Essay: Music x Virtual Instruments

S2163004 Jeannette Shijie Ma

1. An acoustic music supply chain

Few years ago, I attended a course of basic classical musicology. The lecturer introduced us to Sibelius - a score writer program. He told us how efficiently this program helped him in composing. Back to his time as a composition student, the rehearsals of a new composition were extremely expensive and time consuming. It took all his money and time to find players for his graduation work. Piano was the only tool to check his work before rehearsals, and it was not possible to check it with different timbres.

For an orchestral piece, if the composer is the writer, conductors are the performers who bring the piece to live. He or she adjust how each note and phrase sound during rehearsals. As the clarinet players play clarinet, the conductor plays a whole living organism. Some living organism (e.g. a professional orchestra) can be easier to "tune" than some others (e.g. an amateur orchestra). Nowadays a composer might be able to introduce his or her new work easily to anyone (who doesn't even read music) with a well-typed music score which allows computer software to play back. However, conductors still need to rely on instrument performers to express their understanding of the pieces.

When playing classical music as an orchestra member, it is usually good to see ourselves as a part of the whole living instrument. The most important job in this case is to be a well-tuned unit who follows the conductor. However, conflicts could happen when an orchestra member wanted to express his or her own understanding of the music, which is not the same as the conductor's. On the other hand, it could be very difficult for some players to fully understand what kind of sound the conductor wants.

While being an audience of acoustic music, you may have the experience of finding some parts of the performance/recording less favored than the others. For example, when you have your favorite recording of a piano concerto by Rachmaninoff, but you just don't like certain few bars. All you can do with this recording is tolerating these bars and appreciate the others.

In this long music supply chain, we can find interesting stores going on, while other interesting possibilities were basically impossible to happen without technology aid.

When we look at the field of electric music, the supply chain seems much shorter. With the aid of music synthesis technology, new workflow was developed. Nowadays most music producers are a combination of composer and conductor, who write midi compositions, adjust a lot of parameters, and then publish them. Their works can also be performed live by themselves or other people. All these could be amazingly done with only one person, approaching the elimination of all imperfections described by Mondriaan in his art theory.

As music synthesis technology greatly influencing music production and adding new possibilities to it, this new style of production also influenced the trend of technology development. People might already get used to productions using artificial sounds which sound artificial. However, what if acoustic composers also want to join the party using the sounds

that they are familiar with? What if a conductor also wants to express his or her idea of a piece without worrying about tuning the performers? What do different people in acoustic music supply chain want from technology? One of the answers is virtual instrument.

2. Virtual instrument technology

There have been many researches about timbre synthesis which drove the development of virtual instruments. In this part I would like to briefly discuss some related techniques: sample-based synthesis, physical modelling synthesis, and data-based synthesis. For each technique, I will give some application examples and form a discussion on it.

2.1 Sample based synthesis

There are a lot of instrument libraries and plugins available for digital audio workstations. Besides traditional instruments, numerous sound effects of different themes (e.g. factory) can also be downloaded as library or plugins. Most of them could be applied so well that an average audience won't be able to distinguish it from live recordings. Among all these instrument libraries, I found vocal libraries the most interesting ones to discuss about.

Timbre of different singers vary a lot more than instruments produced by different brands. There have been very powerful vocal libraries focusing on one singer (e.g. Clara's vocal library). There are also programs that offer multiple characters (e.g. Vocaloid), and tools to customize your own libraries (e.g. Utau). These libraries breed a unique culture of virtual singer music since the birth of Hatsune Miku (the first Vocaloid virtual singer). Enormous music productions were made with the application of virtual singers [1]. There were even live concerts given by these virtual singers which attracted many audiences.

When comparing the music made by Vocaloid and other big vocal libraries, the later candidates usually sounds more realistic. The artificial sound seems to be an important feature which users liked about Vocaloid. This situation might leave Vocaloid an awkward decision to make: to go more realistic? or keep making robot-like synthesis?

Despite the conflict between music culture and technology development for Vocaloid, sampling-based synthesis is still the most widely used method for virtual instrument, because it has relatively simple concept and higher efficiency [2]. Since it is mostly another way of playing back recordings. The technique is more related to timbre analysis instead of timbre synthesis. To some extend it is actually making robot “playing" instead of “singing".

2.2 Physical modelling synthesis

In order to make robots "sing", another important idea was physical modelling synthesis. It uses mathematical models to mimic the physical feature of an instrument in order to resemble the sound. Unlike sampling, this technique required much more work in model building for each instrument. Non-pitched percussions could be the most extreme example where sampling is much more efficient than physical based synthesis. However, with more timbre and pitch variation and complexity, physical based synthesis shows more its advantage. Timpani is one of the pitched percussions which provides many different timbres for orchestration. It was much more difficult to build a workable sample-based plugin for timpani than a drum kit. However, recent studies made breakthroughs in 3D modelling in synthesis of timpani sound [3].

PianoTech is another excellent application example of physical modelling synthesis. Its latest version successfully resembled very detailed sounds in real pianos. For example, the harmonics resonates at very right end of the keyboard; the sound of keys and pedals; the microphone position effect. These features could be far less possible with a sample-based piano library.

I couldn't find any application example of virtual singer based on physical models. However, there were already been many researches on vocal track model-based speech synthesis [4]. It might be (practically) difficult to apply the model from speech to singing since these two actions behave very different from each other [5]. However, in theory, a well-built physical model of human vocal system could be able to sing as well as it talks. Physical models could especially benefit realistic singing synthesis, since singers produce different timbre at different singing pitch and singing style, with different positions of oral system. There was a new feature of Vocaloid 5 which allows the users to adjust the singing skills of the virtual singers. This function might perform better with physical modelling synthesis than sampling.

2.3 Data training-based synthesis

Artificial intelligence has been a relatively new approach in music production. As it is said: "When you copy one work, it is plagiarism; when you copy a thousand of work, that becomes creativity". Deep learning can help robots to find their own perceived patterns in music samples and generate its own style of playing or singing. Take the Microsoft XiaoIce as an example [6]. This virtual assistant was initially not trained as a singer, but after training on vocal data, she amazed the audiences by her realistic and unique human voice in her performance. Her voice sounds much more natural than similar type of (human composed) virtual singing in Vocaloid, mostly in terms of Chinese pronunciation and phrase connections.

However, it is still a question mark whether majorities of audiences would appreciate this type of realistic virtual singing. Deep learning based virtual instruments can sounds very realistic, but they are also less under control. They are capable to do everything without interferences from human creativity. Thus, it might be less useful as a tool assisting the whole music supply chain, from composer to audience.

3. Conclusion

In my opinion, the future of virtual instrument researches will be likely favor physical modelling synthesis with additional data leaning from timbre analysis. Sample based synthesis could evolve for unique music culture-wise applications instead of pursuing realistic sound.

The development of virtual instruments could also reflect back to music supply chain in the long term. There have always been music lovers who doesn't play any instruments, but good at express themselves' perceptions using timbre judgement. Being imaginary, the future of virtual instrument technology would allow any music consumer in the whole supply chain to customize the music product by their own preference. Despite the issue of copyright, inspiring collaborations could happen under this situation. For example, in an “online orchestra”, a bassoon player could “play” his or her part by just editing the bassoon part. A bass player could do the same in the same music file.

As computer plays for us, everyone can join a conceptual "rehearsal", “performance” or “conducting” regardless of our technique of instrument playing. All we need to know is to communicate our perception (how we want the notes to sound) and assign the tasks. Therefore, social communication should also find its way to develop together with technology. If people still doubt how to describe the sounds they perceive or want, it might be a huge confusion for machine to learn. There are recent researches looking into the perception communication of timbre (e.g. by reading semantic compositional language [7]).

However, who knows who will dominate and lead perception communication? If technology goes faster enough, machines might find themselves a way to communicate to our feelings before we know how to express ourselves. We might then have robots surprisingly know what we want and sing deep into our hearts.

References

1. Kenmochi, H. (2010). VOCALOID and Hatsune Miku phenomenon in Japan. In Interdisciplinary Workshop on Singing Voice.
2. Bonada, J., & Serra, X. (2007). Synthesis of the singing voice by performance sampling and spectral models. IEEE signal processing magazine, 24(2), 67-79.
3. Bilbao, S., & Webb, C. (2012, September). Timpani drum synthesis in 3D on GPGPUs. In Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12), York, United Kingdom.
4. Cook, P. R. (1993). SPASM, a real-time vocal tract physical model controller; and singer, the companion software synthesis system. Computer Music Journal, 17(1), 30-44.
5. Vijayan, K., Li, H., & Toda, T. (2019). Speech-to-Singing Voice Conversion: The Challenges and Strategies for Improving Vocal Conversion Processes. IEEE Signal Processing Magazine, 36(1), 95-102.
6. Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2018). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. arXiv preprint arXiv:1812.08989.
7. Wallmark, Z. (2018). A corpus analysis of timbre semantics in orchestration treatises. Psychology of Music, 0305735618768102.

Motivation

When I was leading tasting panels for my previous study, I strongly experienced that we are training a group of humans to be machine-like precise and accurate. In the end we rely on a bit of statistics and a lot of luck to translate from this group to the larger society. I wondered maybe we can help making this communication process less awkward with new available technologies. Therefore, I decided to come to media technology to learn more about new technologies and ubiquitous sensorial communication.

This might be already mentioned in my media tech application, but I didn't expect to find connection between my research interest and this music course (which I chose merely because I like music). After some class discussions and the timbre synthesis exercises in SSI, I found much interests in timbre perception and synthesis technology (for me it is like some kind of translations between human and machine).

After reading some materials, I chose to focus more on the current virtual instrument technology in which I personally found a wide and interesting discussion space.