Relieving Mental Stress of Speakers using a Tele-Operated Robot in Foreign Language Speech Education

Shizuka Nakamura¹, Miki Watanabe¹,², Yuichiro Yoshikawa¹,², Kohe Ogawa¹,², Hiroshi Ishiguro¹,²

¹ Osaka University, Japan
² JST ERATO, Japan

shizuka@lang.osaka-u.ac.jp, watanabe.miki@irl.sys.es.osaka-u.ac.jp,
yoshikawa@irl.sys.es.osaka-u.ac.jp, ogawa@irl.sys.es.osaka-u.ac.jp,
ishiguro@is.sys.es.osaka-u.ac.jp

Abstract

In an effort to relieve the mental stress experienced by speakers in foreign language speech education classes, an experiment using a tele-operated robot was conducted and its effect was evaluated. The robot was utilized in the following class environment: the teacher and classmates were not able to see the actual appearance of the speaker; however, the speaker could visually recognize the tele-operated robot speaking as its proxy, and was aware of the fact that the teacher and classmates also recognized the situation. The results of statistical analysis of subjective evaluation by the speakers to evaluate the effect on stress show that the degree of stress in speakers was lower when conversations with the teacher were conducted indirectly via the robot than doing so directly. A similar tendency is indicated by the results of subjective evaluation by the listeners in this experiment to imagine the case that they themselves are speakers. These results confirm that it is possible to relieve the mental stress experienced by speakers by utilizing a tele-operated robot.

Index Terms: speech education, foreign language, mental stress, tele-operated robot.

1. Introduction

Mental stress affects the acquisition of speaking skills. Generally, speakers under stress have a difficult time talking fluently. This difficulty is compounded in the case where the learners are studying a foreign language. This is because they are constantly under stress if they have no prior experience of speaking fluently. Learners tend to fall into a vicious cycle. In this way, in speech education classes in a foreign language, learners who are easily stressed experience difficulty from the outset. Many Japanese people are said to be such learners.

The typical foreign language class at a Japanese university consists of one teacher and about 35 students. In such a class, there are many occasions on which a student has a conversation with the teacher, in an environment where the classmates see and listen to the speaker in the same room. The purpose of speech education in a foreign language is to improve the speaker’s proficiency in the language. However, given the above-mentioned characteristic of many Japanese people, stress due to the presence of both the teacher and classmates in the same room becomes a serious obstacle to achieving this purpose. Therefore, it is desirable first and foremost to relieve the stress of a speaker.

To achieve this, we required that the teacher and classmates could not see the speaker as a condition of the class environment. The following class environment conditions were maintained: the speaker could visually recognize a tele-operated robot speaking as its proxy, and was aware that the teacher and classmates also recognized this situation. Thus, the speaker had a clear sense of being able to talk well with the teacher. These conditions allow the speaker to gain self-confidence, which leads to a virtuous cycle. They are also important as the scaffolding in social situations. A listener can also visually and audibly learn from the states and utterances of a speaker having a conversation with the teacher. Gaining this kind of experience is expected to help a speaker relieve stress [1]. Therefore, an environment where a listener can visually and audibly recognize a speaker may be desirable. Thus far, there have been some efforts in Japan to use distance learning as a method of speech education in a foreign language utilizing tele-communication technology such as TV-conference and PC-based video chat systems (e.g., [2]). However, they have a mechanism whereby a video image of the speaker is transmitted to the listener in real time. Therefore, they do not meet the environment conditions required in this study.

Humanoid robots have received attention as potential proxies of distant persons in tele-communication. This is because it is expected that a distant person can ostensibly convey nonverbal messages or their social presence through the body of the robot [3], [4]. Conveyance of nonverbal messages is considered to be one advantage of this style of tele-communication as compared with existing tele-communication media such as teleconference and PC-based video chat systems (e.g., [2]). However, they have a mechanism whereby a video image of the speaker is transmitted to the listener in real time. Therefore, they do not meet the environment conditions required in this study.

Humanoid robots have received attention as potential proxies of distant persons in tele-communication. This is because it is expected that a distant person can ostensibly convey nonverbal messages or their social presence through the body of the robot [3], [4]. Conveyance of nonverbal messages is considered to be one advantage of this style of tele-communication as compared with existing tele-communication media such as teleconference and PC-based video chat systems (e.g., [2]). However, they have a mechanism whereby a video image of the speaker is transmitted to the listener in real time. Therefore, they do not meet the environment conditions required in this study.

Humanoid robots have received attention as potential proxies of distant persons in tele-communication. This is because it is expected that a distant person can ostensibly convey nonverbal messages or their social presence through the body of the robot [3], [4]. Conveyance of nonverbal messages is considered to be one advantage of this style of tele-communication as compared with existing tele-communication media such as teleconference and PC-based video chat systems (e.g., [2]). However, they have a mechanism whereby a video image of the speaker is transmitted to the listener in real time. Therefore, they do not meet the environment conditions required in this study.
co-learners and teacher. Second, because the experience of communication with another person is visually recognized by novices as well as the classmates and teacher, the social self-image of the novice is expected to be biased toward that of a person who is able or motivated to communicate with others in English. Such a bias is conjectured from the assumption that the tele-operator of a robot tends to regard its social presence in the same light as its own (cf. [5]). Third, such opportunities to objectively observe the activities of one’s proxy to communicate are expected to help novices discover how their current ways of communication in terms of how they utter should be improved to match with their desired image (e.g., how their preferable actors talk in the film).

Given the difficulty of using such a robot in a speech education class in a foreign language, there are few previous studies of the same kind. As an example that introduced a tele-operated robot into education, one study showed the advantage of the robot in comparison to a video image in a TV-conference system in facilitating talking [6]. Because the subjects were limited to young children (4–8 years) and mental stress was not quantitatively evaluated, the possibility of using tele-operated robots for English education at a university remain unclear. Therefore, a tele-operated robot that satisfied the environment conditions required in the current study was introduced into an English speech education class for university students whose native language is Japanese. The English proficiency of the learners varied widely. The teacher had experience providing English speech instruction for native speakers of Japanese, as well as a knowledge of English phonetics.

2. Experiments

To verify the effect of relieving the stress of speakers by introducing a tele-operated robot, the learners were divided into the following three groups and conversation trials with the teacher were carried out: (i) learners who had a conversation with the teacher indirectly through the robot (hereafter, robot group), (ii) learners who had a conversation with the teacher directly (hereafter, human group), and (iii) learners who listened to a conversation of a classmate with the teacher (hereafter, listening group). To test the following hypothesis, the learners performed subjective evaluations after their conversation trials: the degree of stress experienced by speakers is lower in the case of having a conversation with a teacher indirectly through the robot than in the case of doing so directly. Furthermore, to quantitatively evaluate the influence of introducing the robot on speech, a portion of the acoustic features was analyzed.

2.1. Subjects

One English language teacher and 30 learners participated in the experiment. All learners were university students. Japanese was the native language for the learners as well as the teacher. The English proficiency of the learners varied widely. The teacher had experience providing English speech instruction for native speakers of Japanese, as well as a knowledge of English phonetics.

2.2. Experimental devices

The Telenoid was developed to connect people in a distant place by using tele-communication technology shown in Figure 1. It was designed to appear and behave like a neutral human; one can easily project the other person's presence whether the other is old or young, male or female. Owing to its characteristics, the Telenoid can be utilized for elderly care. The tele-operation system for the Telenoid allows people to feel as if the other person is actually in front of them. The system includes lip-sync [9], head-movement sync, and voice communication functions [3]. Therefore, it allows the operator to become accustomed to talking with others through the Telenoid.

In this experiment, learners (speakers) talked with the teacher through the Telenoid under the setting shown in Figure 1. The learner (speaker) and the Telenoid were in different
rooms to reproduce a real tele-operation situation. Video cameras and IC recorders were set up in each room to obtain video and voice recordings for subsequent analysis.

2.3. Procedures

First, the 30 learners were divided into the above-mentioned three groups ((i) robot group, n = 10; (ii) human group, n = 10; and (iii) listening group, n = 10) at random after the teacher asked the learners to prepare to answer four questions. The questions were on technologies that the learners found in daily life. Next, a speaker from the human group sat on a chair facing the teacher at a position where all learners could visually and audibly recognize in the classroom, and then had a conversation trial on the answers prepared in response to the questions. Subsequently, a speaker from the robot group performed a similar conversation trial with the teacher as shown in Figure 2. In this way, one speaker each from the human group and robot group, in alternation, had conversation trials with the teacher. This was repeated until all speakers finished. Each speaker finished with the phrase “thank you very much”. For the listening group, every listener saw and listened to all conversation trials. Finally, all learners answered the questionnaire on stress and enjoyment.

2.4. Measurements

To verify the effect of robot use on stress relief, a questionnaire survey was conducted as a subjective evaluation by the learners. Approximately ten different kinds of question were used depending on the group. For example, the following item was presented to the speakers in the robot and human groups: did you feel mental stress in the conversation trial that you experienced today? In contrast, the following item, for example, was presented to the listeners of the listening group: if you had a similar conversation trial with a teacher indirectly through the robot that you listened to today, do you think you would feel mental stress? The answer to each item was expressed on a 7-point Likert scale with scores ranging from 1 (do not think so at all) to 7 (definitely think so).

In the analysis to quantitatively evaluate the influence of introducing the robot on speech, we focused on duration as a basic acoustic feature. In this experimental setting, a speaker of the robot group was forced to speak while listening to his/her own voice echo. The actual value of this echo time was around 0.60–0.75 s. For this reason, we focused on the duration of the phrase “thank you very much” was measured for all speakers.

3. Results

3.1. Subjective evaluation of speakers

The speakers evaluated their own mental stress and enjoyment in the conversation trial. A Wilcoxon rank sum test revealed a significant difference between the robot group and the human group. Regarding mental stress in the trial, the speakers in the robot group felt less stress than those in the human group ($P < 0.05$, Wilcoxon rank sum test), as shown in Figure 2. Regarding enjoyment in the trial, the speakers in the robot group enjoyed the trial more than those in the human group ($P < 0.05$, Wilcoxon rank sum test; Figure 3).

3.2. Subjective evaluation of listeners

The listeners imagined how they would feel if they participated in the conversation trial in both conditions and
evaluated the mental stress in the trial. We found a significant difference between the robot group and the human group. The listeners imagined themselves to feel less stress in the robot group than in the human group ($P < 0.01$, Wilcoxon signed rank test; Figure 3).

### 3.3. Influence of introducing robot on speech rate

The average durations of the phrase “thank you very much” in the robot and human groups were 1.10 and 0.90 s, respectively. The corresponding standard deviations were 0.12 and 0.05. The difference in means was significant ($P < 0.001$, t-test). As shown in Figure 4, the duration in the robot group was significantly longer than that in the human group. That is, the speech rate in the former group was significantly lower than that in the latter group.

### 4. Conclusions and future work

To verify the effect of relieving the mental stress of speakers by introducing the tele-operated robot in speech education in a foreign language, a conversation trial with the teacher was conducted in an English language class for university students whose native language is Japanese.

Subjective evaluations by the learners revealed that the degree of stress in speakers is lower in the case of having a conversation with a teacher indirectly through the robot than in case of doing so directly. A similar tendency was indicated by the results of subjective evaluations by listeners who imagined the case that they themselves were speakers in the experiment. These results indicate the possibility of relieving the mental stress of speakers by introducing the tele-operated robot.

Quantitative analyses of the influence of introducing the robot on speech showed that the speech rate of the robot group was significantly lower than that of the human group. The tendency is likely to be influenced by the voice echo. In the future, to clarify the cause of this influence, it will be necessary to carry out an additional experiment by using an echo cancellation function, and an analysis of other acoustic features. Moreover, it will also be necessary to analyze whether the mental stress of speakers was relieved even in an environment where there might be a voice echo.

Apart from relieving mental stress, in the future, it will be important to conduct a long-term experiment on the following issue: how does the environment of speech education class in a foreign language by using the robot contribute to improving the proficiency of learners and making the scaffolding of social situations? It will be necessary to experimentally evaluate and ascertain the possibilities and limits of speech education in a foreign language by using the robot.

### 5. Acknowledgements

This study was supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (No. 25770148, 25220004 and 24680022) and ERATO Ishiguro Symbiotic Human-Robot Interaction Project.

### 6. References


