# Data Integration, Annotation, and Transcription Methods for Sign Language Dialogue with Latency in Videoconferencing

# Data Integration, Annotation, and Transcription Methods for Sign Language Dialogue with Latency in Videoconferencing

**Mayumi Bono[1&2], Tomohiro Okada[1], Victor Skobov[2], and Robert Adam[3]**
[1] National Institute of Informatics, [2] SOKENDAI (The Graduate University of Advanced Studies),
[3] Heriot-Watt University
[1] 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN, [2] Shonan Village, Hayama, Kanagawa 240-0193 JAPAN, [3] Edinburgh, Scotland EH14 4AS
{bono, tokada-deaf, vskobov}@nii.ac.jp, R.Adam@hw.ac.uk

**Abstract**

This article aims to explain how latency is captured in sign language dialogue via videoconferencing and how recorded data are integrated and annotated using an annotation tool (ELAN). First, we present two examples of the analysis to clarify basic theoretical issues that affect turn-taking via videoconferencing systems focusing on the sequence structure of 'greetings' and 'encounters.' Videoconferencing dialogues often begin with the participants greeting each other, which may be delayed because of the nature of online communication or the technical specifications of each individual's device. Next, to discuss sequential issues with videoconferencing dialogue, we introduce how the fundamental adjacency pair, such as question (first pair part: FPP) and answer (second pair part: SPP), appears to each participant on their computers with latency. This research shows that recording videoconferencing dialogues with latency is useful for next-generation data collection in vision-sensitive sign languages, as well as audio-centred spoken languages with gestures.

**Keywords:** latency, videoconferencing, sign language dialogue

## 1. Introduction

This article aims to explain how latency is captured in sign language dialogue via videoconferencing and how recorded data are integrated and annotated using an annotation tool (ELAN). Since the start of the coronavirus disease 2019 (COVID-19) pandemic, online conferencing has become a part of daily life for many people. This lifestyle change applies to hearing people and Deaf people. How have Deaf individuals, who essentially communicate in three-dimensional space, experienced this shift? To address this question, the present study recorded online conversations between Deaf people using the videoconferencing tool Zoom.

Before the coronavirus disease 2019 (COVID-19) pandemic, Deaf people would meet in so-called Deaf spaces, where they could communicate using sign language—thus, they formed their own society (Kusters, 2015). The pandemic forced Deaf people to meet online, and the Deaf community, which values face-to-face communication, was inspired to extend Deaf space into two-dimensional spaces such as videoconferencing. The long COVID-19 pandemic facilitated human familiarity with and adoption of videoconferencing systems in daily life, resulting in a stable world where Deaf people worldwide can communicate across spatial and distance barriers. Deaf people have been using videoconferencing before COVID-19, and it has been reported that they have unique linguistic and ethnographic ways of integrating such new technologies into their lives (Keating and Mirus, 2003). Before Corona, the Deaf who participated in online communication were a small group of people with strong computer skills, and their use was not stable and continuous. The increase in use and adaptation of online communication in the wake of the coronavirus disaster raises long-term observation needed theoretical questions in Communication Studies regarding the effects on how Deaf people, who have essentially communicated in three-dimensional space, communicate with others in two-dimensional digital space via sign language[1].

In terms of linguistic resources for natural language processing research, videoconference recordings of dialogues could be useful for next-generation data collection. Data recording using videoconferencing systems, which do not require participants to meet in person, will prevent the spread of unknown viruses in the future and allow data recording by people from different regions. For example, the geographic distance between the UK and Japan meant that contact between their respective sign languages was impossible in face-to-face situations. However, now that online communication is commonplace, Deaf people in the UK and Japan can meet more easily and frequently than before.

Here, we report the preliminary results of part of the 3-year international joint project 'Understanding cross-signing phenomena in video conferencing situations during and post-

[1] There are already projects documenting the experiences of Deaf communities in the time of COVID-19 for American Sign Language.

https://doi.org/10.6084/m9.figshare.22340830.v1



*Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages*, pages 26–35

COVID-19 in rural areas' [2] between the United Kingdom (UK) and Japan, which began in 2022. The goal of this project is to observe online cross-signing phenomena among non-shared language situations (Bono and Adam, 2023); it consists of two phases. In the first phase (2022/23), data collection was conducted in the respective countries (UK and Japan) using videoconferencing systems. During the second phase (2023/24), Deaf people in Japan and the UK, who do not have a shared sign language, will meet and interact with each other through a videoconferencing system.

In this article, we describe data integration, annotation, and transcription methods for video clips with videoconferencing-specific latency that were designed by the Japanese team during the first phase. First, we present two examples of the analysis to clarify basic theoretical issues that affect turn-taking via videoconferencing systems focusing on the sequence structure of 'greetings' and 'encounters.' Videoconferencing dialogues often begin with the participants greeting each other, which may be delayed due to the nature of online communication or the technical specifications of each individual's device. To discuss theoretical issues with videoconferencing dialogue, we introduce how the fundamental repair sequence, such as question and answer, appears to each participant on their local computers with latency. This research helps to show that recording videoconferencing dialogues with latency is useful as next-generation data collection for vision-sensitive sign languages, as well as audio-centred spoken languages with gestures.

Section 2 describes the methods used to process the delays; section 3 gives an overview of the data collection; and section 4 demonstrates the actual qualitative analysis of the data. This paper is the first report to show how latency is essential for qualitative analysis research on online sign language dialogues.

## 2. Latency in Videoconferencing

From a technical perspective, many videoconferencing systems seek lower latency to more closely resemble in-person conversations. However, depending on internet speeds and computer specifications, latency may be high in an individual's home. Many sociological and conversation analytical studies of video-mediated interactions have focused on the lack of shared space in conversations that occur via videoconferencing systems (Heath and Luff, 1993). Even in spoken conversation, if the space is not shared, it becomes difficult to use gestures such as eye contact and pointing, which can typically be used without difficulty during face-to-face interactions. In the aftermath of the COVID-19 pandemic, Seuren et al. (2021) observed a remote medical interview conducted using Skype [3], which had been the predominant videoconferencing platform before COVID-19—rather than Zoom [4]—using the Conversation Analysis (CA) method. They concluded that conversation participants communicating via videoconferencing platforms behave as though they inhabit a shared reality.

We believe that two issues must be considered here. The first issue is the importance of latency in interactions such as medical counselling, where the goal is 'solving' or 'curing' a problem. During social interactions, in which the explicit goal is achievement of the objective regardless of latency or transmission problems, these problems may be tolerated if the goal is achieved. The second issue arises in situations where Deaf people use videoconferencing systems. When hearing people use videoconferencing systems, they have the option to cease using the video component if latency or video outages occur; however, Deaf individuals do not have that option. Additionally, Zoom has a function that—if the audio transmission ceases for a certain period of time—allows users to increase the audio speed and transmit all speech that can be understood and heard. Conversely, Zoom does not have a function to reduce the video frame rate and transmit language-understandable and readable video in a single transmission. Thus, when latency or video outages occur, the Deaf person must be able to clearly resolve these troubles so that they can follow the conversation.

The 'greeting' and 'encounter' situations in online communication are the first places where latency due to the recipient's internet environment and personal computer specifications can be identified. If latency in the recipient's video transmission is recognised, it will be necessary for the speaker to consider such latency. When discussing delays in online communication, it is important to discuss this system-induced trouble, which can be termed 'basal latency'. Basal latency results in different ways of viewing sequence organisation between oneself and others in a videoconferencing dialogue. In this paper, we focus on basic adjacency pairs such as question (first pair part: FPP) and answer (second pair part: SPP) and raise theoretical issues regarding sequence organisation in CA (Schegloff, 2007).

## 3. Data Collection

The details of data collection during the first phase have been published elsewhere (Bono and Adam, 2023). This section introduces the method of data collection, focusing on latency and

[2] https://www.ukri.org/news/uk-japanese-collaboration-to-address-covid-19-challenges/

[3] https://www.skype.com/en/

[4] https://zoom.us/

components of the analysis detailed in Sections 5 and 6. Participants were selected from three geographically distant regions in Japan: Hokkaido, Shikoku, and Okinawa. Three participants were selected from each of the abovementioned regions, and then divided into groups A, B, and C for each region (see Table 1). Dialogue pairs were composed of one participant from each group and the other participant from one of the remaining two groups—for example, the 'Hokkaido (HK) and Shikoku (SK) pair', the 'SK and Okinawa (ON) pair', and the 'HK and ON pair' in Group A.

| Region | Group A /ID | Group B /ID | Group C /ID |
|---|---|---|---|
| Hokkaido | HK-A | HK-B | HK-C |
| Shikoku | SK-A | SK-B | SK-C |
| Okinawa | ON-A | ON-B | ON-C |

Table 1: Regions, groups, and identifications (IDs)

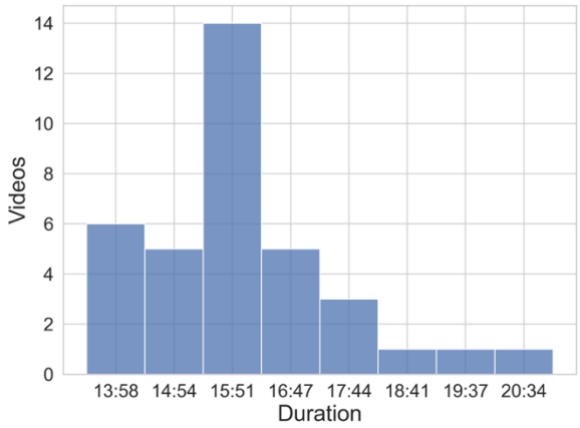



Figure 1: Video Duration Distribution

The online dialogue was recorded locally on the participants' computers using the recording function in Zoom at three sites: the locations of both participants and the monitoring staff (Zoom host). Using the 'hide non-video participants' feature in Zoom, the monitoring staff faded from the Zoom view of the participants as the conversation/experiment began. However, the monitoring staff actually participated in the Zoom call to gauge and monitor the participants' dialogues. The reason for recording at each site was to avoid missing any discussion of latency issues during online communication that might have affected the turn-taking process (Seuren et al., 2021). By recording at three sites, it was possible to process and analyse the timings of various communication phenomena; this allowed the researchers to determine how each participant saw their recipient's image and identify any differences in the way they might subsequently view each other.

After the monitoring staff member turned off their camera and appeared to have left the session, the participants commenced their online dialogue. At the appropriate time, as the conversation was ending (e.g., as indicated by topic shifts; approximately 15 minutes), the monitoring staff member would turn on their camera to terminate the ongoing dialogue. Figure 1 illustrates the distribution of the video durations, showing that most dialogues concluded within approximately 15 minutes but sometimes continued for up to 20 minutes.

## 4. Latency in Analysis

Latency has a noticeable impact on the conversation process: a certain degree of latency can make the conversation impossible. Thus, this study tracked latency during the data collection process.

### 4.1 Capturing Latency in Zoom

Latency has a noticeable impact on participant satisfaction with the conversation process. If the delay reaches 400 ms, the conversation will become unacceptable for participants (ITU-T, 1996). Garg et al. (2022) reported that participants were able to adapt to higher latency, but they exhibited increased fatigue and frustration associated with higher cognitive load during visual tasks. In the context of data collected from sign language dialogues held via videoconferencing, latency tracking and reporting are essential for future conversation analyses.

The built-in tools for latency tracking and reporting in Zoom have an ambiguous description [5] and unclear export capabilities; a requirement for participants to use these tools would add unwanted complexity to the recording process. For post-collection latency measurement, we chose a three-way setup—two participants and a monitor—as shown in Figure 2.

Using this setup, the delay between the two participants could be fully observed only by a monitoring party. The observation was also shifted along the absolute timeline because the observer had its delay. Nonetheless, this observation added context to each participant's recordings, allowing us to synchronise them within the absolute timeline.

Zoom has a function that allows conversations to be recorded and stored in the cloud or in the local memory. The difference between the two options is crucial: if a participant records to the cloud, a

[5]Available at: https://support.zoom.us/hc/en-us/articles/202920719-Accessing-meeting-and-phone-statistics

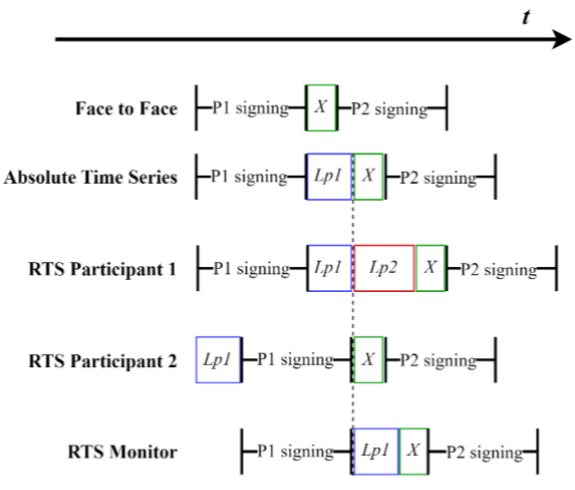



Figure 2: Three-way latency conversation schema, adapted with permission from Hosoma and Muraoka (2022)

delay will be added to his camera view, and the video quality will be reduced. Recordings stored in local memory have superior quality and no delay; thus, this storage is a critical requirement for post-collection latency computation.

The synchronisation is performed by calculating the time shift between the participants' and monitor's records. The participants' recordings are trimmed accordingly, after which they begin simultaneously in the absolute timeline and are effectively synchronised with the monitor's record. They may then be used to measure latency between participants.

The latency and synchronisation time shifts were calculated using cross-correlation within SciPy[6]. For this purpose, we reduced each video to a one-dimensional signal by calculating the Euclidean distance between each frame and an average frame of the entire video.

Participants' recordings were compared with the received version in the other recordings. Each corresponding piece of the frame with the participant's view was cropped to the view size prior to calculation. For synchronisation with the monitor's record, we collected a small portion at the same video position (250 frames). A sliding window of 120 frames was used to determine participant latency at each frame.

### 4.2 ELAN Integration

ELAN Software, which is used to annotate the sign language corpus, has a built-in function that allows time series to be displayed along the video timeline. We utilised this functionality to display the calculated latency in the recordings, as illustrated in Figure 3. The output facilitates comprehension of the delay and reaction time.

Delayed annotations may be created by adding latency to the start and end times in the annotation within the absolute timeline. This addition may be done automatically using the Python *pympi-ling* module.

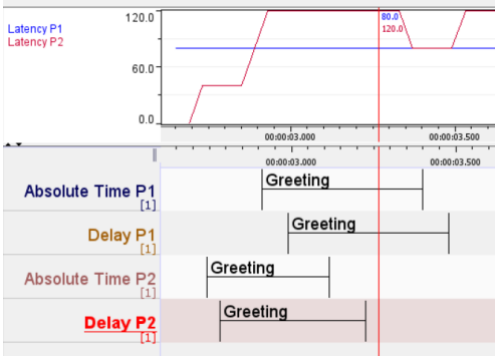



Figure 3: Latency display in ELAN

## 5. Analysis of "Greetings" and "Encounters" in Videoconferencing

### 5.1 Analysis 1: Sequential "Hi" or floating "Hi" (Hokkaido–Okinawa)

Analysis 1 focuses on dialogue of greeting scenes between Hokkaido and Okinawa in Group B (hereafter HK for the Hokkaido participant and ON for the Okinawa participant). Observing the results annotated with ELAN in Figure 4, [7] a sequential relationship can be identified in the monitoring view (recorded in Tokyo) and the Hokkaido view, where HK says 'Hi' first; ON then responds, 'Nice to meet you'. [8] Conversely, in the Okinawa view, it appears that ON said the words 'Nice to meet you' first, whereas HK said 'Hi' almost simultaneously (with a delay of

[6] Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlation_lags.html#scipy.signal.correlation_lags

[7] This ELAN annotation is a preliminary step before the ELAN integration adjustment method is applied based on absolute time, as described in Section 4.2. In this context, M-view means monitoring view, HK-view means Hokkaido view, and ON-view means Okinawa view on the ELAN tiers' names. Because the absolute time has not been adjusted, analysis between the different participant's views is impossible. Therefore, we compare the results between the same participant's views.

[8] Schegloff (2007) does not apply the concept of adjacency pairs to greeting sequences, so we follow this here and describe them as a 'sequential relationship' rather than adjacency pairs. We describe the concept of adjacency pairs in Section 6 more detail.

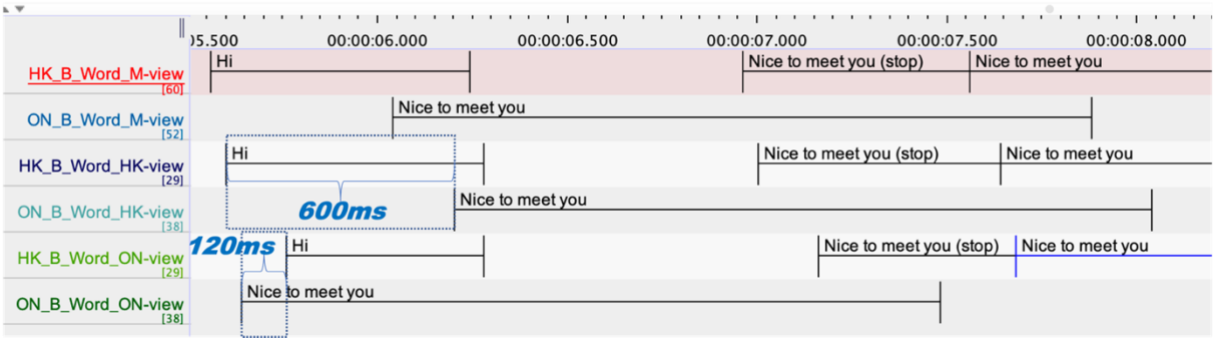



Figure 4: Analysis 1: Sequential "Hi" or floating "Hi" (Hokkaido–Okinawa)

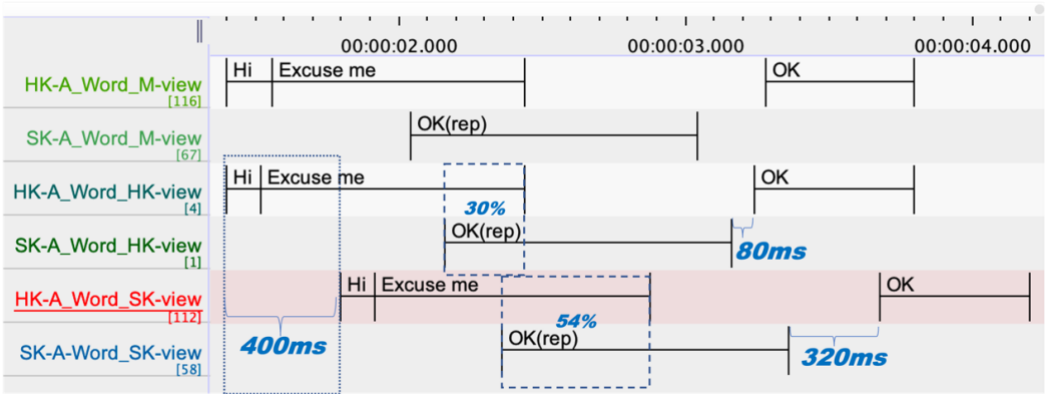



Figure 5: Analysis 2: Showing a positive attitude (Hokkaido–Shikoku)

approximately 120 ms from the 'Nice to meet you' by ON).

Next, in the transcript based on the CA notation, we described these differences (Excerpts 1 and 1'). [9] Theoretically, in CA, the monitoring and Hokkaido views indicate that the 'Hi' uttered by HK is in a sequential relationship with the 'Nice to meet you' next uttered by ON (see lines 01 and 02 in Excerpt 1). Conversely, in the Okinawa view, HK's 'Hi' completely overlaps with ON's 'Nice to meet you'. In this scenario, it appears that ON's 'Nice to meet you' is uttered first; HK then responds, 'Nice to meet you' (see lines 01 and 03 in Excerpt 1'). Accordingly, HK's 'Hi' is considered to be floating in the sequence structure. Subsequently, it appears that HK says 'Nice…Nice to meet you' in line 03 following and imitating ON's greeting in both Excerpts 1 and 1'. Thus, in a dialogue occurring via videoconferencing, HK and ON may hold completely opposite perceptions of who issued the first greeting.

As a part of the ELAN annotation in Figure 4, HK should feel that ON is responding 600 ms after the onset of his 'Hi' utterance. However, ON would have felt as though he had initiated his salutatory utterance 120 ms earlier than HK's 'Hi'. Simply adding these together, ON's salutatory utterance is conveyed to HK with a delay of 720 ms. How does a delay of > 0.7 seconds (sec) affect the interaction? Analysis 2 continues the observation by examining another case.

**Excerpt 1 (Monitoring and Hokkaido view)**
01 HK: H[i
02 ON: [Nice to [meet you
03 HK: [Nice…Nice to meet you

**Excerpt 1' (Okinawa view)**
01 ON: Ni[ce] to meet [you
02 HK: [Hi]
03 HK: [Nice…Nice to meet you

## 5.2 Analysis 2: Showing a positive attitude

The data examined in Analysis 2 are derived from the beginning of the third dialogue experiment (Figure 5). It is an encounter, rather than a greeting, and HK initially apologises for his own connectivity problems. Similar to the data in Analysis 1, there is minimal latency between the monitoring view (recorded in Tokyo) and the Hokkaido view, but the recipient's video transmission exhibits latency in the Shikoku view.

Simple observation of the beginning of 'Hi' uttered by HK in the Hokkaido view and Shikoku view indicates a basal latency of 400 ms between them. Further analysis reveals that SK's 'No worries (OK (rep) [10])' overlaps with the final 30% of HK's

[9] The transcript of Excerpt 1 does not use the word glosses of the signs separated by slashes because this analysis does not aim to show temporal relations; it uses the Japanese translation.

[10] The signal of (rep) added after the word gloss means that the sign expression is repeated. Thus, [OK] is repeated several times here.

‘Excuse me’ utterances (‘Excuse me’: duration 920 ms, overlap time: 280 ms) in the Hokkaido view. In the Shikoku view, this percentage increases to 54% (‘Excuse me’: duration 960 ms, overlap time: 520 ms). SK’s action in the Shikoku view, which overlaps by more than 50% with HK's utterance and responds to it, may be assumed to indicate a positive attitude towards the recipient. Accordingly, SK is repeatedly and quickly expressing ‘No worries’ to HK.

After SK's reply with repeated OK, HK closes the sequence by saying ‘Alright’ (sequence-closing 3rd). However, there is another difference between the Hokkaido and Shikoku views: in the Hokkaido view, HK closes SK's ‘No worries’ with ‘Alright’ without a pause (after a short gap of 80 ms). Conversely, in the Shikoku view, the transmission of HK’s ‘Alright’ is delayed, and the sequence appears to terminate after a lengthy pause of 320 ms. Although this difference is minor, subtracting the actual gap of 80 ms from 320 ms results in a latency of 240 ms, indicating that HK’s response, ‘Alright’ (sequence-closing 3rd), was not transmitted at the appropriate time. Thus, the influence of basal latency is present in these interactions. In the Shikoku-view, because of latency caused by the system, HK’s reaction in line 04 has a weak relationship with the previous sequence, which is also floated from the fundamental sequence organisation.

**Excerpt 2 (Hokkaido view)**
01 HK: Hi/Excuse-[me/ (*Hi, Excuse me*)
02 SK: [OK (rep) (*No worries*)
03 *(gap: 80 ms)*
04 HK: OK (*Alright*)

**Excerpt 2’ (Shikoku view)**
01 HK: Hi/Excu[se-me/ (*Hi, Excuse me*)
02 SK: [OK (rep) (*No worries*)
03. *(long pause: 320 ms)*
04 HK: OK (*Alright*)

In Excerpts 2 and 2’ formed as a CA transcript, Excerpt 2 in the Hokkaido view sequentially appears better than Excerpt 2’ in the Shikoku view; SK’s response in line 02 terminally overlaps HK’s apologies in line 01. Then, after an 80 ms gap, HK expresses ‘Alright’ (sequence-closing 3rd). In Excerpt 2’, however, SK gives responses in line 02 with a positive attitude; there is no rapid sequential feedback from HK. In summary, this encounter is smooth for HK, whereas it is slightly awkward for SK.

## 6. Analysis of Sequence Organisation with Latency

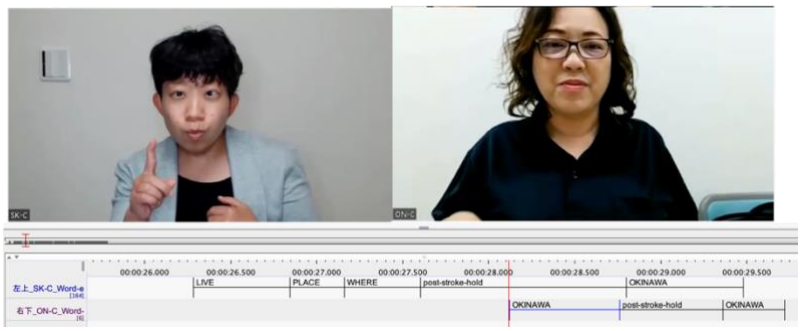
As mentioned in footnote 7, Schegloff (2007) does not apply the concept of adjacency pairs to a sequence of greetings. Therefore, we should not analyse greetings or encountering; we should focus on the contents of the conversation sequence after greetings to understand what occurs in an online dialogue with latency from the perspective of sequence organisation.

In Analysis 3, we focus on differences in the appearance of a simple question–answer adjacency pair between the two views. Analysis 4 shows how the theoretical issues raised in Analysis 3 may be treated in terms of the repair sequence (Kitzinger, 2013; Schegloff et al., 1977).

Recently, several researchers, mainly the language and cognition research group at the Max Planck Institute, have applied comparative and quantitative analysis to repair sequences, especially other-initiated repair (OIR), in several languages as a universal and fundamental system of human communication that transcend differences across cultures and communication modality, in spoken, signed, and tactile conversations (Bono et al., 2023; Byun et al., 2018; Dingemanse and Enfield, 2015; Dingemanse, Kendrick and Enfield, 2016; Dingemanse, Torreira and Enfield, 2013; Floyd et al., 2016; Haakana et al., 2021; Hayashi et al., 2013; Kendrick, 2015; Manrique and Enfield, 2015; Manrique, 2016). This article focuses on more fundamental issues on CA such as adjacency pairs in Analysis 3, and self-initiated self-repair sequence not OIR in Analysis 4.

### 6.1 Analysis 3: Question-answer adjacency pairs

The data in Figures 6 and 7 were obtained from the first session, 26 s after the beginning. SK asks ON, LIVE/PLACE/WHERE, ‘where do you live?’ with questioning facial expressions. After the question, she maintains her hand shape and holds it in signing space, which is annotated as ‘post-stroke-hold’, while looking at the recipient. The concept of post-stroke-hold arises from Gesture Studies (McNeill, 1996; Kita et al., 1998; Kendon, 2004). In spoken conversation, post-stroke-hold functions to hold a topic in discourse, whereas it has several grammatical functions in sign language. Here, SK holds the conversational floor and connects her sequence-closing third, OKINAWA ‘Okinawa (I see)’, to line 03 in Figures 6 and 7. Sequence-closing thirds (SCTs) are placed in the third position of question-answer adjacency pairs by the person who asks a




01 SK-C: Where do you live?
(0.5 s gap)
02 ON-C: (I live in) Okinawa
03 SK-C: Okinawa (I see)
04 ON-C: (I live in) Okinawa

Figure 6: Q–A adjacency pair (Shikoku view)


01 SK-C: Where do you live?
(0.2 s gap)
02 ON-C: (I live in) Okinawa
(0.4 s gap)
03 SK-C: O[kinawa (I see)
04 ON-C: [Okinawa

Figure 7: Q–A adjacency pair (Okinawa view)

question to evaluate the answer provided by the interlocutor and close the current adjacency pair.

During SK's post-stroke-hold, ON answers, OKINAWA, '(I live in) Okinawa'. There is a difference in the gap before answering between the views of Shikoku and Okinawa. In the Shikoku view, the gap is 0.5 s, whereas it is 0.2 s in the Okinawa view. We do not consider this a large difference in an adjacency pair.

The theoretical issue is the explanation for repetition of ON's answer in line 04. In the Okinawa view (Figure 7), the explanation is visible in SK's SCT, 'Okinawa (I see)', which arrives slightly later. There is a 0.4-s gap between

lines 02 and 03. Consequently, ON repeats the answer in line 04. We add more detailed sign movement annotations, prep (preparation), str (stroke), and ret (retraction) to SK's SCT (Kikuchi and Bono, 2013). From the detailed annotations, we observe that when ON begins the repetition, SK continues to prepare for OKINAWA as the SCT. In this context, we consider SK's reaction to ON's answer to be slightly delayed; subsequently, ON repeats her answer again in the Okinawa view. This is an example of self-initiated self-repair by ON (Schegloff et al., 1977; Kitzinger, 2013). ON notices her answer is not conveyed to the recipient, then tries her answer again.

In contrast, in the Shikoku view, SK's reaction is less delayed. SK begins the SCT, 'Okinawa (I see)', immediately after ON's answer. There is no gap here. This is the shortest time to close the sequence. Our question here is how ON's repetition in line 04 appears to SK.

First, some sign language linguists insist that repetitions constitute a form of grammar, such as stress in sentence, for Deaf people (Covington, 1973). A repetition in answer position appears to be part of the answer to the question; thus, ON does not place any emphasis on her answer by repeating it. Second, we observe that ON tends to repeat some expressions in the overall data. It is possible that the repetition is her signing characteristic. We plan to conduct more quantitative analysis comparing other signers in our corpus.

In Analysis 4, we discuss online-communication-specific issues related to the repair sequence in ON's repetition.

### 6.2 Analysis 4: Self-initiated self-repair for a frame-out issue

Figure 8 shows one of the dictionary forms of OKINAWA. In line 02 of Figure 6 and Figure 7, ON's two fingers for answering OKINAWA '(I live in) Okinawa' are frame-out, as shown in Figure 9. Her signing scale is excessively large. In line 04 of Figure 6 and Figure 7, ON reduces her signing scale. This is a successful frame-in, as shown in Figure 10. As the evidence that ON consciously modified her signing scale, after the question–answer adjacency pair, she adjusts the camera position to be captured the upper space of her signing.

This is an example of self-initiated self-repair. In an in-person setting, this type of repair initiation related with frame-out issue does not occur,


Figure 8: An example of dictionary form of OKINAWA (English translation of caption: Hold up the index and middle fingers of the right hand and twist upwards from the temple.) Japanese Federation of the Deaf (2010: 242)

Figure 9: OKINAWA (frame-out, big)

Figure 10: OKINAWA (frame-in, small)

because the signing space is completely opened to between signer and recipients. In online communication, signers monitor how their own signings are viewed by recipients. Occasionally, the signings are frame-out and should be adjusted. This is an online-specific phenomenon.

In the Okinawa view of Figure 7, ON's modification matches as the second pair part (SPP), answering, of the question–answer sequence because SK's SCT in line 03 is delayed. So, line 03 and line 04 are produced almost simultaneously. In the Shikoku view of Figure 6, however, ON's repetition is not placed the second pair part. because SK's SCT in line 03 is not delayed. Because of that, ON's repetition in line 04 floats from the ongoing conversational sequence.

In addition, we notice that some Deaf people tend to increase repetition in online communication more than in-person communication in some small observations of our data-set. At this moment, we plan to compare this type of phenomenon in online and in-person quantitatively for future works.

## 7. Discussion

Levinson (2016) modelled the cognitive mechanisms of turn-taking in everyday human conversation. He estimated intervals of 200 ms to conceptualise one's thoughts, 75 ms to retrieve the lexicon, and 325 ms to encode the form before taking a turn to speak for a total of 600 ms. However, when the timing of turn-taking was measured from actual linguistic data collected worldwide, the start of the response turn was normally distributed with a peak approximately 200 ms after the end of the recipient's turn. He points out that to achieve this, humans plan their own speech production while anticipating their opponent's speech; they also anticipate the end of the turn and follow signals that provide clues to the end of the turn.

Our research question is as follows: What changes would ensue if videoconferencing systems were introduced to the turn-taking process supported by the highly organised human cognitive mechanisms? This is a general question that is common to both spoken dialogues and signed dialogues occurring via videoconferencing systems. Future studies of online communication should consider how recipients accept system-induced latency when basal latency occurs, and how they subsequently interact with each other. Online sign language interaction is an ideal research target to approach this problem because it uses only a video channel without a speech channel.

A limitation of this study is that it is difficult to ascertain whether and how the conversation participants themselves notice and perceive the minute differences in the conversation sequence due to this latency. However, conversation analysis is a research method that analyses how the other party followed the next action in response to a previous action in order to understand the state of awareness of the conversation participants themselves, etc. We will continue to collect data and propose a theory of turn-taking and repair sequences in online communication.

## 8. Conclusion

The technological development of videoconferencing systems, such as Zoom, prioritises the enhancement of usability primarily for hearing people. However, some usability innovations have also been implemented to support the Deaf minority. Although

videoconferencing systems and everyday conversations are not required to be completely equivalent, phenomena including which participant 'greets' the other first or reactions that convey a positive attitude towards the other's utterance, and how the repetitions appear to the remote recipient, as demonstrated in this article, can be significantly inhibited by latency. The sense of accomplishment and satisfaction during a conversation is obtained through a series of interactions with the recipient. We hope that analyses of this nature will be utilised in future efforts to develop video transmission technology.

Thus far, we have merely established the data collection method and data annotation environment. In future studies, we intend to qualitatively and quantitatively analyse the recorded data, then continue the exploration of how Deaf people living in the visual world were forced to confront communicative and cognitive challenges during the COVID-19 pandemic.

## 9. Acknowledgements

We thank our research collaborator; Dr. Ryosaku Makino, who developed the concept of basal latency in online communication with us; Dr. Keiko Sagara, who took part as an interviewer in some sessions; and Prof. Yutaka Osugi, who connected us with the coordinators; the coordinator-in chief, Ms. Megumi Kawakami; area coordinators, Mr. Kazuhiro Naka (Hokkaido), Mr. Ryuji Kondo (Shikoku), and Ms. Eriko Shiroma (Okinawa); and the nine Deaf participants. This study is part of a wider UK–Japanese social science and humanities project seeking to address global challenges presented by the COVID-19 pandemic, with support from the Japan Society for the Promotion of Science (JSPS) International Joint Research Programme JRP-LEAD and UKRI (UK Research and Innovation, UK).

## 10. Bibliographical References

Bono, M., Sakaida, R., Ochiai, K., and Fukushima, S. (2023) Intersubjective Understanding in Finger Braille Interpreter-mediated Interaction: Two Case Studies of Other-initiated Repair. *Lingua*, Elsevier. https://doi.org/10.1016/j.lingua.2023.103569

Bono, M., and Adam, R. (2023) Online cross-signing project between the United Kingdom and Japan: First phase of data collection, *Online Proceedings of JSAI-isAI2023*. (published to only conference audience)

Byun K., de Vos, C., Bradford, A., Zeshan, U., and Levinson, S. C. (2018). First encounters: Repair sequences in cross-signing. *Topics in Cognitive Science*, 10(2), 314-334. https://doi.org/10.1111/tops.12303

Covington, V. C. (1973). Features of stress in American Sign Language. *Sign Language Studies*, 2, 39–50. https://doi.org/10.1353/sls.1973.0017

Dingemanse, M. and Enfield, N. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics*, 1(1), 96-118. https://doi.org/10.2478/opli-2014-0007

Dingemanse, M., Kendrick, K. H., Enfield, N. (2016). A coding scheme for other-initiated repair across languages. Open Linguistics, 2(1), 35-46. https://doi.org/10.1515/opli-2016-0002

Dingemanse, M., Torreira, F. and Enfield, N. (2013). Is "Huh?" a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. PLOS ONE, 9(4):e94620. https://doi.org/10.1371/journal.pone.0094620

ELAN (Version 6.6). (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Floyd, S., Manrique, E., Rossi, G., and Torreira, F. (2016). Timing of visual bodily behavior in repair sequences: Evidence from three languages. *Discourse Processes*, 53(3), 175-204, https://doi.org/10.1080/0163853X.2014.992680

Garg, S., Srivastava, A., Glencross, M. and Sharma O. (2022). A study of the effects of network latency on visual task performance in video conferencing. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3491101.3519678

Haakana, M., Kurhila, S., Lilja, N., and Savijärvi, M. (2021). Extending sequences of other-initiated repair in Finnish conversation. In Lindström, J., Laury, R., Peräkylä, A., and Sorjonen, M. (Eds.), *Intersubjectivity in Action: Studies in language and social interaction*. John Benjamins, 231-19. ISBM-10: 9027209405, ISBM-13: 978-9027209405

Hayashi, M., Raymond, G., and Sidnell, J. (2013). *Conversational repair and human understanding*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511757464

Heath, C. and Luff, P.K. (1993). Disembodied conduct: interactional asymmetries in video-mediated communication, G. Button (Ed.), *Technology in Working Order: Studies of Work, Interaction, and Technology*, Rank Xerox Research Centre, London, UK, 35-54.

Hosoma, H. and Muraoka, H. (2022). How can latency in telecommunication affect action sequence analysis? *The Japanese Journal of Language in Society*, 25(1), 230-237. https://doi.org/10.19024/jajls.25.1_230

International Telecommunication Union (ITU), Telecommunication Standardization Sector

(ITU-T) One-way transmission time. Recommendation G.114, (05/2003).
Japanese Federation of the Deaf (2010). *Watashi tachi no shuwa: Gakushu jiten (Our sign language: Encyclopedia for learning JSL)*, Japanese Federation of the Deaf Publisher.
Keating, E., and Mirus, G. (2003). American Sign Language in Virtual Space: Interactions between Deaf Users of Computer-Mediated Video Communication and the Impact of Technology on Language Practices. *Language In Society*, 32(5), 693–714. http://www.jstor.org/stable/4169299
Kendon, A. (2004). *Gesture: Visible Action as Utterance*, Cambridge University Press.
Kendrick, H., K. (2015). Other-initiated repair in English. *Open Linguistics*, 1, 164-190. https://doi.org/10.2478/opli-2014-0009
Kikuchi, K., and Bono, M. (2013). Sougokoui ni okeru shuwahatuwa wo kijyutu suru tameno anote-shon mojikashuhou no teian (Proposed new annotation and transcription scheme for signed utterances in interaction). *Shuwagaku kenkyuu (Japanese Journal of Sign Language Studies)*, Vol.22, pp.37-63. (written in Japanese) https://doi.org/10.7877/jasl.22.37
Kita, S., van Gijn, I., and van der Hulst, H. (1998). Movement Phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth and M. Fröhlich (Eds.), Gesture and sign language in human-computer interaction, International Gesture Workshop Bielefeld, Germany, September 17-19, 1997, *Proceedings. Lecture Notes in Artificial Intelligence* (Vol. 1317, pp. 23-35). Berlin: Springer Verlag.
Kitzinger, C. (2013). Repair. In J. Sidnell and T. Stivers (Eds.), *The Handbook of Conversation Analysis*, 229–256. NJ: Wiley-Blackwell. https://doi.org/10.1002/9781118325001
Kusters, A. (2015). *Deaf space in Adamorobe: An ethnographic study of a village in Ghana*. Gallaudet University Press.
Levinson, C. S. (2016). Turn-taking in human communication – Origins and implications for language processing, *Trends in Cognitive Sciences*, 20(1), 6-14. https://doi.org/10.1016/j.tics.2015.10.010
Manrique, E. and Enfield, N. J. (2015). Suspending the next turn as a form of repair initiation: Evidence from Argentine Sign Language. *Frontiers in Psychology*, 15 September 2015 | https://doi.org/10.3389/fpsyg.2015.01326
Manrique, E. (2016). Other-initiated repair in Argentine Sign Language, *Open Linguistics* 2(1), https://doi.org/10.1515/opli-2016-0001
McNeill, D. (1996). Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press.
pympi-ling: a Python module for processing ELAN's EAF and Praats TextGrid annotation files. (2013-2021). https://pypi.python.org/pypi/pympi-ling
Schegloff, E. A., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53 (2), 361-382. https://doi.org/10.2307/413107
Schegloff, E. A. (2007). *Sequence Organization in Interaction, A Primer in Conversation Analysis*, 1. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791208
Seuren, L. M., Wherton, J., Greenhalgh, T., and Shaw, S.E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, 172, 63-78. https://doi.org/10.1016/j.pragma.2020.11.005

## 11. Language Resource References

Bono, M., and Adam, R. (2023) Online cross-signing project between the United Kingdom and Japan: First phase of data collection, *Online Proceedings of JSAI-isAI2023*. (published to only conference audience)