Sponsoring Committee: Professor Juan Pablo Bello, Chairperson
Professor Robert Rowe
Doctor Eric J. Humphrey

AUTOMATIC MUSIC TRANSCRIPTION IN THE DEEP LEARNING ERA:

PERSPECTIVES ON GENERATIVE NEURAL NETWORKS

Jong Wook Kim

Program in Music Technology
Department of Music and Performing Arts Professions

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
Steinhardt School of Culture, Education, and Human Development
New York University
2019

ABSTRACT

The problem of automatic music transcription (AMT) is considered by many researchers as the holy grail of the field, because of the notorious complexity and difficulty of the problem. Meanwhile, the current decade has seen an unprecedented surge of deep learning where neural network methods have achieved tremendous success in many machine learning tasks including AMT. The success of deep learning is largely enabled by the ever-increasing amount of available data and the innovation of GPU hardware, allowing a deep learning model to enjoy the increased capacity to process such scale of data. While having more data and higher capacity translates better performance in general, there still remains the question of how to design an AMT model that can effectively incorporate the inductive bias for the task and best utilize the increased capacity.

This thesis hypothesizes that an effective way to address this question is through the use of generative neural networks. Starting with a simplified setup of monophonic transcription, we learn the effectiveness of convolutional representation and the roles of dataset choices in data-driven models for music analysis. In the subsequent chapters, we examine the applications of deep generative models in music analysis and synthesis tasks, by introducing a WaveNet-based music synthesis model that learns a multi-dimensional timbre representation and a music language model applied in an adversarial manner to improve a piano transcription model. Finally, we combine the analysis and synthesis methods to develop a multi-instrument polyphonic music transcription system. From these observations, we conclude that deep generative models can be used to improve AMT in many ways, and they will be a crucial component for further advancing AMT.

To Nayoung.

ACKNOWLEDGEMENTS

I am truly grateful to many wonderful people who believed in me along this long journey. This dissertation would not have been able to come into existence without their support and guidance.

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Juan Pablo Bello. I am incredibly lucky to have you as my doctoral advisor; thank you for being a dependable teacher, an incredible researcher, and a welcoming friend to me. To my committee members and readers — Prof. Robert Rowe, Dr. Eric Humphrey, Prof. Johanna Devaney, and Prof. Brian McFee — I deeply appreciate taking your time to read through my awkward sentences and giving insights to make them better. And to all professors with whom I have had the pleasure of working with: thank you for your classes, chats, and smiles.

I have been fortunate enough to become friends with almost all of MARL's PhD students and postdocs, and I am thankful to every one of them. To the MARL-doctors — Taemin, Areti, Jon, Aron, Braxton, Finn, Eric, Uri, Rachel, and Finn — thanks for showing the ways that I can follow, including but not limited to the coffee shops and bars. To the MARL-doctor-to-be's — Andrea, Andrew, Marta, Peter, Yu, Ho-Hsiang, Willie, Dirk, Tom, Jason, and Chris — it was so much fun sharing office and hanging out with you, and I will miss the occasional beer sessions. To the MARL postdocs — Brian, Justin, Mark, Charlie, Ron, Vincent, Claire, Magdalena, Hitomi — thank you for the insights during the lab meetings and collaborations, and also for sometimes bearing with my unscholarliness.

And to my Korean friends: thanks for being on KakaoTalk whenever I had silly memes to share with, but more importantly for being always welcoming and cheering for me every time I visited Korea, especially that time when you guys gladly spent a whole day at my wedding.

I am honored to have been a recipient of Samsung Scholarship, which is another factor that made all this possible; thank you for your support, networking, and gifts from Leeum. I would also like to thank everyone I worked with during my brief industry experiences — NCSOFT, Kakao, Pandora, and Spotify — for allowing me to learn and achieve what I could not do in schools, and of course for all the free foods and swags.

To my parents who have always believed me and prayed for me, I can't express enough gratitude for your limitless love and unwavering support. Your passion in education made me grow from an aspiring teenager to a respectable scientist. Now that there are no more degrees for you to worry about, please take it easy and enjoy your 60s!

Finally, to my wife Nayoung, you are the foremost reason why I am writing these words now. I cannot believe how lucky I am to have met you and plowed through this journey mixed with joys and tears with you. Thank you for being my best friend, a fellow researcher, a delightful travel partner, a world-class cook, a witty comedian, and a lovely cheerleader who makes me a better person every day.

TABLE OF CONTENTS