International Journal of Electronic Commerce Studies
Vol.13, No.1, pp.001-032, 2022
doi: 10.7903/ijecs.1933

A DEEP LEARNING BASED INNOVATIVE ONLINE MUSIC PRODUCTION FRAMEWORK

Sung-Shun Weng
National Taipei University of Technology
wengss@ntut.edu.tw
Hung-Chia Chen
National Taipei University of Technology
honga.chen@gmail.com

ABSTRACT

In this research, we constructed an online music production process framework to explain the operation of the industry and conduct comparative analysis, competitive analysis, and innovative business model analysis for the current music industry situation. This research can provide strategic suggestions for future scholars and industry development. In this study, through the concept of the internet, we propose the framework process of a deep learning music production (DLMP) system, and each work process and the main work content of each module in the system are described. We used upstream, midstream, and downstream industry roles for comparison and used an innovative operating model to describe the overall industry transformation. Through this research, we can more clearly see a blueprint for the future development of music production and can provide operators with a competitive advantage strategy.

Keywords: deep learning, music production, innovative operation model

1. INTRODUCTION

The music production industry has undergone several technological developments, and the industrial structure has undergone tremendous changes. How to use technology and the development of policies are important issues for the sustainable development of the music production industry [1]. From the rise of MP3 technology in the mid-1990s to the rise of cloud technology and the internet, music production has moved from the age of analog production to the age of digitalization and cloud collaboration.

The music production industry compared with the music industry has different needs and market perspectives. The music industry attaches importance to the overall music content sales and packaging and channel strategies, while the music production industry attaches importance to the creation of music content, the presentation of music art, and the use of production technologies. In the past, there were few analyses on the music production industry that proposed an industry innovation framework. This research analyzed and integrated key data processing technologies currently in the music production industry, and we included the most popular deep learning algorithm technology in recent years.

Through the use of deep learning technology to build an online music production platform framework, we combined new technologies with a digital music production framework on the internet to evaluate the relationship between the value chain and the innovation model to provide a reference for industrial cooperation development.

In past research, sound processing technology was mostly a single technology discussion, and few works described the integration of the technologies that can be used in the cloud in the process of music production. In this study, we found a new industrial operation framework through the perspective of a cloud network. This is also the direction of industry transformation that the music production industry will face in the future.

This research first discusses the internet and cloud collaborative work literature, and we found that the process and characteristics of the music production industry can transfer the operation mode to the cloud. Next, we propose the key processing technologies needed to transfer the music production process to the cloud, including the data format technology of MIDI and MusicXML, the technology of converting AUDIO into the MIDI file type, the noise processing technology applied in the context of mobile recording, and the deep learning technology used in music analysis and recommended applications. Finally, through the integration of these key technologies, we propose a new cloud music production industry work structure.

In addition, there have been many issues in the way of sharing revenue in the music production industry. Among the variety of roles in the music production industry, only a few have been allocated reasonable royalties; however, the music production industry is a combination of many different roles. How to share revenue more fairly to everyone involved is a very important issue in the construction of a cloud digital music production platform in the future. The framework of this study also provides a reference basis for the future development of sharing revenue technology.

2. LITERATURE REVIEW

2.1 Internet

Today, internet networks are a common information service platform. Internet networks were first developed in the 1980s. Now, an enormous international network is connected via various networks and has become a global internet network. The internet has been the incubator for new network technologies, far surpassing the expectations of worldwide network suppliers. Two basic reasons underlie the internet's success: the internet satisfies the needs of its users, and its technologies were developed by solving real problems [2]. The emergence of the internet has also had a huge impact on the relationship of the music industry supply chain [3, 4].

Through internet networks, the manufacturing industry can elevate its efficiency, quality, innovation, cooperation, marketing strength, and information flow, and can increase cash flow. Internet networks create more business opportunities (a minor effect), are able to change the relations of industrial value chain and industrial ecology, and are relevant to international technology development. Although most parts of digital music production have been digitalized, the industry is still at the beginning stage of exploring how to apply this to the internet. Digital music production is similar to other traditional industries that have an innovative development future. In this study, we researched and analyzed how the internet network affects the integration and innovation of digital music production.

2.2 Music Data Processing Technologies

2.2.1 MIDI and MusicXML

The MIDI format is the technology most often used in computational music [5, 6]. In 1981, Dave Smith, an engineer, first introduced the standard of MIDI (musical instrument digital interface). MIDI is a digital representation of music, used especially for communication between electronic musical instruments, including computers. The MIDI message associates an integer number to an individual note (e.g., central C is 60), so it is suitable for a wide range of studies [5]. The channel message includes orders that are transited via any of the channels on the specific MIDI instrument (electronic instruments or electric instruments), which include the note's characteristics, expression characteristics, and performance modes. The system message orders are transited via all channels on the MIDI instrument [7] to control the MIDI messages on all instruments, which includes the system reset and requirements of tuning. There are Exclusive Messages, Common Messages, and Real-time Messages [8].

A newer technology, MusicXML, offers the same possibilities and more, but requires different techniques for data retrieval. MusicXML is a royalty-free format that implements all the features offered by XML technology: data structure, modularity, extensibility, and the possibility of querying and interaction through XML family technologies for conducting automatic analysis of music through adopting a statistical approach [5]. MusicXML attempts to provide a common document type definition that is well designed from the musical, human, and computer perspectives [9]. By MusicXML, the interactive score is easily published on the internet, and it is convenient for musicians who use different music software to exchange their information and collaborate [6].

A large number of non-expressive data files in formats like MIDI and MusicXML [9] are available on the internet, and they are used by many musicians as a standard communication tool for ideas and pieces [10]. Knowledge discovery from data (KDD) includes operations designed to obtain information from complicated data sets [11]. For digital music, the data may comprise all the recording files and individual notes of MIDI records (scales, strength, length, and channels), which truly record the bars, instruments, and notes of the music. These can be used as source for the training data and prediction and recommendation system data.

2.2.2 AUDIO to MIDI

During digital music production, direct audio recording is the majority except for performance inputs via a MIDI controller. This type of non-structured multimedia audio file will not be able to be value added until the data is analyzed and converted. Every original sound of the music is a period signal that is composed of a single or multiple reference frequency sinusoidal waves. Each sound is an individual sinusoidal wave, and the function is

$$s(t) = A\, sin(2\pi f t).$$

A indicates the sound intensity, which is a scale of waveform vibration. For example, for stressed (strong) beats, the magnitude is large, and for unstressed (weak) beats the magnitude is small in music bars. f is for identifying notes, called the pitch or frequency, which indicates the times of the wave repeat and the repeat cycle. t is the time duration of a note according to the score requirement, which refers to the speed and beats:

$$T = \frac{1}{f} \;,\;\; s(t + T) = s(t).$$

When audio is converted to a MIDI signal, notes, pitch, beats, speed, dynamics of the audio are converted and analyzed. The beats and speed are set at the beginning of the music production to obtain the digital information easily. The dynamics can be allocated to MIDI by the equality of the ratios according to the volume analysis.

(1) Note or Solmization. One time frequency between two notes is an octave. The seven keynotes of C major (C-D-E-F-G-A-B) are Do, Re, Mi, Fa, So, La, and Ti (numbers 1 to 7 represent the notes of numbered musical notation). A standard tuning pitch 440 Hertz is A4, and A5 indicates an octave higher. There is the possibility of infinite extents higher and lower. Digital marks of an octave begin from the note C and finish at the note B (C-D-E-F-G-A-B). For example: D4 is D upward C4, and B3 is B downward C4.

(2) Pitch (frequency of notes): Music is composed of all kinds of frequencies of notes. In physics, vibration generates sound, which is presented as the symbol f or ν, and Hertz (Hz) is the unit of frequency in the international system of units. In western music, an octave (ration 2:1) is divided into twelve particular frequencies of notes also known as twelve-tone equal temperaments (12 equal temperaments). There is a mathematical relationship among the steady frequency and frequency of repetition. For example,

$$f = \frac{71(times\ of\ repetition)}{15(time\ duration)} = 4.7Hz.$$

A4 is the basic note, and its standard pitch is 440 Hz. The numerical value of each note corresponds to the note symbol on MIDI after computing, and all of the notes can be represented as integer multiples of the central A(A4). The distance is marked as "n". A plus "n" indicates a note that is higher than A4, but, on the contrary, a minus "n" indicates a note that is lower than A4.

$$f = 2^{\frac{n}{12}} \times 440\ Hz.$$

For example, the note "C5" is three and a half notes (A4 → A♯4 → B4 → C5) away from A4, close to A4 and higher than C of A4, and so its value of n is +3. The frequency of the note is

$$f = 2^{\frac{3}{12}} \times 440\ Hz \approx\ 523.25\ Hz.$$

According to the formula, if two notes differ from an octave or two octaves, the offset-frequency is an integer multiple as well. Under the twelve-tone equal temperament, "n" is absolutely 12-multiple (±12κ，κ means a total of octave) and the formula second-order equation difference can be simplified to

$$f = 2^{\frac{\pm 12\kappa}{12}} \times 440\ Hz = 2^{\pm\kappa} \times 440\ Hz.$$

The following Table 1 shows the MIDI standard corresponds with the frequency and solmization:

Table 1: MIDI standard corresponds with the frequency and solmization.

Solmization	Frequency/Hz	Solmization	Frequency/Hz	Solmization	Frequency/Hz
C4	261.63	C5	523.25	C6	1046.5
D4	293.67	D5	587.33	D6	1174.66
E4	329.63	E5	659.25	E6	1318.51
F4	349.23	F5	698.46	F6	1396.92
G4	391.99	G5	783.99	G6	1567.98
A4	440	A5	880	A6	1760
B4	493.88	B5	987.76	B6	1975.52

Converted relation between MIDI note number ρ and f (frequency):

$$\rho = 69 + 12 \times log_2\left(\frac{f}{440}\right).$$

C4 is defined as MIDI note number 60, and to increase/decrease a number follows by a semitone upward/downward [12].

(3) Chord: multiple notes are played at the same time during music recording. Each instrument plays different sounds and also influences the strength of the harmonic. If the comparison value of the frequency of two notes is very close to ratios of small whole numbers, these two notes play music harmoniously together. For example, if the ratio of frequency of the two notes is 2:3 (440 Hz and 660 Hz respectively), this is called a perfect fifth. Because their harmonics are overlapping, the sound of the two notes is harmonious, which indicates that the fundamental note should be the small whole numbers in mathematics language. Once the relation with ratios of every scale is comprehended, scale combinations of the sound will be decomposed to a single note that can be analyzed by the frequency ratio chart [12] (Table 2).

Table 2: Frequency ratio chart.

Scale Difference(n)	name	Approximate frequency ratio
0	Perfect unison	1:1
1	Minor second	16:15
2	Major second	9:8
3	Minor third	6:5
≀	≀	≀
10	Minor seventh	16:9
11	Major seventh	15:8
12	Perfect octave	2:1

Sound is a period vibration, and the waveform determines the timbre. To assume the waveform of a note on the piano is a sin wave. There are a total of 88 keys, and the frequency of an individual note composes a geometric sequence according to an equal temperament (with 12 equal temperaments). The scale coefficient is [12-14]

$$2^{\frac{1}{12}} = 1.059463.$$

We can obtain a geometric relationship with each single note (Table 3).

Table 3: Geometric relationships with each single note.

Scale Difference(n)	$2^{\frac{1}{12}}$	Approximate Value of Rational Number	White Key	Solmization
0	1		1	Do
1	1.05946			Do#
2	1.12246		2	Re
3	1.18920			Re#
4	1.25992		3	Mi
5	1.33484	1.333333	4	Fa
6	1.41421			Fa#
7	1.49831	1.5	5	So
8	1.5874			So#
9	1.68179		6	La
10	1.78180			La#
11	1.88775		7	Ti

Multi-notes of the chord compose the waveform:

$$f(x) = sin(x) + sin(a^n x) + \cdots.$$

For example, the major third chord of C major is Do+Mi+So (1+3+5), and its waveform is as follows:

$$a = 1.05946,$$
$$f(x) = sin(x) + sin(a^4 x) + sin(a^7 x).$$

2.2.3 Noise Processing

The music recording process uses mobile applications in the cloud and the recording environment has to meet mobile characteristics; therefore, external noise problems should be considered to ensure a better quality recording file. The categories of sound enhancement are based on the number of channels (single channel, stereo channel, or multi-track channel) for recording. Although the multi-channel speech enhancement is better than that of single channel speech enhancement [15-17], the single channel speech enhancement is still a significant field of research due to its simple implementation and ease of computation. Single channel speech enhancement uses only one microphone to collect noisy speech data [15-19].

Frequency spectrum weighting is an efficient, fast, and simple enhancement technology for processing the single channel speech frequency [18, 20, 21]. Subtractive-type noisy speech enhancement algorithms are also used, in addition to the basic spectral subtraction algorithm [15]. The other most notable algorithms are spectral over-subtraction (SOS) [20], parametric spectral subtraction (PSS) [21], spectral subtraction based on cross correlation [22], non-linear spectral subtraction (NSS) [23], multi-band spectral subtraction (MBSS) [24], Wiener filtering (WF) [25], iterative spectral subtraction (ISS) [26], extended spectral subtraction (ESS) [27], and spectral subtraction based on perceptual properties (SSPP) [28].

The spectral subtraction is one of the most well-known and computationally efficient methods for effectively suppressing the background noise from noisy speech as it involves a single forward and inverse transform. The first comprehensive spectral subtraction method, proposed by Boll [29-31] is based on the non-parametric approach, which simply requires an estimate of the noise spectrum and can be used for both speech enhancement and speech recognition.

Suppose the noise interference input signal is $y\ (n)$, clean speech signal is $s\ (n)$, and noise signal is $\omega\ (n)$, then the signal formula is shown as the following [32]:

$$y\,(n) = s(n) + \omega\ (n)\,. \qquad (1)$$

Hereby, the original clean speech signal is considered as the input signal is destroyed to reduce noise:

$$|s(t)|^2 = |y(t)|^2 - |\omega(t)|^2. \qquad (2)$$

For reducing the speech signal, Boll [31] modified the formula of the basic frequency spectrum subtractive method:

$$|\hat{s}(t)|^2 = |y(t)|^2 - \lambda|\widehat{\omega}(t)|^2. \qquad (3)$$

λ is computed as follows, where SNR (Signal-to-noise ratio) indicates the decibels (dB):

$$\lambda = \lambda_0 - \frac{3}{20}\mathrm{SNR} \quad -5\ \mathrm{dB} \le \mathrm{SNR} \le 20\ \mathrm{dB}. \qquad (4)$$

In reality, the noise frequencies in a noisy environment are irregular. For inspecting the influence on the noise of a speech signal, Berouti proposed a method of changing a speech signal to multi-band [20]. The identity is changed as follows:

$$|\hat{s}_i(t)|^2 = |y_i(t)|^2 - \lambda_i|\widehat{\omega}_i(t)|^2. \qquad (5)$$

λ_i is computed as follows:

$$\lambda_i = \begin{cases} 4.75 & \mathrm{SNR_i} \le -5\mathrm{dB}, \\ 4 - \frac{3}{20}(\mathrm{SNR}) & -5\mathrm{dB} \le \mathrm{SNR_i} \le 20\mathrm{dB}, \\ 1 & \mathrm{SNR_i} \ge 20\mathrm{dB}. \end{cases} \qquad (6)$$

The identities below are nonlinear frequency spectrum subtractions for reducing the high SNR-minus instantaneous power frequency spectrum and raising the low SNR-minus instantaneous power frequency spectrum. $|s_i(t)|$ indicates the enhanced speech signal, $|y_i(t)|$ indicates the noise speech signal, and $l_i(t)$ indicates the estimation of the decision of the noise signal:

$$|s_i(t)| = |l_i(t)| \times |y_i(t)|,$$
$$l_i(t) = \frac{\omega_i(t)}{|\hat{y}_i(t)|}\,, \qquad (7)$$
$$|\omega_i(t)| = |\hat{y}_i(t)| - |\widehat{\omega}_i(t)|.$$

To eliminate musical noise, Berouti also proposed a spectral subtraction with the over-subtraction method, asserting that the subtraction quantity of noise spectral amplitude should be inversely related to SNR. Audio with louder speech should be subtracted from softer audio:

$$|\hat{s}_i(t)|^2 = \begin{cases} |\hat{s}_i(t)|^2 \\ = |y_i(t)|^2 - \alpha(\text{SNR})|\omega(t)|^2, |\hat{s}_i(t)|^2 > \beta|\omega(t)|^2, \\ \beta|\omega(t)|^2, \qquad \qquad otherwise, \end{cases} \quad (8)$$

where $0 < \beta \ll 1$ is a constant and $\alpha(\text{SNR}) \geq 1$ can be calculated as follows:

$$\alpha(\text{SNR}) = \alpha_0 + \text{SNR}\frac{(1-\alpha_0)}{\text{SNR}_i}, \quad (9)$$

where $\alpha_0 \geq 1$ and $\text{SNR}_i \geq \text{SNR}$; therefore, when the speech signal is weak (i.e., low SNR), the $\alpha(\text{SNR})$ increases. The amplitude of the noise spectrum is over-subtracted, and the musical noise is eliminated by using $\beta|\omega(t)|^2$ in place of the subtracted result.

2.2.4 Deep Learning in the field of music production

Deep learning, a hot topic in artificial intelligence [33, 34], has been applied in many projects, including automatic processing and the recognition of texts, languages, voices, images, and so on. In 2004, Geoffrey Hinton, a professor in the field of artificial neural networks (ANNs) at the University of Toronto, developed the term, "deep learning" for artificial neural networks. Through the simulation of biological multi-layer neural networks, scholars have developed many layers, architectures, and initializations of artificial neural networks, such as convolutional neural networks (CNNs) [35], recurrent neural networks (RNNs) [36, 37], and long short-term memory (LSTM) [38] over the past 30 years. In this study, we concentrated on deep learning in the field of music production.

In 1998, Hörnel, D., and Menzel, W. explained how the hybrid music-harmonization system was calculated using the feedforward network algorithm HARMONT [39] in the Multiscale Neural Network Model. The system analyzed Bach's music through music symbols and neural-like algorithms, learned features in complex musical styles (including the goals to produce choruses based on musical rules), captured the features of music, and finally automatically produced four choruses and melodies of a similar style. In addition, MELONET, the multi-scale neural model system [40], could learn J. Pachelbel’s harmony-based melodic variation by supervised and unsupervised training to identify musical structures and predict melodies [41].

Later, in 2002, Douglas Eck and JÄurgen Schmidhuber successfully completed a composition of blues music through a LSTM (long short-term memory) algorithm. The project studied the staves and chords of blues songs (see Figure 1) in two parts. The first part of the study confirmed that LSTM, without relying on melody, could easily learn the structure of the chords and produce a new piece of music; the second part confirmed that LSTM could learn the structure of the chords and melody and, then, used this structure to generate a new song. The study proved Mozer [42], who disagreed on the criticism that that RNN was too general to catch the specificity of the style [43].

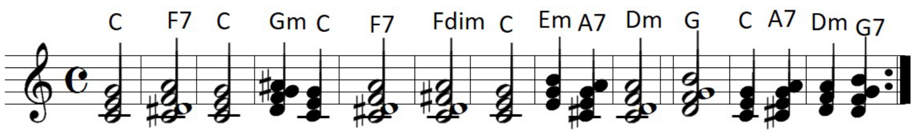

Figure 1: Chords for training data.

In 2012, Boulanger-Lewandowski, N. et al. used Piano Roll (Figure 2) and an RNN chord music symbol sequence of high-dimensional time distribution to analyze a classical piano MIDI file collection, a 1200 folk song collection in ABC chord format, CCARH (Center for Computer Assisted Research in the Humanities) formatted orchestral piano classical music in the MuseData library, and 382 pieces of four chords from the JSB (Johann Sebastian Bach) choir. In addition, they converted the 88 keys of piano to 88 binary units from A0 to C8, used quarter notes as quantized units of notes in the bar, and applied this to chord music generation and music editing. To conclude the project of Boulanger-Lewandowski, N. et al., the recurrent temporal RBM (recurrent temporal restricted Boltzmann machine, RTRBM) [44] and RNN-RBM models performed well for the learning and accuracy analysis of music [45].

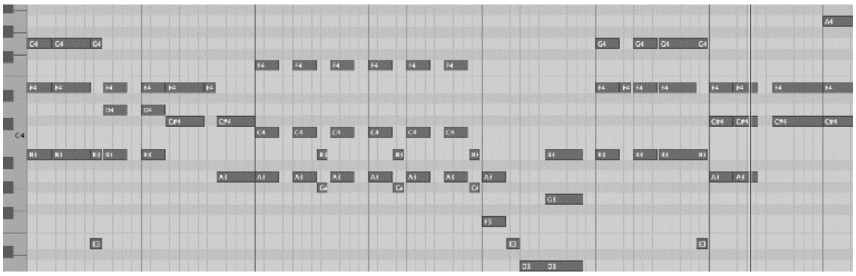

Figure 2: Piano Roll.

In 2016, Gaetan Hadjeres and Francois Pachet proposed the DeepBach model to mimic the four-part chorale by Johann Sebastian Bach. DeepBach also provided an interactive music creation application allowing non-specialists to enhance the notes and tempo or terminate and redefine the chorale and melody. The project defined a chorale as a couple (V, M) composed of voices and metadata, including V1–V4 for soprano, alto, tenor, and bass; and M as a collection of metadata lists (F and S): S for the subdivision of indexes of the beat and F for a fermata symbol. As Bach chorales contained only simple time signatures, the project discretized time with sixteenth notes, which meant that each beat was subdivided into four equal parts. The numbers on the scale were the 88 keys of piano. All variables are displayed in Figure 3 [45].

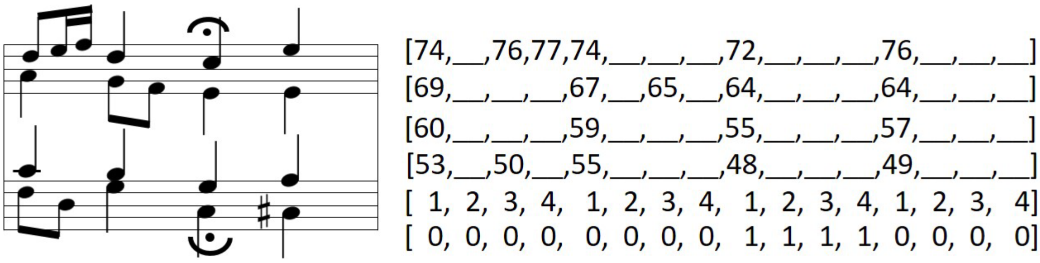

Figure 3: Sample of an extract from a Bach chorale.

The experiment was claimed successful as music experts voted for the system to produce compelling and coordinated music of Bach style, although the authors did mention the existing problems of plagiarism.

In 2017, when Iman Malik built his musical models, he explained that music, the sequences of notes, contained many dependencies. For example, in order to predict how a specific note changed at a given time, it depended on the note before and after it. Iman Malik proposed bi-directional LSTMs with memory gating developed through the concept of bi-directional RNN [46]. For inputs, the project used a binary vector, “note on” was encoded as [1, 1], “note sustained” as [0, 1], and “note off” as [0, 0]. Next, the note pitch was encoded as a number ranging from 0–128 in a matrix, representing the MIDI pitch number, as the first dimension. For the outputs, the columns represented pitch with only 88 notes, and the rows represented the timestep. The note velocity was encoded into its corresponding [pitch, timestep] index. The input and output representation are shown in Figure 4:

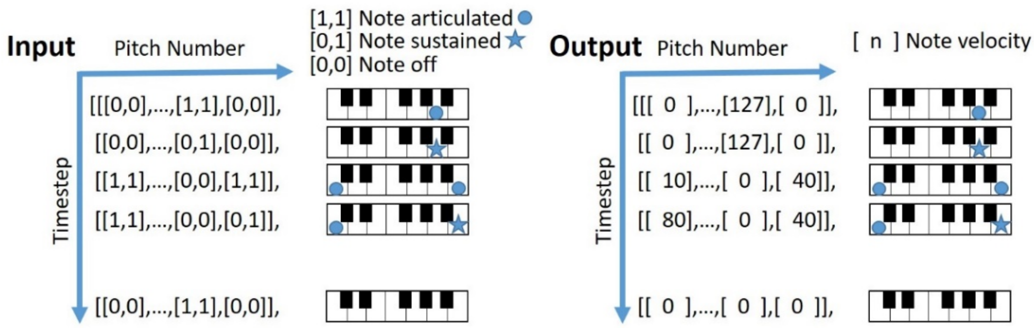

Figure 4: The input and output representation.

Iman Malik finally confirmed that his musical model, StyleNet, could produce human-like performances through the evaluation of two Turing tests [47].

At present, Google Magenta provides a computer program that allows users to create melodies with TensorFlow and train the program models to generate music via the Melody RNN model, and to operate the user-friendly MIDI interface to play with it interactively [48]. In addition, in 2017, Google also introduced NSynth, a neural audio synthesizer system using WaveNet Autoencoders, to synthesize music as well as inspire and aid the creative process [49].

With the research efforts of many scholars over the years on deep learning, a large scale of music can be analyzed for its distribution of notes in (including staves, Piano Roll, the digital matrix, and MIDI number), dynamics, and time series. Deep learning also recommends how the music should be played and, therefore, creates a musical piece with excellent quality.

Breakthroughs in deep learning have been observed in music creation and production and deep learning will play a decisive role in music production in the near future. Deep learning is recognized as an innovative field that could help the music production industry change and evolve in the future[50, 51].

2.3 Media Streaming

Streaming is a technology of network data transmission [52] that uses compressed video files and sends them to the client buffer on the computer, and controls the way the video is played through time stamps [53-55].

The technology of streaming is divided into three categories: HTTP streaming [56, 57], RTSP (Real Time Streaming Protocol) streaming [58], and Clientless streaming by a transport protocol [53]. Online real-time and on-demand playback are the two main modes. Video is typically transmitted online via RTP (Real-time Transport Protocol) streaming (True Streaming's RTP transmission protocol). RTP uses a UDP (User Datagram Protocol) transmission protocol to avoid sound or video delay [54, 59, 60], to modify the transmission speed through an RTCP (real-time transmission control protocol), and even change the load type. RTP provides traffic and congestion control services [60, 61]. In addition, OTA (over-the-air download) is used to deliver the Java player to the client, so that streaming media on mobile devices can improve their efficiency [53, 55].

Online playback technology improves the usability of music. However, in the actual data transmission process, problems may occur due to changes in the internet environment. First, the file may be damaged due to transmission errors, so an error control mechanism must be added to compensate for damaged data. Secondly, bandwidth changes require scalable streaming files. This allows clients to access media data of any size or quality under any bandwidth and device.

Another streaming technology is the protocol for music output and input, ASIO (Audio Stream Input and Output)[62], which is one of the audio API (Application Programming Interface) standards defined by Steinberg (German company). When using a microphone or a digital sound library for editing and performance through a MIDI controller, the delay of audio monitoring must be reduced during tracking, recording, and playback. If the sound is delayed, the musician will not be able to perform properly or control his emotions during the recording.

The most commonly used controller at present is VST (Virtual Studio Technology)/VSTi (Virtual Studio Technology Instrument) imported by Steinberg, plugin programs-sound sources and effects. To obtain the best VST effect, the sound effect interface must support the ASIO specification [62]. ASIO abandoned centralized hardware control of the operating system (to reduce the latency (minimum time)), multi-track, and multi-channel audio processing technology.

Currently, the delay time of the MME driver (Microsoft Multimedia Environment) is 200 to 500 milliseconds, DirectSound is 50 to 100 milliseconds, and the Mac OS sound manager is 20 to 50 milliseconds. Under ASIO, the delay of the buffer can be adjusted to between 1 and 10 milliseconds according to different settings and operating systems, thereby bringing instant effects during recording and music production.

The current sound database is stored on the local hard disk, loaded into the memory through the production application, and controlled and used by VSTi software. If the sound database is stored in the cloud, it must be downloaded and monitored through online streaming. Users use the buffers supported by ASIO to access the data, and control the tone database and editing functions through the music production application. After completing the calculation in the cloud, the results will be sent to the user via streaming. Local MIDI control and recording is done through the ASIO protocol (Figure 5).

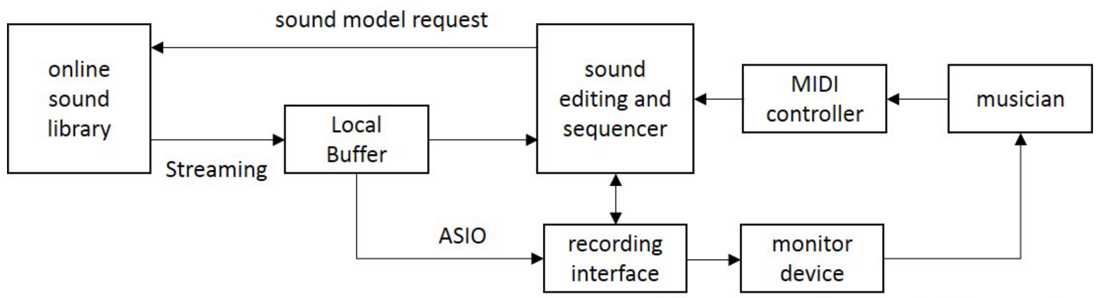

Figure 5: Applying mode of Streaming and Audio Stream Input and Output (ASIO).

2.4 Digital Music Production Industry

The digital music production industry has been booming since digital music use first rose. The structure of the digital music production today is as shown in Figure 6.

While the tools for music production have been replaced by digital technology, the process and work distribution have not changed; however, there is a better possibility of speed, convenience, flexibility, variety, and, in particular, publication. Digital music products arrive in the market fast due to online e-commerce and audio streaming technology development.

In the current technological development and knowledge economy background, the structural evolution of the music industry will also have an impact on the relationship between the music production industry, particularly the protection of copyright and the relationship between profits [51, 63, 64].

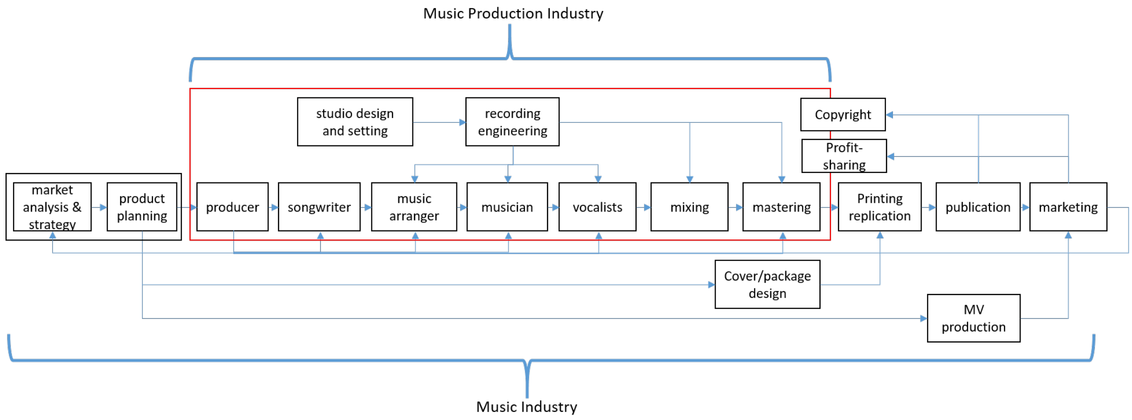

Figure 6: The music industry and music production industry structural diagram.

2.4.1 Digital Sequencer

The arranger is an important part of music production. With the advancement of technology, the software has more powerful functions. The arranger directly uses instrument and effect plugins of the virtual sound module Virtual Studio Technology Instrument (VSTi) to process the audio postproduction [65]. By musical structure planning and arrangement with MIDI controlling devices, multi tracking and transferring after arrangement, which do not need to be transferred to the external tape devices, are completed on the computer. This reduces the sound attenuation and destruction during the conversion between digital and analog signals. The real sound of the instrument or voice is recorded on the computer via the recording interface and then it is mixed together with music tracks that have been made by the virtual software to be output as musical files. A high speed CPU, large size memory, huge data storage space, and fast access data are required of an arranger's working station.

With MIDI devices, all the notes become MIDI data, which is recorded on the computer using a DAW (Digital Audio Workstation) to compose and record. There is close to no delay to play the performance by calling the timbre of the virtual sound source, which is driven through ASIO [65]. Due to mobile devices being universal, music arrangement can be processed not only on desktop working stations but also on mobile devices. Many software companies (Steinberg, Image Line, and Apple) have developed various music production applications for mobile devices. There are also applications for arrangement by artificial intelligence (A.I.) as the technologies of rote learning and artificial intelligence have matured in recent years[66].

Digital arrangement has been substantially changed as mobile devices developed, software/sampling sound/effect plugins were digitalized, controlling devices were diversified, sound identification and editing were elevated, and rote learning/AI (Artificial intelligence) began to be utilized [43]. The advantages of internet networks plus mobilization will be the next focus.

2.4.2 Digital Audio Engineering

Either the arrangement or instrument/vocal recordings must be processed by digital audio engineering, which includes digital recording, digital mixing, and the master process. DAW digital music production software covers most recording, editing, and arranging functions so that the engineer can control different types of functions, recording, adjusting, and adding different effect plugins. All the plugin adjustment parameters can be presented as patterns through digitalization. Software companies offer many sample parameters for the recording engineers to use for adjustment at live recordings. After the mixing of each song is completed for an album, the song goes into the process of mastering.

The primary work of mastering is to obtain better sound performance as a better quality of listening delivers the producer's or singer's original interpretation more directly. Bob Ludwig, a mastering post-production engineer from the well-known Gateway Mastering Studio in the USA, stated, "Mastering is the technical and creative act of balancing, equalizing, and enhancing analogue or digital tapes so that the finished product will have attained the maximum musicality and competitiveness in the open market." Mastering engineers consider how to present the strength of sound and continue the remixer's and singer's ideas of interpretation entirely. Mastering is the last step before publication, and the master file will be sent to CD production or uploaded to the cloud for digital publication. The process of audio engineering and music production is as follows (Figure 7):

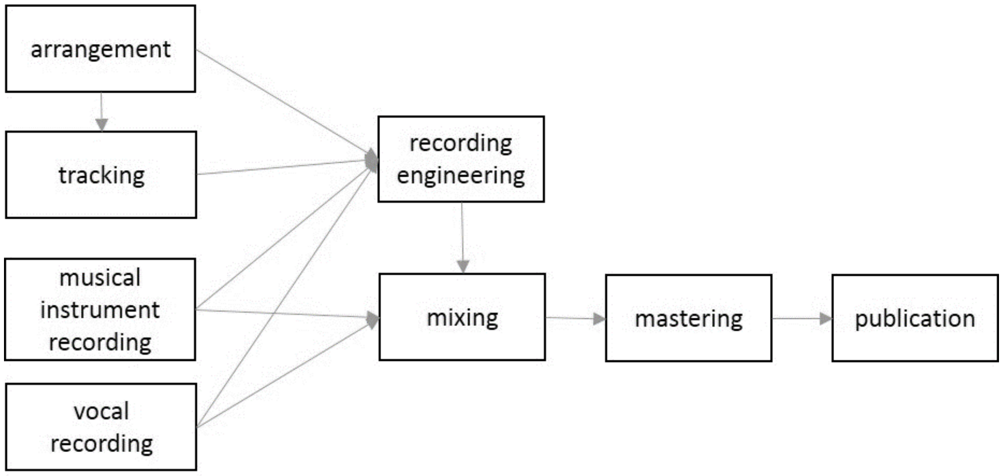

Figure 7: The process of audio engineering and music production.

2.5 Literature summary

At present, the operation mode of the music production industry is primarily carried out in a stand-alone mode. Although the sound library and effect devices have been digitized and many advanced edits can be performed through DAW, the access and editing of sound files requires a great deal of time and hardware costs. In addition, due to regional restrictions, it is difficult for music producers in different regions to cooperate with each other. This chapter summarizes the literature concepts of the music production industry in the cloud collaborative operation mode related to music data processing and audio data network transmission technology as follows (Table 4).

Table 4: Literature summary.

Music Data Processing	Description	References	Research gap
MIDI	MIDI is a digital representation of music, which includes the note's characteristics, expression characteristics, and performance modes.	[5-8]	In the past, research focused on the formulation of MIDI standards and the control of software and hardware peripherals or sound library functions through MIDI controllers.

Table 4: Literature summary.

Music Data Processing	Description	References	Research gap
MusicXML	MusicXML is a royalty-free format that implements all the features offered by XML technology: data structure/modularity/extensibility/ possibility of querying and interaction through XML family technologies for conducting automatic analysis of music, adopting a statistical approach.	[5, 6, 9]	MusicXML can perform data exchange as well as sharing and collaboration functions through the XML standard, but few studies integrate other sound processing technologies into application modes.
AUDIO to MIDI	When audio is converted to a MIDI signal, the notes, pitch, beats, speed, and dynamics of the audio are converted and analyzed. The dynamics can be allocated to the MIDI by equality of the ratios according to the volume analysis.	[12-14]	In the past, scholars' research focused on the processing technology of the conversion of sound frequencies into digital signals and the accuracy of recognition, and there are few application instructions for integrating platforms.
Noise Processing	The spectrum weighting algorithm can quickly and effectively reduce the noise in the sound, which can effectively suppress the noise in the background.	[15-19]	In the past, scholars' research on noise processing focused on the technique of spectrum weighting, and the application direction was mostly background noise processing. In this study, the recording of mobile cloud architecture will become one of the important technologies. In the past, there was less application description regarding this aspect.
Deep learning in the field of music production	Through the collection of a large number of music files and the use of various deep learning algorithms to analyze the characteristics of music and automatically generate high-quality new music fragments.	[39-50]	In the past, scholars' research focused on the application of algorithms and mainly produced music content that could pass the Turing test. It was less focused on the role of deep learning technology in the industry and in the music production process.
Streaming	Streaming is a technology of network data transmission, which uses compressed video files and sends them to the client to buffer on the computer, and controls the way the video is played through time stamps.	[52-55]	In the past, most of the streaming literature discussed video streaming technology and the technology of intelligent judgment of the network environment, and there were few specific frameworks for audio music production applications.
ASIO	ASIO abandoned the centralized hardware control by the operating system to reduce latency (minimum time) and multi-track and multi-channel audio processing technology. Under ASIO, the delay of the buffer can be adjusted to between 1 and 10 milliseconds according to different settings and operating systems, thereby bringing instant effects during recording and music production.	[62, 65]	In the past, more literature compared the delay time and standard setting, and there was less research on integrating different streaming technologies as applied to a network environment.

3. A DEEP LEARNING MUSIC PRODUCTION (DLMP) SYSTEM

3.1 System Framework

As internet technology and mobile smart devices have gained popularity and transmission speed has quickly developed in recent years, the traditional industries are turning these advantages into opportunities. Innovative services in the cloud not only transform the structure of industries but also makes the music industry face a new challenge. Users listen to online streaming music by downloading applications on mobile smart devices. People working in music production from different countries work together for a project via online collaboration. Lately, devices for music production can be controlled by a computer and the recording files are saved on it. After mastering, the music can be published and sold on cloud music publication streaming platforms. Explanations of the deep learning music production (DLMP) system framework (Figure 8) are:

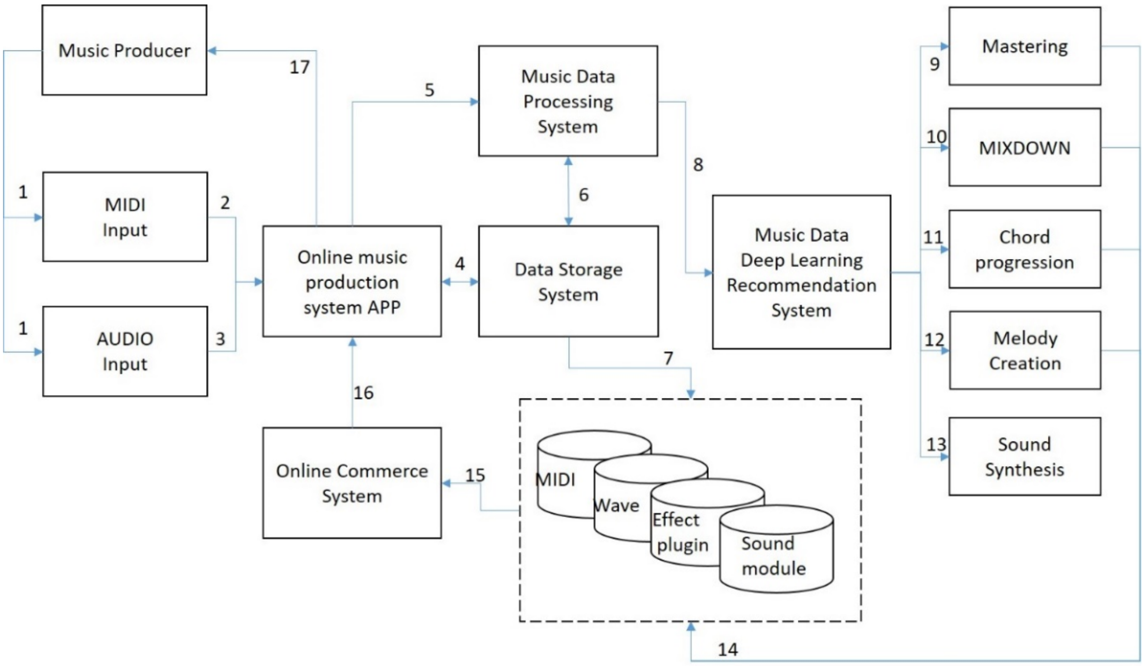

Figure 8: Deep learning music production (DLMP) system framework.

(1) Music producers can use various MIDI controllers to control the sounds loaded from a sound library to perform music performances and can also directly transfer sound files by recording.

(2) Music producers can use the MIDI controller to control the digital sound library data and record the edited MIDI data into the online music arrangement system.

(3) Through direct sound recording, the sound files will be uploaded to the online music production system with streaming technology, and the online music production system will be played back to the production staff for real-time monitoring through streaming. In addition, the environmental sound or noise generated during the sound recording can be edited using noise processing effects.

(4) The recorded files include the data recorded by MIDI and the files of the sound recording, which are stored in separate tracks, and provide editing and modification of complete music production functions. The edited files use intelligent streaming technology to allow the producer to choose to quickly monitor or confirm the complete file.

(5) When editing music content, the editor can perform advanced editing and use the conversion of the music data, such as converting the sound into a MIDI record, or modifying the scale position of the sound through MIDI editing. The processed data will be stored and recorded.

(6) The system can store the processed data and provide production editing, monitoring, and real-time data processing and storage.
(7) The system can store the generated music data in the respective databases according to different types, including the Wave database, MIDI database, Sound Library, and plugin module parameter setting data, and mark the creation, editing, and production owner's record. These records can provide follow-up data exchange and accounting system associations.
(8) After the music data are converted, the data can be input into the deep learning system, and the system algorithm can analyze the music data and produce the recommended information. The algorithms can be analyzed and processed through different data types. For example, after converting a recorded sound file into a MIDI file and a MIDI file into a MUSIC XML file, the recommended content in the music is generated by executing a deep learning algorithm.
(9) Through the deep learning algorithm, the process of processing the sound frequency is applied to the process of music mastering. The frequency of multiple music is adjusted and compared to produce music content that can express the overall balance.
(10) Through deep learning algorithms for mixing work, the system can record the mixing characteristics of the different music types and the frequency characteristics of each instrument and adjust the frequency performance, spatial relationship, timbre and other adjustments. In addition to the overall mixing work, it is also possible to judge and adjust the sound frequency of a single track.
(11) Deep learning analysis of chord progression and chord design in music can provide producers with the function of recommending chord and harmony design suggestions through a large amount of music and analysis. The recommended content can be the content of the previous record, we can choose to directly purchase references or reference learning.
(12) The deep learning algorithm can provide the design and reference content of the producer's melody through music melody learning. The melody must be based on the chord design. The deep learning algorithm can identify the chords and provide the recommended melody design based on the chord changes in real time.
(13) Through the learning of sound frequencies and various effects, deep learning can produce different synthesized sounds. These sounds can be used as new sound libraries for music producers. The synthesized sound can also be the unique creation and synthesis of the music producer, and can also become a musical element that the producer can purchase later.
(14) The music elements generated through the deep learning recommendation system will be recorded in their respective databases for provide instant use or use by future music producers. These generated elements will also record the editors, creators, producers, etc.-related roles for subsequent profit sharing.
(15) The music elements and recommended content in the database can be transformed into music commodities and transactions can be conducted through an online commerce system.
(16) The online music production system combined with the online commerce system allows producers to make their own productions or combine more online resources to make the music more diverse and professional, and enhance the quality of music production. In addition, this can increase the source revenue for music producers.
(17) The online music production system provides producers with quick monitoring of music content or a selection of complete quality monitoring through intelligent streaming technology. When an editor needs to quickly confirm whether the music content is correct, they do not need a return of the highest quality, which can save

on transmission bandwidth. If they require a more detailed confirmation of the quality and that the content is correct, then they need the highest quality file transferred to the production side. In addition, the part that uses the sound library can also be used for fast monitoring to perform and edit, which can save the time of network transmission and data loading.

3.2 Analysis of the system module

The system module of the study includes: an online collaboration module, a music arranger editing module, production commerce module, a MIDI and WAVE database, a sound source and timbre database, a recording/remixing effect plugin parameters database, a deep learning data analysis and recommendation module, a purchase and subscribe module, and an accounting module. Explanations of Figure 9 are as follows:

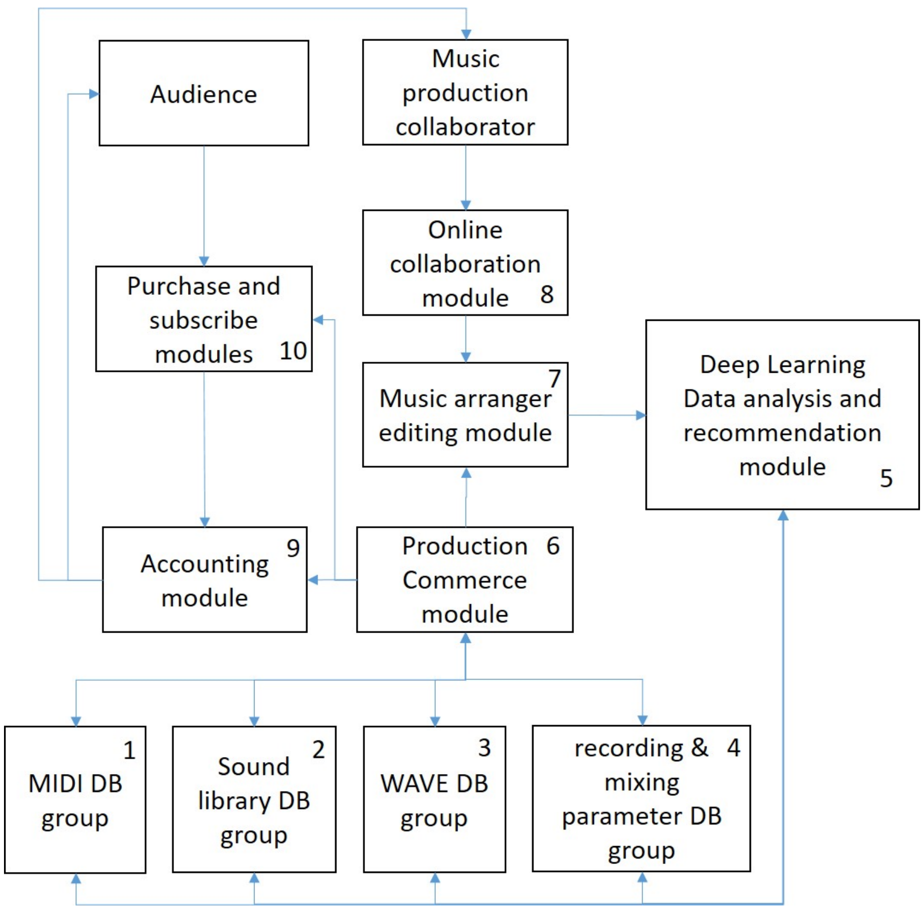

Figure 9: DLMP System Module Analysis.

(1) The MIDI DB Group records all MIDI files in the process of music arrangement. MIDI files record all relevant information of the music, including the original created MIDI files, the data files converted to MusicXML, and the MIDI files converted from the recorded sound files, including the recommended MIDI files generated through the deep learning system.

(2) The sound library DB group includes a database of digital sound library types, providing the need for arrangement in music production, such as drums, piano, and guitar sounds. In addition, this database provides new synthetic sound data generated by deep learning technology, so that producers have more choices of elements in music. The sound library DB group requires huge data space and data read speed. It must be able to access and monitor the sound performance quickly

through intelligent media streaming technology. Through the streaming mechanism that monitors while downloading, it can be adjusted for different network bandwidths and local buffers.

(3) The WAVE DB group is a collection database of sound files. The WAVE data generated after recording will be recorded in the WAVE DB group and provide subsequent editing by the production staff. The WAVE database also stores the information of each track after production. The WAVE file contains huge information for subsequent editing, such as the message converted into MIDI through WAVE through deep learning technology to learn the mixing of the sound frequency distribution and mastering information, and the new sound synthesis data generated after the editing in each sound track. The WAVE DB group also requires huge storage space and reading speed. During recording, the data can be uploaded to the online production system synchronously through the local recording and online streaming applications for monitoring and online collaborative operations. The most important thing is to reduce the transmission delay of sound files in order to provide the state of simultaneous monitoring during recording. In addition, the new WAVE recommendation data set generated by deep learning technology will also be stored in the WAVE DB group.

(4) In digital music production, in addition to using the sounds in the sound library and plugin effects, there are many parameters that need to be adjusted, such as frequency adjustment data, space, or reverb use. These data will be recorded in the recording and mixing parameter DB group. These data record the music producer's sound processing skills and experience. Through the adjustment of the parameters, the sound of each track can be more appropriately represented. The configuration and design of the overall music elements must be rigorously set and adjusted to provide the audience with a perfect musical performance. The parameters of the recording and mixing parameter DB group also provide the use of deep learning technology, which can produce different parameter setting suggestions that can be recommended to the production staff, and combine the reference mechanism to increase the added value of the relevant production staff. Deep learning technology is currently a popular technology used in many fields. In the music field, deep learning can analyze music and design to propose self-made elements or recommend past music elements.

(5) In the deep learning data analysis module, MIDI data type analysis will be performed, the recommended MIDI file will be produced, and the recommended information can also be generated by analyzing the music score graphics. In the sound library type data, deep learning can perform new timbre synthesis calculations, so that the change of timbre is more diverse. In WAVE-type data, deep learning can convert WAVE to MIDI for analysis and, then, present the results through MIDI to WAVE. In addition to directly analyzing the performance of WAVE, this can also perform automatic mixing and mastering, and provide recommended parameters. The application of deep learning technology in the music field includes chord recommendations, melody recommendations, timbre processing, and sound processing types. In the future, we can continue to develop various modules that contribute to the music production process.

(6) In the production commerce module, various business mechanisms for music production elements are provided. The content can create new music elements for the production staff, or suggestion data generated through deep learning technology. Data types include MIDI, WAVE, chord progression, mixing, etc. To create more value for the production staff and distribute profits fairly, the system must record the owner of each part of the data and the proportion of profit

distribution. Music producers can not only create their own content for profit, but also use other people's content to distribute profits to other producers. In addition to improving the quality of the produced content, they can also create multiple interests.

(7) The music arranger editing module can provide the editing functions required for music production, performing music production and recording through split tracks, and can support the use of online effect plugins. The music arranger editing module must provide the use of MIDIs, WAVE recordings, a digital sound library, effect plugins, and also must have a user-friendly operation interface to provide various types of music editing and music element output and input functions. In the music arranger editing module, the commercial products provided by the production commerce module must also be combined to help the music producer obtain more recommended information when creating, so that the production environment has a more convenient and diverse editing space. The content edited in the music arranger editing module can also be directly connected to the deep learning module for data analysis and recommendation, so that the producer can grasp more information and improve the quality of music production.

(8) The online collaboration module provides a collaboration interface in different places through the internet. Due to the convenience of the network and mobiles, music production is not necessarily limited to one place. We can even interact and cooperate with music producers in different countries. Through the online collaboration module, different music producers can be located in different regions or in action, making the creation of music without borders, and fully utilizing the characteristics of the internet. Through the online streaming technology, the recorded sound can be transmitted to the production end for monitoring and online storage in real time. Therefore, other collaborative workers can continue to use the files in subsequent editing. In addition, there may be a great deal of environmental noise in the music recording environment in action. The system must filter unnecessary messages generated during online collaboration, such as environmental sounds and noise processing.

(9) The accounting module provides the processing of all accounts. Different from the past, the accounting module not only handles the transaction of music elements, but also a mechanism for profit distribution through the profits that consumers will obtain through the purchase. In the accounting module, the music production staff includes the purchase of music-related elements and the purchase of music elements produced through deep learning technology. This module can also sell each production element and finished product. Profit sharing after the transaction can help the music industry to increase revenue and can be shared fairly with the relevant production personnel.

(10) The purchase and subscribe modules allow the audience to subscribe and purchase. Through direct sales or the establishment of a subscription system, music producers do not need to distribute profits through centralized units; instead, the online production and online sales can automatically achieve a fair distribution of profits. This is an operation mode that can be achieved through the characteristics of the internet.

4. ANALYSIS AND COMPARISON

4.1 Value Chain analysis

The concept of a value chain is from "Value Chain Analysis (Primary Activity and Support Activity)" presented by Michael Porter [67]. Kogut Bruce mentioned that the value chain equals a series of activities of a product's market value [68]. In addition, Porter also claimed that a series of value chains is the value system, upstream/midstream/downstream suppliers, enterprises, channel distributors, and customers involved [69]. If the industry plans to develop its advantages, it has to determine the advantages from its value system and the need capacities for integrating the value chain system.

As the music production industry is affected by digitalization, the business operation mode has been transformed from CD selling to live performing and endorsed advertising. However, the relations between the value inventors of the industry do not have too many variations. This study proposes that, under the ideas of the structure of DLMP, a model can be set up with the value of every contributor. A relational structure with upstream/downstream companies of the digital music production is as seen in Figure 10.

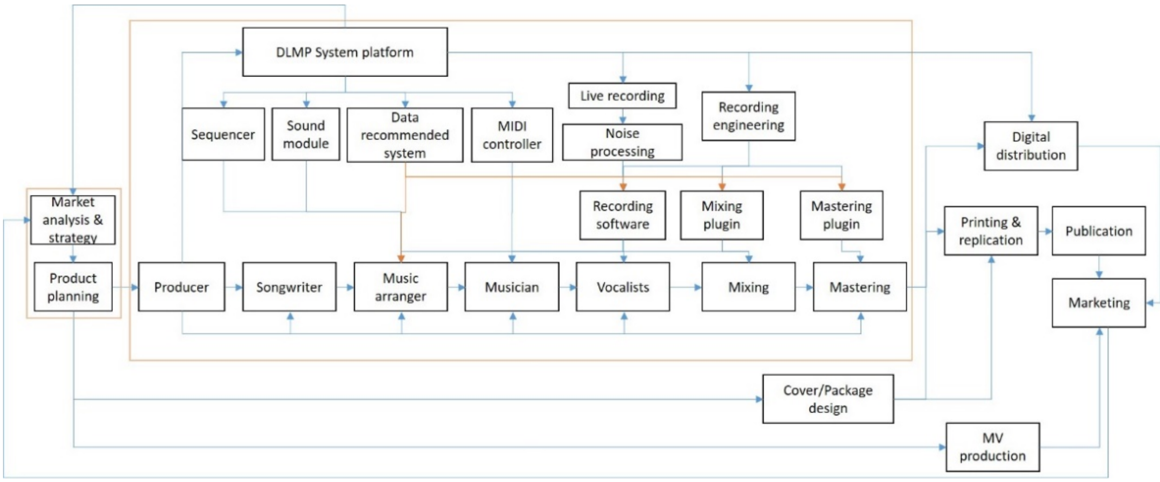

Figure 10: Company relationship structure in the DLMP System Framework.

According to the recent state of the digital music production industry, we analyzed and compared the essence of the value of each affiliated staff. Upstream: record companies, audio companies, and platform providers in the cloud. Midstream: music production staff and audio engineers. Downstream: distributors and buyers (Figure 11).

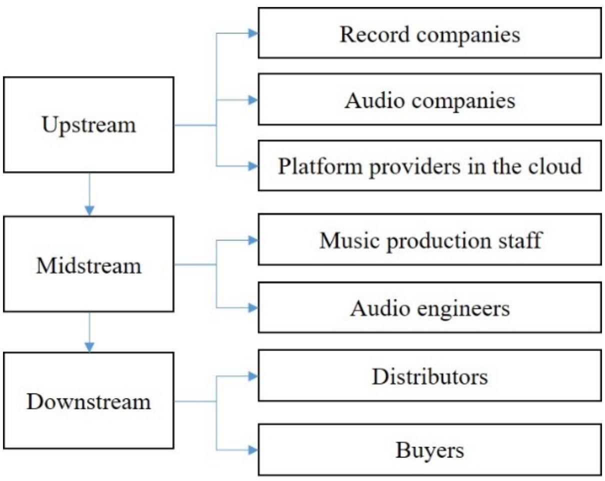

Figure 11: The role of the music production industry value chain.

We describe the current status of the industry and the new value chain role of the industry's structure to analyze the advantages, as shown in Table 5.

Table 5: Analysis and comparison of each staff role value.

Industrial Relation	**Unit**	**Staff Role**	**Key Points of Value Assessment**	**Advantages**	
				Under the Current Industry	**Under the Study Structure**
Upstream	Record Company	Market Analysis	Actual market contact	High	Medium
			Channel and sales information	High	High
			Online operation platform	Listen Habit Playlist	Listen Habit Playlist Production Analysis
		Product Planning	Subjective experience and ability	High	High
			Creativity and integration	High	High
			Visual design ability	High	High
		Marketing	Marketing opportunity integration	High	High
			Endorsement and sales capabilities	High	High
		Financial Management	Interest generation	Low	High
			Profit distribution mechanism	Low	High
	Audio Company	Studio Planning	Equipment and operating capacity	High	Low
			Hardware space capability	High	Low
			Financial and economic ability	High	Low
		Software and Hardware Development	Data digitization	High	High
			Data cloud status	Low	High
			Modularization	Medium	High
			Standardization	Low	High
	Platform Company in the Cloud	Service Planning	Transmission quality	Medium	High
			Online computing performance	Medium	High
			Mobile application development	Medium	High
			Operation system support	High	High
			User interface functions	Medium	High
			Network resource sharing	Medium	High
			Multiple purchase mode	Trial and Subscription	High
			Member Information Analysis	Listen and Subscription Record	High Utilization

Table 5: Analysis and comparison of each staff role value.

Industrial Relation	Unit	Staff Role	Key Points of Value Assessment	Advantages	
				Under the Current Industry	Under the Study Structure
Upstream	Platform Company in the Cloud	Service Planning	Business module	Subscription only	Diversity
			Music production collaboration module	NA	High
			Music production database analysis	NA	High
			Platform module integration ability	NA	High
			Development flexibility	NA	High
			Platform development cost	Medium	High
Midstream	Music Production Staff	Producer	Music skills learning	High	Medium
			Music production experience	High	Medium
			Creative integration skills	High	High
			Communication skills	High	High
			Project management capabilities	High	Medium
			Market insight	High	High
		Songwriter	Observation and perception	High	High
			Emotional sensibility	High	High
			Sound sensitivity	High	High
			Literature and Art Conservation	High	High
			Imagination	High	Medium
		Music Arranger	Music Theory Understanding	High	Medium
			Performance ability	Medium	Low
			Use instrument sound ability	High	High
			Ability to use production tools	High	High
			Sound source	High	Low
			Creative ability	High	High
		Musician	Ability to read sheet music	High	High
			Acting ability	High	Medium
			Music theory	High	High
			Improvisation	High	High
			Cooperation and coordination	High	Medium
		Backing Vocalist and Singer	Pronunciation, listening	High	High
			backing singer's listening ability	High	High

Table 5: Analysis and comparison of each staff role value.

Industrial Relation	Unit	Staff Role	Key Points of Value Assessment	Advantages	
				Under the Current Industry	Under the Study Structure
Midstream	Music Production Staff	Backing Vocalist and Singer	Sound skills	High	High
	Audio Engineer	Engineer	Software and hardware familiarity and operation ability	High	High
			Sound control and parameter adjustment capabilities	High	High
			Software and hardware knowledge	High	Low
			Concept of space and frequency	High	High
			Listening ability	High	High
Down-stream	Distributor	Publication Channel	Physical channel distribution	High	Low
			Virtual channel circulation	High	High
			Quick payment method	High	High
		Sales Channel	Online marketing application	Medium	High
			Media channel integration	High	Medium
			Online platform exposure	High	High
			Marketing activity capability	High	High
	Consumer	Customer: Music	Ability to pay	Low	Medium
			Community influence	Medium	High
			Music and idol preferences	High	High
		Customer: Music Production Staff	Ability to pay	Low	Medium
			Community influence	Medium	High
			Device brand preference and proficiency	Medium	Low

4.2 Innovative Operation Model

A study by Mitchell et al. [70] indicated that outstanding performance enterprises continue reviewing and updating their business models to adapt themselves to the market. The music industry's business models not only show the business logic of an enterprise but also imply its possibilities in the future. A study by Osterwalder et al. [71] also complied with common combinations of the business models to compare and research into their elements. They introduced the elements, four pillars, and nine building blocks to constitute a business model [71, 72]. "A business model is a conceptual tool, including its definition, type of element, attributes, and relationships to other elements, and business logics of an enterprise. An enterprises brings the value to more than one groups of customer, statement of the business structure, creation, marketing, delivering value, partnership network, profits, and in line with revenue" [73].

Research into the business model supports the definition of the participants, roles, responsibilities, market size, business target, core competitiveness, relationship module, revenue module, value exchange map, and key factors of success [74]. This will assist the management to obtain, understand, communicate, design, and analyze, as well as provide a conceptual tool for altering their "business logic" [73]. We state and compare the digital music production industry and the suggested DLMP industry based on the four pillars and nine building blocks (Table 6).

Table 6: Comparison of business model elements.

Four Pillars	Nine Building Blocks	**Digital Music Production**	**DLMP**	Remark
Products	Value Points	. Equipment digitization and software. . Hardware equipment quality improvement. . Obtain product ownership.	. Software and mobile product development. . Have the basic equipment to use the service. . Purchase modular products as needed. Increase the income of music staff.	Propose overall concepts of the product and service to the enterprises
Clients	Target Customers	. Professional music staff. . Business investment.	. Professional music staff. . Non-music expertise. . Enterprise and cloud service developers.	Describe the target customers of the Sales Price
	Distribution Channel	. Online purchase. . Agent. . International business exchange.	. Download or order online. . Agent. . International business exchange.	Describe methods of contact between the enterprises and variety groups of customers
	Relationship	. One-time purchase. . After-sales service and warranty. . Purchase of external expansion modules. . Education and training services. . license. . Agent and Distribution.	. Establishment of a collaborative relationship without borders. . Music product relationship building. . Cloud hardware and software maintenance. . Purchase model of plugin expansion module. . Online learning. . Authorization mode. . Application Developer and Platform Relationship.	Explain the types of connection between the enterprises and variety groups of customer
Fundamental Management	Value Types	. Equipment grade determines value. . Producer technology determines value. . Personal learning experience determines value.	. Not limited by time and space. . Deep learning technology provides added value. . Cloud-based distributed computing model. . Gather the wisdom of everyone. . Quick and convenient collaborative value-added.	Describe activities and resource arrangement
	Core Abilities	. The ability of the device to handle sound quality. . Computer execution capability and hardware space. . Post-processing capability of the processor. . Physical space and sound technology. . Overall software and hardware operation ability.	. Application development on different platforms. . Cloud computing processor capabilities. . Cloud data space and management capabilities. . Deep learning algorithm capabilities. . Recommended accuracy. . Cloud business mechanism design. . Profit distribution and accounting mechanism.	Necessary abilities of executing the business models

Table 6: Comparison of business model elements.

Four Pillars	Nine Building Blocks	**Digital Music Production**	**DLMP**	Remark
Fundamental Management	Partners	. Technology provider of hardware and software manufacturers. . Limited music production staff partners. . Record companies and platform business partners.	. Developers of hardware and software manufacturers. . Global Music Production Staff Link. . Cloud platform vendor maintenance partner.	Be able to propose an effective cooperation agreement internet with other enterprises
Finance	Costs	. Software purchase. . Hardware purchase. . Software and hardware upgrade and update costs. . Space decoration. . Staff cost. . Maintenance cost.	. Software package purchase. . Corresponding hardware purchase. . Recommended content purchase. . Music staff production costs. . System platform construction and maintenance costs.	Overall business models by financial results
	Revenues	. Physical record revenue. . Online music purchase income. . Online streaming membership fee income. . Lead singer and song copyright income. . Platform advertising revenue. . Music staff one-time income.	. Physical record revenue. . Online music purchase income. . Online streaming membership fee income. . Unrestricted income of lead singer and song copyright. . Music staff distributes profit income. . Music staff intellectual property income. . Platform advertising revenue.	Describe the revenue sources and methods of profit

5. DISCUSSION

In this study, we built a new music production industry operating framework through literature research, making full use of the advantages of the internet and digital trends to depict the relationship of the future music production industry's industrial value chain. By analyzing the current key processing technology of music data to construct the DLMP system framework, we provide a blueprint for industrial development.

This study shows that the DLMP system framework can bring many positive industrial values to the music production industry and can also enhance the industry's competitive ability in the innovative operating model.

5.1 Research Contributions and Limitations

(1) We explained the key technology and discussion of music data processing and provide the direction of future researchers' technology development.

(2) We proposed the DLMP system framework and detailed system module analysis to provide the basis for the development direction of industrial system application developers.

(3) The value chain was used to compare the current industry with the industry status changed by DLMP and to provide a clear assessment of the advantages and disadvantages of the industry.

(4) We provided a sustainable development model combining deep learning and digital technology in the music production industry.

(5) As the DLMP music production platform has not yet been developed, many technologies are also limited to a single application method. The challenges for

integrating this technology and the user behaviors of future industrial workers are currently less fully explained and empirical. The limitations of this research are also the directions of future research.

5.2 Practitioner implications

The practitioners in the framework of the DLMP system are divided into four different roles:

(1) System environment provider role:
In the application of mobile cloud systems, the cloud services and online control provided by system manufacturers play very important roles, including the user rights, online resource allocation, audio and video server technology, computing capabilities, and storage and reading capabilities. The architecture proposed in the study allows system vendors to build the direction of the future system platform technology research and development.

(2) Music production hardware and software provider roles:
In the past analogy era, the majority of the equipment needed in the music production process used hardware. After digitization, many digital software programs gradually replaced hardware devices; however, most of them were stand-alone versions. In the DLMP framework proposed in this study, many software and hardware solutions must be able to be digitalized in the cloud. Therefore, in the role of music production software and hardware providers, we must think about the future software development model and provide functions, such as online collaboration and be able to combine data collection to provide subsequent data analysis capabilities.

(3) Music content management role:
In the music production industry, the content of music includes the management of the sound database, the management of the content of the materials produced by the music production, and the management of the finished music. Unlike the past music content management methods, music is not only a finished product, but also includes the data management generated in each step of the production process. These are the basis for more data analysis and copyright sharing profits in the future.

(4) Music producer role
The most direct impact that music producers will face is the change in the production process, although the operation mode of the DLMP system framework can improve the music production process and solve many problems in the music production industry, such as a lack of easy cooperation, improvement of the copyright distribution model, management of huge material, and enhancement of the value of music content; however, the familiar working methods of music producers are limited by the cost of using technology. Perhaps greater industry impact and policy forces are needed to allow music producers to start thinking about the perceptions that must be changed.

6. CONCLUSION

This study described the development and sustainability of the DLMP framework, and provided detailed system processes and descriptions of key system modules. This research also analyzed the value chain and an innovative business model and proposed key technologies for online collaborative operations and online music production, as well as built a new industrial chain structure. In addition, this research framework provided a conceptual description of the impact of the industry through the use of deep learning technology to provide strategic thinking for future industry developers and music producers. The main conclusions of this research are as follows:

(1) From observing various market trends, the internet and collaborative operation modes appear to be the future development direction of digital music production. This system operation mode requires a great deal of technical intervention, whether it is network transmission, data reading, voice processing, data conversion, etc. Application manufacturers have a huge development space, and they must invest in technology that can help with the development of the industry.

(2) The application of deep learning requires a clearer application method in music production. The music production industry is a highly creative industry and a unique way in which people can express emotions through art. Although deep learning can be carried out by studying past music creation and can present high-quality content that passes the Turing test, we should think about how to retain the unique human emotions and artistic presentation, rather than replace it by deep learning and, thereby, find and develop computer and human creativity in a coexistence mode of application.

(3) The music production industry is currently in a challenging new situation. The innovative operation model can help the application development and product strategic direction of all stakeholders in the industry chain. Although there are many key technologies in the framework of this research that need to be further advanced, the DLMP framework built through this research is expected to give the industry a more macro development blueprint and method of thinking, so that technology can assist in the development of the music industry rather than restrict or replace aspects.

7. REFERENCES

[1] S.-S. Weng and H.-C. Chen, "Exploring the role of deep learning technology in the sustainable development of the music production industry," *Sustainability,* Vol. 12, No. 2, p. 625, 2020.

[2] A. H. Weis, "Commercialization of the internet," *Internet Research,* Vol. 20, pp. 420-435, 2010.

[3] G. Graham, B. Burnes, G. J. Lewis, and J. Langer, "The transformation of the music industry supply chain: A major label perspective," *International Journal of Operations & Production Management,* Vol. 24, pp. 1087-1103, 2004.

[4] P. M. C. Swatman, C. Krueger, and K. van der Beek, "The changing digital content landscape: An evaluation of e-business model development in European online news and music," *Internet Research,* Vol. 16, pp. 53-80, 2006.

[5] R. Viglianti, "Musicxml: an xml based approach to musicological analysis," in *Digital Humanities 2007: Conference Abstracts*, pp. 235-37, 2007.

[6] Y. Wan-jun, "From midi to musicxml-the development of computer music score information exchange format," *Entertainment Technology,* pp. 45-49,

2014.

[7] D. B. Williams and P. R. Webster, Experiencing Music Technology, 3th ed. USA: Schirmer Cengage Learning, 2008.

[8] I.-H. Lee, "The Research of Application of Digital Sound Technology into Music Composing," master, Department of Music, National Taipei University of Education, 2006.

[9] M. Good, "Musicxml for notation and analysis," *The Virtual Score: Representation, Retrieval, Restoration,,* Vol. 12, pp. 113-124, 2001.

[10] A. Kirke and E. R. Miranda, "A survey of computer systems for expressive music performance," *ACM Computing Surveys,* Vol. 42, pp. 1-41, 2009.

[11] E. Begoli and J. Horey, "Design principles for effective knowledge discovery from big data," in *2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture*, ed, pp. 215-218, 2012.

[12] W. Joe. (2005). Note names, MIDI numbers and frequencies[Online]. Available: https://newt.phys.unsw.edu.au/jw/notes.html

[13] H. Cheng. (2013). Why keys are arranged in a geometric sequence[Online]. Available: https://wenku.baidu.com/view/ca76790cb52acfc789ebc9bb.html

[14] Wiki. Frequency[Online]. Available: https://en.wikipedia.org/wiki/Frequency

[15] N. Upadhyay and A. Karmakar, "Spectral subtractive-type algorithms for enhancement of noisy speech: an integrative review," *International Journal of Image, Graphics & Signal Processing,* Vol. 5, No. 11, 2013.

[16] Y. Ephraim, H. Lev-Ari, and W. J. Roberts, "A brief survey of speech enhancement," *The Electronic Handbook,* Vol. 2, 2003.

[17] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement," *The electrical engineering handbook,* pp. 12-26, 2006.

[18] Y. Ephraim and Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE,* Vol. 80, pp. 1526-1555, 1992.

[19] Y. Gong, "Speech recognition in noisy environments: A survey," *Speech Communication,* Vol. 16, pp. 261-291, 1995.

[20] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, ed, pp. 208-211, 1979.

[21] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Transactions on Speech and Audio Processing,* Vol. 6, pp. 328-336, 1998.

[22] M. Bhatnagar and P. C. Loizou, "A cross-correlation technique for enhancing speech corrupted with correlated noise," *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat.*

No.01CH37221), pp. 673-676, 2001.

[23] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars," *Speech communication,* Vol. 11, No. 2-3, pp. 215-228, 1992.

[24] S. Kamath, P. Loizou, and U. States, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise an event-based acoustic-phonetic approach for speech segmentation and e-set recognition," *in Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Orlando, USA,* Vol. 4, p. 7803, 2002.

[25] M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab, and F. E. Abd El-samie, "Speech enhancement using an adaptive wiener filtering approach," *Progress In Electromagnetics Research M,* Vol. 4, pp. 167-184, 2008.

[26] S. O. Non-member and T. Shimamura, "Reinforced spectral subtraction method to enhance speech signal," *Evaluation,* pp. 242-245, 2001.

[27] P. Sovka, P. Pollak, and J. Kybic, "Extended spectral subtraction," *in European Signal Processing Conference (EUSIPCO--96,* pp. 963-966, 1996.

[28] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on Speech and Audio Processing,* Vol. 7, pp. 126-137, 1999.

[29] S. F. Boll, "Supression of noise in speech using the saber method," *Ieee,* Vol. 3, pp. 606-609, 1978.

[30] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing,* Vol. 27, pp. 113-120, 1979.

[31] S. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.*, Vol. 4, pp. 200-203: IEEE, 1979.

[32] H. T. Yeh, J. S. Chiou, and T. J. Zhou, "A karaoke system with real-time media merging and sharing functions for a cloud-computing-integrated mobile device," *Mathematical Problems in Engineering,* Vol. 2013, 2013.

[33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* Vol. 521, No. 7553, pp. 436-444, 2015.

[34] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks,* Vol. 61, pp. 85-117, 2015.

[35] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks,* Vol. 8, No. 1, pp. 98-113, 1997.

[36] M. I. Jordan, "Serial order: A parallel distributed processing approach,"

Advances in psychology, Vol. 121, pp. 471-495, 1997.

[37] J. L. Elman, "Finding structure in time," *Cognitive science,* Vol. 14, No. 2, pp. 179-211, 1990.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* Vol. 9, No. 8, pp. 1735-1780, 1997.

[39] H. Hild, J. Feulner, and W. Menzel, "Harmonet: A neural net for harmonizing chorales in the style of JS Bach," in *Advances in neural information processing systems*, pp. 267-274, 1992.

[40] D. Hörnel and P. Degenhardt, "A NeuralOrganist Improvising Baroque-Style Melodic Variations," 1997.

[41] D. Hörnel and W. Menzel, "Learning musical structure and style with neural networks," *Computer Music Journal,* Vol. 22, No. 4, pp. 44-62, 1998.

[42] M. C. Mozer, "Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing," *Connection Science,* Vol. 6, No. 2-3, pp. 247-280, 1994.

[43] D. Eck and J. Schmidhuber, "A first look at music composition using lstm recurrent neural networks," *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale,* Vol. 103, 2002.

[44] I. Sutskever, G. E. Hinton, and G. W. Taylor, "The recurrent temporal restricted boltzmann machine," in *Advances in Neural Information Processing Systems*, pp. 1601-1608, 2009.

[45] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," *arXiv preprint arXiv:1206.6392,* 2012.

[46] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing,* Vol. 45, No. 11, pp. 2673-2681, 1997.

[47] I. Malik, "Neural Translation of Musical Style," master, COMPUTER SCIENCE, University of Bristol, 2017.

[48] T. Kubo. (2017). Next Music Production by Google Magenta[Online]. Available: https://www.slideshare.net/takahirokubo7792/tech-circle-23-next-music-productionby-google-magenta

[49] J. Engel *et al.*, "Neural audio synthesis of musical notes with wavenet autoencoders," *arXiv preprint arXiv:1704.01279,* 2017.

[50] B. L. Sturm *et al.*, "Machine learning research that matters for music creation: A case study," *Journal of New Music Research,* Vol. 48, No. 1, pp. 36-55, 2019.

[51] S.-S. C. Weng, Hung-Chia, "Exploring the role of deep learning technology in the sustainable development of the music production industry.," *Sustainability,*

Vol. 12, No. 2, p. 625, 2020.

[52] S.-R. Jen, Q.-t. Guo, S.-p. Wang, Y.-r. Yang, C.-h. Yan, and G.-j. Zhou, Introduction and Application of Multimedia 3th ed. Taipei: Flag Publishing, 2008.

[53] T.-H. Tu, "Design and Implementation of Embedded Network Audio Broadcasting Receiver System," master, Electrical Engineering, Southern Taiwan University of Science and Technology, Taiwan, 2010.

[54] Y.-T. Liu, "Design and Implementation of Adaptive Streaming Media Player based on HTTP Live Streaming Protocol," master, Electricl Engineering, National Dong Hwa University, Taiwan, 2012.

[55] S.-H. Lu, "Design and Implementation of Mobile Multi-Media Streaming Service and Management System – with Museum as a Case Study," master, Computer Science and Information Engineering, Shu-Te University, Taiwan, 2012.

[56] T. Berners-Lee, R. Fielding, and H. Frystyk, "Hypertext transfer protocol–http/1.0," *Network Working Group,* pp. 1-122, 1996.

[57] W. Contributors, "Transmission Control Protocol," *Wikipedia, The Free Encyclopedia.,* 2014.

[58] H. Schulzrinne, A. Rao, and R. Lanphier, "Real Time Streaming Protocol (RTSP)," *Request for Comments (RFC) 2326,* 1998.

[59] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "Rtp: a transport protocol for real-time applications," *Request for Comments (RFC) 3550,* pp. 1-89, 2003.

[60] W. Goralski, *The illustrated network: how TCP/IP works in a modern network.* Morgan Kaufmann, 2017.

[61] J. Ott, J. Chesterfield, and E. Schooler, "Rtp control protocol (rtcp) extensions for single-source multicast sessions with unicast feedback," *Request for Comments (RFC) 5760,* pp. 1-66, 2010.

[62] W. Wcat. (2011). Asio, Asio 4All, Ks, Was Api[Online]. Available: http://www.360doc.com/content/11/0910/12/7558399_147244236.shtml

[63] J. B. Meisel and T. S. Sullivan, "The impact of the internet on the law and economics of the music industry," *info,* Vol. 4, pp. 16-22, 2002.

[64] P. Preston and J. Rogers, "Social networks, legal innovations and the “new” music industry," *info,* Vol. 13, pp. 8-19, 2011.

[65] Y.-h. Lo, "From the application of computer and audio technology to explore the concept of digital music creation and practice," master, Music, National Taiwan Normal University, Taiwan, 2009.

[66] J. Braasch, "A cybernetic model approach for free jazz improvisations," *Kybernetes,* Vol. 40, pp. 984-994, 2011.

[67] M. Porter, "Creating and sustaining superior performance. Competitive Advantage," *NY: Free Press.* pp. 167*,* 1985.

[68] B. Kogut, "Designing global strategies: Profiting from operational flexibility," *Sloan Management Review,* Vol. 27, pp. 27-38, 1985.

[69] L.-w. Liu, "The Influence of Textbook Digitalization on the Value Chain of College Textbook Publishing Industry," master, Information Management, National Taiwan University of Science and Technology, Taiwan, 2009.

[70] D. Mitchell and C. Coles, "The ultimate competitive advantage of continuing business model innovation," *Journal of Business Strategy,* Vol. 24, pp. 15-21, 2003.

[71] A. Osterwalder and Y. Pigneur, "Clarifying business models : origins , present , and future of the concept," *Communications of the association for Information Systems,* Vol. 15, pp. 1-125, 2005.

[72] A. Osterwalder and Y. Pigneur, "An ontology for e-business models," *Value creation from e-business models,* Vol. 1, pp. 65-97, 2004.

[73] J.-L. BAI, "Explore the Taiwan online music business model," master, Business Administration, National Chengchi University, Taiwan, 2007.

[74] A. G. Pateli and G. M. Giaglis, "Technology innovation-induced business model change: a contingency approach," *Journal of Organizational Change Management,* Vol. 18, pp. 167-183, 2005.