4R－8

Evaluation of Media Synchronization Methods for Multimedia Communication

Jun Sato†, Koji Hashimoto‡, Michiaki Katsumoto§, Hideki Mori† and Yoshitaka Shibata‡

†Dept. of Information and Computer Sciences, Toyo University
‡Faculty of Software and Information Science, Iwate Prefecturel University
§Communication Research Laboratory, Ministry of Posts and Telecommunications

1 Introduction

We have proposed multimedia communication protocol architecture based on the synchronization reference model to realize several applications such as video-on-demand and multimedia conference. In this paper, we developed the multimedia conference system as a possible example of our proposed architecture and evaluated the performance of media synchronization methods to verify the usefulness of our protocol architecture.

2 Unified Multimedia Transmission Protocol

Multimedia services have large styles depending on the target applications and should be relized based on a unified model, not separately. Figure 1 illustrates a unified multimedia transmission protocol architecture which provides multimedia services uniformly for various applications.

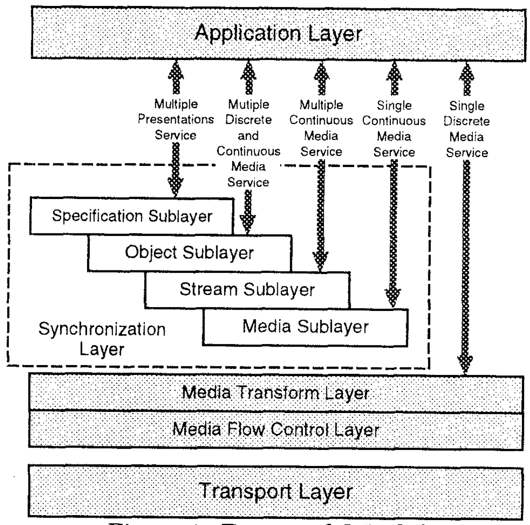

Figure 1: **Protocol Model**

This architecture contains the functions required to provide multimedia services to users and organized by three functional layers: the synchronization layer which takes various meia synchronizations, the media transform layer which performs appropriate media format conversion depending on the used computers and networks, and the media flow control layer which transmits packets over the information networks[1]. Furthermore, in order to support large styles of several multimedia services uniformly, we applied the synchronization reference model[2] in the synchronization layer.

Media Layer: Intra-media synchronization within single continuous media stream such as a audio or a video is performed.

Stream Layer: Lip synchronization between more than two continuous media streams is performed.

Object Layer: Scene synchronization among different types of media streams such as audio, video, image and text is performed based on the presentation scenario.

Specification Layer: Multiple presentations are handled to provide sophisticated multimedia service. The detail of presentation control functions will be explained in following section.

3 Presentation Control Functions

It is considerable that more than two presentations may be provided concurrently in the same application. For an example of the multimedia conference, the live presentations organized by participations face images and voices are exchanged between these participations and stored presentations may be also provided as the reference data of the conference concurrently. In this section, we introduce presentation control functions which handles multiple presentations as shown in Figure 2.

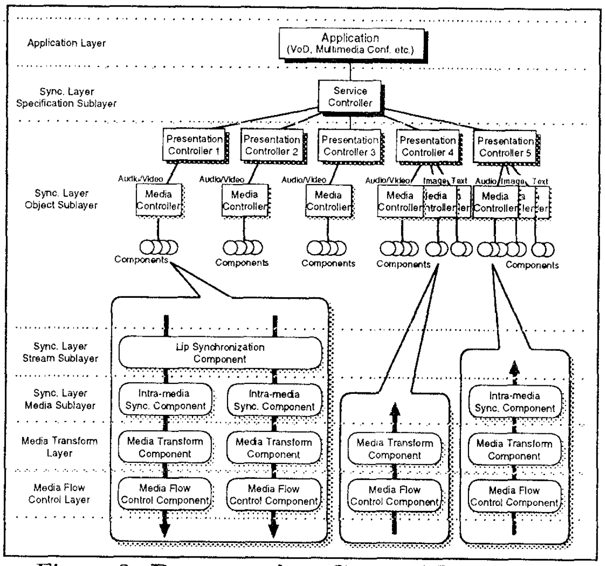

Figure 2: **Presentation Control Functions**

Here, we newly introduce three controllers: service controller(SC), presentation controller(PC) and media controller(MC) to handle multiple different types of presentations independently. The SC controls multiple presentations in multimedia service. The PC controls multiple media streams which consist

---

マルチメディア通信におけるメディア同期制御法の評価
佐藤 純, 橋本 浩二, 勝本 道哲, 森 秀樹, 柴田 義孝

one presentation. The MC performs the media transmission. Media transmission are realized by several components such as lip-synchronization component, media transform component, which are managed by the MC. These components are selected by the MC according to user's QoS requirements and the characteristics of transmitted media stream.

4 Media Synchronization

Multimedia presentations are characterized by the integration of several continuous and descrete media streams. In this section, several synchronization methods including intra-media synchronization and inter-media synchronization are introduced.

4.1 Intra-Media Synchronization

The intra-media synchronization refers time relation of single continuous media stream such as a audio or a video. Logical data units(LDUs)[2], which are a sequence of information units consisting continuous media stream, such as video frames, are adjusted at suitable point on the time axis.

4.2 Inter-Media Synchronization

Two inter-media synchronization, the lip synchronization and the scene synchronization have been proposed. The lip synchronization takes synchronization between more than two continuous media streams such as audio and video. The scene synchronization takes synchronization amoung continuous and descrete media streams and organizes one presentation.

5 Implementation and Evaluation

We implementatd the multimedia conference system to evaluated the performance of synchronization methods and presentation control functions. In the prototype system, while the live audio/video presentations are provided between participations, stored multimedia presentation are also provided as the reference data concurrently as shown in Figure 3.

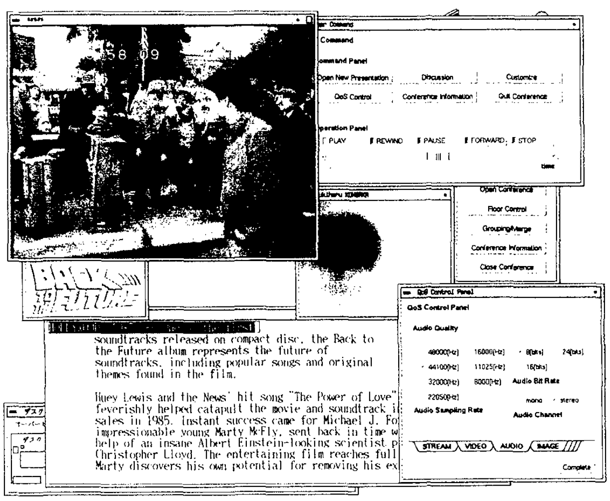

Figure 3: **User Interface of Prototype System**

We evaluated a performance of lip synchronization method which takes synchronization between audio and video based on our prototype system. In the evaluation, we used full color 320 × 240[$pixels$] Motion-JPEG compressed video with 30[fps] and 44.1[KHz] stereo μ-law audio streams on the 100[bps] Fast-Ethernet. The lip synchronization was carried out on every 1.0[sec] and the time difference between audio output and video output was measured in the three cases: 1) synchronization is carried at both sender and receiver, 2) only at sender, and 3) no synchronization.

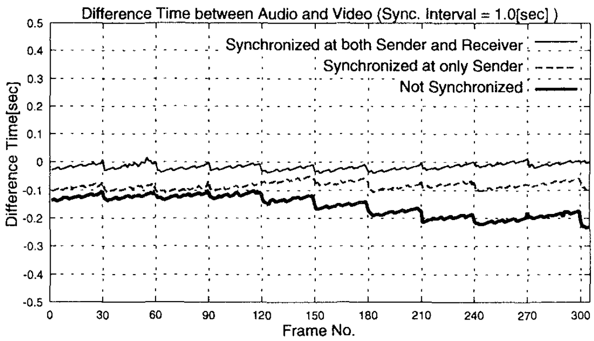

Figure 4: **Evaluation of Lip Synchronization**

Figure 4 indicates the result of the evaluation of lip synchronization. In the case where synchronization was taken at both sender and receiver, time difference of audio output time and video output time could be kept between −0.03[sec] and 0.01[sec]. However, when the synchronization was taken only a sender or the synchronization was not taken, video output time was gradualy delayed from audio output time. Because, video transmission takes the time for software JPEG decompression more than audio transmission. So, it is required to take lip synchronization at both sender and receiver to integrate audio and video streams.

6 Conclusions

In this paper, we developed the multimedia conference system as a possible example of our proposed unified multimedia transmission protocol architecture and evaluated the performance of lip synchronization method in the prototype system. As a result, we could efficiently provide both stored and live presentaions concurrently by same operation in the multimedia conference and verify the usefulness of lip synchronization method. In the future, more practical conference will be held among three locations, CRL(Tokyo), Toyo University(Saitama) and Iwate Prefectural University(Iwate) in Japan.

References

[1] J. Sato, K.Hashimoto, Y.Kohsaka, Y.Shibata and N. Shiratori, "Compressed Video Transmission Protocol Considering Dynamic QoS Control", Proc of Int'l Conf. on Parallel Processing(ICPP-98) Workshop, pp.95-104, 1998.

[2] G. Blakowski and R. Steinmetz, "A Media Synchronization Survey, Reference Model, Specification, and Case Studies," IEEE J. Select. Areas Commun., Vol. 14, No. 1, pp.5-35, January 1996.