Development and practical evaluation of communication skill training teaching materials using gesture and speech recognition

Our research involved the development of a software-based gesture and speech recognition system which helps improve Japanese students' abilities to effectively communicate in English. It is difficult for Japanese students to get enough speaking practice and feedback in EFL classes, so we created a Virtual Interviewing System which focused on verbal communication. However, the importance of non-verbal cues, such as gestures, and body language in an English conversation is about 45-60%.

The new system has 3 components: facial- (FR), speech- (SR) and gestural recognition (GR), each giving important feedback for verbal and non-verbal communication. We built upon existing SR, and increased the accuracy for Japanese-accented speakers. With FR, our system was able to correctly identify and distinguish between identical twins. Finally, we improved upon the GR component by adding a Hidden Markov Model, which allowed the detection of key shapes made when gesturing, and increasing the accuracy.

Research field: Educational technology

Keywords: Nonverbal communication
1. 研究開始当初の背景
(1) English is very important as a global language in business and technology, but Japanese test-takers are scoring the lowest on proficiency exams such as The Test of English for International Communication (TOEIC) and the Test of English as a Foreign Language (TOEFL)\textsuperscript{3}. The Organization for Economic Co-operation and Development (OECD) reports that teacher-student time is an important variable in good and effective teaching\textsuperscript{4}. However, Japan’s large class sizes form an obstacle to the development of communicative competence\textsuperscript{5}: a 45-minute class with 40 students allows one minute of teacher-monitored speaking time each.

(2) Recent advances in technology and affordability will allow us to use a motion sensor camera to teach and assess NVC, thereby giving an opportunity to help students learn “total communication.”\textsuperscript{6} Users will perform better on interview tests such as EIKEN, TOEFL, and TOEIC, thereby increasing their communicative competence\textsuperscript{7}, and ultimately lead to Japanese academics and businesspersons being better-prepared for overseas business trips and presentations.

2. 研究の目的
(1) Our proposal was for the development of a software-based gesture and speech recognition system which will help improve Japanese students’ abilities to effectively communicate verbally and non-verbally in English. Our development and usage of the Virtual Interviewing System for assessing student’s language interviews has been successful in helping students gain speaking confidence\textsuperscript{8}. This was our first step in helping students develop all-around communicative ability.

(2) Research has placed the importance of non-verbal cues, such as gestures, tone of voice and body language in an English conversation at 60-70%, making non-verbal cues more important than words. Should our proposal be successful, our system will empower students to master both verbal and non-verbal communication by effectively creating a virtual one-to-one learning environment. Furthermore, both students and researchers will have access to valuable databases for assessment and analysis.

3. 研究の方法
Over the research period, students had their language interviews and presentations assessed automatically by the system for practice, and comprehensively by the teacher during curriculum-related testing. Students answered questions based on language necessary in various situations from an extensive list of categories: giving personal information, meeting people, going to restaurants, shopping, giving presentations, participating in meetings, and so on. Students were given feedback on their quality and quantity of gestures, and on their spoken answers. The students’ responses were tagged with the appropriate identifiers (year, class, individual student number) and sent automatically to a server from which the instructor will retrieve and score appropriately in accordance with a pre-set marking rubric (scoring guide).

4. 研究成果
(1) The system is designed around 3 components: facial recognition (FR), speech recognition (SR) and gestural recognition (GR). Each component helps give important feedback about the quality of both verbal and non-verbal communication.

(2) First, with the SR component, we were able to build upon existing SR platforms (CMU Sphinx developed by Carnegie Mellon University & JULIUS developed by Nagoya Institute of Technology), and increase the accuracy for speakers with Japanese-accented speech.

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese students</td>
<td>88%</td>
<td>84%</td>
<td>80%</td>
<td>88%</td>
<td>80%</td>
</tr>
<tr>
<td>No overseas study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese students</td>
<td>90%</td>
<td>80%</td>
<td>70%</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>With overseas study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Native Speaker</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 – phoneme modification results

Figure 1 shows the results of our modified SR software. In this experiment, the phoneme dictionary which contained the sounds to be identified were altered to detect the sounds made by a typical Japanese student [ie. Katakana English]. The Japanese students’ speech was similar to that of the English native speaker. In a parallel study we highlighted the fact that, while a machine might not correctly identify the words spoken, accented speech is not the same as poor pronunciation.

(3) With regards to the FR component, we improved the FR accuracy to compensate for
such variables as poor lighting, and distance from the camera. During our experiments, our FR system was able to correctly identify and distinguish between identical twins. Figure 2 shows the twins,

![Figure 2 – subjects (identical twins)](image)

and Figure 3 shows the scatter graph, with the grey data showing the differences between subjects and the black data showing the similarities.

![Figure 3 – data identifying the correct twin](image)

(4) Finally, we improved upon the GR component by adding a Hidden Markov Model, which allowed the detection of key shapes made when gesturing, and increased the gesture accuracy.

![Figure 4: Circular gesture made with 1-hand](image)

Figure 4 shows a typical circular gesture made with one hand to denote a shape when saying a word like, round or globe. Figure 5 shows how the Hidden Markov Model processes the gesture

![Figure 5 – Hidden Markov Model process](image)

In this research project, our goal was to develop a system where students could get feedback on both their verbal and non-verbal communication. As a result of our research, we were able to create and develop each of the 3 components: SR, FR and GR.

<引用文献>


Cooper, Todd, and Tsukada, Akira. "Using Technology to Teach 'Total Communication'". JALT 2013, 39th Annual International Conference on Language Teaching and Learning & Educational Materials, Kobe Convention Center Portopia (Hyogo Prefecture, Kobe City), 2013.

Cooper, Todd, Tsukada, Akira, and Matoba, Ryuichi. "Teaching 'Total Communication' through Technology: Creation of Software to Develop Verbal and Nonverbal Communicative Skills in English". COCET, Kobe, 2013.

Cooper, Todd, Tsukada, Akira, and Yamaguchi, Akihiko. "Teaching 'Total Communication' through Technology". COCET, Kobe, 2013.

Cooper, Todd, Tsukada, Akira, Naruse, Yoshinori, and Matoba, Ryuichi. "Teaching 'Total Communication' through Technology: Creation of Software to Develop Verbal and Nonverbal Communicative Skills in English". COCET, Kobe, 2013.