（千葉大学学位申請論文）

A Study of Non-linguistic Information and its Application for Learners and Interpreters

2008 年　1 月

千葉大学大学院自然科学研究科

情報科学専攻　認識情報科学

田中紗織

A Study of Non-linguistic Information in Japanese Sign Language and its Application for Assisting Learners and Interpreters

Saori Tanaka

2008

Contents

List of Tables

List of Figures

Acknowledgment

Support for this work was provided by a grant for science research from JSPS: the Japan Society for the Promotion of Science and by Grants-in-Aid for Scientific Research (B). My thanks go to my advisers Yasuharu Den, Akira Ichikawa, Harry van der Hulst for their helpful suggestions and important discussions; to Kuniaki Uehara, Kaoru Nakazono, Masafumi Nishida, Yasuo Horiuchi, and Yuji Nagashima along with their students for their feedback and suggestions regarding engineering logistics involved in the experiments in this work; to Sun Tsutiya for his important feedback regarding the conceptual framework for my thesis. I also would also like to thank all subjects who generously agreed to participate in my experiments. I am indebted to them for their patience and willingness to go through the complicated and repetitive tasks that were required of them in the experiments. Special thanks go to many people who gave me helpful feedback: Diane Lillo-Martin, William Snyder, Ronnie Wilbur, Carlos Gussenhoven, Christopher Miller, Onno Crasborn, Sandra Wood, Gaurav Mathur, Martha Tyrone, Soya Mori, Harumi Kimura, Yasuhiro Ichida, Yukiko Kawaguchi, Bert Kimura, Toshie Hatano, Nozomu Hirayama, Manabi Miyagi, Tomoko Ohsuga, Angela Quinlan and Ursula Bellugi. I also want to thank my partner, Yosuke Matsusaka for his support and being my colabolator to the work on which Chapter 4 is based.

Abstract

The aim of this study is to elucidate communicative functions of Japanese Sign Language (JSL) and to apply the results of this elucidation to the assistance of JSL learners and interpreters. This thesis focuses on the following aspects of non-linguistic information conveyed by the communicative use of JSL: (1) production mechanism of prominence, (2) prosody for skill evaluation, (3) learning effect of segmentation, and (4) non-verbal information for online sign interpretation.

Since non-linguistic information in JSL is produced by a set of sequence of movement of body parts, it is necessary to establish the basic unit of movement by which we can measure the relevant physical properties. Chapter 2 reviews the preceding studies that deal the issue of segmentation of a stream of human movements.

In chapter 3, we examine the physical correlate of acoustic prominence in emphasized signs, with a view to providing "phonetics" of JSL for learners and interpreters. The results obtained show that it is not only the emphasized lexical movements but also the transitional movements preceding the lexical movements that change the value of non-linguistic parameters such as duration, speed and distance.

In chapter 4, we verify the most effective timing unit to understand JSL words by an experiment on a computer assisted sign language learning (CASLL)

system. The result indicates that focusing on the longer unit including the preceding transitional movements has an effect on second-language learning.

In chapter 5, we study the relation between the prosody and the meaning of JSL sentences, since sign language sometimes carries prosodic information such as facial expressions, pause and nodding. We also examine how the required skill for interpreters is related to the skill for understanding the meaning of prosody. The results indicate that the skill for recognizing prosody can be useful when estimating how experienced the interpreter is in interpreting JSL into spoken Japanese and in working expeditiously on the interpretation.

In chapter 6, we characterize the function of non-verbal information in the interpreter-mediated conversations by using videophone equipments, in order to evaluate the effectiveness of facial expressions and nodding for hearing people. The preferred configurations had the common characteristics that the hearing subject could see the face of her/his primary conversation partner, that is, the deaf subject.

Based on these experimental analyses, we conclude that non-linguistic information has a significant function in facilitating communication in JSL, and argue for the effectiveness of non-linguistic information in JSL for assisting JSL learners and interpreters.

和文抄訳

本研究では，日本手話のコミュニケーション機能を明らかにし，手話学習者と手話通訳者に対する支援に応用することを目的とする．本論文では，日本手話のコミュニケーション機能に関する非言語学的情報のうち，(1) プロミネンスの産出メカニズム，(2) 分節化による学習効果，(3) プロソディによるスキル測定，(4) オンライン手話通訳サービスにおけるノン・バーバル情報に注目する．

日本手話の非言語学的情報は手話動作から産出される連続的な信号であるため，対応する物理量を測定するには，動きの基本単位を定義することが必要である．このためまず第 2 章では，人間の連続的な動作を分節するという問題を扱っている先行研究について論じる．

次に各章で (1) から (4) までの非言語学的情報の分析とヒューマンインタフェース上での効果の検証を行う．

第 3 章では，日本手話の「音声学的」知識を学習者と通訳者に提供するために，強調された日本手話動作で物理パラメータ（プロミネンス）がどのように変化するかについて明らかにする (1).

第 4 章では，コンピュータによる手話学習システム上で，手話単語を認識するための最も有効な時間的単位について検証する (2).

第 5 章では，顔の表情やポーズ，うなずきなど手話のプロソディと手話文の意味の関係について調査する．これに加えて，通訳者に必要なスキルがプロソディを認識するスキルにどう関与しているのかについて検証する（3）.

第 6 章では，聴者にとっても，顔の表情やうなずきなどの情報が有益である

ことを示すために，オンライン通訳通信機器において，通訳者・ろう者・聴者の3者の画面構成を変化させながら，通訳内容の質の評価を行う (4).

これらの実験的分析を通して，手話の非言語学的情報がコミュニケーションを円滑にする機能があることを示し，手話学習者や手話通訳者を補助するために有効であることを実証する.

Chapter 1

Introduction

1.1 Background

Scientific research can be roughly divided into two types of research: the basic and the applied. Focusing on the substance of phenomena and establishing a systematic theory belongs to the former group. On the other hand, the latter group consists of discovering a problem in daily life and exploring a solution with some kind of system. In sign language research in Japan, there are also these two types of research. One is based on some theoretical research that has developed, especially in the United States, and applying this methodology to the analysis of Japanese Sign Language (JSL) data. The other is developing a system by using engineering methodology to recognize or generate JSL for making an automatic translation system between JSL and spoken Japanese.

The "third type" of research is concerned with connecting the basic and the applied research and human interface development can be categorized in this type. In the research field of engineering, human interface (HI) is defined as a contact point between the human and the computer. Although the main field of the HI development is engineering, the research for modeling HI are highly interdisciplinary: cognitive science, psychology, body mechanics, linguistics, and so on. One of the reasons why so many kinds of researchers are involved in the development for HI is that considering what user is supposed to use the interface is very important for the modeling. For example, if we want to develop a usable mouse for a computer, we need to know about the users of it; their age, their grasping power, their sex, and so on.

This study uses the methodology of HI development and especially focuses on the issue "How we can develop an useful application for the learners and the interpreters in JSL".

In order to develop the more usable and communicative application for the learners and the interpreters in JSL, this study examines not only the symbolic information, such as hand shape and position in sign, that have been discussed in the field of sign language linguistics, but also the non-linguistic information, such as the quality of hands movement, the facial expressions, and the postures of body. Those dynamic properties work in a well structured way in the real-time conversations among the native JSL signers. For the second language learners of JSL, however, it is very difficult to acquire the rule, since they need to learn how they work not by textbooks but through their own eyes. It might be easier to imagine the case that a second language "speaker" faces difficulties to follow the real conversations among the native speakers, but the case for a second language "signers" is the same as well. This thesis reports the experimental results about the effect of the non-linguistic information for such second language signers.

1.2 Goal of this thesis

In order to develop more communicative application for the JSL learning and interpretation, this study discusses the basic and applied researches especially in (1) production mechanism of prominence, (2) prosody for skill evaluation, (3) effect of segmentation learning, and (4) non-verbal information for sign interpretation.

1.3 Structure of this thesis

This thesis consists of the following chapters.

Chapter 2: Review of Relevant Study

In order to study the non-linguistic information in JSL, it is first necessary to establish the basic units by which we can measure the relevant physical properties ergonomically. In this chapter, some studies are reviewed from the point of view of the segmentation from a stream of human movements.

Chapter 3: Production of Prominence in JSL

In this chapter, experimental results about prominence in JSL is discussed. We investigated which physical properties and which parts of a sign are distinctive cues for the production of prominence in Japanese sign language. The data in this experiment came from three native signers who produced three sentences, five times each in two modes: emphatic and non emphatic. From the experiment, it turned out that not only the emphasized lexical movements representing sign but also the transitional movements proceeding the lexical movement changed the value of dynamic parameters such as duration, speed and distance.

Chapter 4: Development of e-learning program for JSL learners

In this chapter, the Computer Assisted Sign Language Learning (CASLL) system is introduced. We propose a new learning program and compare it with the existing learing program implemented in the CASLL system. In the existing learning program, users learn sign words and then try to select the appropriate Japanese translations in a natural conversation expressed by two native signers. In the proposed program, users try to segment each word from a stream of signing by manipulating a control knob on the bottom of a movie screen, and then do the same tasks in the existing learning model. The end of the segmentation task is to know how continuous signs are articulated in the natural discourse. Ten Japanese learners participated in the experiments. Five subjects learned the existing word learning program and the other five subjects learned the proposed segmentation learning program. The mean accuracy rate of the proposed program was higher than that of the existing program. The result has indicated that focusing on transitional movements has an effect for learning JSL as a second-language.

Chapter 5: Evaluating Interpreter's Skill by Prosody in JSL

In this chapter, we developed a scale for evaluating ability to interpreter prosody in JSL, called MPR (Measurement of Prosody Recognition), in order to know whether interpreters can recognize the difference of meaning using only the prosodic difference. We then conducted two experiments: one is to study the relationship between the interpreter's experiences and the performance score on MPR (Experiment1), and the other is to investigate the specific skills that can be estimated by MPR (Experiment2). The data in Experiment1 came from four interpreters who had more than 1-year experience as interpreters, and from other four interpreters who had less than 1-year experience. The mean accuracy of MPR in the experienced group was higher than that in the less experienced group. The data in Experiment2 came from three high MPR interpreters and three low MPR interpreters. Two hearing subjects and three deaf subjects evaluated their skills in terms of the speech or sign interpretation skill, the expeditiousness, and the subjective sense of accomplishment for the ordering pizza task. The two experiments made it clear that MPR is useful to estimate how the interpreter is experienced enough to interpret sign language into spoken Japanese and to work on the interpretation expeditiously.

Chapter 6: Spatial Configurations for Online Sign Interpretation

This chapter discusses the design of configurations of videophone equipment aimed at online sign interpretation. Situations that occur in interpretation service were classified into three types: on-site interpretation, partial online interpretation, and full online interpretation. Simulation experiments of sign interpretation were performed and the qualities of the configurations were assessed. Although on-site interpretation is preferred by users, it was found that a thoughtful design of the configuration improved the availability of online interpretation. The preferred configurations had the common characteristics that the hearing subject could see the face of his/her primary conversation partner, that is, the deaf subject. The results imply that hearing people who do not understand sign language receive non-verbal signals that give additional information.

Finally this thesis ends with the discussion and conclusions based on these results.

Chapter 2

Review of Relevant Study

2.1 Basic research for segmentation of human movements

In order to study the non-linguistic information in JSL, it is first necessary to establish the basic units by which we can measure the relevant physical properties ergonomically. In this section, some studies are reviewed from the point of view of the segmentation of a stream of human movements.

Several studies in the fields of linguistics, engineering, and gesture recognition have dealt with the issue of the segmentation of a stream of human movements. In a study of sequentiality in American Sign Language (ASL), Liddell measured sign movements visually by counting videotape fields at 60 fields/sec [40]. The data was collected from a deaf woman of deaf parents who narrated two stories in ASL at a normal signing rate. The author defined Movements (M) and Holds (H) as the two basic units of a sign and counted such units on the basis of two indicators: image quality and hand location. According to Liddell's coding scheme, a blurry image indicates that the hand is moving, otherwise, it is holding. The hand's location at the reference point, which is identified from the sign's citation form, indicates that the hand is still, otherwise, it is moving.

Wilbur and Zelaznik used zero velocity to determine the onset and offset of

Figure 2.1: “smart”, “children”, “smart” in ASL

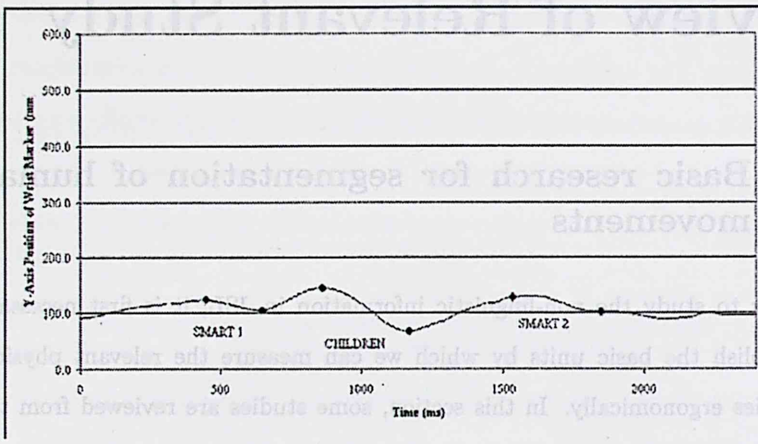

Figure 2.2: Place of Marker on the Wrist of Signer

each sign [5, 7]. They used an optical 3-D motion capture system (WATSMART) to study kinematic properties of prominence as it relates to phrasal position and emphasis. They used several cameras to detect the location of infrared light-emitting diodes on the thumb and index finger of the dominant hand for thirteen subjects in four relevant sign contexts. They concluded from their analysis that peak velocity is the best indicator of emphasis.

To measure undershoot of signs produced at different rates in ASL, Mauk determined sign boundaries on the basis of position data [9]. He used an opto-electronic motion capture system, Vicon 250 from Oxford Metrics. Subjects produced several sign utterances while wearing seven reflective spheres, which were used to track sign movements. A marker on the wrist was used to identify the location of the hand and delimit sign boundaries on the basis of movement along the x-, y- or z-axis. For example, the beginning of the sign SMART in ASL corresponds to a minimum along the x-axis, which is in the anteroposterior direction. The beginning of the sign CHILDREN corresponds to a minimum location of the wrist marker along the y-axis, which is in the vertical direction.

Studies of Japanese sign language (JSL) recognition in the field of engineering have also dealt with the issue of segmentation of the sign stream to detect its lexical components. Ohira et al. used data gloves with two subjects and measured changes in hand shape, hand speed and the direction of hand movement in 40 sentences to detect sign boundaries in JSL [13]. In their study, a speed envelope was also used for the detection of a " sign unit ", which is a repetition movement in the sign (See Figure 2). This method was 89 percent accurate in detecting sign boundaries. Akita et al. proposed a detection method for the lexical components of the sign stream which used angular jerk of human arm movement [23]. For their kinematic analysis of human volitional movement, they used the well-known minimum angular jerk model [4].

Akita et al. presumed that the transitional movement between lexical signs could be represented by the model of minimum jerk; and by extension, maximum jerk could be used to detect the lexical signs themselves. They used a magnetic 3-D tracking system (FASTRAK) to collect sign data from four signers who each wore two sensors on their right wrist and elbow. They tried to detect only lexical movements and not the transitional movements between lexical signs, and the false alarm rate for phrase-internal transitional movements by this method was 3.5 percent, which is a good rate for sign data.

Takata et al. analyzed candidate factors for segmentation in natural one-to-one dialogue between three Japanese deaf subjects and three hearing subjects [53]. They found several factors that were indicators of the boundaries of lexical

Table 2.1: Summary of Previous Study

	Liddell (1984)	Wilbur et al. (1997)	Mauk (2003)	Ohira et al (1995)	Takata et al (2000)	Akita et al (2003)	Uehara et al (2002)
Object for Analysis	Duration of H and M in ASL	ASL phrasal prominence	Undershoot at different signing rates in ASL	Segmen-tation for JSL recognition	Factors for segmentation of JSL	Segmen-tation for JSL recognition	Extraction method for primitive motion
Apparatus	Analog video camera	NDL Canada WATSMART®	Vicon, U.S.A., MOTION CAPTURE®	VPL research Inc, U.S.A., DataGlove®	Digital video cameras	Polhemus. U.S.A. FASTRAK®	MotionAnalysis, U.S.A., HiRes®
Sampling rate	60 Hz	120Hz	60Hz	30Hz	100Hz	120Hz	60Hz
Boundaries for segmentation	H or M	Zero velocity	Position minima or maxima	Minima of speed envelope	Head nodding, eye blinking, etc.	Minima of joint angular jerk	Speed minima
Basic unit	H and M in a word	A word between zero-velocity	A word between position minima or maxima	Signing unit between minima of speed envelope	A word between segmentation factors	A word between minima of jerk	Primitive motion between speed minima
Physical properties	Duration	Duration, peak velocity, size	Duration and mean location value				

signs: a nod, a blink, a hand position in neutral space, contact between the hand and face or contact between the two hands. In addition, they discovered that signers tend to produce transitional movements or just hold the hands still while these factors are occurring.

In their study of gesture recognition, Uehara and Shimada divided human motion data into segments at several breakpoints where velocity changes [22]. They employed this method to tag human motion in order to browse the content of multimedia data efficiently. To detect the time series of 3-D motion data, they used an optical motion capture system consisting of six infrared cameras with a sampling rate of 120Hz. A subject wearing eighteen markers performed some actions and the cameras recorded the subject ' s actions as video images. In that study, Uehara and Shimada defined the breakpoints on the basis of three states which comprise human motions: onset, transition and termination. Speed changes from zero to a positive value at onset, increases or decreases at transition, and changes from positive to zero at termination. They used onset and termination as boundaries of basic units, called "primitive motions". Table shows a summary of previous studies. Although each study has its own set of research questions, determining the boundaries of motion streams is a key issue for all of them. Two basic types of data capture equipment are used: videotape recorders and 3-D optical or magnetic motion tracking systems.

While videotape recorders can provide the duration of the basic unit [40] or the occurrence rate of possible sign boundaries [53], optical and magnetic motion tracking systems can provide much more information almost instantly from 3-D positional data with high temporal and spatial resolutions without interlacing of data samples. High temporal and spatial resolutions make the calculation of speed and acceleration accurate, and non-interlaced scanning makes each data sample clearer so that researchers can code the movements easily.

As segmentation methods for our study in JSL, we adapted both the magnetic motion tracking system (FASTRAK) and the videotape recorders. The former method will be described in Chapter 3, and the later one will be introduced in Chapter 5.

Chapter 3

Production of Prominence in JSL

3.1 Introduction

Consider the following two utterances: "I ordered···COFfee" - "I ordered coffee" with higher F0 and higher intensity on the first syllable of "coffee" in the first sentence, compared to "coffee" in the second sentence, and, in the first sentence, with a pause between "ordered" and "coffee". In the first example, the speaker has placed an emphasis on the words "coffee" without adding a lexical intensifier or changing word order. The manipulation of the physical properties of a string of speech sounds to achieve a relatively stronger acoustic impression is called "prominence". The full set of manipulable physical properties for prominence includes F0, intensity, and duration. Takeda[42] describes that the pause insertions preceding the word that has higher intensity, higher F0, and vowel lengthening can often be often observed in Japanese speech. Clearly, such manipulations of the sound structure around the prominent word further enhance the acoustic impression of this unit. There are two sides to prominence: the speaker's manipulations and the perceiver's acoustic impression.

In sign language we suppose that there are similar effects around a unit of sign that is realized with strong visual impression and we can apply some techniques of speech analysis to sign-language analysis. In order to study the physical properties of sign-language prominence, it is necessary to first establish a basic unit with which we can measure the relevant physical properties in a word; such as segment or syllable in spoken language. In sign articulators, hand, torso, head, face are the main productive media and the development of faster technology, as Tyrone[30] points out, now allows sign language researchers to study real-time articulations of these media with great accuracy[5, 36]. In the hand movements, the most movable media, several physical properties are considered as the distinctive cues of prominence. For example, the signer can give variety to the quality of the hand movements by space, time, fastness and quickness. In order to measure the quality of the movements, we can use distance, duration, speed, acceleration, and deceleration of a unit in a word.

The goal of this study is to investigate which set of physical properties is the distinctive cue for the production of prominence and which unit in a word is able to have the most distinctive prominence in a word of Japanese Sign Language (JSL). To this end, this study reports definition of several units in a word and categorizations of the physical properties on each unit by Factor Analysis.

3.2 Subjects

3 deaf people (all females) between the ages of 30 and 50 participated as subjects in this study. All subjects were right-handed with more than 20 years of sign language experience.

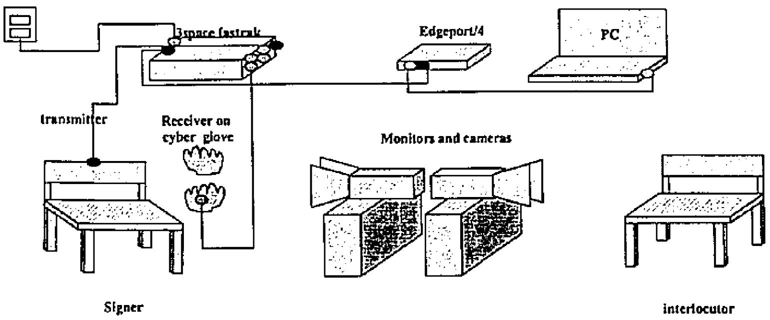

Figure 3.1: Lyout of equipment

3.3 Method

3.3.1 Apparatus

We collected position data by using Polhemus FASTRAK R. The sampling rate of the tracking system is 120Hz. We used 3 receivers, which were on the back of the hand, the wrist, and the head. The sampling rate for each receiver is 40Hz. In the current report, we only look at the data from the back of the right hand. A receiver on the back of the signer ' s right hand detected the position of the hand every 0.025 sec.

3.3.2 Data collecting

Each subject saw Japanese subtitles on a monitor, and after the subject read the sentence, an interlocutor appeared on the monitor and asked questions in Japanese Sign Language. The subject was to answer the questions by signing the sentence, repeating it 5 times. The subtitles were presented so that the signers would all produce the same sentences. We measured only the signing data produced by the 3 subjects. Japanese subtitles were represented in the order used for JSL.

3.3.3 Materials

To prepare the material, two native signers determined the order of words to be presented in subtitles. In order to make the signing during the recording sessions as natural as possible, we asked interlocutors (native signers) to ask questions to the subjects. At the first session, subjects produced 3 sentences 5 times each. The sentences, signs, and translations are as below. At the next session, to see the different effects of emphasized signing, an adjective was represented as bold and bigger text in each sentence, and signers expressed three sentences with emphasis.

(1-1-Q) Interlocutor

Stimulus: Onsen dou-datta?

Signs: ONSEN DOU?

Translation: hot-spring how ?

" How was the hot spring? "

(1-1-A) Subject

Stimulus: ano ohuro hiroi/ **hiroi** are.

Signs: ANO OHURO HIROI ARE

Translation: that bathtub huge it

" That bathtub is huge. "

(1-2-Q) Interlocutor

Stimulus: Shinkansen notta?

Signs: SHINKANSEN NORU OWARI?

Translation: Shinkansen take finish?

" Did you take a Shinkansen bullet train? "

(1-2-A) Subject

Stimulus: Ano densya hayai/ **hayai** are.

Signs: ANO DENSHA HAYAI ARE

Translation: That train fast it

" That train is fast. "

(1-3-Q) Interlocutor

Stimulus: Sono kaban motouka?

Signs: SONO KABAN MOTU?

Translation: That bag take?

" Shall I take that bag? "

(1-3-A) Subject

Stimulus: Ano kaban omoi/**omoi** are

Signs: ANO KABAN OMOI ARE

Translation: That bag heavy it

" That bag is heavy. "

3.3.4 Speed calculation

$$\Delta s(t_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3.1}$$

$$v(t_2) = \Delta s(t_2)/\Delta t \tag{3.2}$$

The formulas for calculating speed based on the position data are given in (3.1) and (3.2). Times t_1 and t_2 are two sequential points in time separated by $\Delta t = 0.025 sec$. Position at t_1 is (x_1, y_1, z_1); at t_2, (x_2, y_2, z_2) . $\Delta s(t_2)$ is the distance between positions at 1 t and at 2 t in three-dimensional space; $v(t_2)$ is the average speed between positions at t_1 and at t_2 .

Formulas (3.1) and (3.2) show the basic idea of calculating speed from position data; we smoothed the plot by applying a moving average that employed a three-point window of dimension x, y, z and then calculated speed. $\Delta s(t_{2,3,4})$ is the distance between the average of positions at $t_{1,2,3}$ and the average of position at $t_{2,3,4}$ in three dimensional space. See formulas (3.3) and (3.4).

$$\Delta s(t_{2,3,4}) = \sqrt{\frac{(x_4 - x_1)^2}{3} + \frac{(y_4 - y_1)^2}{3} + \frac{(z_4 - z_1)^2}{3}} \quad (3.3)$$

$$v(t_{2,3,4}) = \frac{\Delta s(t_{2,3,4})}{\Delta t} \quad (3.4)$$

3.3.5 Definition of phase types

When we saw the movie images while paying attention to the movement of the signer ' s hand, we could establish certain types in the sequence of movements. For example, in the movie image of sentence (1-1-A), (1-1-A) Stimulus: ano ohuro hiroi are.

Signs: ANO OHURO HIROI ARE

Translation: that bathtub huge it

" That bathtub is huge. "

There are 4 different signs in this sentence; ANO, OHURO, HIROI, and ARE. Once we recognized these signs in the movie file, we could see transitional movement between the 4 signs. Sometimes we could also observe that the signers hold the hand still just before the emphasized sign. We defined the label of phase types for each sign as Table.3.2. In this study, we only focused on adjective areas: HIROI " huge " , HAYAI '" fast " and OMOI " heavy " .

3.3.6 Delimitation of each phase type

We delimited the movement-type quantitatively by using speed(Fig.3.3). The speed is not always constant in an utterance. The receiver on the back of the signer ' s right hand is changing position with fluctuating high speed and low speed in three-dimensional space. However, the speed of hand movement is reduced and then increased at each point where a signer changes the direction of the hand movement in the process of producing a sign. The speed minima show up at the point in time where the signer changes the direction of the wrist,

label	τ	π	λ
phase type	transitional phase	Pause phase	Lexical phase
			HIROI, HAYAI, OMOI

Figure 3.2: Definition of phase types

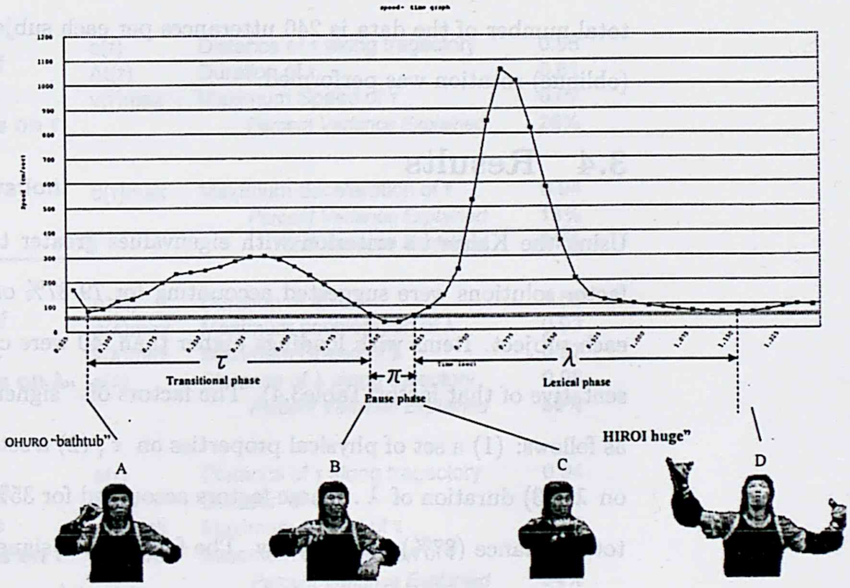

Figure 3.3: Delimitation of each phase type

elbow or shoulder. We paid attention to these speed minima in order to choose one minima as a boundary that divides τ and λ. By observing the movie image and speed graph, we checked which direction-change point coincides with each speed minimum, and then selected a speed minimum as the beginning position of λ. After we picked one of the minima as a boundary between τ and λ, we defined the threshold for calculating duration of π. We defined the threshold as 20cm/sec.

3.3.7 Factor Analysis

This study served as a pilot study to assess the production of prominence in emphasized-mode signing. A common factor analysis by principal component analysis was performed using SPSS for each subject. The original data for this study came from the 3 native signers who produced 3 sentences, 5 times each in 2 modes. We used 8 types of parameter for each subject (Table.3.4), and the total number of the data is 240 utterances per each subject. A non orthogonal (oblique) rotation was performed.

3.4 Results

Using the Kaiser ' s criterion with eigenvalues greater than 1.0, two or three factor solutions were suggested accounting for 79-87% of the total variance in each subject. Items with loadings higher than .40 were considered to be representative of that factor(Table3.4). The factors of " signer K " were interpreted as follows: (1) a set of physical properties on τ, (2) a set of physical properties on λ, (3) duration of λ. These factors accounted for 35%, 33% and 19% of the total variance (87%), respectively. The factors of " signer I " were interpreted as follows: (1) a set of physical properties on λ, (2) a set of physical properties on τ, (3) a deceleration of τ. The factors accounted for 40%, 26% and 13% of total variance (79%). The factors of " signer Y " were interpreted as follows:

Signers	Factors	Item Label	Parameter Types	Rotated loadings
K	1. A set of kinematic properties on τ	s(τ)	Distance of τ along trajectory	0.56
		Δt(τ)	Duration of τ	0.77
		d(τ)max	Maximum deceleration of τ	-0.92
		v(τ)max	Maximum Speed of τ	0.95
			Percent Variance Explained	**35%**
	2. A set of kinematic properties on λ	a(λ)max	Maximum acceleration of λ	0.73
		s(λ)	Distance of λ along trajectory	0.89
		v(λ)max	Maximum Speed of λ	0.92
			Percent Variance Explained	**33%**
	3. Duration of λ	Δt(λ)	Duration of λ	0.92
			Percent Variance Explained	**19%**
			Total Variance Explained	**87%**
I	1. A set of kinematic properties on λ	s(λ)	Distance of λ along trajectory	0.8
		Δt(λ)	Duration of λ	0.75
		v(λ)max	Maximum Speed of λ	0.75
		a(λ)max	Maximum acceleration of λ	0.74
			Percent Variance Explained	**40%**
	2. A set of kinematic properties on τ	s(τ)	Distance of τ along tragectory	0.98
		Δt(τ)	Duration of τ	0.81
		v(τ)max	Maximum Speed of τ	0.67
			Percent Variance Explained	**26%**
	3. Deceleration of τ	d(τ)max	Maximum deceleration of τ	0.94
			Percent Variance Explained	**13%**
			Total Variance Explained	**81%**
Y	1. A set of kinematic properties on λ	Δt(λ)	Duration of λ	0.86
		a(λ)max	Maximum acceleration of λ	0.93
		v(λ)max	Maximum Speed of λ	0.97
		s(λ)	Distance of λ along trajectory	0.98
			Percent Variance Explained	**50%**
	2. A set of kinematic properties on τ	s(τ)	Distance of τ along tragectory	0.94
		Δt(τ)	Duration of τ	0.71
		v(τ)max	Maximum Speed of τ	0.97
		d(τ)max	Maximum deceleration of τ	0.81
			Percent Variance Explained	**34%**
			Total Variance Explained	**84%**

Figure 3.4: Results of Factor Analysis

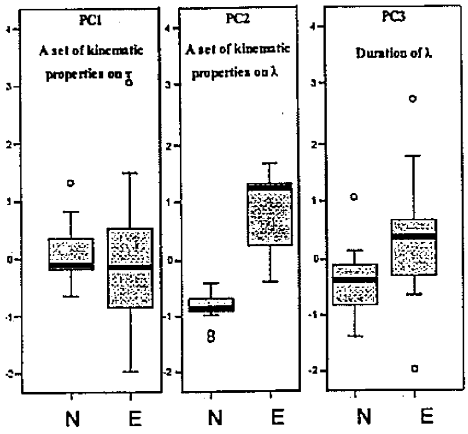

Figure 3.5: Results of signer K

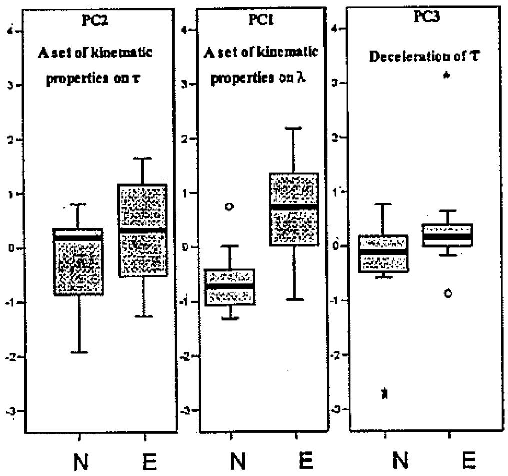

Figure 3.6: Results of Signer I

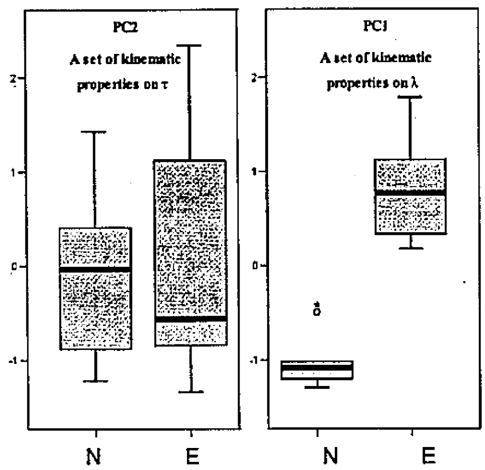

Figure 3.7: Results of Signer Y

(1) a set of physical properties on τ, (2) a set of physical properties on λ. The factors accounted for 50%, 34% of total variance (84%). The loadings for the items assigned to each of the three dimensions, listed in descending order, appear in Table 1. Each of the loadings is sufficiently high, indicating that there is a good fit of the items to the factors. Fig.3.5-3.7 show differences of some sets of physical properties between emphasized signing and non-emphasized signing. N means non-emphasized signing and E means emphasized signing. Each mode is categorized by each factor on each principal component axis. The box length is the interquartile range. Star marks indicate cases with values more than 3 box lengths from the upper or lower edge of the box. Circle marks indicate cases with values more than 1.5 and 3 box lengths from the upper or lower edge of the box.

3.4.1 Pause insertions and transitional phase

In the productions of "signer K", the set of physical properties on τ in emphasized signing had more variations than those in non-emphasized signing. We found 5 pause insertions after τ in emphasized signing, while there were 4 pause insertions after τ in non-emphasized signing. The average-duration of pause phases in emphasized signing was 0.39sec, while that in non-emphasis signing was 0.28sec. Likewise in the productions of signer Y, the set of physical properties on τ in emphasized signing had more variations than those in non-emphasized signing. We found 6 pause insertions after τ in emphasized signing, while there were 5 pause insertions after τ in non-emphasized signing. The average-duration of pause phases in emphasized signing was 0.66sec, while that in non-emphasis signing was 0.37sec. In signer K and Y, the pauses were often inserted in "HIROI" and "OMOI" with relatively small values of physical properties on τ in each signing mode. Although there was no pause insertion in the productions of signer I, the values of physical properties were relatively larger in "HAYAI" in all three signers.

3.5 Discussion

From the results of factor analysis in "signer K" and "signer I", we noticed that there were several levels of factor in a same phase, which could realize qualities of visual impression comprehensively. The results of "signer K" indicate that there are two kinds of prominence on the lexical phases. The set of acceleration, distance and speed was of primary prominence and duration had a secondary prominence. Primary prominence has more sharp difference than secondary prominence between emphasis mode signing and non-emphasis mode signing. From this fact, we can assume that the spatial extensions of the lexical phases with the faster and quicker movements could realize more distinctive visual impression than the durational extensions. The results of "signer I" indicate that there are two kinds of prominence on the transitional phases. The set of distance, duration and speed of transitional phase had a primary prominence and the deceleration of transitional phase had a secondary prominence. From the movie images of the transitional phases in "signer I", we noticed that the quality of movements and the forms of trajectory in the transitional movements were calm straight movements in non-emphasized mode signing, while they were quick arc movements in emphasized signing. The visual quality of arc movements, quickness, could be realized by the factor of higher deceleration as secondary prominence. From these observations, we will try to create prominence by animation synthesis and investigate more about the meaning of each type of prominence in further research.

3.6 Conclusions

In this section, we investigated the production of prominence based on delimitation of sequential hand-movements into 3 kinds of phases; transitional phase, pause phase and lexical phase. Although each type of prominence had different sets of physical properties in 3 signers, prominence on lexical phase was most distinctive through 3 subjects and 3 examples. We also observed the variations of physical properties on the transitional phases and the pause insertions between transitional phases and lexical phases. These facts indicate that there is a similarity between a syllable in spoken language and a lexical phase in sign language, because both units can be the most prominent place in a word. We also observed that there was both primary and secondary prominence on the lexical phases in "signer K" and on the transitional phases in "signer I" and those could characterize the visual impression comprehensively. These facts indicate that there are more varieties of the set of physical properties contributing to the production of prominence in JSL. The knowledge of prominence of JSL is expected to apply in the education field in JSL.

Chapter 4

Development of e-learning Program for JSL learners

4.1 Introduction

In order to present some contents of native sign language by using the technology of computer networks, a lot of research has been done all over the world in the field of educational engineering. In these works, there are two types for their aims; one is to support the deaf people themselves in order to fill the social gaps between the deaf people and the hearing people, and the other is to disseminate the knowledge of sign language in a society so that the circumstances around the deaf people are improved.

In the case of first group, the research of animation generation by natural language processing technology has been progressing in Greek Sign Language[12]. In Japan, based on the knowledge that the speed for playing JSL images depends on the level of proficiency for JSL, a system for playing JSL images in five speed level has been developed[15].

Meanwhile, in the case of the second type of research , a remote communications system to connect a class of American Sign Language to the terminal of students has been developed[38]. In Japan, the self learning system of finger

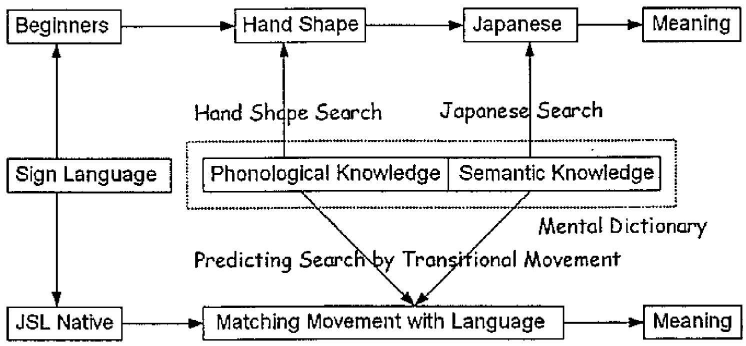

Figure 4.1: Recognition Model for Native Signers and Beginners

spelling with the function of feedback[21], the JSL database with search function based on the linguistic knowledge of native Japanese signers[48] and the teachware of JSL have been developed[14].

The aim of this study is to develop Computer Assisted Learning System (CASLL) and then evaluate the learning programs implemented in this system. Compared to other relevant studies, the distinctive aim for this study is to support the reading task of natural sign which is one of most difficult task for hearing learners.

In this chapter, we organize the reasons why the reading task is so difficult for the second-language learners while referring to the linguistic knowledge that has been revealed in recent years. After that, we evaluate the proposed learning method against the existing learning method.

4.2 Predicting model by transitional movement

One of the required skills for the learners of JSL is the need to learn the knack of reading each word that is expressed in the natural speed discourse. For this knack, this study focuses on the transitional movement which is traveling to the next lexical movement. We also present the "segmentation learning" by which learners study how to segment the continuous sign movement. It is based on the time point when the first lexical movement ends and moves to the next one.

"segment" is to delimit the continuous linguistic information based on a unit. It is called double articulateness that language can be segmented into morphemes and phonemes which are smaller units than morphemes[3]. The continuous sign movement can be segmented into morphemes, words and clauses. This study focuses on a unit, "the transitional movement + the lexical movement" and evaluates the effect of segmentation learning methods for the reading task.

The following data is the reason why we define the segment as "the transitional movement + the lexical movement". While the timing-points when the native signer recognized a word in "the transitional movement + the lexical movement" was about 15% of the whole unit, in the case of JSL interpreters, it was 26%. In the case of JSL beginners, they were not able to recognize it until 60% of the unit[52]. Furthermore, the earlier the deaf person acquires JSL the faster he/she can recognize a word in the recognition experiment[50]. Regarding this data, the closer the subject is to native, the more he/she is able to recognize the word at the first timing-point on the unit. We present the recognition model which can explain the difference of the timing-points between native signers and beginners (Fig.4.1).

The equivalent of phonemes in sign language consists of hand-shape, place and movement and those components structure a word4.1. For example, if a index finger goes straight up to one's temple, that movement represents "think" in JSL. In this case, the phoneme of "think" consists of hand-shape "1", place "temple", and movement "straight line". Those simultaneous comportments of phoneme structure the word "think" in JSL.

Describing the temporal sequence from the change of these comportments to be fixed as a word, (1) start to move from the end point of a proceeding word, (2) start to change the hand-shape, (3) fix the hand-shape (4) fix the place and (5) start to move in a word. There is an experimental report that the recognition of hand-shape and place precedes that of movement (Fig.4.1). From this knowledge, the native signer's time to recognize a word seems to be at stage (1), while a beginner's time seems to be after stage (2).

The delay in recognition time seems to be caused by the beginners cognitive tendency;they pay attention to the changing of hand-shape, and when it is fixed, they try to translate it into spoken Japanese(the upper process in fig.4.1). On the other hand, native signers seems to recognize a word directly by predicting the change of phoneme components(lower process in fig.4.1).

We implemented the segmentation learning on CASLL so that the beginners can detect the timing point when the signer starts to move his/her hands to the next word. We designed the program with which beginners learn from the more long-range movement sets as units, while they pay attention not to the starting point of changing hand-shape but to that of movement.

4.3 Design of segmentation learning

In this section, we will sort out the problems of the existing learning style compared to the characteristics of continuous sign in natural discourse.

In natural discourse by native signers, it is observed that hand-shape and position changes under the influence of those in the proceeding word, while it didn't occur when each word is carefully pronounced [49] [1]. It is also reported that hand-shape slightly changes its feature when two words form a phrase [51]. When a signer produces number "1" as a single word, he/she holds the

[1] In American Sign Language, it is also observed that the position of word is getting closer to the proceeding word position when the speed of articulation is increased[26]

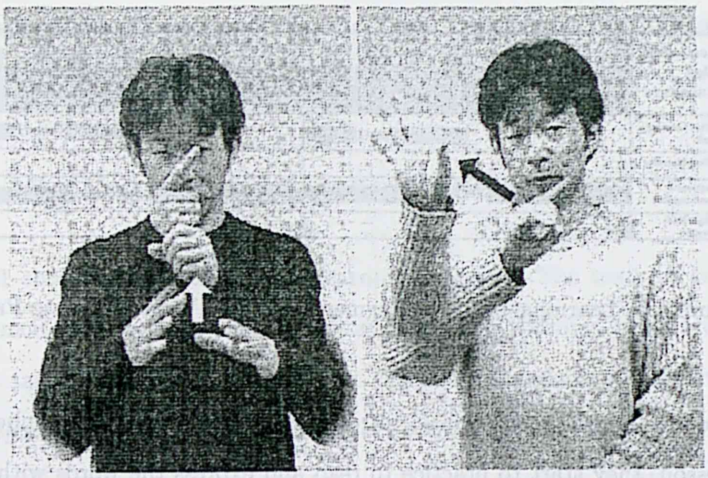

Figure 4.2: Examples for Handshape Assimilation
From Ichida(2005)p.16

other four fingers tight except for the index finger. However, when it forms a phrase with the word which five fingers open softly, the other four fingers except the index finger are closed lightly(see right picture of Fig.4.2). This fact is caused by the fact that the hand shape with five fingers open effects the anteroposterior word, and then the feature of "open" is added to the four fingers except the index finger. This phenomenon seems to be equivalent to the phoneme assimilation in spoken language, the term of "assimilation" is used in sign language research[51]. The phenomenon is related to efficient articulation and is often observed in natural speed discourse in native signers.

Meanwhile in the learning situation of beginners, there are many students who face difficulty reading words in natural discourse, even though they have already learned each word. It is not uncommon to see the number of students are decreasing by more than half in a class, as the levels of a course advances. One of the reasons for this problem seems to be the general learning method used in classes or groups of JSL. In some classes, students are often required to memorize

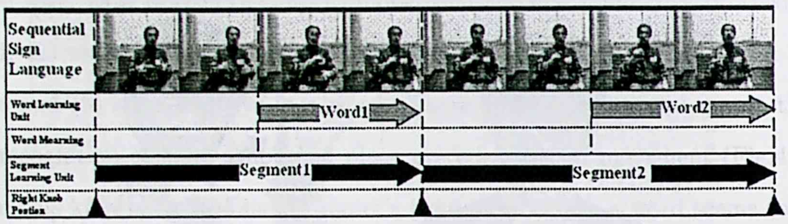

Figure 4.3: Learning Units for Segmentation and Word Learning Programs

all the vocabulary for greetings, introducing their family, and numbers. After the lesson, they start to practice to read and express the words with students in pairs. By such kinds of learning style, it is difficult to read sentences with many phoneme assimilations in natural discourse.

Based on these problems, we defined three points as necessary functions for the design of proposed learning program; (1) students can check the images as many times as they want to, (2) students can learn a word expressed carefully by comparing it to the same word expressed with other words continuously and (3) students can be aware of the segment boundaries including transitional movement.

We also implement the word-learning program that users mainly study JSL vocabulary on CASLL, and then evaluate how much each user can recognize the words in natural discourse. Based on the results, we verify the effectiveness of proposed program including the segment task before the vocabulary task.

Feg.4.3 shows the relations in each leaning program and continuous sign images in natural discourse. In the word learning that is more conventional in JSL education, users mainly focuses on the lexical movement without transitional movement in the segment: the short arrows in fig.4.3. Compared with this, we design the segmentation learning which makes users focus on the long arrow in fig.4.3. We expect that the users can study sign production including the natural assimilations between words, and it will improve their skill for reading as a result. We will describe the details in each learning program.

4.4 Two learning programs and system

Table.4.1 show the process in each learning program. The texts in boldface are common tasks in each one. In the segmentation learning, there are three different tasks before the common tasks. Those three tasks are comportments of the segmentation learning, the proposed method in this study.

In the word learning, users study only lexical movements without paying attention to transitional movements, and then memorize the Japanese translations. After that, users challenge the reading task to answer the meaning of word in each question. On the other hand, in the segmentation learning, the users count the stroke of movements in each example of natural discourse. This task is aimed at making them focus attention in each component of movement, and they should become used to the natural speed with many assimilations. After the counting task, they try to correct the order of word images which are randomly-aligned in the order of natural discourse. In the "word images", a signer produces each word separately. At the upper side of the word image, there is discourse image in which a native signer is signing (Fig.4.4). The re-realignment task is aimed at relating the words and the sequential signs. The users should be conscious of the difference between the words and the segments including transitional movement and lexical movement. After the re-realignment task, users perform the segment task in which they try to match the knob at each segment in the short piece of discourse image(Fig.4.5). We designed these three tasks so that users can be conscious of each segment including transitional movement. After these three tasks, they perform the tasks for the word learning: the word learning task, and the examination for checking how many word

Table 4.1: Learning Processes of Two Learning Programs

	Word Learning Program		Segmentation Learning Program
1	**Word Learning Task**	1	Counting Task
2	**Word Test**	2	Realignment Task
3	**Reading Task**	3	Segment Task
4	Realignment Task	4	Word Learning Task
5	Translation Test	5	Word Test
		6	Reading Task
		7	Translation Test

they correctly memorized. In the following section, we will explain about the detail of each learning program.

4.4.1 Outline of two learning programs

- **Word Learning Program**

1. Word learning task

   Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

2. Word test

   Users are required to select the correct Japanese translation for each which is randomly-presented word. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

3. Reading task

   Users are required to read each word presented in the image of natural discourse which length of about 5-10 seconds and select the meaning of it.

4. Realignment task

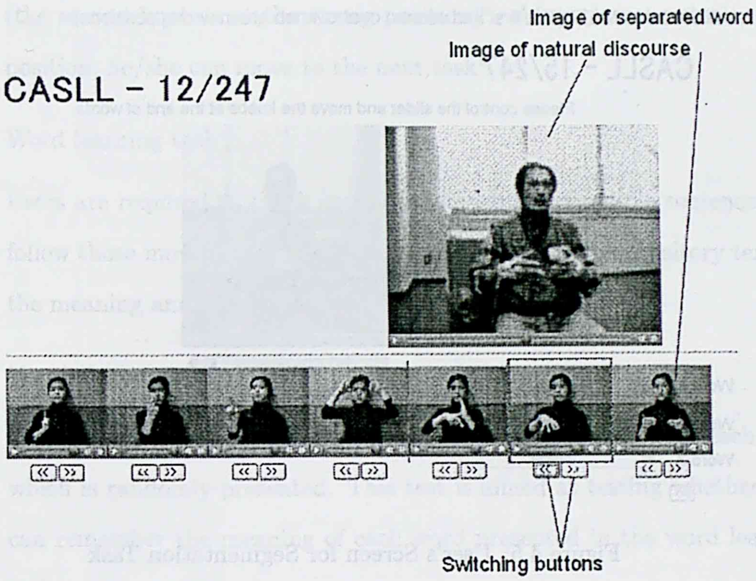

Figure 4.4: User's Screen for Reorder Task

Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. Each image has Japanese subtitles.

5. Translation test

   Users are required to select a correct translation that is appropriate for the example image by a native signer.

- **Segmentation Learning Program**

1. Counting task

   Users are required to count the stroke of movements in each example of natural discourse. They can select the number of stroke from choices.

2. Realignment task

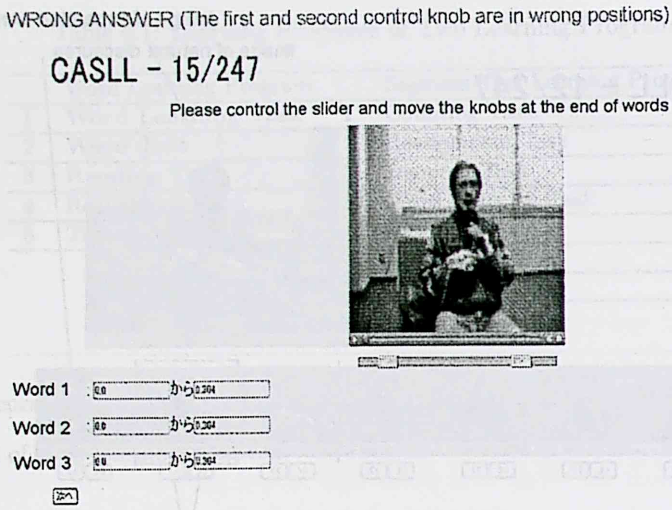

Figure 4.5: User's Screen for Segmentation Task

Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. In order to match the order of word images to the sentence in discourse image, they need to manipulate the buttons for switching the images without Japanese subtitles. At this point, they have not learned the meaning of each word, so they can only check how each stroke structures the word in natural discourse.

3. Segment task

Users are required to select the length of each segment in a short piece of image cut from natural discourse. As in Fig.4.5, they need to manipulate the knobs on the slider with referring to the discourse image, and set the position of knob at at the last frame when the hands start to change the shape and translate to the next word. If the user doesn't set it at the right position, the arrow message is shown up saying "you have wrong answer (the second knob is at the wrong position)". After the user corrects the position, he/she can move to the next task[2].

4. Word learning task

   Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

5. Vocabulary test

   Users are required to select the correct Japanese translation for each word which is randomly-presented. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

6. Reading task

   Users are required to read each word presented in the image of natural discourse that length is about 5-10 seconds and select the meaning of it.

7. Translation test

   Users are required to select a correct translation that is appropriate for the example image by a native signer.

4.4.2 System overview

Fig.4.6 shows the system overview of CASLL. Once the scripts with the procedure of two learning programs has been sent to the server, the server finishes the preparation for judging the answers; generation of hints; and generation of the forms of buttons, knobs, and images. Once the user finished each task with displaying images from the server, each answer will sent to the server. The server will display hints according to the user's answer, and present the form of

[2]The correct answer was set with the 2 frames margin in the vicinity.

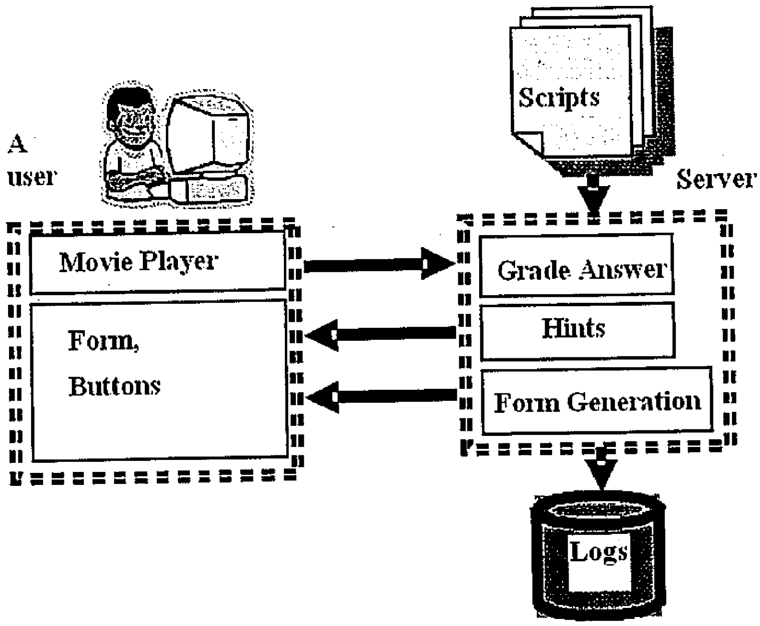

Figure 4.6: System Overview

next task when the user's answer is correct. The hints include the information about the number of correct answers per question, and the wrong position number of the knobs in the segment task. The user's record will be stored in the server in order to evaluate for the next analysis. We designed CASLL based on the assumption that users have Internet Explore 6.0+ (Microsoft), QuickTime 7+ (Apple), and ADSL or cable or large-scale network at universities. From the experiments that we have done before, the users can perform the tasks in CASLL with these environments.

4.5 Evaluation experiment

4.5.1 Procedure

The aim of this experiment is to present a self-learning program by which a user can be thoughtful about hihe/sher choice of tasks. For this aim, we compare the scores of two subject groups; the word learning group and the segment learning group. We evaluate the effect of the segment learning for the reading tasks in CASLL. The users can see the question images as much as they want and perform the tasks until they reach each correct answer. In this situation, if the learning effects of two learning programs for the reading task were not so different, we could not say that there was a special learning effect in the segment learning program. If so, the effect of the reading task would be only the result of the word learning. Therefore, for this experiments, we will analyze the proficiencies of vocabulary in two programs. If we could confirm that the proficiencies are high enough in both two programs, we would use it as a baseline and then compare the scores of reading task in both programs. For this comparison, we would use the percentage of number of questions that the users get correct at the first trial, and the number of trials for reaching each correct answer.

4.5.2 Subjects

The subjects were 10 beginners who have less than two years learning experience at some JSL learning clubs. In these subjects, 5 beginners performed the word learning program and the other 5 beginners performed the segment learning program. The range of their age was from 20 to 38 years old, and the male-female ratio was 3:2 respectively.

4.5.3 Learning program

Although the learning program at the server can be restructured by writing the script, we used some images cut from a image of natural discourse. The length of each image was about 7 seconds. The total number of images were 6, and those included 53 basic words that were used in a natural discourse. The length of whole image was 49 seconds and the average speed of signing was 1.54 [word/sec]. That speed was quite fast for the users based on the results of Questionnaire that all users answered after the experiment. For those 6 images, we made the script for the two learning programs in Table.4.1, and implemented the tasks by which each user perform the 6 trials. We also added a function by which the users could answer each question 5 times and after 5 times, they could select "give up" and go next task.

4.6 Result and discussion

4.6.1 Level of proficiency for vocabulary

$$\text{(the number of question - the number of wrong answer)} \div \text{the number of question} \times 100 \tag{4.1}$$

From the expression (4.1), we calculated the proficiencies of 53 words in both learning programs. These were 98 in both learning programs that were close to perfect. As a result, it was confirmed that the proficiencies of the vocabulary in both learning groups were similarly high.

4.6.2 Analysis for score in reading task

In order to calculate the score of the reading task, we substituted the number of reading questions and the number of wrong answer in the reading task to (4.1). As a result, the average in the word learning group was 67.0 and that in the segment learning group was 54.8. The former average was significantly higher

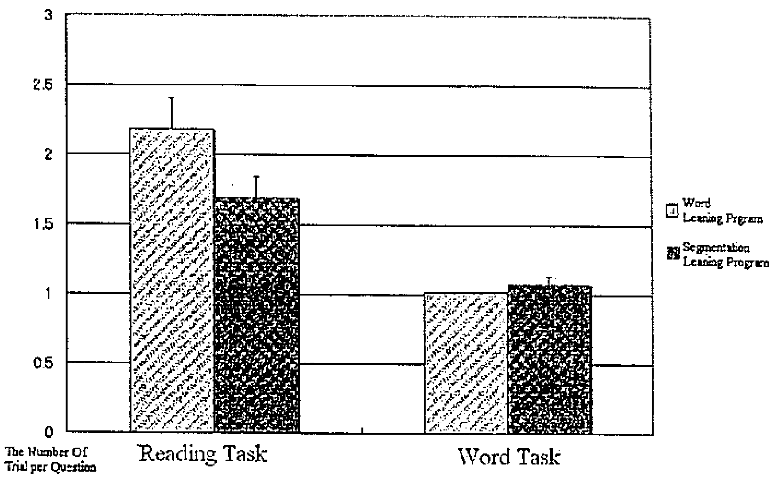

Figure 4.7: Mean Number of Trials for Reading Task and Word Task

than the later one (t(8)=1.97,p<.05).

4.6.3 Analysis for number of trial

Although this system allows the users to skip the question with the giving up function after the fifth trials, they can also answer the questions as many as they want without the giving up function. Therefore, at least in principle, they can continue the trials to answer a question until they reach the correct answer. If he/she answers correctly at the first trial, he/she can have higher score, so we analyzed the number of trials they had in both the reading task and the word task. If the segmentation learning had more effect for the reading task than the word learning, the average number of trials of segmentation users would be less than that of the word users. The result was shown in Fig.4.7.

In Fig.4.7, the two bars on the left side shows the result of reading task. The mean trial number of reading task in the segmentation users was less than that in the word users. This means that the segmentation users reached correct answers with less trials than those by the word users. The mean trial numbers of reading task were 2.18 in the word learning group, and 1.69 in the segmentation learning group; SD were 0.49 and 0.32 respectively (t(8)1.84, p=.051). The less trial numbers needed by the segmentation users seems to reflect the higher score in the reading task.

In Fig.4.7, the two bars on the left side show the result of word task. Both groups of users reached the correct answers mostly at each first trial. The mean trial numbers of word task were 1.02 in the word learning group, and 1.07 in the segmentation learning group; SD were 0 and 0.11 respectively. From this result, we found that all user in both learning groups answered almost perfectly for the random questions in the word task.

The difference of learning program in both learning groups were the following; the segmentation users had the counting task, the realignment task, the segment task, the word learning task, the vocabulary test and then the reading task. The first three tasks are the original component of the segmentation learning. On the other hand, the word users had the word learning task, the vocabulary test and then the reading task. The word users seems to find it difficult to read the meaning of each word in natural discourse immediately after the word learning task and the vocabulary test. Additionally, the word users had a perfect vocabulary skill in 53 words, so even if they had more trials for the word learning task, the score of the reading task would be same. Compared with them, the segmentation users saw the natural discourse image over and over again, and had the counting task, the realignment task and the segment task, so the users seems to find it easy to recognize the relation between movement and meaning in the early stages. From the fact that the levels of proficiency for vocabulary are highly enough in both learning programs, the segmentation learning showed the effectiveness in reading the signing words in

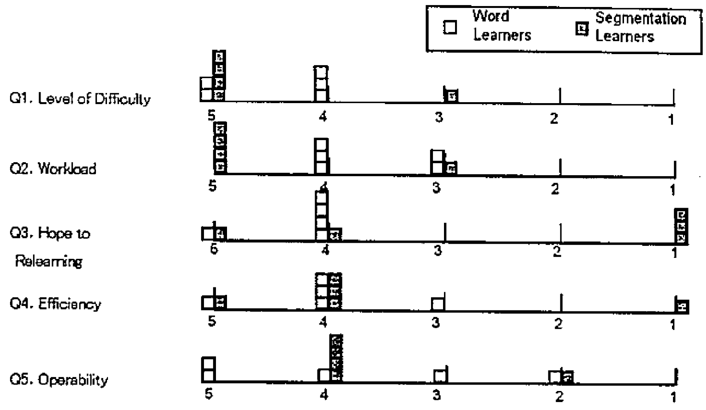

Figure 4.8: Results of Questionnaire

natural discourse.

4.6.4 Questionnaire

From the above analyses, we found the effectiveness of the segmentation learning, but there might be room for improvement in the program which was a first edition. We studied the possible way for improvement from the analysis of questionnaires. Fig.4.8 shows the result of questionnaires that all user answered after their learning programs. They were asked to rate their impression based on five questions from Q.1 to Q.2 on the grade of five scale each. Q.1 was "Were the questions difficult for you?" and almost all user rated more than 4 (difficult). It indicated that the challenge levels were high for all user in both programs. Q.2 was "How did you feel about your workload?". For this question, two word users rated 3 (right quantity) and three rated 4 (it was heavy). On the other hand, one segmentation users rated 3 and four rated 5(it was too heavy). The extra duration of the segmentation leaning seems to make the users feel workload to be heavier. Q.3 was "Do you want to learn by this same learning program again?". Four word users rated 4 (yes, when I have time), and one rated 5 (definitely yes), while three segmentation users rated 1 (no, never), one rated 4 (yes, when I have time) and one rated 5(definitely yes). The evaluation result varied greatly in the segmentation learning. Q.4 was "How was the efficiency of this program compared to other learning methods by which you had learned before". One segmentation user rated 1 (highly inefficient) but more than half of users answered "efficient" and "highly efficient". Q.5 was "How was the operability of your computer for this program ?", and more than half users rated "good" and "very handy". The user who answered "a little bit difficult to operate" described her impressions that her computer screen was too small to see all word image when she attempted the realignment task. From the above results of operability, we found that the problems for the interface would be improved easily if we changed it to the adjustable display system with user's screen size. Additionally, we found some room for improvement in the curriculum of the segmentation learning. For the purpose of comparative experiments, we needed to design the segmentation learning program with a number of questions, but the evaluation of Q.3 seems to be improved if we control the number of trials that the segmentation users study at once. Furthermore, presenting a kinder explanation of each of task and meaning at the beginning of the program would be help to reduce the user's mental burden for the number of questions.

4.7 Conclusion

In this chapter, we reported the results of an experiment comparing two learning programs. One is the segmentation learning program, in which we implemented the segment task to recognize the movement units including each transitional movement and the word learning task. After the evaluation experiments for this program, we analyzed their scores of reading task. The other is the word learning program, and we analyzed their scores of reading task after the vocabulary test without the segmentation task.

As the results, the level of proficiency for vocabulary were high enough in both programs, and the mean score of reading task in the segmentation users were significantly higher than that in the word users. The segmentation learning program is different from the word learning program only because the former including the segment task to recognize the unit of "the transitional movement + the lexical movement".

The results of this comparative experiment indicated that the scores of reading task for natural discourse will be increased by making the beginners focus on the continuous movement pattern of which they cannot be conscious with the existing word learning method.

In order to set the segmentation learning program as the main program for CASLL, we need to modify the design of segmentation learning program based on the subject's evaluations. Since it is difficult for the beginners to match some knobs on the segment point of "the transitional movement + lexical movement", the first task now is to contrive a way to present an instruction with some hints in order to lead the users to the right answer. We will reorganize the curriculum to increase the user's willingness to study and evaluate it again while relying on the operability of CASLL.

We are also going to verify the function for the segments of non-manual signal(NMS) which are larger than "the transitional movement + the lexical movement". In JSL, NMS has a function as a syntactic marker to distinguish different particles and also as a morpheme maker to distinguish between adjectives and adverbs. For example, a signer can distinguish between a genitive case and a nominative case by changing the timing of nodding, and he/she can distinguish between an adjective and an adverbs by facial expressions. We assume that the segment by NMS is related to recognize the meaning of sentence as a whole rather than the selection of a word. We are trying to implement a function to learn NMS markers and then test the effectiveness of it. Although this is our first step for the development of a leaning program, it is expected that we can present the segmentation learning program as a content of e-learning and also we can develop the program by ourselves, while relying on the operability of CASLL and improving the efficiency of programs. Additionally, we expected that CASLL could help to organize the information of JSL and apply it to some other softwares and machine interfaces.

Chapter 5

Evaluating Interpreter's Skill by Prosody in JSL

5.1 Introduction

The demand of interpreters between Japanese Sign Language(JSL) and spoken Japanese has increased these days, but the methodologies to train such interpreters are still limited as compared to the other types of interpreters between spoken languages. Technically, some studies for animation synthesis have been performed in order to propose alternative tools for interpreters[19, 39]. However these technologies are only applicable in restricted situations.

In order to consider simultaneous JSL interpreters who work across a broad range of situations, some other types of technique are needed. As one of such technique, this study focuses on a quantitative indicator for evaluation of the general skills of sign-language interpreters in order to train them according to their level of skill.

The required skills for JSL interpreters include the expression skill, the translation skill and the practical skill (5.1). Usually those skills are combined qualitatively and it is difficult to quantify each skill individually. In order to certificate the official license for JSL interpreters, therefore, their skills have been

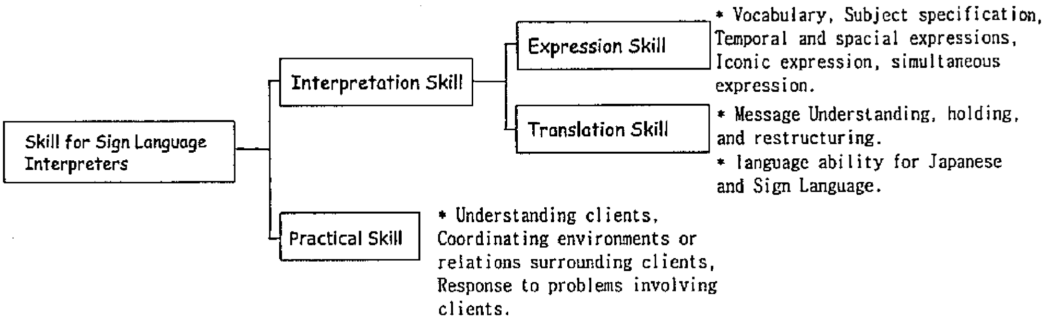

Figure 5.1: Required Skill for JSL Interpreters. The figure was translated from Japanese figure in [1].

subjectively evaluated by well-experienced instructors.

For people who wanted to hire JSL interpreters, it is very difficult to know the level of their candidates and how long they need to train them for reaching the necessary level to conduct their service autonomously. Although some methodologies of kinematic modeling for well-experienced interpreters might be useful to elucidate the skill for the expression skill in JSL[24, 46], the expression skill in JSL is just a component of overall skills and it seems still difficult to evaluate other necessary skills for actual interpretation services. Some researchers have shown that the highly skilled interpreters have distinctive practical skills, with which they coordinate the order of utterance between a deaf and a hearing clients. They also ask questions frequently in order to interpret each utterance more precisely[10, 27, 29]. Therefore, the quantitative indicator that is useful for recruiters of JSL interpreters, needs to be very simple to use, and applicable for evaluating the general skills of interpreters including both their practical and interpretation skills.

This study explores the possibility that the skill for recognizing prosody can be an indicator for other general skills of interpreters.

In the interpretation between JSL and spoken Japanese, one of the most difficult tasks is to recognize the case relation or the modification relation in

JSL. For representing the case and modification relations, spoken Japanese has particles and inflected forms of a word, respectively. In JSL, however, there are no such changes. Instead, there are durational changes between two signs, facial changes, and the insertion of pauses to distinguish the case and modification relations.

This para-linguistic information is called prosody, and it is closely related to the syntactic structure of utterance. [45] describes a correlation between rhythmic pattern and syntactic structure based on a linguistic theoretical analysis called prosodic phonology[28]. It has been also discovered that paralinguistic information plays an important role in achieving real-time and easy understanding and smoothly controlling conversation[2, 31]. Because prosody is conveyed by communicative use of JSL, we assume that the skill for recognizing prosody is related not only to the expression skill but also some other skills. These other skills include the interpretation skill and the practical skill of coordinating the conversation between the deaf and the hearing clients.

In this study, we introduce the unique test called MPR (Measurement of Prosody Recognition) to measure the skill for recognizing prosody, and study the relationship between the interpreter's experiences and the performance score on the test (Experiment-1). We also investigate the types of skill that can be estimated by MPR (Experiment-2).

5.2 Data preparation for MPR

The ingredients of prosody in spoken language are rhythm, prominence, and intonation. These prosodic cues aid in the syntactic processing of sentences, in disambiguation, in focusing particular elements, and in indicating the discourse function of utterances[47]. Linguistic researches in the area of sign language has also been progressive in describing the prosodic correlation of rhythmic

Table 5.1: Example Sentences for Nominative and Genitive cases

Nominative/Genitive	Noun	Predicate
SISTER	hat	like
	earing	
	glasses	
	pin	
	lip rouge	
	bag	
	clothes	
	car	
	book	
	shoes	

Table 5.2: Example sentences for Adverb and Adjective Expressions

Adjective/Adverb	Noun	Predicate
REAL	sister	cheerful
		stubborn
		genial
		smart
		mature
		austere
		serious
		amusing
		sedate
		gentle

pattern and body movement[36]; semantic focus such as eyebrow raising or chin drawing[6]; adjective emphasis as increasing speed, distance or duration [43, 44]

In order to establish a scale of the skill for recognizing prosody and its meanings, it is needed to prepare stimula by which subjects can distinguish between correct and incorrect answers using only the prosodic pattern. In most cases in JSL, however, if the meaning of sentences are different, the signer changes the word order to suit the meaning. Therefore, we decided to prepare two types of sentences in which the word-order is natural enough to understand, and the pattern of prosody for representing each meaning is different. In the following section, we will introduce how we prepared the data for MPR.

5.2.1 Subject

The subject was forty years old female, who was born deaf and had learned JSL since she was at a preschool for deaf children. She works as an interpreter between JSL and American Sign Language and has sufficient knowledge for the production of JSL.

5.2.2 Video recording of test data

We asked the signer to produce three words including "sister", "noun X" and "like" in this order, and to express the sentence in two different meanings as "sister likes X" and "I like sister's X" with the same order of the words (see Table1).

We also prepared the sentences including "real", "sister" and "adverb Y" and asked the signer to produce it in two different meanings as "the sister is really Y" and "the real sister is Y" (see Table2).

In order to minimalize the effect of spatial distance between the contiguous words, we used five words with short inter-word distance and another five words with a long inter-word distance. We recorded the production images by a video

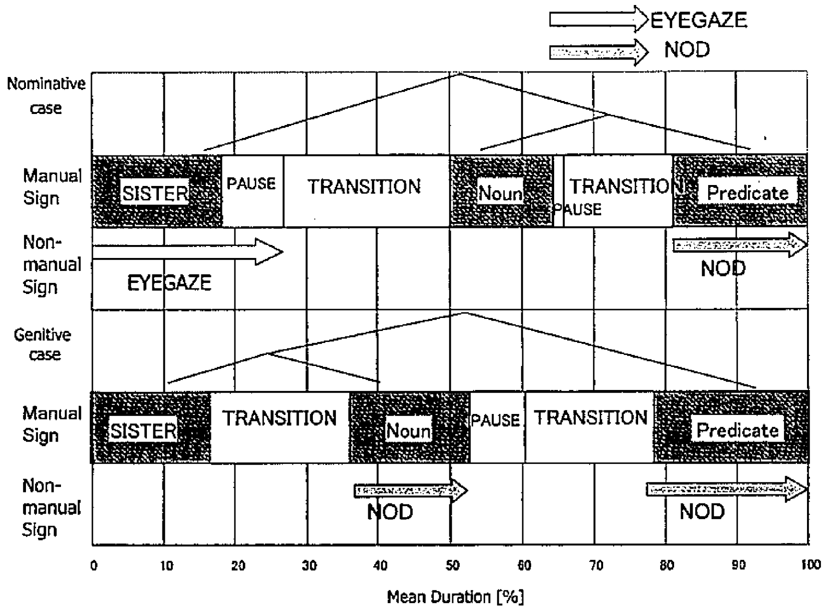

Figure 5.2: Tree Structure of the duration in Nominative and Genitive Cases: A significant difference was shown between the first transitional movements($t(9)=6.10$, $p<.05$) and between the pauses after the noun($t(9)=6.10$, $p<.05$)

camera. The video frequency was 30Hz, and we converted the images to avi for the analysis.

5.2.3 Data analysis

The duration of each movement unit(MU) was measured using the sign language analysis tool, MAT[11]. We classified MU into three types in terms of the lexical movement, the transitional movement and the pause. The lexical movement is a hand movement which has a meaning as a sign. The transitional movement is a hand movement which is inserted between two signs. The pause is the duration in which the hand is visually at the same position.

We counted each movement as the frames of the movement representing a

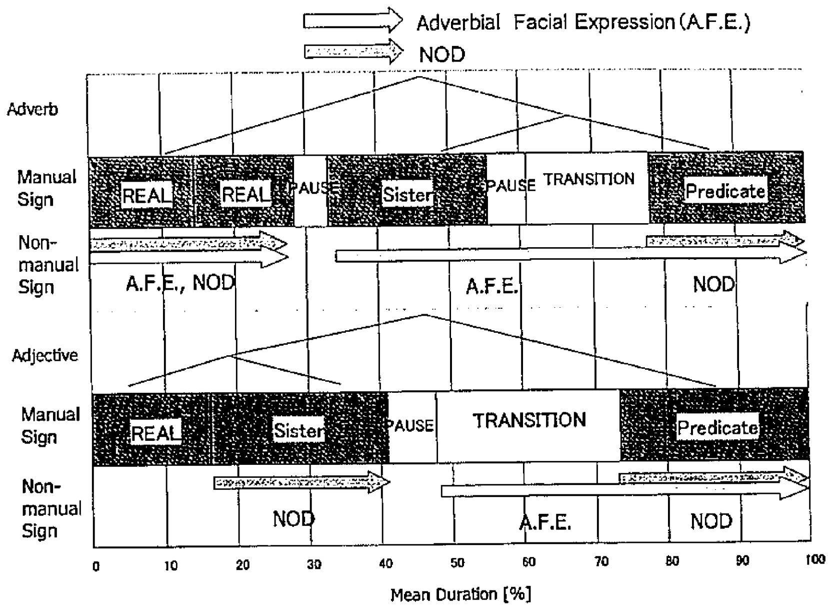

Figure 5.3: Tree Structure of the duration in Adverb and Adjective Expressions: The signer used adverbial facial expressions which are applicable for "really" but there were no such expressions for "real". Other prosody including pauses were placed differently in the two expressions.

word, those of movement traveling from one word to the next word and those of movement in which hand-shape and orientation were respectively stable at the same position. All frames were normalized by dividing by the entire duration of each sentence.

5.2.2 shows the results of analysis for the nominative and genitive cases. From the results of the t-test, it was shown that the first transitional movements and the pauses of the nominative cases were significantly longer than those of the genitive cases ($t(9)=6.10$, $p<.05$, respectively). For the nominative case, the signer also used eye gaze and placed the pause with the last word . For the genitive case, she placed the pause with the second and the last words. The results showed that the signer was changing the timing of producing each word with a variety of prosody to match the meaning.

5.2.2 shows the results of analysis for adverb and adjective expressions. The signer used adverbial facial expressions which are applicable for "really" but there was no such expression for "real". Other prosody including pauses were placed differently in the two expressions.

By using the data, we made up "MPR" and explored whether the length of experience for interpreters effects the score or not. We describe the procedure of the experiment-1 in the following section.

5.3 Experiment1

The aim of the Experiment-1 is to see whether the performance score of MPR is related to the length of interpreter's experience.

5.3.1 Subject

We arrange the subjects into two groups; one consisting of four interpreters who have more than 1-year experience (long-experience group), and the other consists of four interpreters who have less than 1-year experience (short-experience group). The reason for classifying the subjects based on the period of 1-year experience was due to the author's experience that it took at least one year to express and understand the pattern of prosody precisely after acquiring a skill to have a conversation with native signers without an interpreter.

5.3.2 Method

MPR is an examination method in which the subject watches a movie and then writes each answer on his/her answer-sheet. It includes two kinds of task. Firstly, there is meaning task in which the subject marks a right answer that is more appropriate for the sentence image. Secondly, there is the segment task in which subject writes a slash " / ", where he/she feels the three words should

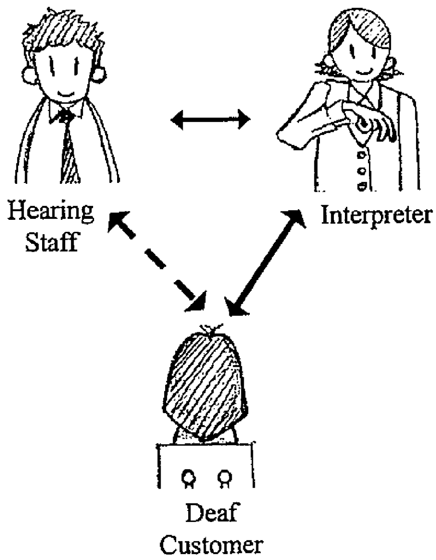

Figure 5.4: Configuration of the three parties. The media between the interpreter and the hearing staff is speech, while that between the interpreter and the deaf customer is sign language. The hearing staff and the deaf customer converse via interpreter mediation.

be segmented.

In each task, there are two types of question: type-1 is to recognize the difference of nominative and genitive, and type-2 is to recognize that of adverb and adjective. Both types contain twenty questions and each one has two kinds of task (the meaning task and the segment task). The total number of items is therefore eighty. The accuracy of each task will be calculated as the number of correct answers divided by twenty.

For this Experiment-1, we asked each subject to have a seat in a quite room and look at the monitor on a personal computer. The size of movie on the monitor was 780*420 pixels, which was 16.5*11 cm. We gave an instruction to each subject to mark his/her answer on the answer-sheet by looking at the movies which were played one after another in 2 seconds.

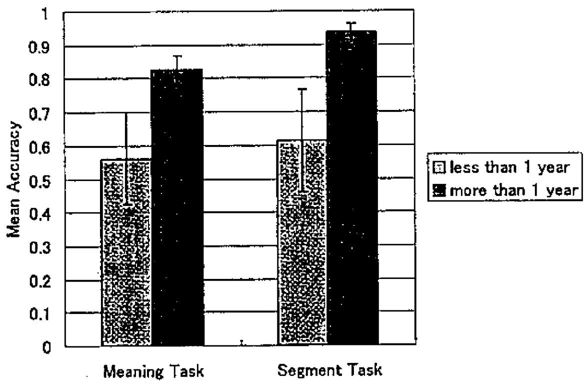

Figure 5.5: Type-1: the task is to recognize the difference between adverb and adjective. The error bars show standard error of the mean. A significant tendency was shown in the meaning task($F(1,7)=5.98$, $p<.1$).

5.3.3 Results

For type-1, the mean accuracies of the subjects who had more than 1-year experience were 0.82 for the meaning task and 0.93 for the segment task. The corresponding mean accuracies of those who had less than 1-year experience were 0.56 and 0.61 respectively. For type-2, those who were more experienced had mean accuracies of 0.93 for the meaning task and 0.95 for the segment task. Those who were less experienced scored 0.76 and 0.78 respectively. From the results of ANOVA, it was shown that the mean accuracy of the meaning task in the subjects who had more than 1-year experience tends to be higher than that of those who had less than 1-year experiences for type-1 ($F(1,7)=5.98$, $p=.06$) and also for type-2 ($F(1,7)=5.98$, $p=.08$) . The results showed that skill for recognizing the meaning of prosody is related to the length of experience. It therefore implies that MPR reflects some kinds of skills for interpreters.

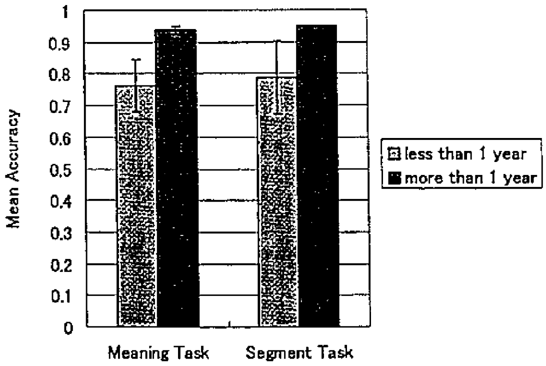

Figure 5.6: Type-2: the task to recognize the difference between nominative and genitive. The error bars show standard error of the mean. A significant tendency was shown in the meaning task($F(1,7)=5.98$, $p<.1$).

5.4 Experiment2

The aim of Experiment-2 is to investigate what kinds of skill can be estimated by MPR. In order to achieve this objective, we compare the evaluation scores in the situation of interpretation between the high and low-scores MPR groups.

5.4.1 Subjects

The subjects were six interpreters, in which three subjects had more than 80 pecent correct answers for MPR (high MPR group) and three subjects who scored less than 80 percent each (low MPR group). All subjects in high MPR group had more than 1 year experience for interpretation work, and those in the low MPR had less than 1 year experience. Two hearing subjects and three deaf subjects also participated in Experiment-2 in order to evaluate the skill of the interpreters.

5.4.2 Method

We prepared "the pizza task, i.e. a situation where a deaf customer orders some pizza from a hearing staff member through an interpreter. The reason why we selected "the pizza task" in this Experiment-2 was that it included some parts of free conversation in a formalized sequence of procedures and we thought that these characteristics in a situation close to a real interpretation scenario.

The hearing staff member was given a piece of paper containing questions about the customer's name, the fax number, the size of pizza, the kinds of sauce, the kinds of pizza crust, the topping to add and the topping to take away. We gave an instruction to each subject to finish the session in 5 minutes and we put a timer device in the room so that the subjects could check the time. The configuration of the three parties in one session was that a hearing customer and an interpreter were sitting in parallel, with the deaf customer sitting opposite the interpreter (5.4).

After each session, the deaf customer and the hearing staff member evaluated the skill of the interpreter on a scale of -to +5. The order and the combination of each interpreter was randomized. Each interpreter had three sessions and they were evaluated by three deaf customers and two hearing staff members. On the answer sheet, there were four questions in order to evaluate the interpretation and practical skills of each interpreter :

Q.1. Was the interpreter's sign(or speech) easy to understand?

Q.2. Did you feel any concern that your order didn't come through properly ?

Q.3 Was the interpreter expeditious?

Q.4 Did you succeed in ordering (or taking the order of) the pizza ?

5.4.3 Results

In the mean evaluation-score by three deaf subjects, the score of each of the high MPR group was higher than that for low MPR group in these subjects. Especially, the evaluation score for Q.3 tended to be higher in high MPR group than that in low MPR group (t(4),p=.056 : see 5.7). In the mean evaluation-score by two hearing subjects, there were significant difference between these two groups in all questions (t(4),p<.05, respectively: 5.8).

In order to know the degree of all the interpreter's skill from the view point of correlation between MPR and the evaluation-scores, we analyzed these correlations. Because the accuracy of each task in MPR was highly correlated (5.3), principal component analysis (PCA) was performed for each accuracy, including varimax rotation. Only one component was extracted and it took up 94.02% of the total variance (5.4).

We calculated the correlation between the component-1 and the mean evaluation score of each question in each subject. According to the results, a moderately strong correlation was found between the principal component for accuracy of MPR and the mean evaluation score by the deaf subjects in Q.2(r=.78, t(4)=2.53, p<.1), Q.3(r=.77, t(4)=2.43, p<.1) and Q4(r=.77, t(4)=2.43, p<.1).

Furthermore, a strong correlation between the principal component for the accuracy of MPR and the mean evaluation score by the hearing subjects was shown in Q.1(r=.94, t(4)=5.94, p<.01), Q.2(r=.96, t(4)=7.17, p<.01), Q.3(r=.99, t(4)=14.2, p<.01), and in Q.4(r=.86, t(4)=3.49, p<.05).

5.5 Discussion

The results of Experiment-1 indicated that MPR could measure some kinds of skills for interpreters. Similarly the results of Experiment-2 showed that MPR could be an indicator for a skill which was mainly evaluated by the hearing

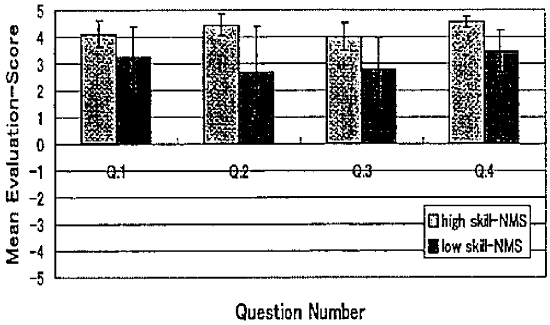

Figure 5.7: Mean Evaluation-Value by Deaf Subjects:the error bars show standard error of the mean. A significant tendency was shown in Q.3 (t(4),p<.1)

Table 5.3: Correlation among tasks. Task1 is meaning task and Task2 is segment task. Type-1 is to recognize the difference of nominative and genitive, and type-2 is to recognize that of adverb and adjective.

	TASK1-type1	TASK1-type2	TASK2-type1	TASK2-type2
TASK1-type1	1			
TASK1-type2	0.98	1		
TASK2-type1	0.9	0.95	1	
TASK2-type2	0.82	0.9	0.96	1

subjects. The hearing subject made the evaluation mainly by the interpreter's voice. Therefore, one possible skill that can be estimated by MPR is considered to be the translation skill from sign language to spoken language. In contrast, the deaf subject made the evaluation by the interpreter's behavior including signing. In this case no significant difference between the high MPR group and the low MPR group was seen. Neither was there any strong correlation between the principal component for the accuracy of MPR and the mean evaluation score, except a moderately strong correlation in Q.2, Q.3 and Q.4 that seems to be related to the practical skill for interpreters.

The stimulus of MPR is sign language and the reaction of that is based on Japanese, so the translation skill from spoken Japanese to sign language, which is mainly evaluated by deaf subjects, might be difficult to estimate by MPR.

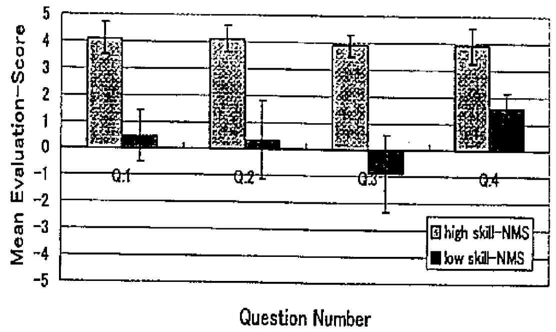

Figure 5.8: Mean Evaluation-Value by Hearing Subjects:the error bars show standard error of the mean. Significant difference was shown in all question(t(4),p<.05, respectively).

Table 5.4: Principal Component Loadings.

	Principal Component 1	Variance
Task1-type1	0.95	0.91
Task1-type2	0.99	0.98
Task2-type1	0.98	0.97
Task2-type2	0.95	0.9
Percent variance explained		94.02
Total variance explained		94.02

Further research are needed in order to confirm this , but we can say that MPR might at least be useful to estimate the translation skill from sign language to spoken Japanese. Additionally, the result of that the expeditiousness evaluated by the hearing subjects was remarkably different between the high and low MPR groups (Q3 in 5.8). Also a strong correlation was found between the principal component for the accuracy of MPR and the mean evaluation score by the hearing subjects. Furthermore we found a moderately strong correlation in Q.2, Q.3 and Q.4 by the deaf subjects as well. These practical skills including expeditiousness seem possible to estimate by MPR. The mean time to accomplish the task for the high MPR group was shorter than that of the low MPR group (3:51 and 4:34, respectively). The reason why MPR and the

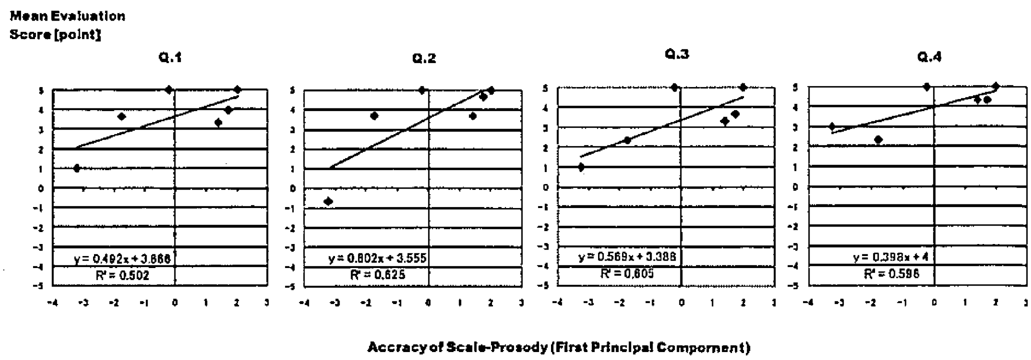

Figure 5.9: Correlation between the mean accuracy of MPR and the Evaluation Score by the Deaf Subjects: Moderately strong correlation was found in Q.2(r=.78, t(4)=2.53, p<.1), Q.3(r=.77, t(4)=2.43, p<.1) and Q4(r=.77, t(4)=2.43, p<.1)

practical skills for interpreters were correlated might be based on the volatility of information which is a remarkable feature of dialogue. [2] makes the following observations about the volatility of information.

- A dialogue cannot be recorded as text or images.
- The recipient has to remember all the information he/she has received.

Firstly, from the view point of the volatility of information in real time conversation, we could interpret that a high skill to recognize prosody would make it possible for interpreters to have enough time to work a sequence of procedures smoothly. This is because they could catch up on the volatile information without asking questions back or making the deaf or the hearing subjects anxious. This would not occur in all situations however. Because the pizza task consisted of very simple conversation, the interpreters did not need to translate any delicate nuance or fuzzy comment as are often observed in debates. We need further research of this point in a complex and debate type conversation.

Secondly, the skill of the expeditious interpretation also seems to be more important in the online interpretation, because the network connection force

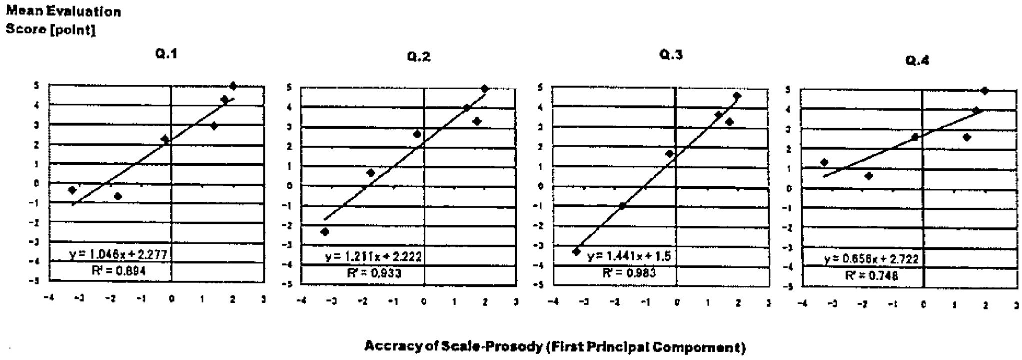

Figure 5.10: Correlation between the mean accuracy of MPR and the Evaluation Score by the Hearing Subjects: Strong correlation was found in Q.1(r=.94, $t(4)$=5.94, p<.01), Q.2(r=.96, $t(4)$=7.17, p<.01), Q.3(r=.99, $t(4)$=14.2, p<.01), and in Q.4(r=.86, $t(4)$=3.49, p<.05).

a delay[35]. Technically, it has been possible to conduct the experiment of online interpretation between JSL and spoken Japanese[33, 34], this will also be confirmed by further research.

Finally, in order to confirm the estimatable skills by MPR, we will develop other scales for general skill and then compare the score of MPR with it too. We will also analyze how long experiences are required for the acquisition of each skill and then apply the knowledge for educating interpreters based on their skill development.

5.6 Conclusion

In this study, we have examined two experiments: Experiment-1 tests the use of MPR determine the relationship between the score and the length of interpreter's experience. Experiment-2 which provides further evidence of the specific skill that could be measured by MPR. Our research indicated a possibility that MPR could be useful to estimate whether or not the interpreter is experienced enough to interpret from sign language to spoken Japanese. Furthermore, the practical skills such as working on the interpretation expeditiously could also be estimated. In order to confirm the estimatable skills by MPR and its situation, we will use different tasks and experimental apparatus. We will also develop other scales of general skill for JSL interpreter and then compare the score of MPR with the score of those scales.

Chapter 6

Spatial Configurations for Online Sign Interpretation

6.1 Introduction

With the growth of digital telecommunication networks and the spread of videophones, online sign interpretation services have become a reality. Generally, sign interpretation is performed by three types of participants: deaf people, interpreters, and hearing people. For easily usable sign interpretation services, the interchange of information among the three types of participants is required to be carried out smoothly. For smooth interchanges, arranging the relevant objects — video screens, video cameras, and people (users) — effectively is important. In this study, we mainly target the interpretation situation in which there is one of each type of participant. We will call it the trilateral interpretation.

The arrangement of videophone-related objects has been studied for the interpretation of speeches or lectures [32, 18], which we will call mass interpretation. However, the arrangement of videophone-related objects that is suitable for trilateral interpretation is expected to be different from that for mass interpretation. In a mass interpretation, there are generally several listeners.

Furthermore, in the case of mass interpretation, most communication is limited to one-way communication from the speaker to listeners, and communication in the opposite direction is limited to the question or discussion period. On the other hand, in trilateral interpretation, the balance of communication directions is nearly maintained and turn-taking frequently occurs.

The arrangement of videophone-related objects for trilateral interpretation service has not been studied. In this study, we discuss these issues, in particular, configurations of video screens and users[1].

Although the original motivation of interpretation is to communicate between a deaf person and a hearing person, they do not communicate with each other directly in real time. Instead, their communication is mediated by the interpreter. We call the conversation between the deaf person and the hearing person the "principal motivated conversation" and the people involved in it the "principal conversation partners."

In the interpretation service, the deaf person and interpreter mainly use sign language and the hearing person and the interpreter mainly use spoken language. However, it is well known that nonverbal information, which is not included in spoken or sign language, for example, pausing, nodding, facial expressions, and attitudes, also plays an important role in the establishment of communication[37, 25, 17]. In particular, note that the principal communication partners, who are motivated to communicate, can only communicate nonverbally between themselves. Therefore, much attention should be paid not only to the intelligibility of voice or sign but also to ease of exchanging nonverbal information, for smooth trilateral interpretation.

Here, we will classify situations of the trilateral interpretation service based on the positional relationship among three people. The first situation is where

[1]In our experiments, the video camera configuration was fixed; the video camera was mounted near the video screen, facing in the same direction as the screen.

three people are present in a place. They start communication without using a network. That is called "on-site interpretation." The second situation is where a videophone terminal is set in a shop and a deaf person has come in to perform some errands. If there is no one present who can use sign language, the shop staff can call an interpreter who is in a remote location waiting to help via the videophone. This situation is called "partial online interpretation." Another situation is where all three people are in different locations. In this case, two videophone connections are sufficient to support the conversation. This situation is called "full online interpretation."

Then, we defined the spatial configurations of videophone equipment for each configuration. In particular, the configuration in which participants can exchange nonverbal signals was presented.

Using these configurations, we performed simulations of sign interpretation services. For the conversation task, we chose "ordering a pizza," in which a deaf customer orders a pizza from a hearing pizzeria clerk mediated by a sign interpreter. After the task was finished, the subjects gave their subjective assessments.

As a first result, the accomplishment times of on-site interpretation were found to be significantly shorter than those of other configurations. This result suggests that partial and full online interpretation provides less smooth communication compared with on-site interpretation, which is a predicted result. However, we found by analyzing the evaluation results in detail that a well-designed configuration improved the feasibility of online interpretations. That is, from the evaluations of hearing subjects, configurations in which they could see images of the deaf person were found to be preferable compared with configurations in which they could not see the deaf person. This implies that hearing people who do not understand sign language receive nonverbal signals that enable online communication to be similar to on-site communication. From these results, we obtained approximate guidelines for designing configurations of videophone equipment for an online sign interpretation service.

6.2 Spatial configuration for sign interpretation service

In this section, spatial configurations of videophone equipment and related people are defined for each configuration.

The classifications and configurations are shown in Fig. 1. Here, "I" denotes the interpreter, "**H**" the hearing person, and "**D**" the deaf person. Lines surrounding these letters indicate that corresponding people are in the same physical location. The solid-line arrows indicate on-site communication and the dotted-line arrows indicate online communication.

6.2.1 On-site interpretation

In "on-site interpretation," the preferred configuration is one in which the hearing person and the interpreter are side by side facing in the same direction (i.e., in parallel), and they are both facing the deaf person (Fig. 1(a)). This configuration is called **3SS-E** (**Three** people **S**haring a **S**pace and the principal conversation partners facing **E**ach other). In this configuration, the principal conversation partners that are motivated to communicate are facing each other, so exchanging nonverbal signals between them is easy. We consider that this is the reason the configuration is preferred.

In the following subsections, we are going to define the configurations for online interpretations by considering the issue of nonverbal signals.

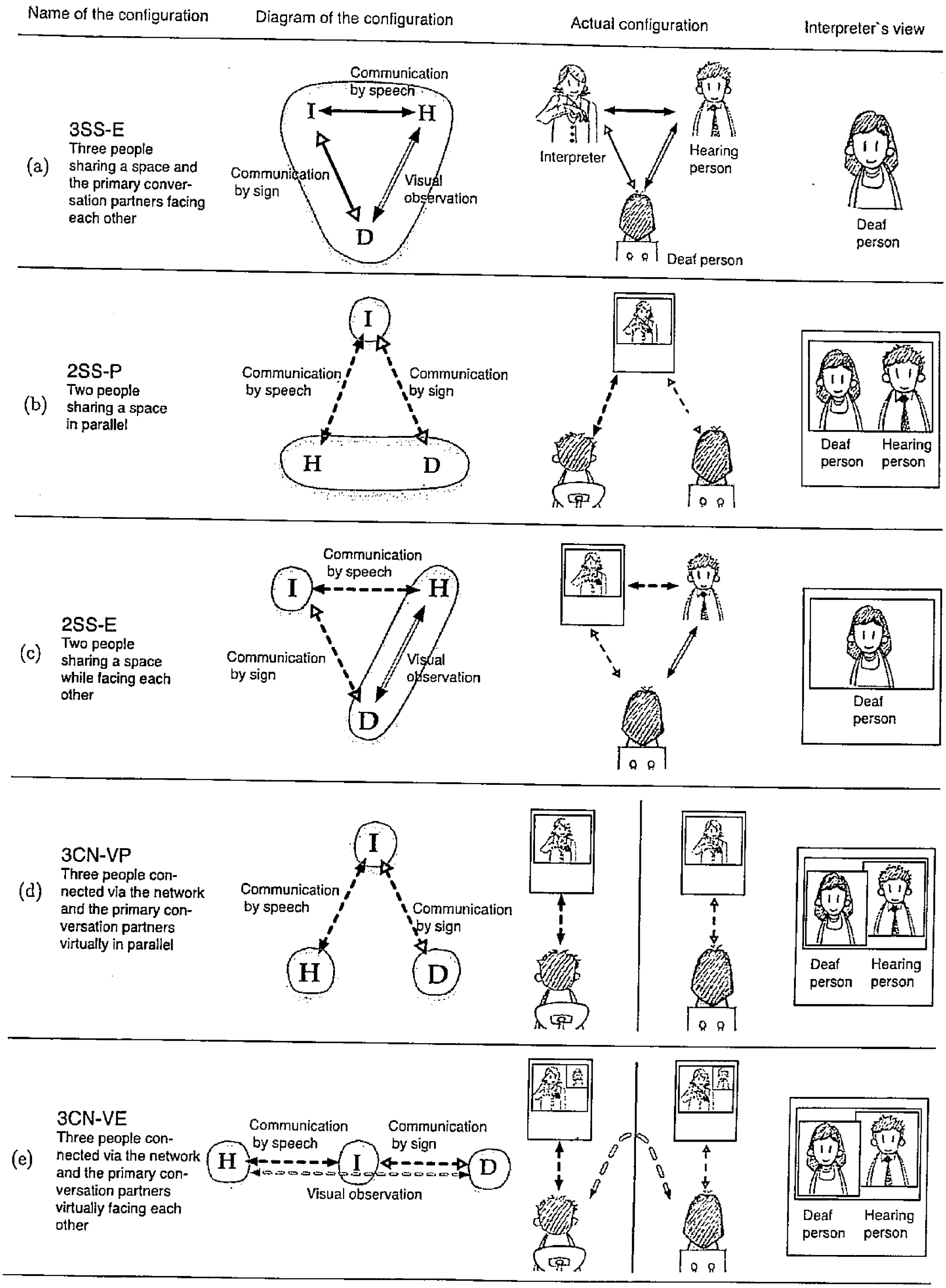

Figure 6.1: Classification of equipment configurations.

6.2.2 Partial online interpretation

In the "partial online interpretation," the deaf and hearing persons communicate with the interpreter by a single videophone connection. For this situation, we considered two types of configurations. In one configuration, the deaf and hearing persons are side by side (in parallel) facing the interpreter via the network (Fig. 1(b)). This is called **2SS-P** (**T**wo people **S**haring a **S**pace in **P**arallel). The other configuration has the deaf and hearing persons facing each other with the deaf person facing the interpreter via the network (Fig. 1(c)). This is called **2SS-E** (**T**wo people **S**haring a **S**pace while facing **E**ach other). This configuration corresponds to the partial online version of **3SS-E**.

6.2.3 Full online interpretation

In the "full online interpretation," the deaf person communicates with the interpreter by one videophone connection and the hearing person communicates with the interpreter by another videophone connection. If a conventional (point-to-point connection) videophone system is applied for the full online interpretation straightforwardly, the deaf and hearing persons cannot see each other because the two connections are isolated. However, by slightly modifying the videophone system, we can enable them to see each other in small windows.

For this situation, we investigated two configurations. In one configuration, the principal conversation partners do not see the image of their partner (Fig. 1(d)). Here, we can consider the principal conversation partners to be virtually sitting in parallel. This is called **3CN-VP** (**T**hree people **C**onnected via the **N**etwork and the principal conversation partners **V**irtually in **P**arallel). In the other configuration, the principal conversation partners can see each other in small windows (Fig. 1(e)). Here, we can consider the principal conversation partners to be virtually facing each other. This is called **3CN-VE** (**T**hree people Connected via the Network and the principal conversation partners Virtually facing Each other). In this configuration, the hearing person faces both the deaf person and the interpreter while the deaf person faces both the hearing person and the interpreter. Note that we cannot do such a thing in the actual world. This configuration can be called artificial.

6.3 Methods

Based on the classification of situations and equipment configurations stated in Sect. 2, we performed simulations of sign interpretation services. The task requiring interpretation was placing an order for home delivery of a pizza. The reason for selecting this task is that it includes some free conversation in a formalized sequence of procedures. We thought that characteristics of the conversation would enable us to compare and evaluate configurations that were as close as possible to actual interpretation situations.

Deaf persons were assigned to the role of customers and hearing persons were assigned to the role of pizzeria clerk. Subjective evaluations were provided by all three groups of subjects.

6.3.1 Experimental system

For the online interpretation experiments, two IP videophone terminals VP1500 (NTT East) and one software phone on a personal computer were used. They communicated via the FLET's .Net IP network. The distance between subjects who were facing each other was set to 1.5 meters. We used an external 20-inch LCD video monitor because the built-in video monitor of the VP1500 has a diagonal size of eight inches, which was too small to see from that distance. Furthermore, we used an external small video camera (CK-300S, Keyence), which was attached underneath the external video monitor because unless the camera was set near the video monitor, the line of sight of the subject was not directed at the camera.

6.3.2 Subjects

Three groups of subjects took part in the experiments. The characteristics of subjects are shown in Table 6.1.

None of the interpreters were certificated JSL (Japanese Sign Language) interpreters. Three of four interpreters graduated from a professional school of sign language interpreters and one of them studied and trained in a workshop for sign language interpreters for one year. Furthermore, all of them have experience in the operation of an interpretation service at a private company for about one year.

6.3.3 Procedure of sign interpretation experiments

Various combinations of deaf person, interpreter, and hearing person were made and pizza-ordering experiments were performed for five kinds of configurations. Each interpreter performed the task once for each of the configurations. However, for the **3SS-E** configuration, the interpreter performed the task three times because the results for this configuration were also used for another purpose. The combinations were made carefully to avoid any bias or order effect.

The deaf persons were given a sheet of paper containing the necessary information: a fictional name, address, and fax number and a description of the desired pizza. The hearing persons had another piece of paper on which to write key information for the customer's order: the customer's name, address, fax number, pizza size, kind of sauce, kind of pizza crust, and toppings to add or remove from a menu item. Subjects were instructed to finish the session within five minutes. A timer was placed in each room to enable the subjects to check the elapsed time.

After each pizza-ordering task was finished, the time taken to accomplish

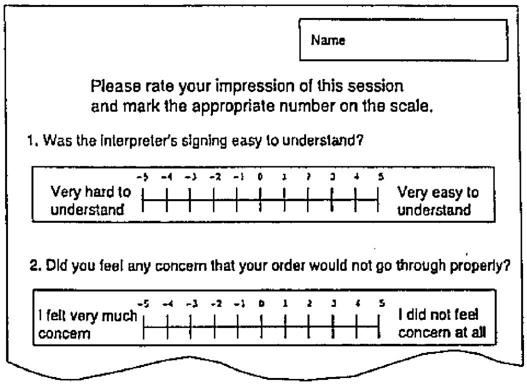

Figure 6.2: Evaluation sheet.

the task was recorded and the three subjects were asked for their assessments.

6.3.4 Evaluation methods

Evaluations by deaf subjects

Deaf subjects who acted as customers evaluated the quality of the conversation. The evaluation sheet is shown in fig 6.2. The five evaluation items are shown below.

Q-D.1 Was the interpreter's signing easy to understand?

Q-D.2 Did you feel any concern that your order would not go through properly?

Q-D.3 Was the interpreter expeditious?

Q-D.4 Did you succeed in ordering the pizza?

Q-D.5 Did you feel the hearing pizzeria clerk was friendly?

Evaluations by hearing subjects

The hearing subjects who acted as pizzeria clerks evaluated the quality of conversation according to the following five items. Items of the answer sheets were the same as those for deaf subjects.

Q-H.1 Was the interpreter's speech easy to understand?

Q-H.2 Did you feel any concern that you might miscommunicate with the customer?

Q-H.3 Was the interpreter expeditious?

Q-H.4 Did you succeed in taking the pizza order?

Q-H.5 Did you feel the deaf customer was friendly?

Evaluations by interpreters using NASA-TLX

The interpreters gave their assessments by using NASA-TLX[8]. NASA-TLX is a subjective workload assessment method developed by NASA (National Aeronautics and Space Administration) Ames Research Center. The evaluation procedure consisted of the following three steps.

1. Evaluation of workloads on six subscales: Evaluate the workload on six subscales using values from 0 to 100
2. Comparison of subscale weights: Evaluate the weights of the subscales by the pair-comparison method
3. Unification of workloads: Calculate the weighted workload (WWL) by taking the average of the weighted values of the subscales

And the six subscales are: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration.

In the original NASA-TLX, the explanations of the procedure and evaluation items are written in English and are too rigid and technical. For example, the description of Mental Demand in the original NASA-TLX is as follows: *How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?*

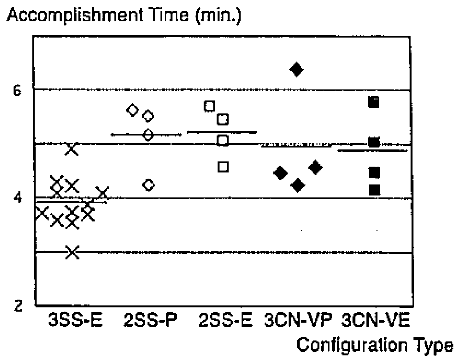

Figure 6.3: Accomplishment times under five configurations.

Japanese versions of NASA-TLX were translated by some researchers and widely used[41]. However, these translations are verbatim and still rigid and technical. Consequently, using the translated version for nontechnical people, we were anxious about introducing an unexpected effect, and carrying out fair assessments became difficult.

We translated all the words and terms in the descriptions into intuitive, simple Japanese sentences and used them for our experiments. Translated descriptions of subscales are shown in Table 6.2 [2].

6.4 Results

This section gives an overview of the experimental results. First, accomplishment times of trials for five configurations are shown. Then, results of assessments under five configurations classified according to subject groups are shown.

6.4.1 Accomplishment times

The time taken to accomplish the task can be regarded as a comprehensive measure of the feasibility of the communication interpretation system because,

[2] Note that the sentences in this Table are translated into English again trying to keep the nuance of the Japanese sentences.

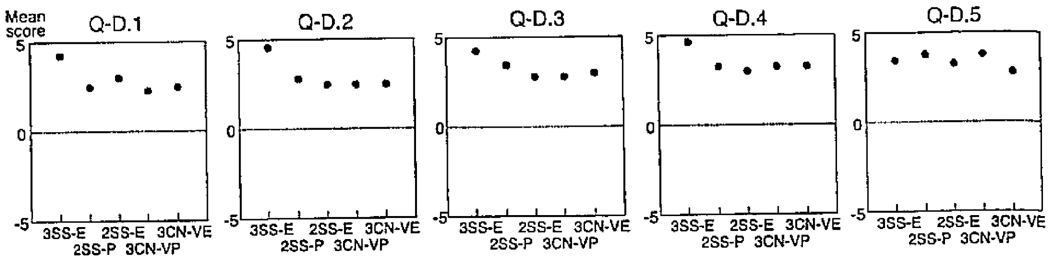

Figure 6.4: Mean scores on five evaluation items assessed by deaf subjects for five configurations.

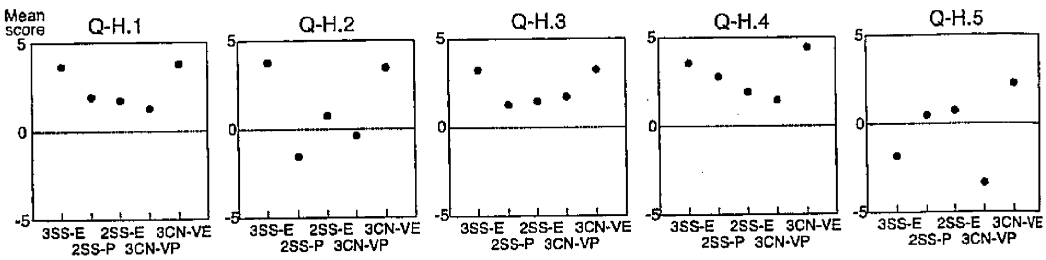

Figure 6.5: Mean scores of five evaluation items assessed by hearing subjects under five configurations.

if the system makes accomplishing the tasks easy, conversation proceeds without disturbance.

The accomplishment times under the five configurations are shown in Fig. 6.3. Markers correspond to the accomplishment times for each trial and short crossbars indicate average times for each configuration. The accomplishment times for the on-site interpretation configuration (**3SS-E**) were significantly shorter than the other four configurations. Among the other four configurations, no obvious differences were observed. The reasons are assumed to be as follows: (1) the number of trials was very small (four times) and (2) this task tends to yield fluctuations in accomplishment times because it irregularly requires repeated asking and confirming of the customer's name or address, which takes a certain amount of time.

6.4.2 Results of evaluations by deaf subjects

The results of the subjective assessments by the deaf subjects are shown in Fig. 6.4.

Overall, no obvious differences were observed among the configurations. The reason is assumed to be that the deaf people were tolerant of impairments to the progression of the sign conversation. The tolerance of deaf people is also exemplified by the findings that they are tolerant of a delay in video images during sign conversations conducted via videophones[20]. This idea is also supported by the finding that all scores were always high (Compare Fig. 6.4 with Fig. 6.5.)

6.4.3 Results of evaluations by hearing subjects

The results of subjective assessments by hearing subjects are shown in Fig. 6.5.

The mean scores for evaluation items Q-H.1, Q-H.3, and Q-H.4 were always high, and the fluctuations were small for all configurations. This is because the intelligibility (Q-H.1) and the expeditiousness (Q-H.3) of the interpreter were independent of configurations, and the senses of achievement (Q-H.4) were not very different between trials because they were carried out until the order was finished. For the remaining questions (Q-H.2 and Q-H.5) significant differences in the answers were observed. These issues are discussed in section 5.

6.4.4 Results of evaluations by interpreters

Variations in WWLs are shown in Fig. 6.6. No obvious differences among the configurations were observed. The effect of configurations from the viewpoints of interpreters will be discussed using the individual results of subscales in Sect. 5.

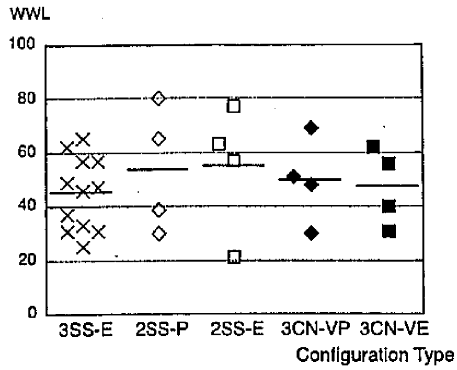

Figure 6.6: Weighted workload values for interpreters.

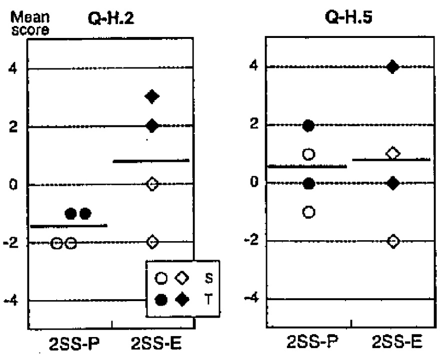

Figure 6.7: Comparison of subjective assessment by hearing subjects (partial online interpretation).

6.5 Discussion

This section discusses configurations of equipment classified by situations of interpretation, that is, partial, online, and full online. For each situation, configurations are discussed using assessments by hearing subjects and interpreters. Unfortunately, clear results were not obtained from the assessments by deaf subjects, so their assessments will not be discussed.

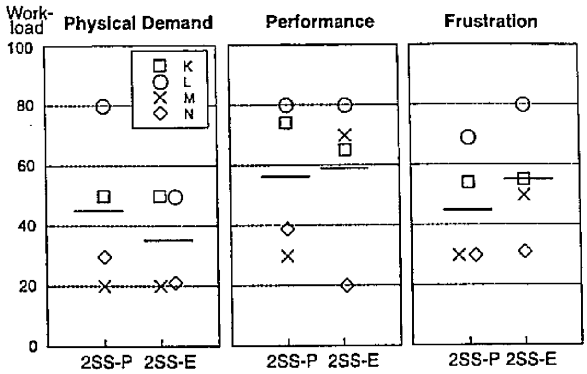

Figure 6.8: Comparison of workload for interpreters (partial online interpretation).

6.5.1 Comparison within partial online interpretation situations

results assessed by hearing subjects The results obtained for subjective assessments by hearing subjects, questions Q-H.2 (Concern about the miscommunication) and Q-H.5 (Friendliness with the partner), are shown in Fig. 6.7. All of the raw data are shown in these graphs.

From the viewpoint of hearing subjects, the differences between **2SS-P** and **2SS-E** can be summarized as follows. The former is the configuration in which the interpreter is seen in the video monitor in front of the subject and the deaf subject is sitting beside him or her. The latter is the configuration in which the deaf subject is sitting in front of the subject and the voice of the interpreter can be heard from the videophone speaker.

The results for Q-H.2 suggest that three of four scores for **2SS-E** were better than those for **2SS-P**. This means that nonverbal information obtained by visual observation of one's principal conversation partner prevented the hearing subject from being concerned that he/she might miscommunicate with the partner. On the other hand, the results for Q-H.5 suggest that there were no differences between the two configurations. This finding implies that if the principal conversation partners are sharing the space, the visual observation of the partner does not make a further impact on the feeling of friendliness.

results by interpreters Instead of examining WWL, let us study each score on the subscale because no obvious differences were observed in the WWLs that we obtained through those procedures. The scores for three individual subscales (Workloads)—Physical Demand, Performance, and Frustration—are shown in Fig. 6.8. The scores of these subscales suggested clear tendencies in the six kinds of subscales.

From the viewpoint of the interpreter, the difference between **2SS-P** and **2SS-E** can be summarized as follows. The former is the configuration in which both deaf and hearing subjects are seen on a video monitor in front of the interpreter; the latter is a configuration in which the deaf subject is seen on the video monitor and the voice of the hearing subject can be heard from the speaker.

As shown in Fig. 6.8 two of four subjects felt that the workload for Physical Demand was heavier in configuration **2SS-P** (the other two subjects felt that there was no difference between the two configurations), while all subjects felt that the workload for Frustration was heavier in configuration **2SS-E**. Deciding which Performance is heavier is difficult. These results imply that nonverbal information obtained by visual observation of one's principal conversation partner imposes much physical stress but causes less frustration.

6.5.2 Comparison under full online interpretation situations

results assessed by hearing subjects The results of subjective assessments by hearing subjects, questions Q-H.2 (Concern about the miscommunication)

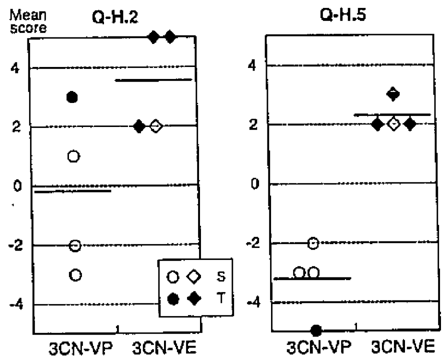

Figure 6.9: Comparison of subjective assessments by hearing subjects (full online interpretation).

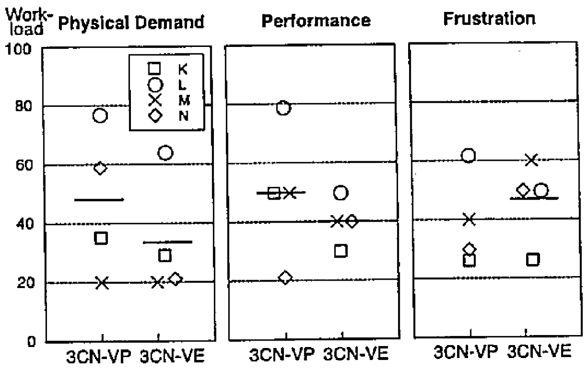

Figure 6.10: Comparison of workload for interpreters (full online interpretation).

and Q-H.5 (Friendliness with the partner), are shown in Fig. 6.9. All of the raw data are shown in these graphs[3].

From the viewpoint of the hearing subjects, the differences between **3CN-VP** and **3CN-VE** can be summarized as follows. The former is the configuration in which only the interpreter is seen on the video monitor. The latter is the configuration in which the interpreter is seen on the video monitor and the deaf subject is seen in a small window.

[3]In these configurations, through our inattention, there was an imbalance in the number of trials for hearing subjects; that is, subject S performs 3CN-VP three times and 3CN-VE once.

Results for Q-H.2 and Q-H.5 indicate that scores of **3CN-VE** tend to be higher than those of **3CN-VP**. This suggests that the nonverbal information introduced by features of configuration **3CN-VE** brought the feeling that subjects' intentions were carried out properly and provided a feeling of friendliness with the deaf subjects.

results by interpreters The scores for three individual subscales (Workloads)—Physical Demand, Performance, and Frustration—are shown in Fig. 6.10.

From the viewpoint of the interpreters, there were no differences between **3CN-VP** and **3CN-VE**; that is, in both of these configurations, the interpreter could see the faces of both deaf and hearing subjects on the video screen (Figs. 1(d) and (e)). Here, we investigate individual scores.

Looking at Fig. 6.10, interpreters had difficulty seeing the difference in workload caused by Frustration, but they felt there was less workload caused by Physical Demand and Performance in **3CN-VE** than in **3CN-VP**. This result is congruent with results of hearing subjects shown in Fig. 6.9[4]. The reason that a difference between the configurations was seen even though there was no difference in the view of the evaluator can be elucidated as follows. The interpreter experienced an indirect effect, which arose from the progress of conversation or the mental situation of the other subjects (hearing subjects, most likely.)

6.6 Conclusion

Configurations of videophone-related objects for a sign interpretation service were studied. Situations encountered in interpretation services were classified into on-site interpretation, partial online interpretation, and full online inter-

[4]Note that the workload is preferable if the value is smaller because the directions of the evaluation axes of Figs. 6.9 and 6.10 are opposite of each other.

pretation. Then, two kinds of configurations were introduced for each of partial and full online interpretations. In particular, a well-designed configuration was presented in which subjects could exchange nonverbal signals by locating objects to see faces of each of the principal conversation partners. Simulation experiments of interpretation services were performed and qualities of the configurations were assessed. As a result, we found that a configuration in which subjects could exchange nonverbal signals improved the feasibility of partial and full online interpretation. Consequently, we obtained approximate guidelines of designing configurations of videophone equipment for online sign interpretation.

In order to confirm the effectiveness of the guidelines, we will continue to evaluate the quality of conversation by different tasks. We will also analyze the interactive behavior by using the conversation data, which was video-taped through all trial in this study. In addition to the subjective evaluations, these findings would contribute to establish the different types of evaluation scales for the quality of conversation.

Table 6.1: Characteristics of deaf subjects.

Symbol	Gender	Age	Hearing acuity	Years of sign experience
Hearing impaired subjects				
A	F	56 – 65	Deaf	10 —
B	F	46 – 55	Deaf	10 —
C	M	56 – 65	Deaf	10 —
D	M	26 – 35	Difficult	4 – 6
E	M	18 – 25	Deaf	4 – 6
F	M	26 – 35	Difficult	10 —
Interpreter subjects				
K	F	26 – 35	Normal	10 —
L	M	26 – 35	Normal	7 – 9
M	M	26 – 35	Normal	1 – 3
N	F	18 – 25	Normal	7 – 9
Hearing subjects				
S	M	46 – 55	Normal	none
T	F	26 – 35	Normal	none

Table 6.2: Descriptions of subscales in NASA-TLX, which were used in our experiments.

Original title	Description of subscale
Mental Demand	Did you have trouble in thinking?
Physical Demand	Were you tired physically?
Temporal Demand	Did you get flustered?
Performance	Did you perform well? Are you OK?
Effort	Did you make an effort?
Frustration Level	Did you become irritated?

Chapter 7

Discussion

For an overview of the difference between the static information and the non-linguistic information, which has been discussed in this study, reclassifying the verbal and the non-verbal communications are useful. In the following sections, a new classification will be proposed based on the results of our experiments.

7.1 Existing classification of VC and NVC

In the research field of social psychology, it has been said that we can distinguish between verbal communication (VC) and non-verbal communication (NVC) by considering whether it consists of speech, or it is independent of speech[16], [25]. From this definition, speech itself would be classified as VC and eye gaze would be classified as NVC. In other words, we can say that VC is produced by the organic working in the oral cavity and is perceived by the acoustic sense, and NVC is produced by the body and perceived by visual sense (see fig.7.1).

However, if we distinguish between VC and NVC by the organs for production and the sense for perception in spoken language, all sign language would be classified as NVC, since the main productive media of sign language are hand, torso, head, face, and those movements are percepted by visual sense. Obviously, there are linguistic properties in sign language, so we need to consider

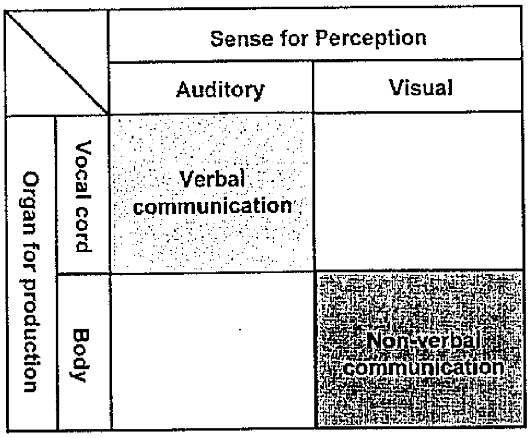

Figure 7.1: traditional classification of VC and NVC

other types of criteria to explain both spoken and sign languages comprehensively.

7.2 New classification of VC and NVC

Considering the cognitive process of VC and NVC, the information seems to be perceived by the sensory organ first. It is then compared to existing knowledge such as symbols and images. It is then finally processed as the propositional or the expressive information which produces a reaction as the linguistic or the emotional response. This is basic processing both in sign and spoken languages, so we can use the three axes to classify VC and NVS; the level of sense for perception (X axis), the level of comparison to the existing knowledge (Y axis), and the level of processing for propositional/expressive information (Z axis).

In VC, there is linguistic information, and also paralinguistic information such as prosodic signal, pitch, pause, and emphasis. Fig.7.2 and Fig.7.3 show the classification of these properties in each language. Fig.7.4 shows the classification viewed from the common cognitive processing after the perception

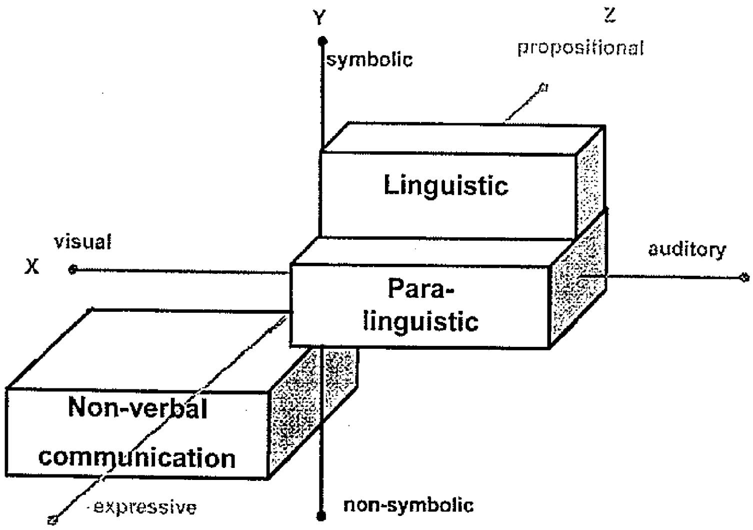

Figure 7.2: Classification of Speech in Three Axes

level. By Fig.7.4, we can see how some media used the three properties in each language. The components of VC and NVC can be classified as Table.7.5. In Table.7.5, this study has discussed about para-linguistic information (Chapter3-5) and non-verbal information (Chapter 6). For the future work, we will continue to evaluate our applications and modify the classification itself based on the results of further analysis.

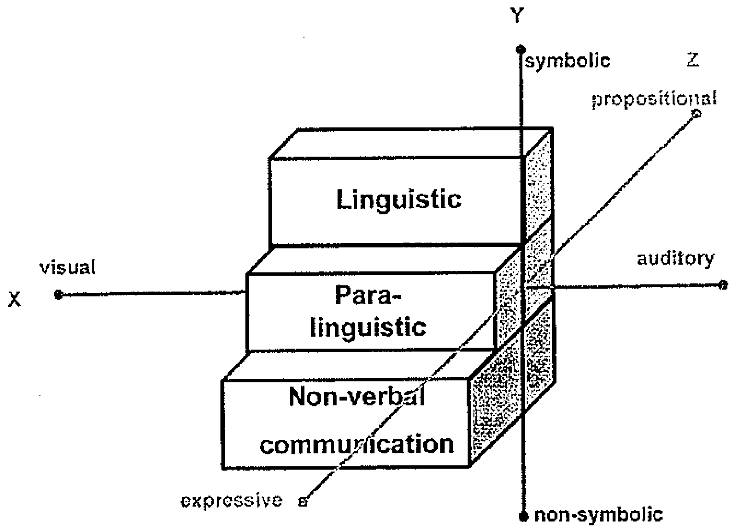

Figure 7.3: Classification of Sign Language in Three Axes

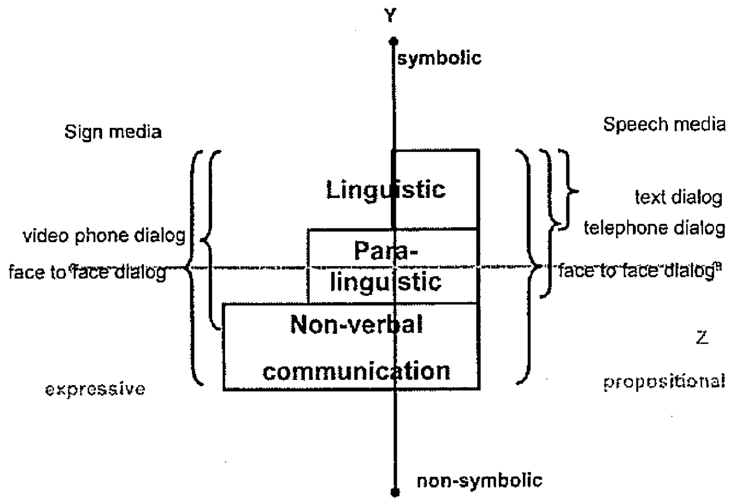

Figure 7.4: Classification Viewed from Y-Z Coordinate

language type	Feature of components				name of components	example
spoken language	verbal	linguistic			speech	phoneme, word, sentence
		para-linguistic	vocal	segmental	accent	pitch accent, stress accent
				super-segmental	prosody	pitch, duration
			non-vocal		timing effect	pause, timing of utterance
	non-verbal	human	active		gesture	gaze, body touch, posture, facial expression
			spacial		proxemics	inter-personal distance, seating position
		non-human	intentional		artifact	clothing, makeup, accessory, road sign
			non-intentional		physical enviroment	furniture, lighting, temperature
sign langauge	verbal	linguistic			sign language	handshape/place/movement, word, sentence
		para-linguistic	manual	super-segmental	prosody	speed, duration, distance
				time series pattern	timing effect	transitional movement, pause
			non-manual	head, facial expressio	facial prosody	head, jaw, cheeks, lip, eyebrows and eyes movement
	non-verbal	human	active		gesture	nod, body touch, posture
			spacial		proxemics	inter-personal distance, seating position
		non-human	intentional		artifact	clothing, makeup, accessory, road sign
			non-intentional		physical enviroment	furniture, lighting, temperature

Figure 7.5: Classification of VC and NVC in Sign and Speech

Chapter 8

Conclusion

In order to develop more communicative applications for JSL learning and interpretation, this thesis has been discussed the basic and applied researches, especially in (1)production mechanism of prominence, (2)prosody for skill evaluation, (3)learning effect of segmentation, and (4)non-verbal information for sign interpretation. Sign language is a type of language that dose not have written language. That was one of the reasons why educating non-native signers is so difficult. By the knowledge of non-linguistic information, however, it is possible to convey the comunicative use of JSL to non-native signers.

We explored the four sides of non-linguistic information and had the following results.

(1) the hands movement, speed and duration were totally involved in the generation of prominence which is correlated with emphasized signs. The knowledge of prominence of JSL is expected to apply in the education field in JSL.

(2) the learning program focusing on the segment unit with the transitional movement were more effective than the existing word learning method.

(3) The score of MPR, which evaluate the recognition skill for prosody, is correlated with the interpreter's general skill, such as the speech interpretation skill and the skill to work on the interpretation expeditiously.

(4) A configuration in which subjects could exchange nonverbal signals improved the feasibility of partial and full online interpretation.

Based on our classification of comunicative compornents in each media, we will continue to evaluate the applications and modify the classification itself. This approach would also help us to rearch the substantive function that is needed for communication beyond the defference of speech or auditoral media.

Bibliography

[1] *Theory and Practices for Sign Language Interpreters(in Japanese).* Japanese Federation of the Deaf, 1998. Committee of instructors for JSL interpreters.

[2] Ichikawa A. Dialogue language and person with disabilities. *IEICE - Transactions on Information and Systems archive*, (6):1312–1319., 2004.

[3] Martinet A. *Elements of General Linguistics.* University Of Chicago Press, 1972.

[4] Rosenbaum D A, Loukopoulos L D, Meulenbroek R G J, Vaughan J, and Engelbrecht S E. Planning reaches by evaluating stored postures. *Psychological Review*, 102:28, 1995.

[5] Wilbur R B. Stress in asl : Empirical evidence and linguistic issues. *Language and Speech*, 42(2).

[6] Wilbur R B. Metrical structure, morphological gaps, and possible grammaticalization in asl. *Sign Language and Linguistics*, 2:217–244, 1999.

[7] Wilbur R. B. and Zelaznik H. N. Kinematic correlates of stress and phrase position in ASL. *Talk presented at the 1997 meeting of the Linguistic Society of America*, 1997.

[8] Hart J C and Bittner A. Development of NASA-TLX: Results of empirical and theoretical research in human mental workload. *British Journal of Social Psychology*, pages pp.139–183, 1988.

[9] Mauk C. *Undershoot in Two Modalities: Evidence from Fast Speech and Fast Signing*. PhD thesis, University of Texas at Austin, 2003.

[10] Roy C. A sociolinguistic analysis of the interpreters role in simultaneous talk in a face to face interpreted dialogue. *Sign Language Studies*, 74:21–61, 1992.

[11] Mutoh D and Nagashima Y. Conversation analysis support system for sign video images. *Human Interface*, 3(3), 2001.

[12] Efthiniou E, Sapountzaki G, Karpuzis K, and Fotinea S E. *Developing an e-Learning platform for the Greek Sign Language*, pages 1107–1113. Springer, 3118 edition, 2004.

[13] Ohira E, Sagawa H, Sakiyama T, and Ohki M. A segmentation method for sign language recognition. *IECE Trans. Information and system*, (1):49, 1995.

[14] Compass: Video for JSL Learning. *http://www.inform-inc.co.jp/video.htm*.

[15] Isono H, Takiguchi Y, Katsumata M, and Nakama M. Preferred reproduction speed of sign language image and audiovisual system via the internet (in Japanese). pages pp.5–8, 2006.

[16] Daibo I. *Communication of gesture (in Japanese)*. Science company, 1998.

[17] Naito I, Murakami H, Katoh N, Minagawa H, and Ishihara Y.

[18] Naito I, Murakami H, Katoh N, Minagawa H, and Ishihara Y. Basic considerations regarding multi-point video conferencing for the hearing impaired. *TR of IEICE SP 2003-110*, pages pp.41–46, 2003.

[19] Morimoto K., Kurokawa T., Isobe N., and Miyashita J. Design of computer animation of Japanese Sign Language for hearing-impaired people in stomach X-ray inspection. In *Computers Helping People with Special Needs -ICCHP 2004*, pages 1114–1120, 2004.

[20] Nakazono K, Nagashima Y, and Terauchi M. Evaluation of effect of delay on sign video communication. In *Computers Helping People with Special Needs-ICCHP 2006.*

[21] Tabata K, Kurota T, Murakami M, Manabe Y, and Chihara K. Prototype design for sign language education system (in Japanese). In *Proceedings of Japanese Association of Sign Linguistics*, number 27, pages pp.34–35, 2001.

[22] Uehara K and Shimada M. *Extraction of Primitive Motion and Discovery of Association Rules from Human Motion Data*, volume 2281, pages 592–602. Springer, 2002.

[23] Akita M., Fukumura N., and Uno Y. A detection method of sign word sequences based on the angular jerk of human arm movements (in Japanese). *Technical report of IEICE*, page 7, 2003.

[24] Akita M., Fukumura N., and Uno Y. A detection method of sign word sequences based on the angular jerk of human arm movements (in Japanese). *Technical report of IEICE*, pages 7–11, 2003.

[25] Argyle M. *Bodily Communication. 2nd edition.* Methen and Co.Ltd., 1988.

[26] Claude M. *Undershoot in two modalities: Evidence from fast speech and fast signing.* PhD thesis, Univ. of Texas, 2003.

[27] Metzger M. *Sign Language Interpreting: deconstructing the myth of neutrality.* Gallaudet University Press, 1999.

[28] Nespor M and Vorgel I. *Prosodic Phonology.* Foris Publications, 1986.

[29] Shirasawa M. technique of sign language interpretation and its psycholinguistic process. In *Introduction to Psychology for Sign Langauge (in Japanese)*, pages 149–175. Toho Shobo, 2005.

[30] Tyrone M. *Overview of Capture Techniques for Studying Sign Language Phonetics*, page 101. Springer, 2002.

[31] Hirayama N, Horiuchi Y, and Ichikawa A. Timing structure of JSL. *Trans. Human Interface Society*, 3(3):9–14., 2001.

[32] Katoh N, Naito I, Murakami H, Minagawa H, and Ishihara Y. Visual communication with dual video transmissions for remote sign language interpretation servicesl. *2002 IEEE International Conference on Systems*, page ., 2002.

[33] Kaoru Nakazono. Frame rate as a qos parameter and its influence on speech perception. *Multimedia Syst.*, 6(5):359, 1998.

[34] Kaoru Nakazono. Digital encoding applied to sign language video source. *IEICE - Transactions on Information and Systems archive*, (6):1893, 2006.

[35] Kaoru Nakazono, Yuji Nagashima, and Mina Terauchi. Evaluation of effect of delay on sign video communication. In *ICCHP*, pages 659–666, 2006.

[36] Braem B P. Rhythmic temporal patterns in the signing of deaf early and late learners of swiss german sign language. *Language and Speech*, 42(2).

[37] Richmond V P, McCroskey J C, and Argyle M. *Nonverbal Behavior in Interpersonal Relations.*

[38] Lehman R and Conceicao S. Involving the deaf community in distance learning, using blended technologies and learning objects. *IEEE Electronic Journal*, pages pp.3–4, 2001.

[39] Kawano S, Izumi C, Kurokawa T, and Morimoto K. Japanese JSL translation and searching display conditions for expressing easy-to-understand sign animation. In *Computers Helping People with Special Needs -ICCHP 2006*, pages 667–674, 2006.

[40] Liddell S. Think and believe: sequentiality in American Sign Language. *Language; Journal of the Linguistic Society of America*, 60(2):373, 1984.

[41] Miyake S and Kumashiro M. Dsubjective mental workload assessment techinique—an introduction to NASA-TLX and SWAT and a proposal of simple scoring methods—. *The Japanese Journal of Ergonomics*, 29(6):399, 1993.

[42] Takeda S and Ichikawa A. Analysis of prominence in spoken Japanese sentences and application to text-to-speech synthesis. *Speech communication*, 14:171, 1994.

[43] Tanaka S and Hulst H. Non-lexical phase and emphasis in Japanese Sign Language -a quantitative approach using speed information-. In *Theoretical Issues in Sign Language Research*, 2004.

[44] Tanaka S, Nishida M, Horiuchi Y, and Ichikawa A. Production of prominence in Japanese Sign Language. In *INTERSPEECH*, pages 2421–2424, 2005.

[45] Tanaka S, Nishida M, Horiuchi Y, and Ichikawa A. Meaning and prosody in sentence of japanese sign language. *Human Interface*, 9(1):13–17., 2007.

[46] Kadous M W. Learning comprehensible descriptions of multivariate time series. In *Proc. 16th International Conf. on Machine Learning*, pages 454–463. Morgan Kaufmann, San Francisco, CA, 1999.

[47] Sandler W. Prosody in two natural language modalities. *Language and Speech*, 42(2), 1999.

[48] Fukuda Y. Compilations of the electronic dictionary of Japanese Sign Language (a second edition) and its instruction manual (in Japanese). *IEICE technical report. Welfare Information technology*, (66):pp.39–44, 2005.

[49] Fukuda Y, Kimura H, Ichida Y, Fukushima K, Izumi Y, Sekine T, Suzuku K, Nakajima N, Kazugau A, and Norimatsu M. Expression through hand and finger actions used in ″words in continuous signing″ : A study in reference to the dialogue in Japanese Sign Language by the deaf persons (in Japanese). *IEICE technical report. Speech*, 97(587):pp.81–85, 1961.

[50] Fukuda Y, Torigoe T, Tsuchiya M, and Ichibashi Y. A study on word recognition in jsl (in Japanese). SP94-112, 1995.

[51] Ichida Y. Phonology and numeral system in jsl (in Japanese). *Sign Language Communication*, (57):pp.8–16, 2005.

[52] Ichikawa Y, Fukuda T, and Seki Y. A study on jsl recognition by methodology of cognitive science (in Japanese). Number 22, 1996.

[53] Takata Y, Nagashima Y, Seki Y, Muto D, Lu S, Igi S, and Matsuo H. Analysis of segmented method for japanese sign language recognition (in Japanese). *Human Interface Society*, 2(3):239, 2000.