* from video *y* shows a document *z* at position *p*".

X

XML

Short for ***Ex**tensible **M**arkup **L**anguage,* a specification developed by the W3C. XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

ANNEX 2: GLOBAL MEETING DESCRIPTORS DTD

```
<?xml version="1.0" encoding="UTF-8"?>

<!--	**************************************
	DTD of the global meeting descriptors
	************************************** -->

<!ELEMENT MEETING (GlobalInfos, Room, Participants, Agenda, Document+)>

<!--	********************
	Global Informations:
	******************** -->

<!ELEMENT GlobalInfos (Title, Date, Time, Location, Description)>
<!ATTLIST GlobalInfos
	startTime CDATA #REQUIRED
	endTime CDATA #REQUIRED
>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT Date (#PCDATA)>
<!ELEMENT Time (#PCDATA)>
<!ELEMENT Location (#PCDATA)>
<!ELEMENT Description (#PCDATA)>

<!--	**************************************************************
	Room layout:
	The meeting room is compound of several seats. Each seat should
	have at least one closeup camera and one microphone in front.
	Overview cameras capture a set of seats or the overall meeting room.
	************************************************************** -->

<!ELEMENT Room (Seat+, CloseupCamera+, Microphone+, OverviewCamera*)>
<!ELEMENT Seat EMPTY>
<!ATTLIST Seat
	IDSeat ID #REQUIRED
	Position CDATA #REQUIRED
>
<!ELEMENT CloseupCamera EMPTY>
<!ATTLIST CloseupCamera
	VideoFilePrefix CDATA #REQUIRED
	Type (usb | firewire | CDATA) #REQUIRED
	SeatRef IDREF #REQUIRED
>
<!ELEMENT Microphone EMPTY>
<!ATTLIST Microphone
	AudioFilePrefix CDATA #REQUIRED
	SeatRef IDREF #REQUIRED
>
<!ELEMENT OverviewCamera EMPTY>
<!ATTLIST OverviewCamera
	VideoFilePrefix CDATA #REQUIRED
	Position CDATA #REQUIRED
	Type (usb | firewire | CDATA) #REQUIRED
	ListSeatRef IDREFS #REQUIRED
>

<!--	*************
	Participants
	************* -->

<!ELEMENT Participants (Participant*)>
<!ELEMENT Participant (Expertise?, Bio?)>
<!ATTLIST Participant
	Name CDATA #REQUIRED
	Role (president | expert | consultant) #REQUIRED
	Email CDATA #IMPLIED
	PhotoFileName CDATA #IMPLIED
	SeatRef IDREF #REQUIRED
>
<!ELEMENT Expertise (#PCDATA)>
<!ELEMENT Bio (#PCDATA)>

<!--	**********************************************************
```

```
       The Agenda:
       It is composed of several topics and  refers to an agenda
       document. Each topic can refer to multiple documents.
       ********************************************************** -->

<!ELEMENT Agenda (Topic*)>
<!ATTLIST Agenda
       DocRef IDREF #REQUIRED
>
<!ELEMENT Topic (Description)>
<!ATTLIST Topic
       IDTopic ID #REQUIRED
       startTime CDATA #REQUIRED
       endTime CDATA #REQUIRED
       DocsRef IDREFS #IMPLIED
>

<!--   *****************************************
       Documents:
       Each document refers to a list of topics.
       ***************************************** -->

<!ELEMENT Document (Description?)>
<!ATTLIST Document
       IDDoc ID #REQUIRED
       Authors CDATA #REQUIRED
       Title CDATA #REQUIRED
       Version CDATA #IMPLIED
       PDFFilename CDATA #REQUIRED
       SourceFormat (pdf | doc | ps | ppt | CDATA) #REQUIRED
       TopicsRef IDREFS #IMPLIED
>
```

ANNEX 3: SAMPLE MEETING RECORDING

The 23rd of July, a 20 minutes long meeting about the specifications of the SMM application has been recorded at the EPFL. Each of the 5 participants has been recorded with a camera and a USB microphone. The audio files have been partially transcribed. This annex shows some samples of the corresponding annotations and of the meeting documents. First, it presents the global meeting descriptors in XML format, as specified by DTD in annex 2. Then, it shows the meeting invitation sent to every participants. The meeting agenda is visible in the second part of the invitation. Ideally, the invitation should be processed automatically by the SMM application to extract the agenda topics and then to link them with the thematic episodes of the discourse annotations. We also reproduce a meeting database schema that was discussed during the meeting. Finally, the annex presents an extract of the speech transcription of the meeting.


```xml
<?xml version="1.0" encoding="UTF-8"?>
<MEETING>
    <GlobalInfos startTime="0" endTime="1232.1194999999">
        <Title>SMM meeting IM2.AP,DS</Title>
        <Date>23rd July 2002</Date>
        <Time>10:02:53</Time>
        <Location>EPFL- LSP Laboratory</Location>
        <Description> The goal of this meeting was first to test the capture environment and the synchronization
                between the audio and video streams, and second to discuss about the audio/video capture and
                streaming, and of the database schema.
        </Description>
    </GlobalInfos>
    <Room>
        <Seat IDSeat="s1" Position="A"/>
        <Seat IDSeat="s2" Position="B"/>
        <Seat IDSeat="s3" Position="C"/>
        <Seat IDSeat="s4" Position="D"/>
        <Seat IDSeat="s5" Position="E"/>
        <CloseupCamera VideoFilePrefix="2_a" Type="firewire" SeatRef="s1"/>
        <CloseupCamera VideoFilePrefix="2_b" Type="firewire" SeatRef="s2"/>
        <CloseupCamera VideoFilePrefix="2_c" Type="usb" SeatRef="s3"/>
        <CloseupCamera VideoFilePrefix="2_d" Type="usb" SeatRef="s4"/>
        <CloseupCamera VideoFilePrefix="2_e" Type="usb" SeatRef="s5"/>
        <Microphone AudioFilePrefix="2_1" SeatRef="s1"/>
        <Microphone AudioFilePrefix="2_2" SeatRef="s2"/>
        <Microphone AudioFilePrefix="2_3" SeatRef="s3"/>
        <Microphone AudioFilePrefix="2_4" SeatRef="s4"/>
        <Microphone AudioFilePrefix="2_5" SeatRef="s5"/>
    </Room>
    <Participants>
        <Participant Name="Rolf Ingold" Role="president" Email="rolf.ingold@unifr.ch" SeatRef="s1"/>
        <Participant Name="Stéphane Sire" Role="expert" Email="stephane.sire@epfl.ch" SeatRef="s2"/>
        <Participant Name="Sebastian Gerlach " Role="expert" Email="sebastian.gerlach@epfl.ch" SeatRef="s3"/>
        <Participant Name="Pier Donini" Role="expert" Email="pier.donini@epfl.ch" SeatRef="s4"/>
        <Participant Name="Denis Lalanne" Role="consultant" Email="denis.lalanne@unifr.ch" SeatRef="s5"/>
    </Participants>
    <Agenda DocRef="d1">
        <Topic IDTopic="t1" startTime="0" endTime="34.682">
            <Description>Rolf Ingold introduces the meeting and presents the points to be discussed.</Description>
        </Topic>
        <Topic IDTopic="t2" startTime="34.682" endTime="172.107">
            <Description>Issues regarding audio-video capture</Description>
        </Topic>
        <Topic IDTopic="t3" startTime="172.107" endTime="729.065">
            <Description>Playback of multiple streams</Description>
        </Topic>
        <Topic IDTopic="t4" startTime="729.065" endTime="1212.032" DocsRef="d2">
            <Description>Database schema</Description>
        </Topic>
        <Topic IDTopic="t5" startTime="1212.032" endTime="1232.1194999999">
            <Description>Conclusion and next meeting</Description>
```

```xml
        </Topic>
    </Agenda>

    <Document IDDoc="d1" Authors="Rolf Ingold" Title="Réunion de coordination du projet smart meeting minutes"
PDFFilename="agenda23smm" SourceFormat="pdf"/>
    <Document IDDoc="d2" Authors="Pier Donini" Title="SMM DataBase schema"
PDFFilename="smmDBSchema" Version="0.4" SourceFormat="pdf" TopicsRef="t4"/>
</MEETING>
```

INVITATION

RÉUNION DE COORDINATION DU PROJET SMART MEETING MINUTES

Comme prévu dans notre agenda, je vous rappelle que nous avons retenu

le mardi 23 juillet, à 9h30 à l'EPFL

pour enregistrer un "smart meeting". Cet enregistrement de 15-20 minutes doit servir ensuite à commencer le prototypage de l'application (essentiellement la BD, quelques annotations et les premières interfaces d'affichage).

Nous avons déjà effectué un premier test le 5 juillet qui a donné 15 GB de données audio et video non compressées. Ce deuxième test devrait nous permettre d'améliorer la qualité (2 micros du précédent test étant plutôt de mauvaise qualité), et à améliorer notre performance :)

Cette fois-ci, Rolf Ingold remplacera Roger Hersch comme moderateur.

L'objet de la discussion est de discuter des choix en cours pour la capture audio et video, pour la diffusion audio et video, et pour le schema de la base de données.

Si vous avez envie d'emmener un document pour l'utiliser pendant la réuniom, n'hésitez pas, merci de l'imprimer avant en 6 exemplaires pour distribuer à chacun. Nous devrions avoir au moins 2 documents: l'agenda (cf ci-dessous) et le schéma de la BD.

J'ai ajouté Abdel à la liste des participants, au cas où il souhaiterait se joindre à nous, il peut dans ce cas servir de candide comme Denis. Au total nous devrions donc être 5 ou 6 personnes suivant la participation d'Abdel.

A mardi donc,

Stéphane

Participants

Rolf Ingold (modérateur)

Stéphane Sire

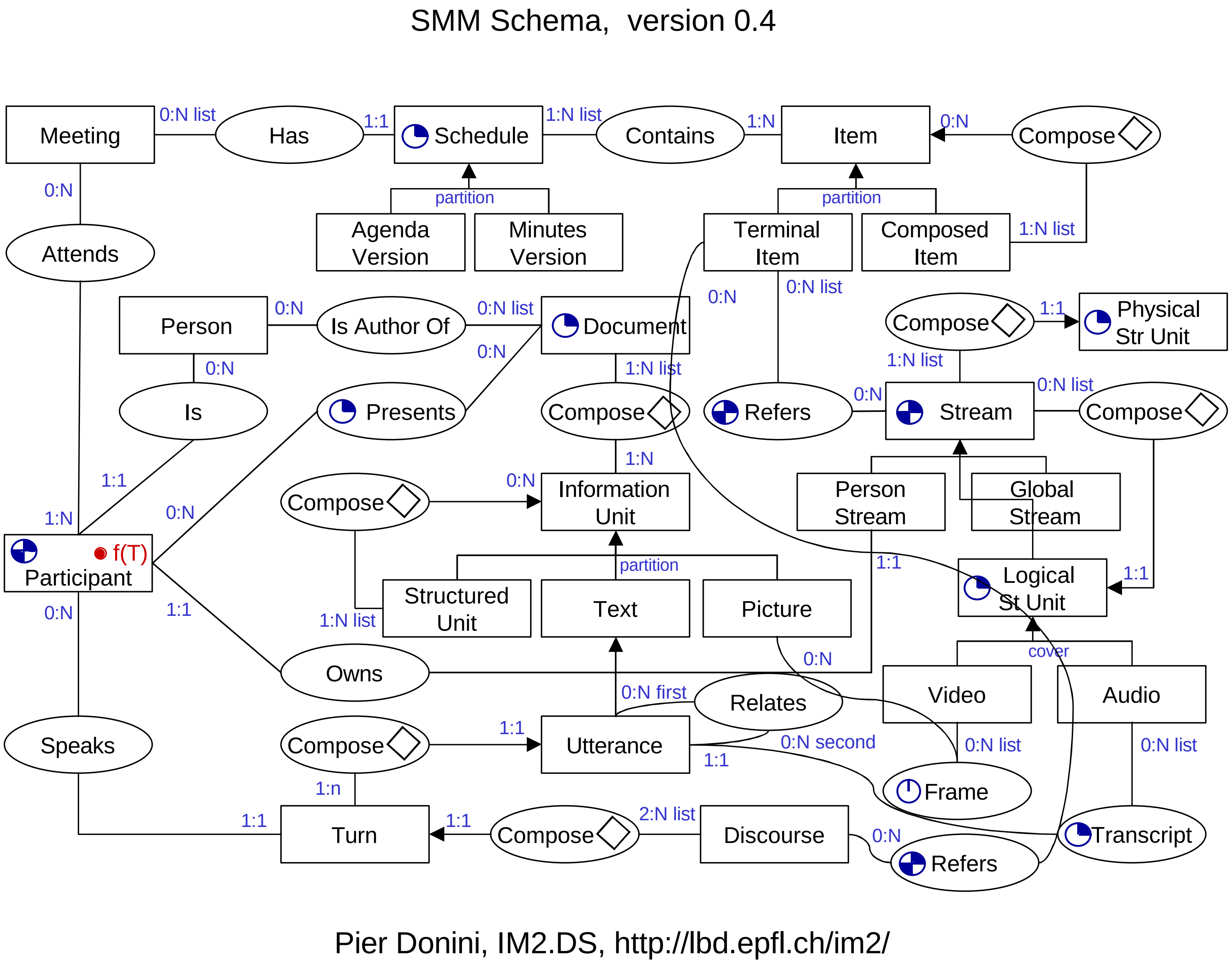
Denis Lalanne

Sébastien Gerlach

Pier Donini

Ordre du jour

1. Welcome / Introduction (Moderator)
2. Issues regarding audio-video capture (intro by Stéphane)
3. Playback of multiple streams (intro by Sebastian)
   - Direct Show based solution / DPlayer (Sebastian)
   - Quicktime based solution / SMIL player (Stéphane)
4. Database schema (short presentation by Pier)
5. Conclusion / next meeting



The document above is an example of a document brought to a meeting by one of the participants. It has been printed and distributed to all the participants before the meeting.

The following document extract is a textual transcription of the 23rd of July meeting (using Transcriber). This transcription is not complete (it would be too longin this report). However, it presents some significative parts of the discourse structure and an interesting transcription of the dialog about the 3rd agenda topic: “playback of multiple streams”.


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Trans SYSTEM "trans-13.dtd">
<Trans scribe="Denis Lalanne" audio_filename="Sound2_2" version="6" version_date="020808">

    <Topics>
        <Topic id="to1" desc="Introduction"/>
        <Topic id="to2" desc="Issues regarding audio-video capture"/>
        <Topic id="to3" desc="Playback of multiple streams"/>
        <Topic id="to4" desc="Database schema"/>
        <Topic id="to5" desc="Conclusion"/>
    </Topics>

    <Speakers>
        <Speaker id="spk1" name="Rolf Ingold" check="yes" type="male" dialect="native" scope="local"/>
        <Speaker id="spk2" name="Stéphane Sire" check="yes" type="male" dialect="native" scope="local"/>
        <Speaker id="spk3" name="Denis Lalanne" check="yes" type="male" dialect="native" scope="local"/>
        <Speaker id="spk4" name="Sébastian Gerlach" check="yes" type="male" dialect="native" scope="local"/>
        <Speaker id="spk5" name="Pier Donini" check="yes" type="male" dialect="native" accent=""
scope="local"/>
    </Speakers>

    <Episode>
        <!--Introduction -->
        <Section type="report" topic="to1" startTime="0" endTime="34.682">
            <Turn startTime="0" endTime="34.682" speaker="spk1" fidelity="high">
                <Sync time="0"/>
                  Voilà, je vous souhaite la bienvenue à cette nouvelle séance de coordination dans le cadre du
                  projet Smart Meeting Minutes. Et puis, comme vous le savez donc le but c'est de faire un peu le
                  point sur l'état d'avancement des différents points qui concernent la mise en place de notre
                  infrastructure.
                <Sync time="22.547"/>
                  Euh, alors pour commencer on avait prévu d'étudier le cas de l'environnement de capture audio-
                  video et je vais peut-être donner la parole à Stéphane pour qu'il nous dise où c'est qu'il en est.
            </Turn>
        </Section>

        <-- Issues regarding audio-video capture -->
        <Section type="report" topic="to2" startTime="34.682" endTime="172.107">
            <Turn speaker="spk2" mode="spontaneous" fidelity="high" startTime="34.682" endTime="111.845">
                <Sync time="34.682"/>
                  D'accord. Alors je vais vous décrire les acquisitions récentes que nous avons fait et la
                  technologie sous-jacente et puis éventuellement si vous avez des questions on pourra entamer
                  une petite discussion.
                   …
                   …
            </Turn>
            …
            <Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="169.969" endTime="172.107">
                <Sync time="169.969"/>
                   Parfait merci.
            </Turn>
        </Section>

        <-- Playback of multiple streams -->
        <Section type="report" topic="to3" startTime="172.107" endTime="729.065">
            <Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="172.107" endTime="188.169">
                <Sync time="172.107"/>
            Si il n'y a pas d'autres commentaires je propose de passer au point suivant qui concerne cette fois si la
            problématique du playback. Euh, là j'aimerais passer la parole à Sébastian pour qu'il fasse le point
            notamment avec une solution possible basée sur DirectShow, si je suis bien informé.
            </Turn>
            <Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="188.169" endTime="240.185">
                <Sync time="188.169"/>
```

```xml
        Tout à fait. Alors en effet quand on a un enregistrement comme celui qu'on fait ici avec 5 flux
        vidéos et 5 flux audio, les players conventionnels sont en général faits pour gérer des
        situations.......
</Turn>
<Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="240.185" endTime="241.378">
    <Sync time="240.185"/>
        Oui Stéphane.
</Turn>
<Turn speaker="spk2" mode="spontaneous" fidelity="high" startTime="241.378" endTime="256.434">
    <Sync time="241.378"/>
        Alors j'ai une question par rapport à cette solution c'est est-ce que l'on pourrait l'uti.......
</Turn>
<Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="256.434" endTime="273.465">
    <Sync time="256.434"/>
        Alors ce genre de solution peut passer sur le réseau sous forme d'un contrôle activeX. Dans la
        mesure ........
</Turn>
<Turn speaker="spk2" mode="spontaneous" fidelity="high" startTime="273.465" endTime="275.291">
    <Sync time="273.465"/>
        Dans dans un browser j'entends...
</Turn>
<Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="275.291" endTime="277.825">
    <Sync time="275.291"/>
        Dans un browser tout à fait...
</Turn>
<Turn speaker="spk2" mode="spontaneous" fidelity="high" startTime="277.825" endTime="280.359">
    <Sync time="277.825"/>
        Mais faudra installer le contrôle activeX donc cela...
</Turn>
<Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="280.359" endTime="284.492">
    <Sync time="280.359"/>
        Il faudra installer le contrôle activeX mais ceci est un processus entièrement automatique.
</Turn>
<Turn speaker="spk2" mode="spontaneous" fidelity="high" startTime="284.492" endTime="285.138">
    <Sync time="284.492"/>
        D'accord.
</Turn>
<Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="285.138" endTime="287.685">
    <Sync time="285.138"/>
        D'accord, je crois qu'il y a Pier qui veut aussi intervenir.
</Turn>
<Turn speaker="spk5" mode="spontaneous" fidelity="high" startTime="287.685" endTime="295.66">
    <Sync time="287.685"/>
        On peut imaginer de sortir un plug-in pour les browsers qui soient pas activeX compatibles.... ou
        c'est....
</Turn>
<Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="295.66" endTime="298.567">
    <Sync time="295.66"/>
        Pour autant que l'on reste sur un plate-forme microsoft. Oui.
</Turn>
<Turn speaker="spk5" mode="spontaneous" fidelity="high" startTime="298.567" endTime="299.163">
    <Sync time="298.567"/>
        Ouais.
</Turn>
<Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="299.163" endTime="302.48">
    <Sync time="299.163"/>
        Parce que la technologie fondamentale directShow n'est pas cross-plateforme.
</Turn>
<Turn speaker="spk5" mode="spontaneous" fidelity="high" startTime="302.48" endTime="304.306">
    <Sync time="302.48"/>
        Oui.
    <Sync time="303.784"/>
        D'accord.
</Turn>
<Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="304.306" endTime="306.84">
    <Sync time="304.306"/>
        Alors je..
    <Sync time="305.126"/>
        Ah il y a peut-être encore Denis.
</Turn>
<Turn speaker="spk3" mode="spontaneous" fidelity="high" startTime="306.84" endTime="312.906">
```

```
      <Sync time="306.84"/>
      Oui je voulais savoir si le format vidéo pour le broadcasting était différent du format de capture en
      fait?
    </Turn>
    <Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="312.906" endTime="333.105">
      <Sync time="312.906"/>
        Ceci est surtout une question de la bande passante réseau que l'on a à disposition. La capture
        ici, si on compte tous les flux ensembles, on a à peu près 30 mégabytes par seconde. La
        diffusion d'une telle quantité de données sur un réseau est peu réaliste donc effectivement on est
        obligé de transcoder les vidéos, de les compresser, pour pouvoir les diffuser sur le réseau par la
        suite.
    </Turn>
    <Turn speaker="spk3" mode="spontaneous" fidelity="high" startTime="333.105" endTime="339.03">
      <Sync time="333.105"/>
        On a déjà une idée du format vidéo, à peu près... pour le broadcasting ou?
    </Turn>
    <Turn speaker="spk4" mode="spontaneous" fidelity="high" startTime="339.03" endTime="358.724">
      <Sync time="339.03"/>
        Bon, si on utilise un player basé sur DirectShow, on a très peu de limitation sur le format utilisé
        donc des critères techniques comme par exemple la facilité de faire des cycles à l'intérieur de la
        vidéo, donc de sauter d'un endroit à l'autre peuvent entrer en compte pour le choix du format qui
        sera retenu. Evidemment la bande passante de la compression étant importante aussi.
    </Turn>
    <Turn speaker="spk3" mode="spontaneous" fidelity="high" startTime="358.724" endTime="360.247">
      <Sync time="358.724"/>
        Ok.
    </Turn>
    <Turn speaker="spk1" mode="spontaneous" fidelity="high" startTime="360.247" endTime="362.632">
      <Sync time="360.247"/>
        D'accord. Oui Stéphane? tu as encore une question?...
    </Turn>
    <Turn speaker="spk2" startTime="362.632" endTime="436.0">
      <Sync time="362.632"/>
        Là j'ai quand même une remarque qui me semble importante..... smil....
    </Turn>
    <Turn speaker="spk1" startTime="436.0" endTime="450.199">
      <Sync time="436.0"/>
        Oui. En tous cas, moi je souhaitais disons dire que je partage le soucis de Stéphane d'essayer
        d'avoir un environnement qui soit le plus ouvert possible, le plus multi-plateforme possible. Cela
        me semble essentiel dans le cadre de notre projet.
      <Sync time="449.677"/>
        Oui Pier?
    </Turn>
    …
  </Section>

  <-- Database schema -->
  <Section type="report" topic="to4" startTime="729.065" endTime="1212.032">
    <Turn speaker="spk1" startTime="729.065" endTime="741.137">
      <Sync time="729.065"/>
        Bien alors je crois que l'on va passer au dernier point de l'ordre du jour d'aujourd'hui à savoir le
        schéma de base de données et là j'aimerais savoir Pier s'il y a du nouveau. Tu ne nous a
        toujours pas remis un nouveau document.
    </Turn>
    …
    …
  </Section>

  <-- Conclusion -->
  <Section type="report" topic="to5" startTime="1212.032" endTime="1232.1194999999">
    <Turn speaker="spk1" startTime="1212.032" endTime="1226.157">
      <Sync time="1212.032"/>
      Bien merci, alors je crois qu'on arrive au terme de notre réunion. Je vous donne rendez-vous pour la
      prochaine fois dans trois semaines environ. On fixera encore le détail par email je pense. Et puis eh
      bien je vous remercie encore d'avoir participé.
    </Turn>
    <Turn speaker="spk3" startTime="1226.157" endTime="1227.647">
      <Sync time="1226.157"/>
      De rien.
    </Turn>
    <Turn speaker="spk4" startTime="1227.647" endTime="1228.057">
```

```
            <Sync time="1227.647"/>
            Merci.
        </Turn>
        <Turn speaker="spk2" startTime="1228.057" endTime="1228.355">
            <Sync time="1228.057"/>
                A la prochaine.
        </Turn>
        <Turn speaker="spk3" startTime="1228.355" endTime="1232.1194999999">
            <Sync time="1228.355"/>
                Merci.
        </Turn>
    </Section>

  </Episode>
</Trans>
```

ANNEX 4: SAMPLE QUERIES

Following are sample queries that could be submitted to the SMM application. They are grouped by categories.

1. History

What were the previous meetings of “IM2 IP heads”?

When is the next meeting of “IM2 IP heads”?

In which other meetings “A”, “B” and “C” did participate?

Which meetings took place in “meetingRoom A”?

Which meetings treated “Topic A”?

2. Situation

Where was taking place the meeting?

What is the title/label (IM2 IP heads, IM2 kick off, IM2.AP, etc) of the meeting?

At what time (date) did the meeting start (take place)?

How long was expected the meeting to last? and how long did it finally last?

3. Participants

Who were the participants?

Who was the president/moderator of the meeting?

What was the expertise of each participant?

I want to see the bibliography and profile of participant A.

What was the participation terms and conditions (knowledge required, documents to read, etc)?

Where were they placed around the table?

4. Turn-taking

I want to see the turn-taking flow of the overall meeting

(M spoke from frame 0 to frame 452; AV1 spoke from frame 453 to 535; etc).

I want to see the meeting transcription.

5. Documents

What were the documents distributed before/during/after the meeting?

What document was to read/see before the meeting?

Which documents relate to which specific participant?

6. Agenda

What was the main theme of the meeting?

What was the agenda of the meeting?

I want to see the president introduction to the meeting

I want to see the structure of the meeting, which topics were treated.

What was the conclusion for the overall meeting?

7. Topics

What were the different topics treated during the meeting?

What did "participant A" said about topic 1?

What were the questions opened about topic 1?

What was the conclusion for topic 1?

What solutions have been chosen concerning topic 1?

Who have accepted solutions X1 concerning topic 1?

Which decisions has been taken concerning topic 1?

8. Task type

What was the task type corresponding to topic 1?

See below for more details on task type.

…

For each of the following work tasks (a subset of work tasks from group research by J.E, McGrath, 1984, "Groups: Interaction and Performance"), we give some possible queries:

1) Planning tasks

   What were the planning tasks and the schedule/events decided?

2) Brainstorming and group creativity

   About which subject members brainstormed?

3) Negotiation

   Were there some negotiations?

   On what issues group members disagreed on? I want to see the corresponding sequence illustrating it…

   Which members disagreed and on what subject?

   What were the tradeoffs being made and what were the criteria (dimensions) used to make up a decision?

4) Decision making tasks

   What were the decisions being made?

   Which criteria were chosen to take the decision?

5) Competitive performance

   What were the competitive issues?

   Which members were competing against each other and on which corresponding issues?

6) Dissemination of information

What information has been disseminated and by who?

“These tasks should be thought of as a continuum, not discrete tasks. So, a complete meeting can contain several of these types of tasks”.

(http://www.nist.gov/speech/test_beds/mr_proj/data_collection/scenarios.html).