Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 8, Number 1, pages 68–70, January 2019

A Study on Effective Skill Passing Method for Uta-Sanshin

Tsugushi Nagahama
Electricity and Communication System Engineering
National Institute of Technology, Okinawa College
Okinawa, Japan
ac184602@edu.okinawa-ct.ac.jp

Kei Miyagi
Electricity and Communication System Engineering
National Institute of Technology, Okinawa College
Okinawa, Japan
k.miyagi@.okinawa-ct.ac.jp

Chikatoshi Yamada
Electricity and Communication System Engineering
National Institute of Technology, Okinawa College
Okinawa, Japan
cyamada@okinawa-ct.ac.jp

***Abstract*—In the "Okinawa 21st Century Vision", the succession and reconstruction of traditional culture are required, one of which is Uta-Sanshin. However, it is thought that many instructive and esoteric expressions of the leader, a decrease in the number of players accompanying the aging, and difficulty in understanding the score are interfering with succession and reconstruction. In this research, we aim at establishing the traditional support technology which can extract and identify singing skills in Uta-Sanshin using deep learning and understand the basics of skill in an easy-to-understand manner.**

Keywords—Deep laerning, Skill passing, Formant

I. はじめに

沖縄の産業発展を阻害する要因として、下請けの仕事が多く低賃金労働から抜け出せないため、優秀な人材を確保する事が難しいという構造的な問題がある。沖縄が自立的発展を遂げるには、沖縄の強みを活かした産業を開拓する必要があり、昨今では、様々な学術分野において ICT の統合化による新産業・新サービスなどの新たな価値の創出や地域活性化が求められている[1]。特に、県民が望む沖縄の姿を示した『沖縄２１世紀ビジョン[2]』において伝統文化の継承や復興が求められており、その代表的なものの一つに歌三線がある。しかし、その技能は口伝式による伝承が一般的であるため、指導者の感覚的かつ難解な表現が多い。また、高齢化に伴う担い手の減少や、図１に示すように歌唱用楽譜（工工四）の分かりづらさなどが継承や復興の妨げになっていると考えられる。関連技術として、三線を弾くための支援アプリケーションなどがすでに提案・開発されている。しかし、最も重要かつ習得が難しいとされる歌三線独特の歌唱技能の伝承法に関する研究はほとんどなされていない。そこで、歌三線と ICT の異分野融合により、技能の基本を伝え知伝え知ることのできる明確な手法の確立が必要だと着想した。特に、東京五輪は、我が国の文化や伝統等の価値を世界へ発信して文化芸術立国を実現する上で絶好の機会となっている[3]。そこで、2020 年に向けて、歌三線と ICT を融合した沖縄型イノベーションの実現を目指す。また、伝統芸能と ICT の融合のあり方を研究する事により、学術分野の異なる人材が協働する事によるイノベーション創出の可能性について考究する。

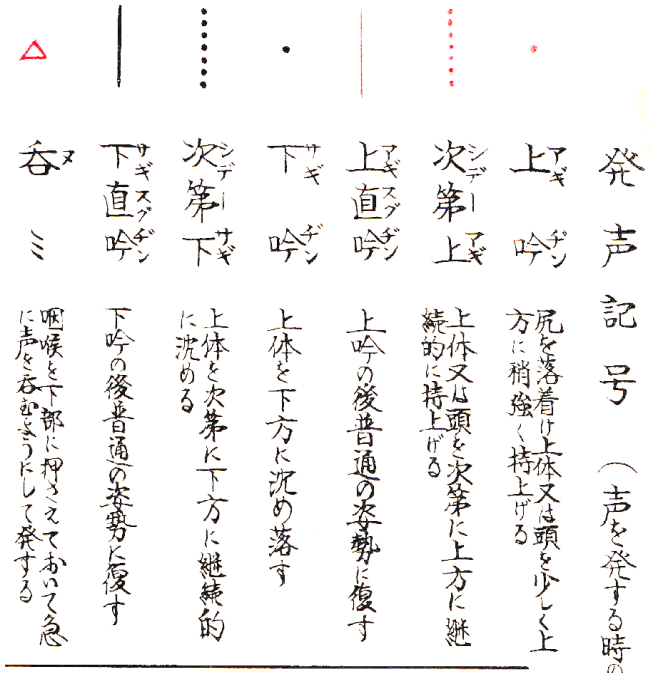

図１：歌唱用楽譜（野村流　工工四上巻より）

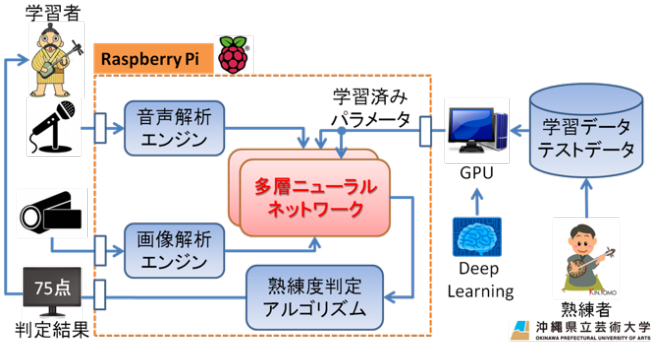

図２：伝承支援システムの概略図

II. 技能伝承システムの構成

本研究で提案する技能伝承線システムの概略図を図2に示す。熟練者の歌声の特徴を事前に抽出してディープラーニング（多層ニューラルネットワーク）により学習させ、熟練者の歌声と学習者の歌声を判別するシステムを構築する。学習は GPU を用いて行い、学習済みパラメータを取得し、プロトタイプの構築を目的として、多層ニューラルネットワークをラズベリーパイ上に実装する。最終的には、リアルタイム処理を目的として、ハードウェア（FPGA）に実装することを目標としている。加えて、ディープラーニングにより熟練者の歌声と一般人の歌声の違いを可視化するために、歌三線の歌唱法を分析するにあたって音声スペクトルの一種であるフォルマントに注目し、LPC（Linear Prediction Coding）分析を行う。分析する音源としては、沖縄の伝統的な音楽として最も広く知られているものであり、祝宴の際などに歌われる「かぎやで風」を使用した。

A. フォルマントの抽出

歌声の特徴を抽出するにあたって注目したのがフォルマント[4]である。フォルマントは声道の特性を表しており、声道には特定の周波数の音を大きく表れるポイントが存在する。そのポイントがフォルマントとなる。フォルマントは複数存在するが、声の特徴を決定づけるために必要なものは、振動数（周波数）が小さい方から５つ目までのポイントであるとされている。この周波数をフォルマント周波数と呼ぶ。第１・第２フォルマントは母音の特徴を表しており、第３・第４・第５フォルマントで声質を決定する。歌三線独特の発声法である、喉を閉じる、絞るといった表現に深く関係しているのではないかと考え、このフォルマントに注目した。

B. LPC 分析

LPC（Linear Prediction Coding）分析は、Z 変換により声道の特徴をモデル化するもので、声道を音響管に見立てた時の特徴量を表すことができる。線形予測器のフィルタ係数は、LPC 係数と呼ばれ、音声合成フィルタを構成するために用いられる。この LPC 係数を求めることを LPC 分析または線形予測分析と呼ぶ[5]。音源から声帯の特性と、声道の特性を分離することができ、図３に LPC 分析を行い、声道の特性のみを抽出した結果を示す。図３ではかぎやで風を歌った音源を使用し、「あ」という発声を LPC 分析により分離された音声信号の声道の特徴のみを抽出している。左から順に第１・第２・第３・第４・第５・第６フォルマントとなっている。フォルマント周波数を導出することで歌声の特徴を数値化することが可能となる。

III. フォルマントによる学習

ディープラーニングによる特徴抽出を行うにあたって本研究では、FPGA 上に実装することを考慮し、学習時間の短縮化を行う必要がある。そこで、分析する対象をフォルマント周波数固有の値に限定することで、計算量の削減、すなわち学習時間の短縮化が図れるのではないかと考えた。従来研究との差分として、文献[6]より、音声認識の分野において、音声信号をスペクトログラムに変換し、畳み込みニューラルネットワーク（CNN: Convolutional Neural Network）[7]による学習を行った結果、高い識別率を得ることができたという報告がある。しかし、これらは３６時間と膨大な学習時間がかかる。[5]本研究では，学習データ数に限りがあることに加え，FPGA 上に実装することを考慮し，学習時間の短縮化を行う必要がある。

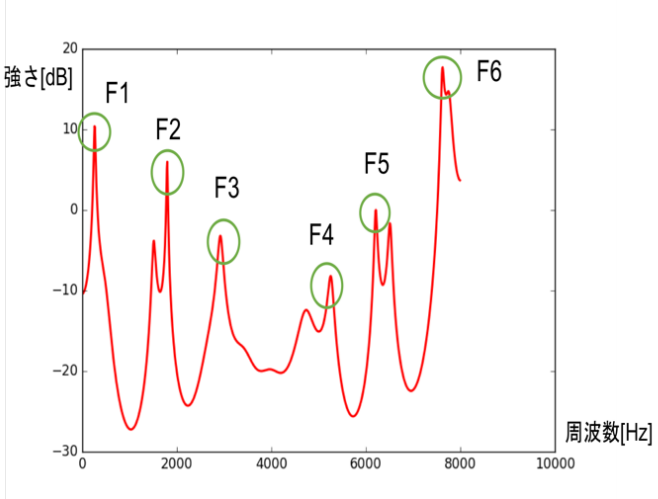

図３：フォルマント周波数の導出

そこで分析する対象をフォルマント周波数固有の値に限定することで，計算量の削減，すなわち学習時間の短縮化が図れるのではないかと考えた。従来研究との差分として，文献[5]で示されている手法と，提案手法の識別率・学習時間の比較を行った。

A. ニューラルネットワークの構成

フォルマント周波数の値を学習するにあたり，使用したニューラルネットワークはMLP（Multilayer Perceptron）である。MLPは神経細胞を模したパーセプトロンを多層化したニューラルネットワークであり，全結合型のもっとも単純なディープラーニングの構成である。入力層，出力層，中間層が4層の全部で6層の構成となっている。スペクトル画像を対象としたCNNと比較し，入力数を大幅に削減する事ができる。

B. 評価方法

ニューラルネットワークの構成はCNNとMLPとなっている。各ネットワークの構成は，Sonyが開発したNeural Network Consoleを用いて事前に最適化を行い，その構成を用いている。入力のノード数は画像データが64×64の大きさ数値データは特徴を決める第5フォルマントまでの情報となっている。学習データは一般人の歌声を400，熟練者の歌声を630の合計1030，テストデータは一般人の歌声を70，熟練者の歌声を200の合計270ものデータを用意した。それぞれ学習し2値分類を行い，識別率を比較した。

ディープラーニングのフレームワークは Chainerを使用した。本研究によって得られる知見の多くは他のフレームワークにも適応可能である。また、Chainer は、FPGA 上への実装を想定する上で、高位合成にも対応しているため、ハードウェア化が容易であるという利点がある。

C. 評価結果

評価結果をTABLE Ⅰに示す。従来のスペクトログラム画像を用いる方式と比べ，提案方式（フォルマント周波数）でも同程度の識別率86%を達成できることが確認できた。また、学習データをフォルマント周波数に限定する事により，ニューラルネットワークの入力ユニット数を大幅に減らす事が可能になり，学習時間を短縮化する事ができた。画像データの方は1枚につき64×64の4096のデータを入力しているのに対し，数値データの方は1つのデータに対し，フォルマント周波数の5つの情報のみを入力しているためだと考えられる。

TABLE I.　評価結果

学習データ	識別率（%）	学習時間（SEC）
スペクトログラム（画像）	90	840＝14(MIN)
フォルマント周波数（数値）	86	278＝4.5(MIN)

IV. おわりに

本研究ではLPC分析により歌声の特徴抽出を行い，その結果を学習させ，熟練者の歌声と一般人の歌声の識別を行い，従来手法との比較を行った。実験結果より，従来方式と同等の識別率となり，学習時間の短縮化した。この結果より，本研究は必要な情報のみを抽出し効率よく学習を行うことが出来たと考えている。今後は、学習したデータをもとに習熟度を評価するアルゴリズムを検討しRaspberry PiやFPGAを用いて歌三線における技能伝承を支援するシステムのプロトタイプを開発する。

今後の展望として将来的には、提案する技能伝承支援システムを歌三線の教育や沖縄県の産業発展のために活用したいと考えている。例えば、本提案システムを教育現場に教材として提供できれば、AIによる本格的な歌三線の学習支援を実施出来る他、歌三線向け歌唱アプリや無形文化財（人間国宝の技能）の永久保存、国内外の様々な歌唱様式への発展等、様々な応用・発展が期待される。

謝辞

本研究は，公益財団法人電気通信普及財団および一般財団法人東熱科学技術奨学財団の助成を受けて実施した。厚く御礼申し上げます

参考文献

[1] 総務省,”平成28年度版情報通信白書”, 2016

[2] 沖縄県,”沖縄21世紀ビジョン基本計画【改定計画】”, 2017

[3] 文部科学省,”平成28年度文部科学白書”, 2016

[4] 高橋純 他, ”歌い手のフォルマントについての一考察：ベル・カント唱法と科学的研究を比較して”，京都市芸術大学紀要, Harmonia(47), 47-63, 2017-03

[5] NVIDIA , “DEEP CONVOLUTIONAL NEURAL NETWORKS FOR SPOKEN DIALECT CLASSIFICATION OF SPECTROGRAM IMAGES USING DIGITS”, NVIDIA Deep Learning Day 2016 Sprig