Application of Information Teaching in Vocal Music Assisted Practice in Mobile Internet Era

Shengnan Liu
Xu Wang (wangxu1550@163.com)
School of Education of Shandong Women's University

Research Article

Keywords: CNN, Data Fusion, Mobile Internet, Vocal Music

Posted Date: August 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1935020/v1

Application of Information Teaching in Vocal Music Assisted Practice in Mobile Internet Era

Shengnan Liu[1,a] Xu Wang[1,b*]

[1]*School of Education of Shandong Women's University, Jinan 250300, Shandong, China*
[a]*Email: 307028435@qq.com*
[b*]*Email: wangxu1550@163.com*
[*]*Corresponding Author*

Abstract

Vocal music practices using the internet and mobile computing rely on the distinguishable pitch, note, and musical library information for ease of understanding. For addressing this problem an Assisted Data Processing Model for Tutor Function (ADPM-TF) is projected in this article. The proposed model learns the user requirements and their improvements based on training hours. The learned inputs are validated for the mishandled tutor sessions for identifying errors. The tutor assistance based on independent and joint (fused) data is recommended by the neural network learning based on previous session user performance/ rating. This process is streamlined depending on the user preference and switching ability over various sessions. The performance is validated using the metrics assistance ratio, error, processing time, and fusion rate.

Keywords—CNN, Data Fusion, Mobile Internet, Vocal Music

Introduction

Vocal music is one of the genres of music that use human voice and tones as a central part. Vocal music is nothing but singing that use instruments to improvise the song. Vocal music practices such as bass, tenor, alto, and soprano are the main vocal practices that are performed by singers [1]. Online vocal music practices improve the performance rate of singers which reduces the energy consumption rate is unnecessary practices. Internet technology provides various sets of services and interactions that enhance the efficiency of an application. Internet technologies are used for vocal music practices that provide instructions and advice to the students [2]. Internet technology motivates students and improves their skills of students. Internet technology provides a certain set of services to the teacher that helps to differentiate the meaning among practice notes [3]. Internet technology provides various tools and applications to get vocal practices. Internet technology also provides feedbacks service that collects learners' points over practices [4, 5].

Vocal music information processing is a process that records music with a certain set of features and functionalities. Information processing uses both hardware and software devices to manage and manipulate vocal music into sound signals [6]. Information processing is also used to identify the features that are available in vocal music. The music information process improves the security level and reduces the error rate in the vocal music education system. A multimodal algorithm is used for vocal music information processing that retrieves vocal-related information for an application [7, 8]. Multimodal information classifies the music information for the data analysis process. The fusion approach is used here to identify the features of music information that are extracted by the classification process. The fusion approach categorizes the information and provides necessary data for the further process [9]. Music information processing systems provide an appropriate set of data for the vocal music management process. The humdrum toolkit is used for music information processing that provides various sets of training and lesson sessions for the music students. Humdrum finds out the necessary data for data processing that enhances the efficiency of the system [10].

Machine learning (ML) techniques are widely used in various fields for detection, prediction, and analysis process. ML techniques improve the accuracy rate in the detection process which enhances the performance of an application [11]. ML technique is a type of Artificial Intelligence (AI) system that is mainly used for the prediction process. ML techniques first train the given dataset and identify the abnormal data for the further analysis process. ML techniques and algorithms are used for music data processing that requires an accurate set of data for the analysis process [12]. A Gaussian algorithm is used for music data processing systems that use cloud computing to train datasets. The Gaussian algorithm reduces the error rate in the prediction process which improves the performance rate in a data processing system [13]. ML-based hidden Markov model is used for music data processing systems. Markov model identifies the necessary data by using the feature extraction process. The feature extraction method is used to extract the important features that are available in a given set of data. Markov model improves the effectiveness and reliability of the music data processing system [14, 15].

Related Works

Yu et al. [16] introduced a hierarchical attention network for singing melody extraction (HANME). A convolutional recurrent neural network (CRNN) is used here to extract the important features that are available in singing melody extraction. CRNN extracts the attention-aware features from the database. Temporal content vectors are extracted by using the long-term features extraction process. The proposed HANME method improves the efficiency and feasibility of the system.

Sharma et al. [17] proposed a singing-adapted short-term objective intelligibility (STOI) based on automatic evaluation of song intelligibility. The proposed method is mainly used to find out the noisy speech and notes that are available in a song. Voice-specific features are used here to identify the intelligibility of singing vocals. STOI improves the performance rate and efficiency rate in the evaluation process. The proposed evaluation method maximizes the accuracy rate in an evaluation process that enhances the effectiveness of the system.

Nakatsuka et al. [18] introduced a compatibility estimation method for vocal and accompaniment in audio tracks. Joint embedding techniques and self-supervised techniques are used for the estimation process. The joint embedding technique first trains the dataset of both vocal and accompaniment that provides an appropriate set of data for the analysis process. Embedded features are extracted by using the joint embedding technique. The proposed method achieves a high accuracy rate in the evaluation process that improves the performance of the system.

Hongdan et al. [19] proposed a deep learning approach-based intelligent music genre analysis method. The feature extraction process is used here that extract the important features that are presented in audio tracks. A classification technique is also used here that uses the data which are extracted from the extraction process. The classification process provides the necessary set of data for the music genre analysis process. The proposed method maximizes the accuracy rate in the classification process which enhances the feasibility of the system.

Mori et al. [20] introduced a decoding method that decodes peak emotional responses to the music. Machine learning (ML) approaches are used here to decode the emotions of users. The computational acoustic technique is used in the decoding process based on a certain set of emotional features. Features such as tears, anger, happiness, and chill are identified by the computational acoustics technique. The proposed method achieves a high accuracy rate in the decoding process that provides an actual set of data for further process.

Castillo et al. [21] proposed a new analysis method for hip-hop music. The proposed method is mainly used to identify the necessary set of data from hip-hop music. The analysis method produces an optimal set of data for the designing and implementation process. The proposed method provides an appropriate set of data for sound engineers. Sound engineers design hip-hop music that enhances the efficiency and reliability of the system.

Sassi et al. [22] introduced a multi-criteria decision-making method for an online music recommendation system. The aggregation technique is used here to identify the relationship between ratings and contexts. K-means algorithm-based predictive model is used for recommendation systems that find out the genres of music. Contextual information provides a necessary set of data for the decision-making process that increases the accuracy rate of the decision-making process. The proposed method improves the effectiveness and feasibility of the system.

Mirbeygi et al. [23] proposed a robust principal component analysis (RPCA) based music separation method. A random singular value decomposition algorithm is used here to identify the key values of the optimization process that produce a feasible set of data. The proposed method reduces the computation cost and time which reduces the complexity rate of the system. The proposed RPCA-based method improves the quality and feasibility of the music separation method.

Wang et al. [24] introduced a music auto-tagging method based on machine learning (ML) techniques. Convolutional neural network (CNN) and recurrent neural network (RNN) approaches are used here to find out the characteristics of both lyrics and music. The proposed method used a classification process to identify the relationship between audio and lyrics of a song. CNN extracts the important set of data that is necessary for the auto-tagging process. The proposed method improves the performance rate and feasibility of the system.

Mavaddati et al. [25] proposed a new voice separation method based on the learnable decomposition technique. The decomposition method is used here to find out the singing vocal signals of a song that produce an appropriate set of data for the further analysis process. Both vocal and non-vocal notes are separated by the voice separation method. The proposed method achieves a high accuracy rate in the voice separation process which improves the effectiveness of the system.

He et al. [26] introduced an improved wavelet model for sound production image processing systems. The proposed model is mainly used to identify the cracks and damages that are presented in sound images. High-quality binary images are used here for the identification process that provides actual details about images. The proposed wavelet model reduces the latency rate in the classification process which enhances the compatibility of the system. The proposed method improves the efficiency and reliability of the sound production image processing system.

Chen et al. [27] proposed a lightweight multi-stage music separation network by using a convolutional neural network (CNN). Hybrid convolution-based CNN is implemented here to perform the feature extraction process. Important features and patterns are identified by the feature extraction process. The proposed method improves the performance rate and effectiveness rate of the system. When compared with other traditional methods, the proposed method achieves a high accuracy rate in the music separation process.

Lai et al. [28] introduced a new robust principal component analysis (RPCA) method for the monoaural voice separation process. A recurrent neural network (RNN) algorithm is used for the RPCA method that identifies the important patterns and notes of the song. RNN trains the extracted set of data that improves the performance and efficiency of the system. The proposed method reduces the computational cost which enhances the feasibility of the system.

Assisted Data Processing Model for Tutor Function

The distinguishable features of note, pitch, music library information, and musical object observation are used to provide better vocal music teaching to the students that are accessed based on assimilated music data. With the students' memory analysis, the acquired data processing is used to retard multiple teaching based on fusion process and learns the user requirement information(Refer to Fig. 1).

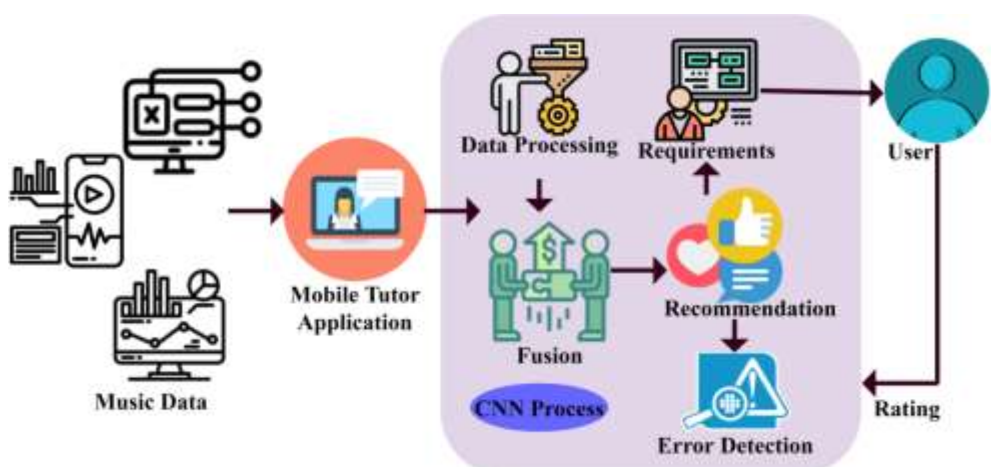

Fig. 1 Proposed ADPM-TF

The proposed tutor function model provide feasible teaching to the students based on adequate processing, data accumulations, and the fusion method of student performance analysis at different instance. At the time of information teaching in vocal music assisted practice to the students is analyzed for improving the user preference and switching ability over various sessions. The identified error in data processing and fusion are jointly classified using conventional neural network learning. This learning is to address the errors in vocal music teaching in mobile internet applications due to input musical data accessing and fusion processing improving the student interest in vocal music practices. Massive vocal music teaching through mobile applications of tutor assistance to the students based on independent and joint (fused) data is recommended by the neural network based on the previous session's user rating/performance. The students were given feedback for the vocal music tutor and their teaching based on ratings or performance. The user rating helps to analyze the best tutor on the mobile internet through the CNN process. The data processing and fusion are accessed continuously by identifying mishandled tutor sessions with previous user preferences and ratings for the available tutor's performance based on vocal music teaching.

In this CNN process, the identifying errors are suppressed for improving the recommendation ratio in vocal music teaching using the assistance availability in adequate processing retards the multiple teaching hours. The proposed model is to recommend vocal music teaching to the students between independent and fused data based on tutor assistance. The error identification performs until detecting the mishandled tutor sessions in the previous sessions' user preference and it reduces the processing time. The input musical data is processed through digital multimedia applications for the students for better vocal musical data learning.

Training Assessment

The vocal music practices rely on pitch, note, musical library information and objects using mobile internet is responsible for teaching. The vocal music data accessed through mobile tutor applications is based on acquiring data processing. The input musical data processing has been accepted by students and teachers gradually and the multiple data fusion has been massively used in vocal music teaching practices related to representation, practical teaching, tutor assistance, etc. In vocal music teaching practices, the mobile tutor applications can be used by students for learning vocal music data $[V_M(d)]$ is given as

$$V_M(d) = \frac{MT^{app} \times \left[D_{p_{max}} - D_{p_{min}}\right]}{MT^{app}} \tag{1}$$

Instead,

$$T_a = \frac{1}{\sqrt{2\pi}}\left[\frac{\left(\frac{D_{p_{min}}}{D_{p_{max}}}-\frac{MT^{app}}{D_p}\right)}{2[V_M(d)-S]}\right] \tag{2}$$

The tutor assistance is computed by the number of vocal music practices observed at various D_p and MT^{app} data processing and fusion through mobile internet applications. In this case, some errors occurred in vocal music teaching due to independent and joint (fused) data for the mishandled tutor session errors in D_p and MT^{app}. The training assessment process is illustrated in Fig. 2.

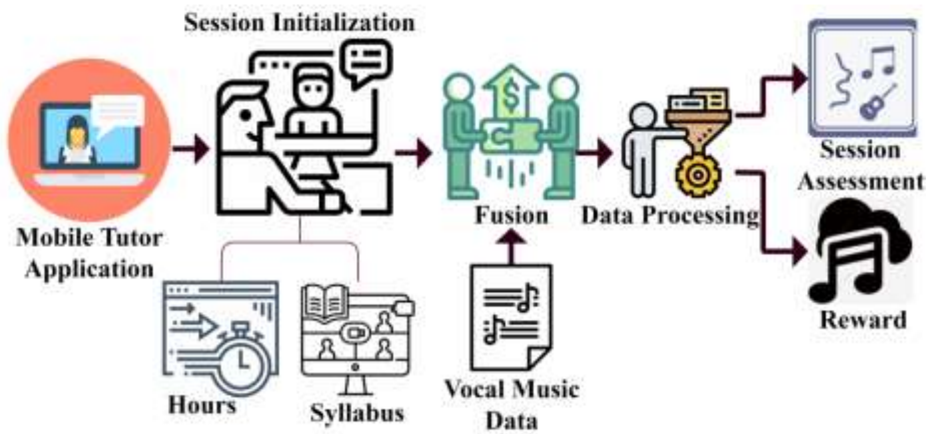

Fig. 2 Training Assessment Process

The mobile tutor application discloses session initialization as hours and syllabus. The required data is fused for the data processing feature providing session assessment and rewards. This session assessment improves the student interests and the rewards are used for promoting the tutor improvements (Refer to Fig. 2). Therefore, the identified errors impact the vocal music teaching in different teaching practices instance, for which the teaching of vocal music is computed as shown in Equations (3) and (4)

$$t[V_M(d)] = \frac{A_p}{\left(\frac{D_{p_{min}}}{D_{p_{max}}}-F_d\right)^2} \tag{3}$$

Such that,

$$F_d = \sum_{i=1}^{S}\left(\frac{V_M(d)-S}{T_a}\right)^{D_p} \tag{4}$$

From the above equation (3) and (4), the variable $t[V_M(d)]$ represents the teaching of vocal music for students based on acquired data processing following the maximum adequacy processing A_p and the multiple data processing (fusion) observation F_d. Here F_d is a fusion for learning input musical data whereas A_p is the assistance availability. For which the teaching practices, based on user requirements, data accumulations, and training hours depend on various S are required. The proposed model is based on the vocal music teaching and $t[V_M(d)]$, the data processing and fusion of vocal music practices are performed through user requirements is given as follows

$$T_a\left[S, t[V_M(d)]\right] = \sqrt{\left[\frac{t[V_M(d)]}{S}\right]_1^2 + \left[\frac{t[V_M(d)]}{S}\right]_2^2 + \cdots + \left[\left(1 - \frac{S}{V_M(d)}\right) A_p\right]_i^2} \; , S \in D_p \tag{5}$$

The vocal music practices based on the user requirements, representation, practical teaching, and tutor assistance are performed using the mobile tutor applications.

Based on the mobile tutor application and data processing based on the conventional neural network process, the music data is shared with the mobile tutor application for teaching the vocal music information to the students. The outcome of the CNN process is to pursue the data processing and fusion concurrently and identify the mishandled tutor sessions. The user requirements depend on adequacy in processing and D_p observation for vocal music teaching instances through $t[V_M(d)]$ implication and S based teaching practices to the students. The first step of the CNN process is to identify the errors based on data processing and fused instance and $t[V_M(d)]$ is processed. In this proposal, two consecutive processing of error identification at different time intervals say E_H and E_D are serving as the input music data for the independent and joint data processing and fusion classification process. Based on the computation, the errors in data processing and fusion are jointly classified and it is modeled as

$$\left.\begin{array}{c} \left.\begin{array}{r} E_H = ob \\ E_D = 0 \end{array}\right\}, as\ the\ first\ \ vocal\ music\ teaching \\ and, \\ \left.\begin{array}{r} E_H = t[V_M(d)] \\ E_D = \frac{MT^{app}}{A_p} \end{array}\right\}, as\ the\ sequential\ vocal\ music\ teaching \end{array}\right\} \tag{6}$$

Similarly,

$$\left.\begin{array}{c} E_H + E_D = S, as\ the\ first\ teahing\ practice \\ where, \\ E_H + E_D = t[V_M(d)] + \frac{MT^{app}}{A_p}, as\ the\ sequential\ training\ practices \end{array}\right\} \tag{7}$$

Therefore, in the joint classification process the sequential vocal music teaching practice of $E_H + E_D = t[V_M(d)] + \frac{MT^{app}}{A_p}$ is handled for identifying mishandled tutor sessions.

The tutor's assistance is based on data processing and fusion observation instances. The data processing and fusion analysis depends on T_a assessment and its joint classification process of $J_c\{T_a[S, t[V_M(d)]]\}$ is performed by the CNN process is given as

$$J_c\{T_a[S, t[V_M(d)]]\} = MT^{app} - U_R * S * E_D \tag{8}$$

Where,

$$\left.\begin{array}{c} E_H(U_R|S) = R[S + (U_R * E_D)] \\ and \\ d_a(S|A_p) = R[(U_R * E_D)] \end{array}\right\} \tag{9}$$

Based on the equation (8) and (9), the variables U_R and E_D represent the recommendation based on user requirements and recommendation based on error detection. This recommendation R is the output based on error detection and user requirement observation in vocal music teaching (i.e.) E_H, E_D with user ratings based on tutor assistance performance and teaching practices. Fig. 3 presents the learning process representation for first and sequential sessions.

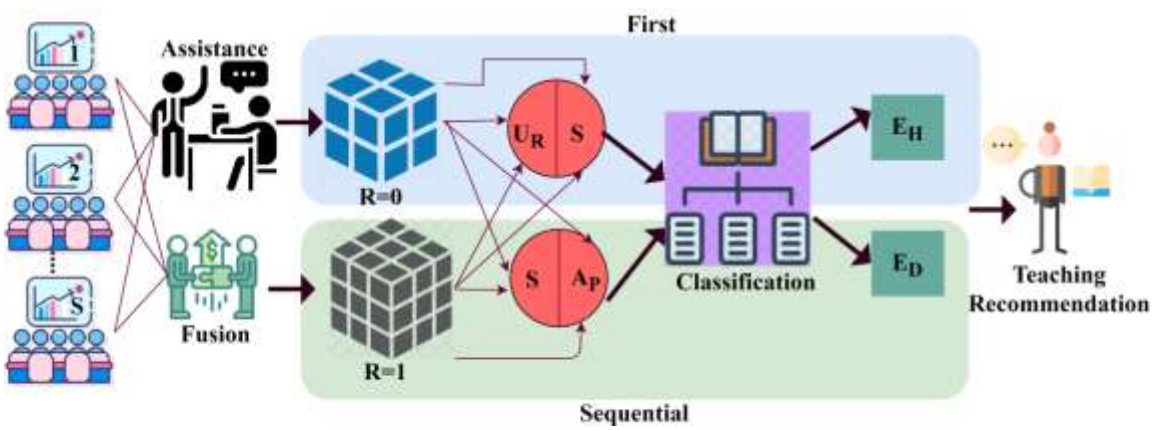

Fig. 3 Learning Process for First and Sequential Sessions

The session inputs are provided with A_p and $F_d \forall V_M(d)$ such that R classification is performed. If $R = 0, then$ $[U_R|$ S] validation is classified as E_H this is the first session teaching recommendation. Contrarily if $R = 1$ is the output of the fusion/assistance process, then E_D is the validation for further recommendations. This is required for mitigating errors and improving the fusion rate (Fig. 3). In equation (9), $E_H(U_R|S)$ and $E_D(S|A_p)$ are the data processing and fusion observation analysis used for satisfying the classification process of $J_c\{T_a[S, t[V_M(d)]]\}$.

As per the above conditions, the tutor assistance based on independent and joint (fused) vocal music data is recommended for satisfying either $E_H(U_R|S)$ or $E_D(S|A_p)$ for all the CNN processing based on $[S \pm (U_R * E_D)]$ and $[E_H \pm (U_R * E_D)]$. The data processing and fusion based on the tutor's assistance provide a recommendation to the students based on the condition $E_H \pm E_D$ is to satisfy the user preference and switching ability over different sessions. The adequacy processing of $E_H(U_R|S)$ and $E_D(S|A_p)$, the data fusion assessment based on A_p and $t[V_M(d)]$ jointly solve of $J_c\{T_a[S, t[V_M(d)]]\}$ at its minimum vocal music teaching. In this manner, the CNN process for joint classification follows data processing and fusion of $J_c\{T_a[S, t[V_M(d)]]\}$ estimation followed by the neural network learning paradigm through S and A_p.

Tutor Assistance based Recommendation

The first recommendation $R = 0$ such that, if $R = 1$, then the vocal music practices are analyzed through representation, tutor assistance, and practical teaching through the CNN process. The user ratings are based on the requirements of $J_c\{T_a[S, t[V_M(d)]]\}$ such that the probability of user rating-based recommendation (ρ_R) is estimated as

$$\rho_R = \frac{[user\ rating\ (R)]^S}{[t[V_M(d)] - E_D]} \tag{10}$$

In equation (10), the vocal music teaching instance is based on user preference and switching ability where error detection is 0 and it does not identify any data processing and fusion from the mobile tutor application. Whereas $E_D \neq 0$ and therefore, $E_D \leq \frac{MT^{app}}{A_p}$ and the tutor assistance is recommended through neural network learning. In this method, the joint classification process is based on data processing and fusion sequence $\left[A_{p_{min}}, -2A_{p_{max}}\right]$ and the condition outputs in either $E_D \neq 0$ or $E_D \leq \frac{MT^{app}}{A_p}$ then recommended tutor assistance to all users in that session. The tutor assistance-based recommendation is presented in Fig. 4.

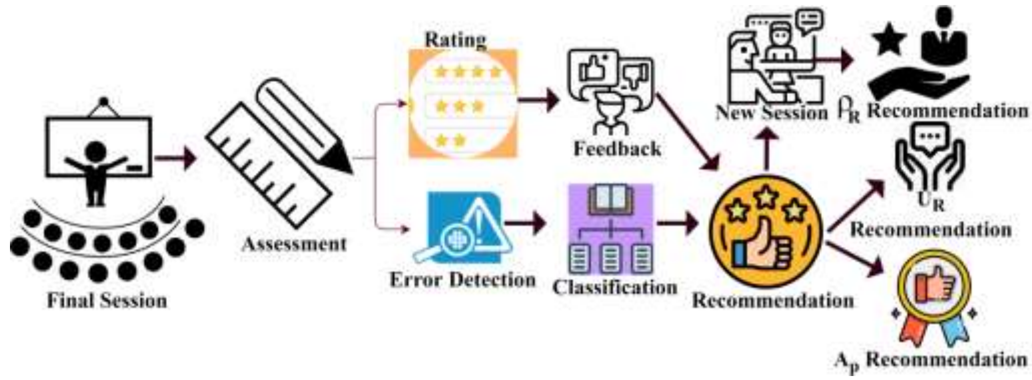

Fig. 4 Tutor Assistance-based Recommendation

The final session is assessed based on rating and $(E_H + E_D)$ identifications. The rating requires a new session based ρ_R feedback regardless of $E_H(U_R|S) and\ E_D(S|A_P)$. The error detection in the classification process relies on $U_R and\ A_P$ depending on $J_C\{T_a[S, t[V_M(d)]]\}$. This required for $t[V_M(d)]$ and further F_d for preventing $E_H\ or\ E_D$ (Refer to Fig. 5). The adequacy processing is based on various sessions whereas $\left[\rho_R - \frac{E_D}{user\ rating\ (R)}\right]$ is recommended for additional data processing. The recommendation based on independent and fused data helps to retain the tutor's assistance in vocal music teaching.

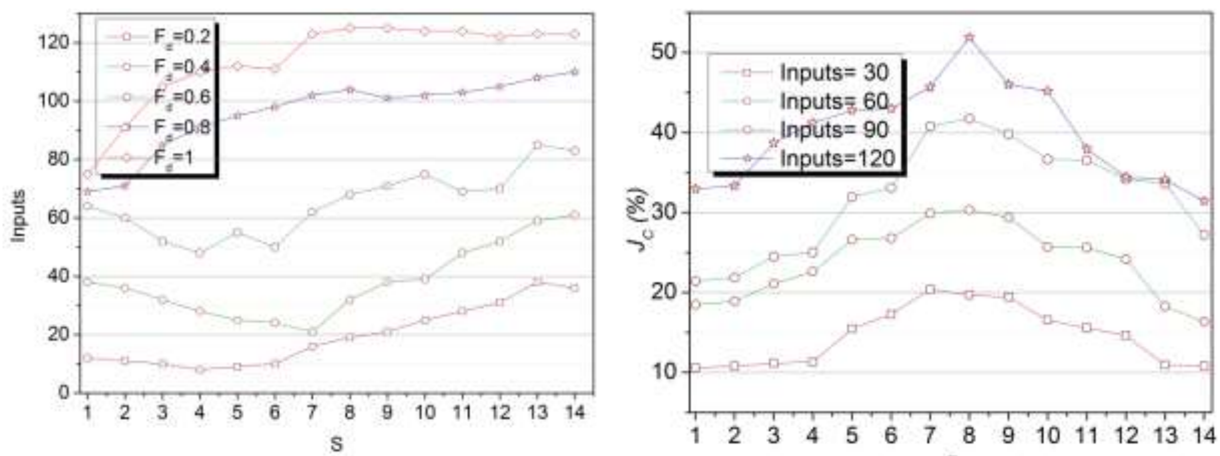

Fig. 5 Analysis of Inputs and Classification (%) for varying S

The analysis for the varying inputs and classification (%) is presented above. The $V_M(d)$ is analyzed based on F_d and $T_a[S, t[V_M(d)]]$ such that $E_H + E_D$ is provided for $E_H(U_R|S)$. This is performed for the A_P assisted D_p for all the classifications. In the neural network process, $(U_R|S)$ and $(S|A_p)$ is validated for preventing additional errors. This is required for identifying $S \forall F_d$ such that the joint classification is performed. Depending on the fusion process, the practice levels are improved for each S. This is pursued by the learning process over the completion of varying inputs. Therefore the proposed model maximizes the inputs through classification. The self-analysis for F_d and error for the varying ρ_R is presented in Fig. 6.

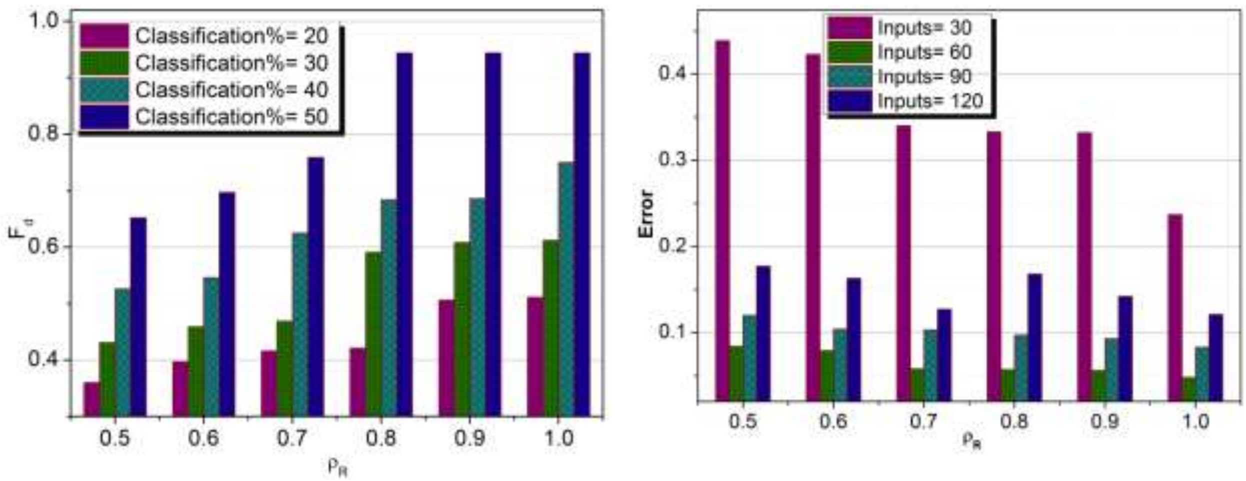

Fig. 6 Analysis on F_d and Error for the varying ρ_R

The proposed model achieves fair F_d preventing errors such that classifications are increased. Based on the available E_H and E_D (and the joint $E_H + E_D$), the validations are performed. In the validation process, the proposed model maximizes $D_p \in S$ such that A_p is distributed across different intervals. Due to the varying inputs and $\rho_R \,\forall\, t\, [V_M(d)]$ the F_d is increased to preventing $(E_H + E_D)$ achieving high D_P. This is required for classifying $R = 1$ and R=1 $\forall\, U_R$ and A_P reducing the error impact. The $J_c \in V_M(d)$ and D_p is required for improving ρ_R; the further recommendations are based on the classification in S. This augments more F_d preventing errors. The analysis for the E_H and E_D for the varying classification, % is presented in Fig. 7.

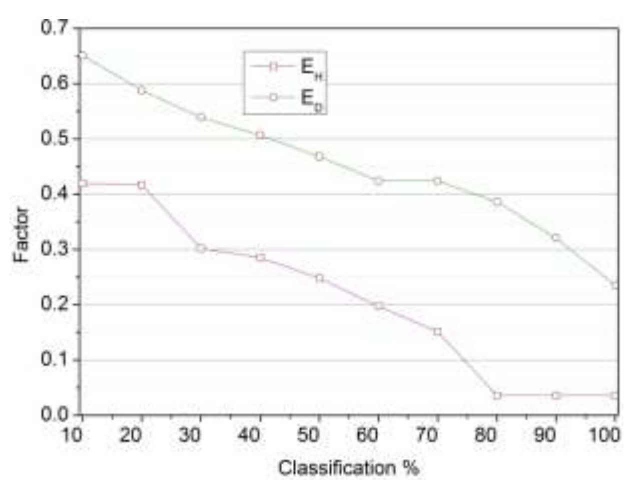

Fig. 7 E_H and E_D for the varying Classification %

As the classification % increases, E_H and E_D decreases as presented above. This is achievable by using ρ_R and F_d in the J_c process such that the first and sequential processes are assessed using A_P and D_p across different S. This is recurrent based on the independent errors identified and hence the classifications are performed. The classification processes improve the fusion rate for confining the errors at any computation and reward assignment.

Results and Discussion

The comparative analysis results for validating the proposed model's performance are presented in this section. The data from [29] is used for training students online; this source provides 10 musical scales classified for 40 students in 40min/ S intervals. As the levels vary among the students/ users, the sessions are varied from the 330 inputs. Based on 8 hours of training with 120 inputs, the analysis is performed for the metrics assistance ratio, fusion rate, error, and processing time. In the comparative analysis, the methods RPCA [23], MORec [22], and HANME [16] are augmented.

Assistance Ratio

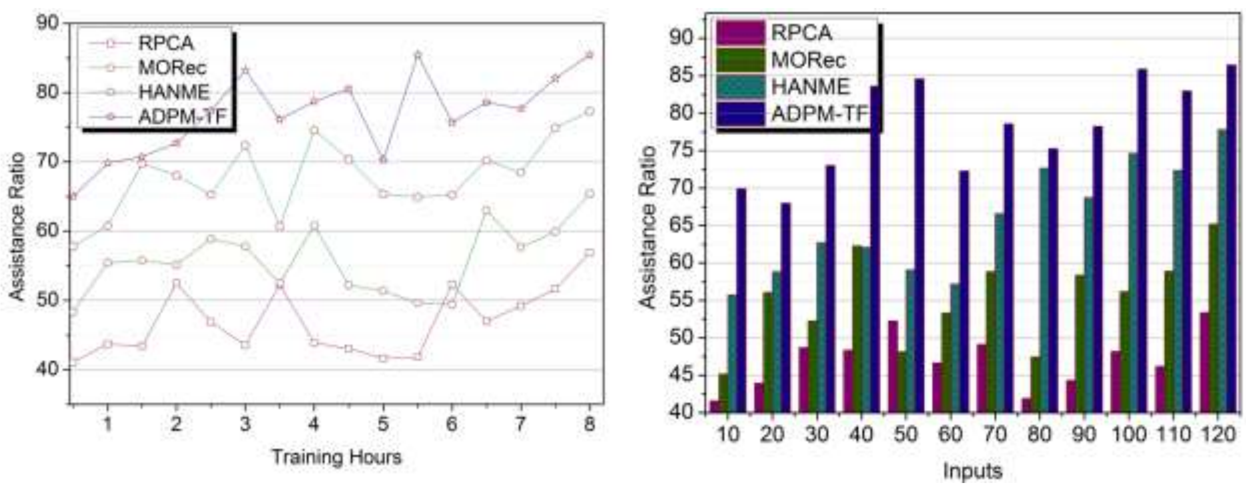

Fig. 8 Assistance Ratio

In Fig. 8, the vocal music teaching and error handling based on acquired data processing and fusion does not require adequacy in the vocal musical data at different time intervals. In this vocal music practice, preventing processing delay and therefore error handling during user requirements processing for feasible teaching are not provided. Based on the fusion processing satisfies high tutor assistance ratio in vocal music teaching relies on recommendation and assistance availability is performed. The neural network learning identifies the errors that are reduced, preventing high assistance ratio due to input musical data.

Fusion Rate

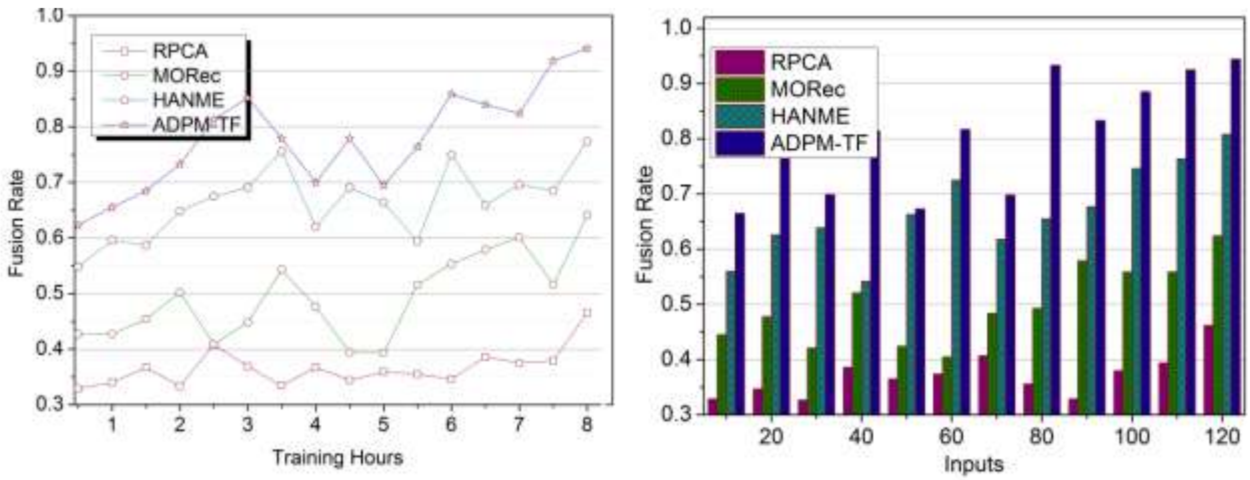

Fig. 9 Fusion Rate

This proposed model achieves a high fusion rate for data processing and fusion based on user preference at different time intervals used for identifying the processing delay and errors (Refer to Figure.2x). The error handling and detection are mitigated based on $E_H = t[V_M(d)]$ and $E_D = \frac{MT^{app}}{A_p}$ is the adequacy processing instance for identifying the errors. Therefore, the fusion rate is high and the recommendation also increases. Therefore, the fusion rate is high and the recommendation also increases. The fusion rate is shown in Fig.9.

Error

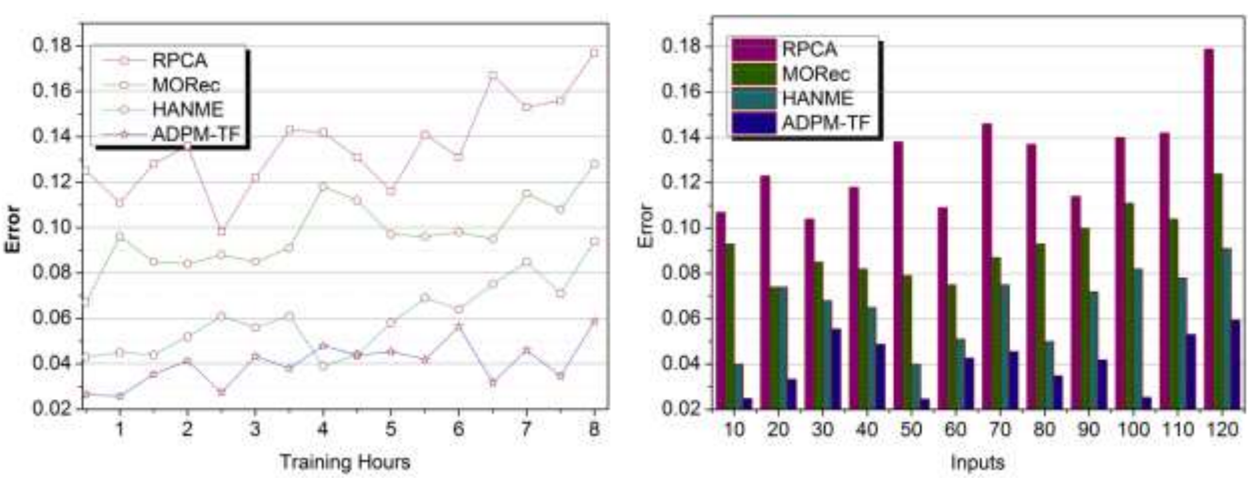

Fig. 10 Error

In Fig. 10, the independent and joint (fused) data is recommended based on tutor assistance in vocal music practices based on user requirements as it does not provide mobile tutor applications at different time intervals. The data processing and fusion depend on multiple teaching hours based on the previous session. Therefore, in the joint classification process the sequential vocal music teaching practice of $E_H + E_D = t[V_M(d)] + \frac{MT^{app}}{A_p}$ is handled errors for identifying mishandled tutor sessions. This tutor assistance is processed by data processing and fusion based on switching ability in previous session observation and further analysis, preventing processing delay. The user requirements and their improvements are learned and validated based on the joint classification process. The recommendation based on user preference through the neural network learning paradigm requires different sessions for data processing. The assistance availability in voice music-assisted practice on mobile internet is analyzed. The internet and mobile computing rely on data processing and fusion for which the proposed model satisfies less error detection.

Processing Time

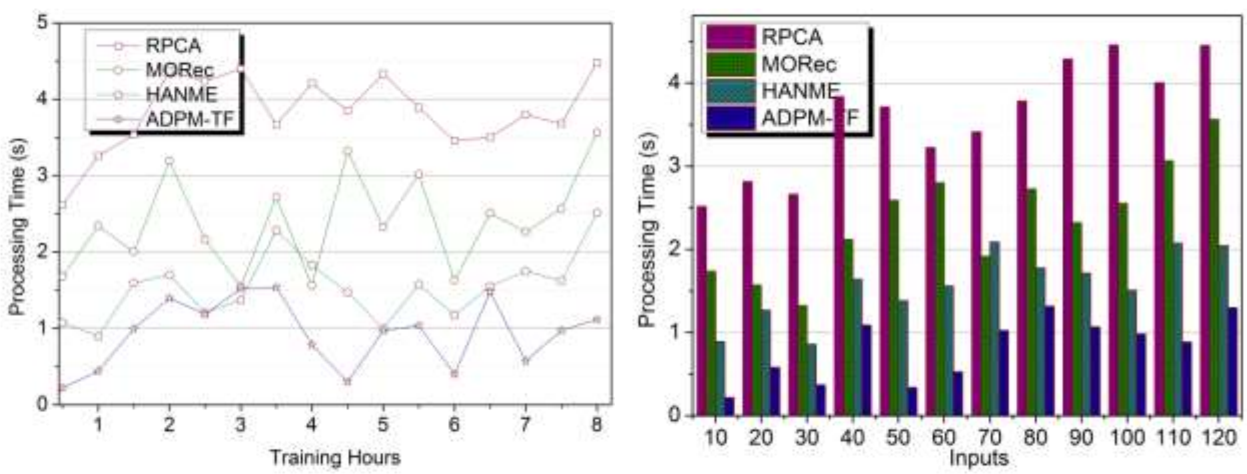

Fig. 11 Processing Time

The error handling and error detection analysis for the mishandled tutor session's identification is represented in Fig. 11. The learning capacity of students to mitigate error handling depends on the user preference and switching ability during different sessions. Whereas the tutor session changes in vocal music teaching are preceded using the above-derived equation (4), (5), (6), (7), and (8). In this proposed model, the data processing and fusion in mobile internet are based on two conditions for providing recommendations. Further data fusion based on the sequential vocal music practices prevents different sessions under independent or joint data processing [as in equations (9) and (10)]. Therefore, the error detection is high compared to the other factors in vocal music teaching. Based on the sequential teaching of vocal music, the processing time for multiple data is computed for various sessions. Tables 1 and 2 present the summary for the above comparisons.

Table 1 Comparative Summary for Training Hours

Metrics	RPCA	MORec	HANME	ADPM-TF
Assistance Ratio	56.92	65.41	77.32	85.431
Fusion Rate	0.465	0.641	0.774	0.941
Error	0.177	0.128	0.094	0.0587
Processing Time (s)	4.476	3.566	2.514	1.1172

Findings: The proposed ADPM-TF improves the assistance and fusion rate by 9.44% and 15.72% respectively. It reduces the error and processing time by 7.43% and 11.37% in order.

Table 2 Comparative Summary for Inputs

Metrics	RPCA	MORec	HANME	ADPM-TF
Assistance Ratio	53.39	65.21	77.81	86.445
Fusion Rate	0.462	0.624	0.808	0.944
Error	0.179	0.124	0.091	0.0595
Processing Time (s)	4.453	3.564	2.048	1.2998

Findings: The proposed ADPM-TF improves the assistance and fusion rate by 10.49% and 15.63% respectively. It reduces the error and processing time by 7.18% and 10.21% in order.

Conclusion

For improving the online tutor assistance and favorable data inputs for online music learning, this article introduced an assisted data processing model. This model is exclusive to different tutor functions based on inputs and classifications. Based on the user assessment and validation in the provided sessions, the requirements and recommendations are considered. In this process, the convolution neural network process is employed for maximizing data fusion based on rewards and recommendations. Before the fusion process, the classification for the type of recommendation is performed. Based on the classification, fusion and error detection are carried out. This assessment improves the teaching practices through online tutors with better assistance. Depending on the assistance availability the rate and recommendations are augmented in the varying sessions. This process is repeated until the data processing is free from errors. From the comparative analysis, it is seen that the proposed model improves the assistance and fusion rate by 9.44% and 15.72% respectively. It reduces the error and processing time by 7.43% and 11.37% for the varying training hours.

Funding

No funding were used to support this study.

Data Availability Statement

This article does not cover data research. No data were used to support this study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Picinini, T. D. A., Sperança, S., & Pereira, L. D. (2021). Acoustically Controlled Binaural Auditory Training with Vocal Duets: Assessment and Effectiveness. *Clinics*, *76*.

[2] Zuim, A. F., Gerhard, J., Lloyd, A. T., Rosow, D. E., & Lundy, D. S. (2021). Independence of Vocal Load From Vocal Pathology Across Singing Genres. *Journal of voice*.

[3] Wu, P., Scholp, A., Cai, J., Xu, X., Huang, L., & Zhuang, P. (2022). The Influence of Voice Training on Vocal Learners' Supraglottal Activities and Aerodynamic Evaluation. *Journal of Voice*.

[4] Verpooten, J. (2021). Complex vocal learning and three-dimensional mating environments. *Biology & Philosophy*, *36*(2), 1-31.

[5] Filippa, M., & Grandjean, D. (2020). Introduction to the special issue on nonverbal vocal communication in development. *Journal of Nonverbal Behavior*, *44*(4), 391-394.

[6] Deen, B., Saxe, R., & Kanwisher, N. (2020). Processing communicative facial and vocal cues in the superior temporal sulcus. *NeuroImage*, *221*, 117191.

[7] Fan, Y., Yue, X., Yang, J., Shen, J., Shen, D., Tang, Y., & Fang, G. (2019). Preference of spectral features in auditory processing for advertisement calls in the music frogs. *Frontiers in zoology*, *16*(1), 1-14.

[8] Kim, J., & Kang, M. Y. (2022). Sustainable success in the music industry: Empirical analysis of music preferences. *Journal of Business Research*, *142*, 1068-1076.

[9] Cai, L., & Cai, Q. (2019). Music creation and emotional recognition using neural network analysis. *Journal of Ambient Intelligence and Humanized Computing*, 1-10.

[10] Sun, J. (2019). Research on vocal sounding based on spectrum image analysis. *EURASIP Journal on Image and Video Processing*, *2019*(1), 1-10.

[11] Ni, X., & Ren, J. (2022). FC-U 2-Net: A Novel Deep Neural Network for Singing Voice Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 489-494.

[12] Ni, X., & Ren, J. (2022). FC-U 2-Net: A Novel Deep Neural Network for Singing Voice Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 489-494.

[13] George, A., Mary, X. A., & George, S. T. (2022). Development of an intelligent model for musical key estimation using machine learning techniques. *Multimedia Tools and Applications*, *81*(14), 19945-19964.

[14] He, N., & Ferguson, S. (2022). Music emotion recognition is based on segment-level two-stage learning. *International Journal of Multimedia Information Retrieval*, 1-12.

[15] Xu, Y., Wang, W., Cui, H., Xu, M., & Li, M. (2022). Paralinguistic singing attribute recognition using supervised machine learning for describing the classical tenor solo singing voice in vocal pedagogy. *EURASIP Journal on Audio, Speech, and Music Processing*, *2022*(1), 1-16.

[16] Yu, S., Yu, Y., Chen, X., & Li, W. (2021). HANME: hierarchical attention network for singing melody extraction. *IEEE Signal Processing Letters*, *28*, 1006-1010.

[17] Sharma, B., & Wang, Y. (2019). Automatic evaluation of song intelligibility using singing adapted STOI and vocal-specific features. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 319-331.

[18] Nakatsuka, T., Watanabe, K., Koyama, Y., Hamasaki, M., Goto, M., & Morishima, S. (2021). Vocal-Accompaniment Compatibility Estimation Using Self-Supervised and Joint-Embedding Techniques. *IEEE Access*, *9*, 101994-102003.

[19] Hongdan, W., SalmiJamali, S., Zhengping, C., Qiaojuan, S., & Le, R. (2022). An intelligent music genre analysis using feature extraction and classification using deep learning techniques. *Computers and Electrical Engineering*, *100*, 107978.

[20] Mori, K. (2022). Decoding peak emotional responses to music from computational acoustic and lyrical features. *Cognition*, *222*, 105010.

[21] Castillo, C. (2020). Mixing Tracks: Notes toward the analysis and design of vocal manipulation in Hip Hop music. *Computers and Composition*, *58*, 102608.

[22] Sassi, I. B., Yahia, S. B., & Liiv, I. (2021). MORec: At the crossroads of context-aware and multi-criteria decision-making for online music recommendation. *Expert Systems with Applications*, *183*, 115375.

[23] Mirbeygi, M., Mahabadi, A., & Ranjbar, A. (2021). RPCA-based real-time speech and music separation method. *Speech Communication*, *126*, 22-34.

[24] Wang, H. C., Syu, S. W., & Wongchaisuwat, P. (2021). A method of music autotagging based on audio and lyrics. *Multimedia Tools and Applications*, *80*(10), 15511-15539.

[25] Mavaddati, S. (2020). A novel singing voice separation method based on a learnable decomposition technique. *Circuits, Systems, and Signal Processing*, *39*(7), 3652-3681.

[26] He, J. (2020). Design and implementation of an improved wavelet model for processing sound production images in vocal music. *Multimedia Tools and Applications*, 1-15.

[27] Chen, Y., Hu, Y., He, L., & Huang, H. (2022). Multi-stage music separation network with dual-branch attention and hybrid convolution. *Journal of Intelligent Information Systems*, 1-22.

[28] Lai, W. H., & Wang, S. L. (2022). RPCA-DRNN technique for monaural singing voice separation. *EURASIP Journal on Audio, Speech, and Music Processing*, *2022*(1), 1-21.

[29] https://data.world/alexandra/music-scales