

## 保育士養成教育向け歌唱教育支援のための口の形状追尾システム

## Mouse-Shape Tracking System for supporting singing education to pre-school teacher education

小林 翔太†  
Shota Kobayashi中平 勝子†  
Katsuko T Nakahira

## 1 はじめに

ピアノ弾き歌い教育は多くの保育者養成機関で取り上げられるが、こうした技能の伝達には学習者個人の技能レベルを考慮した対面による個別指導が必要となる。しかし、教員数と講義時間の関係により個人に割り当てられる時間は少なく、教育の質が低下するという問題が引き起こされている。

こうした教員数と指導時間数の関係から引き起こされる問題に対して、特に保育者養成機関においては中平ら[1]により ICT を併用した学習がより効果的であると示唆されている。また、保育士に対する歌唱指導は、正確な歌唱のみならず表情を豊かにすることが要求され、[1]中の学生に対するアンケートにおいても、表情について大切な点を示してほしいという傾向が見られた。表情の付き方に関しては、顔の部位(目、鼻、口)の動きの関係をとることで、表情を一意に検出することができるという先行例[2]がある。

自身の歌唱時の表情を数値で表すことで客観的に観察することは、歌唱の表現においても意味があると期待できるため、学生の歌唱練習時における表情つけや発声に伴う顔の動きを記録し、系統的に分析することで学生の弱点を指摘するための支援システムを提案したいと考えている。本稿ではその基礎調査として、特に発声量とも関係があり、歌唱時に大きく動く口の動きに着目し、手本との口の動きを比較することで表情のつき方の差異を取得しそれを分析する。

## 2 提案システムのデザイン

本稿において提案するシステムは、保育者養成機関の学生に対し、教員数と指導時間の不足を補い、歌唱時に正確な表情付けを行う技能を習得させることを目的としているため、次のようにシステムのデザインを行う。

従来の教授方法では学生は講義によりグループレッスン又は対面による個別指導をうけ、学生自身による指導書の閲覧や個人訓練による自主学習を行う。その後、実技試験などにより教員から評価を受けることで技能を修得していく。しかし、自主学習の評価は主観的なものになる場合が多く、自主学習の質に差が出てしまう。また、中平らによる技能者育成の過程分析[3]によれば、動きの反省にビデオカメラを利用したことで予想以上の成績をおさめ、「選手の会得した技術の数量化、およびビデオを取り入れた対面指導は非常に効果的であった。」という考察がされている。

そこで本稿では従来の技能習得までの流れの中に、学習者の練習・技能を数量的に評価し、学習者が自身の練習・技能を客観的に知ることができる支援システムを追加する。

教員数と指導時間の不足を補い、自主学習を支援するシステムとなるため、学習者の練習を自動で分析・評価し、それを学習者にフィードバックすることが求められる。練習の分析・評価には、より豊かな表情付けをできているかを分析するため、学習者の歌唱練習中の動画像を解析することで分析・評価を行

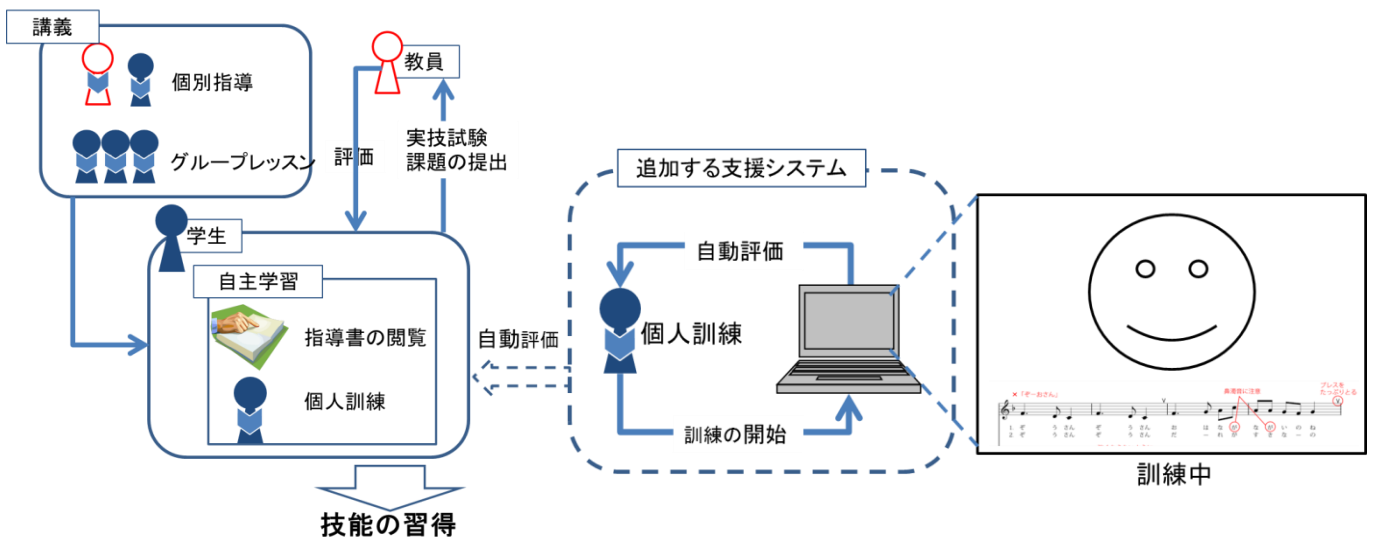


図1 提案システムのデザイン

† 長岡技術科学大学  
Nagaoka University of Technology

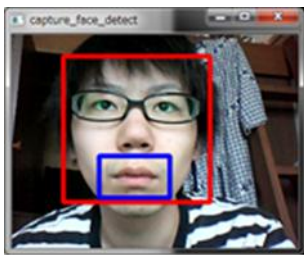


図 2 検出された口領域



図 3 口を開けた場合の認識結果



図 4 領域拡張後の認識結果

う。このとき、練習の分析・評価をリアルタイムで行うことで学習者は練習中に自身の弱点を意識させることができる。また、練習中の動画像を保存できるようにすることで、手本用の動画像と比較することも可能となり、練習中の評価とあわせることで学習の質向上が期待できる。

以上のことから、1) 練習用楽曲の楽譜の表示、2) 練習用楽曲の音楽再生、3) 学習者の練習撮影、4) 歌唱時の表情の分析・評価、5) 手本用動画像と学習者の練習動画像の再生、を支援システムの開発項目とする。

本稿では歌唱時の表情の分析・評価における分析と評価を行う領域の中でも、特に歌唱時の表情付け、発声量に関係のある口領域の検出方法について述べる。

## 2 口領域の検出

手本との口の動きを比較するためには、練習中の学習者の口がある領域を認識させる必要がある。既存の手法を組み合わせることにより、歌唱中に手本と比較するための口領域の認識を行えるか検討を行った。口領域は1) 顔領域の検出、2) 口領域を探索する領域の指定、3) パターンマッチングによる口領域の検出、4) 口を大きく広げた際の口領域の拡張、の順序により検出を行う。

オブジェクトの検出等にはインテルが公開しているオープンソースコンピュータビジョンライブラリの OpenCV を用いる。また、練習は画面を見て行うことを想定しており、撮影は正面から行われるとする。

まず、背景と学習者の分離を行うため、正面顔の学習データを用いて顔領域の検出を行い、映像から顔の領域を検出する。次に口領域を探索する領域を指定する。正面から顔を撮影した場合、通常口は顔の半分より下の位置にあるため、検出された顔領域中の下半分を探索領域とする。

指定された探索領域内で、口領域の検出用データを用いて、口の領域を検出する。図2に検出された口領域を示す。口を閉じている場合は、顔の角度にほぼ関係なく正常に口領域を検出することができる。しかし、図3のように、大きく口を開けた場合には正常に口領域を検出することができない。これは、口の形が大きく変化し、用意されたパターンから外れてしまったためである。口の形状が用意されたパターンから外れる場合は、顔の角度による変化と、口の動きによる変化があげられるが、

画面に表示するコンテンツに練習中に注視する箇所(楽譜、パラメータなど)を表示することで、顔の角度を一定区画内に収めることで問題を解決する。

口の動きによる変化に起因する口領域の未検出問題は次の様に解決する。口を開く動作は下あごの動きによるものであり、上唇の位置が変化しないことを利用し、口領域の拡張を行う。口の開きの最大長は2横指～2横指半であるため、人体の標準的な長さ[4]と日本人頭部寸法<sup>‡</sup>を参照し、図4のように正常に検出された際の口領域を下方へ1.4倍拡張し、それを口が大きく開いた際の口領域として検出する。

## 3 まとめ

本稿では、歌唱練習時における口の動きと手本動画の口の動きを比較し、表情付けの差異を取得することで学生の弱点を指摘するシステムを設計し、その基礎的部分として、比較対象である歌唱練習中の口領域の検出方法について提案した。

実際のシステムの運用に際して、練習室にPC・カメラといった大きな装置を持ち込むことはできず、また、学生のメディアリテラシーを勘案すると、直感的に操作することができるデバイスが望ましい。以上の点からシステムを実装するデバイスにタブレット PC を考えており、性能の低いデバイスでも動作ができるようなシステムの軽量化が今後の課題である。

### 参考文献

- [1] 中平 勝子, 赤羽 美希, 深見友紀子, "ブレンデッドラーニングを取り入れたピアノ弾き歌い指導の改善", 日本教育工学会論文誌, vol. 34, Suppl., pp. 45-48, 2010.
- [2] A. D. Dharmawansa, K. T. Nakahira, Y. Fukumura, "Creating More Realistic e-Learning Environment with Student Real Time Facial Features", International Conference on Intelligent Decision Technologies, vol.2012, pp.54-55, 2012
- [3] 中平 勝子, 横山 淳一, "マルチメディアを利用した企業内旋盤加工技能者育成の過程分析", 教育システム情報学会全国大会講演論文集, vol.31, pp.419-420, 2008
- [4] 社団法人人間生活工学研究センター, "人間特性基盤整備事業成果報告書", 2006

<sup>‡</sup> AIST 人体寸法・形状データベース  
<http://riodb.ibase.aist.go.jp/dhbodydb/index.php.ja>  
 2012年6月30日