PRELIMINARY ANALYSIS OF SELF-CONFIDENCE LABELS OF SECOND LANGUAGE UTTERANCES IN AN ONLINE ASSIGNMENT TO DEVELOP A LEARNING SYSTEM WITH FEEDBACK ADAPTATION

Omori, Yuga
Kyushu University

Tamura, Mika
Kyushu University

Ikeda, Daisuke
Kyushu University

出版情報：INTED2023 Proceedings, pp.6532-6541, 2023. IATED
バージョン：

PRELIMINARY ANALYSIS OF SELF-CONFIDENCE LABELS OF SECOND LANGUAGE UTTERANCES IN AN ONLINE ASSIGNMENT TO DEVELOP A LEARNING SYSTEM WITH FEEDBACK ADAPTATION

Y. Omori, M. Tamura, D. Ikeda

Kyushu University (JAPAN)

Abstract

This study focuses on self-confidence of the second language (L2) learners' speaking. The L2 self-confidence is important because it affects L2 competence and the L2 willingness to communicate. The confidence has been measured by questionnaires or interviews, but they are not suitable to measure dynamic confidence frequently. Our approach is to develop a learning system that has a machine learning model to predict learners' confidence. To our knowledge, there is no dataset of L2 utterance and self-confidence to achieve this. Therefore, we conducted an experiment to collect these data from 14 international students in an online Japanese course at Kyushu University. This paper reports some findings from a preliminary analysis of the collected data. Our developed prototype system for L2 speaking collected approximately 4,500 data which consist of utterance audio and 4-point confidence labels. Various labeling patterns suggest that a predicting model should be flexible for each learner's confidence. These results provide us with a direction to create the model, such as an investigation of some variables that could affect the confidence labels.

Keywords: Second Language Learning, Learning system, Speaking, Self-confidence, Web Application.

1 INTRODUCTION

Among the four basic language skills, listening, speaking, reading, and writing, speaking is vital in the language classroom [1] and also positioned as an activity that provokes learners' anxiety [2]. It is already known both theoretically and practically that psychological aspects, such as anxiety and self-confidence, affect language learners. From a theoretical perspective, there are the following two studies: a hypothesis that negative affects disrupt learners' language acquisition [3], [4]; a model of willingness to communicate using a second language (L2) [5]. Several practical studies have reported the effects of language classroom anxiety on language learners [6], [7] and a relationship between the achievement of oral interviews and self-confidence [8]. Thus, it is important to consider the learner's psychological aspects.

To this end, we start with the following question: how can teachers help the psychological aspects of the L2 learners? A practical study has suggested that teachers can assess learners' psychological aspects, and they can develop activities that consider the aspects including L2 self-confidence [9]. L2 self-confidence is an important factor which is listed in the above theoretical studies. As ways to assess the confidence, they mentioned the questionnaire and interview, which have been widely used in related studies. Note that these methods are able to assess static confidence at the time of the survey or at a certain point of time. It is likely that the confidence changes depending on some situations [5]. If we want to assess learners' dynamic confidence, we need to increase the frequency of these surveys. However, it is difficult to do that because teachers have limited time to support learners and frequent surveys are burdensome for them. To capture the dynamic changes, we need an alternative method, such as a language learning system that automatically predicts learners' one.

To develop such a system, we believe that machine learning is promising. However, there is no dataset of L2 utterance and self-confidence to our knowledge, so we first have to collect such data to create a model or relationship between them. We also need to see the characteristics of these data. Thus, this study aims to create a self-confidence dataset and provide some findings from a preliminary analysis of the dataset. To achieve this, we used our prototype system for speaking Japanese in an online Japanese course at our university. Our prototype collected L2 utterance audio and L2 self-confidence labels from international students. From these data, we found the labeling patterns labeled differently for each student.

2 RELATED WORK

In this section, we will introduce L2 self-confidence and related datasets from previous studies.

2.1 L2 self-confidence

2.1.1 Definition

Before we look at several methods to collect data on self-confidence, we should clarify the definition of self-confidence we will use here. [10] defined self-confidence in using L2 as following: “low anxious affect and high self-perceptions of L2 competence.” That is, L2 self-confidence is a complex concept that has both cognitive and affective aspects. It also suggested that this confidence influences L2 proficiency both directly and indirectly.

Whereas these constructs of the confidence as above reflect personal characteristics, [5] suggested that self-confidence varies with some situations. It made a model of L2 Willingness to Communicate (L2 WTC) and listed dynamic confidence as one of those factors. It is called state self-confidence and is described as follows: “(...) it is likely that some situations will entail more confidence than others, primarily depending on characteristics of prior L2 contact (...) in these specific situations.” It also consists of the two aspects. However, it is suggested that some situations can increase or decrease the confidence. As the situations, they suggested that several factors appear to be particularly relevant. For example, topics of communication will affect language use [11]. Thus, the L2 self-confidence that influences the L2 WTC is not always constant but is seen as continually changing with the situations.

2.1.2 Approaches to measurement

Here, we summarize how learners’ L2 self-confidence has traditionally been measured. In several studies of investigating L2 proficiency or L2 WTC, questionnaires [8], [9], [12] and interviews [13] have been conducted to measure the L2 self-confidence. Most of these studies’ contexts ask about confidence, which is supposed to influence the proficiency and the WTC, to measure them. Some of the studies (e.g., [14], [15]) separately measure anxiety and self-perceptions, which are components of the confidence. These methods allow us to measure the confidence at the time of the survey or at a certain point of time. Interviews are particularly effective tools for investigating the relationship with the proficiency or the WTC. It is because they allow us to directly ask the learners about the materials for the cause of their self-confidence. However, these traditional methods are not suitable to measure the state self-confidence which changes dynamically with the above situations because the frequent measurement is a burden to the L2 learners. If we apply the dynamic changes to feedback generated by teachers and learning systems, an alternative method is required.

2.2 Related datasets

In order to determine the settings of our experiment to collect L2 self-confidence data, we look at the ones of the previous studies from the following four perspectives (see Table 1): (1) Raters: Who labeled the data; (2) Labels: What the raters labeled; (3) Materials: What is used to estimate the labels; and (4) Context: Under what situation was the labels and the materials collected.

The first thing to note is who labeled materials. These raters are broadly categorized as either material providers or otherwise. Which rater is appropriate depends on the context of studies. For example, on [16] and [17], self-confidence is labeled by the material providers. These confidences are referred to as the degree of speaker’s one perceived by listeners. In this case, it is appropriate that non-speakers label the confidence. When we select raters, we have to consider the context of the study.

The second is what raters labeled. A particularly important is the scale of labels. It can be seen from Table 1 that the labels, labeled with confidence and similar factors, are different scales. For example, [17] used 7-point Likert scale from 1(Not Confident) to 7(Very Confident) and [18] used a 5-point scale like the semantic differential scale. If a study is in a digital context, a slider or radio button can be utilized to label it. In the case of [19], the slider enables the learners to report detailed self-efficacy from 0 (low) to 100 (high).

3 METHODOLOGY

To collect L2 self-confidence labels with L2 utterances, we conducted an experiment. Here we will show three aspects of the experiment: setting, our prototype system, and procedure.

Table 1. Settings of the collected data in the previous studies.

Authors	*Raters*	*Labels*	*Materials*	*Context*
Ishimaru *et al.* [20]	Learners	Self-confidence (Yes/No)	Learning data & Eye gaze	Multiple-choice questions (MCQ) on learning system
Cooper *et al.* [18]	Learners	4 affective states (5-point)	Learning data & Sensor data	Learning system for geometry
McQuiggan *et al.* [19]	Learners	Self-efficacy (101-point)	Physiological response etc.	MCQ on learning system
Krajewski *et al.* [16]	Annotators	Self-confidence (10-point)	Speech audio	Lectures of participants
Dinkar *et al.* [17]	Annotators	Self-confidence (7-point)	Video	Review of movies
Forbes-Riley & Litman [21]	Annotators	Certainness (One of four labels)	Spoken audio	Tutoring dialogues of physics

3.1 Setting

3.1.1 Label

We set up this experiment based on the related studies in the previous section. Firstly, L2 self-confidence in this study was defined as follows.

(i) *Learner's perception of proficiency in L2 speaking*. We should treat L2 self-confidence as what has both cognitive and affective aspects. However, it is suggested that L2 anxiety in a digital context is different than in a traditional classroom context [9]. In this experiment, we will focus on L2 self-perception which has a common definition outside of L2, such as [22].

(ii) *Dynamic state that changes finely*. We treat the confidence as a dynamic state influenced by several factors (e.g., topics in L2 communication). For learning systems to generate feedback adapted to learners' confidence frequently, we have to observe the finely tuned changes of the confidence. Therefore, we will collect self-perception of the learner's L2 utterance as self-confidence label. In this study, the confidence will be asked on a 4-point semantic differential (SD) scale.

3.1.2 Environment

We conducted an experiment to collect labeled data on a Japanese course for international students at Kyushu University in Japan. The target course was "Active Japanese" held from April to July 2022. In this course, the students learn Japanese as L2 for daily communication. Due to COVID-19, this course was conducted online, and the students participated in it via a videotelephony software. We chose the students attending this course as the participants of our experiment to collect the above confidence labels.

3.1.3 Participants

Twenty-five students attended this course, and 14 of them participated in our experiment (Undergraduate: 5, Graduate: 8, Research student: 1). Henceforth, these 14 students will be referred to as participants. They had studied Japanese for half to three years before attending this course. Their mother tongues include Chinese, Korean, Malay, Indonesian, Spanish, Shona, and French. The reasons for attending this course were varied such as to communicate with Japanese, to pass the Japanese-Language Proficiency Test, and to find employment. According to the Common European Framework of Reference (CEFR), their Japanese proficiency was approximately A2 or B1 judged by the teacher of this course. Of them, three participants were not able to enter Japan at the start of this course.

3.2 Our prototype system

To collect the utterance audio and the self-confidence labels, the participants used a prototype system that we developed as an online assignment to practice speaking Japanese. This prototype was modified to its usability [23] and the function to collect learners' operational logs [24]. Table 2 shows the configuration of our prototype.

Our prototype, a web application running in web browsers, can be accessed from learners' PCs or smartphones. The learners can use our prototype anytime and anywhere by logging in with their ID and password. They need to prepare only the microphone as the only necessary peripheral (e.g., microphones built into learners' devices or external microphones on their earphones).

Next, we will explain the configuration of the web page for speaking practice and the procedures for the practice (see Figure 1). The page is composed of two areas: question area where example sentences and instructions are displayed, and operation area where the learner operates some buttons and other controls. In the former area, an option to have a pre-prepared synthetic voice, generated by [25], read out the displayed text is available. This option allows the learner to listen to the voice as many times as needed. In the latter area, as described in more detail below, several buttons, voice players, and messages are shown/hidden depending on the progress of the speaking practice.

Then, we will show the procedures for the practice. Learners can first record their utterances after pressing the Record button. From the audio player that appears after recording, the learners can listen to their recorded utterances. To label the L2 confidence in their utterances, our prototype shows four buttons, from *Not confident* to *Confident*. After the learners press one of the four buttons, they are asked by the prototype whether the utterance is to be assessed by the teacher or not (Save/Reset). This feature allows the learners to practice their speaking as many times as they are satisfied. Some questions display a model answer when the Save button is pressed. In this experiment, labeled utterance audio is collected by our prototype in either case. The utterance audio is generated as Waveform Audio File Format (WAV). Our prototype collects downsampled WAV files (Sampling Frequency: 16kHz, Bit Depth: 16bit).

Table 2. The configuration of our prototype.

	Tool / Language
Client	HTML / JavaScript
Web flamework	Django (Python)
Database	MySQL
Web server	Nginx

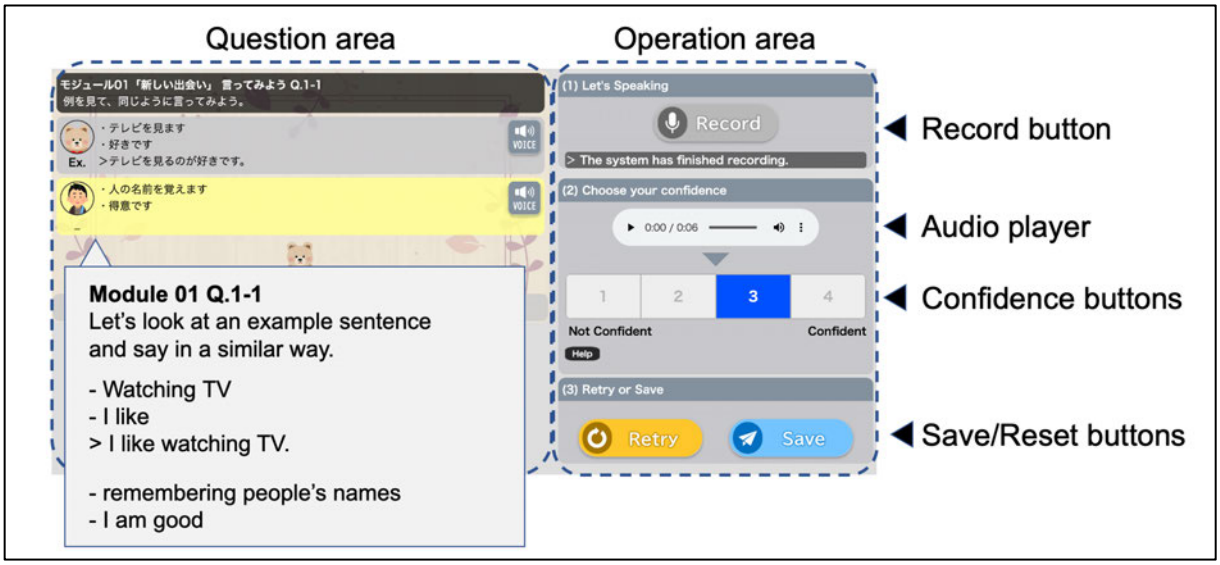

Figure 1. A page of our prototype for speaking practice.

3.3 Procedure

This experiment was conducted with the approval of the Experimental Ethics Committee on the Faculty of Information Science and Electrical Engineering at Kyushu University. We conducted this experiment over one semester, approximately 15 weeks. During this experiment, we had never met the participants in the classroom.

We will first introduce the preparations to provide our prototype for the participants. We explained the experiment and the prototype to the participants in the first lecture. At this time, we also explained what should be labelled as follows.

> *To what degree are you confident that your utterance conveyed what you were supposed to tell a native Japanese listener? 1(not confident): A native doesn't understand what you mean at all. 4(confident): A native definitely understands what you mean.*

The participants were able to reconfirm this explanation from our prototype at any time. We also explain that labelled values do not affect the grade of this course. After the explanation, we conducted a questionnaire on Moodle, a Learning Management System used by our university. The purpose of this questionnaire is to collect the consent of our experiment and personal information from the participants. After the second lecture, IDs and passwords for our prototype were distributed via e-mail to participants who responded to the questionnaire.

The assignments were added every two weeks, each with a deadline. The participants were able to practice their speaking many times in the assignments until each deadline. Our prototype provides the following three types of speaking tasks: (i) Pattern Practice. The learners have to combine given words or sentences as one sentence like a given example sentence; (ii) Fill in a blank of dialogues. The learners have to speak the appropriate sentence for the blank in the given dialogues; and (iii) One-minute speech. The learners have to speak for one minute on a given topic.

4 RESULTS

4.1 Collected data

Fourteen participants provided 4,641 labeled audio data via our prototype. Of them, 70 data were excluded according to the following exclusion criteria: (1) data with the audio that was not recorded properly, and (2) data with the *Save* audio recorded for less than 1 second. Table 3 shows the breakdown of the remaining 4,571 data.

Next, we will rearrange these 4,571 labels by participants. Here we exclude the data of the two participants because they did on less than half of all assignments. Figure 2 shows the confidence labels of the remaining 12 participants in chronological order. The points plotted in Figure 2 are arranged at equal intervals for each participant, not at intervals of actual elapsed time.

4.2 Grouping of the participants

Prior to our analysis, we need to reach a common understanding on the type of the label's scale because different types have different values available for analysis. There is ongoing debate as to the SD scale is treated as either an ordinal scale or an interval scale. For this preliminary analysis to reveal as much as possible, we shall treat this scale as an interval scale on which means can be computed.

Table 3. Details of the confidence labels collected by our prototype.

Confidence	***Save***	***Reset***	***Total***
1 (not confident)	84	1,277	1,361
2	287	362	649
3	1,097	152	1,249
4 (confident)	1,283	29	1,312
Total	2,751	1,820	4,571

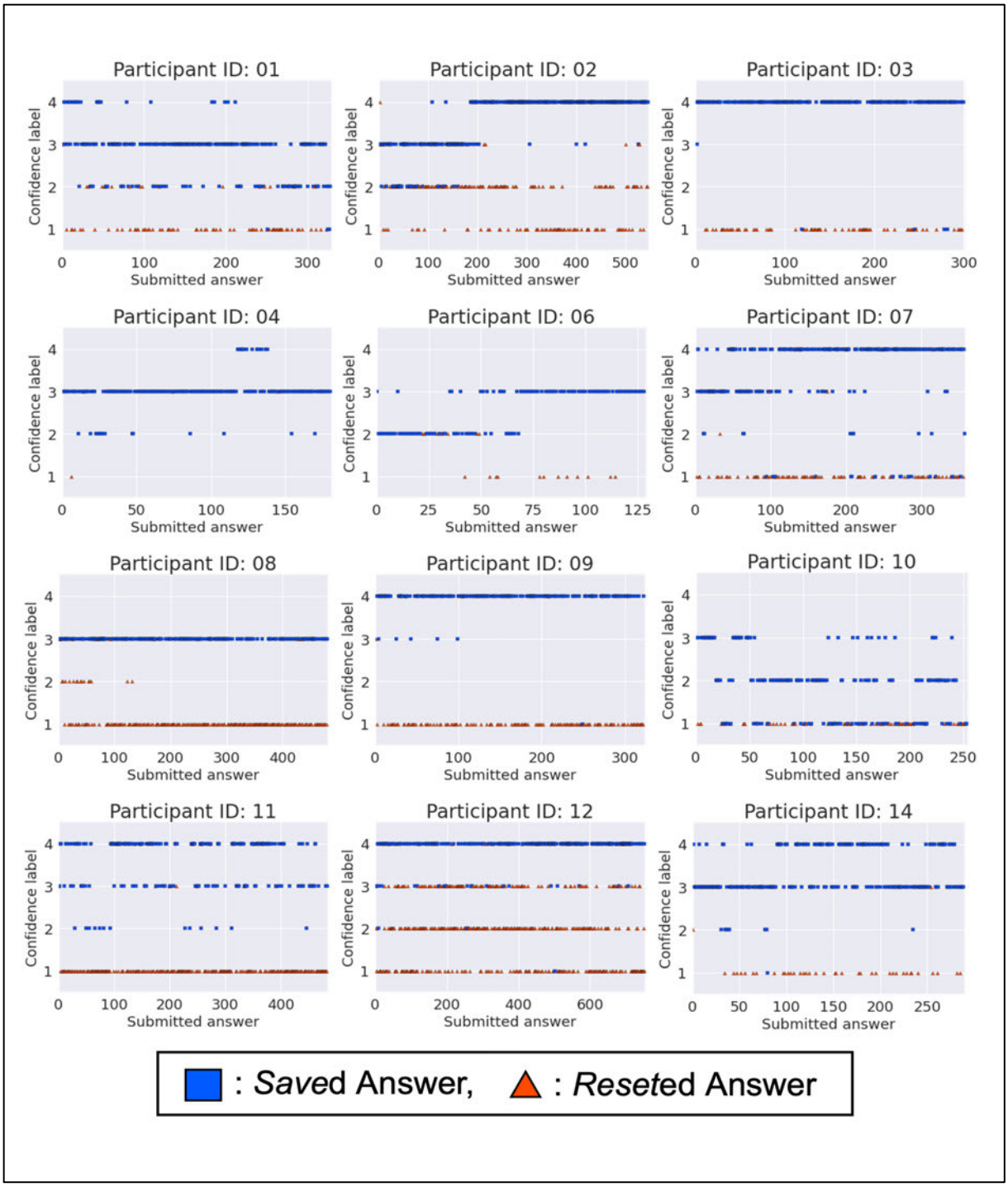

Figure 2. Changes in the confidence labels for each learner

To consider models that predict learners' confidence automatically, we will attempt to group the learners based on the collected labels. Our grouping used statistics of confidence labels. We also calculated entropy to reveal detailed labeling patterns of each learner. Table 4 shows the following three values: Mean, Variance, and Entropy.

Then, we will see some labeling patterns of the learner by classifying each value into binary groups around a threshold value. Each threshold value is the mean value of each column as follows: (1) Mean: 2.57; (2) Variance: 1.14; and (3) Entropy: 1.42. Groups above the threshold value are designated *High* <*value*> and groups below one are designated *Low* <*value*>. Table 5 shows the results that are divided into four groups based on the combination of variance and entropy, and then divided into two mean groups.

Table 4. Values of three indicators for confidence changes of each learner.

Participant ID	CEFR	Size	Mean	Var	Entropy
01	A2	329	2.404	0.815	1.729
02	B1	547	2.921	1.175	1.860
03	A2	301	3.199	1.760	0.866
04	B1	181	2.978	0.166	0.785
06	B1	129	2.411	0.431	1.345
07	B1	358	2.874	1.684	1.627
08	B1	481	1.967	0.970	1.168
09	A2	324	2.712	2.174	1.101
10	B1	254	1.685	0.564	1.472
11	B1	485	2.105	1.751	1.517
12	A2	753	2.607	1.340	1.967
14	A2	290	3.021	0.892	1.565

Table 5. Results of grouping the participants by three indicators.

(Var, Entropy)	Low mean	High mean
(Low, Low)	06, 08	04
(Low, High)	01, 10	14
(High, Low)	-	03, 09
(High, High)	11	02, 07, 12

5 DISCUSSION

We collected L2 self-confidence labels to analyze learners' confidence in the L2 utterances. The labeling pattern of each participant is grouped by the three indicators. This result suggests that these indicators describe certain characteristics of the labeling patterns. In the following, we will discuss the collected data from three perspectives.

First, we begin the discussion about the detail of the confidence labels. The percentages of the collected labels are following: 1: 29.8%, 2: 14.2%, 3: 27.3%, and 4: 28.7%. This result is contrary to our expectation that the labels are concentrated in 2 and 3 because people avoid extreme choices in general, called Goldilocks. We observe from Figure 2 that the above percentages reflect differences in the standard values of each participant's confidence. It could be one piece of evidence that flexibility is needed in models to predict learners' confidence. The differences between the standard values could be influenced by learners' over- or underestimation. For example, [26], who investigated changes in learners' self-esteem in certain lectures, has suggested that the learners were affected positive bias [27] or negative bias [28]. It is difficult to discuss this hypothesis from the limited number of participants, but we should have this perspective for future analysis.

Next, we discuss changes in confidence labels for each participant. Figure 2 shows various patterns of change in confidence labels; staying constant, going up/down between certain answers, and fluctuating. [15] illustrated a trend that the higher the academic year of international students, the higher L2 self-confidence. We interpreted it as an increase in confidence with experience and expected these confidence labels to increase with each utterance. However, the collected labels were contrary to this expectation. This finding suggests that learners' confidence during speaking practice does not necessarily change linearly. It was revealed by asking the participants to label their confidence with each utterance. In addition, we attempted to group the patterns using several indicators to create a model describing these various labeling patterns. Table 5 and Figure 2 show the suggestion that the grouping using entropy as well as mean and variance is available to estimate labeling patterns of the confidence. To estimate detailed changes in the confidence, further indicators and perspectives are necessary. To this end, we first need to investigate the influence of non-learner variables because of the lack of knowledge about changes in confidence of L2 utterance. One of the variables is the attribute of the

assignments implemented in our prototype. For the participants to prepare and review the lectures, we prepared a variety of assignments: topics that match each lecture, with/without model answers, and that take only a few seconds or one minute to speak. That is, the labeling patterns seem to be influenced by the difficulty of the assignments and the learner's level. Our future work will analyze how each element of the assignments affects L2 self-confidence.

Finally, the collected Save/Reset labels also show interesting results. In this experiment, the criteria for selecting Save/Reset labels were left to each participant. The results suggest that some characteristics of individual learners affect the Save/Reset criteria. From the details of the Save/Reset labels, we found the following facts. Figure 2 shows that the number and percentage of Reset vary from participant to participant. Some participants used it heavily, while others did not. In addition, Table 3 shows various combinations of the confidence labels and the Save/Reset labels. Although the confidence labels for Save tended to be higher and for Reset lower, there are 1-Save and 4-Reset combinations. In the 4-Reset audio, there is an utterance with no speaking errors. The Reset label also shows that they are used for causes other than speaking errors. As one of the characteristics, speaking styles can influence the number of Reset labels and thus the number of low-confidence labels. [29] classified L2 learners into two types based on the relationship between the learners and the language learning environments. The differences between them are frequency of corrections during speech due to whether speech content is planned before the speech or not. Therefore, it is interesting to investigate the following: "What are the reasons for Reset?", "What corrections are made to speaking errors?", and "How does the correction affect their confidence?".

6 CONCLUSION

To develop a learning system that predicts learners' confidence, 14 international students who attended an online Japanese course provided us with approximately 4,500 confidence labels. This paper reported some findings from a preliminary analysis of these labels. In conclusion, these labels suggested that a predicting model should be flexible for each learner's confidence.

The details of the collected labels were different for each learner, as the values of the labels did not concentrate on the non-extreme values. This result suggests that different learners have different feelings of the confidence and self-evaluation biases. In addition, the grouping of each labeling pattern suggested that the entropy is available to understand a part of the patterns. However, this grouping alone is not sufficient to predict the detailed patterns of the confidence. For this prediction, we should investigate the relationship between the labels and some attributes of the assignments. These findings may apply to the design of lectures and activities for second language (L2). Furthermore, although this study targets Japanese as L2, it could be adapted to other languages such as English.

This experiment has four limitations as follows. The first is the data used in our analysis. This paper focuses only on the labels and does not treat the utterance data and the logs of learning behaviors that are collected by our prototype system. In future analysis, we will use these data to focus on the contents of the utterances and the operation of our prototype during speaking. The second is the number of participants. We attempted a limited analysis because of the limited number of them. To reveal labeling patterns, more data enable us to analyze the labels from a further perspective. Furthermore, we are planning on adding questionnaires and interviews because this finding suggests the need to survey learners' characteristics. The third is the assignments for L2 speaking. This experiment was conducted in the context of an online assignment designed to prepare for and review the lectures. Thus, the topics, the difficulty, and the number of assignments were not controlled for this experiment. In the future, we will conduct an experiment in which the assignments' contents are controlled in order to reveal the influence of the assignments on the confidence. Finally, the fourth is the components of L2 self-confidence. The confidence in this experiment only consists of the learner's self-perception without the anxiety. When revealing the influence of the confidence on L2 willingness to communicate, the anxiety also needs to be measured in some methods.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP22K00789.

REFERENCES

[1] A. Burns, “Concepts for Teaching Speaking in the English Language Classroom,” *Learn J. Lang. Educ. Acquis. Res. Netw.*, vol. 12, no. 1, pp. 1–11, 2019, Retrieved from https://so04.tci-thaijo.org/index.php/LEARN/article/view/168564

[2] E. K. Horwitz, M. B. Horwitz, and J. Cope, “Foreign Language Classroom Anxiety,” *Mod. Lang. J.*, vol. 70, no. 2, pp. 125–132, 1986, doi: 10.1111/j.1540-4781.1986.tb05256.x.

[3] S. D. Krashen, *Second Language Acquisition and Second Language Learning*. 1981. doi: 10.4018/978-1-59140-723-2.ch007.

[4] S. D. Krashen, *Principles and Practice in Second Language Acquisition*, vol. 17, no. 2. 1982. doi: 10.2307/3586656.

[5] P. D. Macintyre, Z. Dornyei, R. Clément, and K. A. Noels, “Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation,” *Mod. Lang. J.*, vol. 82, no. 4, pp. 545–562, 1998, doi: 10.1111/j.1540-4781.1998.tb05543.x.

[6] M. L. Price, “The Subjective Experience of Foreign Language Anxiety: Interview with Highly Anxious Students,” in *Language Anxiety: From Theory and Research to Classroom Implications*, E. K. Horwitz and D. J. Young, Eds. Prentice Hall, 1991, pp. 101–108.

[7] E. M. Phillips, “The Effects of Language Anxiety on Students’ Oral Test Performance and Attitudes,” *Mod. Lang. J.*, vol. 76, no. 1, pp. 14–26, 1992, doi: 10.1111/j.1540-4781.1992.tb02573.x.

[8] D. O. Pyun, J. S. Kim, H. Y. Cho, and J. H. Lee, “Impact of affective variables on Korean as a foreign language learners’ oral achievement,” *System*, vol. 47, pp. 53–63, 2014, doi: 10.1016/j.system.2014.09.017.

[9] J. S. Lee and J. Chen Hsieh, “Affective variables and willingness to communicate of EFL learners in in-class, out-of-class, and digital contexts,” *System*, vol. 82, pp. 63–73, 2019, doi: 10.1016/j.system.2019.03.002.

[10] R. Clément, Z. Dörnyei, and K. A. Noels, “Motivation, Self-confidence, and Group Cohesion in the Foreign Language Classroom,” *Lang. Learn.*, vol. 44, no. 3, pp. 417–448, 1994, doi: 10.1111/j.1467-1770.1994.tb01113.x.

[11] J. Zuengler, “Encouraging Learners’ Conversational Participation: The Effect of Content Knowledge,” *Lang. Learn.*, vol. 43, no. 3, pp. 403–432, 1993, doi: 10.1111/j.1467-1770.1993.tb00620.x.

[12] E. Ozdemir and M. Papi, “Mindsets as sources of L2 speaking anxiety and self-confidence: the case of international teaching assistants in the U.S.,” *Innov. Lang. Learn. Teach.*, vol. 16, no. 3, pp. 234–248, 2022, doi: 10.1080/17501229.2021.1907750.

[13] Y. Cao and J. Philp, “Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction,” *System*, vol. 34, no. 4, pp. 480–493, 2006, doi: 10.1016/j.system.2006.05.002.

[14] T. Yashima, “Willingness to Communicate in a Second Language: The Japanese EFL Context,” *Mod. Lang. J.*, vol. 86, no. 1, pp. 54–66, 2002, doi: 10.1111/1540-4781.00136.

[15] T. Aoyama and T. Takahashi, “International Students’ Willingness to Communicate in English as a Second Language: The Effects of L2 Self-Confidence, Acculturation, and Motivational Types,” *J. Int. Students*, vol. 10, no. 3 SE-Research Articles, pp. 703–723, Aug. 2020, doi: 10.32674/jis.v10i3.730.

[16] J. Krajewski, A. Batliner, and S. Kessel, “Comparing Multiple Classifiers for Speech-Based Detection of Self-Confidence - A Pilot Study,” in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 3716–3719. doi: 10.1109/ICPR.2010.905.

[17] T. Dinkar, I. Vasilescu, C. Pelachaud, and C. Clavel, “How confident are you? Exploring the role of fillers in the automatic prediction of a speaker’s confidence,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 8104–8108. doi: 10.1109/ICASSP40776.2020.9054374.

[18] D. G. Cooper, K. Muldner, I. Arroyo, B. P. Woolf, and W. Burleson, “Ranking Feature Sets for Emotion Models Used in Classroom Based Intelligent Tutoring Systems,” in *User Modeling, Adaptation, and Personalization*, 2010, pp. 135–146. doi: 10.1007/978-3-642-13470-8_14.

[19] S. W. McQuiggan, B. W. Mott, and J. C. Lester, “Modeling self-efficacy in intelligent tutoring systems: An inductive approach,” *User Model. User-adapt. Interact.*, vol. 18, no. 1, pp. 81–123, 2008, doi: 10.1007/s11257-007-9040-y.

[20] S. Ishimaru, T. Maruichi, A. Dengel, and K. Kise, “Confidence-Aware Learning Assistant.” arXiv, 2021. Retrieved from https://arxiv.org/abs/2102.07312

[21] K. Forbes-Riley and D. J. Litman, “Analyzing Dependencies Between Student Certainness States and Tutor Responses in a Spoken Dialogue Corpus BT - Recent Trends in Discourse and Dialogue,” L. Dybkjær and W. Minker, Eds. Dordrecht: Springer Netherlands, 2008, pp. 275–304. doi: 10.1007/978-1-4020-6821-8_11.

[22] J. S. Shrauger and M. Schohn, “Self-Confidence in College Students: Conceptualization, Measurement, and Behavioral Implications,” *Assessment*, vol. 2, no. 3, pp. 255–278, Sep. 1995, doi: 10.1177/1073191195002003006.

[23] Y. Omori, M. Tamura, and D. Ikeda, “A prototype of an asynchronous interactive learning system for speaking practice,” in *Proceedings of 15th International Technology, Education and Development Conference*, 2021, pp. 4989–4998. doi: 10.21125/inted.2021.1026.

[24] Y. Omori, M. Tamura, and D. Ikeda, “A Field-Test of a Web Application for L2 Speaking Practice under an Uncontrolled Learning Environment for a Whole Semester,” in *Proceedings of the 6th International Conference on Education and Multimedia Technology*, 2022, pp. 97–103. doi: 10.1145/3551708.3551717.

[25] Voicepeak, The text-to-speech software developed by AHS Co.Ltd. and Dreamtonics Co., Ltd., Accessed 27 January, 2023. Retrieved from https://www.ah-soft.com/voice/6nare

[26] K. Takamatsu *et al.*, “Combining Key Performance Indicators and Signs of Changes in Students to improve the quality of education in Institutional Research based on Eduinformatics,” *IIAI Lett. Institutional Res.*, vol. 2, pp. 1–9, 2022, doi: 10.52731/lir.v002.052.

[27] J. Kruger and D. Dunning, “Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments.,” *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1121–1134, Dec. 1999, doi: 10.1037//0022-3514.77.6.1121.

[28] P. R. Clance and S. A. Imes, “The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention.,” *Psychother. Theory, Res. Pract.*, vol. 15, pp. 241–247, 1978, doi: 10.1037/h0086006.

[29] H. W. Seliger, “Utterance Planning and Correction Behavior: Its Function in the Grammar Construction Process for Second Language Learners,” in *Towards a cross-linguistic assessment of speech production*, H. W. Dechert and M. Raupach, Eds. Frankfurt: Lang, 1980, pp. 87–99.