Synlogue with Aizuchi-bot

Investigating the Co-Adaptive and Open-Ended Interaction Paradigm

Kazumi Yoshimura
Graduate School of Letters, Arts and Sciences, Waseda University, Japan
y5mrk.2nn@gmail.com

Dominique Chen
Faculty of Letters, Arts and Sciences, Waseda University, Japan
dominique@waseda.jp

Olaf Witkowski
Cross labs, Cross Compass Ltd., Japan
olaf@cross-compass.com

ABSTRACT

In contrast to dialogue, wherein the exchange of completed messages occurs through turn-taking, synlogue is a mode of conversation characterized by co-creative processes, such as mutually complementing incomplete utterances and cooperative overlaps of backchannelings. Such co-creative conversations have the potential to alleviate social divisions in contemporary information environments. This study proposed the design concept of a synlogue based on literature in linguistics and anthropology and explored features that facilitate synlogic interactions in computer-mediated interfaces. Through an experiment, we focused on aizuchi, an important backchanneling element that drives synlogic conversation, and compared the speech and perceptual changes of participants when a bot dynamically uttered aizuchi or otherwise silent in a situation simulating an online video call. Consequently, we discussed the implications for interaction design based on our qualitative and quantitative analysis of the experiment. The synlogic perspective presented in this study is expected to facilitate HCI researchers to achieve more convivial forms of communication.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); HCI theory, concepts and models.

KEYWORDS

Computer-Mediated-Communication, AI-human interaction, aizuchi, synlogue, overlap, co-adaptation

ACM Reference Format:

Kazumi Yoshimura, Dominique Chen, and Olaf Witkowski. 2024. Synlogue with Aizuchi-bot: Investigating the Co-Adaptive and Open-Ended Interaction Paradigm. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642046

1 INTRODUCTION

This study distinguishes between two types of conversational patterns: dialogue, wherein speakers are expected to exchange completed messages while taking turns (turn taking), and synlogue, wherein incomplete utterances are mutually delegated and complemented in an overlapping manner (turn coupling). In recent years, mainstream SNS and communication tools have assumed the former dialogue model of communication, and interfaces and interactions have been designed accordingly. However, we argue that such a communication style that clearly distinguishes individuals clarifies differences in positions; thus, when people in different positions interact with each other, confrontational relationships tend to become apparent and emphasized [54]. We aimed to explore the potential of synlogues as an alternative design space to address the emerging problems of computer-mediated communication [55–57].

To mitigate these social divisions in today's information environment, we examined the concept of a synlogue that focuses on constructing a co-creative, open-ended relationship with others in conversation. A synlogue is a co-constructive and playful conversation pattern that allows two or more agents to participate in continuous, simultaneous, and active conversations with each other. In contrast to dialogue, which is a traditional turn-taking conversation pattern, synlogue does not require a specific order, thereby facilitating a more fluid, open-ended, and flexible flow of conversation. Mizutani who investigated co-constructive conversation in Japanese and introduced the notion of kyōwa (共話, which the current authors translated alphabetically to synlogue) states that "it is immeasurable how much the warm, cozy (...) communication made possible by the kyōwa way of speaking supports the spiritual life of people today [2]." Other scholars linked this character of kyōwa (synlogue) to the ideas of phatic communion [3] and grooming [4], with reference to the characteristics of co-speech in Japanese [5]. Synlogic conversational forms have been noted in both linguistics and anthropology for their effects on generating psychological safety [39] with sympathetic feelings of warmth, encouragement, security, and satisfaction. We hypothesized that if synlogic interaction could be realized through user interfaces, relationships between humans could also be constructed in a convivial and psychologically secure manner.

This study proposes synlogue as a design concept based on studies of co-construction and cooperative overlaps in conversation analysis, anthropology, and philosophy, and reports the results of a conversation experiment to explore the conditions necessary for achieving synlogic interaction in user interfaces. Through an experiment, we focused on aizuchi, which is the characteristic backchanneling element that underpins synlogic conversation. Building upon previous work on computational models of backchannels, we developed a bot that dynamically utters aizuchi in response to a speaker's utterance and simulated a one-on-one video chat with the camera turned off to investigate how the human perception varies with different setups. Based on a quantitative analysis of

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642046

the participants' subjective evaluations and qualitative analysis of their interview responses, we analyzed whether a bot that uttered only aizuchi could reproduce the characteristics of a synlogic conversation and what processes are undergone of reproducing these features. Based on the results of this analysis, we discussed the subjective elements on the user sides that are considered to be important for realizing synlogic interactions in user interfaces, and explored their implications for HCI design.

The contributions of this study are twofold. First, we proposed the design concept of a synlogue for HCI researchers as a co-constructive paradigm that focused on the co-adaptive and open-ended effects of interaction, which are often overlooked in the dominant turn-taking paradigm. Second, we analyzed the results of a minimal experiment to test whether minimal interaction with the voice interface can bring out the synlogic effects and to discuss what subjective factor is important on the human side for synlogic communication.

2 SYNLOGUE

In this section, we present the idea of a synlogue by first explaining its relationship and difference with the traditional view of turn-taking and the overlaps discussed in conversation analysis studies. We then argue that the synlogic perspective departs from turn-taking epistemology by emphasizing the overlooked aspects of concurrent, thereby overlapping utterances to present it as a novel design concept for communication design.

2.1 Co-construction and Cooperative Overlap

Since its inception by Sacks et al., the modern theory of conversation analysis (CA) has been founded on the assumption that turn taking constitutes the structure of conversation [13]. In this dominant discourse, speech conversation is framed as a co-constructive joint action realized by interlocutors seeking to create common ground [44]. However, the turn-taking perspective assumes a "one speaker at a time" rule, wherein speech overlap, understood as a "more than one at a time" situation, is considered as a problem to be repaired or solved. For instance, in his seminal analysis of speech overlap, Schegloff observed how various "overlap-resolution device" emerged in English conversations [47]. Although he argued that in most cases, overlap was something to be managed and resolved, he noted certain special cases where overlap and simultaneous talk were not problematic, such as terminal overlaps, continuers, choral utterances (laughter and greetings), and conditional access to the turn [47, p5].

Related to the exceptional cases pointed out by Schegloff, Lerner worked on choral coproduction, where more than two interlocutors simultaneously uttered the same speech [45]. He analyzed cases, such as searching for words or a shared reminiscence, where speakers anticipated completing an interlocutor's turn. In such cases, Lerner stated that a sense of co-authorship and co-ownership of experience arises, which is appreciated by interlocutors; however, he carefully considered the position of "one-at-a-time speaking."

Tannen has worked on the co-constructive aspect of overlap. She devised the notion of cooperative overlap in Jewish conversational culture [16, 17]. There is a widespread belief that Jewish people tend to interrupt their interlocutors more often than other ethnic groups. Tannen, in opposition to this popular belief, analyzed lived conversation samples and argued that Jewish Americans do not interrupt to dominate others; rather, they demonstrate a higher level of involvement in the conversation. She referred to this manner of speaking as cooperative overlap and distinguished interruptions intended to dominate the floor [18] from the ones used to cooperate with the interlocutor.

Tannen reported that cooperative overlaps occurred more in casual conversation among friends who shared a "high involvement" style [19]. Such a conversationalist is "a listener talking along with a speaker not in order to interrupt but to show enthusiastic listenership and participation" [18, p53]. Thus, in this study, we focused on this collaborative attitude of speech overlap, which entails cultural specificity (as well as controversy). Next, we discuss the co-constructive conversational perspective in Japanese language, kyōwa, which is the origin of our notion of synlogue.

2.2 Synlogue in the Japanese language

We based the conceptual foundation of our idea of synlogue on the linguistic notion of kyōwa, a Japanese term that literally translates to English as "cooperative conversation," and the current authors translated it as synlogue for the scope of this current paper. Kyōwa was introduced and developed by Nobuko Mizutani [1, 2], an educational linguist who observed the process of Japanese language acquisition by international students in Japan and found that students who had acquired more natural Japanese could speak in a cooperative manner. According to Mizutani, kyōwa is a conversational style wherein speakers actively help each other to complete phrases. A simple example of kyōwa in Japanese is as follows:

> A: Kyo-no Tenki-sah (*Today's weather...*)
>
> B: Un, Samui-kedo, kimochi ii yone. (*Yes, it is cold, but it is nice and sunny.*)

In such an unassuming, everyday conversation, one can find the characteristics of kyōwa. First, Speaker A throws out the unfinished phrase that B receives and completes. Here, A's utterance is finished by "sah," a final particle with a slightly pause-like manner, suggesting that A is letting go of completion. A similar way to end a phrase involves the use of the final particle "nandakedo;" however, in any phrase, the intention of not completing the phrase can be conveyed by placing a pause in the middle of a phrase without finishing it. Here, B took over the subject "weather" from A and continued the phrase. In Japanese, it is possible to construct a phrase by omitting the subject; therefore, B's phrase without the subject is natural. This ability to omit the subject and share it with the conversation partner is a grammatical feature of the Japanese language that facilitates such kyōwa.

Another important point in B's utterance is that B first utters "Un," which is a type of backchanneling called aizuchi. Mizutani defines aizuchi as "something that the listener inserts in the middle of a speech to help the speech progress [1]." Like backchanneling in other languages, aizuchi, such as "un" or "hai," encourages the person to whom it is uttered, by conveying the fact that he or she is listening and that the other person should continue the conversation. However, when comparing a Japanese conversation with that of other languages (American English and Mandarin Chinese), the number of aizuchi, nodding, and half-completed sentences has

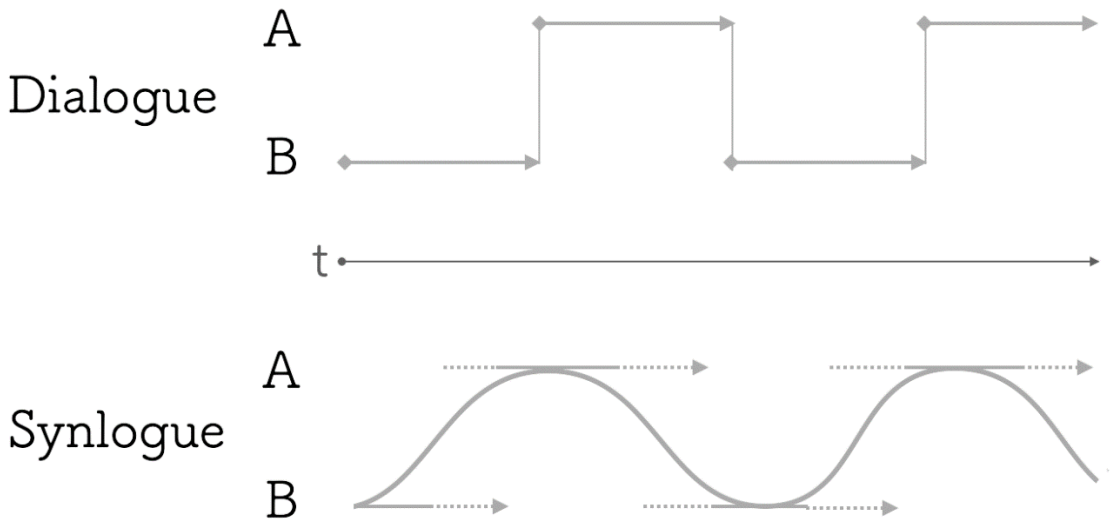

Figure 1: Comparison between dialogue and synlogue. Dotted lines represent overlapping in utterances.Based on Mizutani (1993) [1] and Kawada (2001) [10], extended by Chen (2020) [11].

also been found to be significantly higher [6, 15]. The idea that Japanese conversations contain more evidence of co-construction than other languages has long been debated among researchers of conversation analysis [41, 42].

It is fair to state that the heavy use of aizuchi is natural and even encouraged in Japanese conversations. This differs from the cultural practices, particularly when compared to American English. Mizutani reported that when native Japanese speakers use excessive aizuchis in conversations with Americans, the latter may misunderstand that the Japanese interlocutors agree with them, resulting in miscommunication. Mizutani noted that in American English, the excessive use of aizuchis may offend the other party, who may feel that they are being treated like children [2]. This may also be related to the cultural mindset wherein overlaps and interruptions are avoided in American English, as they are considered to interfere with the speaker's speech.

Although kyōwa has been mostly discussed in Japanese literature as a co-constructive speech style based on active cooperation among participants, aizuchi, a key component of kyōwa has been studied in English literature as well in relation to the broader notion of backchanneling [5, 7–9]. We argue that the collaborative aspect of kyōwa is not exclusive to the Japanese language and culture, by highlighting its similarity with the aforementioned work by Tanen for the cooperative overlap in Jewish American conversation, in addition to other cultures and fields of research that are presented in the following subsections.

Figure 1 is a reconstruction of Mizutani's figure [1] that demonstrates the difference between the flows of a kyōwa (synlogue) and Taiwa (dialogue) in, with added straight lines (showing the turn-takings) and curved lines (merging of turns).

2.3 Synlogue in Cultural Anthoropology

The alphabetical term synlogue used in this paper was first introduced by the cultural anthropologist Junzo Kawada in the context of his research on the Mosi people in Western Africa, independent of the discussion of kyōwa in linguistics. Kawada, who is known for having extensively studied West African cultures, observed an important culture of the Mosi people: a nocturnal meeting of people who enjoyed folktales.

> "It proceeds with the intervention of Aizuchi, the listener, and various words, sometimes corrected by others in the room, and sometimes with the help of others who fill in forgotten or stumbled upon parts of the text. The listener is not a passive recipient at all but at the same time a 'potential speaker,' participating in the 'realization' of the story through his or her voice, and also becoming the next speaker [10]."
>
> (Kawada, 2001, p.110, translated to English by Chen)

Kawada named this form of cooperative speech synlogue, which he compared with the monologic and dialogic types of speech in Western African cultures. Notably, he first conceived the term polylogue to stress the plurality of the speakers; however, upon discussion with his academic peers, he decided to use the prefix "syn" to emphasize the synchronic and co-constructive aspects of the speech. Moreover, this synlogue of the Mosi people can be compared to a shared reminiscence recognition solicit observed in turn-sharing [45]; however, the Mosi synlogue is not only composed of choral turn-sharing but is continuous, where co-speakers mutually succeed in incomplete utterances.

A similar phenomenon was described by cultural anthropologist Daiji Kimura [12]. Kimura studied the people of the Baka Pygmy in southeastern Cameroon and analyzed the characteristics of speech overlap in men's daily meetings. They sometimes spoke simultaneously, all at once, followed by silence, after which their voices

Table 1: Features of synlogue viewed under synlogic and dialogic perspectives, with examples.

	Synlogic perspective	Dialogic persective
Incompleteness	Unfinished, incomplete messages are accepted as communication units. *Examples: Incomplete utterance with Final Particles (Japanese), half-way chat messages.*	Each speaker must complete their sentence. Dependence to the partner is avoided. *Examples: Unfinished e-mails, or articles not accepted.*
Overlap	Utterances can occur simultaneously, allowing for conversational overlaps. *Examples: Aizuchi, cooperative overlap.*	Speakers take turns to speak: any overlap results in an interruption that "steals the turn". *Examples: An academic talk delivered to an online audience via Zoom, waiting on a parent to finish a long text message.*
Multimodality	Simultaneously conveying information via secondary channels is accepted, at the same time as using the primary channel. *Examples: Aizuchi, nodding, text chats in video meetings, body movement can all be used as the main conversation thread.*	Strong focus on primary channel; any digressions are avoided or ignored as noise. *Example: Making a phone to book a table at a restaurant, writing a telegram to a relative.*
Co-adaptation	Agents adapt (change their behavior accordingly) to each other's outputs. *Examples: Cypher (rap), duo-impro (jazz), brainstorming, small talk.*	Antagonistic setting where any influence from the speaker is avoided. *Examples: Competition, debate.*

began to overlap. Kimura described this situation as a polyphony, and highlighted its similarity to the synlogue observed by Kawada.

Kimura also pointed out that traditional turn-taking conversation analysis [13] is not applicable to a conversational floor that is open to everyone. Citing Zimmerman and West [14], Kimura stated that interruptions in Western societies tended to be avoided as accidental or perceived as a sign of male dominance over women, whereas the Baka Pygmy people were more willing to synchronize their voices to build solidarity. Although further linguistic investigation is needed to compare Japanese kyōwa and West African synlogue, this study attempted to correlate the cooperative communication aspects across cultures under the alphabetical term of Synlogue. As a matter of fact, synlogue, which is composed of the prefix syn ("together with") and logue (used to denote discourse of a specified kind) can be considered a direct translation of kyōwa, which is composed of the Chinese characters Kyô ("together with") and Wa ("conversation").

2.4 Features of Synlogue

Although the observations of synlogues in Western African cultures made by Kawada and Kimura, the phenomenon of kyōwa observed by Mizutani in Japanese conversation, and the cooperative overlap analyzed by Tanen in the Jewish American culture differ in both their respective cultural and historical contexts, they share the important perspective that overlapping is a sign of cooperative engagement, rather than a problem to be solved as traditionally defined in CA [13, 47]. The purpose of this current paper is neither to propose a new notion for the field of linguistics nor to refute the many co-constructive characters of the turn-taking paradigm explored in conversation analysis; our aim is to present synlogue as a design concept for HCI, drawing attention to the collaborative values of overlapping (entailing an alternative perspective of turn), for researchers to design sympathetic and convivial communication.

We explored and presented synlogic characteristics applicable to communication beyond verbal interaction to introduce the concept of turn coupling, which underpins the synlogic structure. We summarize these synlogic features with comparative cases from a dialogic perspective in Table 1, which is explained in the following subsection.

2.4.1 Components of Turn-coupling. The fundamental difference between dialogues and synlogues is whether turns are taken in sequence or coupled synchronously. Turn coupling occurs when turns are not taken in order by interlocutors but rather conjointly in a synchronous manner. We note that turn-coupling is distinguished from the notion of turn-sharing, which is defined as "saying the same thing at the same time" [20, 45]. We visualized the flow of turn coupling in synlogic communication in Figure 2 by employing the key features of the synlogue presented in Table 1.

Incompleteness An important factor that enables turn coupling is the incompleteness of each information utterance. If the information block is too complete, there is no space (or affordance) for real-time co-construction. Co-constructive relaying can occur only if previous information is instantly perceived as incomplete, which invites the interlocutor to continue. In Japanese conversation, certain final particles are used to show explicitly that a phrase is incomplete, thus inviting the partner to join the turn and continue [5]. This property of incompleteness yields other synlogic features, such as overlap, multimodality, and co-adaptation.

Overlap As demonstrated by cross-cultural studies of aizuchi and other backchanneling behaviors [6, 15] and their characteristics in Japanese conversations [5], synlogic interactions often incite overlap among multiple interlocutors. When overlaps occur, turns are perceived by agents as if they are coupled, thus generating a

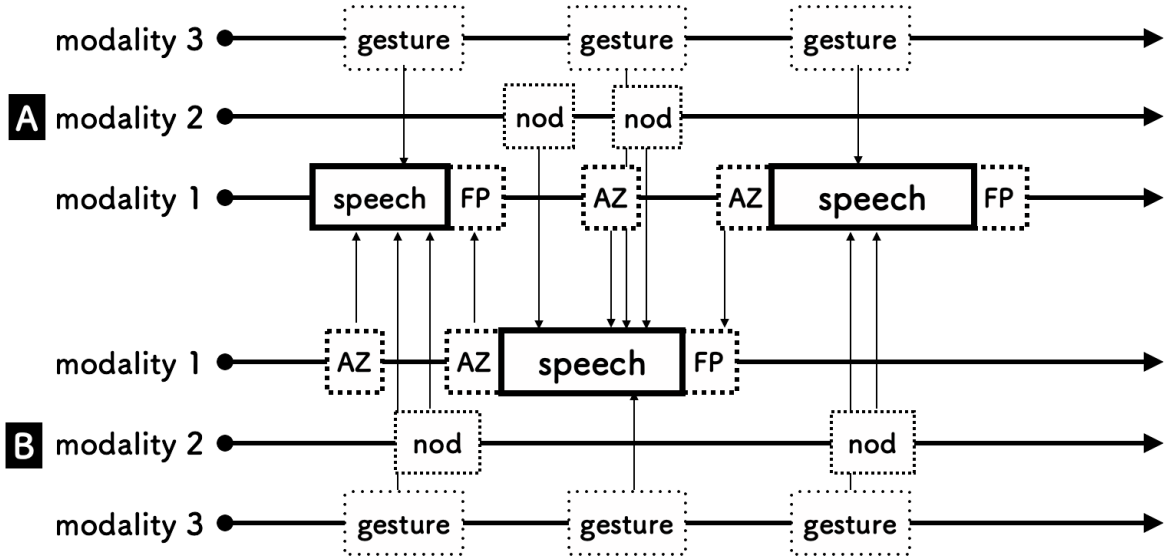

Figure 2: Elements of turn-coupling in synlogic conversations. Each agent utters incomplete information (speech ending with final particles [FP]) using multiple modalities, which overlap with each other. Aizuchi [AZ] and other backchannelings (nod, gesture) are emitted as probes for co-adaptation (shown with arrows).

sense of synchronous co-construction. In addition to oral communication, that is, face-to-face conversation, the activating effect of information overlap has been observed in real-time text chat situations [58]. The study, which quantitatively compared transfer entropy with other subjective qualities such as social presence, indicated that the higher the concurrency of information, the more constructive the communication.

Multimodality Under a synlogic framework, communication is not limited to a primary communication channel. This is in contrast to the dialogic perspective, which tends to focus solely on the primary channel content, that is, speech utterance, as the conveyer of information, discarding any extraneous signal as noise. For synlogues, subchannels (e.g., aizuchi, nodding, gestures, emotion, or other examples of backchanneling) are considered crucial in establishing synchronicity among participants and are viewed as the primary content contributing to the overall effectiveness of the interaction.

Co-Adaptation We argue that, through micro-interactions based on incomplete, multimodal, and overlapping information, synlogic partners co-adapt to each other in turn-coupling. On the conceptual level, the synlogic co-adaptation is comparable to dialogic co-construction in such instances wherein co-ownership of the floor occurs, as in turn-sharing analyzed by Lerner in [45, p16]. The synlogic perspective further focuses on the phatic aspect of the communication [3], wherein mutual influences guide the process of co-adaption among participants.

To utter unfinished sentences is an act that demands that one accept their own vulnerability, and accepting the other's overlap with one's own voice calls for a certain frame of mind to welcome the alien components into one's subjectivity. Reminiscent of Haraway's notion of becoming-with others, or sympoiesis [23], the open-ended attitude of synlogues escapes from the rational and dialectic teleology of an adversarial dialogue. The playful attitude can then be interpreted as a will to embrace unexpected consequences (such as enjoying going off track in small talk or free brainstorming) and a more social purpose, for example, to make kinship.

To understand the unique quality of synlogue, especially in comparison with dialogic perspective, its co-adaptive consequence needs to be explored further. How the open-ended attitude of a synlogue is achieved? Is this something one willingly aims to achieve? Alternatively, is this feature acquired without consideration? To investigate these questions, we turn next to relevant literature in philosophies of reflection, language and joint actions, in addition to computational conversational analysis.

2.4.2 Co-Adaptation and Open-Endedness. First, we consider how the features of synlogue contribute to the generation of co-adaptive and open-ended relationships. The process of synlogic interaction evokes Schön's analysis of reflection-on-action and reflection-in-action [21]. Through this concept, dialogue can be understood as a series of reflection-on-actions applied to each exchanged information block. In synlogic communications, reflection-in-action appears to occur within the interactions of the participants: participants in a synlogue treat each other's microexpressions (incomplete, overlapping utterance, and backchannelings) as co-owned resources for co-construction. In Schön's example of an architect who draws a sketch, each brush stroke of the pencil incites feedback to which her consciousness adapts on a microscale of time. We note that Schön's discussion of reflection-in-action targeted the creative process of individuals, but here we apply the concept of a collective relationship: in turn-couplings, synlogic partners serve each other's fragments of expressions to reflect upon within a few hundred μs. This idea also resonates with studies on the effect of aizuchi as a tool for internal information processing [22], wherein,

one might emit an aizuchi or other form of backchanneling, not simply to show support to the partner, but also to disclose internal cognitive processing both to the partner and the emitter.

In a prosodic analysis of backchannel responses in Japanese and English conversations, Ward and Tsukahara, with regard to kyōwa, noted that "conversations seem to involve something special, something present above and beyond the goals and actions of the two participants" and speculate that "the 'reflex' aspects of some conversation behaviors (i.e. backchanneling), that is, the way they involve direct links between stimulus and response that apply automatically and unconsciously [24]." Synlogues might then be a communicational mode wherein participants interact with a lesser sense of autonomous will and responsibility, the dominant components of the notion of individuality in Western modernity. In other words, particularly when compared to the dialogic presumption of individuality, synlogue is a communication style wherein the relations of the interlocutors are mediated through their structural characteristics by design [59].

Recent discussions on the philosophy of language have critically examined the problem of individuality. Kokubun conducted relevant research that questioned the idea of will and responsibility through an archaeology of the middle voice (diathesis) that existed in ancient Greek and Sanskrit before the dichotomy of the active and passive voices that are familiar to us today [25]. Statements articulated in the middle voice are not active ('I broke the window') or passive ('The windows were broken by someone'), but only describe a change in the subject's status ('The window broke'). From this comparison, we realize that the active–passive dichotomy invokes a world controlled by instruction and representation, whereas the middle voice exhibits greater affinity to a world where meanings emerge within relations [26].

In a discussion on joint action, Miki proposed the notion of concessive joint action, arguing that the classical idea that joint action requires a shared goal is excessively strong [48]. She presented examples of partners who conceded to each other and ended up achieving a goal other than the one initially set; the identity of the cooperation changed throughout its deployment. Another study on freely improvised joint action (FIJA) [49], symbolically emphasized through the case of musical improvisation, depicted the autotelic nature of such communication that "target the continuation of the present state or activity." Synlogues involving an a priori collaborative attitude also share the open-ended and self-referential character of concessive joint action and FIJA. These concepts of joint action help to understand the co-adaptive aspects of synlogic interactions.

We argue that synlogic situations wherein subjectivities become entangled afford the unexpected emergence of meanings in a playful and open-ended manner.

3 RELATED WORKS

In the fields of HCI and Social Robotics, studies have been conducted to achieve synlogic conversation in telecommunication and interaction with robots by inserting spoken verbal aids and nodding movements into conversations. Yano and Ito (1996) attempted to compare the differences in the emergence of synlogic aspects in audiovisual and audio-only communication settings and found that more natural synlogues were generated in the former. Okato et al. examined the prosodic cues of aizuchi used in natural speech to create a mechanical speech response system that produced natural aizuchi [9]. Mori focused on the speech waveform of natural aizuchi and showed that randomly generated aizuchi provided a natural impression to humans interacting with a machine system [8]. Ward and Tsukahara proposed a prosodic method to predict backchanneling timing based on low-pitch detection [24, 43]. Morency et al. achieved improved accuracy of backchanneling predictions by employing a hidden Markov model [40]. However, as a recent review of turn-taking conversational systems demonstrates, very few studies have been conducted on cooperative overlap in computational research [53].

In designing social robots that interact with humans, Nakamichi et al. and Iwabuchi et al. designed aizuchi behaviors to be performed simultaneously with nodding motions [28, 29]. Arimoto et al. reported that the "nodding" behavior of a bystander robot improved its social presence in telecommunications [30]. Park et al. proposed a storytelling robot signaling attention and listening using nodding as a backchannel [31], and Murray et al. devised a social robot using a data-driven model to create a natural experience close to human-human conversations [32]. Lala et al. designed a humanoid device capable of attentive listening with backchanneling and response generation from a turn-taking perspective [50]. However, these studies focus on designing natural human-computer interaction, and thus do not target the co-constructive synlogic aspects of overlaps and backchanneling as the current paper does.

Other studies relevant to our discussion of synlogue aimed to explore the active roles played by the listeners of the conversation. Malisz et al.'s study on the active listening corpus [51] and Healey et al.'s analysis of collaborative bodily gestures [52] foreground the active role of the listeners and their co-constructive contribution to the conversation generation, even under a dialogic perspective with 'one-speaker-at-a-time' imperative. This work informs the synlogic characteristics of multimodal information overlap in speech conversation.

Besides speech interaction, Kojima et al.'s research in the realm of computer-mediated communication (CMC) encompasses two significant studies: the perceptual crossing experiment [61] and text chat [58]. These studies highlight that simultaneous input from multiple parties increases the transfer entropy of communication, implying that co-adaptation takes place through a micro-timescale information overlap. Furthermore, the investigation by Reddy et al. on the topic of unsupervised human-machine co-adaptation [60] provides a pertinent example of synlogic ad hoc coordination between a human and an algorithm, where prior knowledge or intention is not shared between the two parties.

Compared to these related studies, the current study makes a different contribution in that it focused on spoken aizuchi and attempted to analyze aizuchi dynamically uttered by the system from a synlogic perspective.

4 SYNLOGUE EXPERIMENT WITH AIZUCHI-BOT

In synlogic conversation, backchanneling using multiple modalities, such as aizuchi, nodding, gestures, and facial expressions, plays an important role, similar in importance to or greater than

the semantic content of speech. In particular, in Japanese-spoken conversation, aizuchi, which is uttered more frequently than in English or Chinese, plays an important role in realizing a co-creative conversation process. In this experiment, we focused on aizuchi, which is an important and minimal unit of speech for realizing synlogues, and developed a bot that dynamically responded to the speaker's utterances with aizuchi only. By combining the results of these quantitative and qualitative analyses, the purpose of this experiment is to clarify how the insertion of minimal utterance, such as aizuchi, can generate synlogic conversations, and what the subjective factors cause such conversations in the interactions between humans and user interfaces. Through the analysis and considerations, we hope to discuss what are the important factors for incorporating synlogic communication into any communicative user interface. Because this experiment was limited to voice, the discussion was limited to voice.

4.1 Design of Aizuchi-bot

To conduct this conversation experiment, we developed an aizuchi-bot that dynamically utters aizuchi in response to speaker utterances.

4.1.1 Implemention of When to Utter Aizuchis. In this experiment, we employed the approach proposed by Okato et al. [33] because it is relatively easy to implement aizuchi. Aizuchi tends to be uttered at the end of a prosodic phrase, such as a pause in the speaker's speech. Okato et al. reported that 93% of aizuchis were uttered within 0.1 to 0.3 s after the end of phrases, and aizuchis that are delayed longer than 0.3 s were perceived as slow [33]. In addition, speech tended to be terminated after an average of 0.25 s, when the fundamental frequency (F0) dropped by 20% from its average value [33]. Based on these results, Okato et al. proposed a method to predict the end of the next prosodic phrase from the prosodic information of the speaker's utterance, and to predict the appropriate timing when aizuchis should be uttered.

In the current version of the aizuchi-bot, the prediction of prosodic phrase endings was used to determine the timing of the aizuchi-bot. However, Okato et al. reported that the prediction of aizuchi using this method had an accuracy of 52%. In the actual implementation, we found that in many cases aizuchi was inserted in the middle of the utterance. Therefore, in addition to Okato et al.'s method, we added a logic to the aizuchi-bot that assumed that speech was still in progress and canceled aizuchi if a speaker's utterance was detected during the time between the prediction of the end of a prosodic phrase and the aizuchi's strike.

The implementation logic of the aizuchi-bot is summarized as follows.

- The aizuchi-bot always acquired the fundamental frequency (F0) of the participants' voices picked up by a microphone and stored them in an array.
- Once the F0 value was obtained, the average value of the F0 array at that time was calculated.
- Aizuchi was uttered 0.5 s (0.25 s + 0.25 s, these are shown below) after the F0 value of below 80% of the average value was detected.
  - If the acquired F0 was less than 80% of the calculated average values, the end of the phrase of the participants' speech was predicted to occur 0.25 s later.
  - As aizuchi must be uttered within 0.3 s of the end of a prosodic phrase, the pause length was set as 0.25 s constantly.
- If the microphone detected the speaker's utterance during the 0.5 s between the prediction of the end of the prosodic phrase and the time when the aizuchi was uttered, it was assumed that the speaker's utterance was still ongoing and the aizuchi was canceled.
- The F0 array is initialized when aizuchi is uttered.
- To prevent frequent aizuchi utterances, the microphone was not used for 2 s after the utterance of aizuchi.

4.1.2 Varieties of Aizuchis. We prepared aizuchi using two patterns: a synthetic voice and a human voice. For the synthetic voice, we used "Kyoko," a Japanese synthetic voice available on Mac OS, and for the human voice, we used the voice of the first author, which was recorded in advance. Both the synthetic and human voices are female voices with a similar voice pitch. Ogawa and Saito found that in voice communication situations where visual information was not available as a cue, a variety of spoken aizuchi types were rated significantly higher than a single type in terms of familiarity, activity, pleasant impression of the conversation, conversation maintenance skills, conversation satisfaction, and evaluation of conversational behavior toward the listener [34]. Therefore, we designed the aizuchi-bot to randomly select multiple types of aizuchi to strike rather than a single aizuchi style. However, certain types of aizuchi typed in daily life are context-independent, such as "un" and "hai," while others are context-dependent, such as "hee (oh)" and "sounanda (I see)." For such aizuchi to be uttered, the content of the conversation partner's speech must be understood, and the appropriate aizuchi must be selected according to the context. However, if aizuchi is to be uttered after contextual understanding in real-time, processing will take time. Aizuchi must be uttered within 0.3 s from the end of the prosodic phrase of the utterance; thus, the aizuchi may be delayed from the appropriate timing if contextual understanding is included in the processing. Therefore, only context-independent aizuchi were used in this study. In addition, we prepared a single inflection pattern for the synthetic voice and multiple types of inflection patterns for the human voice. The list of aizuchi and inflection patterns adopted in the aizuchi-bot is presented in Table 2.

4.2 Testing Conversations with Aizuchi-bot

To develop a design for the experiments, we first conducted a preliminary conversation test using the aizuchi-bot. In the experiment, we initially wanted to investigate whether the mere intervention of a bot that only uttered aizuchi in a conversation between humans produced synlogic effects. Therefore, we conducted a conversation test assuming a situation wherein the aizuchi-bot intervened in a conversation between two humans. In this conversation test, aizuchi-bot dynamically inserted aizuchi voices such as "un" or "hai" during a conversation between two people on a free topic.

Table 2: Varieties of aizuchis of the aizuchi-bot.

	Patterns of aizuchi	Patterns of inflection
For synthetic voice	"hai"	constant
	"un"	constant
	"hmm"	constant
For human voice	"un"	naturally, cheerfully, quietly, or joyfully
	"hai"	naturally, cheerfully, or seriously
	"unun"	naturally
	"haihai"	naturally

However, the conversations did not progress well on this test. The two speakers reported that they felt that the aizuchi-bot interrupted their conversations. When we listened to the audio recording of the conversation to investigate why this was happening, we found that, during the conversation, the human listener, as well as the aizuchi-bot, unconsciously and constantly uttered aizuchi. The speaker timed the conversation to match the aizuchi uttered by the listener rather than the aizuchi uttered by the aizuchi-bot. As a result, the timing of aizuchi uttered by the aizuchi-bot did not match the speaker's speech, and it seems that the speaker felt that the aizuchi-bot was an obstacle to the conversation.

The reason that the human listener prioritized the human response was believed to be that the human who responded appropriately was more involved in the conversation than the bot that only uttered aizuchi. Moreover, as the human conversation was conducted face-to-face, there was visual feedback. The authors hypothesized that the aizuchi of the human participants may have been prioritized because of these reasons. To avoid this situation in the experiment, the bot should be involved in the conversation to the same extent as the human, not only with aizuchi, and provide visual feedback to the same extent or, conversely, eliminate feedback from the human listener.

Because the primary purpose of this experiment was to investigate whether small interventions by voice feedback such as aizuchi from a bot could produce a synlogic effect, we decided not to increase the degree of its involvement in the conversation, but to eliminate feedback from the human listener. We would like to conduct future experiments with an aizuchi-bot in multi-person conversations after modifying it to enable responses other than aizuchis.

4.3 Design of Experiments

The experiment was conducted with 14 undergraduate and graduate students (seven males and seven females) aged 18–23 whose first language was Japanese. The participants were recruited by posting a notice on an online university system. The experiments were conducted face-to-face, and each participant was given a reward of 2,000 yen per hour.

In this experiment, participants were asked to speak freely about a particular topic to an online partner who had turned off the camera based on the insights gained from the conversation test described above. To reproduce "a situation wherein participants talked with a person who turned off the camera in an online call," a PC running the aizuchi-bot was placed in front of the participants with the image shown in Figure 3 (left) displayed in full screen, and the participants were asked to talk to the PC (Figure 3, right). During the experiment, the researcher moved out of sight of the participants to avoid affecting the task. The topic was presented in the form of a card, and the participants were provided 1 min to think about what they were going to say, after which they were asked to speak freely with no time limit on the conversation.

The conversation task was conducted in three different patterns: (1) a case wherein the online conversation partner was muted and completely silent (a pattern of silence), (2) a case wherein the online conversation partner only uttered aizuchis with a synthetic voice for responses (a pattern with synthetic voice aizuchis), and (3) a case wherein the online conversation partner only uttered aizuchis of a human voice for responses (a pattern with human voice aizuchis). After the conversation in each pattern was completed, the participants were asked to respond to a questionnaire regarding their subjective evaluation of their impressions of the conversation in each pattern. Because the experiment was conducted using a within-subjects design, the order in which each pattern was performed was switched for each subject. This ensured that familiarity with the task did not affect the results of the experiment.

For subjective evaluation, a questionnaire was created with a total of 23 items: 15 of the 16 evaluation items used by Tsuzuki and Kimura in their analysis which clarified the factors that cause differences in psychological characteristics toward conversation partners in different modes of media communication [35] (only the item "effective in gathering information" was excluded because it could not be used under the current experimental conditions), 3 items from the "the conversation satisfaction" index [36, 68], and 5 items added by the current authors. All questionnaire items are presented in the appendix (A.1). For each item, data were collected using the semantic differential scale method with a 5-point scale from "not at all applicable" to "very applicable" or from "I agree very much" to "I disagree very much."

The items we added were:

1. The content of the conversation has expanded off-topic.
2. Feeling more mentally relaxed.
3. Feeling anxious.
4. Easier to come up with things to talk about while talking.
5. Can speak smoothly.

After all the conversation patterns and answers to the questionnaire were completed, semi-structured interviews were conducted. The interview questions were as follows.

For each pattern we considered the following:

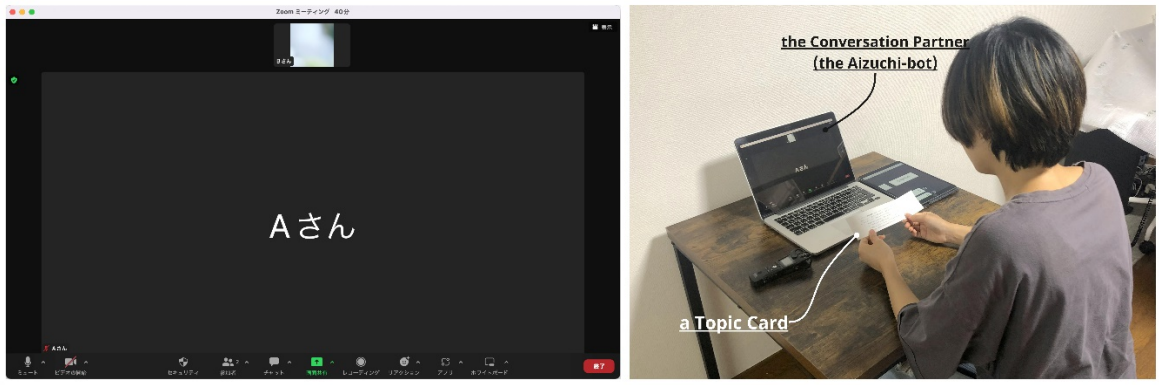

Figure 3: Screens shown on a PC during the experiment (left) and the experimental scene (right).

1. What made it easier for you to speak without any voice, with synthesized voice aizuchis, and with human voice aizuchis?
2. What did you find difficult to talk about in the following situations: without any voice, with synthesized voice aizuchis, and with human voice aizuchis?
3. In each pattern, what kind of presence did you feel the conversation partner was?
4. When did the conversation digress or ideas expand?
5. Types and sounds of aizuchi: Depending on the type of aizuchi ("un" or "hai"), the inflection and tone of the aizuchi, and the tone of the words, what did you notice?

5 RESULTS

The data obtained in this experiment were subjected to statistical analysis for subjective evaluation by our questionnaire and qualitative analysis of the interview data using the modified grounded theory approach (M-GTA) proposed by Kinoshita [37]. The M-GTA is a qualitative research method suitable for analyzing interview data, applying the theoretical and content characteristics of the original version of the Grounded Theory Approach (GTA) [38] to generate explanatory models of human behavior, and critically continuing and reorganizing its tradition. The original version of the GTA was proposed as an objectivistic and inductive research method in qualitative research to analyze data and generate theories as rigorously as quantitative research. In the GTA, however, the meaning of data in grounded-on-data analysis needs to be considered, which is then interpreted and selectively judged by the researcher. Therefore, it has been criticized that the subjective viewpoint and existence of the researcher should not be abstracted from the standpoint of objectivism. In contrast, the M-GTA was devised following the grounded-on-data principle of the GTA, in which analysis is based on correspondence with data, although it was taking the position of qualitative research that the complexities of humans in society are understood through interpretations and the contexts are important [37]. In the M-GTA, it is named the "modified version" in that it does not code the meaning of the data by dividing it into small pieces from the beginning, but allows interpreting the data by going back and forth between the data and the concepts to understand of their contexts, and that the clarification of the analysis method and process, which have been considered issues in the GTA, are presented in detail in its methodology. The M-GTA is used as an analytical method for generating substantive grounded theories and identifying and predicting social interactions, particularly in the context of various human services, such as public administration [66], nursing and healthcare [67] in Japan.

In this experiment, M-GTA was used for qualitative analysis because we aimed to analyze the interviews to determine which processes did (or did not) generate synlogic conversations in each condition. For the interview data, only the responses to the conversation with the conversation partner with the aizuchi-bot (synthetic voice bot and human voice bot) as aids were employed for analysis, with the aim of clarifying the process of generating a synlogic conversation using the aizuchi-bot. The results of each analysis are presented below.

5.1 Statistical Analysis of Subjective Evaluation Data

For the subjective evaluation data of the questionnaire, we conducted a paired t-test using the silence pattern as the baseline and comparing it with the pattern with the synthetic voice aizuchi and the pattern with the human voice aizuchi, respectively. Table 3 presents a selection of items for which significant differences were obtained in the comparison results between the silent pattern and synthetic voice bot, and between the silent pattern and human voice bot. A paired t-test comparing synthetic voice aizuchi with human voice aizuchi was also conducted; however, because no significant differences were obtained for any of the items, they were omitted from the table.

The 5 items for which significant differences were obtained for both the human voice and the synthesized voice were: "relieve loneliness," "feel close to the partner," "compassion can be realized," "be easy-going," and "you can be interested in conversation." In other words, these items were improved by the use of aizuchi compared with when the conversation partner was completely silent. Regardless of whether the voice of the aizuchi was mechanical or human, it was found that the mere insertion of a short utterance such as "un" or "hai" into the conversation greatly relieved the sense of loneliness (SV: $t(13) = 3.17$, $p < .01$, Cohen's $d = 0.881$, HV: $t(13) = 4.16$, $p < .01$, Cohen's $d = 1.15$), made the participants feel

Table 3: Results of analysis of subjective evaluation.

Items		Average (silent pattern)	Average (with the synthetic or human voice aizuchi-bot)	t value	p value	Effects size (Cohen's d)
(1) Relieve loneliness	Synthetic voice	1.57	2.93	3.17	p < .01	0.881
	Human voice	1.57	3.29	4.16	p < .01	1.15
(3) Be nervous	Synthetic voice	3.14	2.14	-2.75	p < .05	0.764
	Human voice	3.14	2.71	-1.00	n.s.	0.278
(5) Feel close to the partner	Synthetic voice	1.29	2.43	3.31	p < .01	0.918
	Human voice	1.29	3.14	5.38	p < .01,	1.49
(8) Compassion can be realized	Synthetic voice	1.71	2.64	2.41	p < .05	0.670
	Human voice	1.71	2.64	2.51	p < .05	0.700
(12) Be easy-going	Synthetic voice	2.21	3.57	3.80	p < .01	1.05
	Human voice	2.21	3.21	2.65	p < .05	0.734
(16) Talks are expanded beyond the topic	Synthetic voice	2.57	3.00	2.48	p < .05	0.688
	Human voice	2.57	3.29	1.93	n.s.	0.536
(21) The conversation progressed cooperatively	Synthetic voice	1.86	2.64	1.92	n.s.	0.534
	Human voice	1.86	2.86	2.55	p < .05	0.707
(23) You can be interested in conversation	Synthetic voice	2.21	2.93	3.24	p < .01	0.898
	Human voice	2.21	3.00	2.24	p < .05	0.622

more familiar (SV: t(13) = 3.31, p < .01, Cohen's d = 0.918, HV: t(13) = 5.38, p < .01, Cohen's d = 1.49) and more casual (SV: t(13) = 3.80, p < .01, Cohen's d = 1.05, HV: t(13) = 2.65, p < .05, Cohen's d = 0.734) with the other party. Further, it made them more considerate to the partner (SV: t(13) = 2.41, p < .05, Cohen's d = 0.670, HV: t(13) = 2.51, p < .05, Cohen's d = 0.700) and more interested in the conversation (SV: t(13) = 3.24, p < .01, Cohen's d = 0.898, HV: t(13) = 2.24, p < .05, Cohen's d = 0.622) than the case where the conversation partner was completely silent.

Meanwhile, the two items for which significant differences were obtained only for the synthetic voice aizuchi-bot were "be nervous" and "talks are expanded beyond the topic." The results showed that the aizuchi-bot partner with the synthetic voice was less nervous than that with the silent pattern (t(13) = -2.75, p < .05, Cohen's d = 0.764), and that the content of speech was more expansive (t(13) = 2.48, p < .05, Cohen 's d = 0.688). The only item that was significantly different for the human-voiced bot was "the conversation progressed cooperatively," with the human-voiced aizuchi-bot partner progressing more cooperatively than with the silent pattern (t(13) = 2.55, p < .05, Cohen's d = 0.707). From the above, it can be said that the use of a synthesized voice had a significantly positive effect on relieving nervousness and the spread of topics, while the use of a human voice had a significantly positive effect on the perception of cooperativeness to the partner. The considerations of these results are discussed in section 5.3 with the results of the qualitative analysis in section 5.2.

5.2 Qualitative Analysis of Data from Interview

The data obtained from the semi-structured interviews (the total time: approximately 280 min, the length of transcript text: 119152 characters) conducted in the experiment were analyzed using M-GTA in terms of "the process of generating synlogic conversation by the aizuchi-bot." Since the interviews were conducted in Japanese and the first author is also a native Japanese speaker, the analysis was conducted in Japanese. Through this analysis, 30 concepts were generated, of which 8 categories were identified (Table 4). The resulting diagram is shown in Figure 4. The detailed descriptions of each category and concept are provided in the appendix (A.2). In the following, the categories are indicated in [ ] and concepts in " ".

5.2.1 Storyline: The process of generating synlogic conversation by the aizuchi-bot. The story line is expressed as follows.

When participants began to speak and heard a conversation partner's voice, in many cases, their [perceptions of the conversation partner] changed depending on whether the voice was synthetic or human. In the case of a synthetic voice, participants often "perceived the conversation partner as a machine," and if it is a human voice, they "perceived the humanness in the conversation partner." When they perceived that they were talking with a machine, they felt that the voice was automatically and mechanically uttered, even if they were responded via aizuchi, which increased their "sense of absence of the conversation partner" and made them feel that "the partner was someone who did not listen." Thus, they were in the position of the speaker, and they felt "fixed in the role of speaker." Consequently, they felt a sense of loneliness and thought that if they were talking with "the partner who did not listen", they did not have to talk further. Then, they "lost motivation to continue the conversation." This can be considered a situation of [failure to establish cooperative relationships] with the aizuchi bot. However, when the "sense of absence of the conversation partner" increased, there were cases where participants could talk freely without being aware of the partner, and may enjoy conversing by feeling "the ease of a machine partner." This situation can be considered as the [realization of a casual conversation].

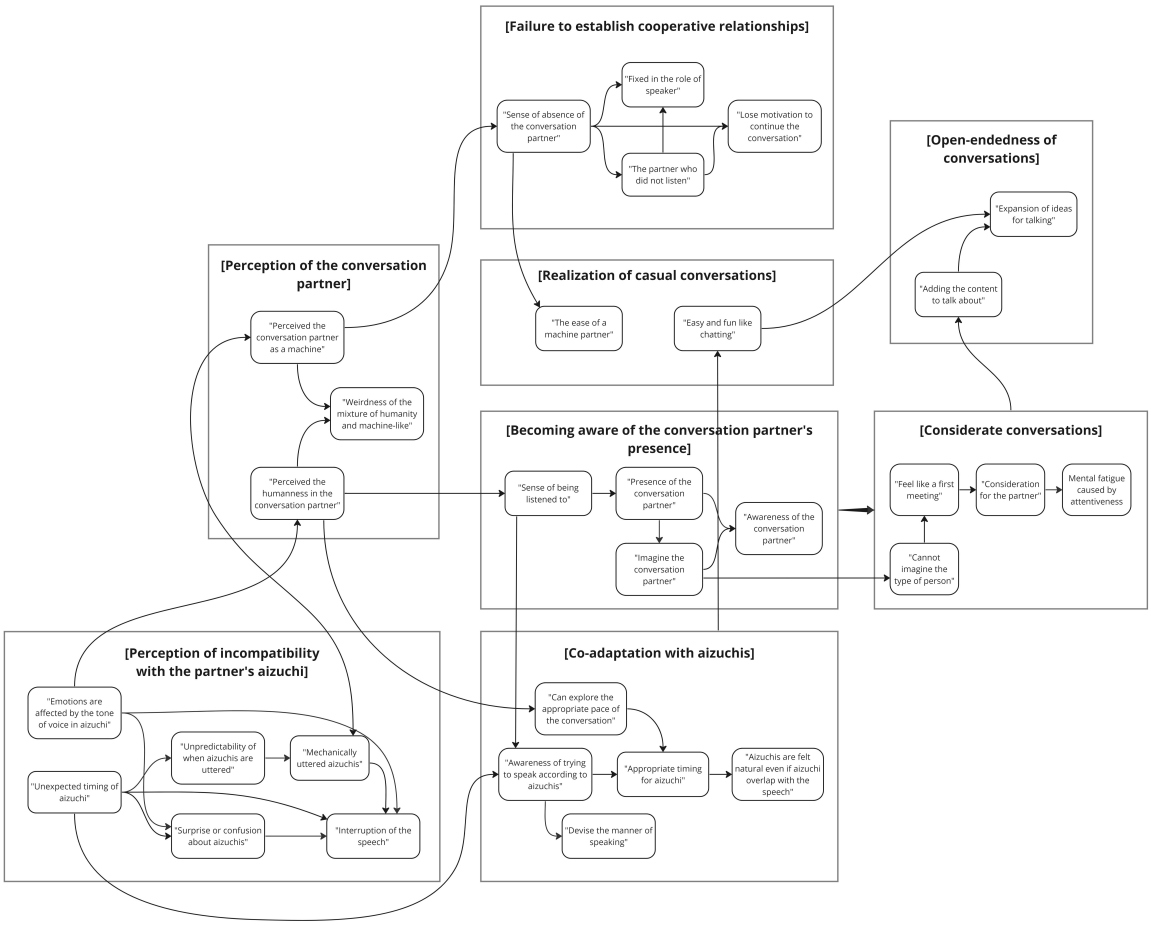

Figure 4: The process of generating synlogic conversation by the Aizuchi-bot.

However, if they "perceived the humanness in the conversation partner," then their "sense of being listened to" grew because of the response of aizuchi. When they felt that they were being listened to, the "presence of the conversation partner" increased, and they began to "imagine the conversation partner" and their "awareness of the conversation partner" increased in the conversation. In the process of [becoming aware of the conversation partner's presence], they imagined the partner, but because the camera was off and they have never seen the partner, they could not have a concrete image of him/her. Therefore, they "couldn't imagine the type of person" and "felt like a first meeting." Consequently, they became concerned about whether they made the partner uncomfortable with their comments, or they felt a desire to entertain the partner. This is how "consideration for the partner" occurs. In this way [considerate conversation] was achieved, but "mental fatigue caused by attentiveness" may be felt. As a result of the realization of [considerate conversations], participants attempted to change the content of the conversation as they spoke, or "adding the content to talk about" in response to reactions, which resulted in "broadening of ideas." In addition, participants attempted to change and "add the content to talk about" while speaking out of consideration for the partner, which resulted in "expansion of ideas for talking." Consequently, [open-endedness of conversations] were realized.

We observed this process when a conversation proceeded smoothly. In reality, however, cases wherein aizuchi was inserted at "unexpected timing," or "emotions were affected by the tone of voice in aizuchi" could occur because the tone of aizuchi was not appropriate to the content of the conversation. In such cases, speech interruption may occur, resulting in errors such as an inability to speak smoothly (disfluency) or silence. When this happens several times, participants may feel the "unpredictability of when aizuchis are uttered" and think that "aizuchis are uttered mechanically." In this case, they [perceived incompatibility with their partner's aizuchi] and gave up attempting to adapt to it. However, if they "perceived the humanness in the conversation partner" and had the "sense of being listened to," they can develop "awareness of trying to speak according to aizuchis" even when such errors occur. In these cases, participants may "devise a manner of speaking", such as changing the speed of their speech or making some pauses in the conversation according to the aizuchis. As a result, the "awareness of attempting to speak according to aizuchis" can result in feeling "appropriate timing for aizuchi," or to a state wherein the

aizuchis were felt as natural even when they overlapped with the speech. In this process, it can be said that the humans attempted to "co-adaptation to aizuchis." Moreover, to the extent that their adaptation behavior also affected the timing of the computed aizuchi utterances, the interaction can be considered a co-adaptation process.

When co-adaptation with the aizuchi-bot occurs, the conversation becomes "easy and fun like chatting," and owing to new topics, they also feel "expansion of ideas for talking." This can result in [realization of casual conversation] and [open-ended conversations].

As described above, "co-adaptation" and "open-endedness," which are characteristics of synlogic conversation, occurred in the process of conversation with aizuchi-bot and we were able to clarify the process by which these characteristics were generated.

5.3 Consideration

Based on the results of the analysis in Section 5.2, it is evident that there were differences in the interactions that occurred, depending on whether the user perceived the aizuchi-bot as human-like. In most cases, the voices used for aizuchi felt machine-like when they were synthetic and human-like when they were human. This allowed us to consider the items in the results of the analysis in Section 5.1, where certain differences in the results between the synthetic and human voices were observed.

In the results of the analysis in 5.1, the only items that exhibited significant differences only with the synthesized voice were "be nervous" and "talks are expanded beyond the topic." In other words, when a synthetic voice was used for aizuchis, the participants felt less nervous and could broaden their conversation. This is consistent with the results of the analysis in Section 5.2, which states, "when the presence of the conversation partner is not strong, participants can talk casually without worrying about the conversation partner, which reduces nervousness, and as a result, the topics of conversation become broader." Considering this characteristic of synthetic voice in terms of synlogic conversation, it can be said that it affected the open-ended nature of the utterance. The main reason for this result for the synthetic voice can be attributed to the "ease of dealing with a machine rather than a human being. On the other hand, the only item that showed a significant difference for the human voice was "the conversation progressed cooperatively." The analysis results in Section 5.2 showed that even if the timing of the partner's response was felt to be inappropriate, the speakers grew awareness of attempting to speak according to the timing of the partner's response in the case that they felt humanness to the partner. Consequently, they could co-adapt to their responses, and the speakers felt that the aizuchis uttered by their conversation partners was natural. This may explain why the participants felt that their partner with the human voice was cooperative, as shown in the results of the analysis in Section 5.1. In fact, when the speaker perceived that a cooperative relationship was established with the partner, he or she felt aizuchis were uttered at the appropriate time, even if aizuchis sometimes overlapped with the speaker's speech. In this case, the overlaps were perceived as natural and the characteristics of synlogic conversations may be realized. Moreover, when co-adaptation with aizuchi was established, open-ended conversation also occurred. Therefore, from the synlogue perspective, it can be seen that co-adaptation and open-endedness are occurred in the conversation if the conversation partner is felt human-like, even if it is the bot which only utters aizuchi.

6 DISCUSSION

Throughout this paper, we have described the design concept of synlogue and its characteristic themes, and presented a minimal experiment to explore the qualities of synlogic communication. Based on our experiment, we found that even if participants felt uncomfortable with the timing of aizuchi, if they could sense the humanness of the conversation partner, they may adapt to the conversation with the aizuchi-bot and perceive it as a natural aizuchi. In this process, co-adaptation and open-endedness that are the features of synlogue are caused. In other words, it is conceivable that the subjective perception of the humanness of a conversation partner may be an important factor to bring out the effect of the synlogic conversation.

6.1 Humanness

As shown above, the experimental results indicate that the subjective perception of the humanness of the conversation partner (in this case, the aizuchi-bot) is an important factor influencing co-adaptation. However, this perception of humanness also heightened the awareness of the conversation partner's presence, inadvertently increasing the speaker's nervousness. As a result, familiarity with the conversation partner diminishes and synlogic conversations are not caused. Thus, in the realm of cooperative communication design, it is important to balance the level of perceived "humanness" so as not to give the feeling of nervousness to people.

The elements contributing to "humanness" in this study encompassed aspects such as the aizuchi-bot's responses sounding human-like and the conversational partner being perceived as adaptable to the speaker's discourse. The former might involve utilizing a human voice in the aizuchi-bot, incorporating inflection in responses, or both. The latter emphasizes the importance of not defying the speaker's expectation that the aizuchi is responsive rather than automatic, thereby reinforcing the belief that the conversational partner is capable of evolving its speech patterns through a process of trial and error. Critical to this aspect is the refinement of aizuchi timing and the employment of contextually relevant inflections. The experiment, however, did not definitively identify which of these elements are essential, leaving scope for further exploration and experimentation.

In addition, "humanness" is the term that emerged as a result of conceptualization based on the interviews in this experiment. Therefore, it is possible that the term depends on the design of the experiment. In this experiment, the term "humanness" may be developed because we use not animal voices but human voices or synthetic voices. In light of the process obtained in section 5.2., it can be interpreted more abstractly. It is possible that the feeling that the partner with whom a person interacts is the person with whom he or she seems to be able to build a cooperative relationship may be a more important factor in causing co-adaptation rather than humanness. In this case, the interaction partner is not necessarily limited to humans, but may also include animals as well.

Table 4: Categories and concepts.

Categories	No	Concepts	Definitions (number of specific examples)
Perception of the conversation partner	11	Perceived the conversation partner as a machine	Feeling that the conversation partner is a machine or AI, mainly when aizuchis are uttered by a synthetic voice (8)
	12	Perceived the humanness in the conversation partner	Feeling of the humanness in the conversation partner when aizuchis are uttered by the human voice (8)
	6	Weirdness of the mixture of humanity and machine-like	Feeling the conversation partner weird because of humanity and machine-like at the same time (2)
Failure to establish cooperative relationships	9	Sense of absence of the conversation partner	Feeling that the conversation partner does not exist when aizuchis are uttered in synthetic voices (6)
	13	Fixed in the role of speaker	A state of being clearly aware that participants are in the position of the speaker (3)
	23	The partner who did not listen	Feeling that the conversation partner lost interests and was not listening to the participant (2)
	2	Lose motivation to continue the conversation	Feeling that the conversation partner with whom participants are having a conversation is not listening to participants, and having less motivation to continue talking (2)
Realization of casual conversations	16	The ease of a machine partner	The ease of not having to worry about or consider the partner's reaction because the conversation partner is a machine rather than a human being (5)
	18	Easy and fun like chatting	A state of being able to achieve the ease and enjoyment of usual conversation (3)
Open-endedness of conversations	1	Adding the content to talk about	A state wherein the subjects is highly motivated to continue the conversation and to add extra content to speech when they have finished talking about what they thought about beforehand (3)
	19	Expansion of ideas for talking	A state wherein ideas expand during speaking and participants talk about something other than what you initially thought of (4)
Becoming aware of the conversation partner's presence	10	Sense of being listened to	Feeling that the conversation partner is listening to participants because aizuchis are uttered (9)
	28	Presence of the conversation partner	Feeling that the conversation partner exists on the other side of the screen (8)
	17	Imagine the conversation partner	Imagining the conversation partner and having an image of their age, gender, and relationship to the participants concretely (3).
	5	Awareness of the conversation partner	A state of being aware of the presence of the conversation partner (6)
Considerate conversations	26	Cannot imagine the type of person	The inability to picture the conversation partner and not being able to imagine the partner (3)
	20	Feel like a first meeting	Feeling that the conversation partner was a new acquaintance (7)
	24	Consideration for the partner	Consideration and care for the conversation partner, such as taking the partner's feelings into account and devising the content of the conversation (6)
	25	Mental fatigue caused by attentiveness	The psychological load and mental fatigue caused by being considerate for the conversation partner (2)
Perception of incompatibility with the partner's aizuchi	22	Emotions are affected by the tone of voice in aizuchi	The state of being emotionally affected by the tone of the partner's voice (5)
	3	Unexpected timing of aizuchi	Aizuchi being uttered at unexpected times, such as in the middle of a word (10)
	29	Unpredictability of when aizuchis are uttered	Feeling that the timing of aizuchi is unpredictable (3)
	7	Mechanically uttered aizuchis	Perceiving that aizuchis are uttered mechanically because of the fixed tone of voice or unpredictable timing, etc. (4)
	27	Surprise or confusion about aizuchis	Surprise or confusion due to the difference from the imagined aizuchi (7)
	4	Interruption of the speech	Interruptions in conversation or thought due to the timing of aizuchi and feeling like noise (6)
Co-adaptation with aizuchis	15	Can explore the appropriate pace of the conversation	Be able to grasp and adapt to the pace of aizuchi while speaking (2)

14	Awareness of trying to speak according to aizuchis	The state of being conscious of speaking in accordance with aizuchi (7)
21	Devise the manner of speaking	Devising speaking style, such as having pauses, changing speaking speed or etc. (4)
8	Appropriate timing for aizuchi	Aizuchi is uttered at a time that participants feel appropriate (6)
30	Aizuchis are felt as natural even if aizuchi overlap with the speech	Do not mind if a conversation is overlapped by the insertion of aizuchis while speaking (4)

Moreover, the study revealed that when the aizuchi's human-like qualities are marred by mechanical timing, it results in a simultaneous perception of both human-like and machine-like attributes, leading to discomfort and apprehension towards the conversational partner. This phenomenon may be akin to encountering an 'uncanny valley' [65] in speech. Hence, it is imperative not to excessively enhance humanness.

Consequently, our objective should not be the complete anthropomorphism of AI to mimic human-like qualities, but rather to identify and incorporate the minimal "humanness" necessary for fostering cooperative communication. Clarifying these enabling factors will pave the way for establishing the essential conditions required to design a voice UI conducive to synlogic conversations.

6.2 Implications for computer-mediated-communication and digital wellbeing

The onset of the COVID-19 pandemic marked a significant shift in communication patterns, leading to a decline in in-person interactions and a surge in remote communication methods like videoconferencing and online chatting. This transition has notably impacted the co-constructive nature of daily communication, a phenomenon documented in various studies [55]. Over the last decade, a growing body of research has focused on the psychological implications of excessive reliance on digital screens, providing a wealth of correlational evidence pointing to its negative effects on mental health (e.g., [57]).

While these studies offer valuable insights, there is a pressing need for more comprehensive research to further elucidate these findings. This gap presents an opportunity to investigate the potential of fostering more engaging and positive online communication experiences to enhance social well-being. The interplay between digital media usage and subjective well-being has been a subject of debate, with studies revealing both beneficial and detrimental effects [56]. Consequently, there is a call for more in-depth research to better understand and improve the conflict-mitigation role of online communications [54], particularly in the context of current synlogic interactions. Given the inherently non-confrontational nature of synlogues, they hold significant promise for the development of more harmonious communication environments within society.

A notable observation from our study is the variability in participants' perception of virtual presence when interacting with a bot programmed to provide aizuchi, or backchannel responses. These findings contribute to the broader discourse on 'social presence' [62], a concept that designates the "the sense of "being with another" [63] which has been explored in social psychology and computer-mediated communication. A recent systematic review of this topic [64] delved into various factors that influence social presence, including cutting-edge media technologies like virtual and augmented reality. While the link between social presence and the emotional quality of communication necessitates further exploration [64], our study highlights a potential downside: the heightened perception of social presence can inadvertently increase user nervousness. This outcome parallels Miki's observations regarding the dynamics of power balance, such as harassment, in concessive joint action [48], yet differs in the nature of its negative impact. Further research is imperative to ascertain an optimal level and quality of cooperative overlap, contributing to the burgeoning field of human well-being in computer-mediated communication (CMC).

6.3 Limitations and Future work

In the analysis of this experiment, we aimed for a comprehensive qualitative analysis of the interviews in order to focus on subjective factors and effects on the human side, and the quantitative analysis was also based on an analysis of subjective evaluation rather than a measurement of the speed of user response. Therefore, the experiment was conducted on a limited number of participants and a specific age group in order to fully conduct a qualitative study. Although the content of the survey was not overly dispersed, it is fact that the number of participants of was too small. As the survey was conducted on young people, a separate experiment must be conducted to investigate whether the same results hold for older age groups.

In addition, the five items that showed significant differences in both human and synthetic voices ("loneliness is eased," "I feel closer to the other person," "I can feel compassion," "I can be at ease," and "I can talk with interest") were significantly improved even with short utterances of "aids" (i.e., a simple "hai" or "un"). However, because the subjects were silent conversational partners, it can be interpreted as simply concluding that those who responded were preferable to those who did not. Because we could not investigate this point in detail in this experiment, a separate experiment and a new study must be conducted to analyze whether the effect was really owing to aizuchi or simply because of the presence or absence of a response.

For this study, we chose to reimplement Okato et al.'s method to generate a predictive aizuchi system [9] because of its simplicity and ease of testing in an experiment. However, more sophisticated and complex systems to generate aizuchis have been studied [24, 40]. Recent relevant studies have mainly focused on generating backchannels at appropriate times, but few studies have been conducted on producing cooperative overlapping speech [53].

The results of this experiment indicate that the use of the aizuchi-bot broadens the topics of conversation and makes conversations

more open-ended. In chat-like conversations, the purpose is not to reach a conclusion, but to enjoy the process of conversation itself. In other words, in chatting among multiple people, it is not appropriate to interact with an agent whose conversation content is guided by the AI, but rather the effect of increasing the open-endedness of the conversation is important. Therefore, it is considered that using aizuchi-bot to support generating synlogic conversations will lead to more enjoyable chats in conversations among multiple people. However, the results of the test operation described in section 4.2 imply that there is a case that the speakers feel that the aizuchi-bot's responses are obstructive because they prioritize other human responses over the aizuchi-bot's responses. In the future, it is necessary to consider the possibility of supporting conversations among multiple humans by increasing the degree of involvement of the aizuchi-bot in the conversations, which could not be verified in this study. In this case, it is also necessary to analyze how the bot's utterance other than aizuchi, such as speech, supports to generate synlogic conversation, and what degree of involvement is desirable.

7 CONCLUSION

This study proposed the theoretical concept of synlogue based on previous work in linguistics and anthropology. Building upon prior studies in computational-model of backchannels, we validated the concept using aizuchi bots. We found that even a bot that only uttered short aizuchi such as "un" and "hai" could be used to find synlogic conversational features such as co-operation and open-endedness. We used M-GTA to clarify the process by which such synlogic interactions occurred in conversations with aizuchi-bots. Consequently, it was found that different interactions occurred depending on whether or not the user perceived human-like qualities in the conversation partner. In the case of perceiving humanness or the presence of the other party in the conversational partner, open-endedness occurred after a process of co-adaptive interaction; however, in the case of perceiving machine-likeness in the conversational partner, such a process was not followed in the present experiment. Moreover, it was determined that it is necessary to establish a more empathic relationship with the other party; therefore, the perception of humanness and the presence of the other party could be important factors.

In proposing the design concept of synlogue, which should be understood as a design space that focus on the co-constructive aspect in both verbal and non-verbal communications, and not as a brand new linguistic model, we do not aim to be prescriptive nor conclusive: we intended to offer a generative approach for designers and researchers to explore co-constructive convivial and open-ended relationships through the synlogic qualities of *incompleteness*, *overlap*, *multimodality* and *co-adaptation*.

ACKNOWLEDGMENTS

This paper is supported by JSPS KAKENHI Grant Number JP21K18344 and JP21H03768 to D. C., and Grant 0470 from the Templeton World Charity Foundation, Inc. to O. W. The authors thank the anonymous reviewers for their critical and insightful comments. We would also like to thank Lana Sinapayen, Noboru Yasuda, and Hiromichi Hosoma for their precious advice and input.

REFERENCES

[1] Nobuko Mizutani. 1988. Aizuchi-ron (in Japanese) [on backchannels]. Nihongogaku (Japanese Linguistics), 7(13), 4–11.
[2] Nobuko Mizutani. 1993. Kyoowa'kara 'taiwa'e (in Japanese) [from 'kyoowa'to 'taiwa']. Nihongogaku (Japanese Linguistics), 12(4), 4–10.
[3] Bronisław Malinowski. 1923. The problem of meaning in primitive languages. The meaning of meaning (eds) CK Ogden & IA Richards. London: Kegan Paul, Trench and Trubner. I, 935.
[4] Dunbar, R. I. M. 1996. Grooming, gossip, and the evolution of language. Harvard University Press.
[5] Sotaro Kita and Sachiko Ide. 2007. Nodding, aizuchi, and final particles in japanese conversation: How conversation reflects the ideology of communication and social relationships. Journal of Pragmatics, 39(7):1242–1254. Nodding, aizuchi, and Final Particles in Japanese Conversation
[6] Senko K. Maynard. 1993. Kaiwabunseki (in Japanese) [Conversation analysis] Research Series of Contrast between English and Japanese, 2. Kuroshio. ISBN: 978-4874240717
[7] Scott Saft. 2007. Exploring aizuchi as resources in japanese social interaction: The case of a political discussion program. Journal of Pragmatics, 39(7):1290–1312. https://doi.org/10.1016/j.pragma.2007.02.010
[8] Hiroki Mori. 2013. Dynamic aspects of aizuchi and its influence on the naturalness of dialogues. Acoustical Science and Technology, 34(2):147–149. https://doi.org/10.1250/ast.34.147
[9] Yohei Okato, Keiji Kato, Mikio Kamamoto, and Shuichi Itahashi. 1996. Insertion of interjectory response based on prosodic information. In Proceedings of IVTTA'96. Workshop on Interactive Voice Technology for Telecommunications Applications, pages 85–88. IEEE. https://doi.org/10.1109/IVTTA.1996.552766
[10] Junzo Kawada. 2001. Kôtô Denshô Ron (in Japanese) [On Oral Traditions]. Heibonsha. ISBN: 978-4582763898
[11] Dominique Chen. 2020. Mirai wo tsukuru kotoba: wakari aenasa wo tsunagu tame ni (in Japanese) [Language to construct the future: to mediate misunderstandings]. Shinchosha. ISBN: 978-4101042411
[12] Daiji Kimura. 1995. Speech overlap and long silence among the baka pygmy. Journal of African Studies, 1995(46):1–19. https://doi.org/10.11619/africa1964.1995.1
[13] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735. https://doi.org/10.2307/412243
[14] Don H. Zimmerman, and Candace West. 1975. Sex roles, interruptions and silences in conversation: In thorne, b. & henley, n.(eds.), language and sex: Difference and dominance (pp. 105-129).
[15] Sheida White. 1989. Backchannels across cultures: A study of americans and japanese. Language in society, 18(1), 59–76. https://www.jstor.org/stable/4168001
[16] Deborah Tannen. 1981. New york jewish conversational style. International Journal of the Sociology of Language., 30, 133–150. https://doi.org/10.1515/ijsl.1981.30.133
[17] Deborah Tannen. 1983. When is an overlap not an interruption? one component of conversational style. In The first Delaware symposium on language studies, volume 4, pages 119–129. University of Delaware Press Newark, NJ.
[18] Deborah Tannen. 1994. Gender and discourse. Oxford University Press.
[19] Deborah Tannen. 2005. Conversational Style: Analyzing Talk among Friends. Oxford University Press.
[20] Stefan Pfänder, and Elizabeth Couper-Kuhlen. 2019. Turn-sharing revisited: An exploration of simultaneous speech in interactions between couples. Journal of Pragmatics, 147:22–48. https://doi.org/10.1016/j.pragma.2019.05.010
[21] Donald A. Schön. 1987. Educating the Reflective Practitioner. Jossey-Bass Publishers. https://doi.org/10.1002/chp.4750090207
[22] Yoshiko Kawabata, and Toshihiko Matsuka. 2021. Aizuchi as a sign of internal information processing and its interpretations by listeners. In 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE. pages 380–385.
[23] Donna J. Haraway. 2016. Staying with the trouble: Making kin in the Chthulucene. Duke University Press.
[24] Nigel Ward, and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in English and Japanese. Journal of pragmatics, 32(8):1177–1207. https://doi.org/10.1016/S0378-2166(99)00109-5F
[25] Koichiro Kokubun. 2017. Chu-do-tai no Sekai: Ishi to Sekinin no Kokogaku (in Japanese) [The world of the Middle Voice: an archaeology of will and responsibility]. Igakushoin. ISBN: 978-4260031578
[26] Francisco J. Varela. 1989. Autonomie et connaissance: essai sur le vivant (in French) [Principles of biological autonomy]. Paris: Seuil.(Original work published 1979). ISBN: 978-2020100304
[27] Hiroyuki Yano, and Akira Ito. 1996. Toward constructing a dialogue management model for kyōwa. In Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA. IEEE. pages 489–494. https://doi.org/10.1109/ROMAN.1996.568886
[28] Daisuke Nakamichi, Shuichi Nishio, and Hiroshi Ishiguro. 2014. Training of telecommunication through teleoperated android 'telenoid' and its effect. In The

23rd IEEE International Symposium on Robot and Human Interactive Communication, pages 1083–1088. IEEE. https://doi.org/10.1109/ROMAN.2014.6926396
[29] Yusuke Iwabuchi, Ikuma Sato, Yuichi Fujino, and Norihito Yagi. 2019. The communication supporting robot based on 'humanitude' concept for dementia patients. In 2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech), pages 219–223. IEEE. https://doi.org/10.1109/LifeTech.2019.8884049
[30] Tsunehiro Arimoto, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2014. Nodding responses by collective proxy robots for enhancing social telepresence. In Proceedings of the second international conference on Human-agent interaction (HAI '14). Association for Computing Machinery, New York, NY, USA, 97–102. https://doi.org/10.1145/2658861.2658888
[31] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. Telling Stories to Robots: The Effect of Backchanneling on a Child's Storytelling. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17). Association for Computing Machinery, New York, NY, USA, 100–108. https://doi.org/10.1145/2909824.3020245
[32] Michael Murray, Nick Walker, Amal Nanavati, Patrícia Alves-Oliveira, Nikita Filippov, Allison Sauppe, Bilge Mutlu, and Maya Cakmak. 2022. Learning backchanneling behaviors for a social robot via data augmentation from human-human conversations. In Conference on Robot Learning, pages 513–525. PMLR.
[33] Yohei Okato, Keiji Kato, Mikio Yamamoto, and Shuichi Itahashi. 1999. Inritsu jouhou wo mochiita aizuchi no sounyuu (in Japanese) [Giving 'aizuchi' Using Prosodic Information]. Jouhoushori gakkai ronbun-shi (Information Processing Society of Japan), 40(2): 469-478. http://id.nii.ac.jp/1001/00012826/
[34] Kazumi Ogawa and Kazushi Saito. 2007. Denwa bamen ni okeru aizuchi no shurui no oosa ga inshou ni ataeru eikyou (in Japanese) [Impact of the variety of aizuchis on impressions in telephone situations]. Nihon shinrigakkai taikai happyou ronbunshuu, 71. https://doi.org/10.4992/pacjpa.71.0_1PM017
[35] Takashi Tsuzuki and Yasuyuki Kimura. 2000. Daigakusei ni okeru media communication no shinrigakuteki tokusei ni kansuru bunseki: taimen, keitaidenwa, keitai mail, denshi mail no hikaku (in Japanese) [An Analysis of the Psychological Properties of Media Communication Among University Students]. Ouyou shakaigaku kenkyu (Journal of Applied Sociology), 42, 15-24. https://rikkyo.repo.nii.ac.jp/records/1756 oai:rikkyo.repo.nii.ac.jp:00001756
[36] Masanori Kimura, Yukiko Iso, Akiko Sakuragi and Ikuo Dibo. 2005. Sansyakan kaiwa bamen ni sikaku media ga hatasu yakuwari: egao to unazuki no kyoushutsu oyobi sorera no koudou matching ni chumoku shite (in Japanese) [The role of visual media in triadic communication: The expressions of smile and nodding, and their behavior matching], Taijin shakai sinrigaku kenkyu, 5, 39-47. https://doi.org/10.18910/4818
[37] Yasuhito Kinoshita. 2020. Teihon M-GTA: jissen no rironka wo mezasu shitsutekikenkyu houhou-ron (in Japanese) [Revised edition of the book M-GTA]. Igakushoin. ISBN: 978-4260042840
[38] Barney G. Glaser, and Anselm L. Strauss. 1967. Discovery of grounded theory: Strategies for qualitative research. Aldine Publishing Company.
[39] Amy C. Edmondson, Roderick M. Kramer, and Karen S. Cook. 2004. Psychological safety, trust, and learning in organizations: A group-level lens. Trust and distrust in organizations: Dilemmas and approaches, 12(2004), 239-272.
[40] Louis-Philippe Morency, Iwan De Kok, and Jonathan Gratch. 2008. Predicting listener backchannels: A probabilistic multimodal approach. In International Workshop on Intelligent Virtual Agents, vol 5208, pp. 176-190. IVA. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-85483-8_18
[41] Tsuyoshi Ono, and Eri Yoshida. 1996. A study of co-construction in Japanese: We don't finish each other's sentences. Noriko Akatsuka, Shoichi lwasaki and Susan Strauss (eds), Japanese/Korean Linguistics, 5, 115-130.
[42] Takashi Suzuki, amd Mayumi Usami. 2006. Co-constructions in English and Japanese revisited: A quantitative approach to cross-linguistic comparison. Gengo Jōhōgaku Kenkyū Hōkoku, 263-276.
[43] Nigel Ward, and Wataru Tsukahara. 1999. A responsive dialog system. In: Yorick Wilks (eds) Machine Conversations. The Springer International Series in Engineering and Computer Science, vol 511, 169-174. https://doi.org/10.1007/978-1-4757-5687-6_14
[44] Herbert H. Clark, and Susan E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), Perspectives on socially shared cognition, pp. 127–149. American Psychological Association. https://doi.org/10.1037/10096-006
[45] Gene Lerner. 2002. Turn-sharing: The choral co-production of talk-in-interaction. Chapter 9 of The Language of Turn and Sequence, ed. C. Ford, B. Fox, and S. Thompson, 225-256.
[46] Deborah Tannen. 2007. Talking voices: Repetition, dialogue, and imagery in conversational discourse, Vol. 26). Cambridge University Press.
[47] Emanuel A. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. Language in society, 29(1), 1-63.
[48] Nayuta Miki. 2022. Concessive Joint Action: A New Concept in Theories of Joint Action. Journal of Social Ontology, 8(1), 24-40. https://doi.org/10.25365/jso-2022-7307
[49] Pierr Saint-Germier, Cédric Paternotte, and Clément Canonne. 2021. Joint Improvisation, Minimalism and Pluralism about Joint action. Journal of Social Ontology, 7(1), 97-118. https://doi.org/10.1515/jso-2020-0068
[50] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, pp. 127-136. Saarbrücken, Germany. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-5516
[51] Zofia Malisz, Marcin Włodarczak, M, Hendrik Buschmeier, Joanna Skubisz, J., Kopp, S., & Wagner, P. 2016. The ALICO corpus: analysing the active listener. Language resources and evaluation, 50, 411-442. https://doi.org/10.1007/s10579-016-9355-6
[52] Patrick G. T. Healey, Nicola Plant, Christine Howes, and Mary Lavelle. 2015,. When words fail: Collaborative gestures during clarification dialogues. In 2015 AAAI Spring Symposium Series.
[53] Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: a review. Computer Speech & Language, 67, 101178. https://doi.org/10.1016/j.csl.2020.101178
[54] Danica Radovanovic, and Massimo Ragnedda. (2012). Small talk in the digital age: Making sense of phatic posts. Proceedings of the WWW'12 Workshop on 'Making Sense of Microposts', pages 10–13.
[55] Robin Dunbar. 2022. How many friends does one person need? In How Many Friends Does One Person Need? Harvard University Press. ISBN: 978-0674057166
[56] Dong Liu, Roy F. Baumeister, Chia-chen Yang, and Baijing Hu. 2019. Digital communication media use and psychological well-being: A meta-analysis. Journal of Computer-Mediated Communication, 24(5):259–273. https://doi.org/10.1093/jcmc/zmz013
[57] Jean M. Twenge, Jonathan Haidt, Andrew B. Blake, Cooper McAllister, Hannah Lemon, and Astrid Le Roy. 2021. Worldwide increases in adolescent loneliness. Journal of Adolescence, 93:257–269. https://doi.org/10.1016/j.adolescence.2021.06.006
[58] Hiroki Kojima, Dominique Chen, Mizuki Oka, and Takashi Ikegami. 2021. Analysis and design of social presence in a computer-mediated communication system. Frontiers in Psychology, 12:641927. https://doi.org/10.3389/fpsyg.2021.641927
[59] Peter-Paul Verbeek. 2015. Beyond interaction: a short introduction to mediation theory. interactions 22, 3 (May - June 2015), 26–31. https://doi-org.waseda.idm.oclc.org/10.1145/2751314
[60] Siddharth Reddy., Sergey Levine, and Anca Dragan. 2022. First Contact: Unsupervised Human-Machine Co-Adaptation via Mutual Information Maximization. Advances in Neural Information Processing Systems, 35, 31542-31556.
[61] Hiroki Kojima, Tom Froese, Mizuki Oka, Hiroyuki Iizuka, and Takashi Ikegami. 2017. A Sensorimotor Signature of the Transition to Conscious Social Perception: Co-regulation of Active and Passive Touch. Frontiers in Psychology, 8, 239858. https://doi.org/10.3389/fpsyg.2017.01778
[62] John Short, Ederyn Williams, and Bruce Christie. 1976. The Social Psychology of Telecommunications. London: John Wiley & Sons. ISBN: 978-0471015819
[63] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a more robust theory and measure of social presence: review and suggested criteria. Presence: Teleoperators and Virtual Environments, 12, 456–480. https://doi.org/10.1162/105474603322761270
[64] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A systematic review of social presence: definition, antecedents, and implications. Front. Robot. https://doi.org/10.3389/frobt.2018.00114
[65] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. IEEE Robotics & automation magazine 19, 2 (2012), pp. 98–100. https://doi.org/10.1109/MRA.2012.2192811
[66] Ayako Hashizume, and Masaaki Kurosu. 2013. Role of Kansei Experience for the Active Use of ICT among the Elderly. International Journal of Affective Engineering, vol.12, no.2, p. 111-117. https://doi.org/10.5057/ijae.12.111
[67] Ryo Odachi, Tomoko Tamaki, Mikiko Ito, Taketoshi Okita, Yuri Kitamura, and Tomotaka Sobue. 2017. Nurses' Experiences of End-of-life Care in Long-term Care Hospitals in Japan: Balancing Improving the Quality of Life and Sustaining the Lives of Patients Dying at Hospitals. Asian Nursing Research, Volume 11, Issue 3, 207-215. https://doi.org/10.1016/j.anr.2017.08.004</bib>
[68] Masanori Kimura, Masao Yogo, and Ikuo Daibo. 2005. Kanjou episode no kaiwa bamen ni okeru hyoushutsusei halo kouka no kentou(in Japanese) [Examining the expressivity halo effects in conversational situations of emotional episodes]. Kanjou shinrigaku kenkyuu (The Japanese journal of research on emotions). 12(1): 12-23. https://doi.org/10.4092/jsre.12.12
[69] Kaori Okamoto, and Susumu Takahashi. 2006. Shinmitsudo no chigai oyobi communication keitai no chigai ga media communication kan ni oyobosu eikyou (in Japanese) [The impacts that are provided to the views of communication by differences in intimacy and forms of media communication]. Jikken shakai shinrigaku kenkyuu (The Japanese journal of experimental social psychology). 45(2): 85-97. https://doi.org/10.2130/jjesp.45.85
[70] Takeshi Fujiwara, and Ikuo Daibo. 2013. Kanjou ga kaiwamanzokudo ya taijininshou ni ataeru eikyou (in Japanese) [The impacts of emotions on conversational satisfaction and interpersonal impressions]. Shinrigaku kenkyuu(Japanese Journal of Psychology). 84(5): 522-528. https://doi.org/10.4992/jjpsy.84.522

A A APPENDICES

The appendices are presented below.

A.1 Questionnaire Items Used for the Subjective Evaluation

The questionnaire used for the subjective evaluation is as follows. This questionnaire items consisted of some from existing studies [35, 36, 68] for reference and some added on our own. Items 1.1 to 1.15 are 15 items from the 16 evaluation items used by Tsuzuki and Kimura in their analysis which clarified the factors that cause differences in psychological characteristics toward conversation partners in different modes of media communication [35]. Only the item "effective in gathering information" was excluded because it could not be used under our experimental conditions. Items 1.16 to 1.20 are five items added by the authors. Also, items 2.1 to 2.3 are three items used by Kimura et al. in their investigation of the effects of smiles, nods and behavior matchings on conversational satisfaction in triadic conversations [36]. We used the same items because the study also concerned backchannels such as nodding and facial expressions which are the same as aizuchis. These three items were selected from the items used in a study that examined the effects of observers' judgments that the more active the speaker's expression of emotion is, the more successful communication is [68]. All items from existing studies were devised based on previous studies referred to in the respective references, and they were also used in some other studies which analyzed correlations or factors in intimacy or conversational satisfaction by form of media communication [69, 70].

The answer choices were provided on a 5-point scale from "not at all applicable" to "very applicable" for question 1, and from "I disagree very much" to "I agree very much" for question 2.

1. For each of the items, how much of the following applies to you in the (silent / synthesized voice aizuchis / human voice aizuchis) pattern of conversation?
   - 1.1 Relieve loneliness.
   - 1.2 Be fun.
   - 1.3 Be nervous.
   - 1.4 Communication intentions are conveyed quickly.
   - 1.5 Feel close to the partner.
   - 1.6 Be easy to open up.
   - 1.7 Have a hard time.
   - 1.8 Compassion can be realized.
   - 1.9 Be formal.
   - 1.10 Be easy to communicate your intentions.
   - 1.11 Can talk about personal stories.
   - 1.12 Be easy-going.
   - 1.13 Have a purpose.
   - 1.14 Can concentrate.
   - 1.15 Get tired.
   - 1.16 Talks are expanded beyond the topic.
   - 1.17 Feel more mentally relaxed.
   - 1.18 Feel anxious.
   - 1.19 Be easy to come up with things to talk about while talking.
   - 1.20 Can speak smoothly.

2. Please describe your overall impression of the conversation in the pattern of (silent / aizuchi with the synthesized voice / aizuchi with the human voice interaction).
   - 2.1 The conversation progressed cooperatively.
   - 2.2 It was hard to have conversations.
   - 2.3 You can be interested in conversation.

A.2 Explanation of Categories and Concepts

This appendix defines each category, the concepts that comprise it, and the relationships among the concepts by quoting statements made by the experimental participants in the interviews. The concept names and definitions for each category and the number of specific examples are listed in Table 4. In the following, the categories are indicated in [ ] , the concepts in " " and definitions of categories and concepts in < >. The speaker's subject ID is indicated in parentheses at the end of each quotation phrase.

a. [Perception of the conversation partner]

This category indicates <what kind of existence the conversational partner is perceived as depending on the type, inflection, timing of the voice of aizuchi>. It consists of three concepts: (11) "Perceive the conversation partner as a machine," (12) "Perceive the humanness in the conversation partner," and (6) "Weirdness of the mixture of humanity and machine-like." When synthetic voice is used for the conversational partner's aizuchi, participants often (11) "perceive the conversation partner as a machine." This concept is defined as <feeling that the conversation partner is a machine or AI, mainly when aizuchis are uttered by a synthetic voice>.

> "When it was a machine sound, after all, it was easy for me to think, 'The partner I'm talking to is definitely not a human but a machine."' (2211KF)
>
> "Since it's a synthetic voice, I know it's a machine, so [. . .]" (2213EM)
>
> "It was like Siri. It was like Alexa. I felt like I was talking to a gadget." (2413FF)

On the other hand, the interviewees stated that they (12) "perceive the humanness in the conversation partner" when they heard a human voice or sensed intonation in their online conversational partner's aizuchi. We defined this concept as <feeling of the humanness in the conversation partner when aizuchis are uttered by the human voice>.

> "It sounded like a human voice, so it wasn't machine-like. It made me to think that I talk with a real person, . . . so it was easy to talk to." (2314NM)
>
> "The third one (human voice) is basically very easy to talk to, and after all, a human voice is more human-like and [. . .]" (2413FF)
>
> "If anything, certainly, I think that the more intonation, the more human-like I felt." (2415SM)

As described above, in many cases, the type of conversational partner is perceived differently depending on the type of voice or inflection, but even aizuchi uttered by a human voice may be perceived as mechanical because the timings of aizuchis don't match with the speech. In this case, the human-like quality because of human voices and the machine-like quality because of the timings are percieved at the same time, resulting in a sense of (6) "Weirdness of the mixture of humanity and machine-like." This concept

was defined as <feeling the conversation partner weird because of humanity and machine-like at the same time>.

> "I had never heard realistic human voices uttered from machines, so I was a little scared, and I was not used to hearing them." (2211KF)
>
> "I'm not really accustomed to the human voice aizuchis intervening quite frequently in my speech." (2411MM)

b. [Failure to establish cooperative relationships]

This category indicates <a situation in which a cooperative relationship is not established with the aizuchi uttered from the conversational partner>. This category consists of four concepts: (9) "Sense of absence of the conversation partner," (13) "Fixed in the role of speaker," (23) "The partner who doesn't listen," and (2) "Lose motivation to continue the conversation." When the speaker perceives the machine partner as a machine, (9) "sense of absence of the conversation partner" is increased. This concept is defined as <feeling that the conversation partner does not exist when aizuchis are uttered in synthetic voices>.

> "In that respect, I thought the pattern of the silence and the mechanical voice were pretty similar, and it was pretty easy to recognize more strongly that 'I guess I don't have a conversation partner,' so [. . .]" (2211KF)
>
> "I felt that I probably don't have (a conversation partner) in the case of the artificial one." (2814TF)

When the speaker does not feel the presence of the partner in the conversation or feels that the aizuchi uttered is mechanical, one feels as if one is speaking to (23) "the partner who doesn't listen." This concept is defined as <feeling that the conversation partner lost interests and was not listening to the participant>.

> "The tone of her voice made me think she wasn't interested in what I said and [...] I was thinking that she probably wasn't listening." (2214MF)
>
> "I thought, 'If she was a human being, she probably wasn't listening and wasn't interested at all."' (2414AM)

When a speaker feels that the conversation partner is not present, or that the partner is not listening, the speaker feels that he or she is (13) "fixed in the role of speaker." This concept is defined as <a state of being clearly aware that participants are in the position of the speaker>.

> "(Because I felt like I was talking to a machine,) I felt a bit like I was talking alone." (2814TF)
>
> "(In the pattern of synthetic voice,) I thought that there was a strong sense of duty to speak what I planned to talk about beforehand. There is a strong sense of being restrained by the role of speaking." (2313OF)

When a speaker feels the presence of the conversation partner or feels that the other person is not listening to him or her, (2) "lose motivation to continue the conversation." This concept is defined as <feeling that the conversation partner with whom participants are having a conversation is not listening to participants, and having less motivation to continue talking>.

> "I lost the will to speak, and I didn't really know what I was talking for." (2214MF)
>
> "Even if I thought the partner is listening to me a little bit, I felt like, 'It's just a computer program' (and I lost my motivation to speak)." (2315SM)

c. [Realization of casual conversations]

This category indicates <a situation in which the participant is in an easygoing mood and enjoys conversing with the conversation partner>. It consists of two concepts: (16) "The ease of a machine partner" and (18) "Easy and fun like chatting." When the participants felt that their conversation partner was a machine, they sometimes felt (16) "the ease of a machine partner" in talking because the partner was not a human. This concept was defined as <the ease of not having to worry about or consider the partner's reaction because the conversation partner is a machine rather than a human being>.

> "In that aspect, there was a certain carefreeness to talk, which I think led to the ease of speaking with the last one, synthetic voice." (2213EM)
>
> "I don't know if there is such a thing as a talking machine, but the conversation partner seemed more like that and I didn't give much consideration to the partner." (2215TM)

In addition, when the participants had a natural impression of their conversation partner's delivery, they felt that the conversation was more like a normal conversation, and they sometimes felt that it was (18) "easy and fun like chatting." This concept can be defined as <a state of being able to achieve the ease and enjoyment of usual conversation>.

> "I think that the patterns with aizuchi creates a more conversational atmosphere and enhances the chatting-like atmosphere, so I don't think anyone concentrates on talking during the chat, so I could naturally talk without concentrating and relax in a good way." (2213EM)

d. [Open-endedness of conversations]

This category indicates <the state in which the conversation diverges into a variety of topics that deviate from the topic due to the expansion of ideas that occur during the conversation>. This category consists of two concepts: (1) "Adding the content to talk about" and (19) "Expansion of ideas for talking." In the experiment, In the experiment, (1) "adding the content to talk about" was an action resulting from the desire to entertain the conversation partner. This concept is defined as <a state in which the subject is highly motivated to continue the conversation and to add extra content to speech when they have finished talking about what they thought about beforehand>.

> "With aizuchis, the partner would often say 'un un,' so when I was at a loss for a word, it made me feel like, 'I'll try to keep talking."' (2211KF)
>
> "When I ran out of something to talk about, I felt more like 'what should I talk about?' (difficulty in coming up with something) without aizuchis than with." (2311IF)

When the participants tried to add to the content of their speech as described above, or when they became excited during talking because of the ease of chatting, (19) "expansion of ideas for talking" that deviated from the given topic was sometimes generated. This concept is defined as <a state in which ideas expand during speaking and participants talk about something other than what you initially thought of>.

> "By being able to relax and talk a bit, I was able to broaden my ideas for talking because I didn't have to focus too much to talk." (2213EM)

> "The second pattern of conversation (human voice) led to a different content than what I really planned to talk about, so [...] I think the flow of the conversation was different from what I had intended." (2414AM)

e. [Becoming aware of the conversation partner's presence]

This category indicates <the state of being aware of the presence of the conversation partner while talking>. It consists of four concepts: (10) "Sense of being listened to," (28) "Presence of the conversation partner," (17) "Imagine the conversation partner," and (5) "Awareness of the conversation partner." When the participants felt that the conversation partner was human and his or her aizuchis were inserted between conversations, the (10) "sense of being listened to" was improved. This concept is defined as <feeling that the conversation partner is listening to participants because aizuchis are uttered>.

> "Well, in the case with aizuchi, I felt like he or she was listening to me, [. . .] so I kept talking." (2313OF)

> "Because aizuchi's response to what I said was, uh, well, it would be an exaggeration to say that she approved me, but . . . it's because I felt that she understood or recognized what I said." (2814TF)

When they felt that they were being listened to, the (28) "presence of the conversation partner" was promoted. This concept was defined as <feeling that the conversation partner exists on the other side of the screen>.

> "In the case of the voice saying 'un' in a normal human voice, I felt the most like the conversation partner was there." (2814TF)

> "Because the human voice pattern was the first task of this experiment and I felt the partner was listening to me, I was nervous when I was talking with the human voice." (2411MM)

When participants are able to sense the presence of the conversation partner, a concrete image of what kind of person he or she is conversing with is generated, and they became to be able to (17) "imagine the conversation partner." This concept is defined as <imagining the conversation partner and having an image of their age, gender, and relationship to the participants concretely>.

> "I imagined that the partner was not someone who was much older than me, but someone with whom I could talk in a frank manner." (2213EM)

> "When I heard the voice, it sounded a bit like my own, and I imagined a woman who is as old as me." (2313OF)

When the presence of the conversation partner is felt, or when the imagination of what kind of person is talking with is generated, (5) "awareness of the conversation partner" seems to develop. This concept is defined as <a state of being aware of the presence of the conversation partner>.

> "In the first and second patterns, I just talked and didn't think about the partner at all, but in the third pattern (the human voice pattern), I began to try communicating with the partner exactly what I was talking about." (2413FF)

> "The third one (the human voice pattern) gave me the feeling that I was being listened to. I guess it could be called a kind of pressure, but it gave me the feeling that I was being listened to." (2213EM)

f. [Considerate conversations]

This category indicates <the state in which the awareness of the presence of the conversation partner causes the state in which the speaker tries to entertain the conversation partner or cares not to make the partner feel uncomfortable>. It consists of four concepts: (26) "Can't imagine the type of person," (20) "Feel like a first meeting," (24) "Consideration for the partner," and (25) "Mental fatigue caused by attentiveness." In the process of becoming aware of the conversation partner, the participants sometimes imagine what the partner is like, but as a result of their imagination, they may not have a concrete image of the partner, or the image may be betrayed. The concept defined as <the inability to picture the conversation partner and not being able to imagine the partner>, and we named it (26) "Can't imagine the type of person."

> "The conversation partner . . . I can't assume the partner who I'm talking to. I couldn't do that. When I started speaking with the partner in mind, I heard a tone of voice that was different from what I imagined." (2413FF)

> "I couldn't guess the partner, but I thought that maybe it was because I couldn't see his or her face." (2415SM)

Because of the unimaginability of the conversation partner, the speakers felt the psychological distance from the partner and sometimes felt that they (20) "feel like a first meeting." This concept can be defined as <feeling that the conversation partner was a new acquaintance>.

> "It was kind of like a situation where I introduced myself to someone whom I met for the first time and I was going to be friends." (2215TM)

> "As with talking about my personal story, I was nervous about suddenly talking to someone I had never met before for a few minutes." (2311IF)

When the participants feel psychological distances from their conversation partners, they try to entertain them and avoid causing them discomfort. It causes (24) "consideration for the partner." This concept is defined as <consideration and care for the conversation partner, such as taking the partner's feelings into account and devising the content of the conversation>.

> "When I was talking, I felt anxious about the possibility that the partner might feel uncomfortable. [...] He or she might think that I was annoying or that I was talking too fast." (2215TM)

> "I tried to entertain the listener and make her laugh, and when it didn't go well, I got tired." (2216NF)

As a result of this concern for the other person, the ease of conversation was reduced and the (25) "Mental fatigue caused by attentiveness" was increased. This concept can be defined as <The psychological load and mental fatigue caused by being considerate for the conversation partner>.

> "(I tried to make her laugh, but) when the tone of her voice was low, I felt like I had to make her laugh more, and it became tiring." (2216NF)
>
> "I felt that it was a little hard for me (to talk to someone I had never met before)." (2311IF)

g. [Perception of incompatibility with the partner's aizuchi]

This category indicates <the speaker's own feeling that the tone and the timing of aizuchis uttered by the conversation partner does not match the content, the tone of voice, and the timing of what the speaker says>. It consists of six concepts: (22) "Emotions are affected by the tone of voice in aizuchi," (3) "Unexpected timing of aizuchi," (29) "Unpredictability of when aizuchis are uttered," (7) "Mechanically uttered aizuchis", (27) "Surprise or confusion about aizuchis", and (4) "Interruption of the speech." It may be caused <the state of being emotionally affected by the tone of the partner's voice>, for example, when a dark tone of voice is used as aizuchis during a conversation, the speaker feels down. The concept defined by this is named (22) "emotions are affected by the tone of voice in aizuchi."

> "I think that the context of aizuchi changes depending on the tone of voice and the content of the conversation, but his or her aizuchis seemed to be listening to the serious conversation, so I got worried if he or she was okay. [. . .] After all, the tone of aizuchi was so different from the tone of my talk (so I felt I was not good at the partner)." (2214MF)
>
> "It contained the tone of the voice uttered when a person couldn't understand what was said, and [. . .] (it made me anxious)." (2413FF)

There are times when the speaker feels that the timing does not match the aizuchi uttered by the conversational partner, and (3) "unexpected timing of aizuchi" is uttered. This concept is defined as <aizuchi being uttered at unexpected times, such as in the middle of a word>.

> "When I thought 'This is when aizuchi should be uttered,' but it didn't, I would get stuck and wonder. On the other hand, there were times when I thought 'Oh, you're going to utter it now."' (2211KF)
>
> "I felt uncomfortable when I heard aizuchi while I'm uttering a word of a sentence, not between clauses" (2215TM)

When aizuchis are often inserted at unexpected times, the speakers feel (29) "unpredictability of when aizuchis are uttered." This concept is defined as <feeling that the timing of aizuchi is unpredictable>.

> "On the contrary, I couldn't imagine it (the timing of aizuchi), so I thought that if aizuchi was uttered again, I would just have to deal with it then." (2215TM)
>
> "Well, maybe . . . I thought it was a program that uttered aizuchi at the random timing, so I thought I shouldn't expect anything like it responded at appropriate timing." (2315SM)

In some cases, the unpredictability of the timing of the aizuchi caused the speakers to perceive the aizuchi as being uttered at random times and (7) to think it was "mechanically uttered aizuchis." This concept is defined as <perceiving that aizuchis are uttered mechanically because of the fixed tone of voice or unpredictable timing, etc.>.

> "The voice is human-like, but something is different. [...] I guess it's because of the timing of aizuchi." (2211KF)
>
> "On the other hand, the artificial one uttered like 'hai', 'hai' all the time (it was a constant tone), so I felt it was a bit . . . a bit unnatural." (2814TF)

When an aizuchi is inserted at an unexpected time, or when an aizuchi is uttered in a different tone of voice than expected, it seems that (27) "surprise or confusion about aizuchis" is caused. We defined this concept as <surprise or confusion due to the difference from the imagined aizuchi>.

> "I was surprised because the tone of the voice was different from the reaction I expected." (2216NF)
>
> "(When aizuchi overlaps with my speech at unnatural timing) I guess maybe it's the confusion." (2311IF)

When the aizuchis were inserted at unexpected times, or when the participants felt confused by aizuchis, (4) "interruption of the speech." occurred. This concept is defined as <interruptions in conversation or thought due to the timing of aizuchi and feeling like noise>.

> "When aizuchis overlapped, I felt as if I was being interrupted by it because of the loudness of his or her voice." (2314NM)
>
> "When aizuchis overlapped, I felt aizuchi's voice was loud and I was disturbed by it." (2315SM)

h. [Co-adaptation with aizuchis]

This category indicates <a state of co-adaptation with the conversation partner's aizuchi, in which the conversation with the aizuchis feels natural by the speaker's devices to adjust speaking speed and pauses in conversation to the aizuchi, who felt that the timing was not right>. This category consists of five concepts: (15) "Can explore the appropriate pace of the conversation," (14) "Awareness of trying to speak according to aizuchis," (21) "Devise the manner of speaking," (8) "Appropriate timing for aizuchi," and (30) "Aizuchis are felt natural even if aizuchi overlap with the speech." When the speaker feels that the partner to whom he or she is speaking is human, he or she tends to try to speak while assuming that he or she <is able to grasp and adapt to the pace of aizuchi while speaking>. In this case, it seems that they feel they (15) "can explore the appropriate pace of the conversation" with their conversation partner.

> "I thought, 'This is about the right timing to talk.' [. . .] If it's a live person and a machine, I think it's the human who can flexibly adapt to me, so [...]" (2213EM)

When participants feel that their conversation partner is listening to them, even if the aizuchi is uttered at an unexpected time, they seem to develop (14) "awareness of trying to speak according to aizuchis." This concept is defined as <the state of being conscious of speaking in accordance with aizuchi>.

> "If anything, I thought 'I' ll wait for aizuchis' more in the case of the human voice pattern." (2311IF)
>
> "I did not stop speaking consciously, but I was imagining in my head when aizuchis would be uttered while talking." (2214MF)

When the participants were conscious of trying to speak in accordance with aizuchi, it was found that they were able to (21) "devise the manner of speaking," such as pausing and changing the speed of their speech. The concept is defined as <devising speaking style, such as having pauses, changing speaking speed or etc.>.

> "Oh, yes. There was one time when I thought I might pause for a moment." (2311IF)
>
> "I might have been conscious of trying to speak in more understandable terms." (2214MF)
>
> "When the response was not so good, I would speak more faster and try to explain more." (2413FF)

When a speaker is speaking in a state in which he or she has developed the awareness of speaking in accordance with aizuchi, he or she perceives that <aizuchi is uttered at a time that participants feel appropriate>. The concept defined by this is named (8) "appropriate timing for aizuchi."

> "In the case of the first pattern (the human voice pattern), I had images that aizuchi was uttered at the point when I thought." (2211KF)
>
> "In the human voice, I didn't think the timing of the 'un' was odd." (2814TF)

If the speaker feels that aizuchs are appropriately timed, the conversation will not be interrupted even if the aizuchi overlaps with the speaker's speech, and (30) "aizuchis are felt natural even if aizuchi overlap with the speech." This concept can be defined as <don't mind if a conversation is overlapped by the insertion of aizuchis while speaking>.

> "Oh, yes. I don't mind too much when aizuchi overlaps my speech. It seems to happen in everyday life." (2415SM)
>
> "In the case of human voice, this kind of thing (overlapping aizuchi to the speech) happens a lot in everyday conversation, and I kind of ignored the aizuchi. It didn't interfere with talking." (2413FF)