Synthesizing Human Characteristics based on String Instruments

Vincent Zhang *

Shaker High School, Cohoes, United States

* Corresponding author email: yzhang68@albany.edu

Abstract. As music technology continues to evolve, new software enables us to digitally produce audio. Although plenty of progresses have been made in music genres (e.g., EDM or hip hop), many string instruments have yet to be thoroughly synthesized. Instruments, such as the cello, require incorporating human characteristics to evoke emotion. On this basis, Nyquist and Audacity are utilized in this paper to analyze and create the synthesized pitch of the cello. Some may argue that computers are incapable of human characteristics in music, but the results of this study clarify the feasibilities at some extent. Based on both visualizing and listening, the output creates a similar sound hardly indistinguishable from a human. It presents the possibility of expanding onto other string instruments including the violin, viola, etc. Instead of the necessity for an entire orchestra, anyone can have their own symphony at their fingertips. From movies, video games, or any form of entertainment, the advancements in this field of research could revolutionize the music industry. These results shed light on guiding further exploration of synthesizing human characteristics in context of string instruments.

Keywords: Vibrato; Crescendo; Ritardando.

1. Introduction

Computational music began in the early 1960s written by Max Mathews. His software, MUSIC, developed the concept of a “unit generator”, the “primitives” of performing sound generation and processing. In the 1970s, computer music expanded into hundreds of universities and research centers by the end of the decade. In 1985, Barry Vercoe, a computer scientist from MIT, developed a version of MUSIC that could compile with any computer that could accommodate the C programming language. Thus, computers around the world could access this new standard synthesis language [1]. By the start of the 21st century, many audio programming languages and audio editors were produced. In the year 2000, Roger Dannenberg and Dominic Mazzoni from Carnegie Mellon University released Audacity, an audio editor that can process all types of audios. Earning multiple awards for its contribution, Audacity has amassed over 100 million downloads since March of 2015 [2]. Recently afterwards, Nyquist, an audio programming language, was also created by Roger Dannenberg. Nyquist is capable of writing plug-in effects when used alongside Audacity [3]. Both software’s are still widely used contemporarily, which will be used in this study. Plenty of progresses have been made in this field of research throughout the recent years. In 2016, an artificial intelligence robot named Artificial Intelligence Virtual Assistant (AIVA) can create compositions that provide uniquely human qualities. AIVA is the first virtual composer to be recognized by SACEM (Society of Authors, Composers, and Music Publishers). From the world’s greatest composers, this robot has analyzed over 30,000 compositions and can formulate music similar to composers including Bach and Brahms. AIVA’s musical pieces are now used in films, advertisements, and game studios [4].

In 2019, members of the Music and Machine Learning Lab applied artificial intelligence to classify different violin techniques [5]. This study uses motion sensors to create a physical model. The technology can identify several different bowing techniques such as détaché, martelé, spiccato, ricochet, sautillé, staccato and bariolage. With 94 percent accuracy, this technology recorded both movement and audio data to output a result. Relying on merely movement or audio data in less consistent results. This study was carried out in hopes of finding the way that computers can improve the practices of musicians. In 2020, Hildur Guðnadóttir produced a cello-playing robot for her composition [6]. Hildur works on movie compositions, including *The Joker,* and uses this to represent a perception of the future. The robot physically holds the bow and change frequencies using the tuning peg. The piece is pre-programmed into the machine that outputs a consistent result of her composition. This is a great example of how technology is capable of playing these instruments.

Working with a physical instrument, however, can become inconvenient. The purchasing and maintenance of the instrument could cost thousands of dollars. Producing the machine that plays the instrument is another hurdle to overcome. The instrument must also be frequently tuned, and the bow must be rosined. In summary, using an instrument present many “variables” that are inconvenient. Rather than relying on a physical instrument, it might be possible to solely rely on synthesizing the instrument. If this can be achieved, composers are able to produce high quality music without the need for an abundance of money. On this basis, this study analyzes and creates the synthesized pitch of the cello in terms of Nyquist. The rest part of the paper is organized as follows. The Sec. 2 will introduce the specific procedure and codes logic. Subsequently, the evaluations will be carried out in Sec. 3. Afterwards, the applications as well as the limitations of this study will be demonstrated. At last, a brief conclusion will be given to summarize the results.

2. Methodology

2.1 Procedure

The procedure of the project involves choosing a composition to recreate, analyzing audio, synthesizing music, and evaluating results. Since the objective concentrates on mimicking human characteristics, it is preferable to devote the project to an expressive piece of music. “The Swan” composed by Camille Saint-Saëns involves changes in tempo, volume, and vibrato uniquely creating a challenge for computers to recreate. In order to synthesize a composition designated for the cello, it is important to understand how the sound is created.

String instruments have tension tuned to a specific frequency on each string. Using the fingerboard, the musician can shorten or lengthen the string affecting the pitch. It can be easily accomplished through Nyquist simply by changing a couple of numbers. The challenge, however, is recreating the techniques used when playing string instruments. Techniques give a sense of emotion to the composition that infamously lacks in computers. Evoking emotion is a fundamental part of any song; without it, the resulting sound will sound dull or “robotic.”

Knowing what should be accomplished, the study will involve referencing the audio of different cellists playing the piece. The goal is to produce music that is indistinguishable from a human and computer. Thus, the next phase involves examining how and when techniques are used to make the piece sound good. By identifying patterns in the playstyle of other cellists, one can establish the most optimal uses of these techniques. Subsequently, Audacity allows us to input and visually interpret the audio. As shown in Fig. 1, amplifying the frequency allows to indicate how soft or loud a pitch is played at a specific time. The graph also offer a way to identify the magnitude and speed of the vibrato used by the cellist. Ideally, this software is used to then compare the synthesized and original audio to formulate similar waveforms.

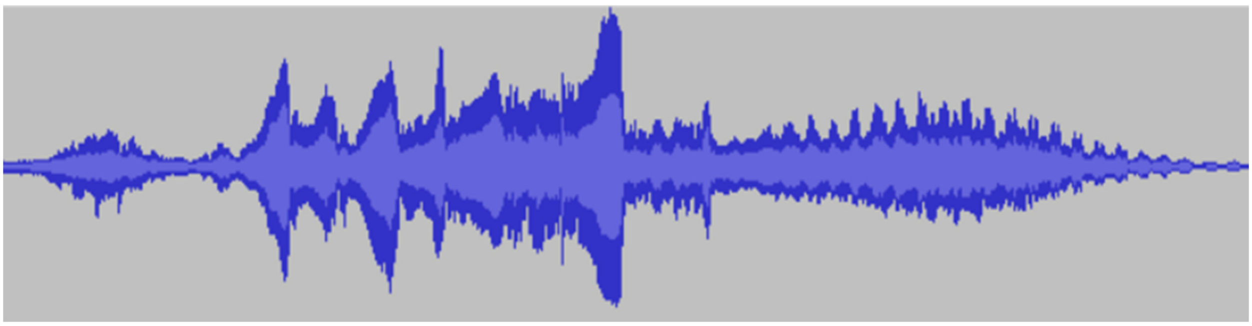

Fig 1. Audio Sample from Audacity.

The next step involves experimentation which is done to determine whether these results can be achieved using these software. This involves determining whether or not a “cello-like” sound can be synthesized. By using built-in functions within Nyquist and reducing its frequency, it is possible to imitate the low tone. Hence, it is possible to recreate the sound of a cello.

2.2 Codes

Fabricating the code in Nyquist consists of using both predefined and newly defined functions. The method that will be used to complete this project consists of creating new functions with parameters accustomed to their uses. For instance, a function that is used to play the note G sharp can have parameters that adjust its length and vibrato speed. Once the structure of the function is defined, we can use the predefined methods, such as bowed-freq and lfo (controls vibrato speed and weight), to match the parameters. Envelope functions and pwl functions are also used to change the loudness of a note at a specific time. Pwl and env functions change the amplitude of the waves affecting its volume. The values for the functions were determined by the wave amplitude recorded from the plots in audacity. By plotting the spectrum, one can record the decibel values at a given time and use the values within the envelope function. In such a manner, the synthesized note will be nearly identical to the audio recording. The spectrum also allows one to record its frequency determining the pitch of each note practically being pitch perfect. Once the basic functions are written, specific functions are made to account for unique scenarios such as a scale. Since creating functions for each note is inefficient, this allows the user to change the pitch to efficiently synthesize a scale.

Although the body of the code is finished, an issue arises regarding the splices of the notes after each line of code. Whenever a function is executed, there is a noticeable small break in the audio known as a grain. To resolve the issue, Nyquist offers the sim function allowing functions to be played at a specific time. As illustrated in Fig. 2, the wavelengths show the result of smoothing the audio. Similar to a score function, sim is designed for audio with multiple notes. By slightly overlapping the times, the grains become less apparent, smoothing the audio. Alternatively, Audacity has a feature that amplifies the sound removing cuts in the audio. The sim function is also used to alter the tempo; as the tempo accelerates, the time in between each note can be shortened to the user's preference. In this case, there is an accelerando which prompted the use of the function. By smoothing the grains and copying the time frames, the program is able to synthesize a more expressive output.

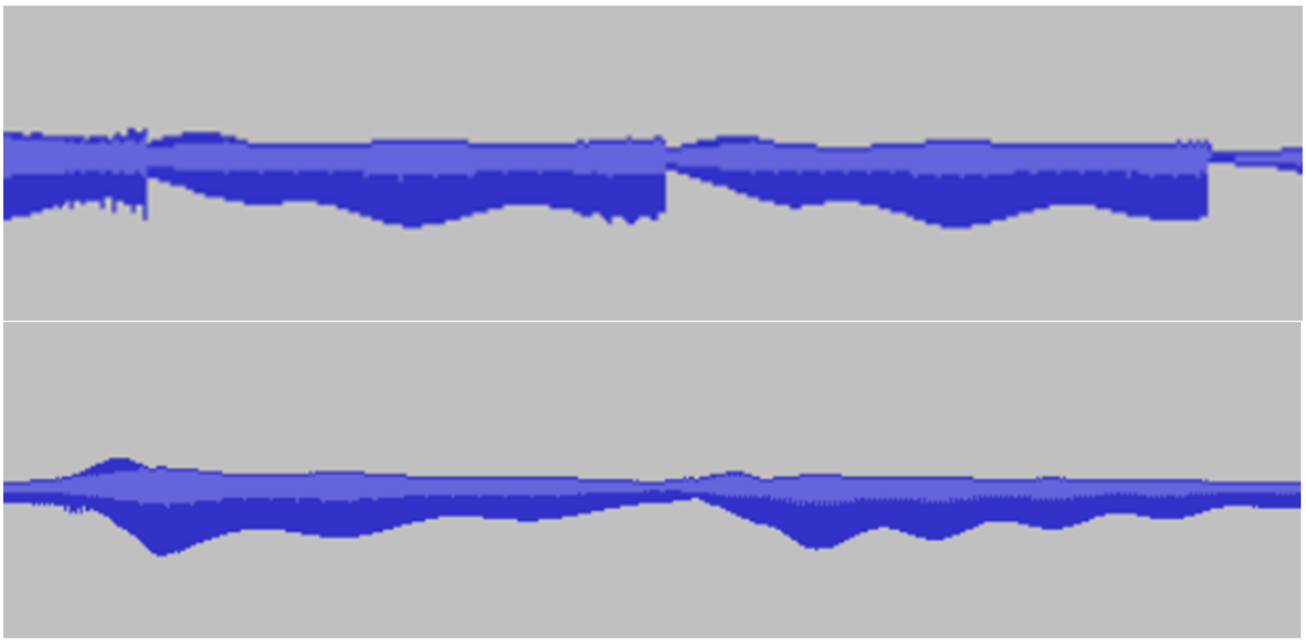

Fig 2. The results of before (upper panel) and after (lower panel) smoothing.

3. Evalutions

In evaluating the results, there are still a few areas that require polishing. The code should be more concise by reducing the need for unnecessary parameters. Rather, it would be more reasonable to create a base function of a note and use other functions to alter pitch, vibrato, length, etc. By approaching this way, the code can be much more organized and easier to write for future purposes.

Another challenge includes having each note sound more realistic. As of now, the code relies on bow-freq, which is a start but does not sound similar enough to a real instrument. Based on other software, it might be possible to output a more realistic tone. There are also a couple miniscule grains in the audio. The grains can be slightly solved by overlapping the notes in the sim function, but it is not perfect and difficult to account for while managing the tempo. The code should also contain more functions for other purposes in the future. For instance, giving a note an accent can be a function stored in the code for other songs. These are some of the problems that can be resolved in the future.

Although the outcome is not perfect, it is believed this work is successful because this study proves synthesizing realistic string instruments is achievable. The code was able to achieve techniques including vibrato, ritardando, crescendo, and other fundamental techniques for string instruments. When comparing the original to the synthesized audio in Audacity, the spectra show a similar output in both scenarios (seen from Fig. 3 and Fig. 4). Taking a closer look, the original audio shows more "randomness" than the synthesized audio because a performer is less concise than a computer. This is visually depicted by the vibrato which stays consistent in the synthesized graph, but less consistent in the original. Despite this, more fine-tuning can be done to adjust these minor problems. These graphs allow entire pieces to be replicated opening the possibility of creating audio that surpasses its original.

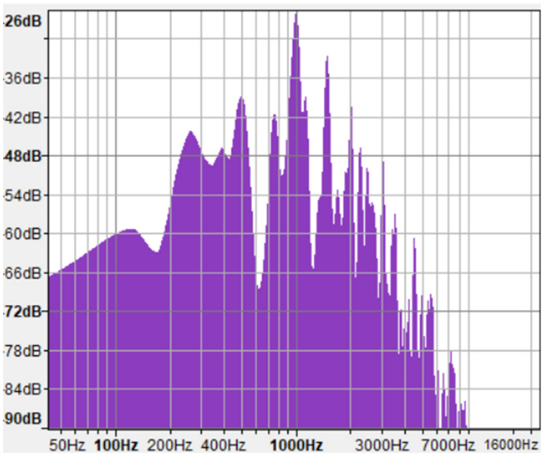

Fig 3. Original Audio

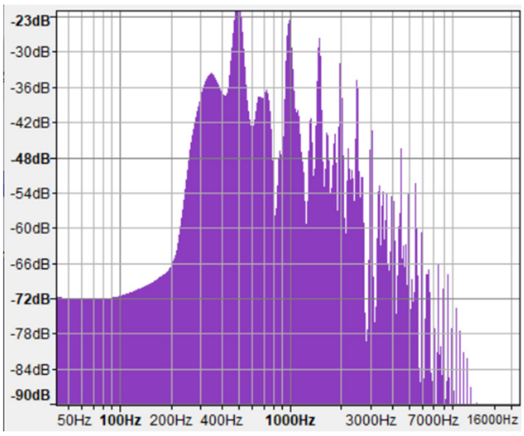

Fig 4. Synthesized Audio

4. Applications

Since this study solely focuses on the cello, these results can motivate the imitation of other instruments. Particularly, the violin, viola, and bass are three string instruments similar to the cello. Once any instrument can be synthesized using a software, the ultimate goal of artificial intelligence is the final step. As artificial intelligence already has the capability to compose music, artificial intelligence now has the potential to synthesize compositions. Access to this technology can allow music producers to own a cheap and efficient method in creating music. The costs of musicians, instruments, and other factors may mitigate the integrity of a producer's work. To put this into perspective, symphony orchestras will cost around 50,000$ to 100,000$ for a couple of rehearsals [7]. This charge rate neglects the possibility of composition changes or other factors that might exceed one's budget. In context of this technology, a singular computer can synthesize an entire orchestra, with any number of instruments, to one's desire without the need of thousands of dollars. Not only does it benefit the producer, it benefits its audience in ensuring a quality result is produced. As stated previously, synthesizing music allows one to perfectly create their desired result. The possibilities are limitless avoiding hindrances, such as human error, to disrupt the outcome. Imitating these human characteristics is an under-examined field that could eventually surpass that of human talent itself. Eventually, artificial intelligence will be capable of playing compositions without the need for human dependency. This technology will expand the knowledge of music and forever change the future of the music industry.

5. Limitations and Prospects

Music technology has already shown its potential in the past, for example the production of Shazam. Shazam Entertainment Limited initially released in 2002 and is now valued at nearly 1 billion dollars [8]. This software relies on AI to record and compare sound waves from its database. With over hundreds of millions of uses, Shazam quickly rose to be major contributor in the music industry. Its profound effect changed the way consumers interact with music benefitting the artist as well. Other artificial intelligence (e.g., AIVA) can now understand and compose emotions. It is only a matter of time for AI to start synthesizing compositions. Along these lines, the outlook of this study aims towards creating an AI that can play not only string instruments, but any instrument by itself while factoring in human emotions, one that is indistinguishable from another human. Instruments such as the flute or clarinet also possess many techniques similar to string instruments that effect the emotions of a composition. The thought of this idea is feasible knowing that other AI have already accomplished similar feats. Instead of applying artificial intelligence to compose emotions, we can instead use it to synthesize emotions.

Releasing this technology would enhance productivity, quality, and creativity for artists in the music industry without the need for much human dependency. Scholars argue that around 20-30% of songs in the top 40 chartbusters relied on the use of AI [9]. From video games, movies, and other entertainment, composers can use this technology as a leverage to create soundtracks. The need for expensive instruments, studios, and equipment can be resolved utilizing a single computer. Demand for string instruments is also at an all-time high. According to the International Music Products Association, around 54% of households has a member that plays a musical instrument. Another survey showed that 85% of people have regretted not learning how to play an instrument [10]. Unfortunately, many of those who have regretted not learning an instrument was due to the lack of resources. According to previous study, around 41% of low-income families were unable to purchase an instrument [11]. Although this technology does not provide a physical instrument, it could better expose computer music to anyone in need. Upon closer examination, it is proven that this project can reach a wide range of people as music is our universal language. If this technology were to be completed, it will undoubtedly change the lives of consumers and artists around the world.

6. Conclusion

In conclusion, this paper investigates the synthesizing of human characteristics in string instruments based on its human counterpart. Specifically, this study demonstrates the possibility of computers evoking emotion and exceeding human talent. According to the analysis, new technology allows one to nearly mimic sounds created by a cello. In addition, this research can expand to imitate any string instrument using the same method. Moreover, these synthesized instruments can be played in uniform creating endless possibilities. Nevertheless, there is much more that needs to be improved. Fine-tuning the base cello sound, creating more functions for other techniques, and smoothing of audio are steps towards accomplishing this goal. In the future, artificial intelligence creates the opportunity to rid of human dependency. Technology will be able to interpret human emotions and evoke them through the means of music. By using original audio as a basis, computers give the chance to rewrite and improve what has already been achieved. In this case, technology can exceed that of human talent. Overall, these results offer a guideline for the future of music technology in synthesizing string instruments.

References

[1] Collins N. Introduction to computer music. John Wiley & Sons, 2010.

[2] Mazzoni D, Dannenberg R B. A fast data structure for disk-based audio editing. Computer Music Journal, 2002, 26(2): 62-76.

[3] Dannenberg R B. The implementation of nyquist, a sound synthesis language. Computer Music Journal, 1997, 21(3): 71-82.

[4] Zulić H. How AI can change/improve/influence music composition, performance and education: three case studies. INSAM Journal of Contemporary Music, Art and Technology, 2019, 1(2): 100-114.

[5] Artificial Intelligence Enables Recognizing and Assessing a Violinist's Bow Movements. Medical Xpress - Medical Research Advances and Health News, Medical Xpress, 3 Apr. 2019, Retrieve from: https:// medicalxpress.com/news/2019-04-artificial-intelligence-enables-violinist-movements.html.

[6] Asprou, Helena. “Watch a Cello-Playing Robot Perform Hildur Guðnadóttir's Eerie Music.” Classic FM, Classic FM, 24 Mar. 2020, Retrieve from: https://www.classicfm.com/discover-music/ instruments/ cello/ playing-robot-hildur-gunadottir/.

[7] Correspondent, Travis Rivers. Symphony for Rent; Bargain Rates (Rehearsals Extra). Spokesman.com, The Spokesman-Review, 2011.

[8] News, Bloomberg. “Music-Discovery Company Shazam Valued at Roughly $1 Billion.” The Wall Street Journal, Dow Jones & Company, 21 Jan. 2015, Retrieved from: https://www.wsj.com/articles/music-discovery-company-shazam-valued-at-around-1-billion-1421780459.

[9] Cope D. Music, artificial intelligence and neuroscience. Handbook of Artificial Intelligence for Music. Springer, Cham, 2021: 163-194.

[10] Carlsbad. Calif. Gallup Organization Reveals Findings of ‘American Attitudes toward Making Music’ Survey. NAMM.org, 2003, Retrieved from: https://www.namm.org/news/press-releases/gallup-organization -reveals-findings-american-atti.

[11] Eira Winrow PhD Research Candidate and Research Project Support Officer, and Rhiannon Tudor Edwards Professor of Health Economics. “Poorer Children Priced out of Learning Instruments but School Music Programmes Benefit the Wider Community.” The Conversation, 3 Sept. 2020, Retrieved from: https://theconversation.com/poorer-children-priced-out-of-learning-instruments-but-school-music-programmes-benefit-the-wider-community-104181.