Pitchy Perfect

Smith, Nicholas
CIS 4930
nsmith78@ufl.edu

Wood, Jeffrey
CIS 6930
Jmwood22@ufl.edu

ABSTRACT
In this paper, we describe Pitchy Perfect, the real-time "vocal tuner" we will work to develop this semester in CIS4930/6930.

Keywords
Pitchy Perfect; Vocal tuner; 3D audio; Real-time feedback; Perfect pitch; Experiment; Spatial Perception

1. INTRODUCTION
Pitchy Perfect is a product that serves to effectively assist musicians, vocalists, and anyone else who wants to be trained to sing in relative pitch. In general, the purpose of Pitchy Perfect will be to aid those who are attempting to train their ear. If you have ever struggled to sing along to your favorite song in the right pitch then you know that being able to sing in the correct frequency range, on command, is a difficult challenge. Relative pitch is different from perfect pitch, however, in that a person with perfect pitch does not need a reference note in order to identify or re-create a particular note.

There are various solutions that exist to achieve this goal, one of which is the PitchPerfector [1]. This product costs anywhere from $67 to $97, depending on the type of software package [1]. This software provides real-time visual feedback on the user's singing and gives instructions on how to diagnose the individual pitch problems [1]. Some notable features include being able to select the singing range, a keyboard and a staff display for visual feedback, and a spectrograph to visualize the vocal attributes [1]. Another existing solution is the Roland VT-12 Vocal trainer [2]. This product costs about $200 and checks the pitch of the vocal input in half-step increments, giving visual feedback with the illuminated pitch meter [2]. This device also provides a variety of onboard exercises for training one's voice [2]. The PolyTune 2 Mini is another existing device and currently costs about $60 [3]. This device is primarily used for tuning instruments, but also serves as a vocal tuner. Like the other devices, the PolyTune 2 has a dependence on a visual component, in that it has an LED display to give visual feedback to the user's pitch [1, 2, 3].

When reviewing all the features of the existing tuners and vocal trainers, it is clear that there is significant room for improvement. A significant problem with these existing solutions is that they can be quite expensive for the casual user [1, 2, 3]. This issue can be addressed by releasing Pitchy Perfect as an inexpensive product, if not completely free. It is also clear that many devices that served as vocal trainers were actually meant for instruments, not a vocal input. Therefore, Pitchy Perfect will be a vocal tuner that can receive vocal data in real time by using the computer's microphone. This implementation technique was chosen because it is inexpensive and will be dedicated to solely accommodating vocal input. Another shortcoming among the existing products that were examined is that they all relied on a visual interface to communicate the data to the user. By incorporating 3D audio technology that works in tandem with a visual display into Pitchy Perfect, it will go above and beyond what is currently available by not depending on the visual interface to be effective.

Pitchy Perfect will be a "tuner" that uses 3D audio technology to train users to improve their ability to have relative pitch. In doing so, a 3D audio display will be used to play a selected reference note to the vocalist on a left-to-right spectrum. If the vocalist is singing a note that is "flatter" than the reference note, the sound will appear to be coming from the left. If the vocalist is "sharper" than the reference pitch, the sound will appear to be coming from the right. The reference note will then sound closer or farther away to the user, depending on how close the vocalist is to the pitch of the reference note. Once the vocalist has adjusted their pitch appropriately, the reference note should be heard equally in both ears and perceived as being located in the center of the left-to-right spectrum. In addition, some key features that will be implemented in Pitchy Perfect are 3D audio, a visual display, which will also aid in the debugging process; real-time feedback; and a function to allow the user to choose a key they would like to practice. If time permits, some training lessons that a user could utilize will be included. In this, the user will be given a sequence of notes or a melody that they must re-create. The user would then be scored on how close they were to each individual note. To get optimal results using Pitchy Perfect, the user should use headphones and a microphone. Success, in regards to this project would be having created a user-friendly interface that effectively tells the user where they are on the spectrum of frequencies within a certain pitch and how to correct the issue to make their vocal accuracy better. It is also imperative that the reference note that is played to the user be spatially correct. In other words, the reference note should be perceived to be approaching from the correct side of the left-to-right spectrum and at a representative distance away, depending on how "flat" or "sharp" the user's singing is. The end goal of Pitchy Perfect is to turn someone who struggles with relative pitch into someone that can effectively sing every reference note correctly.

Challenges expected to exist while developing this software include reducing any noise the microphone may be hearing, other than the vocal input, which could alter the real-time feedback. To overcome this, it is likely that appropriate filters will be utilized to reduce unnecessary noise the program is receiving as input. It may also be difficult to utilize 3D audio in such a way that the user understands how much they need to adjust their pitch to sing the note correctly. However, this issue will be tackled by designing the audio display in such a way that the reference note moves proportionately in space to how much the user has changed their pitch. Additionally, a maximum limit on how far away the reference note will be played will be implemented. This is to assure that the reference note can always be perceived, even if the user is very far off the reference pitch. Another challenge, if time permits and training lessons are included, may be to identify the different notes the vocalist is singing, assuming the notes that are sung are very similar to each other or sung in rapid succession.

Therefore, if time permits, research will also be conducted on how to quickly and effectively separate and analyze various vocal inputs so that real-time feedback can be provided to the user.

2. Experiment

In this experiment, we will measure the time it takes for people of different musical backgrounds to identify that they're off pitch and how long it takes them get back on pitch, both with and without Pitchy Perfect. By doing so, we hope to show that not only can Pitchy Perfect train one's ability to be in relative pitch, but can also teach them to do so quickly.

2.1 Participants

The target user population is both musicians and non-musicians that are working towards improving their ability to sing in relative key. The characteristics that they must have are the ability to spatially locate sounds, recognize when they are not in relative pitch, adjust the pitch of their voice based on the real-time feedback given by Pitchy Perfect, and the capability to hit a buzzer at specific times.

2.2 Stimuli

The sounds we will use for this experiment will be contained in a library of sine waves, each played at a specific frequency. More specifically, we will have one sine wave at the frequency of each key on a standard full-size keyboard. One of the goals of this experiment is to assure that the real-time 3D Audio feedback is what is leading the vocalist to correct their pitch. By using sine waves, we are using a very basic sound and eliminating any help that may come from using the sounds of an instrument. It may be the case that participants find it easier to locate a particular pitch when using a specific instrument as a guide. In fact, before deciding on sine waves, we considered using the sounds of an instrument that most people consider familiar; however, to avoid this or any other bias related to the sounds, we decided to use sine waves.

We will satisfy the 3D Audio component by using real-time 3D Audio feedback. If, for example, the vocalist is singing a note that is "flatter" or "sharper" than the reference note, the reference note will appear to be coming from the left or right, respectively. The sound of the reference note will also appear to change in proximity to the vocalist, depending on how close the vocalist is to the correct pitch of the reference note. Once the vocalist has adjusted their pitch appropriately, the reference note should appear to be coming from the center of the left-to-right spectrum.

2.3 System

The system is going to include the participant sitting at a station with headphones, the Pitchy Perfect program, a timer, and a buzzer that the participant will use. For the purposes of this experiment, we will not use the visual graphical user interface simply because we are working to test the adequacy of the 3D Audio component of Pitchy Perfect and do not want the visual component to affect the results. In the final product, however, we plan to have a visual interface that will work in tandem with the audio component as well. Inputs to the system will be the participants' voices and the buzzer. The output of the experiment will be the amount of time it takes for the user to press the buzzer.

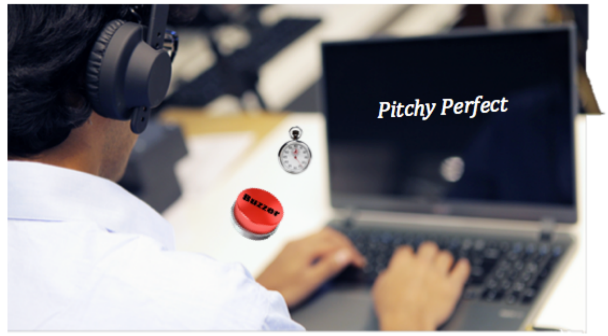

Figure 1: Basic Experiment User/Visual Interface

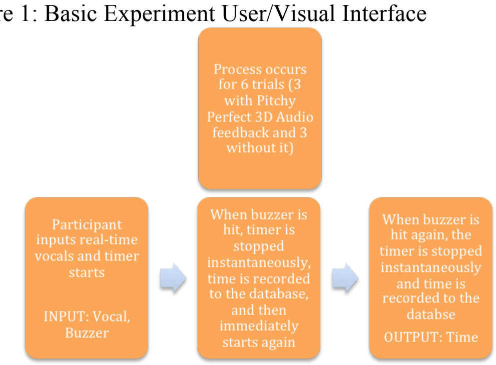

Figure 2: Flow of Inputs and Outputs in Experimental System

2.4 Procedure

We will begin by instructing the participant to sit down at the test station. We will then tell the participants that they will be played a set of sounds and at that point, they should do their best to sing and match the pitch of the reference sound. Once they confirm that this information is understood, we instruct them to press the buzzer at the point that they realize that they are off pitch and to press it again when the system tells them, on the screen, that they are singing the specified note. The headphones are then put on the participants and the specified parameters for each trial run are set.

Once three trials are complete, we tell the participants that they will be played another set of sounds and at that point, they should do their best to sing and match the pitch of the reference sound. We then explain that the system will now tell them how far off they are from the correct reference note by moving the sound along a left-to-right axis in the headphones. Once they confirm that this information is understood, we instruct them to press the buzzer at the point that they realize that they are off pitch and to press it again when they hear the reference sound being played in the center of the left-to-right axis. The use of the buzzer will indicate both the time it takes for the participant to realize that they are not singing the correct pitch and how long it takes for them to correct the issue. The headphones are then put on the participants and the specified parameters for each trial run are set in the Pitchy Perfect program in MatLab. Once three trials are completed, the experiment is completed with that participant.

Thus, two conditions are being tested with this experiment: how long it takes to efficiently correct one's pitch with and without Pitchy Perfect.

2.5 Statistical Treatment

The evaluation metrics will be time (in seconds) it takes for each participant to realize that they are not singing the correct pitch, and how long it takes them to correct their pitch, both with Pitchy Perfect and without. This data will be used to determine the level of success of Pitchy Perfect by comparing the two sets of times collected from each participant. If the time it takes for the participants to press each buzzer using the Pitchy Perfect 3D Audio technology is reduced by an average of at least 10%, we can consider the product a success.

3. Potentially Interested Groups

There are many potential useful applications of Pitchy Perfect in businesses, the consumer sector, and the entertainment industry. Companies that sell instruments, like Fender, Gibson, Tama, Yamaha, to name a few, would be interested in this because they all sell tuners for their instruments and for vocalists. However, none of these companies sell a vocal tuner that utilizes 3D Audio technology. Other potentially interested groups include university acapella groups, singers, music schools, and even Broadway to help their actors improve their ability to hold perfect pitch.

4. Addition to Existing System

A system that Pitchy Perfect can effectively be added to is an existing instrument tuner. The only necessary hardware addition for this system to function correctly would likely be an output port for the user's headphones. To make the system more responsive, additional processing power can also be added. When it comes to the 3D Audio component, it would be useful to establish an average Head Related Transfer Function (HRTF). This would allow for the 3D Audio effects to be more effectively perceived by all users, regardless of the differences in their HRTFs.

5. Future Steps

If there were an unlimited amount of time to work on Pitchy Perfect, we would make the whole system into a mobile application that users would simply plug in and start practicing at any time. Additionally, we would improve the user interface after receiving copious amounts of feedback from the users. Once this is done, we would then improve the source code of the entire project to assure that it is running in the most efficient manner. Aside from core improvements, we would also like to add more functionality to Pitchy Perfect. Such additional functionality includes the capability to detect the key the user is singing in, adding courses so that users can practice several notes and scales, and allowing the user to select which real-time feedback sound they want to hear.

6. User Observation

In order to get some feedback on how the system should function, we sent three emails to three different potential users of Pitchy Perfect. We asked each participant three different questions: 1. "When hearing the note you are trying to sing being played back to you, does a simple sine wave suffice (like you heard in the demo) or would you prefer a different sound (such as a musical instrument playing that note)?" 2. "Were you able to detect a difference in the sound's position as you moved it around and if not what could we do better to help you perceive these differences?" 3. "In a system such as this what elements do you look for in the visual interface?". All three participants said that they were able to detect the sound moving in the correct manner, depending on what pitch they were singing relative to the target pitch. All three participants also said that they did not like to hear the sine wave for the real-time feedback while using Pitchy Perfect. Interestingly, with this question, they did give useful feedback on what they would like to hear instead of the sine wave. The first and second participants said that they would prefer to hear a musical instrument as feedback, while the third just stated that they just generally wanted to hear a different sound. The participants also gave interesting feedback in reference to the visual display. For this question, the first participant said that they would like to see "keys being played by an instrument". The second participant gave a more general response saying that they just wanted something to show how flat or sharp their voice is and some visual representation of the tune they're aiming for. The third participant said that they would like to see a "needle" to show how flat or sharp their voice is, in reference to the target note. These responses had some influence on our design, particularly when it came to the visual display. We created a visual system that effectively shows how far off the singer's voice is, in a manner that people are used to. Most people know how a meter works, whereby if the indicator is in the center, the user is on target. This is the concept we used to visually show if the user is in relative pitch or not. When it comes to the sound heard for real-time feedback, MatLab limited our options, in that it was not possible to use a sampled instrument sound or any sound other than the user's own voice being played back to them without significantly increasing the delay of the real-time feedback. Since it was clear that the performance of the system was more important than the style of feedback, we implemented the user's own voice being pinged back to them as feedback to minimize any delay.

Once Pitchy Perfect was completed, we collected some initial data to see how users perform when using it; compared to when using other non-3D Audio configured tuners. We found that users took slightly longer to correct the pitch of their voice when using Pitchy Perfect when compared to the time it took them to correct their pitch using a standard tuner. On average, it took users an additional 0.13 seconds to sing the correct note using Pitchy Perfect. This could be due to a number of different factors, three of which are the responsiveness of our system, the small sample size involved in the pilot study, or even because the users are not accustomed to using 3D Audio technology so it may take them longer to perceive the audio queues and make the proper adjustment in the pitch of their voice. In any case, we believe that it would be prudent to do an additional user study with a larger sample size to really see how Pitchy Perfect performs when compared to existing industry-standard tuners.

7. Challenges

Challenges we faced while working on Pitchy Perfect included mapping the frequency of the singer's voice to the movement of the slider in the graphical user interface, filtering the inputted data, minimizing any real-time feedback delay, and configuring a real-time feedback source. We also faced small performance issues with MatLab's ability to provide real-time feedback to the user without creating a significant delay. To solve this, however, we cleaned up our code to assure that MatLab was only processing what was necessary.

8. References and Citations

[1] PitchPerfector™. Product Overview. *PitchPerfector™ Discover the Secret to Singing on Pitch Easy & Fun… Like a Computer Game!* http://www.onpitchsinging.com/welcome/

[2] Roland. THE ULTIMATE POCKET VOCAL COACH. *VT-12 Vocal Trainer*, http://www.rolandus.com/products/vt-12/

[3] PolyTune 2 Mini. POCKET-SIZED PERFECTION. *tc electronic*, http://www.tcelectronic.com/polytune-2-mini/