Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/68/20241431

Synthesis of natural sound based on Nyquist

Boya Zhang

Affiliated High School of South China Normal University, Guangzhou 510000, China

The corresponding author's e-mail address: zhangby.ricky2022@gdhfi.com

Abstract. In recent years, sound synthesis has become an important part of daily life with the rapid development of the technology. There are many methods of natural sound synthesis, but each of them has its disadvantages. This paper explored a novel method of natural sound synthesis based on Nyquist. This method has significant advantages in high speed, convenience, flexibility, randomness, and small file size. The research investigates the effectiveness of this method by synthesizing four distinct natural sounds - rain, wind, thunder, and a combination of all three - using unit generators in Nyquist. To further evaluate the result, Mel spectrograms and a questionnaire are applied. The evaluation shows that the synthesized sounds meet expectations and are relatively real and natural. This method can be combined with other technology, such as the large language model, to provide better accessibility. This method can generate natural sounds for many tasks, especially those requiring high speed.

Keywords: Natural sound synthesis, Nyquist language, unit generators.

1. Introduction

Sound synthesis has been an active topic for over a century. In 1896, Thaddeus Cahill invented the Telharmonium, an electrical organ [1]. It is considered the first electronic musical synthesizer. The Voder, the first machine to synthesize human voices, was invented in 1940 [2]. In the latter half of the 20th century, digital methods for sound synthesis appeared, such as mathematical models and signal processing [3, 4]. In recent years, with the development of artificial intelligence, the field has experienced a technological revolution. In 2016, the publication of WaveNet, an autoregressive neural network, illustrated that deep learning is a possible methodology for sound synthesis [5]. Recently, sound synthesis has become an important part of many aspects of daily life, such as music, films, and video games [6]. Natural sound synthesis is becoming more important as a specific topic in sound synthesis. Natural sound synthesis is to generate natural sound using electronic equipment. Natural sound synthesis has many methods, just like other kinds of sound synthesis.

Most of the recent methods of natural sound synthesis are based on AI models. In Visual to Sound: Generating Natural Sound for Videos in the Wild, natural sounds are generated from videos in the wild using three SampleRNN models [7]. Evaluations show that most of the sounds generated from these models can confuse people into thinking they are real, indicating they are highly realistic. Onoma-to-wave: Environmental Sound Synthesis from Onomatopoeic Words proposed a method of environmental sound synthesis from onomatopoeic words [8]. By using a long short-term memory encoder and decoder in the seq2seq framework, natural sound with high realism and diversity can be generated. In Diverse and Vivid Sound Generation from Text Descriptions, natural sounds can be generated from text

DOI: 10.54254/2755-2721/68/20241431

descriptions using a Variation-Quantized GAN and a Transformer model [9]. The evaluation shows that the method can generate high-quality and diverse sound content that corresponds to the text.

Some methods of natural sound synthesis do not use AI models. The Synthesis of Environmental Sound Textures by Iterated Nonlinear Functions, and its Ecological Relevance to Perceptual Modeling illustrates the possible usage of iterated nonlinear functions in synthesizing natural sounds [10]. In Sound Synthesis Based on Ordinary Differential Equations, the authors explored the possibility of sound synthesis using ordinary differential equations [11]. In Efficient Sound Synthesis for Natural Scenes, sounds of waterfalls and oceans are generated based on physical simulation. It successfully applied a mathematical model to sound synthesis since its speed is much higher than the previous frameworks [12].

Although there are plenty of methods to synthesize or obtain natural sounds, each of them has its drawbacks. For example, playing the recordings of natural sounds may be the most efficient way, but it has no randomness and flexibility, takes large storage space, and sometimes requires complex recording methods. On the other hand, generating natural sounds using AI models provides great randomness, flexibility, and accessibility. However, it requires a long time and strong computation power, which reduces its efficiency and makes it unsuitable for tasks that require instant generation. Using mathematical models may solve the problem of low speed, but they are usually too complicated to use and lack flexibility. Therefore, it is important to discover a method that balances efficiency, flexibility, and accessibility. In this research, four natural sounds, including rain, wind, thunder, and their mixture, are synthesized using unit generators in Nyquist. The reason of choosing Nyquist and the procedure of sound synthesis is introduced. The synthesis results are evaluated using Mel spectrograms and a questionnaire. This research discussed the merits and demerits of the synthesis results and provided suggestions for modification. Finally, it discussed the future outlook and possible applications of this technology.

2. Method

2.1. Software

Nyquist is used to generate the sounds for this research. Nyquist is a programming language designed for sound synthesis and music composition. Using Nyquist to generate sounds has many advantages. First, it is an integrated system that can deal with signals, synthesis, and events. This gives Nyquist a strong ability to generate sounds. Second, Nyquist is a programming language. This provides it with a high speed in generating sounds and a small file size. Third, there are lots of unit generators in Nyquist. These unit generators are easy to use and adjust. This ensures that Nyquist is flexible and easy to use. Therefore, Nyquist is suitable for this research.

2.2. Sound Synthesis Method

Four natural sounds are synthesized using unit generators in Nyquist, including noise generators, oscillators, bandpass filters, lowpass filters, frequency modifiers, amplitude envelopes, and feedback delays. Noise generators, bandpass filters, and amplitude envelopes are used to synthesize rain sounds. The sounds of rain are mainly caused by the collision of raindrops with the ground. Each collision provides a pulse of sound. Since there are too many raindrops colliding with the ground in a short time, the rain sound is similar to noises with bandpass filters. Therefore, noise generators and bandpass filters are used to simulate rain sounds. Additionally, amplitude envelopes are used to simulate the change in the strength of rain with time. Synthesizing wind sounds used oscillators, noise generators, frequency modifiers, bandpass filters, and amplitude envelopes. First, a sine wave is generated using an oscillator. Then, its frequency is modified using a noise sound. Next, the sound is passed through a bandpass filter with a center frequency the same as the initial sine wave. This creates an unstable whistle sound similar to the whistle of wind. Finally, wind sounds with different frequencies are combined, and amplitude envelopes are used to simulate the variety in the strength of the wind.

The thunder sounds are synthesized with noise generators, lowpass filters, amplitude envelopes, and feedback delays. First, it generates a noise sound modified by a lowpass filter and an amplitude envelope to simulate the initial explosive sound of thunder. Then, feedback delays with random delays and lowpass filters with lower frequencies are applied to the sound. This simulates the random reflection and loss of the high frequency of the thunder sound. A mixture of the three sounds above is also synthesized. The rain sound is the main sound, the wind sound is an ornament, and the thunder sound joins at the seventh second.

2.3. Evaluation Method

Mel spectrograms are used to evaluate whether the sounds meet expectations. They can show the change of amplitudes in each frequency with time, which gives information about the timbres of the sounds. A questionnaire is used to evaluate the realism of the synthesized sounds. The questionnaire included the four synthesized sounds and four questions about the reality of each sound, each scored 0-10. If the scores are higher, the participants think the sounds are more real. A score of 10 means the sound is exactly the same as reality, and a score of 0 means the sound is totally unreal. The participants will listen to the four sounds, evaluate them, and answer the questions by filling in their scores. The questionnaire was posted on several Chinese social media, such as WeChat, QQ, and Baidu Tieba. The Wenjuanxing website is also used to invite random people to fill out the questionnaire.

3. Results and discussion

3.1. Sound synthesis results and Mel spectrograms

The four sounds described in Section 2.2 are successfully synthesized. Each sound takes less than 0.5 seconds to generate. Three Mel spectrograms are generated to evaluate the rain, wind, and thunder sounds. Figure 1 shows the Mel spectrogram of synthesized rain sound. The frequency is widespread, with a center at 2000 Hz initially. As time passes, the sound becomes louder, and a new center at 1000 Hz is added. This resembles the rain getting larger. Finally, the sounds become smaller. Overall, rain is a widespread noise that changes loudness and frequency over time, making it sound real and dynamic. Figure 2 shows the Mel spectrogram of synthesized wind sound. The spectrogram shows that there are two parts in the sound: wind whistling sound and wind noise. The major frequencies of the wind whistling sounds are around 523 Hz and 679 Hz, which creates an unstable chord. The amplitudes and frequencies of the wind whistling sound have slight random changes, creating an unstable and scary sound. The wind noise is not as widespread as the rain sound. It has a relatively low frequency. The noise makes the wind sound more real. Figure 3 is the Mel spectrogram of synthesized thunder sound. The spectrogram shows that the thunder sound is a low-frequency noise. Both the loudness and the frequency of the sound decrease over time, simulating the traveling and dissipation of thunder. However, the reflection of the thunder sound is not obvious on the spectrogram. Although there are some sudden increases in low frequencies, these changes are too slight and flat, making the thunder sound like one smoothly decreasing noise, not many separate explosive sounds. This defect may reduce the realism of the thunder sound.

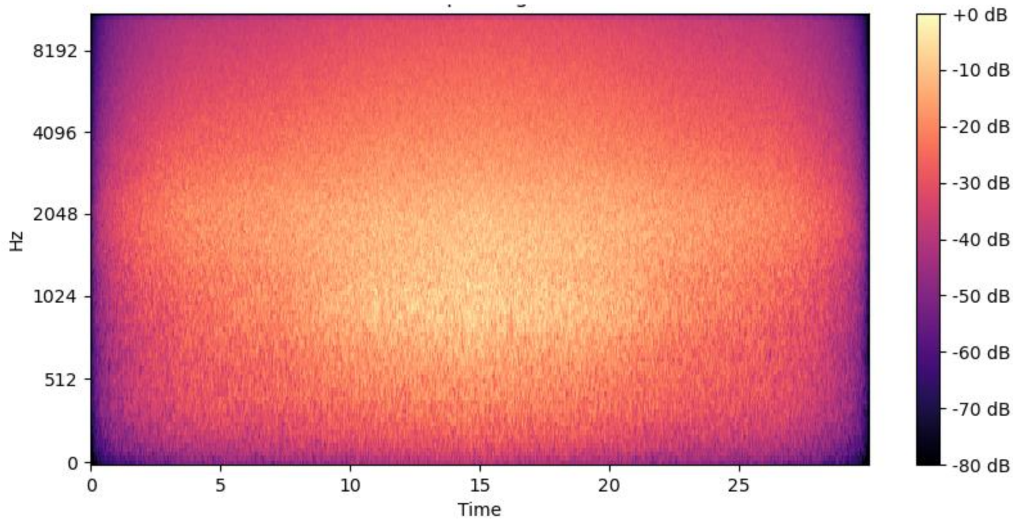

Figure 1. The Mel spectrogram of rain sound (Photo/Picture credit: Original).

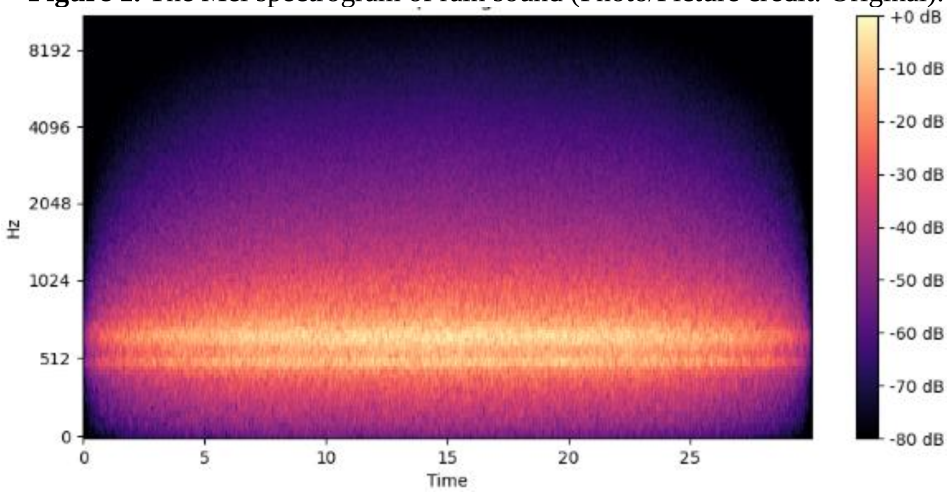

Figure 2. The Mel spectrogram of wind sound (Photo/Picture credit: Original).

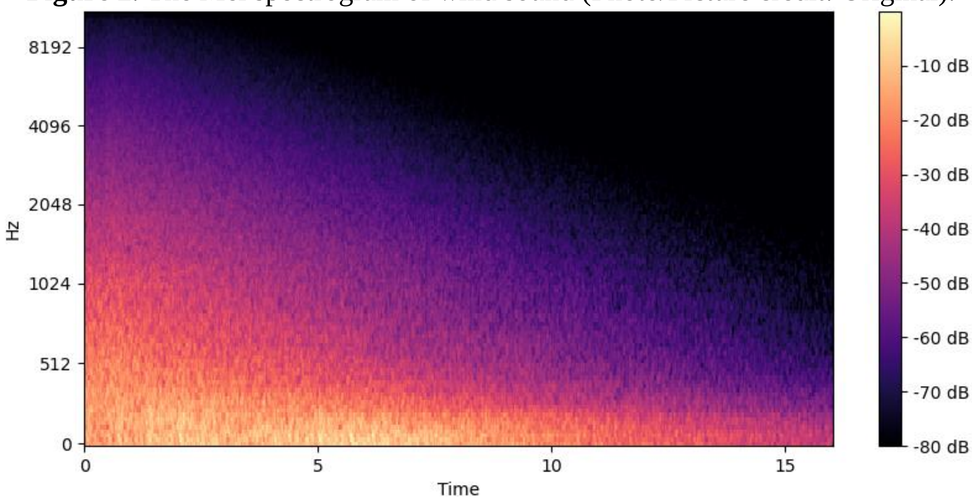

Figure 3. The Mel spectrogram of thunder sound (Photo/Picture credit: Original).

3.2. Questionnaire result

A total of 41 people answered the questionnaire. Figure 4 shows the average score of each sound. The scores of the sounds are between 6.61 and 7.68, showing a relatively high realism of the sounds.

Nevertheless, there are still considerable shortcomings in each sound. Among the rain, wind, and thunder sounds, the wind sound has the highest score of 7.10. Its high realism may be due to its complexity and randomness. However, a potential disadvantage of the wind sound may be that it is too flat and lacks dynamic. The rain sound scores 6.61. One reason it is less realistic is that only the noise is synthesized, without other sounds in the rain. Without the sounds of single raindrops, the sound is more like an artificial noise and less like a natural rain. The thunder sound has the lowest score of 6.17. As Section 3.1 mentioned, the thunder sound is too smooth and continuous, which is different from the thunder in reality. Another reason is that the reverberation of the sound is too thin and weak, making the sound drab and not real. An unexpected result is that the mixture of three sounds received the highest score of 7.68, much higher than both three separated sounds. One possible explanation is that the human ear is more likely to hear the whole and ignore the details. Therefore, when the sounds are mixed, the flaws in the mixture are ignored, which makes people feel more real. However, the causes of this phenomenon may be more complicated and need further research.

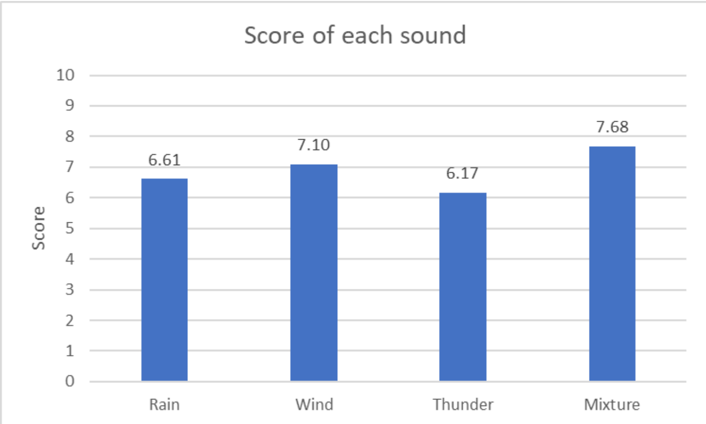

Figure 4. The average score of each sound (Photo/Picture credit: Original).

3.3. Explanations

The Mel spectrograms and the questionnaire show that the synthesized sounds generally meet expectations and are relatively real. There still exist flaws, but there are some possible solutions to improve the sounds. For the rain sound, the sounds of individual raindrops can be added to make it more natural. For the wind sound, more tones of whistling and more unstable noises can be added to make it more unstable and dynamic. For the thunder sound, the unit generators should be better chosen and adjusted to better simulate the rumble sounds of thunder. The high score of the mixed sound indicates that the technology is better used as part of more complex sounds or music. The mixture of three sounds has a higher score than any of the individual sounds, demonstrating the potential for greater realism by combining synthesized sounds with other sounds. By using these synthesized sounds as backgrounds for other sounds or pieces of music, they can give audiences a good experience of realistic nature sounds. To sum up, this research has revealed a strong possibility of synthesizing natural sounds with Nyquist. By utilizing the possible solutions discussed above and combining synthesized sounds with other sounds, greater realism in sound synthesis can be achieved, which can be used for various applications.

3.4. Limitations and prospects

There are still some deficiencies in the research. In the sound synthesis part, the method of synthesizing sounds should be improved to meet the expectations better and become more realistic. Adjustments to the methods and further exploration of new methods are needed to reach that goal. In the evaluation part, the questionnaire is too simple and lacks objectivity. The design of the questionnaire should be improved to evaluate the result better. For example, a real natural sound should be added as a contrast for each synthesized sound. This can make the evaluation more objective and effective. Another way to improve the questionnaire is to include more criteria, such as sound quality and emotion. Through improving the evaluation method, more data can be collected for further improvements.

In the future, more natural sounds can be synthesized using Nyquist, such as waves, leaves rustling, and even animal sounds. Some artificial sounds can also be synthesized, such as cars, walking, and musical instrument sounds. Furthermore, to make this technology easier to use, it can be combined with large language models. By converting text into parameters, this combination can provide an accessible and controllable way of generating different natural sounds. Due to its advantages, this technology has a broad prospect. Its randomness, high speed, and small file size make it suitable for tasks that require instant generation, such as accompaniment of music and drama, video game sound effects, AI music generation, dynamic soundscape generation, and the Metaverse. Since the mixed sound has a high realism, it is also suitable for music production and film soundtracks.

4. Conclusion

In this research, four natural sounds, including rain, wind, thunder, and their mixture, are synthesized using unit generators in Nyquist. The methods used for natural sound synthesis in this research have significant advantages in convenience, flexibility, randomness, high speed, and small file size. Mel spectrograms and a questionnaire are used to evaluate the synthesis result. The synthesis is successful because the sounds and their spectrograms are consistent with expectations. The questionnaire gives a middle-high score on the synthesized sounds, showing that these sounds are relatively real. To sum up, this research shows the possibility and potential of generating natural sounds using Nyquist. In the future, this technology can be applied to many different tasks, especially those that require instant generation. Overall, this is a new and potential technology that needs more exploration.

References

[1] Dunn D 1992 Pioneers of Electronic Art pp 3–4
[2] Dudley H 1940 The Vocoder—Electrical Re-creation of Speech, Journal of the Society of Motion Picture Engineers vol 34(3) pp 272–8
[3] Bilbao S 2009 Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics John Wiley and Sons
[4] Smith III JO 2010 Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects
[5] Van Den Oord A, Dieleman S, Zen H et al. 2016 Wavenet: A Generative Model for Raw Audio, arXiv preprint arXiv:160903499 12
[6] Holmes T 2012 Electronic and experimental music: technology music and culture Routledge
[7] Zhou Y, Wang Z, Fang C, Bui T and Berg T L 2018 In Visual to Sound: Generating Natural Sound for Videos in the Wild, Proceedings of the IEEE conference on computer vision and pattern recognition pp 3550–8
[8] Okamoto Y, Imoto K, Takamichi S, Yamanishi R, Fukumori T and Yamashita Y 2022 Onoma-to-wave: Environmental Sound Synthesis from Onomatopoeic Words, APSIPA Transactions on Signal and Information Processing vol 11 p 1
[9] Li G, Xu X, Dai L, Wu M and Yu K 2023 Diverse and Vivid Sound Generation from Text Descriptions, ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp 1–5
[10] Scipio A D 2002 The Synthesis of Environmental Sound Textures by Iterated Nonlinear Functions, and its Ecological Relevance to Perceptual Modeling, Journal of New Music Research vol 31(2) pp 109–17
[11] Stefanakis N, Abel M and Bergner A 2015 Sound Synthesis Based on Ordinary Differential Equations, Computer Music Journal vol 39(3) pp 46–58
[12] Wang K, Cheng H and Liu S 2017 Efficient Sound Synthesis for Natural Scenes, IEEE Virtual Reality (VR) pp 303–4