Multimedia Document Structure for Distributed Theatre

Jack Jansen
CWI: Centrum Wiskunde & Informatica
Science park 123
1098 XG Amsterdam, the Netherlands
+31 20 5924300
Jack.Jansen@cwi.nl

Michael Frantzis
Department of Computing
Goldsmiths
London SE14 6NW, United Kingdom
m.frantzis@gold.ac.uk

Pablo Cesar
CWI: Centrum Wiskunde & Informatica
Science park 123
1098 XG Amsterdam, the Netherlands
+31 20 5924300
P.S.Cesar@cwi.nl

ABSTRACT
This paper explores the suitability of structured (and declarative) multimedia document formats for supporting a novel type of performing arts: distributed theatre. In distributed theatre, the actors are split between two (or more) locations, but together deliver a single performance mediated by the cameras, the internet, and projection technologies. Based on our efforts to make an actual distributed theatre production happen (the Tempest by Miracle Theatre), this paper reflects on our experience. Our findings are divided into two main areas: workflow and document structure. We conclude that novel types of video-mediated applications, like distributed theatre, require new manners of authoring documents. Moreover, specific extensions to existing document formats are needed in order to accommodate the new requirements imposed by such kind of applications.

Categories and Subject Descriptors
D.3.2 [**Language Classifications**]: Specialized application languages; H.4.3 [**Information System Applications**]: Communication Applications - Computer conferencing, teleconferencing, and videoconferencing; I.7.2 [**Document and Text Processing**] Document Preparation - Languages and systems.

General Terms
Design, Experimentation, Human Factors.

Keywords
Video conferencing, Remote audience, Theatre.

1. INTRODUCTION
In this paper we investigate distributed theatre performances and the technical means to support them. The setting is a theatre play where actors and audience are in multiple locations, and audiovisual streams combined with prerecorded media are used to present a unified experience to the spectators, whether present in one of the theaters or watching from home. When, where and how media and live streams are played back is an integral part of the performance. This should therefore be under control of the artistic director, and will be different for each location at various times during the performance. In addition, the performers in the different locations should be able to act together in a seamless way, as if they are co-located. The requirements can be viewed as combining distributed multimedia playback and telepresence.

Performance artists have creatively made use of physical and, more recently, digital illusionary tools: "there is nothing in cyberspace and the screened technologies of the virtual that has not been already performed on the stage. The theatre has always been virtual, a space of illusory immediacy." [4]. It is not our intention to survey all the various efforts from artists to exploit technology to enhance performances, since [1] and [5] already provide comprehensive overviews. Nevertheless, there are a number of recent pieces that showcase current directions and challenges. "Skype Duet" [10] is a distributed live performance between New York and Berlin. In "Graphic Ships" [6] musicians and a dancer are distributed across locations. Visuals are created based on the movements of the dancer, captured by motion sensing, and create a graphic score from which the live musicians improvise the musical accompaniment. The audience sees the live movement of the local dancer, the visualisation her movement creates and simultaneous projections of the multi-sited musicians themselves.

All these pieces connect two or more locations with the aim of creating a single performance, similar to our experimental performance. Figure 1 shows a scene from "the Tempest" by Shakespeare, performed by *Miracle Theatre*, as experienced from the two theaters, with Ariel (left) in conversation with Prospero (right). In each location there are multiple HD video cameras to capture the scene from different angles. Head-worn wireless microphones capture the audio from each actor. Multiple HD projection screens are embedded in the set to display video

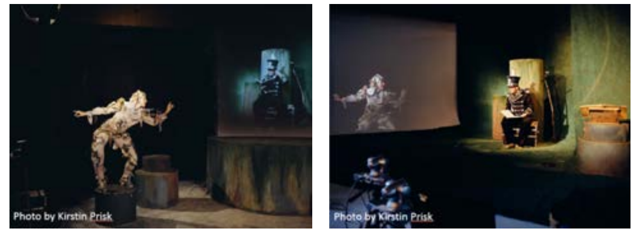

Figure 1 - One scene from "the Tempest", as seen from the two theatres

DocEng '15, September 08 - 11, 2015, Lausanne, Switzerland

ACM 978-1-4503-3307-8/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2682571.2797087

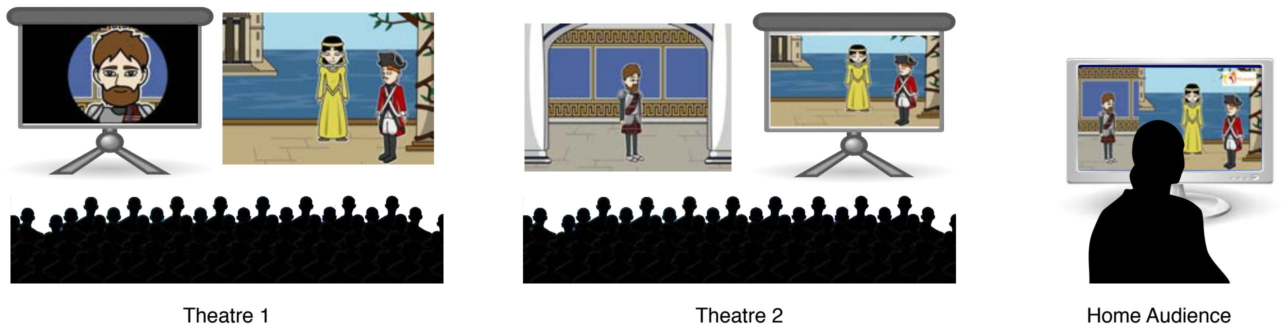

Figure 2 - Schematic representation of locations, audiences and actors

streams from the remote theatre and prerecorded media. Obviously, audio and video need to have low latency and good synchronization to enable fluent interaction between actors [8]. One of the screens is semi-transparent, and this is used in combination with precise timing of prerecorded media and lighting to make "magic" things happen. In addition, the streams and prerecorded media are used to create a single stream for home viewers. Our solution hinges on the fact that the theatre performances we target are scripted: visual layouts and scene transitions and such can be designed before the actual live show.

Figure 2 gives a schematic representation of the setup, and shows that different camera angles and compositions can be used for the different locations. The choice of what to show where, and how to show it (such as the "giant talking head" seen in theatre 1) is a creative choice, made to fit the current scene artistically.

Our setup is intended to be relatively easy to deploy: off-the-shelf hardware, standard broadband internet connections and affordable in personnel cost. While we primarily address distributed theatre we believe the ideas are extensible into other areas where interactions are (somewhat) scripted, such as distributed lectures and classrooms. Our hypothesis is that we can use existing structured multimedia document formats to assist the creative process of designing the visual layouts as well as enabling a single person to manage all locations during the live performance.

2. PROBLEM STATEMENT

The problem statement can be summarized as "*Enabling a creative director to design scene and media changes for multiple locations, and coordinate these centrally during the performance*". The distributed nature of our setting means that the coordination becomes an enabling feature (in stead of merely a convenience, in a single-location setting). The problem statement leads to two sets of requirements: those motivated by the creative process and those from a purely technical media handling perspective.

It emerged that the scripting requirements, and associated model, for the former were distinctive from existing forms of media handling, in particular because of the needs of distributed theatre. Cinema production has the concept of a shooting script which can offer a useful starting point, even though it does not necessarily contain the information of how scenes will be shot and a strict timeline need not be adhered to (the order and content of a narrative can easily be changed before and after capture). In live television, the narrative of events is fixed and strictly governed by the real time constraints of the live action. The distributed theatre model is in some ways closer to videoconferencing, but distinct from it in the fact that the representation can benefit from the larger number of cameras and shots available in television production but it still follows the format of a script, a script that forms an inherent part of the creative theatrical artifact. Our solution will have to provide a good balance between things that can be designed in advance and decisions that should be taken live at showtime.

There is also a technical issue that our solution needs to address. Synchronization and timing need to be handled locally for each location, so that we can maintain lip-sync and correctly time playback of prerecorded media and transitions that need to be synchronized with activity on the stage, or lighting changes and such. However, control on a higher level, the representation script, needs to be centralized. The *representation script* and its associated centralized control interface, *Sync Control*, and the corresponding *Sync Editor*, are not discussed in this paper due to lack of space. The video communication platform, and its innovations, developed to address the challenge of supporting complex use cases for multimedia communication between ad hoc groups is described elsewhere [13].

3. DOCUMENT STRUCTURE

Let us examine how we used structured multimedia documents to enable the requirements set out in the previous section. Reactive declarative multimedia documents have been studied extensively within the document engineering community, for various use cases [9, 12]. Our solution is designed around 3 concepts which are visualized in figure 4: *layouts*, *regions* and *streams*. A layout determines what is seen in one location at a certain point in the play, which can range from a single fullscreen live video region to a complex composition of multiple live regions, prerecorded media (audio, video, image, text) with precise cropping, positioning and begin and end timing, transitions between those, etc. Layouts have symbolic names and layouts with the same name are tailored to each playback location. The home video feed from figure 4 shows an example layout showing two cropped video feeds and a static logo image.

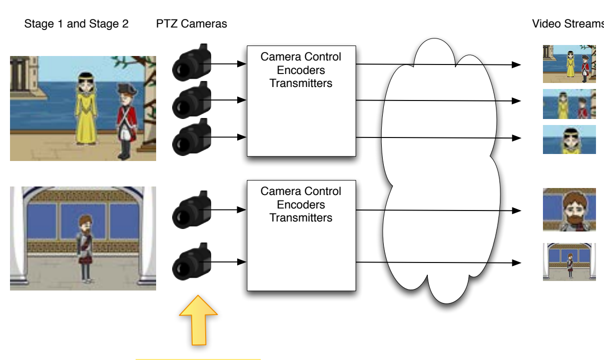

Figure 3 - Video capture

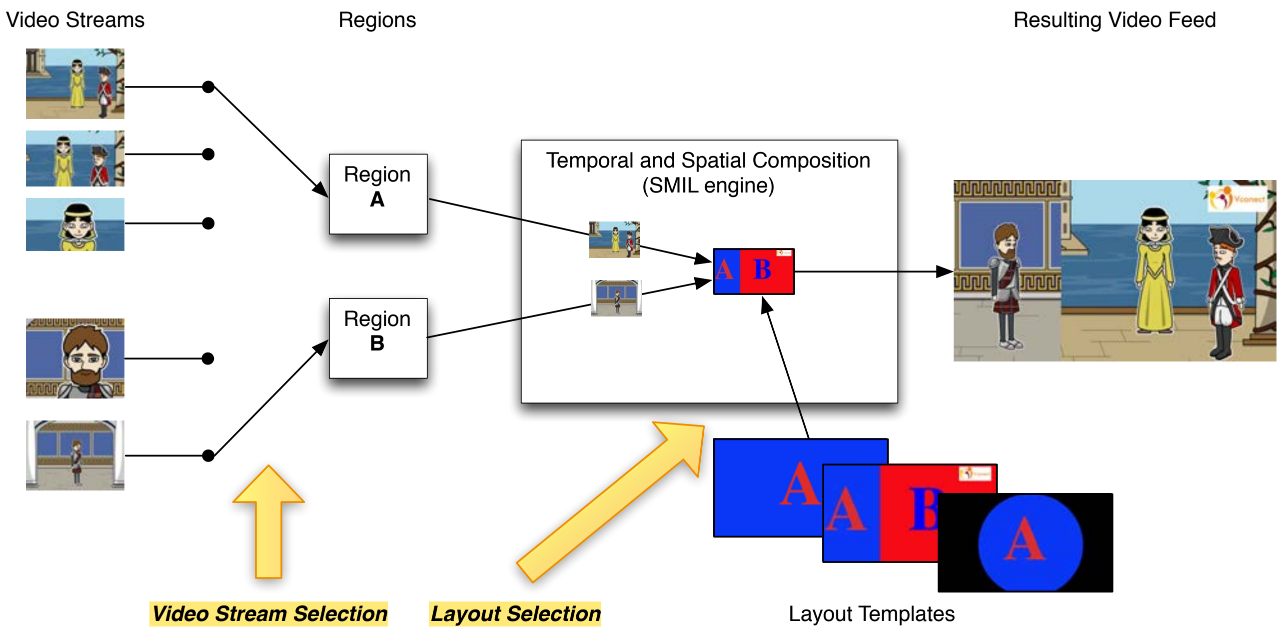

Figure 4 - Video composition (shown for the home audience feed)

Within a layout we can reference regions, which are live stream placeholders and again symbolically named. Regions are long-lived and survive layout switches, so that live streams can continue playing seamlessly during a layout switch. Regions that are not currently visible are on "stand by" so they can start rendering instantaneously when called up during a layout switch or a visual transition within a layout.

Assignment of live video streams to regions is a separate concept from switching layouts, to forestall an explosion of the number of layouts. Also, hard cuts from one camera to another are the most common visual changes. A final concept, *PTZ position*, is used on the video capture side (figure 3). These are per-camera symbolic names such as "*Prospero extreme closeup*" or "*total wide*" and can be called up at will during show time, ideally when the feed from the given camera is not active.

We have implemented streams and regions in a SMIL [3] boilerplate document, and layouts as fragments of SMIL code. These are combined into a single SMIL document per location before document playback.

4. WORKFLOW

The extended document structure, as laid out in Section 3. can support the representation requirements. Nevertheless, there are also a set of requirements for enabling the creation of the documents. Supporting these requirements is extremely important, given that the representation script is an intrinsic part of the theatrical experience. In our specific use cases, creating the multimedia documents is primarily the task of the creative director, who knows what should be shown, and when, and where. There are also technical issues involved, however, such as understanding limitations such bandwidth budgets and the fact that SMIL code needs to be written manually at the moment. This is the domain of the technical director, who closely cooperates with the creative director.

The creative director, while building the representation script, describes the required layouts and PTZ positions to the technical director, who creates the SMIL fragments. These fragments are then assembled into preview documents, one for each location. These preview documents can be played back with an ordinary SMIL player, and contain all prerecorded media and transitions, but use placeholder videos for the live streams. This allows the creative director to check that the composition and transitions work artistically (and that the technical director has understood his intentions correctly).

During on-location dress rehearsal cameras and screens are placed, and the actual PTZ parameters for each shot are determined and recorded. The layouts are fine-tuned, primarily positioning and sizing of items to cater for the physical location of projection screens and such. At this point the final per-location SMIL documents are created. During show time the creative director (or an operator under his instructions) uses the sync control tool to simply step through the layout and camera switches at the right time, for all locations at once. If the need arises special layouts (opening and closing screens and such) can be called up at the press of a single button.

The workflow is similar to theatre lighting: design and fine-tuning happen before the show, and during the performance these prerecorded settings care called up with the press of a button.

5. IMPLEMENTATION

Our implementation consists of a capture and playback engine in each theatre, a playback engine for the home feeds and a single centralized component for sync control (in addition to centralized components for audio and video routing and such [13]).

At show time, the per-location SMIL documents are played back by engines based on the Ambulant [2] SMIL player, with a control module that allows the SMIL code to be modified in a controlled manner using ActiveMQ messages from the central Sync Control component. The playback engines for the theaters also incorporates additional modules to control PTZ cameras and grab/encode/transmit video. The playback engine that creates the feed for the home viewers has modules to do all rendering offscreen and encode and transmit the rendered video stream.

The documents for all locations are structurally the same but different in details, such as which regions are active in which layout and placement of media. Sync Control is isolated from these differences because it communicates with the engines only using symbolic names for streams, regions, layouts and shots. Hereby Sync Control is solely responsible for the global representation script, while the individual engines are responsible for timing, synchronization and layout. The yellow arrows in figures 3 and 4 show how Sync Control can influence the engines.

Changing a multimedia document during playback can have serious consequences for the timegraph and lead to temporal inconsistencies, as we have investigated in earlier work [7]. The scripted nature of a distributed performance has allowed us to use a solution that is similar to the distributed gaming use case from that paper: the multimedia document itself is static and all dynamic changes are implemented through modification of SMIL State variables. *Streams* are implemented through SMIL `<video>` elements that are active throughout the performance, but their rendering position, size and z-order are modified as they are assigned to a region and set to size `(0, 0)` when not assigned to a *region*. *Layouts* are SMIL `<par>` nodes with a begin condition depending on a SMIL variable which is set to activate the layout. They contain all the prerecorded media playback and transitions and such, and also update the region positions (and therefore the live stream rendering positions). This method of changing stream assignments and layouts ensures temporal integrity of the presentation.

6. DISCUSSION

From a creative and artistic perspective the solution as implemented was capable of expressing all the requirements and intentions of the director and crew of the theatrical performance. There were no instances of limitations in the amount of expression offered by SMIL and our associated implementation. The orthogonality of streams and layouts enabled seamless continuation of live streams during layout switches.

The original workflow was based on the principle that the creative director would mark up the representation script which would then become the narrative or theatrical source for media control. However, during the course of the rehearsal it became clear that the capacity to preview and edit on the fly was crucial to a successful production. This brings with it the requirements for flexibility in the document structure to support this process. A limitation of the method of using SMIL fragments was that changing something on the fly required “restarting the world”. Note that fixing this would not necessarily be at odds with [7] as the changes required in the documents on the fly would primarily be about details in timing and layout.

In addition, there was the difficulty in allowing technically inexperienced users to define new layout types and shot positions without the need to communicate with the technical director. A possible direction of future work could be developing a templating system (for example such as [11]) and/or interface which would offer a generic model which could be then be easily adapted for different productions, and involve built in compatibility with preview tools as well as script editing tools.

From the perspective of workload the workflow (and the document structure to support it) performed fine. There were some issues with fine-tuning layout requiring someone to go from one location to the other to see things with their own eyes, but during show time the representation script execution was handled smoothly by a single person.

7. ACKNOWLEDGEMENTS

Part of this work has been done in the Vconect project, funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760. We would like to thank our Vconect partners, especially Doug Williams and Rene Kaiser and Miracle Theatre, especially Bill Scott for enabling us to do this work.

8. REFERENCES

[1] J. Birringer. 2008. Thinking Images: Paul Kaiser and Marc Downie in conversation with Johannes Birringer. In *PAJ: A Journal of Performance and Art* (30) 2, 2008, 17-37.

[2] Bulterman D.C.A., et al. 2004. AMBULANT: A Fast, Multi-Platform Open Source SMIL Player. In *Proceedings of ACM International Conference on Multimedia* (New York, USA, 2004).

[3] D. Bulterman and L. Rutledge. 2004. *SMIL 2.0: Interactive Multimedia for Web and Mobile Devices.* Springer-Verlag, Heidelberg, Germany, ISBN: 3-540-20234-X.

[4] M. Causey. 1999. The Screen Test of the Double: The Uncanny Performer in the Space of Technology. *Theatre Journal* (51) 4, 1999, 383-394. DOI=10.1353/tj.1999.0083.

[5] S. Dixon. 2007. *Digital Performance, a history of new media in theatre, dance, performance art, and installation*. Cambridge MA: MIT, 2007. ISBN: 978-0-262-04235-2.

[6] *Graphic Ships*. 2014. Retrieved 27-May-2015 from https://vimeo.com/83458633

[7] Jansen, J., Cesar, P. and Bulterman. D.C.A. 2010. A model for editing operations on active temporal multimedia documents. In *Proceedings of the 10th ACM symposium on Document engineering (DocEng '10)*. ACM, New York, NY, USA, 87-96. DOI=10.1145/1860559.1860579.

[8] I. Kegel, et al. 2012. Enabling 'togetherness' in high-quality domestic video. In *Proceedings of the ACM international conference on Multimedia* (MM '12). ACM, New York, NY, USA, 159-168. DOI=10.1145/2393347.2393375

[9] P. King, et al. Behavioral reactivity and real time programming in XML: functional programming meets SMIL animation. In *Proc. ACM DOCENG’2004*, pages 57–66. ACM, 2004.

[10] Skype Duet. 2011. Retrieved 27-May-2015 from http://per-aspera.net/en/skype-duet

[11] Soares, L.F.G., et al. 2012. Architecture for hypermedia dynamic applications with content and behavior constraints. In *Proceedings of the 2012 ACM symposium on Document engineering* (DocEng '12). ACM, New York, NY, USA, 217-226. DOI=10.1145/2361354.2361403

[12] Viel, C.C. et al. Go beyond boundaries of iTV applications. In Proceedings of the 2013 *ACM symposium on Document engineering* (DocEng '13). ACM, New York, NY, USA, 263-272. DOI=10.1145/2494266.2494287

[13] D. Williams, et al. 2015. A Distributed Theatre Experiment with Shakespeare. In *Proceedings of the ACM international conference on Multimedia*.