One-Handed Signs: Standardization for Vehicle Interfaces and Groundwork for Automated Sign Language Recognition

Akihisa Shitara†*, Taiga Kasama*, Fumio Yoneyama* and Yuhki Shiraishi*
†Graduate School of Library, Information, and Media Studies, University of Tsukuba, Japan
Email: theta-akihisa@digitalnature.slis.tsukuba.ac.jp
*Faculty of Industrial Technology, Tsukuba University of Technology, Japan
Email: {yonefumi, yuhkis}@a.tsukuba-tech.ac.jp

***Abstract*—In scenes where d/Deaf and hard of hearing (d/Dhh) individuals drive vehicles, they may face issues reliant on not only environmental sounds but also auditory information through communication. Therefore, we investigated the needs in scenarios where drivers drive in a car and proposed an in-car sign recognition standard using one-handed signs to improve communication issues d/Dhh drivers face. Specifically, we focused on one-handed signs in sign language conversations among d/Dhh individuals and selected sign language expressions based on one-handed signs assuming sign language recognition. We selected one-handed signs used in assumed context scenarios driving car situations where the one-handed sign was likely to occur. Additionally, we conducted surveys with d/Dhh individuals involved to assess whether they found these signs natural and acceptable. We also discussed the annotation rules for annotation labels in datasets intended for sign language recognition.**

Keywords–Deaf and hard of hearing; Sign language; Camera; Sensor glove; Recognition.

I. Introduction

In recent years, the Act for Eliminating Discrimination against Persons with Disabilities and the disqualification provisions stipulated in Article 88 of the Road Traffic Act in Japan have made it possible for d/Deaf and hard of hearing (d/Dhh) individuals to obtain a driver's license. In daily life, d/Dhh individuals can drive vehicles under certain conditions, such as wearing hearing aids or cochlear implants, or using magnifying glasses and displaying hearing impairment markers when not wearing assistive devices.

However, issues remain regarding environmental sounds related to emergency information from emergency vehicles and communication issues. Regarding environmental sounds related to emergency information from emergency vehicles, specific examples include the siren sounds of ambulances, fire trucks, and police cars. Hearing individuals are required to promptly clear the way or stop driving when they hear a siren sound. However, if d/Dhh individuals fail to notice the siren sound, they may unknowingly become involved in issues such as delaying rescue time for ambulances and fire trucks or violating stop orders by police cars.

Furthermore, communication problems arise when d/Dhh individuals engage in sign language conversations while driving. They may look away from the road to the passenger seat or be distracted by the rear seat, increasing the risk of accidents. Additionally, when hearing individuals who do not understand sign language are present, they may have to drive without being able to comprehend any information due to the lack of audio cues.

Therefore, we propose an in-vehicle sign language recognition standard focusing on two aspects: the installation location of drive recorders and the one-handed sign that occurs during sign language conversations between d/Dhh individuals. Moreover, with the practical application of speech recognition, voice user interfaces are also being put into practical use. Research on virtual assistants installed in smart homes and smartphones is particularly notable. However, accessibility issues remain for d/Dhh individuals who have difficulty using voice commands. Nevertheless, sign language recognition has been gaining momentum in recent years, and research on sign language interfaces as an application of sign language recognition is also progressing [1] [2]. Accordingly, as we elevate from our previous work on continuous fingerspelling recognition [3] to sign language recognition, we have established the following research topics:

1 Proposing limited sign language by selecting one-handed signs,

2 Evaluation of the one-handed signs by the parties concerned.

Regarding research topic 1, sign language recognition requires collecting information on the hand shapes, movements, positions of both hands, and facial information. However, for ASL, only short sentences have been studied. Similarly, for Japanese Sign Language (JSL), only short sentences in Signed Japanese (SJ), rather than true JSL, have been researched. Therefore, we focused on one-handed signs, which is also observed in sign language conversations between deaf individuals, and decided to use one-handed signs as a step in the process of elevating from continuous fingerspelling to sign language recognition. As it is difficult to assume situations where one-handed signs are likely to occur in daily life, we specifically focused on the scenario of driving a car, which is relatively easy to envision.

Regarding research topic 2, in most cases of data collection for sign language recognition, the parties concerned are asked to express predetermined sign language words or short sentences. However, to the best of our knowledge, there is no mention of whether this is acceptable and natural for the parties concerned. Therefore, all the authors, including three d/Dhh individuals, selected the content, sign language words, and short sentences spoken in the scenario of driving a car. Additionally, after data collection was completed, the parties concerned were asked to evaluate whether it felt natural and acceptable.

II. Related Work

Previous research on fingerspelling or sign language recognition has proposed two types of sensors for recognizing a series of operations in fingerspelling: contact-type sensor gloves and non-contact-type cameras.

1) Image recognition: Several methods have been proposed for recognizing hand shapes based on processing images of fingerspelling as captured by cameras. Mukai et al. [4] reported that a fingerspelling recognition method targeting 41 immobile characters in JSL resulted in an average recognition accuracy of 86%. They used a classification tree and machine learning based on a support vector machine to classify individual images. Hosoe et al. [5] used deep learning for recognition and achieved a recognition rate of 93%, but only for static fingerspelling. Jalal et al. [6] reported a recognition rate of 99% for American Sign Language (ASL) images based on a deep learning algorithm for static fingerspelling (i.e., excluding "J" and "Z"). However, the recognition accuracy could not be considered as sufficient for practical recognition in JF. Additionally, relatively few recognition results have been reported for dynamic fingerspelling (i.e., fingers moving when expressing a character). In a study of dynamic fingerspelling in JSL [7], the identification of hand shapes was performed using a kernel orthogonal mutual subspace method from images of hand regions obtained from distance images, and the classification of movements was performed using decision trees based on center-of-gravity coordinates. These results yielded a 93.8% identification rate. However, the recognition accuracy was insufficient for the practical recognition required for JF.

2) Sensor glove recognition: Several methods have been proposed for recognizing hand shapes based on measurement data acquired by contact-type sensor gloves. These methods can measure finger flexion, hand positions, and directional data. The measurement data are then sent to a personal computer, and a classification algorithm is used to recognize hand shapes. Cabrera et al. [8] paired the Data Glove 5 Ultra [9] sensor glove with an acceleration sensor to acquire information regarding the degree of flexion of each finger and wrist direction. They conducted test classification using 24 static fingerspelling characters in ASL, excluding "J" and "Z." Their neural network was trained using 5 300 patterns and achieved a recognition rate of 94.07% for 1 200 test patterns. Mummadi et al. [10] prototyped a sensor glove with multiple embedded inertial sensors. They collected French sign language fingerspelling data from 57 people and achieved an average recognition rate of 92% with an F1-score of 91%. Kakoty et al. [11] reported on a dataset of one-handed Indian sign language alphabets (C, I, J, L, O, U, Y, W), ASL alphabets (A to Z), and signed numbers (0 to 9), using a radial basis function with 10-fold cross-validation Using a kernel-supported vector machine, they achieved an average recognition rate of 96.7% and reported that the data were converted to speech. Chong et al. [12] placed six Inertial Measurement Units (IMUs) on the back of the palm and on each fingertip to capture their motion and orientations. Ultimately, 28 proposed word-based sentences in ASL were collected, and 156 features were extracted from the collected data for classification. Using the Long Short-Term Memory (LSTM) algorithm, the system achieved an accuracy of up to 99.89%. Notably, 12 people cooperated with us in the data collection experiment, but whether they were deaf or hearing people was unclear. Yu et al. [13] reported on the architecture of a data glove system comprising a stnm32MCU, flex4.5 bending sensor, mpu6050 six axis sensor, Bluetooth transmission module, and cellphone voice application. The system was developed and connected to a Java-based processing software. They reported that their system recognized sign language movements and could output the words to be said using the intelligent voice system. However, the glove does not feature global movement and rotation tracking. Glauser et al. [14] demonstrated a glove's performance in a series of ablation experiments while exploring various models and calibration methods. However, the glove does not come with a global translation and rotation tracking. Realizing a sign language recognition system requires hand orientations and motions. Among the various methods for performing JF recognition, the conductive fiber braid method [15] uses gloves woven with conductive fibers instead of flexion sensors. These gloves can recognize hand shapes and movements as they are directional gyro sensors incorporated into them. However, the recognition rate for JF ("a," "i," "u," "e," "o") based on Euclidean distance has been reported as only 60%.

3) Data collection: Regarding image recognition, several large-scale Continuous Sign Language Recognition (CSLR) benchmarks have been published [16]. For example, we introduced three large-scale CSLR benchmarks: PHOENIX-2014, Chinese Sign Language (CSL), and PHOENIX-2014-T. PHOENIX-2014 is a publicly available German Sign Language dataset and the most famous CSLR benchmark. This corpus is taken from broadcast news regarding the weather. The CSL dataset consists of 100 sign language sentences and 178 words related to everyday life. Fifty signers performed each sentence, resulting in 5,000 videos in total. A matched isolated CSL database containing 500 words is also provided for pre-learning. Each word was performed 10 times by 50 signers. PHOENIX-2014-T annotates the new videos with two annotations: the sign language terms for the CSLR task, and the German translation for the a Sign Language Translation (SLT) task. The vocabulary consists of 1,115 terms for sign language and 3,000 for German. This dataset is available in [17]. However, the data of these three large-scale CSLR benchmarks are insufficient to realize a highly accurate sign language recognition system using deep learning. Further research is being conducted to increase the amount of available data.

Extensive data for image recognition can be obtained from online sources. For example, the Shi et al. [18] dataset contains clips of fingerspelling sequences cut from sign language "in the wild" videos obtained from online sources such as YouTube and dafvideo.tv [19]. The datasets contain 5,455 training sequences from 87 signers of "ChicagoFSWild," 981 development (validation) sequences from 37 signers, and 868 test sequences from 36 signers, without overlapping signers among the three sets. Another dataset, "ChicagoFSWild+," contains 50,402 training sequences from 216 signers, 3115 development sequences from 22 signers, and 1,715 test sequences from 22 signers. Compared to ChicagoFSwild, the crowdsourcing setup of ChicagoFSWild+ enables the collection of considerably more training data while significantly reducing the efforts of experts and researchers.

Danielle et al. [20] expressed privacy concerns regarding contributing to a filtered sign language corpus, using very

TABLE I. QUESTIONNAIRE OF NEEDS SURVEY

ID	question
1	When conveying information such as navigation to a deaf or hard-of-hearing driver, what do you mainly use?
2	While driving, what content do you want to convey immediately to a deaf or hard-of-hearing driver?
3	While driving, how do you call out to a deaf or hard-of-hearing driver?
4	Have you ever felt anxious when talking to a deaf or hard-of-hearing driver while driving?
5	Have you ever felt difficulty when talking to a deaf or hard-of-hearing driver while driving?
6	How do you communicate with a hearing driver when you want to convey something while driving?
7	How do you communicate with a deaf or hard-of-hearing driver when you want to convey something while driving?
8	Have you ever wanted to use sign language to communicate with a hearing driver while driving?
9	Have you ever felt danger when using sign language communication while driving?
10	How do you call out to a deaf or hard-of-hearing person sitting in the passenger seat or back seat while driving?
11	How do you call out to a hearing person sitting in the passenger seat or back seat while driving?
12	Is there any content you want to communicate with the person in the passenger seat or back seat while driving?
13	Have you ever had difficulties communicating with a person who can use sign language sitting in the passenger seat or back seat while driving?
14	Have you ever had difficulties communicating with a person who cannot use sign language sitting in the passenger seat or back seat while driving?
15	What sign language (including pointing, gestures, classifiers, etc.) do you want to use for operating the car navigation system?
16	Apart from the car navigation system, are there any other occasions where you want to use sign language (including pointing, gestures, classifiers, etc.) to operate the car?

expressive avatars and blurred faces, which may affect the willingness to participate. Training on filtered data may improve the recognition accuracy. In the case of camera recognition, the look of the face is also captured; thus, privacy must also be considered. In contrast, sensor glove recognition does not require pictures of the face; thus privacy concerns are reduced and the data can be more simply collected.

III. NEEDS SURVEY

To understand the needs of d/Dhh individuals when driving a car, we collected data using a web-based questionnaire on Google Forms. The study has been approved by the ethics committee. The total number of respondents was 143, consisting of 88 males, 53 females, and 2 unspecified (mean age: 32.01, standard deviation: 13.87). Among the respondents, 121 individuals possess a driver's license. Regarding their identity, 68 people identified as Deaf, 20 as deaf, 35 as hard of hearing, 4 as hearing, 11 as having never thought about it or unsure, 3 as others, and 2 responses were invalid.

A. Questionnaire and Results

The questions of needs survey are listed in Table I. The results showed that regardless of the d/Dhh individual's identity, more than half of the respondents felt anxiety or difficulty when calling out to a d/Dhh driver while driving. Moreover, when asked about the information they want to convey immediately, over 80% of the respondents mentioned highly urgent information. This suggests that it is problematic to have difficulty conveying information that needs to be communicated immediately, such as danger, caution, or the presence of emergency vehicles.

Compared to Deaf individuals, hard of hearing individuals are more likely to use their voice when calling out to the driver while driving (with a significant difference). This suggests that Deaf individuals require a means of communication that does not rely on voice. Additionally, it was found that more than 75% of d/Dhh individuals wanted to use sign language to communicate with the driver while the vehicle is in motion. Over 75% of the respondents also reported having difficulties communicating with a person who can use sign language while sitting in the passenger seat. Furthermore, some respondents mentioned in the free-response section that they had experienced a sense of danger. However, more than 75% of the respondents indicated that they want to communicate with the person sitting in the passenger seat while driving.

IV. PROPOSAL OF VEHICLE INTERFACES STANDARD

The in-vehicle sign language recognition standard is as follows: Since the driver must always face forward, it is assumed that sign language data will be collected using a camera placed near the location where the drive recorder is installed or by using the drive recorder itself as a camera. The driver may also wear gloves while gripping the steering wheel. Moreover, since the driver needs to grip the steering wheel with at least one hand at a minimum, the sign language is limited to one-handed signs.

A. One-Handed Signs

We selected one-handed sign words and short sentences to be evaluated for sign language recognition, under the rule of not using grammar that employs facial expressions and other elements. Specifically, all authors, consisting of two deaf individuals, one d/Deaf individual, and one hearing individual, discussed and made the selections. It is worth noting that all authors possess a driver's license and have driving experience. The number of words chosen was 58, and the number of short sentences was 40 (Table II). The reason for the selection was based on the needs survey, which included three highly urgent pieces of information: "danger," "watch out," and "emergency vehicle," as well as some directions and small talk.

B. Evaluation

After completing the data collection experiment (refer to Section V-A), we immediately conducted the necessity of in-vehicle sign language recognition and the use of one-handed signs inside the vehicle. The evaluation was conducted with the cooperation of 22 participants, with an average sign language experience of 11.3 years (standard deviation of 7.1 years).

The results regarding the perceived necessity of in-vehicle sign language recognition and the use of one-handed signs inside the vehicle are shown in Figure 1. For the question "Did you feel the need for in-vehicle sign language recognition?", the options were categorized as follows: "Strongly Disagree," "Somewhat Disagree," and "Neither Agree nor Disagree" as negative, and "Strongly Agree" and "Somewhat Agree" as positive. The null hypothesis was set as "Negative and positive options are chosen at a ratio of 0.5 each." A binomial test was conducted, and the results showed that the null hypothesis was rejected at a two-sided significance level of 0.05 with p=0.017, confirming a significant difference. Similarly, for the question "Did you feel the need to use one-handed signs inside the vehicle?", the options were categorized as follows: "Strongly

Copyright (c) IARIA, 2024. ISBN: 978-1-68558-163-3

TABLE II. LIST OF EXAMPLE SENTENCES: ORIGINAL (JAPANESE) AND ENGLISH TRANSLATION

ID	sentence	word1	word2	word3	word4	word5
s1	行きたい [I want to go.]	トイレ(w1) [toilet]	行く(w47) [go]	希望(w33) [want]		
s2	トイレや休憩大丈夫？ [Are restrooms and breaks okay?]	トイレ(w1) [toilet]	休憩(w2) [rest]	希望(w33) [want]		
s3	危険，落ち着いて見て [Danger, watch calmly]	危険(w3) [danger]	落ち着く(w4) [calm]	見る(w13) [look]		
s4	コンビニ，行きますか？ [Are we going to the convenience store?]	コンビニ(w6) [convenience]	行く(w47) [go]	か？(w37) [question]		
s5	あそこで止まってください [Please stop over there]	向こう(w25) [over there]	止まる(w44) [stop]	お願い(w48) [please]		
s6	赤信号だよ [It's a red light]	向こう(w25) [over there]	赤(w8) [red]	信号(w10) [traffic light]		
s7	おーい，青信号だよ [Hey, it's a green light]	手を振る(w11) [call]	青(w9) [blue]	信号(w10) [traffic light]		
s8	右見て [Look right]	右(w14) [right]	見る(w13) [look]			
s9	次の信号で右折 [Turn right at the next signal]	次(w16) [next]	信号(w10) [traffic light]	右(w14) [right]		
s10	速くしてもいいよ [You can speed up]	スピード(w43) [speed]	構わない(w38) [may]			
s11	ゆっくりで [Take it slow]	ゆっくり(w21) [slow]				
s12	次はどこ行く？ [Where are we going next?]	次(w16) [next]	行く(w47) [go]	場所(w17) [place]	何(18) [what]	
s13	次に左折するのはどこ？ [Where do we turn left next?]	次(w16) [next]	左(w15) [left]	場所(w17) [place]	何(w18) [what]	
s14	ここは直進 [Go straight here]	そこ(w34) [there]	直進(w19) [straight]			
s15	高速行くか？ [Shall we take the highway?]	高速(w19) [highway]	行く(w47) [go]	か？(w37) [question]		
s16	ガソリン（残量）大丈夫？ [Is the gas (level) okay?]	ガソリン(w24) [gazoline]	大丈夫(w23) [safe]	か？(w37) [question]		
s17	ガソリンスタンドどこ？ [Where is the gas station?]	ガソリン(w24) [gazoline]	場所(w17) [place]	何(w18) [what]		
s18	別の道があるか調べて [Check if there is another route]	別(w27) [another]	道(w22) [way]	調べる(w26) [search]	お願い(w48) [please]	
s19	ヘッドライトつけて [Turn on the headlights]	ヘッドライト(w30) [headlight]				
s20	ルームライト消して [Turn off the room light]	ルームライト(w31) [roomlight]	消す(w56) [turn off]			
s21	緊急車両が来るから右寄せて [Pull over to the right because an emergency vehicle is coming]	サイレン(w5) [siren]	来る(w32) [come]	右寄せ(w29) [right-aligned]		
s22	コンビニ行きたい [I want to go to the convenience store]	コンビニ(w6) [convenience]	希望(w33) [want]			
s23	コンビニは無い [There is no convenience store.]	コンビニ(w6) [convenience]	無い(w36) [none]			
s24	ガゾリンスタンドはない [There is no gas station]	ガソリン(w24) [gazoline]	場所(w17) [place]	無い(w36) [none]		
s25	向こうで事故かも [There might be an accident over there]	向こう(w25) [over there]	事故(w7) [accident]	らしい(w49) [seem]		
s26	おーい，緊急車両，来ている [Hey, an emergency vehicle is coming]	手を振る(w11) [call]	サイレン(w5) [siren]	来る(w32) [come]		
s27	左に止めても構わない [You can stop on the left.]	左(w15) [left]	駐車(w57) [parking]	構わない(w38) [may]		
s28	それは難しい [That's difficult.]	難しい(w39) [hard]				
s29	ちょっと待って [Wait a moment.]	少し(w40) [little]	待つ(w41) [wait]			
s30	速くないか？ [Isn't that fast?]	スピード(w43) [speed]	過ぎ(w35) [over]	違う(w55) [different]		
s31	任せるよ [I'll leave it up to you.]	任せる(w42) [defer]				
s32	おーい，ウィンカー消して [Hey, turn off the turn signal.]	手を振る(w11) [call]	ウィンカー(w45) [blinker]	消す(w56) [turn off]		
s33	そこの前に停めて [Stop in front of that.]	そこ(w34) [there]	前(w12) [front]	駐車(w57) [parking]		
s34	次の交差点で左折だから，左寄せして [It's a left turn at the next intersection, so stay to the left.]	次(w16) [next]	交差点(w58) [intersection]	左(w15) [left]	左寄せ(w28) [left-aligned]	
s35	1つ目の信号で右折 [Turn right at the first traffic light.]	1つ(w51) [one]	目(w50) [eyed]	信号(w10) [traffic light]	左(w15) [left]	
s36	2つ目の交差点で左折 [Turn left at the second intersection.]	2つ(w52) [two]	目(w50) [eyed]	交差点(w58) [intersection]	左(w15) [left]	
s37	3つ目の交差点で右折だから，右寄せして [It's a right turn at the third intersection, so stay to the right.]	3つ(w53) [three]	目(w50) [eyed]	交差点(w58) [intersection]	右(w14) [right]	右寄せ(w29) [right-aligned]
s38	食べるところ調べて [Look for a place to eat.]	食べる(w54) [eat]	場所(w17) [place]	調べる(w26) [search]	お願い(w48) [please]	
s39	ご飯どうする [What shall we do for food?]	食べる(w54) [eat]	何(w18) [what]			
s40	ハザードランプ消して [Turn off the hazard lights.]	ハザードランプ(w46) [hazard lamp]	消す(w56) [turn off]	お願い(w48) [please]		

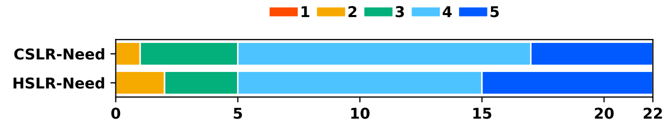

Figure 1. CSLR-Need represents the responses to the question “Did you feel the need for in-vehicle sign language recognition?”
HSLR-Need represents the responses to the question ”Did you feel the need to use one-handed signs inside the vehicle?”
1: Strongly Disagree, 2: Somewhat Disagree, 3: Neither Agree nor Disagree, 4: Somewhat Agree, 5: Strongly Agree.

Disagree,” ”Somewhat Disagree,” and ”Neither Agree nor Disagree” as negative, and ”Strongly Agree” and ”Somewhat Agree” as positive. The null hypothesis was set as ”Negative and positive options are chosen at a ratio of 0.5 each.” A binomial test was conducted, and the results showed that the null hypothesis was rejected at a two-sided significance level of 0.05 with p=0.017, confirming a significant difference.

Additionally, we immediately conducted an evaluation of the one-handed sign words and short sentences that we had selected through discussion. The response data used was from the group that participated in when using the camera on data collection experiment (refer to Section V-A). The number of participants was 10, with an average sign language experience of 11.6 years (standard deviation of 5.76 years). However, we could only evaluate 57 out of the 58 one-handed sign words. Because, due to our oversight, we failed to include the evaluation item related to the one-handed sign word w58 (“交差点 [intersection]”).

The responses to the question “Did you feel any unnaturalness in the one-handed sign expression for each word?” are shown in Figure 2. The options were categorized as follows: “1: Strongly Disagree,” “2: Somewhat Disagree,” and “3: Neither Agree nor Disagree” as positive, and “4: Somewhat Agree” and “5: Strongly Agree” as negative. The null hypothesis was set as “Positive and negative options are chosen at a ratio of 0.5 each.” Using a binomial test, words were selected where the null hypothesis was not rejected at a two-sided significance level of 0.05, and no significant difference was confirmed (p=0.344) when the number of negative responses was 3 or more. The selected words were w2, w4, w6, w7, w10, w16, w22, and w27, totaling 7 words. Subsequently, we proceeded to verify their acceptability.

The responses to the question ”Can you accept the one-handed sign expression for each word?” are shown in Figure 3. The options were categorized as follows: “1: Completely Unacceptable” and “2: Somewhat Unacceptable” as negative, and “3: Neither Acceptable nor Unacceptable”, “4: Somewhat Acceptable,” and “5: Acceptable” as positive. The null hypothesis was set as “Negative and positive options are chosen at a ratio of 0.5 each.” For each short sentence, a binomial test was conducted. The results showed that at a two-sided significance level of 0.05, the null hypothesis was not rejected, and no significant difference was confirmed (In all cases, the number of negative responses was 1 or less).

The responses to the question “Did you feel any unnaturalness in the one-handed sign expression for each short

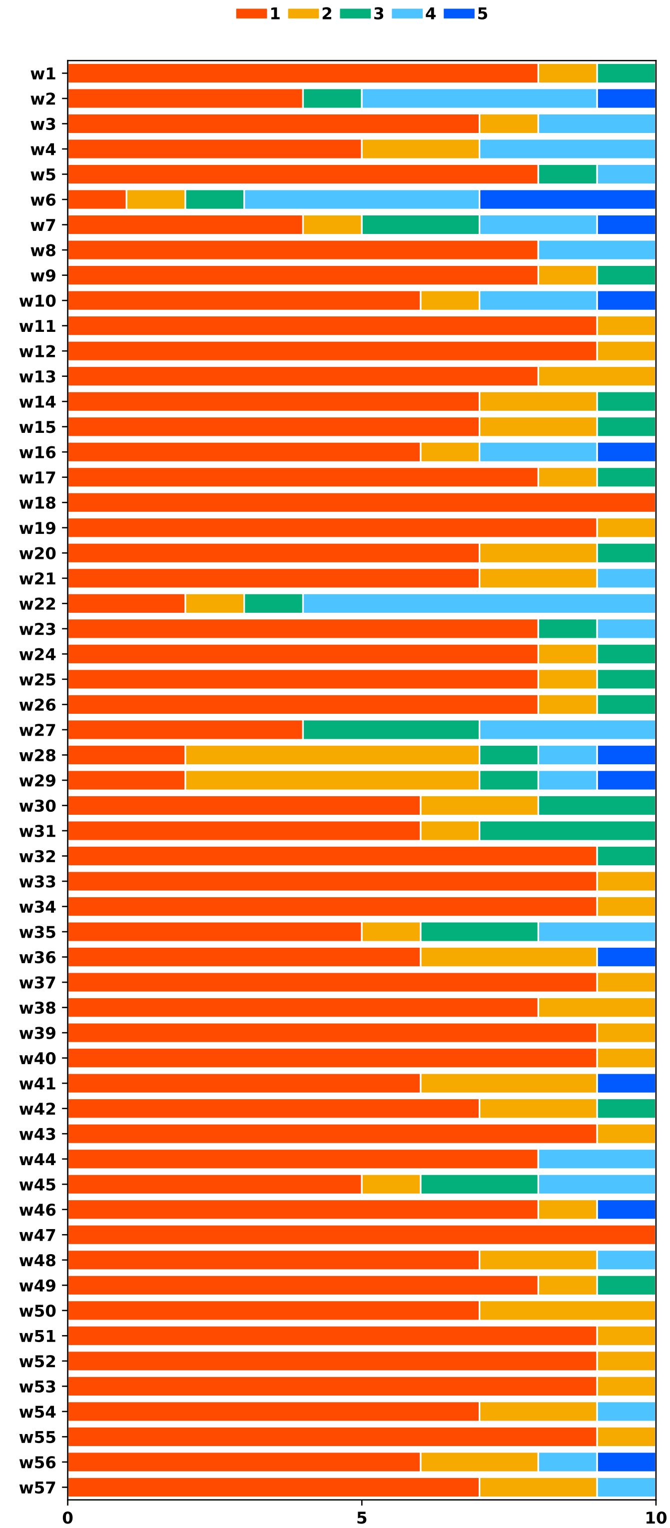

Figure 2. The responses for the question “Do you feel any unnaturalness in the one-handed sign expression?” for each word.
1: Strongly Disagree, 2: Somewhat Disagree, 3: Neither Agree nor Disagree, 4: Somewhat Agree, 5: Strongly Agree.

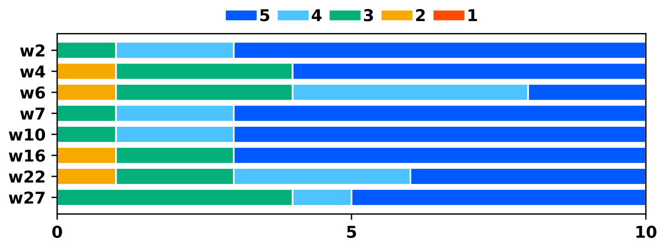

Figure 3. The responses for the question "Can you accept the one-handed sign expression for each word?"
1: Completely Unacceptable, 2: Somewhat Unacceptable, 3: Neither Acceptable nor Unacceptable, 4: Somewhat Acceptable, 5: Acceptable.

sentence?" are shown in Figure 4. The options were categorized as follows: "1: Strongly Disagree" and "2: Somewhat Disagree" as negative, "3: Neither Agree nor Disagree", and "4: Somewhat Agree" and "5: Strongly Agree" as positive. The null hypothesis was set as "Positive and negative options are chosen at a ratio of 0.5 each." Using a binomial test with a two-sided significance level of 0.05, we selected short sentences where the null hypothesis was not rejected, and no significant difference was confirmed (p=0.344) when the number of negative responses was 3 or more. We then proceeded to verify the acceptability of these sentences. The selected short sentences were s3, s4, s9, s12, s13, s18, s21, s24, s27, s32, s34, and s37, totaling 12 sentences.

Subsequently, we proceeded to verify their acceptability. The responses to the question "Can you accept the one-handed sign expression for each short sentence?" are shown in Figure 5. The options were categorized as follows: "1: Completely Unacceptable" and "2: Somewhat Unacceptable" as negative, "3: Neither Acceptable nor Unacceptable," and "4: Somewhat Acceptable," and "5: Acceptable" as positive. The null hypothesis was set as "Negative and positive options are chosen at a ratio of 0.5 each." For each short sentence, a binomial test was conducted. The results showed that at a two-sided significance level of 0.05, the null hypothesis was not rejected, and no significant difference was confirmed. (In all cases, the number of negative responses was 1 or less.)

V. Data Collection and Annotation Rules

To select the one-handed sign data for sign language recognition, we collected data from collaborators and asked them to evaluate the expression of the words. The data used in this data collection experiment is the same as the data evaluated through the questionnaire (refer to Section IV-A).

A. Data Collection Experiment

To simulate the in-vehicle environment, a webcam was placed below the mirror, and the positions of the driver seat and passenger seat were defined in Figure 6. Additionally, assuming that the driver is always facing forward while driving, a whiteboard was used, and the participants were asked to look at the whiteboard. A monitor was placed below the webcam to display the words and short sentences. For each short sentence, data was collected four times, and the process was conducted for both the driver seat and the passenger seat. In this experiment, data was collected from 10 participants using a camera and 12 participants using gloves. When using the camera, participants were asked to continuously express each short sentence four times. However, when using gloves, each short sentence was collected once. The order of expression followed the order of the short sentences. It is worth noting that when using gloves, video data was also recorded for annotation purposes.

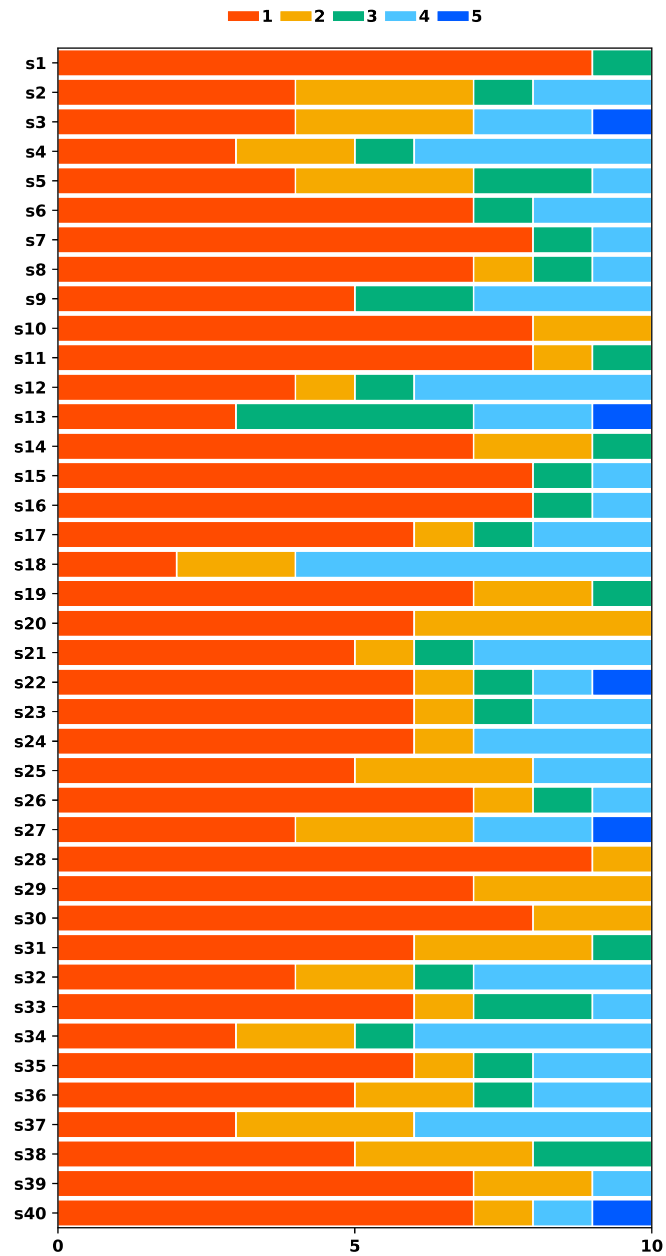

Figure 4. The responses for the question "Do you feel any unnaturalness in the one-handed sign expression?" for each short sentence.
1: Strongly Disagree, 2: Somewhat Disagree, 3: Neither Agree nor Disagree, 4: Somewhat Agree, 5: Strongly Agree.

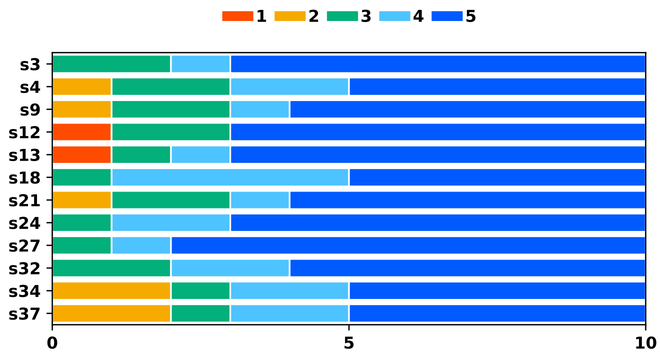

Figure 5. The responses for the question “Can you accept the one-handed sign expression for each short sentence?”
1: Completely Unacceptable, 2: Somewhat Unacceptable, 3: Neither Acceptable nor Unacceptable, 4: Somewhat Acceptable, 5: Acceptable.

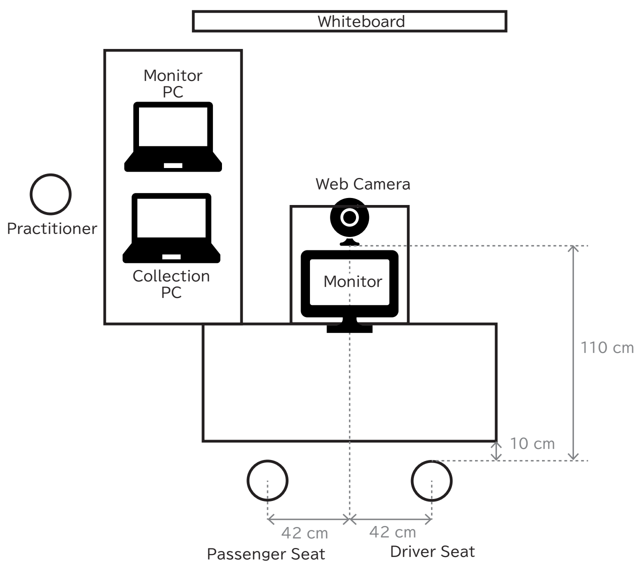

Figure 6. Data acquisition experiment.

B. Annotation Rules

The annotation labels are generally divided into ”word” and ”transition.” The definition of a ”word” is as follows:

For dynamic sign language expressions Start point: The moment when the handshape is determined and reaches the starting position. Even if the handshape is already determined, it is not considered the start point until it reaches the starting position. End point: The moment when either the handshape collapses or the hand moves away from the ending position.

For static sign language expressions Start point: The moment just before reaching the fixed position. (Ideally, the moment of reaching the fixed position is desirable, but it is difficult to judge.) End point: The moment when the hand moves away from the fixed position. (Ideally, the moment of reaching the fixed position is desirable, but it is difficult to judge.)

VI. Discussion

A. Validity of One-Handed Sign Expressions

In this study, as soon as the data collection experiment using cameras was completed, we verified the validity of the expressions by d/Dhh individuals through two questionnaires. However, it is necessary to increase the number of participants and consider participant attributes, such as sign language experience, JSL or SJ, identity, and involvement in deaf school communities to improve the validity and refine the word and sentence expressions.

B. Annotation Rules

“Here” and “There” Initially, we used “there” consistently, but in the short sentences, “there” and ”here” were confused, and the expressions were different, so we decided to distinguish them. “Here”: The handshape is “number 1,” and the expression points to one’s own feet. “There”: The handshape is “number 1,” but since it is a high-context expression in Japanese, it depends on which place is being pointed to. In this case, since the same place is being pointed to, we unified everything into a single ”there” (having only one variation).

“Why” and “Go” Although we did not include ”why” this time, we are concerned that the one-handed “why” expression may overlap with the “go” expression. For future reference, we will describe the differences. “Why/Reason (one-handed)”: While making the handshape “number 1,” twist the wrist downward from one’s chest to the front (aiming for 45 to 90 degrees) twice. “Go”: While making the handshape “number 1,” twist the wrist downward from one’s chest to the front (aiming for more than 80 degrees) once.

“(Number)th” Initially, we tried to distinguish between “(number)” and “th” in “(number)th”. However, this arose from the concern that “(number)th” might be more commonly used in everyday life. Upon checking the videos, we found that “number” and “th” were expressed separately, so we decided to unify each “th” as a single “th”.

C. Differences between Camera and Gloves

In this study, we focused specifically on in-vehicle sign language recognition as one of the one-handed sign interfaces. However, the differences between cameras and gloves, as well as which one is more suitable, have not yet been examined. As a future remaining tasks, it is necessary to evaluate the accuracy of sign language recognition using models trained separately on camera and glove data. Additionally, user studies should be conducted to determine which input interface is preferable for scenarios driving a car.

D. Limitation

In this study, the annotation process has not yet been completed, and the evaluation experiments using recognition have not been conducted. Moreover, as evident from the questionnaire results, it cannot be said with certainty that the validity of the expressions has been ensured by d/Dhh people. Moreover, since facial expressions and other grammatical elements are not targeted, it is assumed to be difficult to recognize JSL (refer to Section IV-A). In addition, the annotation labels are specific to the context of scenarios driving a car; the validity of the annotation labels is not considered high when generalized as everyday context.

VII. Conclusion and Future Works

This study investigated the issues faced by d/Dhh individuals in scenarios driving a car and proposed a solution using one-handed signs to improve communication. The research focused on selecting appropriate one-handed signs based on sign language conversations among d/Dhh individuals and evaluating their validity through surveys with the individuals concerned. The study also conducted a fundamental analysis of one-handed signs in sign language recognition, exploring the potential of using one-handed signs to facilitate the transition from recognizing continuous finger spelling to sign language recognition. The results suggest that introducing one-handed signs, as seen among d/Dhh individuals, could be a viable approach to address communication issues in scenarios driving a car.

Future works are as follows:

1) Expand the dataset: Collect a larger dataset of one-handed signs used in scenarios driving a car to improve the robustness and generalizability of the sign language recognition system.
2) Develop real-time recognition: Implement a real-time one-handed sign recognition system that can be integrated into vehicles to assist d/Dhh drivers in communication and navigation.
3) Conduct user studies: Perform extensive user studies with d/Dhh individuals to evaluate the effectiveness and usability of the proposed one-handed sign recognition system in real-world scenarios driving a car.
4) Refine annotation rules: Further refine the annotation rules for one-handed signs based on feedback from the d/Dhh community and research findings to establish a standardized approach for sign language recognition in scenarios driving a car.
5) Investigate multimodal approaches: Explore the integration of other modalities, such as facial expression or eye-tracking, to enhance the overall communication experience for d/Dhh drivers.

Addressing these future works will advance the assistive technologies for d/Dhh individuals in scenarios driving a car, ultimately improving their communication, safety, and overall driving experience.

Acknowledgment

This work was supported by Woven by Toyota, Inc., JSPS KAKENHI Grant Number #19K11411 and the Promotional Projects for Advanced Education and Research in NTUT.

References

[1] A. Glasser, V. Mande, and M. Huenerfauth, "Accessibility for deaf and hard of hearing users: Sign language conversational user interfaces," in Proceedings of the 2nd Conference on Conversational User Interfaces, ser. CUI '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–3. [Online]. Available: https://doi.org/10.1145/3405755.3406158

[2] T. Kato, A. Shitara, N. Kato, and Y. Shiraishi, "Notification, wake-up, and feedback of conversational natural user interface for the deaf and hard of hearing," International Journal on Advances in Software, vol. 15, 2022, pp. 54–64.

[3] Y. Shiraishi, A. Shitara, F. Yoneyama, and N. Kato, "Sensor glove approach for continuous recognition of japanese fingerspelling in daily life," International Journal on Advances in Life Sciences, vol. 14, 2022, pp. 53–70.

[4] N. Mukai, N. Harada, and Y. Chang, "Japanese fingerspelling recognition based on classification tree and machine learning," in 2017 Nicograph International (NicoInt). New York, NY, USA: IEEE (Institute of Electrical and Electronics Engineers), June 2017, pp. 19–24.

[5] H. Hosoe, S. Sako, and B. Kwolek, "Recognition of jsl finger spelling using convolutional neural networks," in 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), May 2017, pp. 85–88.

[6] M. A. Jalal, R. Chen, R. K. Moore, and L. Mihaylova, "American sign language posture understanding with deep neural networks," in 2018 21st International Conference on Information Fusion (FUSION). New York, NY, USA: IEEE (Institute of Electrical and Electronics Engineers), July 2018, pp. 573–579.

[7] M. Kondo, N. Kato, K. Fukui, and A. Okazaki, "Development and evaluation of an interactive training system for both static and dynamic fingerspelling using depth image," IEICE technical report, vol. 114, no. 512, 2015, pp. 23–28, (in Japanese).

[8] M. E. Cabrera, J. M. Bogado, L. Fermin, R. Acuna, and D. Ralev, "Glove-based gesture recognition system," in Adaptive Mobile Robotics. World Scientific, 2012, pp. 747–753.

[9] "5DT Data Glove 5 Ultra," 2019, URL: https://5dt.com/5dt-data-glove-ultra/ [retrieved: April, 2024].

[10] C. K. Mummadi, F. P. P. Leo, K. D. Verma, S. Kasireddy, P. M. Scholl, and K. Van Laerhoven, "Real-time embedded recognition of sign language alphabet fingerspelling in an imu-based glove," in Proceedings of the 4th International Workshop on Sensor-Based Activity Recognition and Interaction, ser. iWOAR ' 17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 1–6. [Online]. Available: https://doi.org/10.1145/3134230.3134236

[11] N. M. Kakoty and M. D. Sharma, "Recognition of sign language alphabets and numbers based on hand kinematics using a data glove," Procedia Computer Science, vol. 133, 2018, pp. 55–62.

[12] T.-W. Chong and B.-J. Kim, "American sign language recognition system using wearable sensors with deep learning approach," The Journal of the Korea Institute of Electronic Communication Sciences, vol. 15, no. 2, 2020, pp. 291–298.

[13] X. Yu, S. Liu, W. Fang, and Y. Zhang, "Research and discovery of smart dumb gloves," in Journal of Physics: Conference Series, vol. 1865, no. 4. IOP Publishing, 2021, p. 042054.

[14] O. Glauser, S. Wu, D. Panozzo, O. Hilliges, and O. Sorkine-Hornung, "Interactive hand pose estimation using a stretch-sensing soft glove," ACM Transactions on Graphics (TOG), vol. 38, no. 4, 2019, pp. 1–15.

[15] R. Takada, J. Kadomoto, and B. Shizuki, "A sensing technique for data glove using conductive fiber," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, ser. CHI EA ' 19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–4. [Online]. Available: https://doi.org/10.1145/3290607.3313260

[16] H. Zhou, W. Zhou, Y. Zhou, and H. Li, "Spatial-temporal multi-cue network for continuous sign language recognition," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, 2020, pp. 13 009–13 016.

[17] "PWTH-PHOENIX-Weather 2014-T)," 2019, URL: https://www-i6.informatik.rwth-aachen.de/~koller/RWTH-PHOENIX-2014-T/ [retrieved: April, 2024].

[18] B. Shi and et al., "Fingerspelling recognition in the wild with iterative visual attention," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 5399–5408.

[19] "Chicago Fingerspelling in the Wild Data Sets (ChicagoFSWild, ChicagoFSWild+)," 2019, URL: https://home.ttic.edu/~klivescu/ChicagoFSWild [retrieved: April, 2024].

[20] D. Bragg, O. Koller, N. Caselli, and W. Thies, "Exploring collection of sign language datasets: Privacy, participation, and model performance," in The 22nd International ACM SIGACCESS Conference on Computers and Accessibility, ser. ASSETS '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3373625.3417024