

京都大学	博士 (工学)	氏名	于 肆 洋
論文題目	Utilization of Visual Sensing and Face Analysis for Enhancing E-Learning (画像センシングと顔画像解析を利用した e-ラーニングの機能増強)		
<p>(論文内容の要旨)</p> <p>本論文は、e-Learning の高度化のために画像解析を用い、学習中の受講者の状態を認識して教員にフィードバックする機能、および、語学学習における発音習得のために 3 次元的な口形状を提示する機能を実現する研究をまとめたものである。本論文は以下の 5 章からなっている。</p> <p>第一章は序論であり、e-Learning の現状や特性について基礎的な議論を進め、学習者の状態を教師へフィードバックすることが e-Learning の高度化のための重要な課題となっていることを従来の教育理論などを基に示している。さらに、学習者への教材提示やフィードバックに関して語学学習 e-Learning の発音訓練をとりあげ、画像解析を用いた発音方法の提示を課題とすることを述べている。</p> <p>第二章では、e-Learning 中の学習者を画像により観測し、学習者の内部状態を推定して教師にフィードバックする枠組みについて述べている。この手法では、学習者の学習環境 (PC その他) の周辺にカメラを設置し、それによって撮影された映像から頭部の姿勢や動き、表情、注目箇所などを自動計測して、学習者の外部状態として記録する。この外部状態と学習者の内省によって得られた内部状態の対応データを訓練データとして機械学習を行うことにより、学習者の観測データから内部状態を推定するための識別器を得る。学習者の内部状態としては、従来から教育工学において重要視されてきた、学習者の集中度 (concentration-distraction)、学習者にとっての教材の難しさ (difficulty-ease)、学習者の興味 (interest-boredom) の 3 つの内部状態を設定しており、それぞれについて内省データを収集し、SVM (Support Vector Machine) を用いて識別器を構成する。この手法を京都大学の CALL (Computer Assisted Language Learning) システムに適用し、英語教材を学習する 7 人の学生に対して 3000 サンプル (1500 分) を超える学習状況のデータを収集して検証実験を行った。得られたデータを用いて内部状態の識別器を構成した結果、同一被験者内では厳密なマッチング基準で 60% 程度、幅を持たせたマッチング基準で 90% 程度の精度を得ている。これにより、十分なデータが得られれば、画像解析により学習者の内部状態を一定の精度で推定することが可能であることを実証した。また、学習者の振る舞いには個人差があり、他人の類似した外部状態に対して異なる内部状態が対応しうることを確認するとともに、他人のデータで学習させた識別器を適用した場合には精度が劣ることを確認している。</p> <p>第三章では、e-Learning に第二章で述べられた方法を適用するために重要となる、受講者の内部状態に対する内省データ収集の問題を軽減する方法について述べている。第二章の手法を適用するためには訓練データを集める必要があるが、その際に学習者の多くの手間が必要となるため、全ての学習者にその負担をかけることは望ましくない。また、第二章で述べられているように、学習中の振る舞いには個人差があり、全員のデータを集めたものを学習させた識別器 (unified classifier と呼ぶ) では良</p>			

京都大学	博士 (工学)	氏名	于 肆 洋
<p>い精度が得られない．そこで，少数の学習者 (prototype learner) から集中的に集めたデータを用いて内部状態の識別器を複数個構成しておき，その中から新しい学習者に適したものを選ぶ方策を提案している．その方法として，まず，ランダムに少数のサンプルを選び，それに対する識別精度が良い識別器を選ぶ手法を提案し，それによって unified classifier よりも良い精度が得られることを示している．さらに，頻出値を用いる手法も検討している．この手法では，頻出値を選ぶために，サンプルの生起確率分布を推定し，生起確率が大きなサンプルに対してクラスタリングを行うことによって，代表サンプルを得ている．このようにして選ばれた代表サンプルに対して識別精度が良い識別器を選ぶことで，ランダムな選択手法と同等以上の精度の向上が得られることを報告している．さらに，ランダムな選択手法では，精度が悪い識別器が偶然選ばれる可能性を排除できないのに対して，この手法では一意に良い結果が得られる利点がある．また，代表サンプルの傾向と当該学習者のサンプル全体との傾向が似ている (相関が高い) 場合には良い精度が得られる確率が高いことが示されている．これにより，提案する手法によって，新しい学習者の負担を少なくとどめながら，unified classifier よりも精度を上げるができることを実証した．</p> <p>第四章では，語学教育の e-Learning において，画像センシングを用いて発音学習の支援を提供する方法について述べている．正しく発音するためには，口周りの筋肉や舌を正しく動かす必要があるが，従来は発音方法を図示したり，映像で提示したりすることが行われてきた．しかし，唇を突き出す動作 (lip protrusion) などをわかりやすく伝えることが難しかった．そのため，この研究では顔形状を 3 次元計測し，口の周りの形状がわかりやすいように擬似カラー等で強調して提示する手法を提案している．このようにして教師側の手本を学習者へ提示すれば，発音方法の良い教材となるが，それだけでなく，学習者が自分の発音方法を自習したり，教師が学習者の発音方法を評価・指導したりするための良い手段ともなる．実験では，中国語を学習する京都大学の 40 人以上の学生に対し，本手法によって発音の手本を見せた場合と，通常の映像により手本を見せた場合の比較実験を行い，提案手法によって学習者の正しい発音の割合が有意に増加することを示している．</p> <p>第 5 章は結論であり，本論文で得られた成果について要約している．さらに，第二章，第三章で述べてきた学習者の内部状態を認識するために，多様なセンシング手法を検討する必要があるなど，今後の展望について述べている．</p>			

## (論文審査の結果の要旨)

本論文は、e-Learning の高度化のために画像解析を用い、学習中の受講者の状態を認識し教員にフィードバックする機能、および、語学学習における発音習得のために口形状を 3 次元的に提示する機能を実現する研究をまとめたものである。得られた主な成果は次のとおりである。

1. e-Learning 中の学習者の様子をカメラで撮影することによって学習者の振る舞いを計測し、得られたデータに対して機械学習およびパターン認識を用いることによって、学習者の内部状態、すなわち、学習者の集中度、学習者にとっての教材の難しさ、学習者の興味、の 3 つの内部状態を推定する手法を提案した。実際に 3000 サンプル(1500 分)を超える学習状況のデータを収集して検証実験を行い、同一被験者内では厳密なマッチング基準で 60%程度、幅を持たせたマッチング基準で 90%程度の精度を得た。
2. 学習者の内部状態推定に関し、個人差に適応する方法を提案した。すなわち、少数の学習者から集中的に集めたデータを用いて、内部状態の識別器を複数構成し、それらの中から新しい学習者について、負荷の少ない方法でその学習者にあった識別器を選ぶ手法を提案した。さらに、そのためのサンプルや識別器の選択基準について検討し、提案する手法によって単純な選択方法に比べて精度が向上することを示した。
3. 語学教育の e-Learning のために、画像センシングを用いて発話時のネイティブスピーカーの顔形状を計測し、正しい発音を行うための顔の動き、特に口の周りの形状がわかりやすいように擬似カラー等で強調して学習者に提示する手法を提案した。実際に 40 人以上の学習者に対して従来の教材との比較実験を行い、提案手法によって、学習者の正しい発音の割合が有意に増加することを示した。

本論文の内容は、画像センシングを e-Learning の現場に適用する新規性の高い手法を提案したものであり、これからの e-Learning の発展に資する、工学的価値が高い内容となっている。したがって、学術上、実際上寄与するところが少なくない。よって、本論文は博士(工学)の学位論文として価値あるものと認める。また、平成 31 年 2 月 28 日、論文内容とそれに関連した事項について試問を行って、申請者が博士後期課程学位取得基準を満たしていることを確認し、合格と認めた。

なお、本論文は、京都大学学位規程第 14 条第 2 項に該当するものと判断し、公表に際しては、当該論文の全文に代えてその内容を要約したものとすることを認める。