

# 深層学習による琉球古典音楽のリアルタイム推論

非会員 長濱 嗣志\* 非会員 上原 一郎\* 非会員 宮城 桂\*<sup>a)</sup>  
正員 山田 親稔\* 正員 市川 周一\*\*

## Real-time Inference of Ryukyuan Classical Music Using Deep Learning

Tsugushi Nagahama\*, Non-member, Ichirou Uehara\*, Non-member, Kei Miyagi\*<sup>a)</sup>, Non-member,  
Chikatoshi Yamada\*, Member, Shuichi Ichikawa\*\*, Member

(2019年2月11日受付, 2019年6月6日再受付)

The classical music “uta-sanshin” has been sung since the Ryukyu Kingdom period, and its skills commonly depend on folklore method by bush telegraph. Accordingly, there exist much sensibilities and esoteric expressions of the uta-sanshin expert in passing down the skill. Also, the decrease in number of successors accompanying aging and the difficulty in understanding the musical score are hindering the inheritance and the reconstruction of the music. In this paper, we apply the deep learning to Ryukyuan classical music and develop a system that identifies vocalism by real-time processing. The results of the evaluation, compared with the conventional method, show that the execution time is reduced to 98%, and the identification accuracy is improved by 6%.

キーワード：琉球古典音楽, 歌三線, 深層学習, ニューラルネットワーク, 歌唱フォルマント

**Keywords:** ryukyuan classical music, uta-sanshin, deep learning, neural network, singing formant

### 1. はじめに

我が国は、諸外国を魅了する多くの有形・無形文化財を有している。文化芸術立国の実現の向け、これらを維持、継承、発展させることはもとより、日本人自身はその価値を十分に認識した上で、国内外への発信を、更に強化していくことが求められている<sup>①</sup>。しかしながら、少子化や高齢化の進行、ライフスタイルの変化や多様化などにより伝統文化の担い手が減少することによる後継者不足が危惧されている。県民が望む沖縄の姿を示した『沖縄 21 世紀ビジョン<sup>②</sup>』においても伝統文化の継承や復興が求められており、その一つに琉球古典音楽の歌三線がある。三線は、中国福建省で生まれた弦楽器「三弦」を原型とする撥弦楽器であり、15 世紀以降、琉歌と共に琉球王国で発展した。昨今で

は古典音楽や民謡の他、ポップスやクラブミュージックなど様々なジャンルで用いられ、演奏するアーティストも沖縄音楽や沖縄文化圏に留まらない。加えて、歌三線の担い手として 4 名が重要無形文化財認定保持者として指定されており、技能の維持、継承が求められている。しかしながら、その技能は口伝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歌三線独特の楽譜である工工四（くんくんしー）<sup>③</sup>の分かりづらさなどが継承や普及の妨げになっていると考えられる。歌三線を普及・発展させるための取り組みとして、これまでに三線を学ぶためのアプリケーションなどが提案・開発されている。また、岡田らは、歌三線を練習するための教材開発<sup>④</sup>を行なっているが、いずれも三線を弾くための支援に限られており、伝承するにあたって難しいとされる歌唱法についての研究事例は少ない。山口らは、歌三線における発声法の一つである“大掛（ウーガキ）”の音としての仕組みと構造を解析している<sup>⑤</sup>。歌唱時の音声をつりえ解析することにより、“大掛”を掛けている間の所要時間や最大・最小周波数を明らかにしている。しかしながら、歌唱技能の要となる咽喉の使い方や、その伝承方法までは考慮していない。そこで、筆者らは歌三線における技能の本質を分かりやすく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。その実現のためには、歌唱特

a) Correspondence to: Kei Miyagi. E-mail: k.miyagi@okinawa-ct.ac.jp

\* 沖縄工業高等専門学校

〒 905-2192 沖縄県名護市辺野古 905

National Institute of Technology, Okinawa College  
905, Henoko, Nago, Okinawa 905-2192, Japan

\*\* 豊橋技術科学大学

〒 441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

Toyohashi University of Technology 1-1 Hibirigaoka,  
Tempaku, Toyohashi, Aichi 441-8580, Japan

有の音響的特徴量を明らかにする必要がある。関連研究として、オペラや邦楽歌唱、民謡歌唱における音響分析が各所で行われている。特筆すべきは、典型的な良い声の条件として歌唱フォルマント (Singing formant) を有することが示されており、それが「歌声らしい」という聴感的印象に影響を与えているという知見がある<sup>(6)(7)</sup>。しかしながら、フォルマント自体が歌唱技能にどのように関わっているのかと言う研究は少ない<sup>(7)</sup>。技能の伝承を目的として、例えば、歌唱フォルマントの特徴を多変量解析して、定量化かつ可視化できれば、口伝による伝承が一般的であった熟練者の暗黙知を形式知として分かりやすく表現できる可能性がある。本研究では、歌三線における技能伝承を支援するシステムの開発を目指し、歌唱フォルマントと深層学習を用いた分類器により、人が知覚する「歌三線らしい」という聴感的印象を高精度に推論する手法を提案する。歌唱技能の伝承を支援するシステムを実現するためには、音響的特徴量として得られる模範との差異を学習者が即座に把握できることが望ましい。そこで、深層学習における推論のリアルタイム処理を実現するための手法について示す。

## 2. 琉球古典音楽「歌三線」の音響的特徴

歌三線における独自の発声法は、咽喉の使い方が重要とされている。以下に歌三線の楽譜である工工四<sup>(8)</sup>に記載されている声楽譜記号の名称とその発声法の一例を示す。

- 大掛 (ウーガキ)：咽喉を圧迫して母音を発す。
- 呑 (ヌミ)：詰まった音即ち促音を出すようにして、息を呑込んで歌いだす。
- ウチグイ：老の高さで喉をえぐるように圧迫して母音を軽く発す。

これらは、技能としての暗黙知を比喩表現として形式化したものであるが、「咽喉を圧迫」「喉をえぐる」など難解な説明が数多く存在する。これらの独特な発声法が作り出す音色は咽喉の開閉によって変化するとされており、開閉が上手くいかないと同一音色で音色の複雑さが出ないとされている<sup>(9)</sup>。その技能を明らかにするためには、歌三線固有の発声が奏でる音響的特徴を明らかにする必要がある。ここで、一般的な歌声は、話声と比較して発声の高さや強さの変動幅が広く、より複雑で動的な特性を持つことが知られている。また、スペクトルにおける歌声特有の特徴も存在し、中でも Sundberg によって報告された歌唱フォルマントが代表的である<sup>(8)</sup>。フォルマントとは、音響的にある音の音色を特徴づけ、音色の異なる他の音から区別させる周波数成分の集りである。音声などの各周波数帯における強さの分布として表わされる他、その周波数は声道の形状と関係する事が知られている。その中でも、歌唱フォルマントとは、主に男性オペラ歌唱の母音スペクトルの 3kHz 付近に存在する顕著なピーク成分とされる。これは、歌声に聴感的な響きを与える要因として考えられており、日本の伝統芸能における歌唱にも観測されることが報告されている<sup>(9)(10)</sup>。

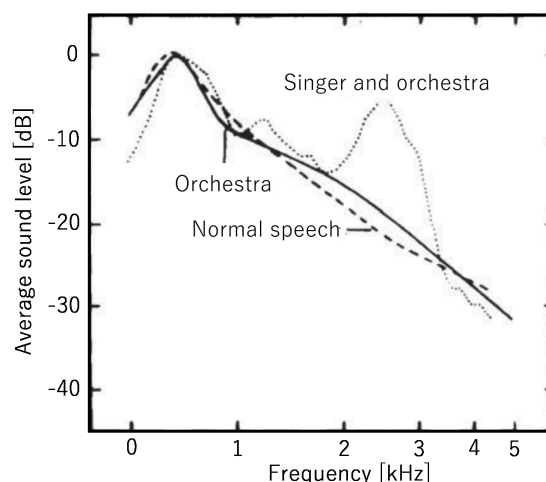


Fig. 1. Long time spectrum in soloist (practice) and when not in symphonic music (dotted line), and in normal speech (broken line)<sup>(11)</sup>.

Fig. 1 は歌手の歌声、その歌い手の話し声、ならびにオーケストラの音の長時間スペクトルを比較したものである<sup>(11)</sup>。オーケストラの音には 3kHz 付近に共振がそれほど顕著に現れないため、その付近の周波数帯域にパワーの盛り上がりが存在することはオーケストラの音以外の音が存在することを示す、すなわち、ベル・カント唱法による歌声が存在することの証拠となる。その結果、歌声はオーケストラの音にマスクされにくいという傾向を生む。同じ人間の話し声では、3kHz 付近のピークが出現していないことから、この特性は個人の普段の状態から出現するものではなく、歌唱時に特異的に出現するものであると考えられる<sup>(7)</sup>。しかしながら、歌唱フォルマントの定義はいくつか存在している。例えば Sundberg は男性オペラ歌手の母音スペクトルに現れる 2.8kHz 付近のピークであるとしており<sup>(8)</sup>、中山らは能楽による歌唱において 3~4kHz 付近に同様のピークが見られる事を示している<sup>(9)</sup>。また、斎藤らはソプラノ歌手の場合、音源の非周期性指標が 3kHz 付近において非常に小さくなる事を示している<sup>(6)</sup>。このように、歌唱フォルマントは有無について議論される事が多いが、その強度や形状が評価された事例は少なく、歌唱の習熟度の評価への応用に関しては十分議論されていない<sup>(12)</sup>。

歌唱フォルマントを習熟度の評価に応用した先行研究としては、フォルマントのピークの鋭さ (Q 値) を定量化することで、熟練度の評価が出来る可能性が示唆されている<sup>(13)</sup>。また、佐久間らは音響特徴量として LPC (Linear Predictive Coding) スペクトル包絡線における高周波帯域の鋭さを Q 値、周波数スペクトルにおける高周波帯域の含有率を SFR (Second Formant Ratio) として定量化し、それらによる習熟度評価手法について検証と考察を行っている<sup>(12)</sup>。声楽発声指導者からの一ヵ月間の指導を通して、歌唱技能の向上とともに、Q 値と SFR 値が増加する事を確認している。筆者らは、これらの知見を参考に、歌唱フォルマントの形状から深層学習により知覚情報を抽出することで、歌唱技能

の習熟度を定量化し、技能伝承に活用できる可能性があると考えた。音声の推論に深層学習を利用した研究としては、Nigelらの研究がある<sup>(14)</sup>。Nigelらは、50時間にも及ぶ会話データ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により6言語、20方言の推論を行っている。筆者らも、これまでに歌唱データからフォルマント周波数を抽出し、全結合ニューラルネットワークに学習させ、推論精度と推論時間の評価を行ってきた<sup>(15)~(17)</sup>。全結合ニューラルネットワークへの入力データとして、声道の特徴を表すフォルマント周波数だけに着目する事で、推論の際に必要な特徴を残しつつ、入力データの次元数を削減している。それにより、Nigelらのスペクトログラム画像を用いる手法に対して推論時間を大幅に短縮した。しかしながら、推論精度は若干劣る結果となっていた。そこで、本研究では推論精度の向上を目的として、歌唱フォルマントの形状(Q値)に着目し、それを深層学習に応用する事で習熟度を含めた音響的特徴を高精度で推論する手法を手案する。3章では、歌唱フォルマントの抽出方法と深層学習への適用方法について述べる。

### 3. 特徴抽出と深層学習による推論

歌三線における技能としての特徴を抽出・推論するシステムの構成をFig.2に示す。まず、学習用データとして、歌三線の熟練者の歌唱データを用意する。次に、音声解析を行い、歌唱フォルマントを抽出する。そこで得られたフォルマントを深層ニューラルネットワークに学習させ、学習済みモデルを得る。続いて、学習者の歌唱をマイクから取得し、音声解析を行い、歌唱フォルマントを抽出した後に推論処理を行うことにより、熟練者との音響的特徴量の類似度を定量的に求める。その結果を学習者にリアルタイムでフィードバックすることにより、学習者は、自らの習熟度を確認することが可能となる。次節では、歌唱フォルマントを抽出するための手法について述べる。

〈3・1〉 LPC 分析 LPC (Linear Prediction Coding) 分析<sup>(18)</sup>は、Z変換により声道の特徴をモデル化するものであり、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。線形予測器の構成をFig.3に示す。 $h_m(m=1\cdots m)$ はLPC係数、 $s(n)$ は線形予測器への入力音声、 $w(n)$ は出力の音源情報、Dブロックは1時刻の遅延を表している。この時、出力 $w(n)$ は、

$$w(n) = s(n) + \sum_{m=1}^M h_m s(n-m) \cdots \cdots (1)$$

として与えられる。線形予測器の出力 $w(n)$ からピッチ周期やパルスの大きさなどを求めることができる。また、伝達関数 $H_{LP}(z)$ は、

$$H_{LP}(z) = 1 + \sum_{m=1}^M h_m z^{-n} \cdots \cdots (2)$$

として与えられる。これらの式に基づいて分析を行うことにより、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る。すなわち、LPCを用いると、与えられた音声信号の周波数スペクトルの概形(スペクトル包絡)を求めることができる。音素を推論するうえでスペクトル包絡のピーク(フォルマント)の位置や形状は重要な手がかりを与えるため、LPCは歴史的に重要な音響的特徴量抽出法としての役割を果たしてきた。Fig.4にLPC分析により抽出したフォルマント周波数を示す。フォルマントはいくつも存在するが、声の特徴を決定づけるために必要なものは振動数が小さい方から5つ目までとされており、その周波数

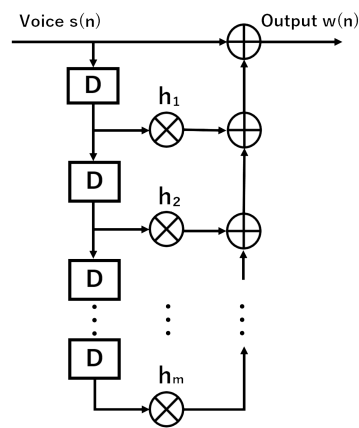


Fig. 3. Construction of linear predictor.

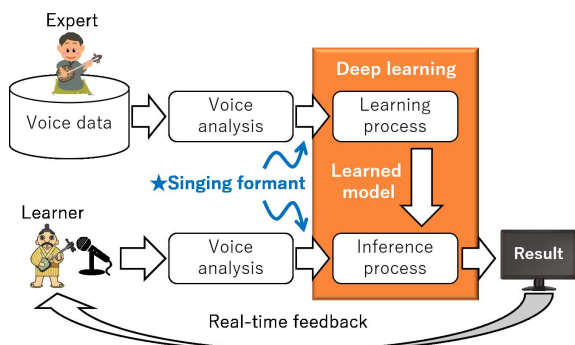


Fig. 2. System configuration

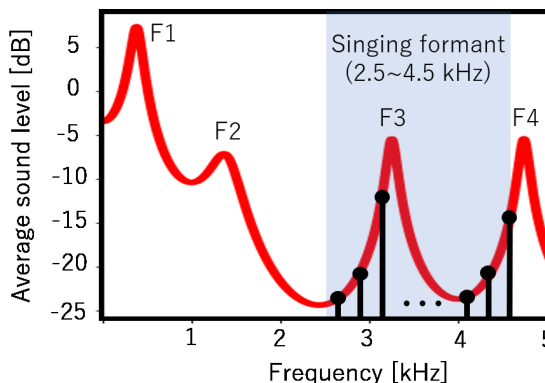


Fig. 4. Singing formant of uta-sanshin.

をフォルマント周波数と呼ぶ。第1 (F1)・第2 (F2) フォルマントは母音の特徴を表しており<sup>(7)</sup>, 第3 (F3) 以上のフォルマントで声質の特徴を表す<sup>(8)</sup>。また, フォルマントは声道の特徴を表すことから<sup>(9)</sup>, これらの値は歌三線独特の発声法である、『喉を閉じる』や、『喉を絞る』といった表現に深く関係しているものと考えられる。本研究では, 2章で述べた通り, 歌唱時の声道の特性を表す歌唱フォルマントが2.8 kHz~4 kHzの領域に現れる点に着目し, Fig. 4に示すように歌唱フォルマントの領域を2.5 kHz~4.5 kHzとして定義する。その領域の平均音レベルを特徴ベクトルとして抽出し, 深層学習により歌唱フォルマントにおける特徴を抽出する手法を提案する。歌唱フォルマントのみを深層学習に適用する事で, 従来のスペクトル画像を用いた手法に比べて入力データの次元数を大幅に減らし, 演算に要する時間を削減する, すなわち, リアルタイム処理が可能となる。

**3・2) 深層ニューラルネットワーク** 音声において, 深層ニューラルネットに基づく技術進展は目覚ましく, 全結合型ニューラルネットワーク (Fully-connected Neural Network: FNN), リカレント型ニューラルネットワーク (Recurrent Neural Network: RNN), 畳み込み型ニューラルネットワーク (Convolutional Neural Network: CNN) の適用, 及び様々な拡張の検討が進んでいる<sup>(10)</sup>。本研究では, RNNに代表されるような時系列情報を用いた予測ではなく, 歌唱フォルマントから得られる形状 (Q値) から音響的特徴として「歌三線らしさ」の推論ならびに習熟度の評価を目的としているため, パターン分類器として一般的に用いられるCNNおよびFNNを利用する。

**3・2・1) 畳み込みニューラルネットワークによる推論** 畳み込みニューラルネットワーク CNNは畳み込み層 (Convolution Layer) とプーリング層 (Pooling Layer), 全結合層 (Fully connected layer) から構成されるニューラルネットワークである。深層学習の分野では画像の推論等に用いられる。Nigelらは, 音声をスペクトログラムとして2次元の画像に変換する事により, CNNを用いた音声の推論を実現している<sup>(14)</sup>。学習データとして用いられるスペクトログラムは音声信号を短時間フーリエ変換 (STFT) し, 周波数スペクトルとして表したものである。以下の式で示すように, STFTの絶対値を2乗することでスペクトログラムを得る。

$$spectrogram(t, w) = |STFT(t, w)|^2 \dots \dots \dots (3)$$

先行研究では音声をスペクトログラムに変換し, 50万枚以上の教師データを学習させることで高い正答率を得ている<sup>(20)</sup>。しかし, 深層学習は一般的に多くの演算処理が必要となるため, GPUで実行する必要が生じる。GPUは, 消費電力も大きいため, 学習時間や推論時間の短縮化は深層学習における課題の一つとなっている。さらに, 歌三線において熟練者の歌唱データを大量に収集する事は困難であるため, 少量のデータから上手く特徴を抽出しなければな

らない。そのためには, 歌唱データの前処理 (データクレンジング) が重要となる。

**3・2・2) 全結合ニューラルネットワークによる推論** 全結合ニューラルネットワーク (FNN)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, 音声認識の分野で広く使われている。本研究では, 良質な学習データを取得するために, 前処理としてLPC分析を行う。そこで得られた歌唱フォルマントをFNNの学習データとして使用する。そのためには, LPC分析で得られた歌唱フォルマントをFig. 4に示すように一定間隔でサンプリングして特徴ベクトルを得る必要がある。ここで, サンプリング数Nを増やせば, 歌唱フォルマントの形状 (Q値)の詳細を取得できる一方で, 入力データの次元数が増えてしまうため, 推論時間は増加する事が予想される。すなわち, 推論精度と推論時間はトレードオフの関係となるため, 4章では, サンプリング数Nを4~64の範囲で評価し, 精度と時間の関係性を明らかにする。その上で, 最適なサンプリング数Nを定め, 従来手法との比較・評価を行う。

#### 4. 評価

本方式の有効性を明らかにするために, CNNにスペクトログラムを適用した手法と, FNNに歌唱フォルマントを適用した手法の比較・評価を行う。深層学習に用いる音源としては, 沖縄の伝統的な音楽として最も広く知られており, 祝宴の際などに歌われる「かぎやで風」を使用した。これらの音源はYouTubeや歌三線の経験者に歌ってもらい取得した。また, 「歌三線らしさ」の特徴を抽出するため, ポップスや演歌などの異なるジャンルを学習させ, 正しく推論できるかの評価を行った。

**4・1) 開発環境と評価条件** 開発環境をTable 1, 評価条件をTable 2に示す。学習データは歌三線の熟練者の歌声 (3曲)を楽器の音が入らないように取得し, 1 [sec] 間隔で区切ったものを700パターン, 同様にポップス (2曲), 演歌 (2曲)の歌唱音源を1 [sec] 間隔で区切ったものを700

Table 1. Development environment.

OS	Mac OS 10.14.2
CPU	Intel Core i5
Framwork	Chainer 4.2.0
Langrage	Python 3.6.0

Table 2. Learning condition of CNN and FNN.

Image size of CNN		64×64
Number of feature vectors of FFNN		4, 58, 16, 32
Epoch number		100
Learning rate		0.001
Batch sizeN		50
Training data	Correct answers	700
	Incorrect answers	700
Test data	Correct answers	100
	Incorrect answers	100

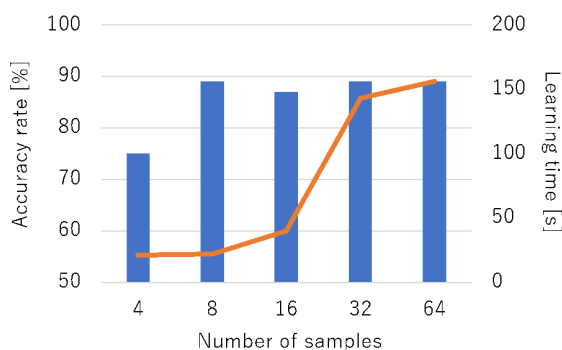


Fig. 5. Evaluation of identification accuracy of FNN.

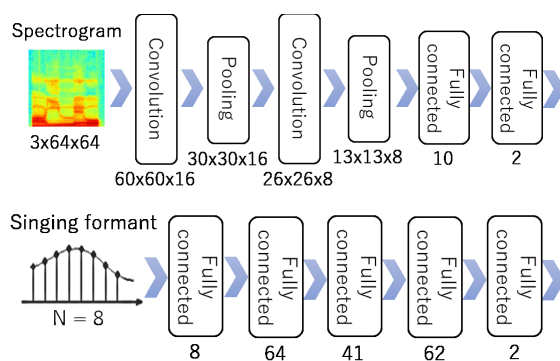


Fig. 6. Optimization of neural networks.

パターン, 計 1400 パターン用意した。テストデータも同様に, 熟練者の歌声 (1 曲) を 100 パターン, その他のジャンルの歌声 (1 曲ずつ) を 100 パターンの合計 200 パターン用意した。それらをスペクトログラムと歌唱フォルマントに変換して CNN と FNN を学習させ, 歌三線の特徴をどの程度正しく推論できるかの評価を行った。なお, ニューラルネットワークは Python を用いて記述し, 深層学習のフレームワークは Chainer を使用した。スペクトログラムの画像サイズは  $64 \times 64$  pixel とし, 歌唱フォルマントから得られる特徴ベクトルはサンプリング数を  $N$  として,  $N=4, 8, 16, 32$  の場合で評価した。特徴ベクトルの取得方法を以下に示す。

- (1) 歌唱フォルマントの周波数範囲 (2.5 kHz~4.5 kHz) を  $N$  等分する。
- (2) 各等分点 ( $n=0, \dots, N$ ) の周波数に対応した [dB] 値を特徴ベクトルとする。

本研究では, 先ず  $N$  の値が推論精度と学習時間に与える影響を評価するために, FNN において歌唱フォルマントのサンプリング数  $N=4, 8, 16, 32$  とした場合の推論精度と学習時間の評価を行った。ニューラルネットワークの構成は 4 層とし, 中間層のノード数は 1024 で統一して評価した。結果を Fig. 5 に示す。棒グラフは正答率, 折れ線グラフは学習時間を示している。評価の結果,  $N=4$  の場合に 73% と精度が低くなったが,  $N=8, 16, 32$  の場合の推論精度は 87%~89% となり,  $N$  の値に関係なく, ほぼ同等の正答率が得られた。一方, 学習に要した時間は,  $N=32$  の場合に 143 [sec],  $N=16$  の場合に 40 [sec],  $N=8$  の場合に 22 [sec] であった。これらの結果から, 正答率・処理時間ともに良好な結果が得られた  $N=8$  を最適値として採用する。

**〈4・2〉ニューラルネットワークの最適化** 評価に先立ち, CNN と FNN のネットワークを Sony が開発した Neural Network Console<sup>(21)</sup> を用いて事前に最適化を行った。Neural Network Console には, Structure Search (ネットワークの自動最適化機能) をイネーブルにした状態で学習を実行すると, 自動的にネットワーク構造を様々に変化させながら, より精度が高く演算量の少ないネットワーク構造を探す最適化が行われる。評価には, 自動的に構成された 50 パターン以上のネットワーク構造から, 最も精度が高いものを選

Table 3. Evaluation of CNN and FNN

	Accuracy [%]	Learning time [sec]	Inference time [sec]
Spectrogram with CNN	89	840	3.07
Singing formant with FNN	95	301	0.06

択して用いた。Fig. 6 に最適化後のニューラルネットワークの構造を示す。CNN は 6 層で構成され, 活性化関数はソフトマックス関数を用いている。一方, FNN は 5 層で構成され, 活性化関数には tanh 関数を用いている。また, 図中の数字は各層におけるノード数を示している。CNN では縦横 64 pixel のカラー画像を入力としているため, 次元数は 12288 ( $3 \times 64 \times 64$ ) 次元であるのに対し, 提案手法では  $N=8$  次元まで減らす事が可能となる。

**〈4・3〉評価結果** 4.2 節で得られたネットワーク構造に基づいて, 深層学習のフレームワーク Chainer と Python を用いてニューラルネットワークを記述し, 正答率と学習時間・推論時間の評価を行った。学習条件は Table 2 で示した通りである。評価の結果を Table 3 に示す。スペクトログラム画像と CNN を用いる手法は正答率 89% であったのに対し, 歌唱フォルマントと FNN を用いる手法は, 正答率 95% を達成できることが確認できた。これは, 歌唱フォルマントに着目することにより, 歌声の特徴を上手く捉えることが出来たためだと考えられる。さらに, 特徴ベクトルを歌唱フォルマントに限定することにより, 次元数を大幅に削減し, 推論時間を 98% 短縮できる事が確認された。特に, 推論処理に関しては, GPU を用いずとも 60 [msec] で実行する事が出来ており, リアルタイム処理にも十分対応できるという知見が得られた。また, GPU のような高い並列性を持つデバイスに依存する必要もないため, システムの低コスト化や低消費電力化にも有用であると言える。

## 5. おわりに

県民が望む未来の沖縄の姿を示した基本構想『沖縄 21 世紀ビジョン』において伝統文化の継承や復興が求められており, 歌三線においても新たな担い手や継承方法を真剣に模索すべき時期に来ている。例えば, 中学校学習指導要領

では、音楽の授業に『我が国の伝統的な歌唱のうち、伝統的な声の特徴が感じ取れるもの』を含める事が明記されている。しかしながら、歌三線が教育現場で普及しているとは言いがたい。その原因として、指導者不足が考えられる。そこで本研究では、歌三線における技能伝承を支援するシステム開発の一環として、歌唱フォルマントと深層学習を用いた分類器により、人が知覚する「歌三線らしい」という聴感的印象を高精度に推論する手法を提案した。加えて、深層学習における推論のリアルタイム処理を可能にする手法について示した。結果として、スペクトログラムとCNNを用いた推論手法に対し、推論時間を98%短縮しつつ、推論精度は6%向上した。今後は、習熟度の評価指標を明らかにするとともに、技能の伝承を支援するための可視化手法について検討を進め、技能向上に結びつけるためのシステム構成を明らかにする予定である。本システムを教育現場に学習教材として提供すれば、AIによる歌三線の学習支援を実施出来るものと期待される。今後の展望としては、歌三線向け歌唱アプリや歌三線風ボーカロイドの開発、重要無形文化財の永久保存、国内外の様々な歌唱様式への発展等、様々な応用・発展が考えられる。

#### 謝辞

本研究は、公益財団法人電気通信普及財団および豊橋技術科学大学・高専連携教育研究プロジェクトの助成を受けて実施した。

#### 文献

- (1) 文部科学省:「平成27年度文部科学白書」(2016)
- (2) 沖縄県:「沖縄21世紀ビジョン基本計画【改定計画】」(2017)
- (3) 野村流師範 伊差川世瑞, 世礼国男:「声楽譜附工四上巻」, 野村流音楽協会発行 (1970)
- (4) 岡田恵美:「沖縄県の小学校高学年を対象とした郷土音楽学習の教材化に向けて」, 琉球大学教育学部紀要 (2015)
- (5) 山口喜七郎・石原 茜・仲村善信:「琉球古典音楽の唱法における発音・音声の研究 野村流「工四」における声楽譜, “大掛”について」, 琉球大学教育学部紀要 (2006-9)
- (6) 齋藤 毅, 辻 直也, 鶴木祐史・赤木正人:「歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析」, 日本音響学会誌, Vol.64, No.5, pp.405-417 (2008)
- (7) 高橋 純・津崎 実:「歌唱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), pp.47-63 (2017-3)
- (8) J. Sundberg: “Articulatory interpretation of the ‘singing formant’”, J. Acoust. Soc. Am., Vol.55, No.4, pp.838-844 (1974)
- (9) I. Nakayama: “Comparative studies on vocal expression in Japanese traditional and Western classical-style singing, using a common verse”, Proc. ICA, pp.1295-1296 (2004)
- (10) 齋藤 毅・辻直也・鶴木祐史・赤木正人:「歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析」, 音響学会誌, Vol.46, pp.405-417 (2008)
- (11) スンドベリ, ヨハン:「歌声の科学」榊原健一, 伊藤みか, 小西知子, 林良子訳, 東京電気大学出版局 (2007)
- (12) 佐久間雄輝・香山瑞恵・池田京子・大谷 真・橋本昌巳・伊東一典:「歌声の音響特徴量の比較に基づく歌唱スキル習熟度の考察」, The 28th Annual Conference of the Japanese Society for Artificial Intelligence (2014)
- (13) 山辺大貴・池田京子・香山瑞恵・大谷 真・橋本昌巳・伊東一典・山下泰樹・浅沼和志:「歌声の心理的印象と音響特徴量との対応付けによる歌唱の習熟度評価に関する基礎的検討」, 信学技報, 応用音響 EA, Vol.112, No.266, pp.61-66 (2012)
- (14) N. Cannings: “Deep Convolution Neural Networks for Dialect Classification

- (15) of Spectrogram Images”, Intelligent Voice, URL: www.intelligentvoice.com  
長濱嗣志・宮城 桂・山田親稔:「歌三線における技能伝承支援システムに関する検討」, 第23回知能メカトロワークショップ, 講演番号 3B1-2 (2018-09)
- (16) 長濱嗣志・上原一朗・宮城 桂・山田親稔:「ディープラーニングを用いた歌唱技能の可視化手法に関する検討」, 電気・情報関係学会九州支部連合大会, 03-1P-05 (2018-09)
- (17) T. Nagahama, K. Miyagi, and C. Yamada: “A Study on Effective Skill Passing Method for Uta-Sanshin”, Proc. The 8th International Workshop on Networking, Computing, Systems, and Software, NCSS2 Poster 220 (2018-11)
- (18) 板倉文忠:「音声の線形予測分析」, 電子情報通信学会技術研究報告, SP, 音声 98, pp.25-32 (1998-9)
- (19) 久保陽太郎:「ニューラルネットワークによる音声認識の進展」, 人工知能学会, Vol.31, No.2, pp.180-188 (2016-3)
- (20) NVIDIA: “DEEP CONVOLUTIONAL NEURAL NETWORKS FOR SPOKEN DIALECT CLASSIFICATION OF SPECTROGRAM IMAGES USING DIGITS”, NVIDIA Deep Learning Day (2016)
- (21) 足立 悠:「ソニー開発の Neural Network Console 入門」, リックテレコム (2018)

長 濱 嗣 志 (非会員) 2018年沖縄工業高等専門学校情報通信システム工学科卒業。現在、同専攻科創造工学システム専攻電子通信システム工学コース在学中。



上 原 一 朗 (非会員) 2015年4月沖縄工業高等専門学校情報通信システム工学科入学。現在、在学中。



宮 城 桂 (非会員) 2008年高知工科大学情報システム工学科卒業。2010年同大学大学院修士課程修了。2014年同大学大学院博士課程修了。同年沖縄工業高等専門学校情報通信システム工学科助教。現在に至る。博士(工学)。自己同期型回路を用いた超低消費電力VLSIの研究に従事。電子情報通信学会会員。



山田親稔 (正員) 2000年琉球大学大学院理工学研究科博士前期課程修了。2004年同大学大学院博士後期課程単位取得満期修了。同年拓殖大学北海道短期大学専任講師。2007年沖縄工業高等専門学校情報通信システム工学科助教。2009年同高等専門学校情報通信システム工学科准教授。2014年ビクトリア大学(カナダ)客員研究員。2015年より、沖縄工業高等専門学校情報通信システム工学科准教授。



現在に至る。博士(工学)。形式的設計検証, リコンフィギャラブルシステムの研究・教育に従事。IEEE, 電子情報通信学会, 各会員。

市川周一 (正員) 1985年東京大学理学部卒業。1987年同大学大学院理学系研究科修士課程修了。1987年新技術事業団創造科学推進事業(ERATO)研究員。1991年三菱電機(株)LSI研究所, システムLSI開発研究所勤務。1994年名古屋大学工学部助手。1997年豊橋技術科学大学工学部講師。同助教授, 准教授を経て, 2011年沼津工業高等専門学校制御情報工学科教授。2012年より, 豊橋技術科学大学大学院工学系研究科教授。現在に至る。理学博士。IEEE (seniormember), 電子情報通信学会(シニア会員), ACM, 情報処理学会, 各会員。



現在に至る。理学博士。IEEE (seniormember), 電子情報通信学会(シニア会員), ACM, 情報処理学会, 各会員。