S-TEL: An avatar based sign language telecommunication system

Tomohiro Kuroda, Kosuke Sato and Kunihiro Chihara

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, 630-0101, JAPAN

{tomo,sato,chihara}@is.aist-nara.ac.jp

ABSTRACT

Although modern telecommunication have changed our daily lives so drastically, the deaf cannot benefit from them based on phonetic media. This paper introduces a new telecommunication system for sign language utilizing VR technology, which enables natural sign conversation on analogue telephone line.

On this method, a person converses with his/her party's avatar instead of party's live video. As speaker's actions are transmitted as kinematic data, the transmitted data is ideally compressed without losing language and non-language information of spoken signs.

A prototype system, S-TEL, implementing this method on UDP/IP, proved the effectiveness of avatar-based communication for sign conversation via a real lossy channel.

1. INTRODUCTION

Although modern telecommunication has changed our social communication style so drastically, the audibly challenged cannot take benefits of them based on phonetic media. In order to associate their isolated community together and to increase the quality of their daily lives, a new communication system for signers is indispensable.

Today the deaf use TTY or facsimile instead of telephone. However, with these character based communication systems, they need to translate their sign conversation into descriptive language and to write down or type in. So, they eager a new telecommunication system which enables them to talk in signs.

Nowadays so many research works on computer aid for signers including telecommunication systems for signers are coming. Some of these research works have developed data compression techniques for video stream of sign language, others have developed script-based sign communication methods that translate signs into descriptive languages to reduce transmitted data. These methods succeeded to compress transmitted data, but the language and non-language information contained in signs is lost due to the above compression or the translation. Unfortunately, they cannot mediate natural sign conversation.

Therefore, Kuroda et al. (1995) introduced a concept of new telecommunication method for sign language integrating human motion sensing and virtual reality techniques. In this paper, a new realized telecommunication system based on this method is introduced.

In this system, a person converses with his/her party's avatar instead of the party's live video. Speaker's actions are obtained as geometric data in 3D space, the obtained motion parameters of the actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. Thus, it realizes optimal data compression without losing language or non-language information of given signs. Moreover, users can hide their private information without giving displeased feeling.

In this paper, the features of Japanese Sign Language are briefly explained in section 2 and foregoing studies on sign communication are mentioned in section 3. In section 4, the avatar-based communication method is introduced, and avatar-based communication and video-based communication is compared from bandwidth viewpoint. Finally, in section 5 a prototype avatar-based communication system, S-TEL, is experimented on UDP/IP by Deaf and sign experts.

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

2. JAPANESE SIGN LANGUAGE

Japanese Sign Language (JSL) is the mother tongue of Deaf for a hundred years in Japan (Komekawa, 1998). As JSL is a visual language, there are some features in comparison with phonetic languages.

- The meanings of signs are defined by hands' feature, position, movement, and direction (Fig. 1).
- Some signs cannot identify from frontal view because of occlusion (Fig. 2).
- Three-dimensional position of signs denotes the relation of the current communication. Persons and their ranks are shown by the position to indicate signs or the direction of upper body (Fig. 3).
- Non-Manual Signals (NMS) like nodding work as modifiers.

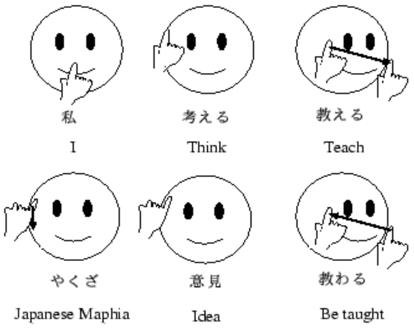


Figure 1. *The meanings are determined by hands figure etc.*

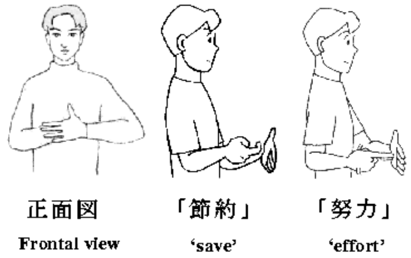


Figure 2. *Some signs cannot be identified from frontal view because of occlusions.*

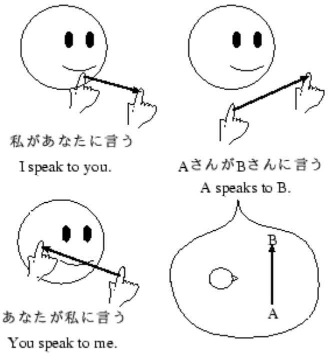


Figure 3. *The position of sign shows character and its rank*

3. FOREGOING RESEARCHES

There are many research efforts on computer aid for signers. However, most of them focus on sign translation or sign education. Only few attempts have so far been made at the communication among signer. These researches on communication among signer can be divided into two groups, which is script-based communication and video-based communication.

4.1 Script-based Communication

Jun et al. (1991) and Ohki (1995) proposed script-based communication system. These systems translate given signs into script language or phonetic language, transmit them, and produce sign animation on receivers terminal. Therefore, they can eliminate bandwidth of transmission so drastically. However, as these script-based systems have translation stage, they cannot forward given sign when their dictionaries have no entry for the given sign or they mistranslate it.

4.2 Video-based Communication

As sign language is visual language, videophone seems usable as telecommunication system for sign language. However, Yasuda (1989) pointed out following 'Eyes Torture' problem of videophones for daily life use.

- User's eyes may break into the other's private space through camera. The other's privacy may be trespassed.

Kamata (1993) experimented with some videophones and cleared that following problems make difficult of sign conversation on videophones

- 2D videophones images cannot show whole sign information, because sign language moved in 3D space.
- The received images do not have enough resolution, view, and frame rate for practical sign conversation.

As sign language includes fast hand motions and occluded postures, general methods for video compression are not suitable for sign communication. Sperling et al. (1985) and Gulska (1990) proposed video compression method for sign language, but their methods cannot transmit sign image sequences that has enough time/space resolution to read on analogue telephone line.

4.3 Bandwidth of Video-based Communication

Kamata (1993) argues that image sequence of QCIF mode ISDN videophones ($176 \times 144\text{(pixel)} \times 15\text{(fps)}$) doesn't always has enough time/space resolution for sign conversation. However, Sperling et al. (1985) says that three bits gray scale image sequence of $24 \times 16\text{(pixels)} \times 15\text{(fps)}$ can visualize 'enough intelligible' ASL. Assuming three bits gray scale image, Kamata's system requires at least 2.2Mbps and Sperling's requires 34Kbps for bi-directional communication. As these results vary so widely, we examined required bandwidth for video-based sign communication.

Firstly, to examine required frame rate, we selected two topics (79 seconds) from NHK sign language news, and measured how long frames each sign word continues. As Tab. 1 and Fig. 4 shows, some words continue less than 1/30 seconds. Moreover, Kanda et al. (1996) cleared that newscasters speak about 70% speed of normal sign conversation. Thus, the conclusion is that video rate (30 frames per second (fps)) is not sufficient for sign conversation. However as there are no faster display for home use, we assume video rate display in following discussion.

Table 1. *The Number of Frames Each Sign Continues. (*1^- *denotes less than 1 frame.)*

	News 1	News 2	Total
Average	5.72	9.36	7.84
Minimum	1^-	1^-	1^-
Maximum	21	30	30

Secondly, to examine required resolution, we selected 10 frames ($640 \times 480\text{(pixels)}$) from NHK sign language news, and measured the width of fingers. The narrowest finger's (pinkie of woman) width was five pixels. Thus, sign image requires 128×96 pixels to identify each finger.

From these discussions, assuming three bits gray-scale image, the required bandwidth for bi-directional video-based sign communication is 2.2Mbps. Therefore, video-based sign communication requires high bit-rate digital channel.

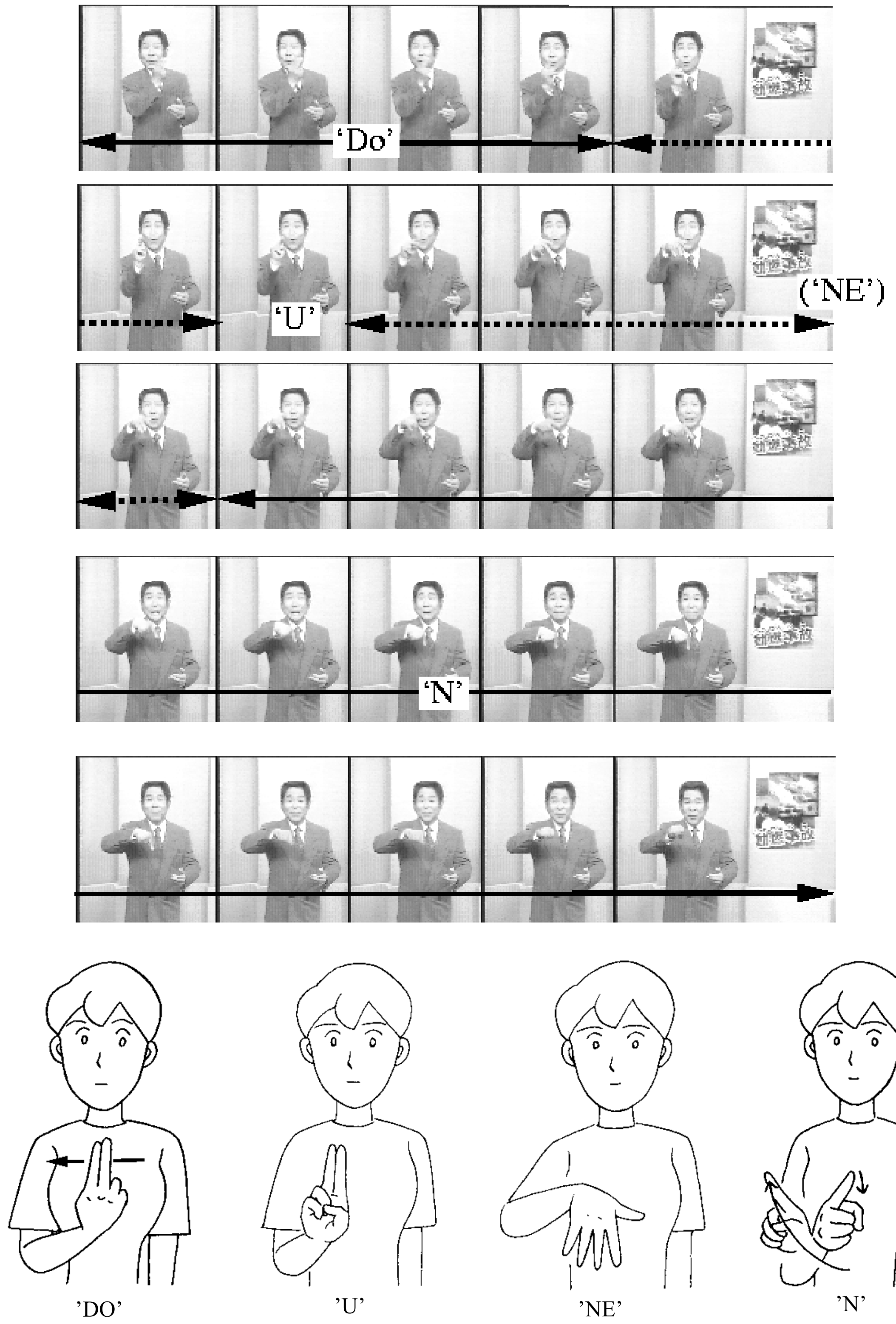


Figure 4. *Finger Spelling 'DOUNEN' (Power Reactor and Nuclear Fuel Development Corporation) from NHK Sign Language News. Finger Character 'NE' continues less than 1/30 seconds.*

4. SYSTEM DESIGN

4.1 Avatar-based Communication

To solve above problems of the previous telecommunication systems, we introduce new telecommunication system for sign language integrating human motion sensing and virtual reality techniques. This method solves natural sign conversation on conventional analogue telephone line.

In this method, a person converses with his/her party's avatar instead of the party's live video. Speakers actions are obtained as geometric data in 3D space, the obtained motion parameters of actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. This avatar-based communication has following advantages. Firstly, sending kinematic data without any translation process, this avatar-based communication realizes data compression without losing language or non-language information of given signs. Secondly, sending 3D motions, receivers terminal can produce signing avatar to increase readability of signs. Thirdly, this method can be applied to conferencing or party talking (Kuroda, 1997b). Finally, visualizing avatar instead of live video, users can hide their private information without giving displeased feelings for his/her party.

This system consists of following components as shown in Fig. 5.

- **Sender** obtains signs as geometric data and sends it. **Motion measuring part** measures hands, arms and upper body motions. **Encoder** encodes and compresses obtained data.
- **Receiver** receives kinematic data of signs and displays avatar. **Decoder** decodes given data into kinematic data. **Avatar producing part** produces avatar from given kinematic data and display it. This part makes use of reader's viewpoint information if needed. Produced avatar can be virtual CG avatar, robot, etc.

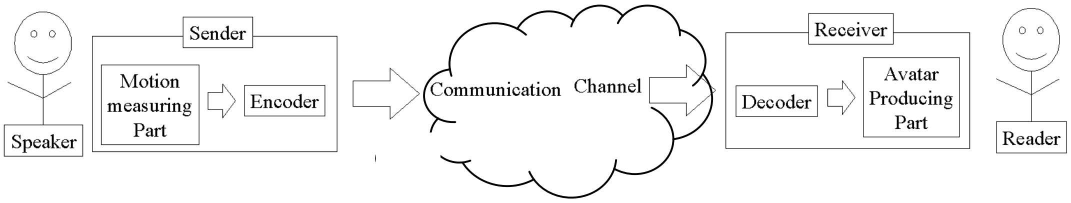
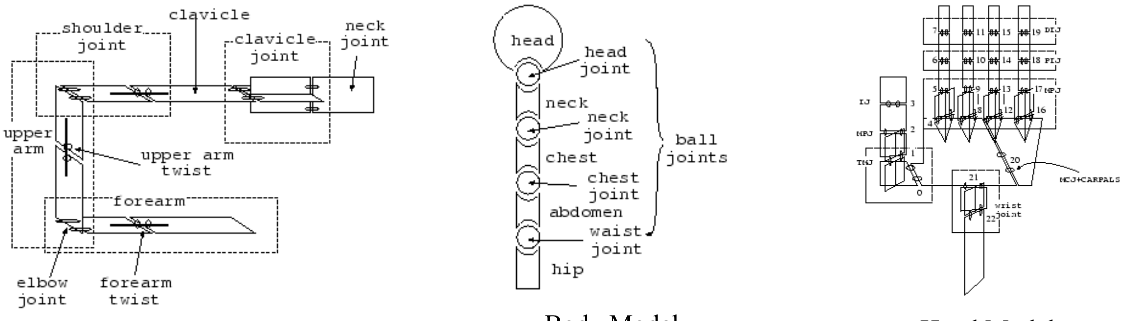


Figure 5. *Overview of Avatar-based Communication*

4.2 Bandwidth of Avatar-based Communication

Avatar must handle whole upper body as signer shows signs with hand, arm and upper body motions. Therefore, we assume skeleton of avatar as shown in Fig. 6. This model has following features.

- The spine consists of 33 or 34 vertebras, and small rotation between these vertebras produces backbone bend. Especially, 5 cervical vertebras and 7 lumber vertebras moves much more than the other vertebras. Thus, this model has joints at both sides of these two parts.
- Some sings like ASL 'why' visualized by shoulder motion. Therefore, this model has joints on neck side of clavicle.



Arm model

Body Model

Hand Model

Figure 6. *Skeleton Model of Avatar*

Suppose finger width is unit length as section 4.3. Finger width for woman is about 2 cm and the sitting woman's wrist move inside a sphere which diameter is 2m. Therefore, seven bits code can denote wrist position, and eight bits code can denote 360^o . This model has 72 degree of freedom and amount of rotation of each joint is as Tab. 2. Thus, 460 bits code can denote whole upper-body posture. Kuroda et al. (1996b) proposed a method to reconstruct upper body motion form position and rotation data of wrists and top of head. Applying this method, 415 bits code can denote whole upper-body posture. From these discussion, the required bandwidth of bi-directional avatar-based sign communication 16Kbps. Thus, avatar-based communication is available on conventional analogue telephone line.

Table 2. *Joints of Avatar*

Part	Degrees of freedom	Rotation	Bits
Hand	42	90	6
Wrist	4	180	7
Clavicle joint	4	90	6
Rotation around clavicle	2	270	8
Others	20	180	7

5. EXPERIMENTS

5.1 Prototype System S-TEL

A prototype, S-TEL, along the design discussed in section 4 is developed as Fig. 7. Kuroda et al. (1996a) cleared that 3D stereo scopic view has no effect on the readability of signs and that 2D CG reflecting readers motion parallax is sufficient to realize practical readability. Therefore, S-TEL uses normal 2D display as shown in Fig. 8.

S-TEL sender composed of Pentium 166MHz PC with Windows95, two CyberGloves and a Fastrak. S-TEL receiver composed of Intergraph TD-5Z workstation (Pentium 100MHz with OpenGL accelerator) with WindowsNT 3.51 and a Fastrak. All software components are built on World Tool Kit Ver. 2.1 for WindowsNT and Visual C++ 2.

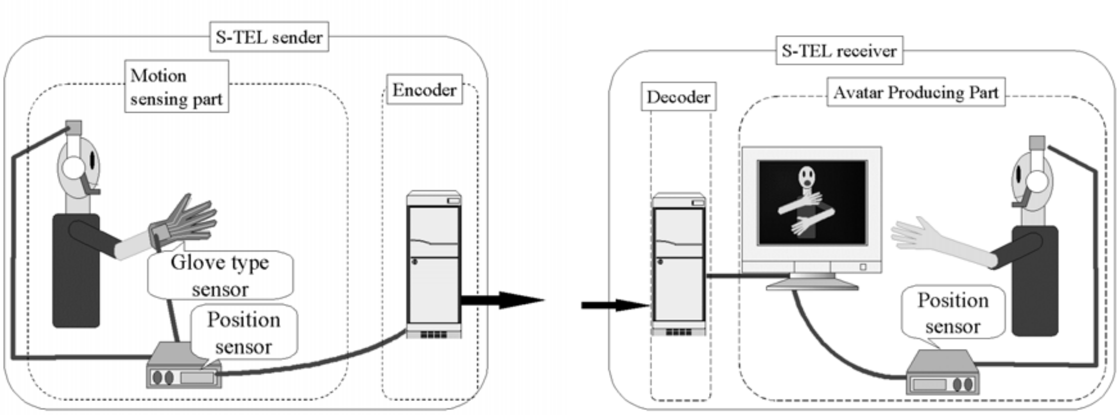


Figure 7. *Design Overview of S-TEL*

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

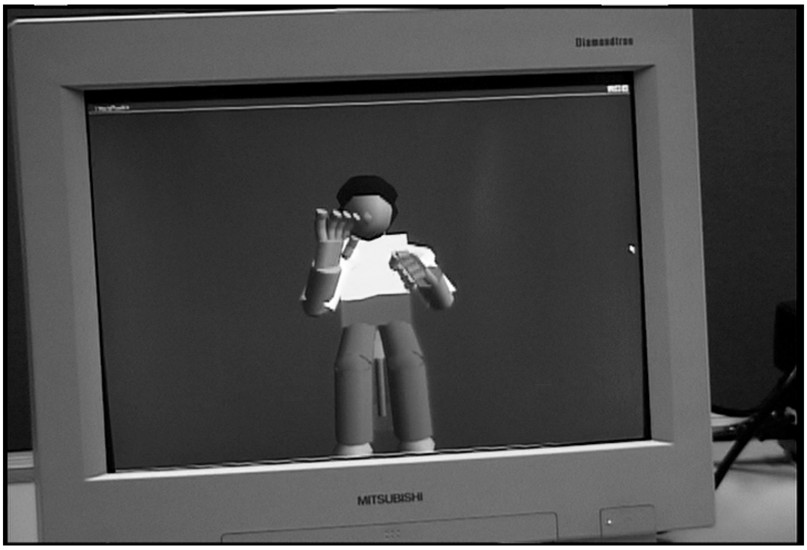


Figure 8. *Produced Avatar of S-TEL*

5.2 Bandwidth of S-TEL

S-TEL obtains signer's action with CyberGloves and Fastrak. CyberGlove obtains 18 finger joint bending as integer and Fastrak obtains position as 3 single float and orientation as single float quaternion (4 degrees of freedom). Therefore, the amount of data of one flame is 120 bytes. Assuming 30 fps, bandwidth of mono-directional S-TEL is 28.8Kbps.

5.3 Experiments

A developed prototype S-TEL was experimented with two types of UDP/IP communication channel.

Firstly, we connected S-TEL by Ethernet. Three deafs, three sign experts and four sign beginners tried to talk freely in sign language on S-TEL.

Secondly, we placed S-TEL sender at NAIST, Ikoma City (Nara) and S-TEL receiver at Kumamoto City (Kuroda et al., 1997a). The distance between two cities is about 700Km. Satellite JCSAT-1 connected two cities. The bandwidth of channel was 4.0Mbps bi-directional. By adding heavy traffic on the communication channel, there is 2% packet loss due to overload of the channel. Speakers were two sign experts and readers were one sign expert and two beginners. All testees are hearing. Testees tried to teach sign language on S-TEL. Testees also tried to talk in signs through NV/VAT Internet video chat, and they compared these two systems.

5.5 Results and Discussions

Experimental results are as follows.

- The realized frame rate of S-TEL is 26.1 fps. S-TEL can perform practical frame rate to read signs on consumer PC.
- 2% packet loss had no effect on the quality of visualized signs. Because, S-TEL detects erroneous flame and keep the continuity of the preceding flame image. This ensures S-TEL works properly on lossy channel.
- Readers could recognize 75% of spoken signs. Users can almost make themselves understood through S-TEL. Speakers signs which readers couldn't recognize are the following cases.
  - ◊ Readers couldn't recognize spoken signs which facial expressions modify, because S-TEL doesn't treat facial expressions.
  - ◊ Readers couldn't recognize finger spelling, because its finger size of the avatar is not sufficient to distinguish fingers.
- Sign experts completed a difficult task to teach 'self-introduction' for beginners over S-TEL. Usually, for beginner, one must teach at hand with utmost care and kindness. It is one proof that S-TEL can provide the environment where users can talk in signs as if they talk at hand.

- Testees said that they prefer S-TEL rather than videophones as sign communication media, because videophones could expose their privacy but S-TEL wouldn't. They also said S-TEL would enrich their daily lives.

These experimental results clear that avatar-based sign communication is effective as users can make themselves understood through S-TEL. Moreover, avatar-based communication is superior to video-based system in bandwidth viewpoint; avatar-base communication is available on lossy and narrow channel.

To increase readability of signs, avatar-based communication should treat facial expressions and visualize hand bigger. However, to make hand bigger causes another problems. Some researchers on sign animation including us did tried to make it bigger. Testees complained that unbalanced avatar makes signs unreadable and that testees sometimes feel hit by avatar when avatar stretches their arm to the front. Thus, to increase readability, avatars hand should expand when signer start to spell. It is needed to develop new method to identify whether spoken signs are finger spelling or not.

6. SUMMERY

In this paper, avatar-based sign communication system, which is an innovative system for sign language telecommunication, is presented. In this method, speakers actions are obtained as geometric data in 3D space, the obtained motions parameters of actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. As avatar-based communication treats speaker's actions as geometric data in 3D space, it allows to talk and read signs naturally through conventional analogue telephone line, increase readability of signs, and protects users' privacy.

The experiments to talk in signs through a prototype system, S-TEL, were performed. These clear the effectiveness of avatar-based communication and the superiority of avatar-based communication over video-based communication as a telecommunication media for sign language.

When S-TEL gets popular among Deaf widely, their isolated community would associate together. S-TEL would increase the quality of their daily lives. Authors are integrating the transmission of facial expressions of signers and the identification of finger spelling into S-TEL in progress.

Acknowledgement: This study is cooperated with Kamigyo Branch of Kyoto City Sign Language Circle "Mimizuku", and Kamigyo Branch of Kyoto City Federation of Deaf, Wide Project, Digital Research Inc. and Japan Satellite Systems Corp.

7. REFERENCES

S Gulska (1990), The Development of a Visual Telephone for the Deaf: Using Transputers for Real-time Image Processing, *In Transputer Research and Applications 3. Proceedings of the Third North American Transputer Users Group*, pp.7-16

X Jun, Y Aoki and Z Zheng (1991), Development of CG System for Intelligent Communication of Sign Language Images between Japan and China, IEICE transactions, **74**, 12, pp.3959-3961

K Kamata (1993), Sign Language Conversation through Video-phone – Experiment at School for the Deaf --, Technical Report of IEICE, ET92-104, pp.37-44, Japanese

K Kanda (1996), Sign Linguistic from Basics, Fukumura Press, Japanese

A Komekawa (1998), Japanese-Sign Language Dictionary, Japan Federation of Deaf, Japanese

T Kuroda, K Sato and K Chihara (1995), System Configuration of 3D Visual Telecommunication in Sign Language, *In Proceedings of the 39th Annual Conference of the Institute of Systems, Control and Information Engineers, ISCIE*, pp.309-310, Japanese

T Kuroda, K Sato and K Chihara (1996a), S-TEL: A Telecommunication System for Sign Language, *In Conference Companion of First Asia Pacific Computer Human Interaction*, pp.83--91

T Kuroda, K Sato and K Chihara (1996b), Reconstruction of Signer's Actions in a VR Telecommunication System for Sign Language, *In Proceedings of International Conference on Virtual Systems and Multimedia VSMM'96 in Gifu*, pp.429-432

T Kuroda, K Sato and K Chihara (1997a), S-TEL: A Sign Language Telephone using Virtual Reality Technologies, *In CSUN's 12th Annual Conference Technology and Persons with Disabilities*, Floppy Proceedings, KURODA_T.TXT

T Kuroda, K Sato and K Chihara (1997b), S-TEL: VR-based Sign Language Telecommunication System, *In Abridged Proceedings of 7th International Conference on Human-Computer Interaction*, pp.1-4

M Ohki (1995), The Sign Language Telephone, TELECOM '95, pp.391--395

G Sperling, M Landy, Y Cohen and M Pavel (1985), Intelligible Encoding of ASL Image Sequence at Extremely Low Information Rates, Computer Vision, Graphics, and Image Processing, **31**, 3, pp.335-391

H Yasuda (1988), TV-phone Now, Spectrum, **1**, 5, pp.88-102, Japanese

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8