Multimedia Radiological Reports: Creation and Playback

Cheryl Schramm and Morris Goldberg

Multimedia radiological reports tie together radiographic images and operations on these images with textual, vocal, or graphic annotation. The result is an animated presentation of the radiological examination report. A temporal representation for a multimedia radiological report that describes the temporal behavior of the animated report is introduced. The temporal behavior specifies the synchronization requirements for the coordinated presentation of the multimedia report components. The temporal behavior is encoded in a temporal semantic network that is a network of temporally ordered nodes, where each node represents some primitive behavior. Automatic creation of the multimedia radiological report consists of the capture, interpretation, and storage of the temporal semantics of the radiologist's reporting session in terms of a set of primitive behaviors. Automatic playback of the multimedia radiological report is the orderly traversal of the temporal semantic network, invoking the behavior associated with each of the nodes in the temporal order described by the interconnecting arcs.
© *1989 by W.B. Saunders Company.*

ONE OF THE MAIN duties of a radiologist is to read radiographs and to report the findings. This report is normally in the form of text but invariably must indirectly reference anatomical features on the radiographs. Direct references are made in face-to-face consultations with physicians or by drawing on tracing paper placed over the radiograph. With the advent of computer-based work stations and integrated voice and data communication systems, it is now feasible to link the radiograph directly with text or voice connections. The result, which we call a multimedia radiological report (MRR), is really a dynamic animated entity that must tie together the set of radiographs, operations performed on

these radiographs (displays, enhancements, etc), and textual, verbal and graphic annotation.

The conceptual model that is used to represent the MRR must be able to accommodate a variety of features and performance requirements, including the following: image processing and manipulation operations on the radiograph; vocal, graphic and textual annotations; automatic creation and playback of the MRR with minimal user input; and variations in work station and user performance.

In this article, we describe a representation for MRRs based on temporal logic and semantic networks. The representation describes the temporal behavior of an MRR. The direct ties between the radiographs, operations, and annotations are given in terms of their synchronization requirements for the MRR playback. Our article is organized as follows. A brief introduction to interval-based temporal logic is given first. The next section describes the MRR representation developed. The final section demonstrates the use of the representation on a set of sample scenarios.

TEMPORAL LOGIC

Interval-based temporal logic[1] expresses knowledge about time in terms of intervals. An interval is a continuous, finite length of time. A temporal relation specifies the temporal ordering between two intervals. The set of 13 primitive temporal relations listed in Table 1 completely describes any possible arrangement of two intervals. Events or activities are linked to intervals so that the intervals represent the occurrences of the activities, and the temporal relations represent the temporal interaction between activities. For example, the statement "Dick called you while you were at work" contains two activities—call and work—whose temporal interaction is described as "call during work" (Fig 1).

Defined relations are disjunctions of two or more primitive temporal relations. They express uncertainty in the exact temporal ordering. For example, suppose that precedes is defined as "before or meets." As shown in Fig 2, "A precedes B" means that Interval A occurs immediately or at some (unspecified) time before Interval B. Reasoning about time uses the

Table 1. Primitive Temporal Relations in Interval-Based Temporal Logic

Relation (I,J)	Definition	Illustration	Inverse (J,I)
Before	I+ < J−	III JJJJ	After
Meets	I+ = J−	IIIII JJJJJ	Met by
During	I− > J− and I− < J+ and I+ < J+	IIII JJJJJJJ	Contains
Overlaps	I− < J− and I+ > J− and I+ < J+	IIIII JJJJJ	Overlapped
Starts	I− = J− and I+ < J+	IIIII JJJJJJJ	Started by
Finishes	I− > J− and I+ = J+	IIII JJJJJJ	Finished by
Equals	I− = J− and I+ − J+	IIIII JJJ	Equals

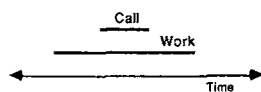Fig 2.   The defined relation precedes.

transitivity property of temporal relations. If two intervals I and J are related to a common interval K, then I and J are related by the transitive relation t where t is a disjunction of one or more primitive temporal relations. Allen[2] produced a transitivity table that contains the transitive compositions between all pairs of interval relations. Interval-based temporal logic is a domain-independent temporal representation of temporal knowledge that can model concurrent activities of uncertain duration and ordering.
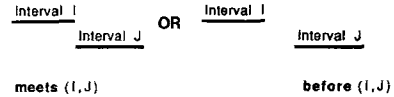
A TEMPORAL SEMANTIC NETWORK

We now introduce a temporal representation for MRRs. A temporal semantic network (TSN) consists of a set of nodes, interconnected by directed two-pole arcs. A node symbolizes an activity. A labeled arc joining a pair of nodes describes the temporal interaction between the two activities. Semantics specialize the general concept of a TSN to a domain, in this case to multimedia radiological reports. The node and arc semantics define what the nodes and arcs mean and how they are to be used, respectively. Network semantics describe how the whole is formed from the parts.

Node Semantics

The temporal behavior of an MRR can be decomposed into a set of primitive behaviors

Fig 1.   Timeline for "Call during Work."

common to all MRRs. These primitive behaviors define the set of nodes that will make up a MRR TSN. The nodes are described in Table 2. Two types of primitive behaviors that can be identified are persistent and ephemeral behaviors. Ephemeral behavior is an action that changes the world state. Although of varying duration, such actions will terminate. Persistent behavior is the effect of an action, and remains indefinitely until another action destroys it. Primitive behaviors can also be divided into three categories: visual, audio and external.

The visual nodes describe the temporal behavior of the image screens. Visual actions update the display on an image screen. The set of visual actions is roughly equivalent to the set of commands available at a radiology workstation. The image processing actions (ZOOM, CONTRAST, TOGGLE, MARKER, LENGTH, and ANGLE) are additive; they add some feature to an already existing display. The LOAD and RESTORE actions are terminating actions; the existing display is destroyed and completely replaced. The importance of this difference lies in the temporal relations that will be used to relate visual action nodes. There is a single visual effect: DISPLAY. This node describes the actual display on an image screen. Because the effect of all visual actions is some display, a DISPLAY node is paired with each visual action node in the MRR TSN.

The audio nodes describe the temporal behavior of the voice system. They are self-explanatory except for the notion of a voice segment used by the PLAY node. A voice examination report is made up of a series of physical voice segments. Physical voice segments are actual storage partitions caused by pauses during the report dictation. Physical voice segments may be subdivided into logical voice segments. Logical voice segments represent points at which the playback process should pause a physical voice segment and artificially synchronize the audio with some other activity. To the playback process, physical and logical voice segments function identically. Physical voice segments are identified automati-

Table 2. Nodes for an MRR Temporal Semantic Network

Node ID	Group	Type	Attributes	Description
Load	Visual	Action	Image file	Clears the image screen and loads the image file into image memory for display
Restore	Visual	Action	Image file	Removes all image operations on display, leaving only the original image
Zoom	Visual	Action	Zoom factor window position	Magnifies a portion of the current image display
Contrast	Visual	Action	Contrast algorithm, window position	Performs a contrast enhancement on a portion of the current image display
Toggle	Visual	Action	None	Toggles the contrast of the image display (from positive to negative, and back)
Length	Visual	Action	Length, line coordinates	Draws a line and labels it with the length
Angle	Visual	Action	Angle, angle coordinates	Draws an angle and labels it with the angle
Marker	Visual	Action	Point coordinates	Draws a marker at the given position
Display	Visual	Effect	Action node ID	Is the display of the associated action
Connect	Audio	Action	Voice file	Establishes a voice session and opens the voice file
Connect	Audio	Effect	Connect node ID	Is the voice session established by the given connect action
Disconnect	Audio	Action	Voice file	Closes the session for the given voice file
Play	Audio	Action	Position, duration	Plays back from the current voice file the voice segment defined by the position and duration
Pause	External	Action	None	Freezes the global display for a default period of time
Begin	External	Action	None	A dummy node to start the MRR TSN
End	External	Action	None	A dummy node to end the MRR TSN

cally by the voice system. The choice of logical segments is an issue to be dealt with by the MRR creation process.

External nodes are introduced to provide external flow control required by some complex reporting scenarios. The PAUSE activity is used to halt the presentation, keeping the display static, for a specified (default) length of time. The other two external nodes—BEGIN and END—are dummy nodes used simply to delineate the endpoints of the temporal semantic network.

Arc Semantics

The arc semantics describe how the primitive behaviors can be combined together. An arc specifies a temporal ordering primitive behaviors associated with the two linked nodes, or more precisely, between the time intervals over which these behaviors occur. For an MRR TSN a set of arcs, called temporal predicates, must be defined such that any temporal interaction that may occur in an MRR can be described by this set. The initial set of temporal predicates is listed in Table 3. The selection of temporal predicates begins with the set of 13 primitive temporal relations listed in Table 1. The primitive temporal relations describe fixed temporal interactions, those in which the comparative positions of all four endpoints are known. For example, as shown in Fig 3, all visual actions culminate in displays that begin as they finish. The ordering is fixed and can be described as "action meets display." There are scenarios, however, that require a less restrictive description for which defined temporal relations are useful. For example, all visual behaviors are of random duration, making it impossible to specify when they will terminate. To describe a visual action started during the playback of a voice segment, three situations may arise, as shown in Fig 4. The defined relation starts during accounts for these three situations.

The particular choice of predicates is simply a

Table 3. Arcs in an MRR Temporal Semantic Network

| Relation (I,J)	Definition	Illustration	Inverse (J,I)
Before	Primitive	III JJJJ	After
Meets	Primitive	IIII JJJJJJ	Met by
Contains	Primitive	IIIIIIIIII JJJ	During
Precedes	Before or meets	IIII JJJJ IIIIIIIII	Follows
Begins	Starts or equals or started by	JJJJJJJ IIII IIIIIIII IIIIIIIIIIIII	Begins
Starts during	During or finishes or overlapped by	JJJJJJJ IIIII IIIIIII IIIIIIIIIIII	Starts before
While	Starts or during or finishes or equals	JJJJJJJJJ IIII IIIIII IIIIIIII IIIIIIIIIIII	Encompasses
Concurrently	Starts during or starts before or begins	—	Concurrently
Independent	"OR" of all primitives	—	Independent

question of suitability to the application, convenience, and manageability. Work with MRRs has shown that the set may evolve to include others that more conveniently describe commonly occurring behaviors.

Network Semantics

Network semantics describe how the network as a whole is to be built and interpreted. The organization and the degree of connectedness are addressed. The MRR TSN is organized into a set of $n + 1$ subnets, n screen subnets and one synchronizer subnet (Fig 5). Each subnet describes sequential temporal behavior only. Arcs joining nodes from two subnets represent synchronization requirements for concurrent behaviors. The organization is based on the idea that the MRR presentation resembles a group of cooperating sequential processes. For instance, on a single image screen, updates are performed sequentially. Assuming that multiple image screens are identical, the screens can be modeled as n sequential processes; hence, the n screen subnets.
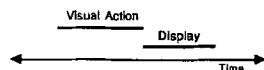

Fig 3. A fixed temporal interaction: meets.
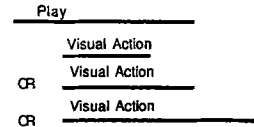
Fig 4. An uncertain temporal interaction: starts-during.

A final synchronizer process acts as a centralized master for the n screen processes, providing the synchronization requirements needed to coordinate and annotate the n screen displays. The synchronizer subnet includes both the voice and external nodes.

The nodes in a MRR TSN are connected according to three guidelines: (1) because the behavior is sequential within a subnet, the nodes are simply chained together, with each node related only to its predecessor and its successor; (2) each node in the synchronizer subnet, whether a voice or external node, represents a synchronization point for the global display. Each synchronizer node is connected to nodes from each screen subnet; and (3) the screen subnets are not linked directly. Instead, the temporal interaction between multiple screens is given indirectly through links to a common synchronizer node. The intent of this "least commitment" strategy is to encourage parallelism in which each screen proceeds independently to the next synchronization point. We note that the network representation of an MRR is not fully connected.

SAMPLE MULTIMEDIA RADIOLOGICAL REPORTS

Interactions between voice and screen displays are basic to the animation of examination reports. The first three examples demonstrate the types of audio-visual interactions possible. These basic scenarios are then extended in the final two examples to handle image sequences and multiple screens.
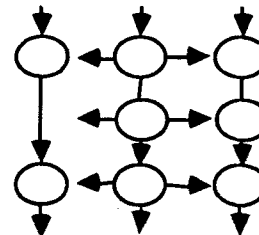
Fig 5. MRR TSN organization.

Scenario 1: Voice Playback With a Single Fixed Display

In this simplest scenario a voice segment is played in conjunction with a completely prepared, stable screen display. The TSN representation given in Fig 6 describes a voice segment being played during the display shown in Fig 7. As in this example, the preparation of the screen display may actually involve a sequence of visual actions. However, the important temporal interaction is between the voice segment and the effect of the final visual action, ie, the final screen display. Only this interaction need be specified in the TSN. In our example, the PLAY activity need not be aware of the LOAD and ZOOM actions.
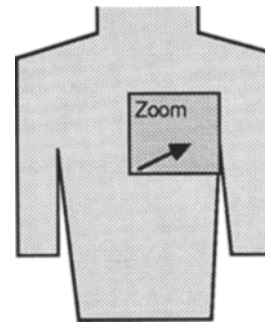
Scenario 2: Voice Playback With Changing Displays

In this scenario, the reporting radiologist starts to dictate while waiting for a screen update to complete. Unlike scenario 1, the temporal interaction is between a voice segment and a visual action. Because of the random durations of the visual actions, the relative start points can be specified but the relative end points are unknown. Figure 8 shows an example in which the radiologist, while waiting for the first image to be loaded, starts dictating the report header: "Reporting for patient John Smith on January 1."

Fig 7. The display described by Fig 6.

Notice that both scenarios 1 and 2 can be easily handled because of the separation of persistent behaviors (eg, displays) from ephermeral behaviors (eg, visual actions).

Scenario 3: Display Updates Overlap Voice Playback

This scenario is the converse of scenario 2: new visual actions are initiated during dictation. The two scenarios are not interchangeable because of the nature of voice. Consider, for explanation, the dictation timeline in Fig 9. Two situations are possible, each involving different temporal interactions.

The first situation is a "don't care" situation. It occurs when the voice segment does not refer to the new display created by this LOAD action. Here, the temporal interaction between the activities is unimportant and can be varied. For example, if the LOAD action is expected to take longer than the given voice segment, it would be more efficient to begin the LOAD action before starting up the voice. The TSN given in Fig 10 describes a "don't care" situation.

In the second situation, some portion of the voice segment does refer to the new display. For instance, suppose the voice segment contains: "Reporting for patient John Smith on January 1,
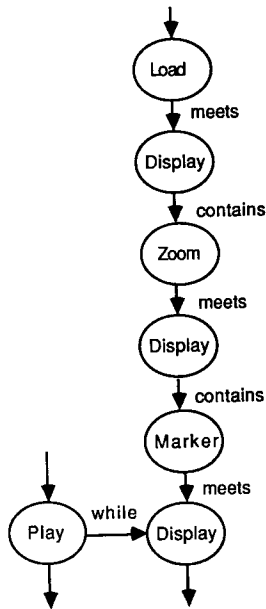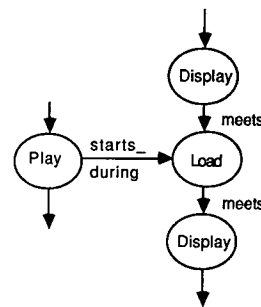

Fig 6. A sample MRR TSN for scenario 1.

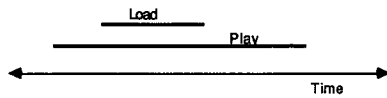Fig 8. A sample MRR TSN for scenario 2.

Fig 9. Dictation timeline.

and on this first image, the frontal view, signs of pneumonia are evident in the upper quadrant of the right lung. . . ." However, as shown in Fig 8, during dictation, the LOAD occurs so quickly that the radiologist does not need to pause. The temporal interaction is complex because: (1) different portions of the same physical voice segment are tied to different visual actions; and (2) the random duration of the visual actions means that the timing during playback may differ significantly from the timing during dictation. This second situation introduces the idea of contextual segmentation. Physical voice segments are subdivided into logical segments such that the temporal interactions are reduced to the simpler scenarios seen above. In our example, two logical segments would be created: (1) "Reporting for patient John Smith on January 1"; and (2) "and on this first image, the frontal view, signs of pneumonia are evident in the upper quadrant of the right lung. . . ." The first logical voice segment is a "don't care" situation. The second is an example of voice playback with a fixed display. The TSN representation is in Fig 11.

Scenario 4: Voice Playback With Image Sequences

A voice segment may refer collectively to a set of screen displays. Because there are a limited number of image screens, the set may appear as a sequence. This scenario differs from contextual segmentation, in which a voice segment refers first to one display and then to another display; a set is used to illustrate a single observation
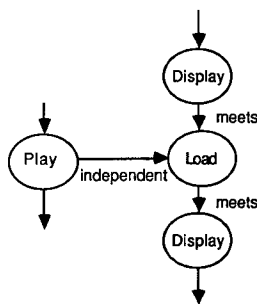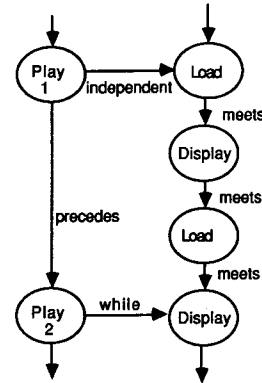

Fig 11. A MRR TSN for contextual segmentation.

described in the voice segment. For instance, the progression of a disease can be traced by comparing a set of equivalent views taken over time. Figure 12 contains the representation for a sequence of three images associated with the following voice segment: "Notice how the infection has spread into the lungs over the past three days." The voice segment is played as the first image is displayed. The remaining two images are sequenced through at some controlled pace, as dictated by the two successive PAUSE nodes.

Scenario 5: Voice Playback With Multiple Image Screens

Multiple image screens involve two types of temporal interactions: the interaction between the synchronizer and the visual behaviors, and

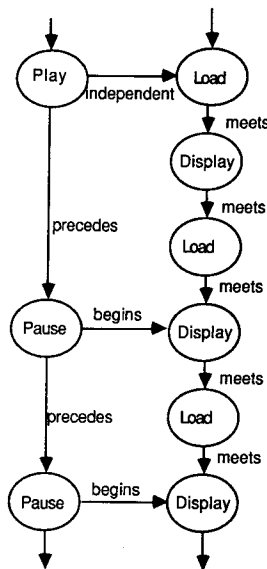Fig 10. An MRR TSN for a "don't care" situation.

Fig 12. A MRR TSN for image sequences.

the interaction between the multiple screens. Each synchronizer node now occurs in relation to a global display that is the combination of the visual behavior of all image screens. The single screen scenarios can be extended by adding a screen subnet for each screen and relating a node in the synchronizer subnet to the appropriate node(s) in each screen subnet. The temporal interaction between multiple screens is given indirectly by a temporal dependency on a common synchronizer node. An exact record of the progress through sequences of intermediate visual activities on the multiple screens is unimportant as long as the final visual activities are all synchronized with the next synchronizer node.

MRR CREATION AND PLAYBACK

MRR creation and playback can be regarded as two processes operating in tandem to perform an automatic two-step transformation of the radiologist's reporting activities into an MRR presentation (Fig 13). We will first discuss the playback process and then conclude with the creation process.

The Playback Process

The playback process automatically coordinates the manipulation of the voice system and n image screens according to the specifications given in the MRR. The playback process has two modes of operation: sequential execution and interrupt execution.

During sequential execution, the playback process performs the unprompted and unaided generation of an MRR presentation. Put simply, the presentation is generated by an orderly forward traversal of the MRR TSN. A central movement down the synchronizer subnet controls the flow of the presentation. Movement to each successive sync node, where movement to a node commences the behavior associated with that node, must be synchronized with parallel movements down each of the n screen subnets. The synchronization requirements are given by the temporal relations extending from the synchronizer node to the nodes in the screen subnets.
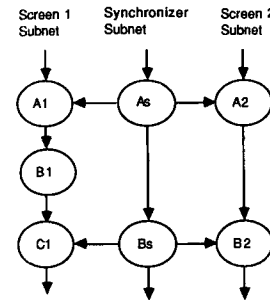
Consider the example in Fig 14. Given an

Fig 14. Playback of a sample MRR TSN.

initial position of (A1, As, A2), the next synchronizer node is Bs. Bs must be synchronized with C1 and B2. Assuming only forward motion during sequential execution, the playback process forward chains along each of the two screen subnets until the related nodes are reached. In this case, Screen 2 is already current but Screen 1 must be updated with B1; the playback process moves to (B1, As, A2). The joint movement to the final position (C1, Bs, B2) can then take place, in the order given by the temporal relations.

Two things must be considered in any implementation of the sequential execution mode for the playback process. First, each synchronizer node will have several related nodes, at least one for each screen subnet. An efficient algorithm is needed to consider all the possible interactions between temporal relations. Second, the software structure should use the parallelism available in the MRR TSN. Concurrent, independent movement down the multiple screen subnets is desirable.

The interrupt execution mode is akin to a supervisory state that handles exceptions and user interactions. An exception is any unexpected behavior that interferes with the MRR presentation, eg, a required file not being available. User interaction allows the user to halt the automatic playback; for example, to pause temporarily, to explore something on his own, or to re-position the playback (wind/rewind).

In both cases, the playback process must plan how to continue with the sequential execution. Sequential execution assumes a certain history, an existing global display. Before re-entering the sequential execution mode, the playback process must ensure that the current global display is compatible with the assumed global display,

Fig 13.   MRR creation and playback.

which is done by comparing the two, and building the assumed display either from scratch or by upgrading the current one. Comparisons are done by retracing the history of the assumed global display as given in the MRR TSN. Rules define the compatibility requirements deciding that two displays are equivalent, not necessarily identical. For example, if the user simply adjusts the contrast to better suit his eyes, a rule can say if the original contrast must be restored or if it can be left as it is. To conclude, the two operating modes highlight the dual nature of the playback process. Sequential presentation is a deterministic, algorithmic procedure. Handling exceptions and user interaction, though, can benefit from a knowledge-based approach to promote flexibility and adaptability.

The Creation Process

The creation process is seen as a two-stage procedure (Fig 15). The first is a "capture" subprocess that records the sequence of the radiologist's reporting activities as a time history. The second is a "generate" subprocess that extracts the MRR from the time history. The capture subprocess is a simple time-stamp generator. The difficult issue here is the selection of events to be captured. The time history must provide the cues needed by the generate subprocess. On the other hand, the cues must be selected so as to not conflict with our primary goal, that of an unobtrusive and automatic report creation facility. To extract the MRR from the time history, the generate subprocess must be able to identify irrelevant exploratory activities to be excluded from the MRR, and to describe the relevant portions of the time history in terms of the known playback scenarios. This subprocess has not yet been fully investigated and is part of the ongoing work. However, it appears that a production system would be an effective vehicle for implementing the MRR generate subprocess.

Fig 15. The MRR creation process.

SUMMARY

An MRR is an animated examination report whose components correspond to image displays, image processing, and voice, text, and graphic annotations. A temporal representation of an MRR describes the animation requirements for its automatic playback. The representation introduced, an MRR TSN captures the temporal nature of the voice-centered MRRs over a distributed multimedia communications system.

An MRR facility must now be built. The playback process will be done first. Completion of the playback process will serve several purposes. First of all, the suitability of the TSN as a representation of an MRR will be demonstrated. The appropriateness of the chosen temporal relations can be tested. Second, the playback process will allow the evaluation of the MRRs through hospital field trials. The reactions of reviewing physicians to an MRR as a useful alternative to the current practice can be gathered. In particular, the multimedia features that add value to the basic examination report can be identified. Finally, the playback process will facilitate the development of the capture process. It can serve as a knowledge acquisition tool for hospital trials geared towards the reporting radiologists. These trials will attempt to find the cues needed to capture a reporting session and the rules that will generate the MRR from the captured time history. The playback process can be used to present reports generated by these rules, thereby evaluating the rules and suggesting further refinements. A second line of research will investigate the interaction of the MRR within a multimedia distributed database. Extensions to include the text and graphic annotations must also be considered.

ACKNOWLEDGMENT

REFERENCES

1. Allen JF: An interval-based representation of temporal Knowledge. Proceedings of the 7th International Joint Conference on Artificial Intelligence, August 1981
2. Allen JF: Maintaining knowledge about temporal Intervals. Communications of the ACM 26:832-843, 1983