Dynamic optimization method for personalized vocal training program driven by reinforcement learning

Yili Fang
[1]Department of Physical, Aesthetic & Labor Education, Zhejiang Shuren University, HangZhou 310000, Zhejiang, China
Corresponding author: Yili Fang (e-mail: Fangyili19951010@126.com).

ABSTRACT This paper proposes a personalized vocal training program dynamic optimization method based on reinforcement learning to address the problems of static solidification and poor adaptability of traditional vocal training programs, as well as the lack of dynamic optimization capabilities in existing intelligent vocal systems, making it difficult to achieve personalized teaching. Based on the theory of vocal music, audio signal processing technology, and deep reinforcement learning algorithms, an end-to-end cloud multi-source data acquisition architecture is built to preprocess and extract features from singing audio, behavioral, and physiological data. A high-dimensional continuous state space and a discrete continuous mixed action space are constructed. Design a multi-objective weighted reward function and temporal enhancement module with improved PPO algorithm as the core, to achieve real-time dynamic adjustment of training content, difficulty, duration, and rest strategy.

INDEX TERMS Learning driven; Vocal training; Reinforcement learning; PPO algorithm

I. INTRODUCTION

A. RESEARCH SIGNIFICANCE

The traditional vocal training model is centered around "one-on-one mentorship", which relies on the vocal teacher's years of accumulated singing experience, auditory discrimination ability, and teaching skills. It can provide real-time correction and guidance for students' pitch accuracy, rhythm, breath, resonance, pronunciation, timbre, and other issues, playing an irreplaceable role in professional talent cultivation. However, the traditional manual teaching model has significant limitations: firstly, high-quality vocal teacher resources are scarce and unevenly distributed, with concentrated teaching staff in first tier cities and professional colleges, making it difficult to match high-level teaching personnel in grassroots and remote areas, resulting in prominent supply-demand contradictions; Secondly, the cost of manual teaching is high, and the fees for one-on-one private lessons are generally high, which most ordinary vocal enthusiasts cannot afford in the long run; Thirdly, teaching efficiency is limited by manpower, and when a teacher faces multiple students at the same time, it is difficult to take into account each individual's differences, resulting in severe homogenization of training programs; Fourthly, teaching evaluation is subjective, and vocal proficiency evaluation relies on teachers' auditory perception, lacking quantitative standards. It is difficult to achieve objective and continuous tracking of training effects and progress evaluation; Fifth, the adaptability of static training programs is poor. Traditional teaching often adopts fixed class hours, fixed tracks, and fixed vocal training processes, which cannot adjust the training content in real time according to the daily status, ability fluctuations, and weak point changes of students.

Under the wave of digitalization, various intelligent vocal devices and online teaching systems have gradually been implemented, realizing basic functions such as audio collection, pitch detection, rhythm comparison, and waveform display, which to some extent make up for the shortcomings of traditional teaching. However, most existing intelligent vocal training systems adopt preset and static training schemes, which can only complete the basic work of "detection+result feedback" and do not have the ability to independently optimize, dynamically adapt, and plan for the long term. The training content, difficulty level, duration, and track selection built into the system are all pre-set and cannot be deeply adjusted based on the physiological characteristics, learning ability, progress speed, weak modules, and emotional state of the students. Some systems simply classify the difficulty according to the "beginner intermediate advanced" level, with a single grading standard, ignoring the multidimensional and uneven development of vocal ability: for example, some students have excellent pitch accuracy but weak breath, and some students have stable rhythm but insufficient resonance. A unified training plan will cause the problem of "repeated practice of advantageous modules and insufficient reinforcement of weak modules", greatly reducing training efficiency, and even causing students to

experience negative emotions such as fatigue and disinterest in learning due to unsuitable training content.

B. CURRENT RESEARCH STATUS AT HOME AND ABROAD

The research on intelligent vocal audio processing and singing recognition technology started earlier in foreign countries. Based on mature audio signal processing technology, countries such as Europe, America, Japan, and South Korea initially focused on basic technology research such as voice separation, pitch detection, and voice synthesis. In the field of singing evaluation, foreign scholars were the first to propose using Short Time Fourier Transform (STFT) and Mel Frequency Cepstral Coefficients (MFCC) to extract singing features and establish automated detection models for pitch, rhythm, and timbre. Some overseas teams have developed intelligent vocal practice software for amateur enthusiasts, which realizes functions such as singing segment comparison and error point annotation. However, due to differences in artistic cognition, foreign research tends to focus more on acoustic parameter analysis and rarely designs training programs based on vocal teaching rules, without involving dynamic optimization of training programs.

Domestic scholars have conducted extensive research on localized education scenarios, applying reinforcement learning to online question banks for primary and secondary schools, vocational education training, language learning, and other fields. For example, adjusting the English word memorization plan based on QLearning algorithm, optimizing the playback order and learning duration allocation of online courses based on deep reinforcement learning. However, looking at existing research, the application of reinforcement learning in the field of education is highly focused on theoretical knowledge learning, and there are very few studies on skill based training that combines body and perception in vocal, instrumental, sports, and other areas.

There is an essential difference between skill based training and cultural course learning: cultural courses mainly focus on memorizing knowledge points and logical reasoning, while state space and evaluation indicators are relatively simple; Vocal training is a continuous action skill, which includes multidimensional information such as acoustic characteristics, physiological status, fatigue status, and emotional status. The evaluation indicators combine quantitative acoustic parameters with subjective artistic expression. The action space includes multiple decision-making dimensions such as track selection, difficulty adjustment, training type, duration allocation, and rest strategy. The modeling difficulty is much higher than that of traditional knowledge learning. At present, there is no mature and complete dynamic optimization scheme for reinforcement learning vocal training, which is also the direction that this study needs to focus on breaking through.

C. RESEARCH CONTENT

Based on the research background and existing problems, this article sets seven core research contents, corresponding to the seven chapters of the entire text, forming a complete research system:

1. Sort out the theories of vocal training, reinforcement learning, and personalized education, complete the construction of the basic theoretical system, analyze the process characteristics, ability dimensions, and optimization needs of vocal training;
2. Design a data collection plan for the entire vocal training process, realizing multi-source data collection, preprocessing, and feature extraction of singing acoustic data, student behavior data, and physiological state data, and constructing a vocal state feature set;
3. Based on vocal training scenarios, complete the overall architecture design of the reinforcement learning model, define core elements such as intelligent agents, environment, state space, action space, and reward functions, and build a reinforcement learning framework that is suitable for vocal training scenarios;
4. In response to the dynamic optimization needs of vocal training, compare multiple reinforcement learning algorithms, complete the selection, improvement, and adaptation of core algorithms, and solve problems such as high-dimensional state space, multi-objective optimization, and continuous action decision-making;
5. Design a hierarchical and multi-dimensional personalized vocal training program system, clarify the execution rules of decision-making items such as training content, difficulty, duration, and rest strategies, and achieve the mapping of algorithm actions to actual training programs;
6. Build a simulation experimental platform and a testing experimental environment, design multiple control experiments, and verify the effectiveness of the model and scheme from dimensions such as optimization effect, training efficiency, personalized adaptability, and stability;
7. Summarize the research results of the entire text, analyze the shortcomings of the research, and propose the technology implementation path and future research directions based on industry development trends.

II. RELATED BASIC THEORIES AND TECHNOLOGIES

A. BASIC THEORY OF VOCAL TRAINING

Vocal music is an art form that uses the human vocal organs as instruments to perform through breath control, vocal cord vibration, resonance regulation, enunciation, and emotional expression. It is also a systematic and progressive skill training system. This section discusses from three aspects: vocal ability dimension, conventional training process, and training influencing factors.

Combining the professional knowledge of vocal teaching with the demand for automated evaluation, breaking down vocal singing ability into seven quantifiable dimensions is also the core basis for subsequent model state evaluation and reward function design:

1. Pitch accuracy: refers to the degree to which the singing pitch matches the standard melody pitch, which is the most fundamental ability in vocal music. Pitch deviation can be classified into small deviations, off key, and staccato, and can be quantified by the difference between the pitch curve of the audio and the pitch curve of the standard score.
2. Rhythm ability: The degree of conformity between singing rhythm, timing, pauses, and standard rhythm, including sub indicators such as speed stability, accuracy of rhythm segmentation, and control of pauses. Rhythm disorder is one of the most common problems for zero foundation students.
3. Breath control ability: Breath is the power source of vocal music, including inhalation depth, exhalation stability, breath endurance, and strength control. Insufficient breath can lead to weak sound, interrupted long notes, and abrupt changes in intensity.
4. Resonance ability: Utilizing resonance cavities such as the chest cavity, oral cavity, nasal cavity, and head cavity to amplify and modify the timbre, including resonance position, resonance uniformity, and timbre fullness, which directly determine the quality of the singing timbre.
5. Articulation ability: The core training content is the combination of articulation and vocalization, especially in ethnic vocal and bel canto vocal music, to control the clarity, rhyme, and spout of lyrics.
6. Tonal stability: The consistency of timbre within the same vocal range and different phrases, including characteristics such as vibrato amplitude, vibrato frequency, and changes in timbre brightness.
7. Singing proficiency: The familiarity of students with the repertoire and practice songs is reflected in behavioral data such as the number of interruptions, repetitions, and reaction time.

The seven dimensions of abilities are interrelated and jointly affect the overall singing level, and there are significant differences in the abilities of different students. This is also a key aspect that personalized training programs need to adapt to.

Professional vocal training follows a standardized process of "warm-up training $\rightarrow$ modular specialized training $\rightarrow$ segmented practice of songs $\rightarrow$ complete singing of songs $\rightarrow$ review and correction", which has three major characteristics of continuity, stages, and cycles:
1. Warm up and vocal training stage: The duration is generally 5–15 minutes, mainly focusing on practicing simple scales and vowels, activating the vocal organs, avoiding vocal cord damage, and gradually finding the state of vocalization. This stage has the lowest difficulty and mainly focuses on adaptive training.
2. Specialized training stage: Focused training is conducted on weak modules such as breath, intonation, rhythm, and resonance, which is the core stage of ability improvement. The training content and difficulty are determined based on the weaknesses of the students.
3. Track practice stage: divided into three levels: clause practice, segmented practice, and complete singing, with gradually increasing difficulty and the highest proportion of training time.
4. Review and adjustment stage: Refer to standard audio and teacher feedback, identify singing errors, record issues, and determine the focus of the next training session.

The traditional training process is fixed in time sequence and content. The core objective of this study is to dynamically adjust the training content, difficulty, duration, module order, and rest interval in the process based on reinforcement learning.

In addition to the inherent abilities of the learners, the training effectiveness is also influenced by various dynamic factors, which will serve as state variables of the reinforcement learning environment
1. Physiological state: Daily vocal cord fatigue, physical fatigue, respiratory status. Excessive fatigue can lead to a significant decline in singing ability, and continuing high-intensity training can easily cause damage to the vocal organs.
2. Emotional state: Positive emotions can enhance singing performance and training focus, while negative emotions can reduce training efficiency.
3. Training duration and intensity: A single training session that is too long or too intense can cause fatigue; If the duration is too short, the training effect cannot be achieved.
4. Training difficulty matching: Difficulty that is too high can easily undermine students' confidence, while difficulty that is too low can cause ability stagnation. Dynamic matching of difficulty is the core of optimization.
5. Historical training records: The effectiveness and changes in weak points of the trainees' training on the previous day and several times determine the direction of long-term training planning.

B. BASIC THEORY OF REINFORCEMENT LEARNING

Reinforcement learning is a trial and error autonomous learning framework, in which an agent adjusts its own strategy based on the reward signals feedback from the external environment through continuous interaction, ultimately achieving long-term cumulative reward maximization. This section introduces the core elements, basic framework, Markov decision process, and classical algorithm classification of reinforcement learning.

The standard reinforcement learning system consists of five core elements, which are also the basis for subsequent vocal model modeling:
1. Agent: The subject that executes decision-making actions. In this study, the agent is referred to as the "vocal training program decision-making system", responsible for outputting decision-making instructions such as training content, difficulty, and duration.
2. Environment: The external interactive object in which the intelligent agent is located. In this study, the environment is a "student+vocal training scenario", which receives the

agent's actions, generates new states, and provides feedback on reward values.

3. State (S): A complete description of the current situation in the environment, which serves as the basis for the agent to make decisions. The state in the vocal scene includes multidimensional information such as the student's singing ability, physiological state, fatigue level, and historical training data.

4. Action (A): The decision-making behavior executed by the agent based on the current state, and in this study, the action corresponds to various adjustment instructions of the training plan.

5. Reward (R): An immediate feedback signal from the environment after the agent performs an action, used to evaluate the quality of the action and guide the agent to optimize its strategy. The rewards are divided into positive rewards (good action effect), negative rewards (poor action effect), and zero rewards, as shown in Table 1 below.

TABLE 1. Multi-source data collection parameters

Data Type	Collection Device	Sampling Frequency	Frame/Window Parameter	Preprocessing Method
Vocal Audio	High-fidelity Microphone	44.1 kHz	Frame:20ms, Shift:10ms, Hamming Window	Spectral subtraction, Silence detection
Behavior Data	Smart Training Terminal	1 Hz	Timing statistics window: 1min	Linear interpolation, Logical verification
Physiological Data	Wearable Sensor	10 Hz	Sliding window: 5 sampling points	Moving average filtering, 3σ outlier elimination

The vocal training process satisfies Markov properties: the current training state (ability, fatigue, emotion) and the currently selected training plan (action) jointly determine the training state and training effect at the next moment, and therefore can be modeled based on MDP.

Traditional reinforcement learning: suitable for low dimensional state spaces, representative algorithms include QLearning, SARSA, Monte Carlo method, and policy gradient method. QLearning is an offline value iteration algorithm that maintains a Q-table to record the value of "state action" pairs. It has stable convergence, simple implementation, and is suitable for small-scale discrete decision-making scenarios.

Due to the high state dimension of vocal training and the fact that some decision items are continuous variables (such as training duration and difficulty coefficient), this study will focus on comparing deep reinforcement learning algorithms.

C. CORE TECHNOLOGIES FOR AUDIO SIGNAL PROCESSING

Vocal data is carried by audio signals, and audio signal processing is the basis for extracting singing features. This section introduces the core technologies used in the research.

Human voice belongs to non-stationary signals, and the overall frequency constantly changes over time, making it impossible to directly use the global Fourier transform. Short time Fourier transform adopts the idea of windowing and framing, dividing continuous audio into short frames. Assuming that the signal is stable in each frame, Fourier transform is performed frame by frame to obtain a two-dimensional time-frequency spectrum, which extracts pitch and spectral features. This study adopts Hamming window frame segmentation, with a frame length of 20ms and a frame shift of 10ms, balancing feature integrity and computational efficiency.

MFCC is an audio feature that simulates the auditory characteristics of the human ear, mapping the linear spectrum to the Mel nonlinear spectrum, which can effectively characterize the timbre, resonance, and vocal state of human voice. This study extracted 12 dimensional MFCC features and combined them with first-order and second-order differential features to construct a comprehensive vocal feature vector, as shown in Table 2.

TABLE 2. Feature composition of vocal training state set

Sub-state Category	Dimension	Feature Description
Singing Ability	12	Pitch, rhythm, breath, resonance, articulation features
Training Behavior	5	Training duration, interruption times, repetition times
Physical Fatigue	3	Heart rate, respiratory rate, vocal fatigue index
Historical Progress	2	Ability change trend, training completion rate
Total Original Features	42	Combined multi-dimensional features

III. MULTI SOURCE DATA COLLECTION AND FEATURE EXTRACTION FOR VOCAL TRAINING

A. OVERALL ARCHITECTURE DESIGN FOR MULTI-SOURCE DATA COLLECTION

Based on the actual scenario of vocal training, design a three-tier data collection architecture for edge cloud: the terminal layer is the collection device, including microphone, motion sensor, touch terminal, and camera; The edge layer serves as the local processing unit, responsible for real-time data collection, preliminary denoising, and caching; The cloud layer is responsible for data aggregation, deep preprocessing, feature storage, and model input. This architecture balances real-time performance (meeting the real-time adjustment requirements of dynamic solutions) and big data storage (meeting the offline training requirements of models).

According to the data source and semantics, the collected data is divided into three categories, which are also the core components of the subsequent state space:

1. Audio singing data: The core data source is collected by high fidelity microphones to extract singing ability features such as pitch accuracy, rhythm, breath, and timbre, including students' singing vocals and practice audio;

2. Student behavior data: collected from training terminals and operation logs, including training duration, track switch-

ing frequency, singing interruption frequency, repeated practice frequency, operation dwell time, etc., used to characterize proficiency, focus, and training preferences;
3. Physiological state data: The lightweight wearable sensor collects data such as heart rate, respiratory rate, and limb movement amplitude, and combines it with the vocal features in the audio to comprehensively judge the fatigue and physical state of the students.

Collection frequency setting: The audio data sampling rate is 44.1kHz, and after frame division, a frame feature is output every 10ms; Second level collection of behavioral data; The sampling frequency of physiological data is 10Hz to ensure the continuity of state perception, as shown in Table 3 below.

TABLE 3. Hyper-parameters of improved PPO algorithm

Hyper-parameter	Value	Parameter Description
Actor learning rate	3×10^{-4}	Learning rate of policy network
Critic learning rate	3×10^{-4}	Learning rate of value network
Batch size	64	Sample batch size for training
Update times per iteration	10	Network update rounds in single iteration
Discount factor (γ)	0.92	Discount coefficient for cumulative reward
PPO clip coefficient (ϵ)	0.2	Threshold for policy update limit
Total training episodes	10000	Total rounds of offline training

B. PREPROCESSING OF TYPED RAW DATA

The original audio data contains environmental noise, device background noise, and background noise. Direct use will significantly reduce feature accuracy. The preprocessing steps are as follows:
Step 1: Audio framing and windowing. Using Hamming window to frame the entire audio segment, with a frame length of 20ms and a frame shift of 10ms, to ensure partial overlap between frames and avoid feature breakage.
Step 2: Noise estimation and denoising. Select the silent segments before the start of training as noise samples, use spectral subtraction to remove noise components frame by frame, and preserve the pure human voice signal.
Step 3: Silent segment detection. Distinguish between vocal and silent segments based on short-term energy and zero crossing rate, eliminate blank frames without singing content, and reduce invalid data calculations.
Step 4: Abnormal Audio Removal: Identify abnormal audio segments such as broken sounds, sudden noise, and equipment malfunctions, directly remove them, and do not participate in subsequent feature extraction, as shown in Table 4.

TABLE 4. Network structure of improved PPO model

Network Module	Activation Function	Function
Shared Feature Layer	ReLU	Feature fusion
Actor-Discrete Branch	Softmax	Output discrete actions
Actor-Continuous Branch	Tanh	Output continuous actions
Critic Value Layer	ReLU	State value evaluation

Behavioral data is structured log data, and the main issues are data loss and logical anomalies. Preprocessing process:
1. Missing value handling: Single behavior data missing due to terminal lag or network interruption is filled with temporal linear interpolation; Multiple consecutive missing segments are marked as invalid training segments.
2. Logical verification: Set rules to verify abnormal behavior, such as obvious abnormal data such as "single training duration exceeding 3 hours" and "track switching exceeding 10 times within one minute", which are judged as misoperations and removed.
3. Structured data organization: Convert unstructured logs into standardized time series tables, sort them by training time, and establish corresponding relationships for "time behavior indicators".

The heart rate and respiratory rate data collected by physiological sensors are susceptible to pulse noise caused by limb shaking. The preprocessing uses sliding average filtering: set the window size to 5 sampling points, replace the current sampling point value with the average value within the window, and smooth the curve. At the same time, a reasonable threshold for physiological indicators is set, and data that exceeds the normal range of the human body is judged as a collection failure and removed, as shown in Table 5.

TABLE 5. Experimental grouping and research objects

Group Number	Scheme Type	Object Classification	Total Training Cycle
Control Group 1	Static training scheme	Beginner, Intermediate, Exam student	4 weeks
Control Group 2	Rule-driven dynamic scheme	Beginner, Intermediate, Exam student	4 weeks
Experimental Group	RL-based optimization scheme	Beginner, Intermediate, Exam student	4 weeks

C. MULTI DIMENSIONAL FEATURE EXTRACTION

Based on preprocessed raw data, quantifiable features are extracted into three categories, all of which are continuous numerical types, providing support for high-dimensional state space modeling.

The audio features are divided into two modules: basic acoustic features and vocal professional features, totaling 28 dimensional features:
1. Basic acoustic characteristics
Short term energy: characterizes the loudness of sound, the strength of breath, and reflects the ability to control breath;
Short time zero crossing rate: Distinguish between human voice and noise, assist in determining the clarity of pronunciation;
Fundamental frequency (pitch) sequence: Extract f_0 frame by frame, calculate the absolute deviation, average deviation, and maximum deviation from the standard pitch, corresponding to pitch accuracy;

MFCC and differential features: 12 dimensional MFCC+12 dimensional first-order differential MFCC, representing timbre and resonance characteristics.
2. Derivative features of vocal music major
Rhythm deviation characteristics: calculate the time difference and average rhythm error between the singing beat and the standard beat;
Stability characteristics of long notes: The amplitude of fundamental frequency fluctuations and energy fluctuations within long note segments reflect the endurance of breath;
Tremol characteristics: Tremol frequency and amplitude, representing singing skills and timbre stability.

Extract 8-dimensional behavioral features from the training log, all of which are temporal statistical features: total duration of a single training session, effective singing duration, number of singing interruptions, number of repeated practice sessions, number of track switches, average number of repeated sessions per session, terminal operation response time, and rest interval time. This set of features directly reflects the proficiency, focus, and training habits of the students.

IV. OVERALL MODELING OF REINFORCEMENT LEARNING MODELS IN VOCAL SCENES

A. OVERVIEW OF THE OVERALL ARCHITECTURE OF THE MODEL

The reinforcement learning model designed in this study is divided into five layers, from bottom to top: data input layer, state perception layer, decision agent layer, action execution layer, and environmental feedback layer, forming a closed-loop interactive system
1. Data input layer: Receive the 22 dimensional standardized state features output from Chapter 3, complete data format conversion, and input them to the state perception layer;
2. State perception layer: Fuse and analyze multidimensional features to identify the current singing ability, fatigue state, weak modules, and training progress of students, and output standardized system states;
3. Decision agent layer: Model core, executes reinforcement learning decisions based on the current state, outputs training plan adjustment actions;
4. Action execution layer: Map the abstract actions output by the intelligent agent into concrete vocal training plans (songs, difficulty, duration, rest, etc.), and distribute them to the training terminal for execution;
5. Environmental feedback layer: Collect singing data, behavioral data, and physiological data of students after implementing the new training program, calculate reward values and new states, and transmit them back to the input end to complete a round of interactive closed-loop.

The system uses "training units" as the interaction cycle, with one training unit defined as 10 minutes of continuous training+state evaluation, and the cycle is set to fit the actual rhythm of vocal training.

B. STATE SPACE MODELING

Corresponding to the seven core competencies of vocal music, including quantitative indicators such as pitch deviation, rhythm error, breath stability, resonance characteristics, articulation characteristics, and timbre stability. The closer the value is to 0, the weaker the ability, and the closer it is to 1, the stronger the ability. This subset is used to determine the current skill level and weak modules of the trainees, and is the main basis for adjusting the training content and difficulty.

Composed of behavioral characteristics, including indicators such as training focus, track proficiency, and practice habits, it is used to determine the acceptance and familiarity of students with the current training content.

Composed of heart rate, breathing, and vocal fatigue index, the higher the value, the higher the fatigue level. The model adjusts the training intensity and rest duration based on this subset to avoid overtraining.

Statistically analyze the trend of ability changes and training completion of nearly 5 training units for long-term planning of the model, in order to avoid shortsightedness in single step decision-making.

The combination of four sub states forms a complete global state, and the state space is a continuous high-dimensional space. Therefore, this study excludes traditional tabular QLearning and chooses deep reinforcement learning algorithms.

C. MODELING OF STATE TRANSITION RULES

The state transition in vocal scenes follows clear patterns:
If the training content that matches weak modules is selected and the difficulty is moderate, the singing ability sub state will develop towards excellence (numerical improvement);
If the training difficulty is too high and the duration is too long, the physiological fatigue state will significantly increase, and the singing ability will slightly decrease;
If low difficulty content is used for a long time, the singing ability sub state tends to stabilize and no longer improves;
4. An increase in rest time will directly reduce fatigue status and have no direct negative impact on ability status.

In the offline simulation stage of the model, a probabilistic state transition model is constructed based on the rules of vocal teaching; In the real online operation stage, the state transition is completely determined by the actual training data of the students, achieving real environment interaction.

V. DESIGN AND IMPROVEMENT OF REINFORCEMENT LEARNING ALGORITHM FOR VOCAL SCHEME OPTIMIZATION

A. COMPARISON AND SELECTION OF CLASSIC REINFORCEMENT LEARNING ALGORITHMS

Comparing the adaptability of mainstream deep reinforcement learning algorithms for the scenario of high-dimensional continuous state+discrete continuous mixed action in this study:

DQN: Only supports discrete action spaces and cannot handle continuous actions such as difficulty and duration, so they can be directly excluded;
DDPG: a classic algorithm for continuous action spaces, using Actor Critic architecture with moderate convergence speed, suitable for continuous decision-making, but lacks support for mixed actions and is prone to unstable training;
3. PPO (Near End Policy Optimization): The current mainstream algorithm in the industry supports both continuous and discrete actions, with strong training stability and high robustness. It can flexibly adapt to mixed action spaces, with simple hyperparameter tuning and high fault tolerance.

After comprehensive comparison, this study selected PPO algorithm as the basic algorithm and improved the algorithm for the multi-objective and long-term characteristics of vocal training scenarios.

B. BASIC PPO ALGORITHM PRINCIPLE OVERVIEW

PPO belongs to the policy gradient algorithm, and its core idea is to limit the magnitude of policy updates to avoid policy crashes caused by a single update being too large. Algorithms are divided into Actor networks (policy networks) and Critic networks (value networks):
Actor network: Input current state, output action probability distribution/continuous action values, responsible for decision-making;
Critic network: Input the current state, evaluate the value of the state, and guide the Actor network to update.

PPO prunes the objective function to constrain the differences between old and new strategies, while ensuring update efficiency and significantly improving training stability, making it very suitable for intelligent training systems that require long-term stable operation.

C. NEURAL NETWORK STRUCTURE DESIGN

Based on the improved PPO algorithm, a dual branch neural network structure is designed, which is uniformly built using fully connected layers and adapted to 22 dimensional input states
1. Shared feature extraction layer: A 3-layer fully connected network with 128, 64, and 32 neurons, respectively, and an activation function of ReLU, to complete high-dimensional state feature fusion and extraction.
2. Actor Strategy Network Branch:
Discrete action branch: fully connected layer+Softmax, outputting probabilities for 7 types of training content and 3 types of training modes;
Continuous action branch: fully connected layer+Tanh, outputting three types of continuous action values: difficulty, duration, and rest interval.
3. Critic Value Network Branch: A 2-layer fully connected network that outputs the estimated value of the current state, with an activation function of ReLU.

The overall network structure is lightweight, balancing computing speed and inference accuracy, meeting the requirements of real-time inference for terminals.

Real time monitoring of the average cumulative reward and state value during the training process: in the first 2000 rounds of training, the agent is in the exploration stage and the reward fluctuates greatly; 2000~6000 rounds, rapid strategy optimization, continuous increase in accumulated rewards; After 6000 rounds, the reward curve tends to stabilize and the algorithm completes convergence.

The converged model can stably output a reasonable training plan based on the student's state, achieve a balance between exploration and utilization, and achieve the expected effect through offline training.

VI. CONCLUSION

This article focuses on the dynamic optimization requirements of personalized vocal training programs, completing the construction of reinforcement learning models, algorithm improvements, data processing, and multiple sets of validation experiments, forming a complete intelligent training optimization system. We have developed a multidimensional quantitative evaluation index and MDP modeling framework for vocal music, which solves the problems of high-dimensional vocal state, mixed movements, and subjective evaluation; Based on the PPO algorithm, the decoupling of mixed actions, dynamic reward weights, and temporal feature enhancement were achieved, which improved the model's decision stability and long-term planning ability. The actual test results confirm that this system can adaptively adjust the training plan based on individual differences and real-time status of students, effectively avoiding overtraining problems and improving training effectiveness in all aspects. At the same time, there are limitations to the research, as it has not yet included the emotional depth characteristics of students and their adaptation to multilingual singing scenes. In the future, the guidance function for artistic emotional expression can be expanded by combining large models, optimizing lightweight models to adapt to mobile devices, and further promoting the practical application of reinforcement learning in the fields of art education and skill training.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

REFERENCES

[1] T. Lan, "Integrating Chinese Operatic Vocal Techniques Into Contemporary Commercial Music Pedagogy: The Lan Tianyang Cross-System Vocal Training Model", Journal of voice: official journal of the Voice Foundation, 2026, doi: 10.1016/J.JVOICE.2026.04.011.
[2] L. S. Thai, D. M. Anh, P. T. Hieu, et al., "Research and development of VR system for training vocal music in Vietnam", Computers & Education: X Reality, vol. 8, Art. no. 100154, 2026, doi: 10.1016/J.CEXR.2026.100154.

[3] Y. Li and J. Liang, "Managing Musical Knowledge Through Virtual Reality: A Case Study of Vocal Training Transformation", International Journal of Knowledge Management (IJKM), vol. 21, no. 1, pp. 1–18, 2025, doi: 10.4018/IJKM.392408.
[4] Y. Xie, "Designing music education curriculum in a multicultural environment: Vocal training in China", International Journal of Music Education, vol. 43, no. 3, pp. 356–371, 2025, doi: 10.1177/02557614231221144.
[5] L. Zong, "Evaluation on the Effect of Enhancing Vocal Music Training Experience with Virtual Reality Technology", International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), vol. 20, no. 1, pp. 1–20, 2025, doi: 10.4018/IJWLTT.382590.
[6] M. Gai, "Implementation of an Innovative Approach to Vocal Training in College: The Case of Artificial Intelligence Technologies NSynth, Sing Like Me, and Flow Machines", The Asia-Pacific Education Researcher, vol. 34, no. 5, pp. 1–11, 2025, doi: 10.1007/S40299-025-01003-Y.
[7] Z. Chen and J. Tao, "Exploring the Effects of Vocal Training Programs on the Skills and Development of Chorus Performers", Journal of voice: official journal of the Voice Foundation, 2025, doi: 10.1016/J.JVOICE.2025.01.016.
[8] S. Prados-Bravo, D. González-Rodríguez, and A. Rodríguez-Esteban, "Evaluating the Inclusion of Vocal Training in Spain's Teacher Education: A Quantitative Analysis", Education Sciences, vol. 14, no. 12, p. 1358, 2024, doi: 10.3390/EDUCSCI14121358.
[9] Y. Shi, "The use of mobile internet platforms and applications in vocal training: synergy of technological and pedagogical solutions", Interactive Learning Environments, vol. 31, no. 6, pp. 3780–3791, 2023, doi: 10.1080/10494820.2021.1943456.
[10] D. Du, "Retraction notice to 'Using mobile apps in vocal training to develop creative thinking' [Thinking Skills and Creativity Volume 44, June 2022, 101020]", Thinking Skills and Creativity, vol. 46, Art. no. 101167, 2022, doi: 10.1016/J.TSC.2022.101167.
[11] S. Doganyigit and O. F. Islim, "Virtual reality in vocal training: a case study", Music Education Research, vol. 23, no. 3, pp. 391–401, 2021, doi: 10.1080/14613808.2021.1879035.
[12] S. Graf, A. Bungenstock, L. Richter, et al., "Acoustically Induced Vocal Training for Individuals With Impaired Hearing", Journal of voice: official journal of the Voice Foundation, vol. 37, no. 3, pp. 374–381, 2021, doi: 10.1016/J.JVOICE.2021.01.020.
[13] "FORECASTING METHOD OF VOCAL TRAINING PREFERENCE BASED ON BEHAVIORAL PSYCHOLOGY", vol. 33, no. S6, pp. 93–0, 2021.
[14] Y. Z., "Application of vocal training in the treatment of voice diseases of music majors", BASIC & CLINICAL PHARMACOLOGY & TOXICOLOGY, vol. 127, p. 215, 2020.
[15] E. Babacan, A. Özçimen, and H. C. Kavşat, "The Problem of Voice Break at Register Transition During Vocal Training and The Approaches Towards The Solution of This Problem", Art Vision, vol. 26, no. 45, pp. 463–475, 2020, doi: 10.32547/ATAUNIGSED.645346.
[16] X. S., "The Reform and Innovation of Vocality Training Mode in Colleges and Universities Under the Diversified Teaching Mode", BASIC & CLINICAL PHARMACOLOGY & TOXICOLOGY, vol. 126, pp. 319–320, 2020.
[17] X. Yang, E. D. Sadika, G. B. Pratama, et al., "An Analysis on Serious Games and Stakeholders' Needs for Vocal Training Game Development", Communication Sciences & Disorders, vol. 24, no. 3, pp. 800–813, 2019, doi: 10.12963/csd.18531.
[18] Voixtek Vr LLC, "Patent Application Titled 'Interactive Training Tool For Use In Vocal Training' Published Online (USPTO 20190266914)", Medical Patent Business Week, 2019.
[19] H. Kawahara, K. Sakakibara, E. Haneishi, et al., "Real-time and interactive tools for vocal training based on an analytic signal with a cosine series envelope", CoRR, abs/1909.03650, 2019.
[20] T. Bano, "Michael Blakemore: Young actors neglect vocal training", The Stage, no. 11, p. 5, 2017.