



Managing turn-taking through beep sounds by Bixby: Applying conversation analysis to human-chatbot interaction*

Ji Young Lah** · Yo-An Lee***
(Sogang University)

Lah, Ji Young and Yo-An Lee. 2023. Managing turn-taking through beep sounds by Bixby: Applying conversation analysis to human-chatbot interaction. *Linguistic Research* 40(Special Edition): 61-87. This paper explores turn-taking between users and Bixby, a Samsung voice assistant. While using a simplified turn-taking system designed to guide users, Bixby does not fully understand and replicate the complex turn-taking system that characterizes human interaction. Utilizing Conversation Analysis (CA), this paper highlights the limitations of Bixby's turn-taking system, which relies on mechanical cues such as ding↓↑ to indicate turn initiation and turn completion. These cues prove inadequate in accommodating the dynamic nature of human interaction, causing confusion in users regarding when and how to talk to the digital assistant. Users often predict turn endings based on their familiarity with human turn-taking, which is often incompatible with Bixby's rigid turn-taking mechanisms. This study incorporates insights from CA and applies sequential analysis to interactions between humans and the voice digital assistant to explicate how Bixby's turn-taking mechanisms become problematic. This information helps explore ways to improve these interactions. (Sogang University)

Keywords conversation analysis, conversational interface, dialogue system, human-computer interaction, repair, turn-taking

1. Introduction

Turn-taking is a fundamental mechanism through which humans manage conversational interaction and perform social actions (Sacks, Schegloff, and Jefferson 1974; Lee 2006).

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A03042244).

** First author

*** Corresponding author

Conversation Analysis (CA) research has empirically demonstrated that people manage turn-taking locally and interactionally (Sidnell and Stivers 2013; Hepburn and Potter 2021). In other words, they determine who should talk and when to change speakership during interactions. To perform these tasks, individuals must be able to recognize when to begin and end their turns in the course of interaction. By tracing the sequential development of turns, participants learn how the current turn is tied to the prior and next turn for its meaning and relevance before acting on them accordingly (Moerman and Sacks 1988).

Recent developments in AI (Artificial Intelligence) research have made it possible to produce digital assistants that are designed to interact with humans. Aware of the complexity of human turn-taking systems and their sequential nature, many human-computer interaction (HCI) researchers have tried to accommodate and emulate the human turn-taking systems (Harris 2005; Moore and Arar 2019; Lee and Goo 2021). However, they have rarely implemented CA's empirical findings in designing conversational interfaces (for CA findings on Amazon Alexa, refer to Porcheron, Fishche, Reeves, and Sharples 2018); this scarcity of application is even more pronounced in the Korean context. Indeed, very few researchers in Korea have applied CA's sequential analysis to conversational interfaces.

Recognizing this gap, the present study is designed to examine turn-taking issues in the interactions between a digital assistant and human users. Turn-allocation is particularly important during these interactions because it determines who talks and when speakership changes. Many big tech companies, including Korean companies, have developed conversational interfaces that are designed to interact with human users without notable success. In light of the current status-quo, the present study examines the interaction between a conversational interface, Samsung's Bixby, and users. By conducting experiments with human users, this study focuses on how turns are allocated and managed and on where problems arise in interactions with Bixby.

The study is designed to achieve several key objectives. Primarily, we strive to discern and evaluate the turn-taking mechanisms embedded in Bixby, identifying unique characteristics that differ from typical human-to-human conversation. Secondly, we describe the process in which human users come to terms with Bixby's unique turn-taking mechanism and how they adapt their interaction accordingly. The findings help us to identify key issues in managing turn-taking matters with a digital assistant and provide insights into possible ways to enhance the user experience.

2. Literature review

Turn-taking refers to the process that determines who talks and when during interactional exchanges. Studies in the field of CA have documented regularities in human interaction (Sacks et al. 1974; Sidnell and Stiver 2013), and some researchers have sought to apply patterns of human interaction to conversational interfaces (Moore and Arar 2019). However, engineers developing dialogue systems have struggled to impose human-like turn-taking rules on these systems. As a result, relative to human interactions, turn-taking in conversational interfaces tends to be rigid and limited (McTear 2020; Ro 2021; Skantze 2021). It is therefore important to understand what current interfaces are capable of when conducting real-time interactions and what aspects of human turn-taking could serve as reference points.

2.1 Turn-taking and repair in human conversation

CA researchers have provided valuable insights into the regular features of turn-taking in natural conversation that determine who talks and how. In CA, a turn refers to the opportunity or right to speak among various possible candidates. CA researchers have uncovered a turn-taking system that explains how people allocate their turns during real-time interactions. This system is designed to explain two features in turn-taking, (1) only one person talks at a time, and (2) speakers take turns (Sacks et al. 1974; Sacks 1992). CA researchers have noted that people are aware of and thus act on the turn constructional unit (TCU) that constitutes a turn.

For the system to work, people must be mindful of the transition relevant place (TRP), a point where the current speaker ends his/her their turn and makes room for other participants to take the floor (Sacks et al. 1974). Additionally, interactants must be aware of the completion of the speaker's turn to know when to initiate their own (Sacks et al. 1974). Moreover, research has shown that to allocate a turn, the preferred sequence is for the current speaker to choose the next speaker, followed by the current speaker choosing the next action, and finally for anyone else in the conversation to self-select and take the floor. CA researchers explain these processes through a lens of people's actions during interactions, which are inherently complex, requiring on-the-spot, spontaneous, and contingent decision-making.

Like turn-taking which involves speakers coordinating their turns during

conversations, repairs are a critical structural feature of interactions. Repairs represent a stable mechanism through which participants in interactions fix problematic turns during ongoing conversations (Schegloff et al. 1977; Kitzinger 2013; Lee and Hellermann 2020). Repair refers to the process of correcting an element that is “repairable or the trouble source” (Schegloff et al. 1977: 363) when issues such as “misarticulations, malapropisms, used of a ‘wrong’ word, unavailability of a word when needed, failure to hear or to be heard, trouble on the part of the recipient in understanding, incorrect understanding by recipients” occur (Schegloff 1987: 210). When trouble occurs in a conversation, a repair can be initiated by either speaker or others while interrupting the progression of the talk.

The nature of repairable content varies from linguistic problems to the pragmatic sense of given turns. Sometimes, repairs occur without the presence of any errors or problems. Repairs are also classified in terms of who initiates and solves them; for example, self-repair refers to the repair by the speaker of the troublesome turn, while other repair denotes a repair initiated and resolved by someone other than the speaker of the trouble-source. These are complex mechanisms whose regularities are well documented and researched in CA. The question is to what extent this turn-taking can be applied to the specific conversational interfaces.

2.2 Dialogue systems modeled on human interaction

Given the crucial role of the turn-taking system in performing conversational interactions between humans, the fact that a multitude of HCI researchers have attempted to apply the systematic traits of human turn-taking to the development of dialogue systems, including the detection of TCU and TRP, is not surprising (Sacks et al. 1974). In terms of turn-taking, it is of utmost importance for dialogue systems to be able to recognize when users finish talking (Jurafsky and Martin 2020). The process of recognizing the end of users’ speech is called endpoint detection (or endpointing) (Jurafsky and Martin 2020: 523). According to Skantze (2021), early spoken dialogue systems detected the endpoint of a users’ turn by calculating the amount of silence at the end of their speech. However, when analyzing 93 telephone calls between humans, Ten Bosch, Oostdijk, and Boves (2005) found that intra-turn pauses are longer than gaps between turns. Therefore, silence turns out to be not a primary indicator in detecting how users finishing their turns (Skantze 2021).

Another HCI approach to detecting the ends of turns is to divide users' speech into inter-pausal units (IPUs). Skantze (2021) defined IPUs as "stretches of audio from one speaker without any silence exceeding a certain threshold, such as 200ms"; thus, a turn can be conceptualized as "a sequence of IPUs from a speaker that is not interrupted by IPUs from another speaker" (Skantze 2021: 3).

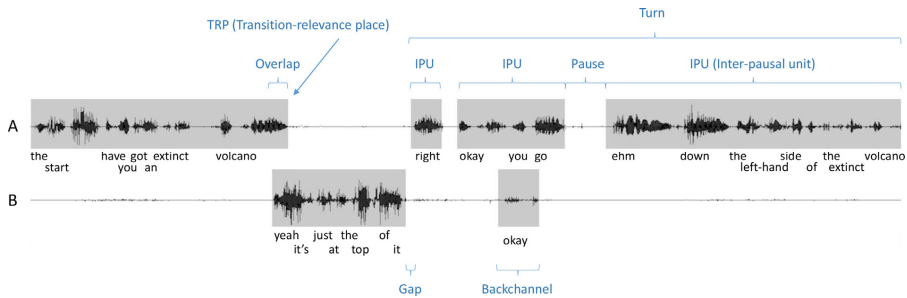


Figure 1. Example of turn-taking from Map Task (Anderson et al. 1991) as cited in Skantze (2021: 3)

For example, as seen from Figure 1, one turn of speaker A consists of two IPUs ("right" and "okay you go") followed by a pause and one IPU "ehm down the left-hand side of the extinct volcano"). In talk-interaction, human utterances can overlap at the TRP because people can project the end of their counterparts' utterances as seen in the end part of the utterance "volcano" from speaker A and the first part of the utterance "yeah it's just a the top of it" from the speaker B. In addition, back-channels such as speaker B's "okay" are considered turns but not barge-in because they serve go-ahead signals similar to Schegloff's continuer (Schegloff 1982). As such, implementing human-like turn-taking system presents an enormous challenge to HCI researchers because machines need to navigate these complex interplays.

Therefore, the systematic structure of human interaction serves as a valuable reference point for understanding the similarities and differences between human-human and human-dialogue system interactions. The challenge for HCI researchers is to determine the appropriate way to implement these complex systems into their research. Some researchers have already taken this approach in the development of dialogue systems (Harris 2005; Moore and Arar 2019). For example, inspired by a formal and qualitative theory of CA, Moore introduced a conversational UX (User Experience) interaction

model for chatbots and virtual agents: a Natural Conversation Framework (NCF) (Moore 2018; Moore and Arar 2019). His team developed the framework by applying multiple key notions of natural conversation through CA such as adjacency pair, repair, and turns. NCF has an independent base adjacency pair sequence such as greeting-greeting, inquiry-answer, and invitation-acceptance/decline. If the base adjacency pair sequence does not finish in two turns, NCF operates expandable sequences that depend on the base sequence, employing repetition, paraphrase, and elicitation. In addition, in cases of communication breakdown, NCF can implement both agent-initiated repairs (e.g., *A: what kinds of food do you like?*) and take user-initiated repairs (e.g., *U: what do you mean?*). Handling these two types of initiation repair is imperative because any utterance can be a source of trouble that leads to miscommunication between parties.

Furthermore, Moore's team designed the NCF model to accommodate generic patterns of conversation so that any platform can employ it: 5 reusable content patterns (1. Inquiry, 2. Open-ended request, 3. Story/instructions, 4. Troubleshooting, 5. Quiz) and 10 conversational management patterns (1. Conversation opening, 2. Offer of help, 3. Capabilities, 4. Repair, 5. Privacy, 6. Sequence closing, 7. Sequence abort, 8. Insult and compliment, 9. Conversation closing, 10. Conversation abort). As such, their NCF demonstrates how conversation features observed via CA such as sequential order and conversation structure can be utilized in the HCI context.

Along with developers of conversational UX design for chatbots and virtual agents, several researchers have examined line-by-line interaction between humans and dialogue systems to specify the contingent and situated interaction (Porcheron et al. 2018; Reeves, Porcheron, and Fischer 2018; Ashktorab, Jain, Laio, and Weisz 2019; Fischer, Reeves, Porcheron, and Sikveland 2019; Albert and Hamann 2021; Lah 2021). Analyses of these interactions have highlighted the need for CA in the HCI context to improve the quality of interactions between humans and dialogue systems. For instance, based on CA, Lah (2021) examined how users interpret a goal-oriented voicebot's non-processing message "There is no schedule that meets the condition." By sequentially analyzing users' varied responses in seven dialogues, Lah found that users respond to the message differently. The author argues that developers and UX designers of voicebots should be aware of the possibility that users' interpretations of voicebot's utterances may not align with developers' and designers' expectations and that these engineers need to embed turn-by-turn progression of conversational interactions in their design.

Albert and Hamann (2021)'s analysis of wake words is another useful resource. They

examined how wake words in smart speaker function in two human-AI speakers. The wake words “Hey, Siri” and “Alexa” summon smart speakers to listen to users’ utterances. Albert and Hamann discovered that, unlike expert users who were found to simplify their interaction cues in previous research, novice users over-design their turns using diverse prosodic and intonational features. This extra interaction work points to asymmetrical expectations between users and UX designers regarding the use of wake words. In the first interaction, when a user summons Alexa, the user produces three different ways of employing the wake work: 1) initial summon with a slight rising intonation; 2) second summon with stress on the second syllable along with flat intonation; and 3) interventional summon with stress and falling intonation when the user wants to interrupt the agent’s action. The researchers suggest that UX designers accommodate a variety of prosodic features sequentially as well as pragmatically within given contexts. Wake words are important in turn-taking as they initiate conversations between virtual agents and users.

As such, more studies should pay particular attention to contingent and situated interactions between humans and dialogue systems such as turn-taking and repairs. This paper is designed to examine how turn-taking mechanisms embedded in a digital assistant, namely Samsung’s Bixby, work in managing turn-allocation with users, what effects they have on users’ responses to the system, and what causes problems. Bixby uses two machine sounds to differentiate its turn taking functions, namely, beginning and closing its turn. Yet, these different functions are not easily recognized or acted upon by users. Tracing this turn taking sequence offers useful pieces of information about how the mechanisms adopted in Bixby influences subsequent exchanges. The relevance and implication of the findings are also discussed.

3. Methods

The data corpus for this study consists of 14 interactions between twelve users and Bixby. Bixby was selected because it performs better than any other conversational interfaces available on the market, including Siri, Google speaker, and Clova. Amazon’s Alexa was not considered because it cannot handle Korean. The twelve participants were recruited for the experiment from a Korean university; three were graduate students and nine were undergraduates. After the study received IRB (Institutional Review Board)

approval, these participants were recruited to participate in the experiment. They were asked to perform a scheduling task using Bixby, which was programmed in a Samsung Android phone. The task involved searching, registering, revising, and deleting a schedule in a linguistics lab booth. For the experiment, the phone numbers of fictitious individuals were registered into the phone along with schedules for two months of activities, such as doctor's appointments, group meetings, and physical exercises.

The task lasted about 10 minutes on average with some participants taking as long as 20 minutes. All participants were given verbal instructions before the experiment regarding the written documents that specified the steps they were to take. Due to the limited functions Bixby can perform, participants were asked to perform three distinctive tasks during the experiment.

Each session was recorded on video to capture the participants' facial, gestural and other non-verbal expressions. Combined, the run-time of the video clips totaled one hour twenty-seven minutes, amounting to 1,222 turn exchanges. These recordings were transcribed automatically using the speech recognition programs, Vrew and CLOVA Note, which were developed by Naver, a large tech company in Korea. In the analysis of the transcribed record, particular focus was placed on how turns were allocated between users and Bixby and what became problematic. The aim of this approach was to elucidate the turn-taking design features that are embedded in the Bixby system and reveal how users respond to and act on them during interactional sequences.

The present paper presented excerpts that display a four-line transcription format. These lines are organized as follows: 1) Korean utterances in the native script, 2) the romanization of the Korean utterance using the Yale system, 3) a morpheme-by-morpheme gloss of the Korean utterance following Leipzig Glossing rules¹, and 4) an equivalent English translation.

4. Results

The dialogue system built in Bixby follows the design principle established for task-oriented dialogue systems that are standard in the field. These systems handle tasks such as scheduling a new event or revising an existing schedule. When interacting with

¹ ACC: Accusative ADV: Adverbial GEN: Genitive INS: Instrumental
 NEG: Negative POL: Polite speech level Q: Question particle TP: Topic marker

Bixby, the user's first task is to summon and activate it before making a request. While this may seem simple, it involves a complex turn-taking task for Bixby. For instance, it must determine when to begin and end its turn by listening to the user's turn. As mentioned earlier, it is important to note that pauses or silences alone are not reliable indicators of turn-transition. For this reason, it is essential for Bixby to have systems in place that ensure regular and consistent turn-exchanges with its users.

This section reports interactional exchanges between Bixby and users in hopes of identifying the turn-taking systems built into Bixby and thereby documenting how users respond to these systems and what is problematic. Although individual users' interactions with Bixby vary to some extent, certain patterns appeared that provide valuable insights into the design principles behind the Bixby system. The goal of the study was to explain how these principles manifest in Bixby's interactions with users. CA and its sequential analysis guided this investigation, which offers a useful complement to engineering solutions to problems with conversational interfaces.

4.1 Bixby's marking for the beginning and ending of its turn

Bixby signals the beginning and ending of its turn using two distinct ding sounds, each performing a different function. The ding sound with a raised tone at the end (ding↑), signifies the closing of Bixby's turn, while the ding↓, with a normal tone, indicates the beginning of the turn. These two different functions have some interactional consequences in engaging users in the interface's turn-taking systems. For example, the closing mark, ding↑, signals that Bixby is ready for the user's response now that its turn has ended. The opening mark, ding↓, serves to secure the floor for Bixby to complete its turn.

To explicate how the system works, let us begin with a summoning sequence. Similar to other digital assistants, Bixby does not initiate contact with users (Reeves 2017). Instead, users summon Bixby by saying "hi Bixby." Bixby responds to this summoning with ding↑, signaling its readiness to receive input from user. The following excerpt illustrates the case in point.

Excerpt 1. [AICA] PRB1 (See Appendix for transcription notation)

1.	U:	하이 빅스비 hai piksupi hi bixby 'Hi Bixby'
2.	B:	딩↑ ding↑
3.	U:	일정 등록해줘 ilceng tunglokhaycwe schedule register 'Register a schedule.'
4.		(1.0)
5.	B:	딩↓ 일정을 언제로 저장할까요↑ 딩↑ ding↓ ilcengul enceylo cecanghalkkayo↑ ding↑ ding↓ schedule-ACC when-INS save-Q-POL ding↑ 'ding When should I save the schedule? ding'
6.	U:	내일 오후 두시 nayil ohwu twu-si tomorrow pm 2-hour 'Tomorrow 2 pm'
7.	B:	딩↓ 제목을 뭐라고 할까요↑ 딩↑ ding↓ ceymokul mwe-lako halkkayo↑ ding↑ ding↓ title-ACC what-as call-Q-POL ding↑ 'ding What title would you like to save it under? ding'
8.		(7.0)
9.	B:	제목은 뭘로 할까요↑ 딩↑ ceymokul mwello halkkayo ↑ ding↑ title-ACC what-INS call-Q-POL ding↑ 'What should the title be? ding'

In line 1, the user begins with a wake word “Hi Bixby” to prompt Bixby to respond. This is followed by Bixby’s ding↑ (line 2), which performs two functions, recognizing the user’s summon while signaling its readiness to receive the user’s request for the service she wants. Therefore, the turn goes back to the user who then makes her request in line 3.

In this type of interaction, human agents would have more options in responding to the summon, for example, “hello” “ok” or “How can I help you?” The first two may initiate greeting sequences, which would produce additional exchanges of turns. However, “how can I help you” would encourage the user to make a request. In this regard, Bixby’s ding↑ is designed to immediately elicit a request.

In line 3, the user requests that Bixby “register a schedule.” Although this is a request,

it requires a follow-up action to gather specific details necessary for fulfilling the request. After one second delay, Bixby responds with a question for the detail in line 5.

This turn by Bixby uses two different ding sounds that have some international consequences. First, it begins with ding↓ signaling that it has secured the turn(s) to follow through with the request. Second, the turn ends with ding↑ to invite the user to answer. Since these two ding sounds mark the turn's beginning and ending clearly, Bixby can take multiple turns between these two ding sounds. Unlike conversational interactions between humans where participants can take their turn at turn transition points, these mechanisms are designed to make speakership changes more predictable. In a way, the ding sounds that bracket Bixby's turns reflect the system's inability to manage turns at every transition point, interactionally. The two ding sounds marks its turn beginning and ending artificially.

Notice here that ding↑ at the end of the turn also signals that Bixby is awaiting the user's turn. In line 6, the user offers a time for the schedule "two o'clock." To this, Bixby organizes a turn resembling the turn in line 5, beginning with ding↓ and ending with ding↑. From a sequential analysis perspective, Bixby's turn in line 7 indicates that it understands the time for the schedule and therefore asks for the next piece of information, the title of the schedule. In doing so, Bixby ends its turn with ding↑, signaling that it is waiting for the user to respond.

Interestingly, the user does not answer for seven seconds in line 8. Let us see how Bixby responds to this silent turn. First, Bixby does not begin with the ding↓ sound that had previously used to mark the beginning of its turns. This may be a device to inform the user that it is repeating the line 7 to prompt the user to answer. Nonetheless, this turn ends with ding↑ to show that it is still waiting for the user's response.

This turn indicates that Bixby uses two ding sounds to manage its turn taking practice, marking the beginning and ending of its turns. The presence of these two ding sounds is an engineering solution to the complex turn-taking tasks that occur at every transition-relevant place after each turn at talk (Sacks et al. 1974). These sounds also make it possible for Bixby to produce multiple turns in between these two ding sounds. The following excerpt illustrates how this arrangement leads Bixby to perform requested functions.

Excerpt 2. [AICA] PRB3

62.	U:	하이 빅스비 hai piksupi hi bixby 'Hi Bixby'
63.	B:	딩↑ ding↑
64.	U:	다음 주 일정 조회 taum cwu ilceng cohoi next week schedule checking 'Checking next week's schedule'
65.		(1.0)
66.	B:	딩↓ 다음 주 일정이 8개 있어요 ding↓ taum cwu ilcengi 8-kay isseyo ding↓ next week schedule 8-unit exist-POL 'ding You have 8 scheduled events next week.'
67.		10월 4일 하루 종일 대체공휴일 10-wel 4-il halwu congil taycheykonghyuil 10-month 4-day one-day all-day replaced-holiday 'October 4 all day replaced holiday'
68.		쉬는 날 개천절 swinun nal kay-chen-cel resting day national-foundation-day 'Day off national foundation day'
69.		10월 6일 오후 6시에 필라테스 10-wel 6-il ohwu 6-si-ey phillatheysu 10-month 6-day pm 6-hour-at pilates 'October 6 at 6 pm, pilates'
70.		일정을 더 알려줄까요↑ 딩↑= ilcengul te allyecwulkkayo↑ ding↑= schedule more tell-Q-POL ding↑ 'Should I read more scheduled items? ding'
71.	U:	=더 알려줘 =te allyecwe more tell 'Tell me more.'
72.		(1.0)
73.	B:	딩↓ 10월 7일 오전 11시에 시장 ding↓ 10-wel 7-il ocn 11-si-ey sicang ding↓ 10-month 7-day am 11-hour-at market 'ding October 7 at 11 am Market'
74.		10월 8일 오후 3시에 팀 미팅 10-wel 8-il ohwu 3-si-ey thim mithing 10-month 8-day pm 3-hour-at team meeting 'October 8 at 3 p.m. Team meeting'

75.		오후 6시에	필라테스	
		ohwu 6-si-ey	phillatheysu	
		pm 6-hour-at	pilates	
		‘ at 6 pm Pilates’		
76.		일정을 더 읽을까요↑	딩↑	
		ilcengul te ilkulkkayo	ding↑	
		schedule more read-Q-POL	ding↑	
		‘Should I read more scheduled items? ding↑’		
77.	U:	아니		
		ani		
		nope		
		‘Nope’		

As in the prior excerpt, the user begins with the wake word, "Hi Bixby" (line 62). In response, Bixby produces a ding↑ to signal its readiness to receive the user's request. In the next turn in line 64, the user makes her request. A silent turn follow, indicating that the user anticipates Bixby's response and refrains from making any subsequent request. That is to say, the user treats ding↑ as a turn constructional unit (TCU). This sequence aligns with the "*wakeword-gap-request*" pattern among Amazon Echo users (Porcheron et al. 2018: 7). After the user's request, Bixby produces ding↓ in line 66 to mark the beginning of its turn, followed by multiple turns in which it lists several scheduled items in lines 66-70. After describing three scheduled items, Bixby asks another question in line 70 to see if the user wants to hear more of the scheduled items. Apparently, more than 3 items are scheduled in the following week. Since this question calls for user action, Bixby's turn closes with ding↑, eliciting the user's response.

As previously noted, the presence of the two ding sounds secures the turns for Bixby to complete its answer; it can produce multiple turns before the ding↑ sound that closes its turn. This has the effect of preventing overlapping turns although it is not clear if the user is aware of these functions. In lines 73-76, a similar phenomenon occurs; Bixby offers to read more, but the user declines the offer in line 77.

In sum, Bixby uses two different mechanical sounds to manage turn allocation and transition. By marking the beginning and ending of its turns, the system displays how it understands the users' requests and completes tasks. While this seems like a reasonable solution to the complex issue of turn-taking, the question of whether and how users recognize this process and act in response is a different matter. In conversational interactions among human participants, turn-taking is locally determined and

interactionally negotiated because people can syntactically and pragmatically process turns to project when they will end. This attention to sequential organization of real-time interaction is artificially controlled through two ding sounds in Bixby's system.

This type of design does not reflect the diverse array of resources users may need when selecting subsequent speakers. For example, Bixby cannot respond to post-positioned wake words such as “register a schedule, Bixby” (Lerner 2003; Reeves 2017). That is to say, this engineering solution to the turn-taking matter can be problematic in two ways. First, users may not be aware of the subtle differences that are enacted in the two ding sounds. Second, these sounds simplify turn-taking matters to such an extent that users may have difficulty noticing and adjusting to them.

4.2 Bixby's absence of turn-closing mark

Regarding these two ding sound patterns, users often struggle to adapt since they are accustomed to the turn-taking systems that are linked and sequenced through meaning-embedded turns. Users begin their turns by projecting the completion point of the prior turn, and this projectability is occasioned not just by syntactic analysis but also by a pragmatic understanding of the content it carries, and scope of said turn.

However, Bixby's ding sounds are just mechanical devices that mark the beginning and ending of Bixby's engagement. For this reason, users need to be aware of whether and when Bixby is engaged or disengaged. In the following excerpt, Bixby disengages from the conversation but the user is not aware of it.

Excerpt 3. [AICA] PTB7

1.	U:	하이 빅스비 hai piksupi hi bixby 'Hi Bixby'
2.	B:	딩↑ ding↑
3.	U:	10월 25일 일정 보여줘 10wel 25il ilceng poyecwe 10-month 25-day schedule show 'Show me the schedule on October 25.'
4.		(1.0)
5.	B:	딩↓ 10월 25일 오전 9시 30분에

6. ding↓ 10-wel 25-il ocn 9-si 30-pwun-ey
 ding↓ 10-month 25-day am 9-hour 30-minute-at
 ‘ding On October 25 at 9:30 a.m.’
 팀 주간 회의 일정이 있네요
 thim cwukan hoyuy ilcengi issneyyo
 team weekly meeting schedule-NM exist-POL
 ‘you have weekly team meeting scheduled.’
7. (1.0)
8. U: ▶ 음 아- 이번 주 토요일 일정 보여줘
 um a- ipen cwu thoyoil ilceng poyecwe
 um a- this week saturday schedule show
 ‘um ah Show me this Saturday’s schedule.’
9. (5.0)
10. U: 하이 빅스비
 hai piksupi
 hi bixby
 ‘Hi Bixby’
11. B: 딩↑
 ding↑
12. U: 이번 주: 토요일 일정 보여줘
 ipen cwu: thoyoil ilceng poyecwe
 this week saturday schedule show
 ‘Show me this Saturday’s schedule’
13. (1.0)
14. B: 딩↓ 이번 주 토요일은 일정이 없네요
 ding↓ ipen cwu thoyoilun ilcengi epsneyyo
 ding↓ this week saturday schedule exist-NEG-POL
 ‘ding You have no schedule this Saturday.’
15. U: ▶ 이번 주 토요일에 (2.0) 하숙 예약해줘
 ipen cwu thoyoil-ey (2.0) haswuk yeyyakhaycwe
 this week saturday-at lodging reserve
 ‘This Saturday (2.0) make a lodging reservation.’
16. (4.0)
17. U: 하이 빅스비
 hai piksupi
 hi bixby
 ‘Hi Bixby’
18. (2.0)
19. U: 하이 빅스비
 hai piksupi
 hi bixby
 ‘Hi Bixby’
20. (2.0)
21. U: 하이 빅스비
 hai piksupi
 hi bixby

22.	B:	‘Hi Bixby’ 딩↑ ding↑
-----	----	---------------------------

In line 1 to 6, the interaction proceeds according to the canonical order described in the prior section. This involves the user’s wake word (line 2), which is followed by the user’s request. In line 5, Bixby makes ding↓ sound, indicating the beginning of its turn, and then it offers the requested schedule information till line 6.

Notably, Bixby’s turn in line 6 does not end with the customary turn-closing mark, ding↑. The absence of the ding sound signifies that Bixby considers the provided information sufficient to fulfill the request, thus not inviting the user to continue speaking. However, the user does not seem to notice the absence of the ding sound and assumes that the conversation is still ongoing. Consequently, she makes a follow-up request in line 8. However, since Bixby has already completed its task, it does not respond at this point (line 9). In line 10, the user’s subsequent turn reveals her realization of the situation and she uses a wake word to re-engage Bixby.

After reengaging Bixby with a wake word, the user repairs her request by repeating the message in line 12. After a short delay, Bixby says “You have no schedule this Saturday” in line 14. Notice again that Bixby produces this turn without adding a ding↑ sound, indicating that it regards this response as a final answer. Unfortunately, the user is not aware of this absence and, therefore, makes another request in line 15. Only after a four second delay in line 16 does the user recognize that Bixby has disengaged again. This prompts her to repeat a wake word in line 17. In the subsequent turn in line 18, the user produces a wake word multiple times until Bixby answers it with a ding↑. The fact that this phenomenon takes place twice is indicative of how difficult it is for users to adapt to this artificial turn-taking system. The following excerpt illustrates another case in which the user is not aware of the absence of the ding↑ that invites the user’s turn. The user’s turn consists of a few repairs that overlap with the beginning of Bixby’s turn.

Excerpt 4. [PICA] PTB8

13.	U:	°어-° 10월- 헤- 하이 빅스비 °e-° 10wel- hey- hai piksupi uh- 10-month he- hi bixby ‘Uh- October- He-Hi Bixby’
14.	B:	딩↑

15.	U:▶	ding↑ 10월:(.) 13일일- 어:- [병- [아이]코.hhh 10-wel:(.) 13-il-il- e:- [pyeng- [aikho.hhh 10-month 13-day-day- e- hos- oh-my .hhh 'October 13th th- Uh- Hos- Oh my'
16.	B:	[딩↓ [2022년 10월 13일은 [ding↓ [2022-nyen 10-wel 13-il-un ding↓ 2022-year 10-month 13-day-TP 'ding On October 13th 2022'
17.		일정이 없네요= lceng-i epsneyyo= schedule-NM exist-NEG-POL 'there is no schedule'
18.	U:	=일정 추가해줘 =ilceng chwukahaycwe schedule add 'Add a schedule.'
19.		(2.0)
20.	U:	하이 빅스비 hai piksupi hi bixby 'Hi Bixby'
21.	B:	딩↑= ding↑=
22.	U▶	=일정 추가해줘 =ilceng chwukahaycwe schedule add 'Add a schedule.'
23.		(1.0)
24.	B:	딩↓ 일정을 언제로 저장할까요↑ ding↑ ding↓ ilcengul enceylo cecanghalkkayo↑ ding↑ ding↓ schedule-ACC when-INS save-Q-POL ding↑ 'ding When should I save the schedule? ding'

In line 13, the user initiates the interaction with the wake word. When Bixby produces ding↑, the user begins to verbalize a specific date but cuts her turn short and initiates repair that ends with “oh my” in line 15. Bixby treats the user's utterance "October 13th" as a completed turn and starts producing an answer signaling the turn-beginning with a ding↓ in line 16 followed by the information that there is nothing scheduled on that day. This turn ended without a ding↑ to signal the completion of questioning sequence.

As in the prior excerpt, however, the user does not recognize the absence of the ding ↑, and she therefore proceeds with another request as if Bixby is still in conversation in line 18. This is followed by a two-second silence in line 19, which prompts the user

to produce a wake word in line 20, indicating that she realizes Bixby has disengaged from the sequence and restarts the conversation.

As previously noted, the positioning of the ding sounds is critical for carrying out conversation with Bixby. Unfortunately, the user is not particularly attuned to these subtle differences embedded in ding sounds. The turn ending ding↑, in particular, poses a challenge because it seemingly indicates a closing and invites the user's turn. However, these ding sounds serve as mechanical devices for turn-organization and do not convey or respond to the semantic meaning of the given turn. Apparently, Bixby is not designed to engage in continued conversation once its purpose of interaction is fulfilled.

4.3 User's failure to differentiate two ding sounds

To this point, the analysis has noted that Bixby produces two different ding sounds and each one performs different functions. Note, however, that the users are not attuned to these differences during their interaction. At times, they may not anticipate these cues due to their subtle difference in sound, as they do not seem to carry explicit pragmatic sense. This failure to understand the different functions leads to some trouble in carrying out the conversations. The following excerpt is a case in point. Here, Bixby asks the user to provide the date of the schedule the user wants to register. Although this turn is followed by a ding↑, the user responds with another wake word.

Excerpt 5. [PICA] PTB7

56.	U:	하이 빅스비 hai piksupi hi bixby 'Hi Bixby'
57.	B:	딩↑ ding↑
58.	U:	병원↑ 일정:(.) 잡아줘 pyengwen:↑ ilceng:(.) capacwe hospital schedule make 'Schedule a hospital visit.'
59.		(1.0)
60.	B:	딩↓ 일정을 언제로 할까요↑ [딩↑ ding↓ ilcengul enceylo halkkayo↑ [ding↑ ding↓ schedule-ACC when-INS do-Q-POL ding↑ 'ding When should I schedule the hospital visit? ding'

61.	U: ▶					[10월- (1.0) [10-wel- (1.0) 10-month 'October'
62.		(1.0)				
63.	U:	십-				
		10-				
		October-				
64.	U: ▶	하이 빅스비↑				
		hai piksupi				
		hi bixby				
		'Hi Bixby'				
65.		(1.0)				
66.	U:	[10월:				
		[10wel:				
		10-month				
		'October'				
67.	B:	[딩↓ 앞으로의 일정을 알려줄게요				
		[ding↓ aphulouy ilcengul allyecwulkeyyo				
		ding↓ front-ADV-GEN schedule-ACC tell-POL				
		'ding I will tell you your upcoming schedule.'				
68.		10월 25일 오전 9시 30분에 팀 주간 회의				
		10-wel 25-il ocen 9-si 30-pwun-ey thim cwukan hoyuy				
		10-month 25-day am 9-hour 30-minutes-at team weekly meeting				
		'October 25th at 9:30 am team weekly meeting'				
69.		10월 27일 오후 7시에 영어회화 수업				
		10-wel 27-il ohwu 7-si-ey yenge-hoyhwa swuep				
		10-month 27-day pm 7-hour-at english-conversation class				
		'October 27th at 7 pm English conversation class'				
70.		10월 28일 오후 2시에 팀 미팅				
		10-wel 28-il ohwu 2-si-ey thim mithing				
		10-month 28-day pm 2-hour-at team meeting				
		'October 28th at 2 pm team meeting'				
71.		일정을 더 읽을까요↑				
		ilcengul te ilkulkkayo↑				
		schedule-ACC more read-Q-POL				
		'Should I read more of the schedules?'				

After the user's wake word (line 56), Bixby responds with a ding↑ inviting the user's request in the subsequent turn. Once the user makes her request in line 58, Bixby asks for a detail in line 60 using two ding sounds, one initiating the turn and the other closing.

Note here that the ding↑ at the end of Bixby's turn indicates that the user's turn should follow. However, the user begins her turn in line 61 that overlaps with the turn

closing ding↑ from Bixby. This indicates that the user sees when Bixby’s turn is finished only for its semantic sense without waiting for Bixby’s ding↑ sound. This indicates that the user is not aware of the fact that Bixby uses the two different ding sounds in managing its turn-taking. As a result, the user’s comment on “October” in line 61 is not responded by Bixby.

In the next turn in line 63, the user repeats her “October” but cuts it off in line 63 and then, produces a wakeword to summon Bixby in line 64. Unfortunately, Bixby does not respond to this wakeword either in line 65 which is followed by the user’s repetition of the word, October. This indicates that the user is not attuned the two different ding sounds to manage turn-taking. The user’s intent is not clear because her turn is not complete. Nonetheless, Bixby announces its intention in line 67 to offer information about the schedule beginning with ding↓ sound. This is followed by multiple turns to show three scheduled events until it closed its turn with another request in line 71. This shows that Bixby treats 66 as a request for informing schedule in October. Again, this excerpt shows that the user is not aware of the turn-taking functions two ding sounds display.

A similar phenomenon occurs in the following segment. Here, the trouble is more readily apparent than in the excerpt above (Jefferson 1987). This excerpt begins with Bixby’s request for confirmation in line 69.

Excerpt 6. [PICA] PTB7

69.	B:	일정을 더 읽을까요↑ ilcengul te ilkulkkayo↑ schedule-ACC more read-Q-POL ding↑ ‘Should I read more of the schedule?’ ding↑	[딩↑ [ding↑ ding↑
70.	U:		[응 [ung ‘Yup’
71.		(2.0)	
72.	U:	하이 빅스비 ↑(.) hai piksupi ↑(.) hi bixby ‘Hi Bixby’	
73.	B:	[딩↓ 앞으로의 일정을 알려줄게요 [ding↓ aphulouy ilcengul allyecwulkeyyo ding↓ front-ADV-GEN schedule-ACC tell-POL ‘ding I will tell you your upcoming schedule.’	
74.	U:	[응 [ung	

75.	U:	‘Yup’ 훗 hwus ‘hoot’
76.	B:	10월 25일 오전 9시 30분에 팀 주간 회의 10-wel 25-il ocn 9-si 30-pwun-ey thim cwukan hoyuy 10-month 25-day am 9-hour 30-minutes-at team weekly meeting ‘October 25th at 9:30 am Team weekly meeting’
77.		10월 27일 오후 7시에 영어회화 수업 10-wel 27-il ohwu 7-si-ey yenge-hoyhwa swuep 10-month 27-day pm 7-hour-at english-conversation class ‘October 27th at 7 pm English conversation class’
78.		10월 28일 오후 2시에 팀 미팅 10-wel 28-il ohwu 2-si-ey thim mithing 10-month 28-day pm 2-hour-at team meeting ‘October 28th at 2 pm Team meeting’
79.		일정을 더 읽을까요↑ ilcengul te ilkulkkayo↑ schedule-ACC more read-Q-POL ‘Should I read more of the schedules?’

The excerpt begins with Bixby offering to read more of the schedule in line 69. Since this turn is a confirmation request, it is followed by a ding↑, inviting the user to confirm. Note, however, that this turn overlaps with the user’s positive response in line 70.

Notable here is the presence of a two second delay in line 71. This is longer than what Bixby normally takes to answer a question. This silent turn is problematic to the user since it is not clear whether Bixby recognized the user’s response in line 70. For this reason, the user treats the silent turn as a problem and therefore, begins with a wake word to start a new exchange. Note, however, Bixby announces its intent to list schedules in line 73. It is likely that the user’s overlap in line 69-70 was not recognized by Bixby. It simply responds to the wake word in line 72 and then, continues the task of telling the scheduled events. If the user had understood the functions of the ding sounds, she would have not needed to re-engage the conversation in line 72, while this conversational interaction proceeds without a serious problem.

In sum, these two excerpts indicate that although the two different ding sounds perform different functions, the user is not entirely aware of the functions they display. While these ding sounds are useful devices for programmers to enable Bixby to manage its turn-taking, their mechanical functions are too subtle to alert users to act accordingly. In other words, while the two sounds are usefully different for engineers, they are not

clear enough to be useful for users. Users may be too practiced in human turn-taking to limit their talk to the simplified system Bixby offers.

5. Conclusion

The present study set out to examine the impact of the digital assistant, Bixby's turn-taking system on user responses and actions in conversational exchanges. Turn-taking is a primordial system that governs turn allocation during interaction. Contrary to the complex human turn-taking documented in CA research (Sacks et al. 1974), we found out Bixby utilizes a simpler turn-taking system with two distinct auditory cues (contouring ding sounds) to demarcate who should talk and when. To some extent, this system is quite useful for engineers because it makes clear when Bixby initiating and closing turns. It is also designed to guide and even, in some cases, oblige users to act in concert with it to accomplish given tasks.

Nevertheless, as the findings reported above indicate, human turn-taking is much more complex than Bixby's system can handle. While Bixby can clearly mark the beginning and ending of its turns with two auditory cues to guide the conversation, our findings indicated a notable challenge: user often found it difficult to notice, or differentiate the two different ding sounds, which leads to re-summoning or repair. Our result highlights that human users are accustomed to a turn-taking system that is locally determined and interactionally negotiated. They are always sensitive to and thus display their understanding of who the speaker is and when the current turn is going to end. Bixby tries to eliminate this complexity by using several mechanical sounds. These sounds can, to some extent, replace basic structural features such as beginning or summoning procedures. Yet, there are other elements that are hard to eliminate, particularly those that are determined during interactions. For example, users respond to turn content and act based on their estimation of when the current turn ends. Sometimes their turns overlap because they are oriented to pragmatic sense of the turn in determining when to respond to the machine.

The challenges we have identified highlight critical areas that programmers should address to enhance digital assistant, and ultimately, the user experience. These include: integrating a more natural turn-taking system informed by CA's findings (see Skantze 2021 for a review of turn-taking in HCI), improving Bixby's capacity to identify and

interpret instances of overlapping speech and silence, and facilitating user-initiated interruptions or “barge-ins” as was already suggested more than 20 years ago (Hirst 1991).

Starting as template-based systems, conversational assistants have improved significantly due to Natural Language Processing (NLP) and Natural Language Generator employing machine learning (ML) techniques such as bidirectional encoders, Reinforcement Learning, and attention-based bidirectional LSTM network (Luo, Lau, Li, and Si 2021). Despite these advancements, the practical applications of conversational assistants still have certain limitations. For instance, the most common tasks assigned to Amazon Alexa consist of weather updates, music playback and control of other devices (Lopatovska, Rink, Knight, Raines, Cosenza, Williams, Sorsche, Hirsch, Li, and Martinez 2019). These tasks typically do not engage users in more conversational exchanges.

Recent developments in generative AI systems have gained increasing attention thanks to the availability of big data, and these systems may solve some problems affecting current programs. Still, CA’s sequential analysis provides a valuable complement to these developments. It enables a detailed examination of the ways in which conversational interactions are deployed, uncovering issues that are often overlooked or taken for granted in HCI. Sequential analysis documents how users comprehend, interpret, and execute social interactions, and contingently select their courses of action by tracing sequential flow of interactions. Therefore, it is crucial for programmers and designers to thoroughly examine turn-taking details within the local and interactional context even within the parameters they have established in the evolving sequence of talk. CA’s sequential analysis is an effective approach to explore the temporal organization of conversational interactions, allowing engineers to identify problems and find solutions that matter to both the participants-machine and users alike.

References

- Albert, Saul and Magnus Hamann. 2021. Putting wake words to bed: We speak wake words with systematically varied prosody, but CUIs don't listen. *Proceedings of the 3rd International Conference on Conversational User Interfaces (CUI '21)*, 1-5. Spain. Association for Computing Machinery.
- Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon

- Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4): 351-366.
- Ashktorab, Zahra, Mohit Jain, Q. Vera Liao, and Justin D. Weisz. 2019. Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-12. Glasgow: Association for Computing Machinery.
- Fischer, Joel E., Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1-8.
- Harris, Allen. 2005. *Voice interaction design: Crafting the new conversational speech systems*. San Francisco, CA: Morgan Kaufmann Publishers.
- Hepburn, Alexa and Jonathan Potter. 2021. *Essentials of conversation analysis*. Washington, DC: American Psychological Association.
- Hirst, Graeme. 1991. Does conversation analysis have a role in computational linguistics?. *Computational Linguistics* 17(2): 211-227.
- Jefferson, Gail. 1987. On exposed and embedded correction in conversation. In Graham Button and John Lee (eds.), *Talk and social organisation*, 86-100. Clevedon: Multilingual Matters.
- Jurafsky, Daniel and James H. Martin. 2020. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Manuscript (3rd draft edition). Stanford University.
- Kitzinger, Celia. 2013. Repair. In Jack Sidnell and Tanya Stivers (eds.), *The handbook of conversation analysis*, 229-256. Malden, MA: Wiley-Blackwell.
- Lah, Jiyoung. 2021. Users' responses to a non-processing message by a dialogue system: HCI. *Korean Journal of Applied Linguistics* 37(Special): 43-74.
- Lee, Yo-An. 2006. Towards respecification of communicative copetence: Condition of L2 instruction or its objective? *Applied Linguistics* 27(3): 349-376.
- Lee, Yo-An and John Hellermann. 2020. Managing language issues in second language storytelling. *System* 93: 1-13.
- Lee, Yo-An and Myoung-Wan Koo. 2021. Applying human interaction to human-computer interface design. *Korean Journal of Applied Linguistics* 37(Special): 7-41.
- Lerner, Gene H. 2003. Selecting next speaker: The context sensitive operation of a context – free organization. *Language in Society* 32: 177-201.
- Lopatovska, Irene, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* 51(4): 984-997.
- Luo, Bei, Raymond YK Lau, Chunping Li, and Yain-Whar Si. 2022. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and*

- Knowledge Discovery* 12(1): e1434.
- McTear, Michael. 2020. *Conversational AI: Dialogue systems, conversational agents, and chatbots. (Synthesis lectures on human language technologies)*. Toronto: Morgan and Claypool Publishers.
- Moerman, Michael and Harvey Sacks. 1988. On ‘understanding’ in the analysis of natural conversation. In Michael Moerman (ed.), *Talking culture: Ethnography and conversation analysis*, 180-186. Philadelphia, PA: University of Pennsylvania Press.
- Moore, Robert J. 2018. A natural conversation framework for conversational UX design. In Robert Moore, Margaret Szymanski, Raphael Arar, and Guang-Jie Ren (eds.), *Studies in conversational UX Design*, 181-204. Cham: Springer.
- Moore, Robert J. and Raphael Arar. 2019. *Conversational UX design: A practitioner’s guide to the natural conversation framework*. New York, NY: Association for Computing Machinery.
- Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1-12.
- Reeves, Stuart. 2017. Some conversational challenges of talking with machines. Presented at *Talking with Conversational Agents in Collaborative Action, Workshop at the 20th ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW '17)*. Portland, Oregon. February 25 - March 1.
- Reeves, Stuart, Martin Porcheron, and Joel Fischer. 2018. This is not what we wanted' designing for conversation with voice interfaces. *Interactions* 26(1): 46-51.
- Ro, Eunseok. 2021. Solving problems in the processing of users’ request in human-machine interactions. *Korean Journal of Applied Linguistics* 37(Special): 75-116.
- Sacks, Harvey. 1992. *Lectures on conversation*. Malden, MA: Blackwell.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696-735.
- Schegloff, Emanuel A. 1982. Discourse as an international achievement: Some uses of ‘uh huh’ and other things that come between sentences. *Analyzing discourse: Text and Talk* 71: 71-93.
- Schegloff, Emanuel A. 1987. Between macro and micro: Contexts and other connections. In Jeffrey C. Alexander, Bernhard Giesen, Richard Munch, and Neil J. Smelser. (eds.), *The micro - macro link*, 207-234. Berkeley, CA: University of California Press.
- Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53(2): 361-382.
- Sidnell, Jack and Tanya Stivers. 2013. *The handbook of conversation analysis*. Malden, MA: Blackwell Publishing.
- Skantze, Gabriel. 2021. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech and Language* 67: 1-26.
- Ten Bosch, Louis, Nelleke Oostdijk, and Lou Boves. 2005. On temporal aspects of turn taking in conversational dialogues. *Speech Communication* 47(1-2): 80-86.

Appendix

Transcription Notations

[The point of overlapping utterances.
(2.0)	Timed silence within or between adjacent utterances.
()	An uncertain hearing of what the speaker said.
(())	Scenic description and accounts
(.)	A short untimed pause.
=	Latching that indicates no interval between adjacent utterances.
-	A halting, abrupt cutoff
.	Stopping/falling intonation contour, not necessarily an assertion.
,	A slightly rising intonation, not necessarily a clause boundary.
↑	Rising intonation, not necessarily a question.
↓	Falling intonation.
◦	Quieter than surrounding talk
:	A prolonged stretch on Stressed syllable, either by increased amplitude or higher pitch or both
∧^	Marked change in pitch: upward or downward.
h	Aspiration, breathing out
.h	Inhalations
ha(h)ve	Smiley voice
<>	Utterance is delivered at slower pace than surrounding talk
><	Utterance is delivered at quicker pace than surrounding talk.
▶	A line of a particular interest in the discussion

(cf. Atkinson and Heritage 1984)

Ji Young Lah

Ph.D. Candidate

Department of English

Sogang University

J827, 35 Baekbeom-ro, Mapo-gu,

Seoul 04107, Korea

E-mail: laney0410@sogang.ac.kr

Yo-An Lee

Professor

Department of English

Sogang University

J821, 35 Baekbeom-ro, Mapo-gu,

Seoul 04107, Korea

E-mail: yoanlee@sogang.ac.kr

Received: 2023. 06. 14.

Revised: 2023. 07. 18.

Accepted: 2023. 07. 21.