Proceedings of the workshop on People in Vision, Language, and the Mind @LREC2022 , pages 17–25
Marseille, 20 June 2022

Cognitive States and Types of Nods

Taiga Mori*†, Kristiina Jokinen†, Yasuharu Den‡
*Graduate School of Science and Engineering, Chiba University
‡Graduate School of Humanities, Chiba University
1-33 Yayoicho, Inage-ku, Chiba 263-8522, Japan

†AI Research Center, AIST Tokyo Waterfront
2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

Abstract

In this paper we will study how different types of nods are related to the cognitive states of the listener. The distinction is made between nods with movement starting upwards (up-nods) and nods with movement starting downwards (down-nods) as well as between single or repetitive nods. The data is from Japanese multiparty conversations, and the results accord with the previous findings indicating that up-nods are related to the change in the listener's cognitive state after hearing the partner's contribution, while down-nods convey the meaning that the listener's cognitive state is not changed.

Keywords: head nod, multimodal interaction, human-agent interaction

1. Introduction

When a speaker is speaking, the interlocuters do not only listen to the presentation, but simultaneously give feedback to the speaker with the help of short utterances, head nods, or sometimes both. Such feedback giving behaviors convey various meanings such as acknowledgement, understanding, agreement and empathy, and they are necessary for smooth interaction. In order to support natural interaction with the user, conversational agents should also exhibit similar behavior with appropriate features and appropriate timing, as well as the capability to recognize the user's behavior to confirm their interest in the ongoing topic or that they have understood what the agent said (cf. Jokinen, 2018).

Many studies have focused on nodding which is generally considered one of the most important and natural feedback signals in human-human conversations. Besides the form and function of nodding in giving and eliciting feedback (see e.g., Navarretta et al., 2012), also the timing when the listener produces a nod is important; for instance, Watanabe and Yuuki (1989) proposed a model to predict listener's nod timing from speech input of preceding utterance, and Yatsuka et al. (1997; 1998) and Watanabe et al. (2004) implemented the model in real and virtual robots.

However, in human-agent interaction studies nods are often defined as vertical head movements in general, and the meaning differences that are conveyed in the forms of the nods are ignored. For instance, it is shown that nods can be classified into two types based on the direction of the initial movement, up-nods and down-nods. Boholm & Allwood (2010) noticed that up-nods and down-nods are likely to co-occur with different vocal feedback expressions in Swedish, while Navaretta et al. (2012) compared the use of up-nods and down-nods in Danish, Swedish and Finnish and reported several differences in the frequency of nods in these languages. It is interesting that although Nordic countries are culturally similar, the study found that e.g., Danes use down-nods much more frequently than Swedes and Finns, whereas Swedes use up-nods significantly more often than Danes and slightly more often than Finns. Moreover, it was observed that up-nods are used as acknowledgement for new information in Swedish. In a closer study of nods in the Finnish language, Toivio & Jokinen (2012) reported that up-nods and down-nods have different functions in the construction of the shared understanding among the speakers, and that up-nods seem to mark the preceding information as surprise or unexpected to the listener, while down-nods confirm the information as expected, and signal the partner to continue their presentation.

Although the distinction between up-nods and down-nods seems to be functionally appropriate in a wide variety of culturally and linguistically different languages, we wish to confirm that the distinction also works in different languages. Thus, in this paper, we investigate how up-nods and down-nods are used as feedback in Japanese conversations and aim to verify if a similar distinction exists in Japanese as in the Nordic languages. Finally, we sketch a model of nod production for conversational agents.

The organization of this paper is as follows. In section 2, we describe our data and method to identify up-nods and down-nods. In section 3, we conduct quantitative analysis and calculate correlations between feedback expressions and the two types of nods. In section 4, we conduct qualitative analysis and precisely examine when and how up-nods are used in conversations. In section 5, we discuss the results of quantitative and qualitative analysis, and based on that, we propose a model of nod production for conversational agents in section 6. Finally, we describe our future work in section 7.

2. Data and Method

2.1 Data

The data is Chiba three-party conversation corpus (Den & Enomoto, 2007). This corpus contains a collection of 3 party conversations by friends of graduate and undergraduate students. Figure 1 shows the settings of the conversation. Participants sat at regular intervals and were recorded by cameras installed in front of each participant and an outside position where everyone can be seen. In addition, each participant's audio was recorded by the headset. In this corpus, the topic of the conversation is randomly determined by a dice such as "angry story" and "stinking story", and the participants freely talked about that. We used all 12 conversations in the corpus for this

study, thus, the total number of participants is 36. The duration of each conversation is 9 and a half minutes, and the total duration of the conversations is 114 minutes. This corpus also contains annotations of morphological and prosodic information, response tokens (Den et al., 2011), head gestures (nod, shake and the others) and so on. We used these existing annotations for the following analysis.

Figure 1: The settings of the conversations

2.2 Identification of Nod Type

According to head gesture annotation, the data contains a total of 2336 nods produced either by the speaker and the listener. We classified them into up-nods and down-nods. As to the definition of the nod type, we followed previous studies and identified them based on the direction of the initial movement. In this study, we used automatic face recognition and automatically classified all nods into the two types. The classification procedure is as follows. First, we conducted face recognition for all frames of videos recorded from the front of participants and estimated the face position in the image. Here, we used OpenCV detector (OpenCV, 2020) learned on frontal face. Second, we smoothed time-series data of vertical face position with moving average filter and normalized it by standardization. The window size of moving average filter is empirically determined to be 7. Finally, we classified all nods into up-nods or down-nods based on whether or not the face is rising in the first 10 frames immediately after the start of the nod. Figure 2 shows examples of trajectories of up-nods and down-nods.

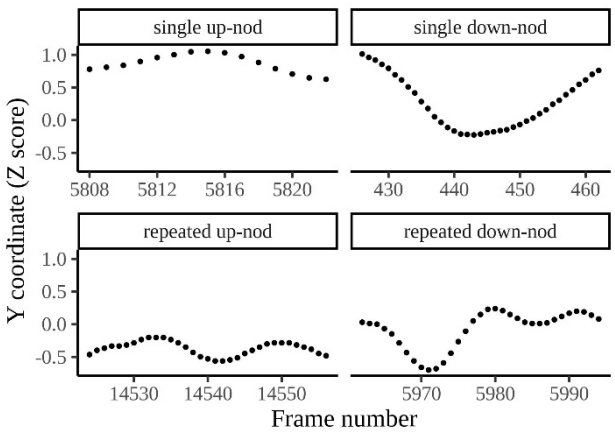

Figure 2: Trajectories of up-nods and down-nods

3. Quantitative Analysis

3.1 Purpose

Previous studies focusing on feedback behaviors in Nordic countries analyzed correlations between the two types of nods and feedback expressions and reported that up-nods are used as acknowledgement for new information in Swedish and Finnish. We also analyzed the correlations between the two types of nods and feedback expressions in Japanese. Our hypothesis is that if up-nods are used as acknowledgement for new information, they should be likely to co-occur with feedback expressions considered as "change of state tokens" (Heritage, 1984). According to Heritage (1984), change of state tokens suggest "its producer has undergone some kind of change in his or her locally current state of knowledge, information, orientation or awareness" (p. 299). Considering Japanese change of state tokens, Tanaka (2010) noted that Japanese particles *aa*, *ee*, *haa*, *huun*, *hee* and *hoo* have similar functions with English change of state token *oh*. Endo (2018) distinguished *a* and *aa* as change of state tokens and noted that *aa* is used when its producer has prior knowledge of the preceding information, while *a* is used when he or she has no knowledge. If the listener acknowledges preceding information as new, he or she would use these tokens in concurrence with up-nods.

3.2 Method

In this analysis, we analyze the correlations between the two types of nods and feedback expressions. First, we defined and extracted feedback expressions from the data. However, this is not so easy because some expressions such as "yes" are used as both an answer as well as feedback. In our data, response expressions are annotated with form tags and position tags defined by Den et al. (2011), and they are useful to determine whether the expression is an answer or feedback. With these tags, we excluded expressions occurred in the first or second pair part of an adjacency pair and unclassified positions such as after a long silence because they are not feedback to other participant's utterance. We also restricted our targets to responsive interjections, expressive interjections and lexical reactive expressions. Second, we extracted the two types of nods overlapping with these feedback expressions. We excluded data if the gap between starting times of the feedback expression and nod exceeds 200 msec because they are likely to be responses to different objects that are temporally adjacent in the speaker's utterance. Finally, we calculated each participant's ratios of the two types of nods with respect to co-occurring feedback expressions. Table 1 shows all feedback expressions co-occurred with up-nods and down-nods in the data. Note that, when consecutive expressions belong to same form, we treated them as one expression (e.g., "maa un" = "maaun").

Expression	Explanation
a (oh)	Expressive interjection to express a surprise or notice.
aa (ah)	Expressive interjection to express a surprise or notice.
aan (ah)	One of the derived forms of *aa*. Perhaps fusion of *aa* and *un*.
ee (really)	Expressive interjection to express a surprise or notice. It expresses stronger

	unexpectedness than *a* and *aa*, and therefore sometimes implies negative meanings such as aversion or disappointment.
haa (oh)	Expressive interjection to express an admiration.
hai (yes)	Responsive interjection to express an acceptance of other's utterance. It is used similarly to *un* but is more formal than *un*.
hee (oh)	Expressive interjection to express a surprise, notice or admiration.
hoo (oh)	Expressive interjection to express a surprise, notice or admiration.
huun (uh-huh)	Expressive interjection to express a surprise, notice or admiration. It is sometimes perceived as a lukewarm reaction.
maane (yeah)	Lexical reactive expression to express an understanding or agreement to other's opinion or assertion. Fusion of *maa* and *ne*. *maa* is also used as filler and therefore sometimes implies hesitation.
maaun (yeah)	Lexical reactive expression to express an understanding or agreement to other's opinion or assertion. Fusion of *maa* and *un*. *maa* is also used as filler and therefore sometimes implies hesitation.
n (yeah)	Responsive interjection to express an acceptance of other's utterance. Abbreviation of *un*.
na (yeah)	Lexical reactive expression to express an agreement to other's opinion or assertion.
naruhodone (I see)	Lexical reactive expression to express an understanding to other's opinion or assertion. Fusion of *naruhodo* and *ne*.
ne (yeah)	Lexical reactive expression to express an agreement to other's opinion or assertion.
oo (oh)	Expressive interjection to express a surprise, notice or admiration. It is used when the provided information is socially or personally desirable.
soo (yeah)	Lexical reactive expression to express an agreement to other's opinion or assertion.
sooka (I see)	Lexical reactive expression to express an understanding to other's opinion or assertion. Fusion of *soo* and final particle *ka*.
soone (yeah)	Lexical reactive expression to express an agreement to other's opinion or assertion. Fusion of *soo* and *ne*.
un (yeah)	Responsive interjection to express an acceptance of other's utterance.
uun ↑ (oh)	Expressive interjection to express a surprise, notice or admiration.
uun ↓ (yeah)	Responsive interjection to express an acceptance of other's utterance. Perhaps one of the derived forms of *un*.

Table 1: All feedback expressions co-occurred with up-nods and down-nods

3.3 Results and Discussion

Figure 3 shows the ratios of up-nods and down-nods with respect to co-occurring feedback expressions. Error bars show standard errors, and "×2" and "×3+" next to the expressions mean "repeated twice" and "repeated more than three times" respectively. First, the figure shows, as we predicted, up-nods co-occurred with change of state tokens *a*, *aa*, *ee*, *haa*×2, *hee* and *hoo* more frequently than down-nods; there is, however, no big difference between them in *aa*×3+, *haa*, *haa*×3+ and *huun*; and the tendency is inversed in only *aa*×2. These results are consistent with our hypothesis. Moreover, comparing *a* and *aa*, *a* co-occurred with up-nods more frequently than *aa*, which is consistent with the difference between *a* and *aa* observed by Endo (2018). Since *aa* is used when the listener has prior knowledge of the preceding informing, it is more likely to co-occur with down-nods than *a*. On the other hand, *huun* and single and repeated *haa* particles do not have clear tendency. As for the character of *huun*, Tanaka (2010) described that it is displaying involvement in ongoing talk without topical engagement. In other words, *huun* is used when the listener acknowledges the information as new but do not have interest in that, and this seems to be applied to *haa* as well. This fact suggests that *huun* and *haa* are not likely to co-occur with up-nods because cognitive change is not big when the information is just new but not interesting.

The figure also shows that *ne* co-occurred with down-nods more frequently than up-nods. As for the character of *ne* as sentence final particle, Kamio (1994) argued from the viewpoint of the theory of territory of information that a part of *ne* ("obligatory *ne*" as Kamio called) is used when the speaker assumes that (1) the information falls into both speaker and listener's territory or (2) that the information falls completely into the listener's territory and partially into the speaker's territory; thus, *ne* is used to seek assent, confirmation and reconfirmation. In other words, *ne* is used by a speaker when he or she assumes that the listener has same level or more detailed information about it. Even though Kamio (1994) argued about only *ne* produced by speakers, this particle is often used by listeners as well when the speaker has used it in the immediate context; for instance, "*Kyoo wa ii tenki da ne* (Today's weather is good, isn't it?)" followed by "*Ne* (Yeah.)". Applying above Kamio's notions (1) and (2) to listener's *ne*, it is assumed that both speaker and listener use ne only when they have same level of information because (2) cannot hold in the speaker side and the listener side at the same time. Therefore, when the listener uses *ne*, preceding information is not new for him or her, and the speaker also does not expect the listener receives the information as such.

Another interesting point is that *un* co-occurred with down-nods more frequently than up-nods when it is single occurrence, but this tendency gradually becomes inversed as the number of repetition increases. In general, single *un* is used as a continuer (Schegloff, 1982) or usual acknowledgement. On the other hand, repeated *un* is used to display one's agreement or understanding to the preceding utterance. Therefore, when the listener uses repeated *un*, he or she may have undergone a change in his or her cognitive state.

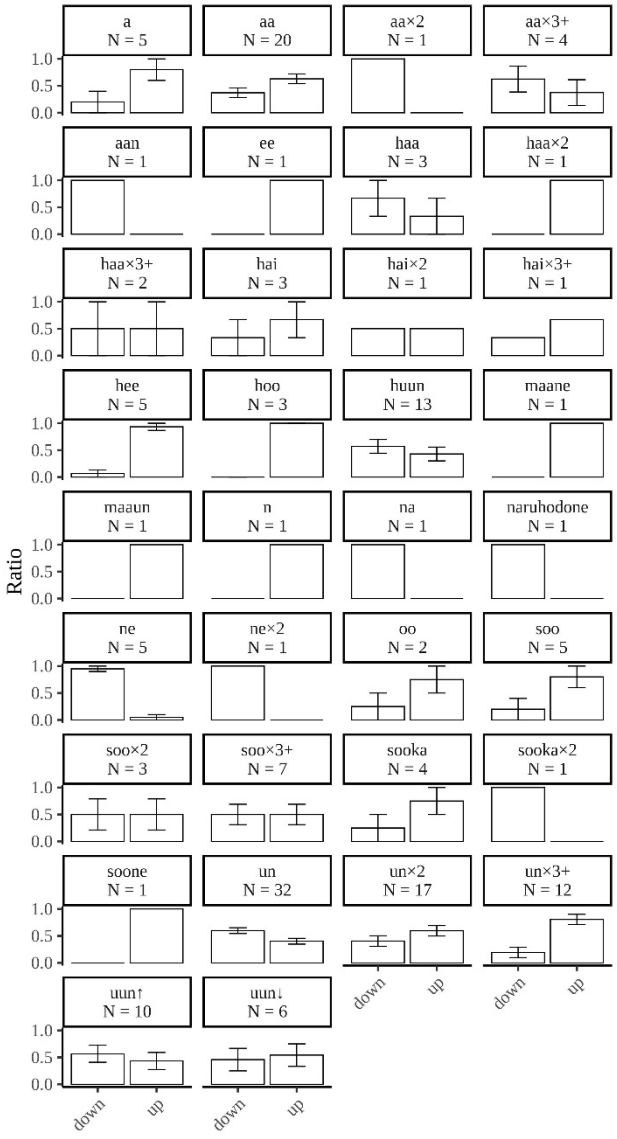

Figure 3: Ratios of up-nods and down-nods with respect to co-occurring feedback expressions

Next, we conducted statistical analysis to confirm significant difference between up-nods and down-nods. We built a generalized liner mixed model (GLMM) to predict a probability of up-nods from the feedback expressions and random intercept of participant. Since dependent variable is the binary values of up-nods and down-nods, we used Bernoulli distribution for probabilistic distribution. Parameters were estimated with Markov Chain Monte Carlo (MCMC). All these procedures were performed with R 4.2.0 (R Core Team, 2022) and the brms package 2.17.0 (Bürkner, 2017; Bürkner, 2018).

Figure 4 shows the estimated probability of up-nods with respect to co-occurring feedback expressions. Error bars show 95% confidence intervals, and expressions whose intervals do not contain 0.5 have significantly higher/lower probability of up-nods. As shown by the figure, *aa*, *ee*, *haa*×2, *hee*, *hoo*, *maane*, *maaun*, *n* and *soone* are significantly likely to co-occur with up-nods. On the other hand, *aa*×2, *aan*, *na*, *naruhodone*, *ne*, *ne*×2, *sooka*×2 and *un* are significantly likely to co-occur with down-nods.

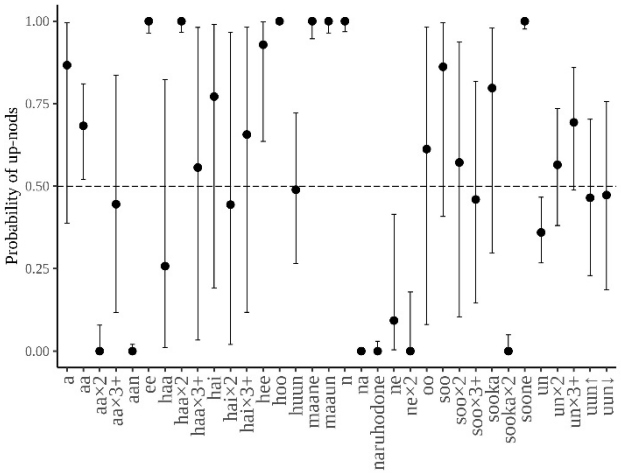

Figure 4: Estimated probability of up-nods

In conclusion, quantitative analysis showed that up-nods are used when the listener has undergone some kind of change in his or her cognitive state such as (1) when he or she receives new information (e.g., *a*, *aa*, *ee*, *hee* and *hoo*) and (2) when he or she understands preceding utterance (e.g., *un*×2 and *un*×3). On the other hand, down-nods are used (3) when he or she has prior knowledge of preceding information (e.g., *aa*, *ne*), (4) when the listener receives new but not interesting information (e.g., *huun* and *haa*) and (5) when he or she uses continuer (e.g., *un*).

4. Qualititative Analysis

4.1 Purpose

In this section, we conduct qualitative analysis of our data to presicely exmine when and how up-nods are used in terms of the type of preceding utterance.

4.2 Analysis

4.2.1 Inform

In the data, one of the positions where listeners use up-nods frequently was within or after the speaker's informing utterances. In excerpt (1), B provides the two listeners, A and C, with an information about her language skill that she can read Latin, Italian and German in line 01. This information may be new for both listeners. In addition, this informing can be heard as positive self-disclosure as well. In general, positive assessments might be more preferred as the response to this information, and in fact, A provides typical positive assesment "*sugoi* (Great)" in line 04. On the other hand, C produces only a particle "*hee* (Wow)" accompanied by an up-nod in lines 06-07, which are emotional expressions of surprise rather than assessment. This C's responses are not treated as problematic by the participants; she shows her surprise with the particle and up-nod, thereby, indirectly assessing A's skill in that it is so great that it deserves to be surprised. In fact, B repeats "*yomiageru dake da ttara* (If only reading aloud)" gazing at C in line 08, which seems to downgrade her skill; she may take A and C's assessments better than she expected. To sum up, in this case, the up-nod is used not only because the information is new, but also because of sequential preference.

(1) chiba0932 8:47-8:56
01 B: *yomu dake da ttara raten go to itaria go to*
If only reading, I can read Latin, Italian and
02 *doitsu go: wa dekiru yo*
German.
03 (0.13)
04 A: *sugoi*[*:*
Great.
05 B: [*hhh hu*
hhh hu
06 C: [*hee*[*:*
Wow.
07 [((up-nod))
08 B: [*yomiageru dake da ttara ne*
If only reading aloud.

In excerpt (2), participants talk about their angry story. Before this excerpt, C has finished his story telling, and A nominated B as next speaker and encourages him to tell his story next in line 01. However, B says he has no story to talk in line 06, and then, C pursues a new topic by proposing a "coming-of-age ceremony" story in line 10. Because A responds to it more strongly than B in lines 13-14, C misunderstands A has a story about the coming-of-age ceremony and encourages him to talk about it in line 15. However, A responds negatively in line 16, and provides an information that he did not even attend it in the first place in lines 20 and 23. After A has just said "*ore mazu i tte nai kara* (I didn't attend the coming-of-age ceremony in the first place)", C says "*a so kka* (Oh, I see)" and produces an up-nod in lines 21-22. In so doing, C seems to recognize that C's prior understanding that A attended the coming-of-age ceremony was wrong. Therefore, this information is not only new to C, but also contradicted with his prior understanding. To sum up, in this case, C's up-nod acknowledges A's new information and shows revision of his understanding at the same time.

(2) chiba0432 7:54-8:13
01 A: *tsugi Kitajima kun ((=B)) oko tta hanashi*
Next, Kitajima ((=B)), tell us your angry story.
02 C: *Wakaba-ku no hanashi*
Story about Wakaba-ku
03 (0.75)
04 B: *Wakaba-ku no hanashi*
Story about Wakaba-ku.
05 (1.08)
06 B: *iya* (0.24) *nai na toku n*i
No, nothing special.
07 C: *ji*[*tsu wa*
Actually
08 A: [*ue*[*e:*[*:*
Gah.
09 B: [*e* [:
Eh.
10 C: [*jitsu wa seejinshiki de* [*mitai na*
Like actually in the coming-of-age ceremony.
11 B: [*hara ga ta* [*tta*
Because I've never
12 *koto nai kara*
gotten angry.
13 A: [*aa a a*
Ah ah
14 [*aa a*[*a seejinshiki de* [*ne*
ah ah ah ah in the coming-of-age ceremony.
15 C: [*a* [*a* [*a tta a tta*
Oh. Oh. Was there? Was there?
16 A: *e nai yo hh* [*hu hu*
Eh, nothing. hu hu
17 C: [*na ha ha nai no ka*
na ha ha Nothing.
18 A: *iya demo* (0.05) [(0.1) *kono hito wa s-*
But this guy s-
19 B: [*ko- ika naka tta tte*
ko- He said he didn't attend.
20 A: *ore mazu i tte nai kara se*[*ejin shiki=*
I didn't attend the coming-of-age ceremony in the first place.
21 C: [*a so kka*
Oh, I see.
22 [((up-nod))
23 A : *=mo- moo* [*kae tte hen kara*
I didn't go back.
24 C : [*tooi mon na*
It's too far, isn't it?

4.2.2 Answer

The other position up-nods were frequently used was in the response to the answer to a question, especially seeking information. However, this is not surprising because we already showed that up-nods are likely to be used as acknowledgement for new information, and the answer to a question of seeking information should be new information for the questioner. In excerpt (4), C asks what club activity B did when she was a high school student in line 01. Even though the final particle "*kke*" seems to be used with consideration for the possibility she has ever heard it before, this question is designed as typical seeking information. After the question, B answers "mandolin" to this question in line 02, and then C says "*a so ka* (Oh, I see)" accompanied by an up-nod in lines 03-04. C may have heard it before and the information may not be strictly new for C, but because it is provided because of C's question, she, as the questioner, has to acknowledge it as new. Therefore, in this case, the up-nod is used not only because the information is new to C, but because C has the responsibility to acknowledge it as such as the questioner.

(3) chiba0332 1:08-1:10
01 C: *nani yatte ta* [*kke*
What did you do?
02 B: [*e mandori*[*n*
Um, mandolin.
03 C: [*a so ka*
Oh, I see,
04 [((up-nod))
05 *mandolin sa re ta n da ne*
you played the mandolin.

Before excerpt (4), B talked her story that she was suddenly asked if she could have an extra lunch box by a strange woman when she was on a train, and refused that offer. Successively, B describes the reason of her refusal that it is unclear whether or not the lunch box has already been opened in line 01. However, both A and C ask "*n*? (What?)" in lines 04-05 after a long silence of one second in line 02. Since these open class questions are typical repair initiator (Schegloff, 1977), A and C may have a

trouble for B's utterance in line 01. Moreover, because the silence in line 02 is long, B also self-repairs her previous utterance by explicitly specifying the subject of the sentence as "*bento* (lunch box)" in line 03. However, this B's self-repair overlapped with A and C's repair initiations and another silence occurs in line 06. B then repairs her original utterance by rephrasing it in line 07. At its possible completion, C says "*aa aa aa aa aa aa* (Ah ah ah ah ah ah)" and simultaneously produces an up-nod in lines 08-09. In this case, although the up-nod is used as the response to an answer, like excerpt (3), it is used to show that C's trouble for the preceding utterance is resolved rather than to acknowledge a new information.

(4) chiba0832 5:26-5:35
01 B: *ai teru ka ai te nai ka mo sa yoku wakan nai jan*
It is unclear whether it is open or not, isn't it?
02 (1.03)
03 B: [*bento*
lunch box.
04 A: [*n?*
What?
05 C: [*n?*
What?
06 (0.45)
07 B: *a aa tto i kkai ake ta ka doo* [*ka*
Ah, um, whether it is open once or not.
08 C: [*aa aa aa aa aa aa*
Ah ah ah ah ah ah.
09 [((up-nod))

4.2.3 Opinion

The next position up-nods were used was in the response to an other person's opinion. Before excerpt (5), C consulted A and B about her students she teaches in part-time job and said that her students look uncomfortable when she talks about a romance in the literature class. In lines 01-03, A offers her opinion to the consultation that teachers are thought not to say such things in Japan. However, C says "*soo na no ka na* (Is that so?)" and disagrees with the A's opinion in line 04. With consideration for this C's disagreement, A adds "*watashi wa omou* (I think)" and "*baito to ka shi teru to*: (based on my experience of part-time job)" to her opinion in line 06 to downgrade the evidence of her opinion from general fact to personal experience. Moreover, A gives the exception of her opinion "very friendly students" to make more concession to C in lines 11-12 and 14-15. In response to this, C finally changes her stance and strongly agrees with A by saying "*so so so so so so soo soo* (Yeah yeah yeah yeah yeah yeah yeah yeah)" and simultaneously producing an up-nod in lines 16-18. Thus, in this case, the up-nod shows not only agreement but also the listener's change of stance form disagreement to agreement..

(5) chiba0132 1:42-2:07
01 A: *n te ka sensee ga soo yuu koto wo yuu tte yuu koto*
I mean, because teachers are thought not to say such
02 (0.343) *ga:* (1.437) *nai koto ni na tteru kara Nihon*
things like in Japan.
03 *to ka da to*
04 C: *so*[*o na no ka na*
Is that so?
05 A: [*sugoku kiki zurai n ja nai* (0.13) *to* (0.548)
It is difficult for students to ask,
06 *watashi wa omou ano* (.) *baito to ka shi teru to*[:
I think, uh, based on my experience of part-time job.
07 C: [*un*
Yeah
08 *un un*
yeah yeah.
09 (1.5)
10 C: [*mada nanka:*
Still something
11 A: [*da kara:* (0.155) *sugoi da kara* (0.227) *kudake*
So, so there are also very friendly students and
12 *ta ko mo ite:*
13 C: *u*[*n*
Yeah.
14 A: [*soo yuu ko wa nani yu tte mo* [*heeki na n da*
such students don't care whatever they are said
15 *kedo:*
and,
16 C: [*so so so so so so*
Yeah yeah yeah
17 [((up-nod))
18 *soo so*[*o*
yeah yeah yeah yeah yeah.
19 A: [*goku hutsuu no sono sensee tte yuu no wa*
there are many and very normal students
20 *sensee na n da* [*tte omoikon deru ko ga kekkoo iru*
who believe teacher is a teacher,
21 (0.149) *de sho*
aren't there?
22 C: [*un un*
Right right.
23 [((up-nod))

4.2.4 Assessment

In our data, up-nods were used as the response to assessments few times. Before excerpt (6), B told her story that she lost her train pass worth 70,000 yen when she was a high school student but her parents did not scold her. In line 01, B expresses her thought that most parents scold their children in such situation and elicits agreements form the listeners. In fact, A provides an agreement to B's thought in line 03. On the other hand, C only accepts A's thought saying "*aa* (Ah)" but does not provide an explicit agreement. The possible reason why the two listeners provide different responses to A's thought is that although agreement is preferred as a response to other person's thought in general, an agreement in this case may be heard as acknowledging A's fault, which deserves to be scolded by her parents. Because of this dilemma, C avoids providing either agreement or disagreement. Moreover, even though A once provided an agreement in line 03, she also provides an assessment "*shoo ga nai* (hopeless)" in line 07. This assessment justifies the fact that A was not scolded by her parents, and therefore, A resolves the dilemma by producing both agreement and assessment. The chage of A's stance is also shown by her use of the conjunction "*de mo* (But)" in line 07. In line 10, C strongly agrees with this assessment saying "*uu un un un u* (Yeah yeah yeah yeah)" and simultaneously producing an up-nod. Even though this agreement contradicts A's thought, it can mitigate A's fault. In addition, an agreement is more preferred in this local context, i.e., after an assessment. Thus, C changes her stance from nuetral to agreement

with A. To sum up, similar with excerpt (3), the up-nod shows not only agreement but also the listener's change of stance.

(6) chiba0832 7:22-7:30
01 B: *hutuu okoru yo ne:*
Most parents scold, don't they?
02 (0.60)
03 A: *so*[*o* [*ne*
Right.
04 C: [*a* [*a*
Ah.
05 B: [*ne*
Yeah.
06 B: *ho*[*nnin:*
The person
07 A: [*de mo maa shoo ga* [*nai kara* [*ne:* [*otoshi cha*
But well, it's hopeless, isn't it? If you lose it.
08 *ttara ne:*
08 B: [*do:no*
Which
09 A: [*ma* [*ho-* [*un*
Well ho- yeah.
10 C: [*uu un un un u*
Yeah yeah yeah yeah yeah.
[((up-nod))

4.2.5 Other

The final position up-nods were used was after the place in which the listener should respond to the speaker regardless of the type of the preceding utterance. In excerpt (7), A talks about box seats on a train in line 01 and invites listeners' responses by producing a silence in the middle of the utterance. However, because she has said only "*bokkusu* (box)" prior to the silince, its meaning is not precisely conveyed to B and C and none of them can respond to it. The design of A's utternce has changed after the silence, and an explanation of "box seats" is added in line 02, assuming the listeners do not know it. In this way, it is clear that A invites the listeners' responses during the silence and because of the lack of responses she understands they do not know "box seats". However, at the same time with A's explanation of "box seats", both B and C provide acknowledgements in lines 03-04. This sueggests that they did not understand what "box" means just after it was produced, i.e., during the silence, but have understood it by the end of line 01. B shows her noticing with a change of state token "*aa* (ah)" accompanied by a down-nod in lines 3-4. On the other hand, C responds to A with repeated "*un* (yeah)" and an up-nod in lines 05-06. Although this "*un*" can be either an answer to the A's question "*wakaru* (you know?)" or delayed response to "box", it seems that the repeated format is designed to compensate for the absence of her response during the silence. Furthermore, the repeated *un* and up-nod can be seen as an account for the absence of her timely response. That is, C also recognizes that she should have responded to A during the silence but she could not because she did not understand what "box" meant. In this way, when the listener did not respond to the speaker at the time he or she should do that, up-nods are used as a display of delayed understanding and an account for the absence of a timely response.

(7) chiba0532 0:59-1:04
01 A: *are tamani: bokkusu* (0.191) *no yatsu wakaru*
Sometimes box (0.191) ones, you know?
02 [*seki ga bokkusu n na tteru yatsu ga aru no*
There are seats built like a box.
03 B: [*aa aa aa un*
Ah ah ah yeah.
04 [((down-nod))
05 C: [*un un un aru aru aru aru*
Yeah yeah yeah there are there are there are there are.
06 [((up-nod))

4.3 Summary of the analyses

In this section, we conducted qualitative analysis and precisely examined when and how up-nods are used in Japanese conversations. First, up-nods are used to achieve multiple interactional actions. When they were used as acknowledgement for new information, they also conveyed that the listener's misunderstanding or trouble for the preceding utterance has been resolved, or that there was a sequential reason why he or she had to use them. This result suggests that the listeners might use not only verbal feedback but also up-nods at the same time in order to achieve these multiple actions. Second, up-nods are used when the listener's cognitive state has changed after hearing the preceding utterance. For instance, when up-nods were used after informing or answering, they indicate that the information provided by the utterance was not only new for the listener but contradicts his or her prior knowledge. In other case, the listener had a trouble understanding the preceding context, and used up-nods to show the preceding utterance resolved the trouble. When up-nods were used as agreement, the listener had a stance unaligned to the speaker's opinion or assessment. In these cases, the cognitive change happening inside the listener might be bigger than when the information is just new or when the listener has a similar opinion or assessment to the speaker; the possibility of using up-nods might also be higher in these cases.

5. Disscussion

In this study, we used both quantitative and qualitative analyses to investigate when and how up-nods and down-nods are used as feedback signals in Japanese conversations and how their usage differs depending on the cognitive state of the listener. As the result of the quantitative analysis, up-nods seem to co-occur with change of state feedback expressions more frequently than down-nods. This result suggests that up-nods are used when the listener did not know the information but comes to understand it by hearing the preceding utterance. On the other hand, down-nods are used with expressions indicating that the listener already knows the presented information, or the listener did not know the information but does not have interest in it, or when the listener uses a continuer. As the result of qualitative analysis, up-nods are used when the listener's cognitive state has changed after hearing the preceding utterance, for instance, if the listener had no prior knowledge about the preceding utterance, had contradicting knowledge about it or when the listener disagreed with or took a neutral stance to the speaker's opinion or assessment before the preceding utterance. Generalizing the results of the two analyses, we conclude that up-nods are related with some kind of

change of cognitive state. In other words, up-nods signal cognitive change in addition to the usual meanings of nods such as “now I know it”, “now I understand it” or “now I agree with it”.

In this study, we can confirm that the distinction between up-nods and down-nods in Nordic cultures can be observed in Japanese. However, new question arises here; why do up-nods have similar meaning in completely distinct cultures? The most likely answer to this question is that up-nods are related with human’s physiological response. This is because if up-nods had been developed from physiological response, it is natural that they are used similarly in distinct cultures. When we are surprised, we sometimes quickly move our head back. This movement may be physiological response to distance oneself from an object when we feel in danger. That is, we think up-nods are copositive movement composed of physiological head back and nods. Moreover, even though nods are used as positive feedback in many cultures, they are also used as negative, especially emotional negative feedback in Mediterranean cultures (Morris, 1977). The fact up-nods are related with the producer’s emotion also supports our hypothesis.

6. Application to conversational agents

As mentioned in the beginning of the paper, in order to support natural interaction with the user, conversational agents should also have a capability to understand and generate appropriate feedback signals, and in particular, they should distinguish the different functions of up-nods and down-nods in different conversational environments. To the best of our knowledge, Wikitalk (Jokinen & Wilcock, 2014), which works in Finnish, English, and Japanese, is the first application to explicitly distinguish up-nods and down-nods as part of the Nao robot’s presentation and feedback strategies. The decisions are based on a rather simple model of the robot’s expectations of the continuation of the dialogue: the robot reacts to unexpected user actions, e.g., requests to stop the conversation, by up-nods signaling surprise, while it reacts to usual inform actions by down-nods.

The findings of the current study can also be applied to conversational agents: this requires that the expectation model is extended with a component that models the partner’s internal cognitive state (such as knowledge, understanding and stance), and on the basis of which the agent can decide on the appropriate type of nod.

Figure 5 is a conceptual diagram of the agent with such a cognitive state update facility. First, the user produces an utterance, and the agent analyses its meaning. Second, the internal state update module calculates a new internal state and calls the feedback module. Third, the feedback module determines the type of nod depending on whether or not the internal state has been changed, and outputs the result to the gesture module. It should be noticed that although the model focuses on the type of nod to be generated, also, the type of possible verbal feedback expression is to be determined in this phase, see the CDM architecture in Jokinen & Wilcock (2014). Finally, the gesture module produces an appropriate nod, and the verbal component produces a verbal expression.

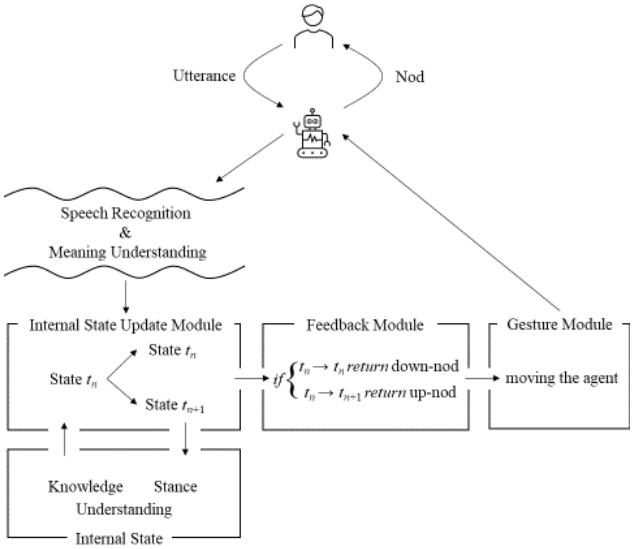

Figure 5: Conceptual diagram of the proposed system

7. Conclusion

Nods are one of the main feedback behaviors in many cultures. Moreover, this study confirmed that they are used in quit similar way in even completely distinct cultures such as Finnish and Japanese. In addition, the fact nods are important in human-human interactions suggests that they are also important in human-agent interactions. Therefore, we also proposed the architecture of the system which has the capability to generate suitable type of nod. In the future work, we aim to build a conversational agent that realizes this model and can evaluate the effectiveness of our model by subjective assessment experiment.

8. Acknowledgement

The study is based on results obtained from project JPNP20006 commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

9. References

Boholm, M. & Allwood, J. (2010). Repeated head movements, their function and relation to speech. In *Proceedings of LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. Valetta, Malta.

Den, Y. & Enomoto, M. (2007). A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. In Nishida, T. (Ed.), *Conversational Informatics: An Engineering Approach*. Hoboken, NJ: John Wiley & Sons, 307-330.

Den, Y., Yoshida, N., Takanashi, K. & Koiso, H. (2011). Annotation of Japanese response tokens and preliminary analysis on their distribution in three-party conversations. In *Proceedings of the 14th Oriental COCOSDA (O–COCOSDA 2011)*, 168-173.

Endo, T. (2018). The Japanese change-of-state tokens a and aa in responsive units. *Journal of Pragmatics*, *123*, 151-166.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action*. Cambridge University Press, Cambridge, 299-345.

Jokinen, K. (2018). Dialogue models for socially intelligent robots. In Ge S. et al. (Eds.), *Social Robotics*. ICSR 2018. Lecture Notes in Computer Science, *11357*. Springer, Cham, 127-138.

Jokinen, K. & Wilcock, G. (2014). Multimodal open-domain conversations with the Nao robot. In J. Mariani, S. Rosset, M. Garnier-Rizet, & L. Devillers (Eds.), *Natural interaction with robots, knowbots and smartphones*. Springer, New York, NY, 213-224.

Kamio, A. (1994). The theory of territory of information: The case of Japanese. *Journal of Pragmatics*, *21*(1), 67-100.

Morris, D. (1977). Man watching. *A field guide to human behaviour*, Elsevier Publishing Projects Ltd., London.

Navarretta, C., Ahlsén, E., Allwood, J., Jokinen, K. & Paggio, P. (2012). Feedback in Nordic first encounters: a comparative study. In *Proceedings of LREC 2012*. Istanbul, Turkey, 2494-2499.

OpenCV. (2020). *Open Source Computer Vision Library*.

R Core Team (2022). R: The R Project for Statistical Com puting. R Foundation for Statistical Computing, Vienna, Austria. Retrieved May 20, 2022, from https://www.R-project.org/

Schegloff, E. A., Jefferson, G. & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361-382.

Schegloff, E. A. (1982). Discourse as an interactional achievement: some uses of ‘‘uh huh’’ and other things that come between sentences. In D. Tannen (Ed.), *Georgetown University Roundtable on Language and Linguistics*. Georgetown University Press, Washington, DC, 71-93.

Tanaka, H. (2010). Multimodal expressivity of the Japanese response particle huun. In D., Barth-Weingarten, E. Reber, M. Selting (Eds.), *Prosody in Interaction. John Benjamins*. Amsterdam/Philadelphia, 303-332.

Toivio, E., & Jokinen, K. (2012). Multimodal feedback si gnaling in Finnish. In A. Tavast, K. Muischnek & M. K oit (Eds.), *Human Language Technologies - The Baltics Perspective: Proceedings of the Fifth International Co nference Baltic HLT 2012 (Frontiers in Artificial Intelli gence and Applications; Vol. 247)*. IOS PRESS. 247-25 5.

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross validation and WAIC, *Statistics and Computing*, *27*, 1413-1432.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P., Paananen, T. & Gelman, A. (2020). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models, *R package version 2.4.1*.

Watanabe, T., Okubo, M., Nakashige, M. & Danbara, R. (2004). InterActor: Speech-driven embodied interactive actor. *International Journal of Human-Computer Interaction*, *17*(1), 43-60.

Watanabe, T. & Yuuki, N. (1989). A voice reaction system with a visualized response equivalent to nodding. *Advances in Human Factors / Ergonomics*, *12A*, 396-403.

Yatsuka, K., Kawabata, K. & Kobayashi, H. (1997). A robot listener for fluent verbal communication. *IEEE RO-MAN 7*, 408-411.

Yatsuka, K., Kawabata, K. & Kobayashi, H. (1998). A study on psychological effects of human-like interface. *IEEE RO-MAN 8*, 89-93.