

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 6日現在

機関番号：53203

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22520612

研究課題名（和文） スピーキング能力向上のための画像音声認識を用いたバーチャルインタビュー教材の開発

研究課題名（英文） Development of a Virtual Interview System with Facial and Voice recognition to Improve English Speaking Ability.

研究代表者

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

研究成果の概要（和文）：外国語教育において、教員と学生の対話時間は教育効果に非常に重要な要素である。大人数のクラスに対する教育に IT 技術が効果的である。本報告では、学生の面接テストを評価するための Virtual Interviewing System (VIS) の開発と実践について述べる。本研究の目的は、学生の将来を決定する就職面接やスピーキングテストのために、学生の実践を支援し英語コミュニケーションスキルを高めることである。

研究成果の概要（英文）：Teacher-student time is an important variable in effective education for the EFL classroom. Technology is an effective way to combat large class sizes. In this report, we would like to discuss our development and usage of our Virtual Interviewing System (VIS) for assessing student's language interviews. The purpose of this research is to help students practice and enhance their communicative skills for that future job interview or speaking test that determines their academic future.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	2,100,000	630,000	2,730,000
2011年度	700,000	210,000	910,000
2012年度	500,000	150,000	650,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：Virtual Interview, 顔認証, 音声認識, スピーキング

1. 研究開始当初の背景

外国語教育において、教員と学生の対話時間は教育効果に非常に重要な要素である。しかしながら、我が国では1クラスが40人程度の学生で構成されるため、個々の学生に対してスピーキング練習に費やすことのできる時間はごく僅かである。すなわち、これまでの英語教育ではリーディング、ライティング、文法が主体でありスピーキング練習があ

まりなされてこなかった。現在、我が国の TOEFL iBT のスピーキングセクションの結果は110カ国中最下位である。

海外留学、入学試験、ビジネス界では昇任条件として、最近注目度が増している TOEIC/TOEFL あるいは英検等の各種試験においてインタビュー形式のテストが課せられている。これは実際に海外に行って活躍するためにはコミュニケーション能力が必

須要件であることを示している。このように、スピーキング能力あるいはコミュニケーション能力に対する教育は急務であるにもかかわらず、クラスを対象とした従来の授業形態ではそれを実現することは困難な状況にある。

また、本校では海外の機関と提携し独自の留学制度を構築している。当初、提携機関では学生が到着する前にカリキュラムを準備するため、当該学生の言語能力を評価するための信頼性の高い方法を必要としていた。まず、エッセイやオンライン多肢選択アセスメントテストを使用した。これらはコミュニケーションレベルの評価には十分ではなかった。そこで、我々は学生のビデオインタビューをアップロードし、これらのテストを置き換えた。その結果、学生のコミュニケーションスキルがより適切に評価され、短い期間内に学習を最大限に引き出すことができた。

2. 研究の目的

コミュニケーション能力（ヒアリングとスピーキング）の向上には教員と学生が一对一で向き合うことが理想ではあるが、上述の背景から我々はIT技術を用い、仮想的に一对一の学習環境を作りスピーキングの機会をできるだけ多く与えるために **Virtual Interview System(VIS)**を開発することを目的とする。

3. 研究の方法

(1)基本システム

VISは、10分程度の時間でクラス全員に同時にインタビューテストを行えるよう、webカメラ等で取得した画像/音声をサーバに記録するシステムを構築する。具体的には、

- ① 各学生はwebカメラとヘッドセットが取り付けられたPCを使用し、これらをクライアントとするサーバを用意する。
- ② あらかじめ教員が質問を記録した動画を多数用意しサーバに置く。
- ③ 質問動画をランダムに学生に提示する。これに対して学生は決められた時間内に回答する。回答を記録した動画をサーバに転送する。
- ④ その後、教員がサーバからクラス全員の回答動画を再生し採点を行う。

(2)音声自動採点機能

音声処理技術を用い、制限時間内の発声単語数、文章数のカウント、リズム、ピッチ、イントネーション、発音等の抽出とネイティブとの比較評価、正誤問題、選択問題、穴埋め問題等の基本問題を自動採点できる機能を構築する。

(3)顔画像処理

画像処理技術を用い、ログイン時と回答中の本人確認のための個人認証、回答中の顔向き、表情等インタビューの評価に大きな影響

を及ぼす非言語コミュニケーションを定量的に評価する機能を構築する。

(4)VISの評価実験

本校国際流通学科の授業で試用実験を行い、より使いやすいシステムへの改良を行う。また、VISの教育効果について検討する。

4. 研究成果

(1)基本システムとその評価

VISは、Webカメラとヘッドセットを搭載したクライアントコンピュータに接続されたネットワークサーバで構成した。プログラミングはon RailsのADOBE FLASHとRubyを用いた。プログラム自体はYouTubeなどの人気ストリーミングビデオサイトのように動作する。インタビューでは、学生は教員によって与えられる質問動画を見た後、難易度に応じて決められた時間内に回答を行う

(図1)。回答は自動的に圧縮記録し、サーバにアップロードされる。試用実験ではクラス全員のファイルが同時に転送できることを確認した。図2は授業風景を示す。教員は管理者のセクションに移動し、グレード、彼らが選択した基準に5点満点で採点できるようになっている。学生は、記録のセクションにログインし、自分の回答動画や採点結果を閲覧することができる。

研究期間内で、ユーザインタフェースに多くの改良を加えてきた。これらの改良は、システムの有効性と使いやすさを高めるため



図1 ベータ2のスクリーンショット



図2 授業風景

に役立てた。図 3, 図 4 は, これまでの 3 つのベータ版のそれぞれの後にとったアンケート結果を示す。ベータ 1 (68 名の学生に調査) はクライアントベース, ベータ 2 (142 名の学生に調査), ベータ 3 (86 名の学生に調査) はブラウザベースである。結果は改良を加える毎に積極的になってきた。また, 本システムをブラウザベースで構築したことにより, 場所や OS を選ばず実行できることも特徴の一つである。

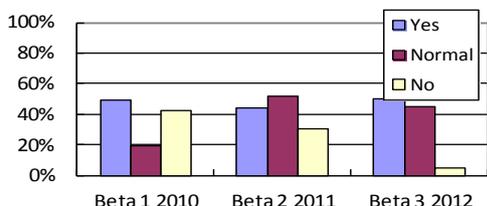


図 3 このシステムは使いやすかったか?

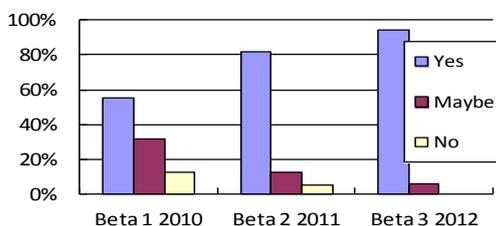


図 4 英語スピーチの練習に効果的か?

(2) 音声自動採点機能

音声認識に関してはロゼッタストーンやニュアンスのドラゴンスピーチなど, 多くの優れた製品が存在する。しかしながら, これらは必ずしも大規模な学術利用に向けてつくられているわけではない。これらの既存のシステムは高価であるか柔軟性に欠け, 本システムにそのまま適用できない。さらに, 生成されたオーディオファイルやビデオファイルを簡単に抽出し処理することが困難である。

当初, フリーソフトウェアである Microsoft の英語認識エンジンとジュリアスの使用を試みたが, 日本人が発声する英語では誤認識が多く発生し使用を断念した。本システムでは, 学生が標準的なアメリカや英語のアクセントを模倣する (正しく発音する) が目的ではない。多少発音は正しくなくても言いたいことが伝えられるためのスピーキングの練習を目的としている。このためには, 認識に用いる音響モデルを日本人用 (非ネイティブスピーカー) にカスタマイズすることが必要と判断した。

カーネギーメロン大学では音声認識エンジン Sphinx を開発しているが, 研究代表者は同大学の研究者と連携し, 日本人のアクセントに適合した音響モデルを開発するためのデータベースの構築に着手した。これと並行

して, 選択問題を英語の発声で解答し, これを Sphinx により自動採点する機能を開発した。現状では, ネイティブスピーカーによる実験では 90% 程度の認識率が得られるが, 日本人では 70-80% の認識率である。

(3) 顔画像処理

ログイン用顔認証システムに関して, 家庭用 DV カメラと蛍光灯照明を用いて基本的なプログラムを作成し, 10 人の被験者が 2 ヶ月間ほぼ毎日認証する予備実験を行った。10 枚の登録画像に対して 1 枚の画像で認証した場合 97%, 3 枚の画像では 100% と満足できる認証率が得られた。しかしながら, 実際の使用を想定し安価な web カメラと LED を用いた簡易照明を用いてクラス 40 人に対して実験を行い 95% と認証率が低下した。そこで, 精度向上を図りゲーム用モーションキャプチャデバイスである Kinect を使用して深度情報に基づくマッチング法を開発した。

一般的に, RGB 画像を用いる 2 次元 (2D) 顔認証は, 虹彩, 静脈, 指紋を用いた認証に比べ認証精度が低いという知見が存在する。これは顔向きや照明方向の違いによって特徴点の位置関係が変わってしまうことに起因する。この課題を解決するため多数の条件変動画像を考慮する等の研究が行われているが計算コストは増大する。

これに対して, 距離画像を用いる 3 次元 (3D) 顔認証は認証率の最も高い虹彩認証に近い認証精度が得られる。3D 顔認証の最大の課題は計測システムが非常に高価なことである。Kinect は安価に深度情報を取得できることが大きな特徴であるが, 320×240 の赤外線画像から簡易的に深度情報を算出するため, 認証精度の低下が予想される。さらに簡易計測のため顔向きによって鼻側面の情報が自己遮蔽によって欠落する部分が生じる。

そこで, 本システムではほぼ正面を向いた顔画像のみを選択的に取得し, Kinect からの距離で正規化した顔マスクを照合に用いることで, 情報の欠落を防ぎ, 顔向きやサイズに対する条件変動の処理軽減を図った。また, 深度情報を用いることは, 従来課題であった顔検出, 照明方向による明るさの違いの考慮, 顔特徴点の抽出等の処理軽減が期待できる。すべてのフレームに対して鼻頭の検出を試みる。ある距離範囲に鼻頭が検出でき, 推定される顔向きがほぼ正面と判断されたら, 顔領域の抽出, 正規化を行い顔マスクを作成する。顔マスクの照合に用いる指標は 2 枚の顔マスクの平均距離とする。平均距離算出時に単純な傾き補正とサブピクセル単位の位置補正を行う。

以上の処理手順で 24 人の被験者に対して, 5 回の撮影で得た 100 セットの顔マスクを

総当たり(7140組)で比較した. 図5は平均距離のヒストグラムを示す. 本人の組み合わせは黒色, 他人の組み合わせは灰色で示す. 図において, 本人の平均距離は, 0.59 mm ~ 1.37 mm(平均 0.953 mm)の間に分布し, 他人の平均距離は 1.30 mm ~ 11.17 mm(平均 3.17 mm)の間に分布した. 両者の重なりは4組であり, 1枚の登録データと1枚の認証データでほぼ認証が可能であることを示している. 本結果より, これまでのエッジを用いる方法に比べ認識率が格段に向上することが確認できた. さらに, 深度情報を用いて顔向きの推定が容易に行うことができ, 頬や口角の詳細な情報が検出できれば, 顔表情の判定も可能である.

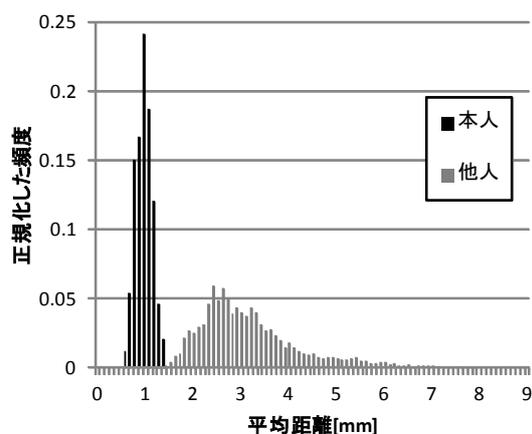


図5 平均距離のヒストグラム

(4) 今後の展望

顔認証に関する研究において, 我々は Microsoft の Kinect や Asus の Xtion Pro などの多次元モーションセンサーカメラを用いることで格段に精度を高めることができた. モーションセンサーカメラはもともとゲーム用に骨格検出が可能であるため, ジェスチャを検出することが可能になる. すなわち, ジェスチャに代表される非言語的コミュニケーションスキルの教育を支援するシステムが開発できると考えている.

非言語コミュニケーションに関しては "7%-38%-55%ルール"が有名である. このルールは感情や態度を相手に伝える, いわゆるコミュニケーションの7%は話し言葉, 38%は声のトーンから, そして55%がボディラングージで構成されるというものである. 一般的に, ほとんどの研究者は非言語コミュニケーションの重要性を50~65%に見積もっている.

モーションセンサーカメラを使用することにより, VIS を非言語コミュニケーションを含めた概念に拡張することができ, ひいてはコミュニケーション全体を向上させるシステムの開発が可能になる.

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4件)

- ① T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, What about Non-verbal Communication? Teaching Total Communication in EFL Lessons, Proceedings of Society for Information Technology & Teacher Education International Conference, 査読有, 2013, pp. 2054-2058
- ② T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, From Interviews to Total Communication: Virtualizing the English Foreign Language Classroom, Proceedings of Society for Information Technology & Teacher Education International Conference, 査読有, 2013, pp. 336-342
- ③ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Tackling Large Class Sizes with the Virtual Interview: Creating One-to-One learning with Technology, The Newsletter of the Speech, Pronunciation, and Listening Interest Section, (on line), 査読有, 2011
<http://newsmanager.commpartners.com/tesol splis/issues/2011-11-15/email.html>
- ④ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Creating a Virtual Interview: Using technology to improve language and interview skills, Proceedings of Society for Information Technology & Teacher Education International Conference 2011, 査読有, 2011, pp. 3165-3166

[学会発表] (計11件)

- ① 今枝 駿, Cooper Todd, 塚田 章, Kinect を用いた 3D 顔認証の基礎的検討, 画像センシングシンポジウム, 2013.6.13, 神奈川 (パシフィコ横浜アネックスホール)
- ② 今枝 駿, Cooper Todd, 塚田 章, モーションセンサー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, 電子情報通信学会 ME とバイオサイバネティクス研究会, 2013.5.24, 富山 (富山県立大学)
- ③ 今枝 駿, 酒匂祐太郎, Cooper Todd, 塚田 章, Kinect センサを用いた 3次元顔認証システムの検討, 平成24年度電気関係学会北陸支部連合大会, 2012.9.1, 富山 (富山県立大学)
- ④ T. Cooper, A. Tsukada, e-Gestures: Developing Educational Software for Learning Body Language, TESOL 2012,

- 2012.3.29, Pennsylvania Convention Center (Philadelphia, USA)
- ⑤ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Virtualizing Total Communication: Speech and Presentation Software Development Using the Kinect and the Virtual Interviewing System, 2012.3.29, Pennsylvania Convention Center (Philadelphia, USA)
- ⑥ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Kinecting to your Listeners: Development of Body Gesture Recognition Software for Language Learning, 2012.3.30, Pennsylvania Convention Center (Philadelphia, USA)
- ⑦ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Development of a Browser-Based Virtual Interview System: Enhancing communication for interviews and standardized tests, Grace Hopper Celebration 2011, 2011.11.10, Oregon Convention Center (Portland, USA)
- ⑧ M. Yago, A. Tsukada, T. Cooper, Applying a Facial Recognition Login Module to a Virtual Interview System, Grace Hopper Celebration 2011, 2011.11.10, Oregon Convention Center (Portland, USA)
- ⑨ T. Cooper, A. Tsukada, A. Yamaguchi, Development of Gesture Learning Educational Software to Improve English Communication, The Council of College English Teachers, 35th Annual Research Conference, 2011.9.3, 京都テルサ (京都府)
- ⑩ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Virtual Interviewing: Using Technology to Enhance Communicative Skills, TESOL 2011, CALL-Interest Section, Electronic Village Fair, 2011.3.17, Ernest N. Memorial Convention Center (New Orleans, USA)
- ⑪ T. Cooper, A. Tsukada, A. Yamaguchi, Y. Naruse, Creating a Virtual Interview System for Self-Study and Standardized Test Preparation, TESOL 2011, CALL-Interest Section, Electronic Village Fair, 2011.3.18, Ernest N. Memorial Convention Center (New Orleans, USA)

6. 研究組織

(1) 研究代表者

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

(2) 研究分担者

塚田 章 (TSUKADA AKIRA)

富山高等専門学校・電子情報工学科・教授

研究者番号：40236849

山口 晃史 (YAMAGUCHI AKIFUMI)

富山高等専門学校・電子情報工学科・准教授

研究者番号：40239877

成瀬 喜則 (NARUSE YOSHINORI)

富山高等専門学校・国際ビジネス学科・教授

研究者番号：00249773