Journal of Namibian Studies, 33(2023): 3413–3431 ISSN: 2197-5523 (online)

The impact of AI on virtual sound filed shaping in the context of digital music

Liu Xi-ya[1]

Abstract

AI has been developed sharply in these years. Especially while the ChatGPT catch the eyes at the year of 2023 by Microsoft. As a musician, I start to concern whether the music AI can replace me in the future. In this case, this paper will explore the impact of the core technology of music mastering AI on the shaping of virtual sound filed in the music production process in the context of digital music. The case of Li's(2014) paper will be used to discuss how the core technology of AI can make more and deeper changes to the steps already implemented at that time to achieve more accuracy, as well as to analyze and explore the concepts mentioned in Owsinski's(2013) paper in conjunction with the research question - Can virtual sound filed constructed by musical mastering AI have the potential to replace human musicians? I will give my conclusion of what I think at last.

Keywords: Music Mastering AI, music production, virtual sound filed.

1. Introduction

1.1. Research Background

In the early 1970's they developed a 2-channel recording system and deployed digital audio transmission systems in their broadcast centers and operations rooms. The first all-digital recording was Ry Cooder's "Bop 'Til You Drop" in 1979, and in 1978 an urgent project led the British record company Decca Records to develop its own two-track digital recording equipment, and Decca released the first recordings in Europe in 1979.

With the help of Sony's Digital Audio Stationary Head and Mitsubishi's Pro-Digi technology, digital recording quickly became mainstream in the 1980s. 1982 saw the introduction of CDs by Sony and Philips, making digital audio popular with consumers.

The core technologies of AI: the first one is machine learning (machine learning and all its components are a subset of AI. (Alpaydin, 2020) Machine learning applies various algorithms to different types of learning methods and analytic techniques, allowing systems to learn

[1] International College, Krirk University, Thailand, xiyaliucia@163.com

from experience and continuously improve automatically, without explicit programming (Jordan, & Mitchell,2015); the second is natural language processing (NLP) (Manning & Schütze, 1999), with the help of NLP techniques, machines can recognize and understand written language and voice commands, including translating human language into language that algorithms can understand (Jurafsky & Martin ,2019); the third is computer vision, which helps computers to view and understand digital images and videos, rather than just recognizing or classifying them (Forsyth& Ponce, 2011); the last is robotics (Siciliano& Khatib, 2016). These core technologies intervene to easily test and detect the acoustic environment, resulting in a figurative sonic image, which is not only very beneficial to beginners in music recording for further analysis and learning, but also allows advanced researchers to cut out some of the more basic steps. For example, the editing of breathing breaths, fixing pitch, and other basic tasks. Although these criteria are very abstract concepts, the ear has little memory for sound, so the use of AI techniques to help us remember musical sensations is undoubtedly a great contribution to basic learners and advanced researchers.

1.2. Research Overview

Music production is a new view that emerged with the rapid development of computers music industry. Musicians use recording technology and computers do music production, converting electrical signals to digital signals and then using computer audio workstations to complete the work of composing, arranging, recording, mixing, and mastering.

1.2.1. The significance of build a virtual sound filed

Virtual sound field production is a significant development in audio technology, as it offers a new level of immersion and realism in audio experiences and has a wide range of potential applications in various fields. In the music industry we often record sounds as direct sounds which is more malleable in later stages, but sometimes to produce a particular sound, we also make sounds that have early reflections or reverberations, which sound more realistic.

In the music production process, musicians pay special attention to the design and adjustment of the virtual sound field, but because in the post-production process, each musician has a different understanding of the subjective thinking and musical aesthetics of the music, the mixing process is often described as a "secondary creation" process, which is also the process of reproducing the details of the real sound field, but This process is also often used to add a lot of exaggerated processing to make a certain part stand out more and make the whole song sound new.

Combined with AI, the process of music production does not reduce the quality of music because of the intervention of computer technology, but rather music recording and mixing requires more of the acoustic environment because the intervention of AI can achieve more of the new technology.

1.2.2. Important factors influencing the establishment of virtual sound field

Virtual sound stage building is the most soulful and central stage of the music production process, and the way a piece of music is built can greatly affect the way it is perceived, depending on the preference of musician, perceived quality and emotional perception. This is a highly complex, multi-dimensional process problem that requires the combination of many different complex sounds, and will also be done in many different ways. The processing and modification of each track is dependent on the mixing of all the other instruments, and each processing and modification needs to be presented in a different way. For example: some equalizer settings are applied in electric guitars, which may be very different from the equalizer settings in acoustic guitars. This process is in a highly complex and non-linear space, heavily dependent on human music perception, preference for music and human auditory and emotional response to music.

In the virtual sound field construction, the sound field is judged by five abstract criteria: one is the "sense of hierarchy", two is the "frequency band distribution", three is the "sense of space", four is the "sense of distance", five is "dynamic distribution". (Owsinski,2013)

The sense of hierarchy is the sound of various instruments is accurately positioned, do not interfere with each other, placed in front of the sound field of instruments and instruments placed behind the sound field has a greater degree of separation, reflecting a strong sense of definition. (Siltanen & Pulkki, 2010)

Frequency listening sense refers to the feeling of each frequency band by size by themselves. For example: high school and low frequency they themselves according to the mutual size ratio and reflect the tone brightness (Siltanen & Pulkki, 2009).

The sense of space refers to the distance, height and position of each instrument, whether into a three-dimensional sense, whether you can feel the size of the instrument in three-dimensional space and the size of the environment, etc (Begault & Wenzel, 2009).

The sense of distance includes horizontal distance and vertical distance. The horizontal distance refers to the restored sound field, the most two ends of the instrument, how far away in the middle. Therefore, in the recording pickup process, we often shake the signal of high frequency from left to right, and then monitor by playback to determine whether the lateral distance expressed by its signal is

consistent with the real sound field. Vertical distance refers to the distance between the front and the end of the instrument on the "stage" when restoring the sound field, also called depth (Tervo, Lokki, & Väänänen, 2009).

Dynamic balance physically reflects the law of change of the total amount of things, that is, the total amount of internal components has a relative balance between the relationship. In music, dynamic balance refers to the balance between the largest dynamic and the smallest dynamic instruments in the whole song, so that the overall sound field of the music can have a sense of hierarchy, distance, frequency band listening and spatial sense. In addition, it can make the whole music can have emotional high and low ups and downs (Kalluri, & Vipperla, 2010).

The process of music mixing entails a step of arranging a wide variety of sound sources in a stereo or mono or even surround track, which may come from different instruments, vocals or orchestral music. However, the techniques and the equipment are using in the different methods by synthesizers, sound processors and mixing desks to mix the song, to build a sound filed. Today, with the advent of technology, computers and mixing software are enough to complete the mixing operations properly that were previously performed. Music engineer is the person to mix the sounding instruments together to create the final sound. The main way they work is to add or subtract a certain amount of volume to each instrument at a certain time, to add or subtract dynamics to each instrument at a certain time, and to add width and depth to a certain extent, in order to enhance the spaciousness of the music.

1.3. Purpose of the study

Music engineers can use the technology to restore the original sound filed or rebuild a special sound filed by using technology. These technologies are made by algorithms. This paper will explore whether different algorithms can replace human musicians in the future by combining the process of establishing a virtual sound field, the role that music mastering AI can play and the application of different algorithms in music mastering AI.

2. Common Algorithmic Rules for AI Models in Music Mastering AI

Music Mastering AI is a technology that uses AI algorithms and techniques to analyze and enhance the sound quality of music recordings during the mastering process. Music Mastering AI can help automate the mastering process by analyzing the audio data of a recording and making adjustments to EQ, compression, and other

parameters to optimize the sound quality. The technology uses machine learning algorithms to learn from previous mastering sessions, which allows it to make more informed decisions and achieve more consistent results.

2.1. Acoustic Model

Acoustic models are used by music mastering AI to simulate the human ear's perception of sound in order to understand the human ear's perception of sound. Acoustic models are currently used in many applications. For examples: (1)Sound feature extraction: AI can learn to extract features of audio signals by analyzing large amounts of audio data, using acoustic models using large amounts of data audio training and optimization. AI can use techniques such as deep learning to train and optimize the acoustic model to improve the accuracy and generalization of the model. There is also a key task in acoustic models such as processing of audio signals. In addition, AI can also use techniques such as adaptive filtering to process the audio signal in real time to adapt to different environmental and scene requirements. Acoustic models are extracted from time-domain features, frequency-domain features and acoustic features (Tzanetakis & Cook, 2002). (2) Model training and optimization: acoustic models are models trained by AI algorithms, which require a large amount of audio data for training. AI can use techniques such as deep learning to train and optimize acoustic models to improve the accuracy and generalization ability of the models. In addition, AI can continuously improve the performance of the model by iteratively training the model. It makes it possible to do data pre-processing, model selection, parameter initialization, loss function, optimization algorithm and regularization, etc (Thakur & Dhiman, 2021). (3)Audio signal processing: AI can use techniques such as deep learning to de-noise, noise reduction, gain, etc., to improve the quality and audibility of audio signals. In addition, AI can also use techniques such as adaptive filtering to process audio signals in real time to adapt to different environments and scenarios needs, which include time and frequency analysis techniques, filtering algorithms, compression techniques, sound feature extraction, speech recognition and audio synthesis (Brownlee, 2019).

2.2. Dynamic range control

Dynamic range control is to make the loudness of the audio signal more balanced, while preventing the signal from being clipped or distorted and adjusting the dynamic range of the audio signal through the music mastering AI control dynamic range algorithm. Dynamic range control data pre-processing: Before the model training, the data needs to be pre-processed, including data cleaning, de-noising, normalization and other operations, in order to improve the training effect of the model. Dynamic range control has the following algorithms: compression algorithm (compression algorithm reduces

the dynamic range of the audio signal by reducing the variability of the audio, making it more consistent and balanced), extension algorithm (enhances the detail and variability of the audio by increasing the dynamic range of the audio signal), noise threshold algorithm (a special dynamic range control algorithm, mainly used to reduce the effect of noise and background noise) limiting algorithm (a more radical dynamic range control algorithm, which restricts the maximum amplitude of the audio signal by forcing it not to exceed a predetermined range), etc.(Hasan and Ali, 2016)

2.3. Equalizer

Music mastering AI can use equalizer algorithms to adjust the frequency response of the audio signal, such as enhancing low or high frequencies, to improve the clarity and good listening experience of the audio signal. Digital filter is a common equalizer algorithm that adjusts the frequency response of audio by the parameters of the filter. Digital filters are usually designed according to the characteristics and requirements of the audio signal, the appropriate filter type and parameters to achieve the best equalization effect. The types are parametric equalizer, digital filter, graphic equalizer, dynamic equalizer, linear phase equalizer, etc. Neural network is an AI algorithm widely used in audio processing, which can be trained to learn the characteristics and response laws of audio signals in order to achieve adaptive equalization. Neural network equalizers usually have high adaptivity and processing accuracy, but requires more training data and computational resources. Genetic algorithm is an optimization algorithm based on the principle of genetics, which can optimize the equalizer parameters by continuous evolution and screening. Genetic algorithm equalizers usually have high optimization accuracy and robustness, but require longer computation time and complex parameter settings (Abdallah and Plumbley , 2020).

2.4. Noise Suppression

Music mastering AI can use noise suppression algorithms to remove noise from audio signals, including background noise, electromagnetic interference, etc., to improve the quality of audio signals. Deep learning is a branch of AI that uses deep neural networks that can extract features in the audio signal, which in turn enables noise attenuation or removal. Noise suppression algorithms based on frequency domain processing can convert the audio signal to the frequency domain and suppress the noise at different frequencies. The noise suppression algorithm based on time domain processing is to directly process the audio signal in the time domain, using filters or other techniques to remove or attenuate the noise. Model-based noise suppression algorithms need to build a model to describe the relationship between the audio signal and the noise, and then use the model to remove or attenuate the noise.

2.5. Compression and limitation

Music mastering AI can use compression and limiter algorithms to adjust the dynamic range and loudness of the audio signal to improve the stability and overall volume of the audio signal. Support Vector Machines (SVM) is a common classification algorithm that achieves classification by dividing hyperplanes in different feature spaces. Hidden Markov Model (HMM) is a probabilistic model for modeling sequential data, which is widely used in speech recognition, handwriting recognition and music information retrieval. In compression and limiter, HMM can be used to build dynamic models of audio signals for more accurate identification and classification of audio signals. Neural Networks (NeNs) are computational models that mimic biological neural systems and enable classification and prediction of data by continuously optimizing weights and biases. In compression and limiter, neural networks can be used to model the audio signal and adjust the dynamic range of the audio signal based on the results of the model prediction. In compression and limiter, AI typically uses deep learning algorithms to train models and uses these models to predict the dynamic range of the audio signal in order to process the audio signal. AI can predict the dynamic range of an audio signal by analyzing and learning from the input signal, and compress or limit the signal based on the prediction. AI can adaptively adjust the compression/limiting parameters based on the dynamic range of the input signal to achieve more accurate audio processing. AI can intelligently adjust the threshold of the compressor/limiter to different types of audio signals by learning the characteristics of the input signal. AI can predict the dynamic range of an audio signal by analyzing and learning from the input signal, and compress or limit the signal based on the prediction. AI can adaptively adjust the compression/limitation parameters based on the dynamic range of the input signal to achieve more accurate audio processing. AI can intelligently adjust the threshold of the compressor/limiter to different types of audio signals by learning the characteristics of the input signal. (Ward & Staley, 2019)

3. The interplay between "technical" standards in human-shaping virtual soundscapes and music mastering AI

3.1. Human musicians are shaping the feel of the mixing music

The virtual sound field cannot be shaped without the sense of hierarchy, frequency listening, space, distance and dynamic distribution of changes, various ways and means, the virtual sound field can be changed from the instrument itself to have orchestration.

Different kinds of instruments have different frequency bands, and the frequency bands of musical instruments have fundamental frequencies and overtones, and different notes have corresponding pitches, so the distribution of fundamental frequencies among instruments and the integration of overtones are one of the means to shape a realistic virtual sound field. The technical tools available for mixing are: volume faders, equalizers, compressors and limiters, expanders and noise gates, reverberators, delayers, etc.

3.1.1. Layering

Layering is the process of positioning the sound of various instruments so that there is a large separation between each instrument in the entire virtual sound field. Imagine the sound field as a 3D space, and define each instrument in terms of "latitude and longitude in the virtual sound field". In a real stage, the instruments have different positions, the main instruments of pop music are probably guitar, bass, drums and vocals, the drums are usually located at the end of the stage, most cases will be at the end of the vocals, so in the processing of vocals and the drum set, the vocals will be more face-to-face than the drums, so the drums will be farther away in space. In a real sound filed, instruments that are farther away will have a longer time for the sound to travel through matter to reach the human ear, and so will be farther away.

There are four ways to change the sense of varying levels: The first is very common- lowering the volume or increasing it, this way will directly change the position of the instruments farther and closer. The second is to add a reverb delay. Adding some reverb or delay can be used to "trick" our brain. This is what we mentioned above, the source of sound multiple times through the material reflection of the sound, or by pushing out the sound and produce the sound effect. Of course, this is the case when there are substances that can reflect sound waves, but if you are in an environment where there are no reflective substances, it is difficult to hear more reflected sound. The third can be added by the sound phase to increase the sense of hierarchy. The last one is the grasp of the sound field from the recording. When recording, if the sound field itself can be recorded, two microphones instead of the human ear placed in the room, can effectively provide a real sound field environment, raise or lower the volume can change the spatial location of each instrument in the real sound field. This is of course one of the more common ways of recording drums, and is the simplest and most direct way to change the spatial location of the entire drum set. (Bittner and Mauch, 2020).

A series of changes are made in order to conform to the aesthetics of human cognitive hearing. As mentioned before, the purpose of establishing a virtual sound field is to restore the real sound field, that from the perspective of the logic of the algorithm rules of AI, AI will

restore the real sound field more easily and readily, theoretically speaking, data analysis of wave forms will be stronger than the ability of the human ear to analyze wave forms, and the ability of deep learning and imitation is indeed Faster than the human ear's ability.

3.1.2. Frequency distribution

The concept of frequency refers to the number of periodic changes completed per unit of time and is a quantity that describes the frequency of periodic motion. The concept of frequency is not only applied in mechanics and optics, but is also often used in quantum mechanics, electromagnetism and radio technology. Whereas sound travels as waves, when measuring the frequency of sound, electromagnetic waves (such as radio waves or light), telecommunication signals or other waves, it indicates the number of waveform repetitions per second. If the wave is a sound, frequency measures the characteristics of the note. Frequency is inversely proportional to wavelength. The frequency f is equal to the velocity v of the wave divided by the wavelength λ:

$$f=v/\lambda$$

The speed of an electromagnetic wave in a vacuum is the speed of light in a vacuum c. The equation becomes:

$$c=\lambda f$$

When a wave travels from one medium to another, the frequency remains the same, while the wavelength and phase velocity change.

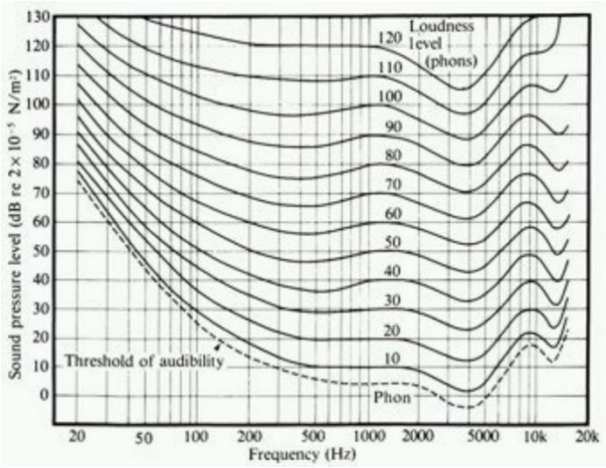
The frequencies are distributed in the interval according to the isophone curve of the human ear (Figure 1):

- Low frequency band (20-120 Hz): there is power bottom noise, bass instruments of the fundamental frequency part. The human ear cannot hear the part that can only be felt, the processing method is generally to do low cut.
- Low and middle frequency band (120-250 Hz): most of the fundamental frequency of the instrument in this band. From the band distribution, just the right amount will have the feeling of power and verve, or warmth, thickness and fullness. Too much will be muddy, too little will be hollow.
- Mid-range (250-2k Hz): This part is rich in overtones and easy to lose when processing, but this part can highlight the timbre as well as the characteristic parts of the instrument. There is a strong sense of presence, giving a powerful, clear, clean effect, too little will make the music dark and dull.
- The middle and high frequency band (2k-6k Hz): according to the human ear's response curve can be seen, this band is the most sensitive area of the human ear, this band more will give people more sense of veneer, the whole music can reflect the

sense of transparency, if too little, will make the whole sound field bright, too thin, too shallow.

- High frequency band (6k-20k Hz): This is the glossy part of the music, the right amount can have a glowing, relaxed feeling, but this part is also the most easily ignored part of the recording. This part contains what we call "airiness".

Figure 1: Equal Loudness Contour



3.1.3. Sense of space

The influence on the sound is related to two factors: the first is related to the real sound field in which the sound is located; the second is related to the position of the sound in the real sound field. Therefore, there are several factors that influence the multi-point recording in the sound field: 1) the location of the sound source in the sound field; 2) the geometric configuration and surface characteristics of the sound field (i.e., the acoustic characteristics of the sound field); 3) the distance between the sound source and the listener; 4) the position relationship between the sound source and the listener; 5) the listener's position in the sound field.

3.1.4. Sense of distance

Horizontal distance and depth distance often interact with each other. When the horizontal distance is constant and the depth distance is increased, the listener will feel that the size of the band is expanding, which is the technique we use to expand the virtual sound stage. However, the horizontal distance is more important than the vertical distance. Usually they can make the horizontal distance equal to the distance of the speakers. It can also be greater or less than the distance of the speakers. But listeners will prefer a treatment greater than the speaker distance. For example, in the case of quartet works, part of the treatment is equal to the size of the real quartet, but a significant part of the treatment is to make it larger than the size of the real

quartet, close to the feeling of a medium-sized orchestra, which will give the listener a sense of satisfaction in listening. In the case of popular music works, it is common to use easily perceptible instruments, such as pedal cymbals, sand hammers, and other soprano instruments, placed on the extreme left and right, so that the lateral width of the virtual sound field is clearly perceived and thus the sound field is significantly widened.

3.1.5. Changing the dynamic distribution

Dynamics are controlled by using compression, limiting and noise gates. A compressor is an automatic level control that requires the use of the input signal itself to determine the output level. This is set by using a threshold and ratio control. Compressors work on a gain ratio, which is measured by the input level versus the output level. Most compressors also have start and release parameters that control how quickly the compressor responds to the start and release of the signal. Many compressors have an automatic mode that automatically identifies the drink, analyzes the waveform, and later sets the attack and release based on the dynamics of the signal. While the automatic mode does work better, it still does not meet the requirements of some mixers for tonal precision on settings.

4. The shaping factors of personalized music

More than just being technically correct, the music must be as interesting as a drama. It must establish a sense of intonation and ebb, while having points of tension and release that engage the listener subconsciously. The first thing a musician must do before doing to build a virtual sound stage is to determine the direction of the song, so that the amount of the above five factors can be determined, and from there, what technical means to use to achieve what musical effect.

All good music, whether it's jazz, classical, pop or some new style we have not heard yet, is characterized by the pulse of the song and how the instrumentation and dynamics are at ease with the intonation, and the mixer's job is to make decisions and choices along the way, and then build the sound stage of the entire song around a center.

4.1. The impact of core technologies of AI on virtual sound field shaping

The human music engineer has developed certain patterns and techniques for shaping the virtual sound filed, and in recent years, the software remixing industry has been coming up with new ideas, both in terms of technical means and the development of Plug-ins to help the music engineer shape the virtual sound filed, even from the emerging electronic music, where the virtual sound filed is shaped

throughout the process of music production, and therefore the professional requirements for human music producers are getting The demand for professionalism from human music producers is therefore increasing.

Looking at the six perceptual shaping goals of AI's core technology above, is there a logical possibility of replacing the human mixer as a profession? What are the advantages and disadvantages of the replacement? What are the current technical difficulties of AI? The author will elaborate on these questions.

4.1.1. AI can produce the layering

The five ways to change the sense of hierarchy, in principle, there is a change in the "volume" of the measurement, for example: volume has to add or subtract dB, reverb has a specific value of seconds of Pre-delay, the sound phase from the left 100 to 0 to the right 100 interval, etc. How much to add and how much to subtract numerically, and how to reach a standard you want, requires a little experimentation, even for experienced mixers, who need to carefully debug each step, and also need to pose each instrument for the whole song. And for the human ear without sound memory, the only thing that can be familiar with the sound is the feeling of the sound, and want to have such a feeling again, need to be in the same acoustic environment, the same set of sound system, the same temperature and humidity in order to strictly standard recovery of the same sound feeling. And the machine learning part of AI, is it possible to directly copy a similar style of music feeling? Is it possible to automatically analyze the reference track and automatically adjust the parameters to achieve the musical feeling the mixer wants?

If we want to analyze the waveform parameters of a certain instrument, we have to strip it, and the stripped waveform is no longer sampled at 44k/s. The sine waveform itself is damaged, and the layering is not as complete as the full audio. Technically speaking, the change of parameters is simple, but this series of analysis and learning has been completely degraded, so it is impossible to completely "copy and paste". With this in mind, AI has a long way to go.

4.1.2. Frequency balance and dynamic distribution of AI

AI can be set up in the way humans require by recognizing signals in a way that the results are potentially more accurate and error-free than humans. Frequency is more digital analysis, from 20-20k Hz data analysis, from the waveform, the amount of each frequency to measure, to control the frequency distribution of each instrument to achieve a balanced state.

However, as mentioned before, the process of frequency balancing requires not only that the strengths of the instrument itself be revealed in the frequency distribution, but also that the stylistic

characteristics of the music be brought out in the process. What frequency balancing can do is to mechanically level out the frequency of the instrument, but it cannot bring out the characteristics of the instrument itself in the sound field that needs to be highlighted in the whole music. Therefore, the frequency balancing that AI can do is relatively mechanical, while the sixth point - the aesthetic part of the mix - is not achievable with current technology.

4.1.3. The spatial and distance sense of AI

Spatial awareness and distance are tied more to human perception. Spatial localization algorithms mean that AI can use localization algorithms to determine the location of each audio signal and adjust the balance of the mix based on their location and orientation. This can be achieved by various localization algorithms such as cross-ear method, wave field equation-based method, sonar signal-based method, etc. Reverberation algorithms are reverberation algorithms that can simulate a variety of different room sizes and shapes, thus adding a sense of space and distance to the audio signal. AI can select the appropriate reverberation algorithm according to the desired reverberation effect and adjust its parameters to achieve the desired effect. Spatial source separation algorithms: Spatial source separation algorithms can separate multiple sources from the mixed signal and add independent spatial information to each source. This can be achieved by using deep learning models or other machine learning algorithms. Stereo and surround encoding algorithms: Stereo and surround encoding algorithms can encode audio signals into stereo or multichannel surround signals for a more realistic audio experience. AI can select the appropriate encoding algorithm and adjust the parameters based on the characteristics of the audio signal and the target encoding format.

4.1.4. Personalization and Creativity in AI

The way AI is personalized and creatively shaped in music mixing is achieved through the training and optimization of algorithmic models. In mixing, AI can understand the semantics and emotions of music by learning the characteristics and structure of the audio signal, and apply this information to automatically generate new sounds and mixing effects. Specifically, AI can personalize and create music mixes in the following ways: Music sample generation: AI can personalize and create music by learning a large number of music samples to generate new music samples. Mixing effect generation: AI can personalize and create music by learning a large number of mixing effects and techniques to generate new mixing effects. Music emotion analysis: AI can analyze the emotion of music by learning the emotion information in the music signal to personalize and create music. Music Adaptive: AI can personalize and create music by learning the characteristics and structure of music signals to self-adapt to the rhythm, volume, and

tempo of music. Musical interactivity: AI can generate new musical effects and remixes by learning the feedback and interaction of users, thus personalizing and creating music.

5. The impact of core technologies of AI on virtual sound field shaping

AI can play a key role in virtual sound field shaping with core technologies such as speech recognition, audio processing, acoustic modeling, machine learning, and deep learning. These technologies can be used to analyze, synthesize, enhance, and virtualize the sound to achieve a spatial sense of sound shaping.

The "depth" of Deep Learning is a good solution to the problems faced by shallow learning: (1) DL's network contains multiple implicit layers and uses multi-stage transformations to describe data features in layers, representing low-level features as abstract high-level features. Therefore, it has better feature representation capability, powerful function fitting capability and generalization capability. Through the hierarchical structure, DL realizes the process of feature abstraction from "point" to "line" to "local" to "whole". (2) The initial parameters of the deep learning model are obtained through training with a large number of sample data, and the feature extraction of the model is done through autonomous learning, without relying on human experience. achieved through unsupervised learning, a characteristic that determines its suitability for processing natural signals and unlabeled data, which cannot be achieved by shallow algorithms. In addition, since DL adopts an underlying data-oriented mechanism, it maximizes the integrity of information and provides a means to effectively express features for underlying data whose features are complex and difficult to extract manually. Pre-training is the bottom-up unsupervised learning. Unsupervised training uses unlabeled data. Parameter tuning i.e. top-down supervised learning. The initial parameters are obtained based on the pre-training learning, and the model is trained using the labeled data, and the parameters of the entire network are "fine-tuned" so that the input and output errors are as small as possible. The "two-step" approach is in fact to group a large number of parameters, find the locally better settings first, and then combine the locally better results for global optimization.

This is possible with the logic of deep learning and top-down unsupervised learning techniques. In virtual sound field shaping, AI can learn sound features and spatial information from large amounts of audio data through techniques such as deep learning, and thus automatically generate spatially rich and creative sound effects. For example, deep learning algorithms can be used to model sound sources at different locations and applied to the simulation of virtual

sound fields to achieve more realistic and three-dimensional sound effects. In addition, AI can also identify and compensate for distortion, noise and time delay in sound by analyzing and processing audio signals in real time, thus improving the quality and clarity of sound and further enhancing the spatial and realistic sense of sound.

6. Conclusion

While AI technology has advanced significantly in recent years, it is not yet capable of fully replacing a human music engineer in the mastering process. Mastering involves a combination of technical expertise, creative decision-making, and an understanding of the music that only a skilled human can provide. However, there are AI tools and software available that can assist music engineers in the mastering process. For example, AI-powered plugins can analyze and optimize the mix for specific frequency ranges, or help to identify and reduce unwanted noise or distortion. These tools can help to streamline the mastering process and improve the overall sound quality of the final product.

AI-powered EQ algorithms can analyze the frequency content of the mix and adjust the levels of different frequency ranges to achieve a more balanced and cohesive sound. AI-powered compression algorithms can analyze the dynamic range of the mix and adjust the levels of the different elements to achieve a more consistent and polished sound. Limiting algorithms can prevent clipping and distortion in the final master by setting a maximum output level.AI-powered noise reduction algorithms can identify and reduce unwanted noise in the mix, such as hiss, hum, or background noise. AI algorithms can analyze the stereo image of the mix and enhance it by widening or narrowing the stereo field to create a more spacious and immersive sound. These algorithms can be used by music engineers to optimize the sound of a mix and create a polished, professional-sounding master. However, these algorithms are only assisting the music engineers. The reason cannot be fully replaced is music is a form of expression and creativity that reflects the personality and style of the artist, and a computer program cannot fully understand or replicate the human emotions and intentions behind the music yet.

Ultimately, while AI technology can be a valuable tool for music engineers, it is unlikely to fully replace the need for a human touch and creative input in the mastering process. However, if the emotion can be recognized, then these algorithms might replace or communicate with humans. At that time, AI is not an assistant, it will a dominate of a song.

Bibliography

[1] Abdallah, S. A., & Plumbley, M. D. (2020). Musical Audio Signal Processing with Deep Learning: A State-of-the-Art Review. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 796-819.

[2] Acoustic model optimization for speaker verification using deep neural networks. Applied Sciences, 9(20), 4413. Retrieved from https://doi.org/10.3390/app9204413

[3] Agrawal, A., Hadjeres, G., & Pachet, F. (2020). A Review of AI Techniques for Music Generation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 249-266. doi: 10.1109/TASLP.2019.2956257

[4] Alpaydin, E. (2020). Introduction to machine learning. MIT Press.

[5] Begault, D. R., & Wenzel, E. M. (2009). Perceptual evaluation of virtual auditory space performance. Journal of the Audio Engineering Society, 57(10), 813-834.

[6] Ben-Tal, O., & Grierson, M. (2015). Exploring Generative Music Techniques for Improved Creativity in Composing and Improvising. Organised Sound, 20(3), 286-295. doi: 10.1017/S1355771815000147

[7] Bittner, R., & Mauch, M. (2020). AI-Assisted Composition and Mixing of Electronic Dance Music. IEEE Multimedia, 27(3), 6-13. doi: 10.1109/MMUL.2020.2995017

[8] Bobby, O. (2017). The Mixing Engineer's Handbook. Media Group.

[9] Brownlee, J. (2019). How to Develop Voice-Enabled Apps and Devices: Alexa Skills Kit, Google Assistant, and Other Voice-Enabled Products. Machine Learning Mastery.

[10] Craig, J. J. (2005). Introduction to robotics: mechanics and control (3rd ed.). Pearson.

[11] Deng, L., & Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3-4), 197-387.

[12] Engel, J., Huang, A., Roberts, A., Donahue, C., & Eck, D. (2019). Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. Journal of Creative Music Systems, 3(1), 1-17.

[13] Fine, T. (2008). The dawn of commercial digital recording. ARSC Journal, 39(1), 1-17.

[14] Forsyth, D. A., & Ponce, J. (2011). Computer vision: A modern approach. Prentice Hall.

[15] Gao, H., Ouyang, M., Zhang, D., & Hong, B. (2011). An auditory brain-computer interface using virtual sound field. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2011, 2011, 1450-1453. doi: 10.1109/IEMBS.2011.6090445

[16] Gibson, D. (2019). The art of mixing: a visual guide to recording, engineering, and production. routledge.

[17] Giri, M., Patnaik, A., & Pandey, R. K. (2020). Time and frequency domain features based acoustic classification of birds using convolutional neural

network. Applied Acoustics, 173, 107636. Retrieved from https://doi.org/10.1016/j.apacoust.2020.107636

[18] Hamilton, R., O'Malley, K., & Campbell, N. (2017). Exploring Artificial Intelligence in Music Composition. ACM SIGCSE Bulletin, 49(2), 82-83. doi: 10.1145/3017680.3017723

[19] Hasan, S. M. R., & Ali, M. A. (2016). A Review on Dynamic Range Compression Techniques in Audio Signal Processing. Proceedings of the 3rd International Conference on Electrical Information and Communication Technology (EICT), 1-6. doi: 10.1109/EICT.2016.7807081

[20] He, Y., Liu, X., & Wu, Z. (2020). A survey on deep learning in audio signal processing. Journal of Signal Processing Systems, 92(5), 707-723. Retrieved from https://doi.org/10.1007/s11265-020-01506-8

[21] Huang, X., & Acero, A. (2001). Spoken Language Processing: A guide to theory, algorithm, and system development. Prentice Hall.

[22] Iwaya, Y., Otani, M., & Tsuchiya, T. (2016). Discrimination of virtual sound fields different in spatial aliasing. The Journal of the Acoustical Society of America, 140(4). doi: 10.1121/1.4964232

[23] Jeon, P., Park, J., Kim, J., & Nam, J. (2019). DeepAudioNet: An Efficient Deep Learning Model for Music Source Separation. IEEE Access, 7, 52383-52392. doi: 10.1109/ACCESS.2019.2914352

[24] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. doi: 10.1126/science.aaa8415

[25] Jurafsky, D., & Martin, J. H. (2008). Speech and Language Processing (2nd ed.). Prentice Hall. Moore, R. E. (2012). A Tutorial on Dynamic Range Control. IEEE Signal Processing Magazine, 29(5), 131-138. doi: 10.1109/MSP.2012.2204182

[26] Jurafsky, D., & Martin, J. H. (2019). Speech and language processing (3rd ed.). Pearson.

[27] Kalluri, S., & Vipperla, R. (2010). Personal audio for enhancing spatial intelligibility of speech in dynamic virtual environments. Journal of the Audio Engineering Society, 58(11), 907-921.

[28] Kim, H. J., Kim, Y. S., & Jang, H. J. (2017). The Design and Study of Virtual Sound Field in Music Production. Journal of the Korea Society of Computer and Information, 22(7).

[29] Lee, S.-W. (2019). Audio Signal Processing for Next-Generation Multimedia Communication Systems. John Wiley & Sons.

[30] 李凯(2012).音乐制作中的虚拟声场设计研究[D].中央音乐学院,北京.

[31] Li, X., Li, Y., Zhou, W., Li, J., & Li, S. (2019). Audio feature extraction based on deep learning: A review. IEEE/CAA Journal of Automatica Sinica, 6(3), 579-593. Retrieved from https://doi.org/10.1109/JAS.2019.1911698

[32] Li, Z., Li, F., & Qin, W. (2018). A new virtual sound field reconstruction algorithm based on the dynamic filter method. Applied Acoustics, 142, 38-47. doi: 10.1016/j.apacoust.2018.07.015

[33] Liu, X., Wang, C., Huang, Y., & Zhou, Y. (2021). A Survey of Deep Learning in Audio Processing. arXiv preprint arXiv:2106.10201.

[34] Liu, Y., & Zhang, X. (2019). Perceptual study on distance perception in virtual sound field reproduction. Applied Sciences, 9(19), 3988. doi: 10.3390/app9193988

[35] Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT Press.

[36] Melo, D. D., Grachten, M., & Lidy, T. (2018). Exploring Generative Music Variations Controlled by Timbre and Pitch via Deep Learning. Frontiers in Digital Humanities, 5, 2. doi: 10.3389/fdigh.2018.00002

[37] Nadig, S. S., Prasad, S., & Prasad, N. R. (2018). Dynamic Range Compression using Artificial Neural Network. Proceedings of the 2018 IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1386-1391. doi: 10.1109/ICACCI.2018.8554865.

[38] Owsinski, B. (2004). The recording engineer's handbook. Hal Leonard Corporation.

[39] Owsinski, B. (2014). The mixing engineer's handbook (p. 312). Course Technology, Cengage Learning.

[40] Rao, K., & Kumar, S. (2020). A review of adaptive filters for audio signal processing. Digital Signal Processing, 103, 102783. Retrieved from https://doi.org/10.1016/j.dsp.2020.102783

[41] 荣隽(2004).计算机音乐音频制作中的虚拟声场效果研究[D].南京艺术学院,南京

[42] Rosenzweig, S., Farbood, M., & Rafii, Z. (2020). Deep Learning-Based Music Generation: A Survey. ACM Computing Surveys, 53(3), 1-37. doi: 10.1145/3390983

[43] Siciliano, B., & Khatib, O. (Eds.). (2016). Springer handbook of robotics (2nd ed.). Springer.

[44] Siltanen, S. A., & Pulkki, V. (2009). Design of a virtual sound field reproduction system for rectangular loudspeaker array. IEEE Transactions on Audio, Speech, and Language Processing, 17(6), 1168-1180. doi: 10.1109/TASL.2009.2018697

[45] Siltanen, S. A., & Pulkki, V. (2010). Perceptual evaluation of a virtual sound field reproduction system. Journal of the Audio Engineering Society, 58(10), 836-848. doi: 10.17743/jaes.2010.0062

[46] Stovold, P. (2020). Intelligent Music Production: A Survey of Recent Developments. IEEE Transactions on Emerging Topics in Computational Intelligence, 4(4), 441-455.

[47] Sturm, B. L. T., Ben-Tal, O., Ramirez, M. A., & Bown, O. (2018). Towards Creative AI: Experiments in Generative Music. Journal of Creative Music Systems, 2(1), 1-19. doi: 10.5920/jcms.2018.01

[48] Szeliski, R. (2010). Computer vision: Algorithms and applications. Springer.

[49] Tervo, S., Lokki, T., & Väänänen, R. (2009). Perceived distance in virtual sound environments. Acta Acustica united with Acustica, 95(6), 980-991.

[50] Thakur, M. K., & Dhiman, A. (2021). Acoustic model training for speech recognition: A review of recent advances. IEEE Access, 9, 43170-43189. Retrieved from https://doi.org/10.1109/ACCESS.2021.3065534

[51] Tylka, J. G., & Choueiri, E. Y. (2019). Domains of practical applicability for parametric interpolation methods for virtual sound field navigation. Journal of the Audio Engineering Society, 67(11), 882-893.

[52] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293-302. Retrieved from https://doi.org/10.1109/TSA.2002.800560

[53] Wang, H., & Chew, E. (2018). Automatic Music Generation for Expressive Performance Using a Hybrid Recurrent Model. ACM Transactions on Intelligent Systems and Technology, 9(4), 1-22. doi: 10.1145/3192950

[54] Wang, H., & Chew, E. (2020). Creating Expressive Music Performances with a Hybrid Generative Model. IEEE Transactions on Affective Computing, 11(3), 436-448. doi: 10.1109/TAFFC.2018.2809426

[55] Ward, D., & Staley, T. (2019). Music and audio engineering: theoretical and practical approaches. Taylor & Francis.

[56] Xie, B., Xie, Y., Yang, F., & Huang, L. (2018). A virtual sound field synthesis method for robustly generating spacious and accurate sound field. Applied Acoustics, 139, 25-37. doi: 10.1016/j.apacoust.2018.03.014

[57] Xie, B., Xie, Y., Yang, F., Wang, M., & Huang, L. (2019). An improved virtual sound field synthesis method based on frequency division multiplexing. Applied Acoustics, 145, 267-277. doi: 10.1016/j.apacoust.2018.10.019

[58] Yao, L., & Wang, D. (2019). A survey on acoustic feature extraction for speech and emotion recognition. IEEE Access, 7, 142378-142401. Retrieved from https://doi.org/10.1109/ACCESS.2019.2944928

[59] Zhang, L., Feng, Y., Liu, J., & Xu, X. (2012). A Novel Method of Dynamic Range Control for Audio Signals Based on a Nonlinear Mapping. IEEE Transactions on Audio, Speech, and Language Processing, 20(1), 344-355. doi: 10.1109/TASL.2011.2159683

[60] Zhang, R., Yang, Y., & Xu, C. (2021). A Survey of Artificial Intelligence in Music: Trends and Challenges. IEEE Access, 9, 77871-77888.

[61] Zhang, X., & Liu, Y. (2019). Perceptual study on vertical localization in virtual sound field reproduction. Journal of the Audio Engineering Society, 67(3), 122-130. doi: 10.17743/jaes.2019.0007

[62] Zhang, Y., & Zhang, H. (2019). Multi-Objective Evolutionary Algorithm for Music Generation with Multiple Influences. IEEE Access, 7, 35040-35050. doi: 10.1109/access.2019.2905722