高品質インターネット　5－19
(2002. 10. 11)

リアルタイム多地点間遠隔コミュニケーションにおけるデジタル音声処理機構の提案

小峯 隆宏[†,‡]　勝本 道哲[‡]　丹 康雄[†]

†北陸先端科学技術大学院大学　情報科学専攻科　〒923-1292 石川県能美郡辰口町旭台１－１
‡通信総合研究所　インターネットアプリケーショングループ　〒184-8795 東京都小金井市貫井北町 4-2-1
E-mail:　†{tkomine, ytan}@jaist.ac.jp,　‡{komine, katumoto}@crl.go.jp

あらまし　我々は、デジタルビデオ（DV）等の高品質な音声・映像の伝送によるリアルタイム遠隔コミュニケーションシステムを研究開発している。遠隔会議や遠隔講義のような遠隔コミュニケーションでは、画面上の相手と自然で円滑な会話が実現できることが大変重要になる。しかし、既存の遠隔コミュニケーションシステムは、多地点接続の場合でのエコーバック等による音声品質劣化や伝送遅延の増大による不自然な会話などいくつかの技術的課題が残っている。本論文では、DV リアルタイム伝送システムを活用して、多地点間遠隔コミュニケーションで必要な音声マトリクスミキシングや配信映像選択をデジタル信号のまま行える、音声・映像処理機構を提案する。
キーワード　遠隔会議, 遠隔講義, 一体感, デジタル音声, デジタルビデオ, DVTS

Proposal for the digital audio processing mechanism in the real-time multipoint telecommunication

Takahiro KOMINE[†,‡]　Michiaki KATSUMOTO[‡]　and　Yasuo TAN[†]

† School of Information Science, Japan Advanced Institute of Science and Technology
1-1, Asahidai, Tatsunokuchi, Nomi, Ishikawa, 923-1292 Japan
‡ Internet Application Research Group, Communications Research Laboratory
4-2-1 Nukui-kitamachi, Koganei-shi, Tokyo, 184-8795 Japan
E-mail:　†{tkomine, ytan}@jaist.ac.jp,　‡{komine, katumoto}@crl.go.jp

Abstract　We have been developing the real-time telecommunication system that transmits high quality audio/video information such as the Digital Video (DV). The realization of natural and smooth conversations through video screens is very important for successful real-time telecommunication including teleconference or distance education. The existing telecommunication systems have some technical difficulties in the case of multipoint connection, such as the deterioration of audio quality with audio echo back and the interruption of conversations with long transmission time-delay of audio/video information. This paper proposes the audio/video processing mechanism that realizes the audio matrix mixing and video selecting with digital signals, on the assumption that using the DV real-time transmitting system performs the multipoint telecommunication.
Keyword　Teleconference, Distance Lecture, Sense of unity, Digital Audio, Digital Video, DVTS

1. Introduction

The history of teleconference began with TV conferences through ISDN. The transmission audio/video quality of those systems was never sufficient in many ways, and the realization of natural and smooth conversations through video screens was once considered almost impossible.

With the spread of the Internet, the concept of teleconference over the Internet came into public mind. In fact, this turned out as the ongoing development of Internet videoconference systems, such as VIC and VAT, which gave birth to several commercial TV conference systems, such as Polycom ViewStation. Commercial servers for telecommunications have already become ready and available.

The existing teleconference systems have some technical difficulties in the case of multipoint telecommunication with natural multi-interaction. These multipoint telecommunications require the audio/video processing mechanism such as audio matrix mixing and video selecting somewhere. The audio/video processing mechanism can causes the deterioration of audio quality with many

Analog/Digital (A/D) or Digital/Analog (D/A) conversions and the interruption of conversations with long transmission time-delay of audio/video information.

In order to achieve real-time multipoint telecommunication with natural and smooth conversations, this paper proposes the audio/video processing mechanism that realizes the audio matrix mixing and video selecting with digital signals, on the assumption that using the DV real-time transmitting system performs the multipoint telecommunication.

Chapter 2 explains the definition of the teleconference we suppose and the "sense of unity" that we consider as the important factor for the realization of natural and smooth conversations through video screens. Chapter 3 describes technical difficulties in the case of multipoint telecommunication with natural multi-interaction and how to improve them. We propose the audio/video processing mechanism that realizes the audio matrix mixing and video selecting with digital signals by making use of Digital Video Transport System (DVTS) that performs the multipoint real-time telecommunication in chapter 4. Finally, chapter 5 concludes this paper by informing our future improvement plans and evaluation methods.

2. Teleconference and "sense of unity"

We define teleconference in this paper as real-time multipoint telecommunication with natural multi-interaction. Especially, we focus the situations that it is important to attend with concentration and to discuss freely, such as a research conference connecting multipoint sites and distance lecture.

As a research of teleconferences, Japan Advanced Institute of Science and Technology (JAIST) and Communications Research Laboratory (CRL) have been conducting joint research on distance lecture. A distance lecture trial was performed as a JAIST intensive lecture program during two weeks on September 2000 and some students took this lecture for two credits, which was the first case among national universities in Japan. In this experiment, some instructors are located at CRL and gave some lectures to JAIST graduate students in a JAIST lecture room by using a real-time bi-directional telecommunication system with natural multi-interaction via Japan Gigabit Network (JGN) [1]. From October 2001, the regular distance lecture trials have been performing on the condition that instructors in a JAIST lecture room give some lectures to JAIST graduate students in a JAIST lecture room and a CRL remote lecture room via JGN (shown in Figure 1).

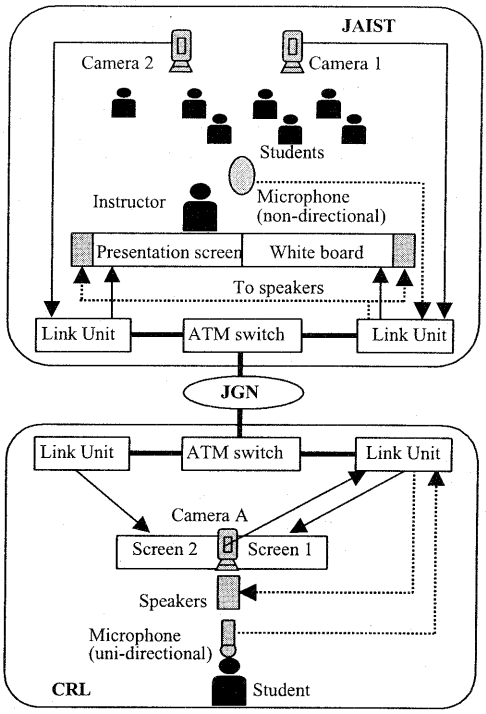

Fig.1 Distance lecture trials between JAIST and CRL

We define "sense of unity" as the keyword to realize natural and smooth conversations through these teleconferences. "Sense of unity" in this paper means a kind of feelings that enable all attending members having a sense of same purpose for this teleconference to spend the same period of time and to attend or study together. The enhancement of "sense of unity" leads attending members in the teleconference to create a tense atmosphere that they have when they attend local conferences or local lectures, and to keep high concentrations for its teleconference. We consider that the improvements of how to transmit audio/video information and how to express them enhance "sense of unity".

3. Technical difficulties

The existing telecommunication systems have some technical difficulties in the case of multipoint connection, such as the deterioration of audio quality with A/D or D/A conversions and the interruption of conversations with long transmission time-delay of audio/video infor-

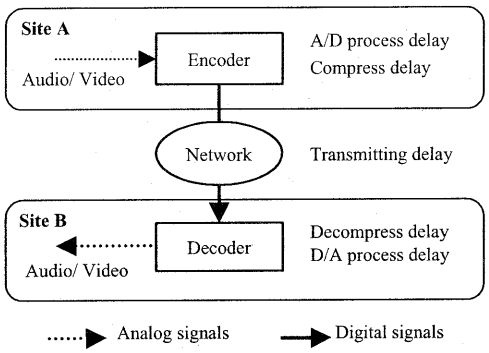

Fig. 2. Flowchart of audio/video information to site B in case of 2 point's telecommunication

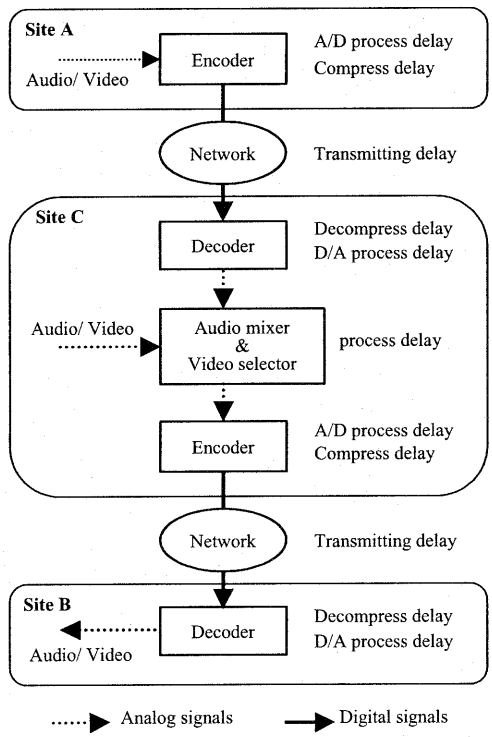

Fig. 3. Flowchart of audio/video information to site B in case of 3 point's telecommunication

mation. Especially, slow reactions from conversation's partners caused frustrating conversation with interruptions by the long transmission time-delay during end-to-end.

Figure 2 shows the flowchart of audio/video information from site A to site B in the case of mutual teleconference between two sites. Analog signals of audio/video information are converted digital signals and compressed at the encoder of teleconference system. The digital signals are transmitted to the decoder of teleconference system through network, and they are decompressed and converted analog signals again. Therefore, we consider that it is effective to select the encoder/decoder sets that can transmit audio/video information with shorten time-delay by some process delay such as A/D or D/A conversions delay and compress/decompress delay.

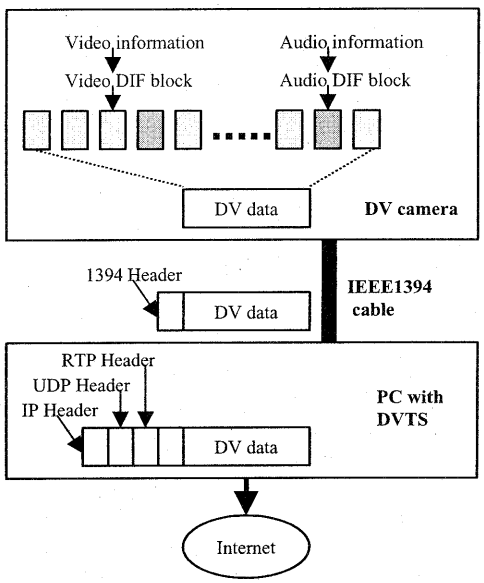

Fig. 4. Data structure of DVTS

Figure 3 shows the flowchart of audio/video information from site A and C to site B in the case of mutual teleconference among three sites. Site C acts as hub of transporting audio/video information from all sites. In this case, it is necessary to use additional encoder/decoder sets and the audio/video processing mechanism such as audio matrix mixing process and video selecting process. We consider that it is effective to reduce the number of A/D or D/A conversions because the process of A/D or D/A conversions has some possibility to deteriorate the quality of audio/video information.

4. Proposal system

We propose the audio/video processing mechanism that improves the technical difficulties indicated in section 3.

At first, we select DVTS as the real-time transmitting system for teleconference with natural multi-interaction because of following reasons [2]-[4].

a. High quality of audio/video information

b. Small time-delay

c. Low cost

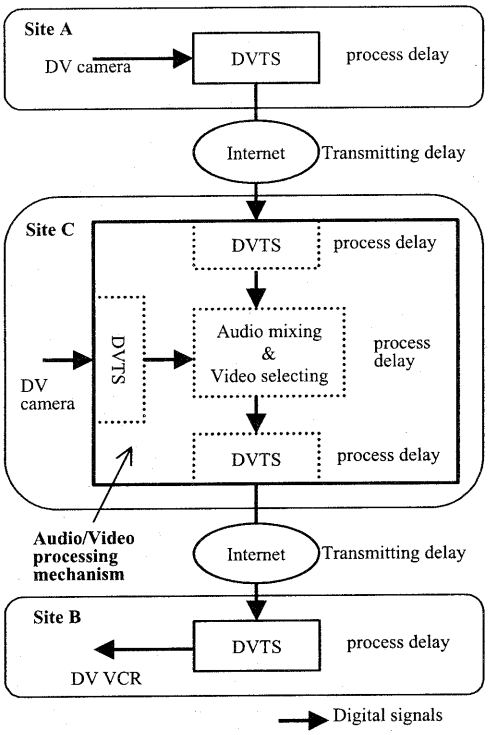

Fig. 5. Flowchart of audio/video information to site B by using our proposal system

DVTS transmits the digital video data over the IP packets without any inter-frame compression, and is also application system on regular or note PCs. DVTS can handle video information of NTSC TV broadcasting quality and audio information of PCM signals with 48 kHz sampling.

Figure 4 shows the data structure of DVTS. Every DV data is constructed with 80 bytes DIF blocks such as the audio DIF block including digital audio information and the video DIF block including digital video information. DVTS carries out real-time mutual signal conversion between RTP/UDP/IP packets through the Internet and DV data through IEEE1394 cable.

This paper presents the audio/video processing mechanism with digital audio/video signals by making use of DVTS. Figure 5 shows the flowchart of audio/video information from site A and C to site B by using this mechanism in case of mutual teleconference among three sites. The digital signals including audio/digital information from DV camera are transmitted to the audio/video processing mechanism from site C directly and from site A through the Internet. They are decomposed into the audio DIF block and the video DIF block, and the output data by some digital processing such as audio matrix mixing or video selecting are transmitted to site B through the Internet. We consider that this mechanism can shorten the time-delay during end-to-end and keep the quality of audio/video information because the number of A/D or D/A conversions reduces well as compared with the case of Fig. 3.

Figure 6 shows overview of audio/video processing mechanism in Fig. 5. This mechanism mainly consists of some input/output blocks, the process of audio matrix

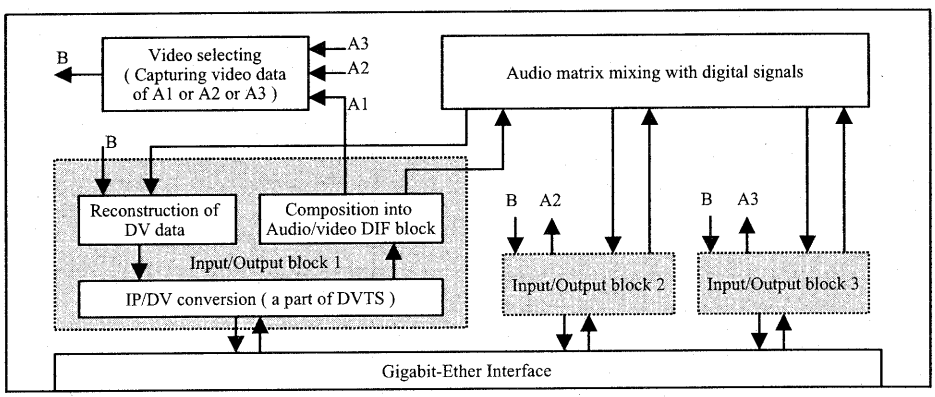

Fig. 6. Overview of audio/video processing mechanism with digital signals

mixing, and the process of video selecting. Each input/output block has three functions of the IP/DV conversion that is a part of DVTS, the composition into audio/video DIF block from DV data, and the reconstruction of DV data from processed audio information and video information. The process of audio matrix mixing calculates the output digital signals by matrix mixing (D1, D2, D3 in Fig.6) against input digital signals that mean PCM audio signals in audio DIF block (C1, C2, C3 in Fig.6). In this case, for example, the value of D1 is addition of the value of C2 and C3 except C1, because there is no need for returning own site's audio. The process of video selecting captures one of input digital video signals (A1, A2, A3 in Fig.6) every frames according to user's choice and transports the output digital signals (B in Fig.6) to every input/output block.

We also consider that this audio/video processing mechanism with digital signal have another potential for dealing with digital information through teleconference via the Internet. For example, there is worth considering for full digital echo canceling mechanism instead of existing echo canceling device with analog audio inputs and digital calculations [5]. This processing mechanism with digital information has possibility that it is easy to attach value added information and install another processing for enhancement of the teleconference.

5. Conclusion

We figured out technical difficulties in the case of multipoint telecommunication with natural multi-interaction and the "sense of unity", which was the important factor for the realization of natural and smooth conversations through video screens, was very important for successful real-time telecommunication including teleconference or distance education.

This paper proposed the audio/video processing mechanism that realizes the audio matrix mixing and audio echo canceling and video selecting with digital signals by making use of DVTS that performs the multipoint real-time telecommunication. This mechanism will be able to keep audio/video quality and to reduce the transmission time-delay by cutting some A/D or D/A conversions process. We also consider that this audio/video processing mechanism with digital signal have another potential for dealing with digital information through teleconference via the Internet.

We intend to develop our proposing audio/video processing mechanism, and the implementation based on software programs on the personal computers is under development at present. For the future, we will evaluate its performance.

Acknowledgement

The authors would like to thank their staffs at JAIST and CRL for their helpful comments and support.

This research is the joint research with TAO using JGN, JGN-G11027 [6].

Reference

[1] T. Komine, A. Machizawa, S. Nakagawa, F. Kubota, and Y. Tan, "Development of high presence video communication system – Trial experiment of the Next Generation real-time remote lecture -," Proc. 16th International Conference on Information Networking, vol.2, 4C-4, Cheju Island, Korea, Jun. 2001.

[2] A. Ogawa, "DVTS (Digital Video Transport System) WWW page," http://www.sfc.wide.ad.jp/DVTS/, 2001.

[3] A. Ogawa, K. Kobayashi. K.Sugiura, O. Nakamura, and J. Murai, "Design and implementation of DV based video over RTP," Packet Video 2000, pp.140-146, May 2000.

[4] K. Sugiura, T. Sakurada, and A. Ogawa, "Demonstration of high quality media transport system using Internet," IPSJ Journal, vol.41, no.12, pp.1321-1326, 2000.

[5] J. Ohga, "Howling, its physics, prediction and suppression," Journal of the Acoustical Society of Japan, vol.56, no.2, pp.115-120, 2000.

[6] Telecommunications Advancement Organization, "JGN (Japan Gigabit Network) WWW page," http://www.jgn.tao.go.jp/, 2002.