研究計畫

壹、研究計畫題目

深樂家：基於深度學習之多樂器音樂作曲與演奏合成系統

Deep Musician: Multi-instrument Music Composition and Performance Synthesis using Deep Learning

貳、研究計畫摘要

自動音樂創作擁有數十年的歷史，因其高自由度與挑戰性，一直被視為邁向通用人工智慧的一個重要里程碑。近年來，**機器學習為各領域帶來革命性的轉變，人工智慧亦將會是未來音樂產業轉型的重要技術**。然而，雖然許多研究者嘗試利用機器學習於自動音樂創作，現有技術距離被實際應用於專業音樂製作仍有所差距。在自動音樂作曲方面，過去研究僅探討旋律、爵士樂譜（旋律與和弦）、四聲部聖詠的生成，但現代流行音樂通常由眾多樂器與音軌組成，擁有更為複雜的結構和織體，因此現有音樂生成模型並無法直接被套用至多樂器音樂作曲。在音樂演奏合成方面，現有樂器合成器仍無法達到人類演奏的細膩度與靈活度，多數合成器僅提供單調的演奏風格，並且無法根據樂譜上的演奏符號（如動態、圓滑線、特殊技巧等）改變其音色，致使音樂創作者難以藉由合成器模擬完整作品的聲響效果，必須在錄音階段方能調整其編曲。有鑑於此，**本計畫旨在應用前瞻深度學習技術，開發協助音樂創作者創作音樂的多樂器音樂作曲與演奏合成系統**。此二年期計畫將分為以下三個階段：（一）開發音樂深度學習模型：我將研究當前神經網路架構並探討如何將音樂知識植入其中，開發適用於音樂資料的深度學習模型，以學習高效率的音樂特徵。（二）開發多樂器音樂作曲系統：此系統將由數個基於 Transformer 神經網路的樂器模型及一個編曲模型組成，每個樂器模型學習編寫特定樂器的音樂，而編曲模型學習選擇適當的樂器並協調各個樂器模型，以產生完整的多樂器音樂。（三）開發音樂演奏合成模型：此系統將結合 Transformer 神經網路與生成式對抗網路，前者學習處理樂譜上的音符與表現符號，而後者學習將其合成為音訊，同時我將導入隱變數模型以模擬多樣的演奏風格。目前，我在鋼琴音樂自動配器以及小提琴與鋼琴演奏合成方面均已獲得良好的初步研究結果。同時，我已蒐集超過百萬首的樂譜並建立當前資料量最大的樂譜資料庫，將作為我所提出的深度學習模型的訓練資料。**我展望此系統與專業音樂製作軟體整合，藉由降低音樂製作的技術門檻，促進音樂創作的普及化。此外，本系統可以提供內容創作者低成本之免權利金音樂素材，在音樂治療與教育上亦可降低個人化課程的製作成本。**

The history of automatic music creation dates back to decades. Due to its high flexibility and difficulty, automatic music creation has long been considered a major milestone towards artificial general intelligence. In recent years, machine learning has been revolutionizing the state of the art in many fields, and **artificial intelligence will also be the critical technology in the future transformation of the music industry**. However, while many researchers have started to investigate applying deep learning to automatic music creation, the current technology is still far from being practically applicable in professional music production. In the aspect of automatic music composition, much prior work only aims at generating melodies, lead sheets (i.e., melodies and chords) or four-part chorales, but modern pop music usually consists of multiple instruments or tracks and possesses more complex structures and textures. As a result, existing music generation models cannot be directly applied to multi-instrument music composition. In the aspect of music performance synthesis, current musical instrument synthesizers cannot reach the same smoothness and flexibility as human performers, and most of them only provide a fixed playing style. In addition, they cannot adjust the timbre according to the expressive markings (e.g., dynamics, slurs and articulations) on a musical score. These together make it challenging for musicians to simulate how the full composition sounds, requiring them to adjust their arrangements only after entering the recording stage. In view of these challenges, **I aim to apply advanced deep learning techniques to develop a multi-instrument music composition and performance synthesis system that could assist music creators in their music creation process**. This two-year project will be divided into three stages: First, I will study current neural architectures and investigate how to equip them with musical domain knowledge for learning efficient music representations. Second, I will develop a multitrack music arrangement system consisting of several instrument models and an arrangement model, where each instrument model learns to compose music of one specific instrument, and the arrangement model learns to choose a proper set of instruments and coordinate the instrument models to generate the full score. Finally, I will develop a music performance synthesis system that combines a Transformer model and a generative adversarial network, where the former learns to process the musical notes and expression markings on a musical score, and the latter learns to synthesize them into audio. In addition, I will introduce a latent variable model to model the variety of playing styles. My preliminary results have shown promising and convincing results in automatic instrumentation for piano as well as performance synthesis for violin. Further, I have collected over one million musical scores and compiled the largest ever sheet music dataset, which will serve as the training data for my proposed deep learning models. **I envision my proposed system to be integrated into professional music production software, which could lower the barrier to entry for composition and facilitate the democratization of music production. Moreover, the proposed system could also provide content creators with low-cost royalty-free music and lower the production cost of personalized courses in music therapy and education.**

參、研究計畫內容

一、緣起

台灣自 1970 年代以來即為華語流行音樂界的發展中心，擁有豐厚的人才庫與完善的產業鏈。自早期校園民歌至新台語歌運動，到各種曲風蓬勃發展的今日，台灣始終位居華語流行音樂界的領導地位。最近十年以來，適逢中國大陸流行音樂產業的崛起，憑藉其巨大的市場優勢，台灣流行音樂產業遭遇前所未有的競爭。然而，近期許多新興台灣音樂創作者積極汲取全球各地曲風，發揮創意納入台灣元素，為台灣流行樂壇帶來許多火花，也展現出與傳統華語流行音樂截然不同的風味。

近年來，人工智慧正在改變各行各業的工作型態，深度學習為各行各業帶來許多革命性的轉變，廣泛應用於自動駕駛、虛擬助理、資料探勘、金融分析、醫學診斷等領域。歐美樂壇更可見許多科技圈與音樂人合作的案例，例如：知名製作人 Alex Da Kid 與 IBM 開發的人工智慧 Watson 共同創作了「Not Easy」單曲[1]；迷幻搖滾樂團 The Flaming Lips 與 Google Magenta 團隊合作，利用人工智慧將水果與氣球轉變為樂器，在 Google I/O 2019 進行了一場前所未有的互動式現場表演[2]（見圖一）；舞蹈家森山開次與 Yamaha 團隊合作，透過人工智慧以舞步控制鋼琴演奏，與柏林愛樂 Scharoun 室內樂團進行跨界演出[3]（見圖二）。

（來源：https://www.youtube.com/watch?v=HGWkQP9lVPw）

圖一　音樂家與科技公司合作，利用人工智慧將水果與氣球轉變為樂器

（來源：https://www.youtube.com/watch?v=tLFe2AzCodk）

圖二　舞蹈家與科技公司合作，透過人工智慧以舞步控制鋼琴演奏

[1] https://www.youtube.com/watch?v=U-e90ELRnnQ
[2] https://www.youtube.com/watch?v=HGWkQP9lVPw
[3] https://www.youtube.com/watch?v=tLFe2AzCodk

2020 年文化內容策進院的台灣文化內容產業調查報告 (文化內容策進院, 2020)指出：「未來 AI 工具的深化應用將改變音樂製作與發行作業，使得音樂製作與發行更為簡單與可負擔，這個發展趨勢將使得全球成千上萬的音樂創作者更能自行創作出高品質且專業程度高的作品」。在未來十年間，**人工智慧是未來音樂產業轉型的關鍵技術，將自創作、表演、教育、行銷等各方面為全球音樂產業帶來巨大的轉變**。2020 年，全球流行音樂產值達 216 億美元 (International Federation of the Phonographic Industry, 2021)，台灣流行音樂產值達 254.71 億元 (文化內容策進院, 2020)，是重要的藝文產業。有鑑於此，強化台灣音樂產業的科技實力有其必要性，以利開拓新的發展方向，接軌國際市場。

以下簡短介紹學界與業界在自動音樂創作與演奏的研究現況。

（一）學界研究現況

自動音樂作曲最早可追朔至 18 世紀風行的「音樂骰子遊戲」，作曲家預先創作上百小節的音樂，並由玩家擲骰子決定其演奏順序。其中，最有名的版本莫過於莫札特於 1792 年所創作的版本，可產生上兆種不同組合。1957 年，伊利諾大學香檳分校開發的 ILLIAC I 電腦所創作的 Illiac Suite，普遍被認為是第一首電腦創作之音樂作品。然而，這些電腦音樂創作系統仍需由音樂家預先編寫好規則與演算法。20 世紀末，隨著人工智慧的發展，自動音樂作曲有了新的研究方向。同時，因為音樂的高複雜度與主觀性，自動音樂作曲至今仍被視為邁向通用人工智慧的重要里程碑。

近期，有鑑於深度學習的高速發展，學界著手探討將深度學習應用於自動音樂創作之可能性 (Briot, Hadjeres, & Pachet, 2020)。在許多國際大型學術會議中，每年均可見自動音樂創作相關之論文，如 AAAI 人工智慧年會（AAAI）、國際人工智慧聯合會議（IJCAI）、國際機器學習會議（ICML）、神經信息處理系統大會（NeurIPS）、國際表徵學習會議（ICLR）、國際音樂資訊檢索會議（ISMIR）。國際間著名實驗室包含中研院由楊奕軒博士領導的音樂與人工智慧實驗室、美國羅徹斯特大學段志尧教授領導的音訊資訊研究實驗室（AIR Lab）、上海紐約大學由夏光宇教授領導的 Music X Lab、韓國科學技術院由 Juhan Nam 教授領導的音樂與音訊計算實驗室（MAC Lab）、新加坡科技設計大學由 Dorien Herremans 教授領導的音訊音樂與人工智慧實驗室（AMAAI Lab）、法國音樂與聲響研究中心由 Philippe Esling 教授領導的 ACIDS Lab。

過去多數研究僅探討旋律、爵士樂譜（旋律與和弦）、四聲部聖詠的生成，這些系統多數運用隱馬可夫模型（hidden Markov model，HMM）、受限玻茲曼機（restricted Boltzmann machine，RBM）、時間遞歸神經網路（recurrent neural network，RNN）等模型。在本計畫著

重的多樂器音樂生成系統方面，較具代表性的方法包含系統包含生成式對抗網路（generative adversarial network，GAN）、Transformer 神經網路、變分自編碼器（variational autoencoder，VAE）。其中，較為著名的系統包含 MuseGAN (Dong H.-W. , Hsiao, Yang, & Yang, 2018)（將於後文詳述）、BinaryMuseGAN (Dong & Yang, 2018)、LakhNES (Donahue, Mao, Li, Cottrell, & McAuley, 2019)、MMM (Ens & Pasquier, 2020)。MuseGAN 與 BinaryMuseGAN 系統利用生成式對抗網路生成以多軌鋼琴捲格式表示的音樂資料。LakhNES 與 MMM 系統利用 Transformer 神經網路生成以類似於 MIDI（音樂數位介面）之格式表示的音樂資料。

（二）業界研究現況

許多大型企業與新創公司對自動音樂創作與演奏系統皆有所涉獵，以下逐一簡介。

- **Sony CSL**：Sony Computer Science Laboratories（CSL）為 Sony 在日本東京與法國巴黎的跨國實驗室。該團隊開發的 Flow Machine 系統，從大數據中學習各種音樂風格，與音樂家合作共同以披頭四樂團的風格譜寫出「Daddy's Car」單曲[4]。近期，Flow Machine 以協助音樂家創作為目標，釋出多項數位音樂工作站的插件。該團隊也開發了可以創作出類似巴哈四聲部聖詠的 DeepBach 系統，其作品通過圖靈測試，即一般大眾無法成功辨別其創作與真實巴哈聖詠[5]。

- **Google Magenta**：Google Magenta 為 Google Brain 的其中一個團隊，開發了多項自動音樂創作與演奏系統，包含 MelodyRNN（旋律生成）、Performance RNN（演奏模擬）、CocoNet（聖詠創作）、MusicVAE（旋律生成）、Music Transformer（音樂生成）、Wave2MIDI2Wave（音訊合成）、GANSynth（音訊合成）、DDSP（音訊合成）等模型。該團隊曾與迷幻搖滾樂團 The Flaming Lips 及流行舞曲樂團 YACHT 合作。

- **JukeDeck**：JukeDeck 為 2012 年於英國倫敦成立之新創公司，致力於應用人工智慧自動創作音樂，其在 2019 年被字節跳動（ByteDance）收購。其系統可接受使用者輸入的標籤及情境產生指定長度之音樂，使用者可藉此避免高昂的音樂素材成本及授權問題。

- **AIVA**：AIVA 為 2016 年於盧森堡成立之新創公司，其開發之 AIVA 為第一個在法國音樂作者作曲家製作人協會註冊為創作者的人工智慧。AIVA 自大數據中學習古典音樂的規則與風格，在 2016、2018 年推出「Genesis」及「Among the Stars」二張古典音樂專輯，並於 2017 年 Nvidia GPU 年度技術大會上推出單曲「I am AI」[6]作為其開場音樂。

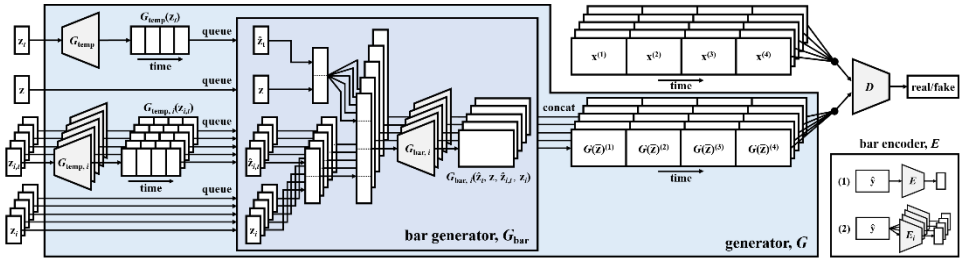
[4] https://www.youtube.com/watch?v=LSHZ_b05W7o
[5] https://www.youtube.com/watch?v=QiBM7-5hA6o
[6] https://www.youtube.com/watch?v=Emidxpkyk6o

- **OpenAI**：OpenAI 為 2015 年成立於美國舊金山之非營利人工智慧研究單位。其團隊所開發之 MuseNet（多軌音樂生成）、JukeBox（音訊生成）模型，利用深度學習模型自數十萬、數百萬首音樂中學習，可以依據使用者之需求，產生各種風格的音樂。

（三）本人在自動音樂創作之研究

本人在自動音樂創作方面已發表數篇論文，以下簡介其中二項研究。

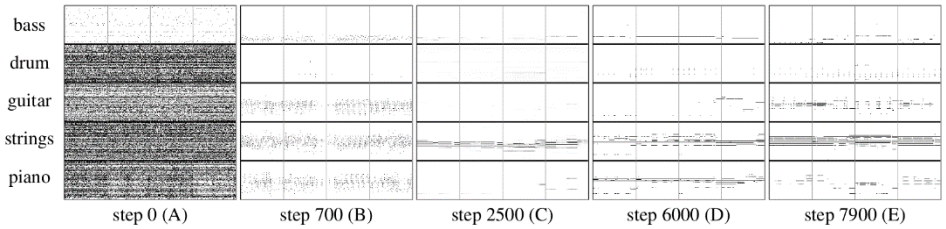
- **MuseGAN**：MuseGAN (Dong H.-W. , Hsiao, Yang, & Yang, 2018)提出創新的多軌序列式生成式對抗網路，為第一個能產生多樂器複音音樂的深度學習模型（見圖三）。有別於大多數文獻中使用之時間遞歸神經網路，此系統使用高度平行化的卷積神經網路（convolutional neural network，CNN），大幅提高神經網路的運算效率。



（來源： (Dong H.-W. , Hsiao, Yang, & Yang, 2018)）

圖三　MuseGAN 之系統架構圖

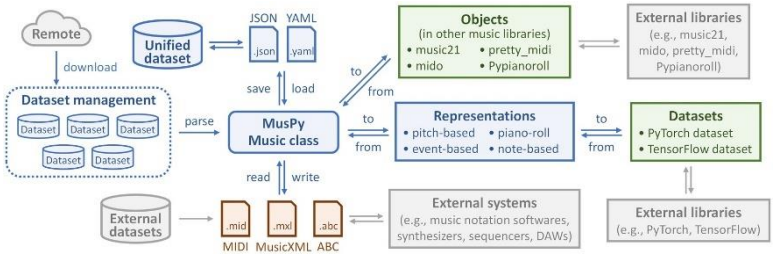
為了訓練 MuseGAN，我們收集了上百萬小節的流行音樂資料，並將其樂器粗略歸類為鋼琴、吉他、貝斯、鼓、弦樂等五項最常見之樂器類別。此系統之訓練目標即為生成上述五項樂器所構成的音樂片段（見圖四）。同時，我們參考常見的音樂創作型態，從而設計出相應之機器學習模型，包含即興樂團模型（jamming model）、作曲家模型（composer model）、混合模型（hybrid model）。除了自動創作模式之外，此系統也支援自動伴奏模式，可以接受使用者輸入其中一個樂器的音樂，並據此生成其他四個樂器的音樂伴奏，可應用於人機互動創作的情境。



（來源： (Dong H.-W. , Hsiao, Yang, & Yang, 2018)）

圖四　MuseGAN 在不同訓練階段所生成的多樂器音樂範例

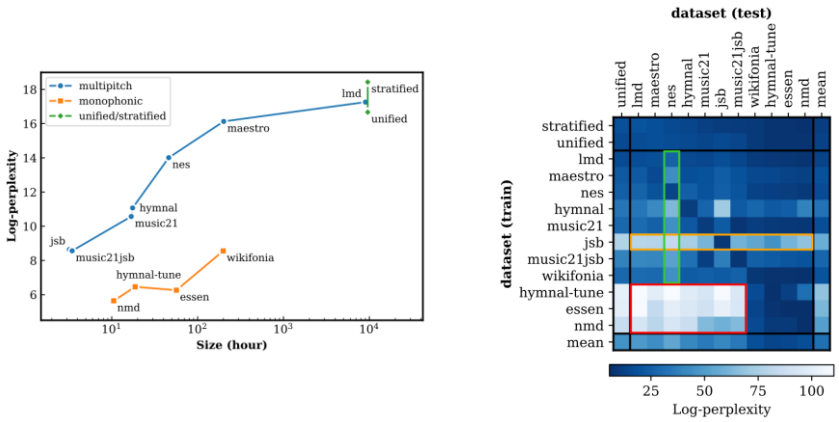
- **MusPy**：MusPy (Dong, Chen, McAuley, & Berg-Kirkpatrick, 2020)是一個開放原始碼的 Python 套件，專門為自動音樂創作系統開發者設計。其內建資料庫管理系統，開發者可以簡單下載常用的音樂資料庫。MusPy 支援各種常見音樂資料格式（如 MIDI、MusicXML、ABC 等），並提供與其他常用 Python 音樂套件（如 music21、mido、pretty_midi、Pypianoroll 等）之介面。同時，我們實現多種常見的音樂表示方法，以利開發者快速獲得適用於不同模型的輸入格式。MusPy 亦提供視覺化、音訊合成、評價指標等工具，以利開發者評估並優化其模型。其系統架構圖見圖五。



（來源： (Dong, Chen, McAuley, & Berg-Kirkpatrick, 2020)）

圖五　MusPy 之系統架構圖

受益於 MusPy 的強大功能，我們得以進行數項過去難以執行的實驗。首先，我們在 MusPy 支援的十一個資料庫上訓練一個長短期記憶神經網路（long short-term memory，LSTM），藉由觀察深度學習模型在不同資料庫上的表現，我們得以了解不同資料集的複雜度與多元度（見圖六左）。此外，我們測試了深度學習模型跨資料庫的可推論性，發現在特定資料庫上訓練出的模型具有較差的推廣能力（見圖六右）。這些結果可作為未來研究者選擇資料庫的參考，避免代表性或多元性不足的資料庫。



（來源： (Dong, Chen, McAuley, & Berg-Kirkpatrick, 2020)）

圖六　深度學習模型在（左）不同資料庫上的困惑度（perplexity）及（右）跨資料庫的可推論性（generalizability）

二、目的

本計畫旨在應用前瞻深度學習技術於多樂器音樂作曲與演奏合成，並藉此探索人類與人工智慧共同創作音樂之可能性。我提出以下三個主要的研究課題：

- 如何利用深度學習模型從音樂資料中擷取高效率的音樂特徵？
- 如何利用深度學習模型學習多樂器音樂中不同樂器間的相互依賴關係？
- 如何利用深度學習模型掌握樂譜上的演奏記號與不同演奏風格？

根據這三項研究課題，我將本研究計畫分為以下三個階段：

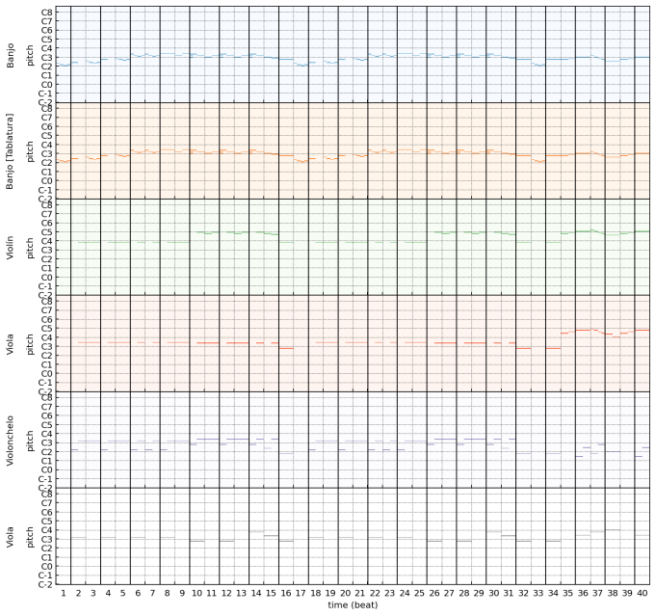
1. **音樂深度學習模型開發：**研究當前神經網路架構並探討如何將音樂知識植入其中，以利神經網路學習高效率的音樂特徵。
2. **多樂器音樂作曲系統開發：**此系統將由數個基於 Transformer 神經網路的樂器模型及一個編曲模型組成，每個樂器模型學習編寫特定樂器的音樂，而編曲模型學習選擇適當的樂器並協調各個樂器模型，以產生完整的多樂器音樂。
3. **音樂演奏合成系統開發：**此系統將結合 Transformer 神經網路與生成式對抗網路，前者學習處理樂譜上的音符與符號，而後者學習將其合成為音訊，同時我將導入隱變數模型以模擬多樣的演奏風格。

三、方法

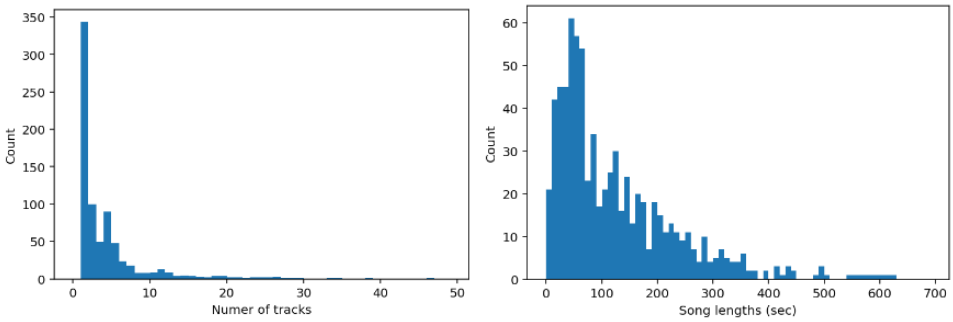
（一）音樂深度學習模型開發

- **蒐集訓練資料**

為了訓練並比較不同深度學習模型在音樂生成上的效果，我自 MuseScore 論壇上蒐集訓練大量資料。該論壇為當前最大的樂譜分享社群，供使用者分享原創及改編樂譜，擁有超過百萬首樂曲的樂譜，且曲風多樣、配器多元。圖七展示其中一個多樂器樂譜的開頭片段，此視覺化由我開發的 MusPy (Dong, Chen, McAuley, & Berg-Kirkpatrick, 2020)及 Pypianoroll (Dong, Hsiao, & Yang, 2018)套件產生。圖八則展示 MuseScore 音樂資料庫之音軌數及總時長的統計數據，可以發現此資料集相當多元。



圖七　MuseScore 音樂資料庫之範例多樂器音樂片段，以多軌鋼琴捲格式展示



（隨機選取一萬首之統計數據）

圖八　MuseScore 音樂資料庫之（左）音軌數及（右）總時長的統計數據

- **比較現有音樂深度學習模型並開發創新音樂深度學習模型**

我將比較在 MultitrackVAE (Simon, et al., 2018)、MuseNet (Payne, 2019)、MMM (Ens & Pasquier, 2020)中使用的多樂器音樂表示方法。透過在 MuseScore 音樂資料庫上使用不同音樂表示方法訓練一個 Transformer 神經網路，我可以了解這些多樂器音樂表示方法的優缺點，從而改善並開發新的多樂器音樂表示方法。同時，我將研究 Transformer 神經網路的不同延伸模型，從中找出或開發最適合音樂資料的深度學習模型。

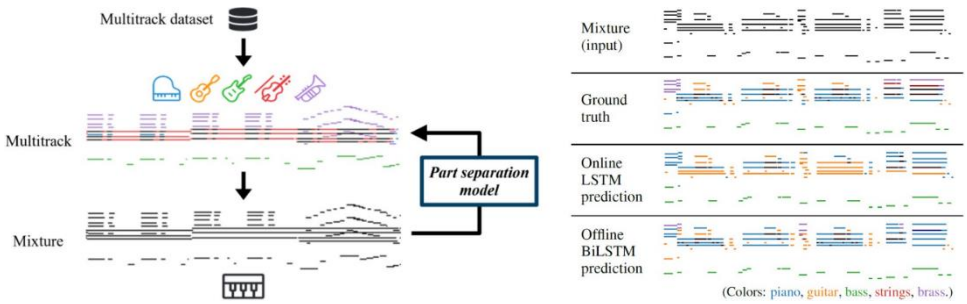
（二）多樂器音樂作曲系統開發

- **蒐集訓練資料與前處理**

我預計主要使用 MuseScore 音樂資料庫，並以其他音樂資料庫做為輔助，以增加較罕見樂器的訓練資料量，相關音樂資料庫包含 Lakh MIDI 資料集 (Raffel, 2016)及 MetaMIDI 資料集 (Ens & Pasquier, 2021)。

- **初步結果—鋼琴音樂自動配器**

我在鋼琴自動配器研究上已獲的初步的結果，我開發的 Arranger (Dong, Donahue, Berg-Kirkpatrick, & McAuley, 2021)系統利用深度學習模型學習自無樂器資訊之樂譜中分離聲部的任務，達到自動配器的應用。我們假想的使用情境有二：一、實現鍵盤手在鍵盤上同時演出多種樂器的應用；二、輔助編曲家進行配器。由於獨奏與配器後的樂譜資料難以取得，我們將多樂器音樂的樂器資訊去除以獲得獨奏樂譜，藉此獲得成對的訓練資料（見圖九左）。我考慮四個不同的音樂資料庫：巴哈四部聖詠、弦樂四重奏、遊戲配樂、流行音樂。我將此問題視為序列多類別分類問題，並比較二種模型：長短期記憶神經網路及 Transformer 神經網路。在四個資料庫上，我提出的模型所產生的配器結果均顯著超越既有模型的結果（見圖九右）。這些研究結果發表於 2021 年的國際音樂資訊檢索會議（ISMIR）。

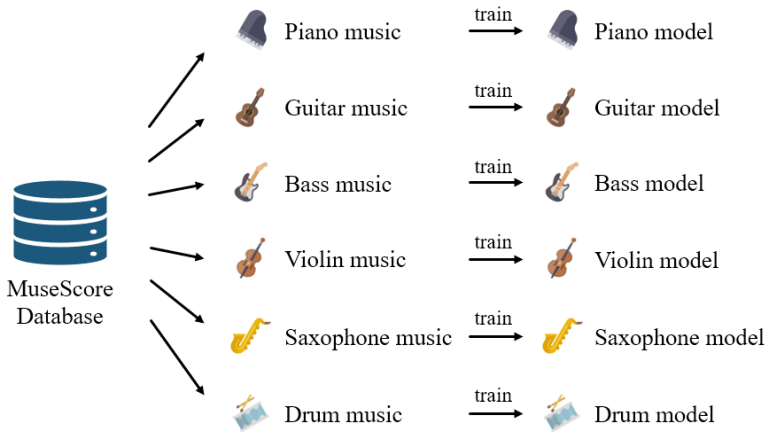


（來源： (Dong, Donahue, Berg-Kirkpatrick, & McAuley, 2021)）

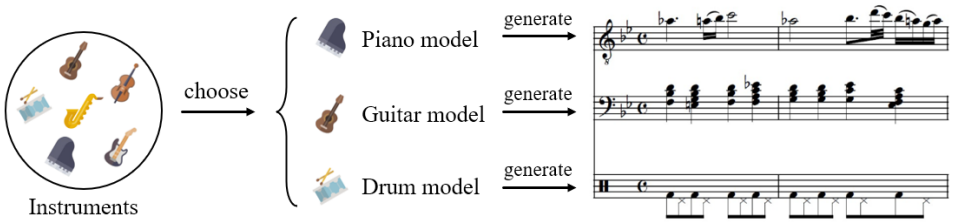
圖九　Arranger 之系統架構圖（左）及在流行音樂資料庫的配器範例（右）

- **開發多樂器作曲系統**

我預計利用 MuseScore 資料庫訓練數個樂器模型，這些樂器模型由 Transformer 神經網路組成，其將學習編寫特定樂器的音樂（見圖十）。接著，我將訓練一個同樣由 Transformer 神經網路組成的編曲模型，其將學習選擇適當的樂器並協調各個樂器模型，以產生完整的多樂器音樂（見圖十一）。基於變分自編碼器的架構，將可透過重構目標函數（reconstruction loss）的方式訓練此編曲模型。由於變分自編碼器為一隱變量模型，此系統將可提供使用者操控隱變量的數值，從而調整系統輸出的結果。



圖十　利用多樂器音樂資料庫訓練數個樂器模型



圖十一　編曲模型選擇適當配器並協調不同樂器模型以生成多樂器音樂

（三）音樂演奏合成系統開發

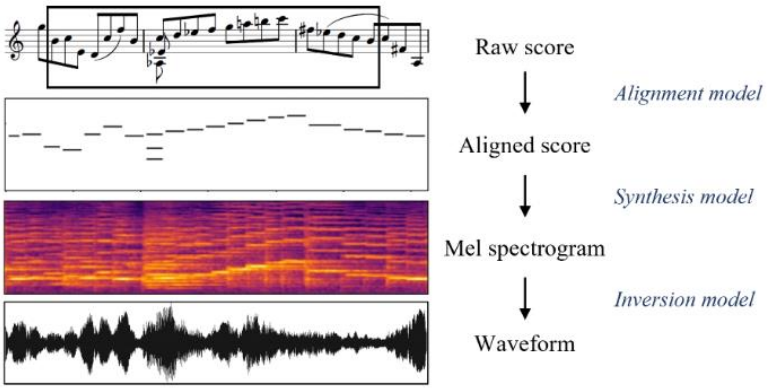
- **蒐集訓練資料與前處理**

現有用於訓練音樂演奏合成系統的資料集多數僅收錄鋼琴音樂，包含許多文獻中使用的 MAESTRO 鋼琴資料庫 (Hawthorne, et al., 2019)。然而，多數樂器擁有較鋼琴更為複雜的發聲原理並具有許多特殊演奏技巧。有鑑於此，我收集巴哈小提琴資料集，其提供超過六小時的巴哈小提琴練習曲與奏鳴曲錄音。同時，並利用動態時間校正（dynamical time warping，DTW）將錄音與樂譜對齊，獲得能夠用於訓練音樂演奏合成系統的對齊配對資料。未來，我將仿照此模式收集其他樂器的資料集，同時我也會

利用 URMP 資料庫 (Li, Liu, Dinesh, Duan, & Sharma, 2018)來訓練特定樂器的演奏合成模型。

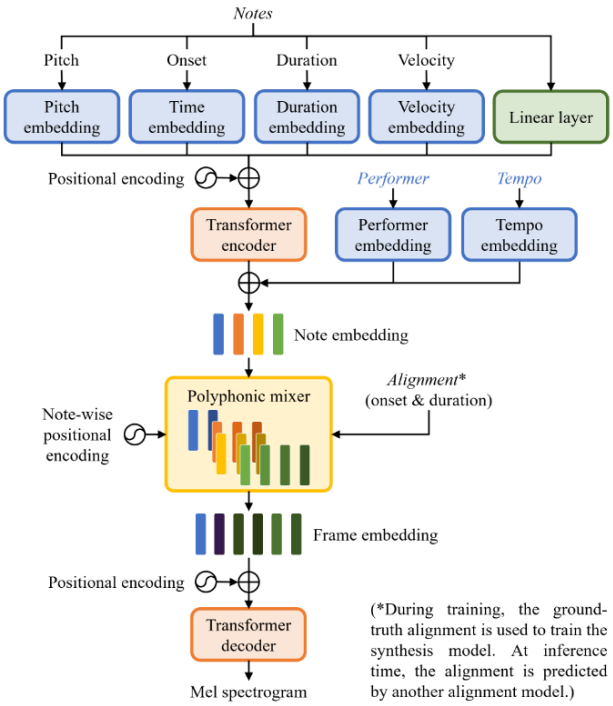
- **初步結果—小提琴與鋼琴演奏合成**

我在小提琴與鋼琴演奏合成的研究已獲的初步的結果，我開發的 Deep Performer (Dong, Zhou, Berg-Kirkpatrick, & McAuley, 2022)系統結合 Transformer 神經網路與生成式對抗網路，提出一個創新的合成器模型。此模型分為三個階段（見圖十二）：首先，對齊模型（alignment model）利用 Transformer 神經網路預測輸入樂譜的動態速度，使其與目標輸出音訊對齊；接著，合成模型（synthesis model）使用 Transformer 神經網路將對齊的樂譜合成為梅爾時頻譜（見圖十三）；最後，反向模型（inversion model）使用生成式對抗網路將合成的梅爾時頻譜逆向轉換為音訊波形。在巴哈小提琴資料庫上，此系統與現有模型達到同等水準。在 MAESTRO 鋼琴資料庫上，此系統在音準、音色、訊噪比等項目皆大幅勝過既有之模型。這些研究結果將發表於 2022 年的國際聲學、語音和信號處理會議（ICASSP）。



（來源： (Dong, Zhou, Berg-Kirkpatrick, & McAuley, 2022)）

圖十二　DeepPerformer 之系統架構圖

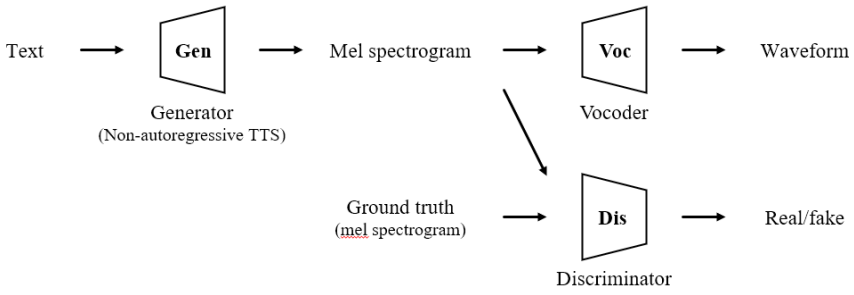


（來源：(Dong, Zhou, Berg-Kirkpatrick, & McAuley, 2022)）

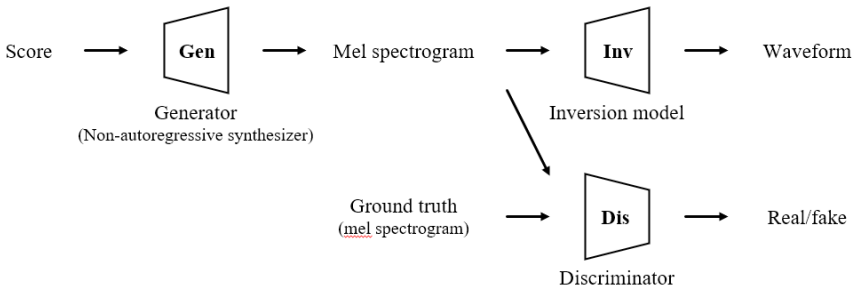
圖十三　基於 Transformer 神經網路之合成模型

- **<u>開發音樂演奏合成系統</u>**

我預計參考 GANSpeech 語音合成系統 (Yang, Bae, Bak, Kim, & Cho, 2021)提出的生成式對抗網路架構（見圖十四），作為音樂演奏合成系統的雛型。GANSpeech 語音合成系統透過引進對抗目標函數（adversarial loss），得以顯著提升合成音訊的銳利度。有鑑於文字語音合成（text-to-speech synthesis）與音樂演奏合成的諸多相似處，我預期此系統將也能在音樂演奏合成上獲得良好的效果。此音樂演湊合成系統將結合 Transformer 神經網路與生成式對抗網路，前者學習處理樂譜上的音符與表現符號，而後者學習將其合成為音訊（見圖十五）。同時，我將在生程式對抗網路的架構下，導入隱變數（latent variable）以模擬多樣的演奏風格。透過操控此隱變量的數值，使用者將可以調整系統輸出的演奏風格，從而提升系統的可操控性。



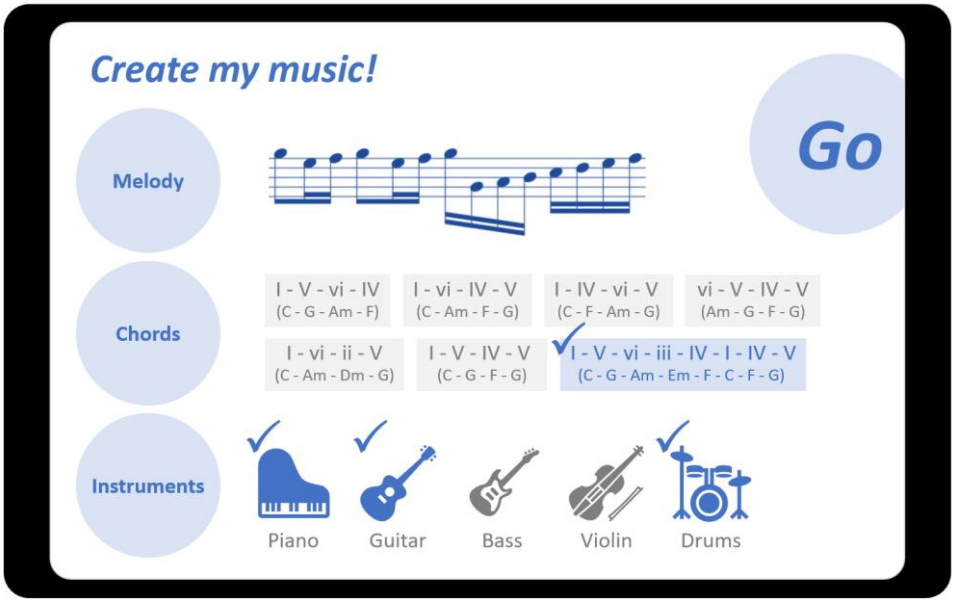
圖十四　GANSpeech 語音合成系統之示意圖



圖十五　基於 GANSpeech 語音合成系統之音樂演奏合成系統

（四）系統整合

最後，我將整合以上所述之多樂器作曲系統及音樂演奏合成系統，建立一個使用者可以與人工智慧互動的軟體。使用者不僅可以操作本計畫開發的多樂器作曲系統，選擇特定的樂器組合與和弦進行，也可以提供一段其創作的旋律，利用此系統為其伴奏。同時，透過本計畫開發的音樂演奏合成系統，使用者將可及時合成音訊，並根據其聲響效果進行調整，進而達成人類與人工智慧的共同創作。為了建立最佳的使用者介面，我預計進行簡單的使用者體驗分析，藉由本系統探索人類與人工智慧各種可能的互動模式。下圖為我初步構想之使用者介面。



圖十六　初步構想之使用者介面

四、期程

以下為本計畫之工作期程，為期二年。根據此期程，此計畫將能發表至少四篇論文，並完成一篇博士論文。此外，本研究之產出包含一套線上音樂創作系統。

工作項目	季度	第一年				第二年			
		第一季	第二季	第三季	第四季	第五季	第六季	第七季	第八季
音樂深度學習模型開發	蒐集訓練資料與前處理	◎							
	比較現有音樂深度學習模型	◎							
	開發創新音樂深度學習模型	◎							
	進行實驗並優化模型	◎	◎						
	撰寫及投稿論文		◎						
多樂器音樂作曲系統開發	資料蒐集與前處理	◎	◎	◎					
	研究多樂器音樂資料表示方法		◎	◎					
	開發多樂器音樂作曲系統			◎					
	進行實驗並優化模型			◎	◎				
	撰寫及投稿論文				◎				
音樂演奏合成系統開發	資料蒐集與前處理	◎	◎	◎	◎	◎			
	重現 GANSpeech 語音合成系統					◎	◎		
	開發音樂演奏合成模型					◎	◎		
	進行實驗並優化模型					◎	◎		
	撰寫及投稿論文						◎		
系統整合	開發互動使用介面			◎	◎	◎	◎	◎	
	架設展示網站							◎	
	進行使用者體驗分析並優化介面							◎	
	撰寫及投稿論文							◎	
	撰寫博士論文							◎	◎

五、重要性

誠如「一、緣起」所述，在未來十年間，人工智慧將會是未來音樂產業轉型的關鍵技術，將自創作、表演、教育、行銷等各方面為全球音樂產業帶來巨大的轉變。本計畫切合音樂產業實際需求，著重開發多樂器音樂作曲與演奏系統，將可降低現有技術與實際專業音樂製作上之差距。對於提升台灣的音樂產業技術能力，有其必要性。本計畫亦有以下幾項潛在影響：

（一）降低音樂創作的技術門檻

由於創作音樂需具備一定的音樂理論知識，並且通常需要學習彈奏不同樂器，因此時至今日，音樂創作仍僅限於專業人士，一般人入門往往需面臨相當高的技術門檻。本計畫提出的多樂器音樂作曲與演奏系統透過簡單的操作介面與使用者互動，藉此將可降低音樂創作的技術門檻，同時可進而促進音樂創作的普及化，使普羅大眾都能創作自己的音樂。

（二）降低音樂素材的生產成本

內容創作者（如 YouTuber、TikToker、Twitch 直播主等）在創作內容時，經常需要使用各種背景音樂、音樂特效、音樂素材以創造特定效果。本計畫提出的自動音樂創作系統可以用於大量創作音樂素材，並以免權利金（royalty-free）的方式提供內容創作者使用，藉此將可避免內容創作者使用有版權的音樂素材所需付出的高額權利金，同時也可以避免因版權問題所導致的潛在法律訴訟成本。

（三）應用於音樂治療與教育之情境

本計畫除了可以應用於協助音樂人創作以外，亦能應用於音樂治療（music therapy）與音樂教育上。在音樂治療與教育的應用中，往往需要為個人量身打造課程，需要耗費相當高的成本。本計畫提出的自動音樂創作系統不只可以降低其製作成本，亦可以進一步發展為互動式的音樂治療或教育軟體。

（四）探索人類與人工智慧的關係

在人工智慧被大量運用於生活中的今日，人類與人工智慧間的關係是未來須要嚴謹面對的課題。本計畫透過人類與自動音樂創作系統的互動，可以探討人類與人工智慧共同創作音樂的可能性，將可為人工智慧於其他領域的應用帶來不同的討論。

六、相關文獻

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2020). *Deep Learning Techniques for Music Generation.* Springer.

Donahue, C., Mao, H. H., Li, Y. E., Cottrell, G. W., & McAuley, J. (2019). LakhNES: Improving multi-instrumental music generation with cross-domain pre-training. *Proc. ISMIR.*

Dong, H.-W., & Yang, Y.-H. (2018). Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation. *Proc. ISMIR.*

Dong, H.-W., Chen, K., McAuley, J., & Berg-Kirkpatrick, T. (2020). MusPy: A Toolkit for Symbolic Music Generation. *Proc. ISMIR.*

Dong, H.-W., Donahue, C., Berg-Kirkpatrick, T., & McAuley, J. (2021). Towards Automatic Instrumentation by Learning to Separate Parts in Symbolic Multitrack Music. *Proc. ISMIR.*

Dong, H.-W., Hsiao, W.-Y., & Yang, Y.-H. (2018). Pypianoroll: Open Source Python Package for Handling Multitrack Pianorolls. *ISMIR Late-Breaking Demos.*

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., & Yang, Y.-H. (2018). MuseGAN: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. *Proc. AAAI.*

Dong, H.-W., Zhou, C., Berg-Kirkpatrick, T., & McAuley, J. (2022). Deep Performer: Score-to-Audio Music Performance Synthesis. *Proc. ICASSP.*

Ens, J., & Pasquier, P. (2020). *MMM: Exploring Conditional Multi-Track Music Generation with the Transformer.* arXiv preprint arXiv:2008.06048.

Ens, J., & Pasquier, P. (2021). Building the MetaMIDI Dataset: Linking Symbolic and Audio Musical Data. *Proc. ISMIR.*

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., . . . Eck, D. (2019). Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. *Proc. ICLR.*

International Federation of the Phonographic Industry. (2021). *Global Music Report – State of the Industry.*

Li, B., Liu, X., Dinesh, K., Duan, Z., & Sharma, G. (2018). Creating A Multi-track Classical Music Performance Dataset for Multi-modal Music Analysis: Challenges, Insights, and Applications. *Proc. MM.*

Payne, C. (2019, 4 25). MuseNet. *OpenAI*. Retrieved from https://openai.com/blog/musenet/

Raffel, C. (2016). Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching. *PhD Thesis*.

Simon, I., Roberts, A., Raffel, C., Engel, J., Hawthorne, C., & Eck, D. (2018). Learning a Latent Space of Multitrack Measures. *NeurIPS Workshop on Machine Learning for Creativity and Design.*

Yang, J., Bae, J.-S., Bak, T., Kim, Y., & Cho, H.-Y. (2021). GANSpeech: Adversarial Training for High-Fidelity Multi-Speaker Speech Synthesis. *Proc. INTERSPEECH.*

文化內容策進院. (2020). 2020 年台灣文化內容產業調查報告 III：流行音樂產業.

肆、預期完成之工作及具體成果

我於 2017 年取得國立臺灣大學電機工程學系學士學位，隨後於中央研究院資訊與科技創新中心擔任將近二年的研究助理，跟隨楊奕軒博士進行音樂生成的研究，並先後發表了二篇論文。隨後至加州大學聖地牙哥分校攻讀博士學位，第一年我獲得電機系提供的博士班第一年全額獎學金，並在第二年轉入電腦科學系跟隨 Julian McAuley 教授與 Taylor Berg-Kirkpatrick 教授進行研究，陸續發表四篇論文（包含三篇第一作者論文）。我目前就讀博士班三年級，已修完所有畢業所需課程並通過博士候選人資格考。今年我獲得了楊信家族基金會所頒發的楊信獎學金（2021 秋季至 2022 春季），僅十位在加州大學聖地牙哥分校就讀的台灣研究生（共約二百人）獲選。現階段，我正跟隨我的二位指導教授研究專門處理音樂資料的深度學習模型，預計於今年夏天前投稿。**依據我的博士論文規劃，在畢業前應能再發表四至六篇論文。**

本計畫之二年期程（詳見「四、期程」）預計能就「音樂深度學習模型」、「多樂器音樂作曲」、「音樂演奏合成」三個研究方向分別發表一至二篇論文，預計共能發表三至五篇論文。同時，所有研究的內容皆會秉持開放原始碼的精神，全數公開釋出給大眾，供所有程式開發者使用。此外，我將架設展示網站供所有人嘗試本計畫開發的線上智慧音樂創作系統。**這些研究將能填補現有技術與實際音樂產業需求間的差距，使學界與業界的科研方向更加切合，共同帶動未來音樂產業的轉型。**

在這二年新冠肺炎疫情的催化下，人工智慧的發展方興未艾，深度學習的應用席捲各大產業。在未來數年間，人工智慧也將自各方面為音樂產業帶來革命性的轉變，而台灣身居華語流行音樂的領導地位，在人工智慧音樂新時代到來之際不容缺席。**我將致力探索人工智慧於音樂創作的應用，此份獎學金將有助於我全心投入我的博士論文研究，向國際展現台灣學子的競爭力，並將當前最前沿的人工智慧音樂科技帶回台灣。**