A real-time singing scoring system based on virtual reality technology

Pei Zhang, Yi Han*, Weidan Liang and Jingyi Chen
School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China.
*Corresponding author’s e-mail: hanyi@whut.edu.cn

Abstract. In this thesis, a real time singing scoring system based on virtual reality technology is proposed. When a user sings a song, the system will analyse and process the user's audio, compare it with the standard audio, give the corresponding score and change the character animation in the scene in real time according to the score level. By giving the user a visually realistic experience, the system creates desensitisation, thus reducing the user's fear of performing on stage and improving their ability to perform on stage.

1. Introduction
In today's life, there is such a group of people who, either because of genetic factors (phobias have a family tendency to run in families), or because of psychological fears, are involuntarily nervous and scared when singing songs in front of people, or because of low self-esteem lacking the courage and confidence to interact, resulting in them being afraid to interact, unwilling to interact and unable to interact in social life. This stage fright and introversion can sometimes have a serious impact on their lives. The use of VR technology to overcome fear of heights has been proposed in a paper by Kitanosono, Iku, as a way of addressing people's psychological fears [1]. In his thesis, Jindrich Adolf explores the application of VR technology to juggling training to avoid the dangers that arise during training[2]. Meanwhile Mina C. Johnson-Glenberg in her thesis suggests that VR technology can effectively improve learning compared to 2D PC [3]. Therefore, 3DMAX modelling technology can be used to build a virtual world similar to the real one, and then signal processing technology can be used to identify people singing in the virtual world and then score them. This not only removes the barriers of time and place, but also helps such people to exercise their singing skills and overcome stage fright and fear. This thesis proposes a virtual reality-based real-time singing scoring system that solves these problems.

2. Performance scoring algorithm design
A speech signal is a non-stationary signal whose characteristics vary over time, but can be considered relatively stable over a short period of time, i.e. short-time smoothness. Thus speech has short-time autocorrelation. This time period is about 15 ms, for which both the statistical and spectral characteristics are for short time periods. This makes it necessary to digitally process the speech signal by first dividing it into short-time segments with sufficient overlap between adjacent frames. In this way each frame is short-time smooth and thus subject to short-time correlation analysis. After
extracting the fundamental period of the audio by means of the short-time autocorrelation function, the Dynamic Time Warping (DTW) algorithm is then used to process the singer's audio fundamental period against the standard audio fundamental period to calculate the optimal path. To implement the comparison function, a set of null controls is used, and the null control data is processed with the standard speech signal by the DTW algorithm to obtain the comparative best path data. Ultimately, the two data are processed to obtain the scoring data.

![Performance scoring algorithm flow chart](image)

The project uses the Windows Multimedia API to buffer the audio stream and send it piece by piece in order to achieve real-time audio acquisition, processing, scoring and writing to a txt file once the audio data is acquired, for real-time speech signal processing.

The scoring module first uses MATLAB to simulate the functions of each module, then the functional module algorithms are converted into C++ code and encapsulated into a dynamic link library (DLL) file to be called in Unity.

As the simulation using MATLAB used some of the functions that come with MATLAB, this led to a large number of modifications to the algorithmic ideas of the module in the process of changing to C++ code in order to achieve the relevant functions. In addition, in order to implement the real-time scoring function it was necessary to process the language data quickly and optimise the algorithm running time, where triple loops in the framing and autocorrelation algorithms resulted in long running times when using the short-time autocorrelation function method for fundamental sound cycle detection. Firstly, to save unnecessary processing time, the data obtained from the standard audio by the short-time autocorrelation function method is stored in a text file, and each time a song is selected, the data in the text file is read directly and then processed by the DTW algorithm. Secondly, in order to remove the triple loop algorithm, the data is stored half a frame at a time during the voice recording, and after the second storage, the data is processed and half of the frames are guaranteed to overlap with each other. This saves a lot of processing time.

3. Unity Development
The project was developed using Unity for scene layout and interaction.

3.1. Scene setting
In terms of scene setting, in order to enhance the immersion of the user, a small model of a singing room was designed and built in this thesis using 3DMAX as a stage for the user to sing. Five audience models were designed and built, and corresponding animations were created for each of them.
3.2. Jukebox design and implementation of singing rating function

In terms of the design and implementation of interactive functions, three songs are used in this system in order to enable the song ordering function. By designing the UI interface to bind the corresponding songs and instantiating the interface by clicking on the corresponding buttons in the scene, the function of playing the corresponding songs and displaying the corresponding lyrics is completed.

Import the dynamic library of singing scores into Unity, develop the corresponding script in C#, design a thread to run the dynamic library; a thread to obtain the scores from the txt file written by the dynamic library every 40s, and display them on the UI interface of song ordering, and modify the audience animation; a thread to time the song, close all threads after the song is finished playing, and calculate the average score of several times, as the final score of this singing and display it on the interface.

3.3. Virtual Reality Functionality Implementation

The system uses the HTC VIVE as the VR headset for this project. The VR component of Unity3D was used to implement the game interaction on the VR platform, i.e. the Unity3D expansion library.
Steam VR, which transforms the PC game program into a VR game that can be displayed on a VR device. The Steam VR plugin was used and developed to allow players to interact with the game through the joystick. The use of Teleporting allows the player to move around with the joystick, making it easier to continue playing in a narrow environment.

4. Experimental result

Connect the HTC VIVE and package this system to run on a computer with a GTX1070 graphics card configuration.

Play the original song recorded into both the system and the comparison singing scoring software, marked with x, to record its final score and the comparison software score. No sound was recorded, marked as y, and its final score and comparison software score were recorded. Five randomly selected experiencers’ abcde to sing the same song using this system and record their final score given by this system and the comparison software. The results of the experiment are shown in Table 1 below.

| TAB | Final Score | Comparison Score | Difference |
|-----|-------------|------------------|------------|
| x   | 97.12       | 98.52            | 1.4        |
| y   | 6           | 2                | 4          |
| a   | 89.63       | 91.65            | 2.02       |
| b   | 83.38       | 84.52            | 1.14       |
| c   | 92.13       | 92.28            | 0.15       |
| d   | 94.45       | 95.16            | 0.71       |
| e   | 88.15       | 90.21            | 2.06       |

Comparing the scores of this system with those of the comparison software shows that the overall error between the two is small. Within the margin of error, the scores of this system are generally in line with the actual situation.

One experiencer was selected to sing the same song and the real-time score and final score were recorded six times respectively, and the above experimental process was repeated five times. The results of the experiment are shown in Table 2 below.

| Frequency | Score 1 | Score 2 | Score 3 | Score 4 | Score 5 | Final Score |
|-----------|---------|---------|---------|---------|---------|-------------|
| 1         | 95.62   | 93.21   | 91.52   | 91.65   | 94.21   | 93.24       |
| 2         | 89.21   | 92.31   | 92.25   | 93.18   | 90.02   | 91.39       |
| 3         | 85.25   | 89.69   | 92.14   | 88.21   | 93.12   | 89.68       |
| 4         | 92.11   | 87.69   | 90.09   | 94.18   | 86.24   | 90.06       |
| 5         | 90.58   | 98.51   | 92.26   | 93.84   | 90.13   | 93.06       |

The analysis of the above data shows that the system can score properly within the margin of error and can be read out properly to give a final score.

One experiencer was selected to sing 3 songs and the final score was recorded. The results of the experiment are shown in Table 3 below.

| Song 1 Score | Song 2 Score | Song 3 Score |
|--------------|--------------|--------------|
| 91.22        | 85.64        | 92.61        |
The analysis of the above experimental data shows that this system can score three songs normally within the error operating range.

5. Summary
The virtual reality-based real-time singing scoring system proposed in this thesis has been experimentally tested to have high immersion and accurate scoring characteristics. This system is a good application of virtual reality technology in song singing scoring, thus it can solve the shortcomings of the traditional singing scoring system of low immersion, improve the user's experience and enhance the user's familiarity with the stage. It allows users to become desensitised through constant exercise, reducing their fear of the stage and enhancing their stage performance. This system not only satisfies the practical need for assisted singing practice, but also provides some ideas for the development and application of virtual reality technology.

References
[1] Iku Kitanosono, Toshiyuki Haramaki, Tsuneo Kagawa, Hiroaki Nishino. “A vr system for alleviating a fear of heights based on vital sensing and placebo effect,” Advances in Intelligent Systems and Computing, v 1036, 2020, p 62-72.
[2] Jindrich Adolf, Peter Kán, Benjamin Outram, Hannes Kaufmann, Jaromír Doležal, Lenka Lhotská. “Juggling in VR: Advantages of immersive virtual reality in juggling learning.” Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST, November 12, 2019.
[3] Mina C. Johnson-Glenberg, Vanessa Ly, Man Su, Ricardo Nieland Zavala, Hannah Bartolomeo, Elena Kalina. “Embodied Agentic STEM Education: Effects of 3D VR Compared to 2D PC,” Proceedings of 6th International Conference of the Immersive Learning Research Network, iLRN 2020, June 2020, p 24-30.