様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　２８　日現在

機関番号：　５３２０３
研究種目：　基盤研究(C)（一般）
研究期間：　2013～2016
課題番号：　２５３７０６８１

研究課題名（和文）ジェスチャ及び音声認識を用いたコミュニケーション能力育成教材の開発と実践評価

研究課題名（英文）Development and practical evaluation of communication skill training teaching materials using gesture and speech recognition

研究代表者

ＣＯＯＰＥＲ　Ｔ・Ｄ（COOPER, T.D.）

富山高等専門学校・一般教養科・准教授

研究者番号：７０４４２４４９

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：日本のクラスサイズではスピーキング練習に時間を割けないことから，我々は仮想的にインタビューを行うVirtual Interviewing System (VIS)を開発してきた．しかし，英語コミュニケーションではジェスチャ等の非言語による情報伝達が45-60%を占める．
当該システムは，非言語を含めた英語コミュニケーション能力向上を考慮して，顔，音声，ジェスチャ認識で構成される．音声認識部では，既存の音声認識エンジンを用い，日本人のアクセントに対する精度向上を図った．顔認識では双子を判別することが可能になった．ジェスチャ認識では隠れマルコフモデルを利用し定型のジェスチャ認識を行った．

研究成果の概要（英文）：Our research involved the developing a software-based gesture and speech recognition system which helps improve Japanese students' abilities to effectively communicate in English. It's 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.

研究分野：　educational technology

キーワード：　nonverbal communication

## １．研究開始当初の背景

(1) English is very important as a global language in business[1] 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)[2]. The Organization for Economic Co-operation and Development (OECD) reports that teacher-student time is an important variable in good and effective teaching[3]. However, Japan's large class sizes form an obstacle to the development of communicative competence[4]: 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."[5] Users will perform better on interview tests such as EIKEN, TOEFL, and TOEIC, thereby increasing their communicative competence[6], 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[7]. 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.

| | Q1 | Q2 | Q3 | Q4 | Q5 |
|---|---|---|---|---|---|
| Japanese students No overseas study | 88% | 84% | 80% | 88% | 80% |
| Japanese students With overseas study | 90% | 80% | 70% | 80% | 90% |
| English Native Speaker | 80% | 80% | 80% | 100% | 100% |

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)

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

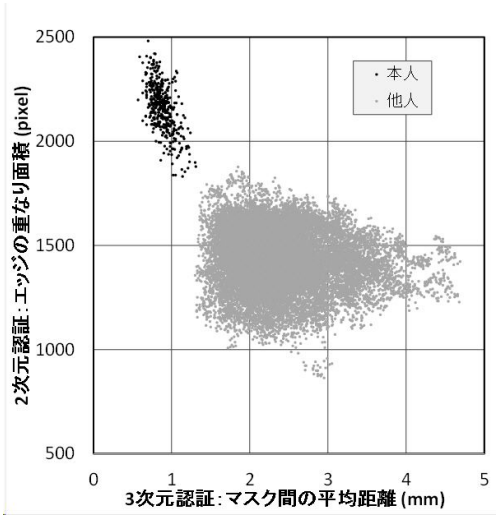


Figure 3 – data identifying the correct twin

(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

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

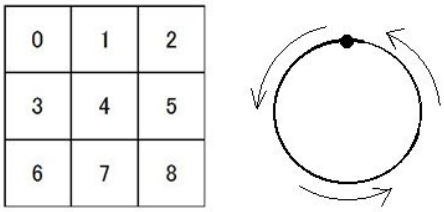


Figure 5 – Hidden Markov Model process

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＞


1. Clark, D. “English –The Language of Global Business?” Forbes.com, Oct. 26, 2012. Retrieved from http://www.forbes.com/sites/dorieclark/2012/10/26
2. ETS (Educational Testing Service). (2011). TOEFL Test and Score Data Summary for TOEFL Internet-based and Paper-based Tests: 2010 Data. ETS Research Report # TOEFL-SUM-10 2011, Retrieved from http://www.ets.org/research/policy_research_reports/toefl-sum-10
3. OECD. (2009). OECD Education at a Glance 2009. Retrieved from www.oecd.org/edu/eag2009
4. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1, 1-47.
5. Cooper, T., Tsukada, A., Yamaguchi, A. & Naruse, Y. (2011). Creating a Virtual Interview: Using technology to improve language and interview skills. In M. Koehler & P. Mishra (Eds.), Proceedings of SITE Conference 2011 (pp. 3165-3166). Chesapeake, VA: AACE.
6. Powers, D., Kim, H.-J., Yu, F., Weng, V., & VanWinkle, W. (2009). The TOEIC® Speaking and Writing Tests: Relations to Test-Taker Perceptions of Proficiency in English. ETS Research Report. Retrieved from http://www.ets.org/Media/Research/pdf/RR-09-18.pdf
7. Lebedev, J. “TESOL 2012 Highlights: Day 2, Part 1” English with Jennifer: a blog for teachers, April 2, 2012. Retrieved from http://englishwithjennifer.wordpress.com/2012/04/02/tesol-2012-highlights-day-2-part-1

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　　5 件）

① T. Matsunaga, T. Cooper, A. Tsukada., Quantifying the Quality of Facial Expressions., Proceedings of Society for Information Technology & Teacher Education International Conference 2016, 査読有, 2016, pp.962-964, DOI 978-1-939797-13-1

② H. Sudo, R. Matoba, C. Todd, A. Tsukada., Effect of the Symmetry Bias on Linguistic Evolution., Artificial Life and Robotics, 査読有, 2016, pp.207-214, DOI 10.1007/s10015-016-0276-7

③ 塚田 章, Cooper Todd., 距離画像カメラを用いた 2D3D 顔認証システム., 自動認識, 査読有, Vol.27, No.5, 2014, pp.30-35

④ T. Cooper, A. Tsukada, Y. Naruse, R. Matoba., Accent versus Mispronunciation: modifying speech recognition software to recognize speech and not a perfect accent in the ESL context., Proceedings of Society for Information Technology & Teacher Education International Conference 2014, 査読有, 2014, pp.1096-1099

⑤ 今枝 駿, Cooper Todd, 塚田 章., モーションセンサデバイスから得られる深度情報を用いた顔認証の基礎的検討討., 電子情報通信学会技術研究報告, 査読有, Vol. 113, No. 61, MBE2013-5、2013, pp.19-24

〔学会発表〕（計　　19 件）

① Ryuichi Matoba, Yusaku Funane, Todd Cooper., Automatic Construction of Emotion Corpus with Emotion Estimation by Naïve Bayes., 22nd International Symposium on Artificial Life and Robotics, 査読有, B-Con Plaza (Oita Prefecture, Beppu City), 2017 年 1 月 19 日－21 日

② 澁谷 和帆, Cooper Todd, 塚田 章。, マウス代用装置における眼球位置座標を用いた頷き検出., 平成 28 年度電気関係学会北陸支部連合大会, 福井工業大学(福井県福井市), 2016 年 9 月 13 日－14 日

③ COOPER Todd, Dowden Andrew Robert., Collaborate, Create and Communicate: Teaching Presentation Skills in Class., 平成 28 年度ＣＯＣＥＴ研究大会, 査読有, 国立オリンピック記念青少年総合センター(東京都渋谷区代々木), 2016 年 9 月 03 日－04 日

④ T. Cooper, A. Tsukada, T. Matsunaga., Pronounce Perfectly or Communicate Productively: Modifying Speech Recognition Technology for EFL students., EdMedia 2016: World Conference on Educational Media and Technology, 査読有, Sheraton Vancouver Wall Centre Hotel (Vancouver, British Columbia, Canada), 2016 年 6 月 28 日－30 日

⑤ Dowden Andrew Robert, COOPER Todd., Presentation's 3 Cs: Cooperate, Collaborate, Communicate., 15th JALT PanSIG International Conference, Innovations in Education, 査読有, Meio University (Okinawa Prefecture, Nago City), 2016 年 5 月 20 日－23 日

⑥ 竹中 宏兵, 今枝　駿, Cooper Todd, 塚田 章., 顔認証システムにおける Kinectv2 と Xtion の精度比較., 平成 27 年度電気関係学会北陸支部連合大会, 金沢工業大学（石川県金沢市), 2015 年 9 月 12 日－13 日

⑦ 澁谷 和帆, Cooper Todd, 塚田 章., Kinect v2 を用いた顔の Roll 推定, 平成 27 年度電気関係学会北陸支部連合大会, 金沢工業大学（石川県金沢市), 2015 年 9 月 12 日－13 日

⑧ 松永 友樹, Cooper Todd, 塚田 章., バーチャルインタビューシステムにおける顔表情豊かさの評価について, 平成 27 年度電気関係学会北陸支部連合大会, 金沢工業大学（石川県金沢市), 2015 年 9 月 12 日－13 日

⑨ COOPER Todd, TSUKADA Akira, NARUSE Yoshinori, MATSUNAGA, Tomoki., Teaching Perfect Pronunciation or Productive Communication: utilizing SR technology in the Kosen EFL classroom., 平成 26 年度ＣＯＣＥＴ研究大会、査読有, 国立オリンピック記念青少年総合センター(東京都渋谷区代々木), 2014 年 9 月 13 日－14 日

⑩ 今枝 駿, Cooper Todd, 塚田 章., Xtion より得られる正面顔を用いた 2D 顔認証., 平成 26 年度電気関係学会北陸支部連合大会, 富山高専（富山県富山市), 2014 年 9 月 11 日－12 日

⑪ 若松 絵里, 坂田 結衣, 由井 四海, 塚田章., Kinect における RGB 画像と距離画像の位置補正., 平成 26 年度電気関係学会北陸支部連合大会講演論文集。富山高専（富山県富山市), 2014 年 9 月 11 日－12 日

⑫ COOPER Todd, TSUKADA Akira, NARUSE Yoshinori, MATOBA Ryuichi., Big Brother is Watching You [get better at EFL interviews and presentations]! Harnessing facial, speech and gestural recognition for education., EdMedia 2014: World Conference on Educational Media and Technology 2014, 査読有, Tampere University (Tampere, Finland), 2014 年 6 月 23 日－26 日

⑬ 今枝 駿, Cooper Todd, 塚田 章., Xtion を用いた 2D3D 顔認証システムについて., 第 20 回画像センシングシンポジウム講演論文, パシフィコ横浜アネックスホール（神奈川県横浜市), 2014 年 6 月 11 日～13 日

⑭ 今枝 駿, Cooper Todd, 塚田 章., Kinect を用いた 3D 顔認証., 第 3 回バイオメトリクスと認識・認証シンポジウム, 日本科学未来館（東京都江東区青海), 2013 年 11 月 26 日－27 日

⑮ Todd Cooper, Akira Tsukada, Ryuichi Matoba, Yoshinori Naruse., Using Tech to Teach "Total Communication., JALT 2013 39th Annual International Conference on Language Teaching and Learning & Educational Materials, 査読有, Kobe Convention Center Portopia (Hyogo Prefecture, Kobe City), 2013 年 10 月 25 日－28 日

⑯ 今枝 駿, Cooper Todd, 塚田 章., Xtion を用いた 3D 顔認証システムの基礎的検討., 平成 25 年度電気関係学会北陸支部連合大会講演論文集, 金沢大学（石川県金沢市), 2013 年 9 月 21 日－22 日

⑰ T. Cooper, A. Tsukada, Y. Naruse、R. Matoba., Teaching “Total Communication” through Technology: creation of software to develop verbal and non-verbal communicative skills in English., 平成 25 年度 COCET 研究大会, 査読有, 京都府中小企業会館（京都府、京都市), 2013 年 9 月 21 日－22 日

⑱ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse., スピーキング能力向上のためのバーチャルインタビューシステム教材の開発., 平成 25 年度全国高専教育フォーラム。教育研究発表会, 査読有, 国立大学法人豊橋技術科学大学（愛知県豊橋市), 2013 年 8 月 21 日－23 日

⑲ 今枝 駿, Cooper Todd, 塚田 章., Kinect を用いた 3D 顔認証の基礎的検討., 第 19 回画像センシングシンポジウム講演論文集, パシフィコ横浜アネックスホール（神奈川県横浜市), 2013 年 6 月 12 日－14 日

６．研究組織

(1)研究代表者
COOPER　T・D（COOPER, Todd）
国立富山高等専門学校・一般教養科・准教授
研究者番号：70442449

(2)研究分担者
塚田　章（TSUKADA, Akira）
国立富山高等専門学校・電子情報工学・教授
研究者番号： 40236849

(3)研究分担者
的場　隆一（MATOBA, Ryuichi）
国立富山高等専門学校・電子情報工学・准教授
研究者番号：　30592323

(4)研究分担者
成瀬　喜則（NARUSE, Yoshinori）
富山大学・教職実践開発研究科・教授
研究者番号：　00249773