Statement of Purpose

Julien Guinot

Music Technology PhD applicant - NYU Steinhardt - Fall 2023

Interested in researching under:

Dr. Juan Pablo Bello (jpbello@nyu.edu), Dr. Brian Mcfee (brian.mcfee@nyu.edu)

Music has long been a focal point of my life and has captivated me with its technical and subjective nature. It has been my main driver both personally and professionally. Recently, this passion has crystallized into a decision to pursue research at the crossroads of artificial intelligence and music. I believe that technological advances in musical AI have had a profound beneficial influence on the creative composition process and the understanding of the deep-rooted cross-cultural phenomenon that music represents. I aim to deepen my knowledge of musical AI and contribute to state-of-the-art research in high-level academia or industry as a professional music ML researcher. Thus my motivation to build upon my master's degree and thesis experience in music information retrieval by undertaking a doctoral degree in music technology at NYU Steinhardt.

Beginning in high school, I taught myself acoustic and electric guitar, keys, and electric bass. I learned to produce, mix and master. Over the years that followed, I added music theory and singing technique to my repertoire. From my third year of undergraduate as a Bachelor of Engineering student in France to my final year of postgraduate studies, I dedicated myself to several musical projects, including four years as a mix-master engineer, vocalist, and vocal coach for a student-produced musical. I presided over the school's DJing student society. Music has, to say the least, played a central role in my life in recent years. It seems evident to me that music should be directly associated with my career, which is why I elected to specialize in acoustics during my engineering studies. Through this, I aimed to get as close as possible to music within a technical curriculum. For the second year of my MS (Engineering), I was accepted for a dual degree in Acoustics Engineering at the University of Adelaide.

I have COVID to thank for my interest in musical AI. Constrained to take a gap year to postpone the start of my degree because of Australian borders closing, I explored a novel field to me: Machine Learning. At this point, my outlook opened up - I *knew* I wanted to combine the technical creativity of ML with the artistic creativity of music in my work. While once I had been discouraged by the thought that I had lost interest in research in acoustics, I am now thrilled to have found my field: I am fascinated by the vast amount of unexplored alleys at the crossroads of music and AI. The questions underpinning my areas of research interest resonate strongly with me as a musician and a scholar: How can technology help the creative musical process? How can AI provide deeper insights into music? These are the questions that drive me and my research.

I remember my father - who pursued a Ph.D. in chemistry - describing the moment he observed a yet unseen phenomenon and realized he *knew* how to explain it. For an hour, he was the sole possessor of a piece of knowledge that no one else had ever had. This singular moment of knowing one has expanded the corpus of human knowledge in a field I am passionate about underpins my motivation for research. I am curious by nature, and this desire to learn has driven me to go above and beyond in past research projects. To be honest, I am in awe of the advances current Music and AI studies are making in key research areas. The admiration I hold for these researchers compels me to conduct research at the level of innovation and standards held at NYU.

By joining the Music Technology doctoral program at NYU Steinhardt, I hope to develop my research experience and explore topical and fascinating aspects of Music in AI in a research program that routinely outputs top-quality research. I am confident that pursuing this doctorate will prove an invaluable step toward my goal of conducting cutting-edge research on ML applied to music in industry or academia. In my master's thesis - Music Automatic Tagging Towards Better Musical Recommendations - I focused on Music Information Retrieval. That said, as an avid learner and an interdisciplinary musician, my research interests span a wide range of topics, many of which align perfectly with the research produced at NYU Steinhardt. I am highly interested in Automatic Music Transcription, (Polyphonic) F0 estimation, generative methods for audio and music, audio source separation, and differential digital signal processing. In addition, I would like to incorporate into future research unsupervised methods and research on meta-learning such as contrastive learning, active learning, and few-shot generalization.

Among the researchers at the Music and Audio Research Lab, I am highly interested in working under the supervision of Dr. Juan Pablo Bello. I have followed his work closely and took the opportunity to correspond with him before submitting this application. His continued work on automatic music and drum transcription [10, 1] paves the way for studies on polyphonic sources. More to the point, His work on few-shot source separation, vocal ensemble F_0 estimation [3, 10] and multi-modal language-music learning, [12] is directly related to two research projects I would like to pursue: prompt conditioned or style-informed musical sample generation, and dual transcription/separation of backing vocal ensembles. I am also interested in working under the supervision of Dr. Brian McFee. As his recent work frequently overlaps with Dr. Bello's, it aligns with my research interests. Moreover, his papers on (self-supervised) representation learning [2, 11] appeal to my desire to study meta-learning and provide context for future improvements on musical classification analysis with representation learning. My specific project ideas for research are centered around but not limited to the application of novel cross-modal generative methods to various tasks, and studying the singing voice:

Generative methods: By far the most publicly-discussed advance in AI recently has been state-of-the-art image generation capabilities from models such as DALL-E, Imagen, and more recently, stable diffusion. The implementation of new generative methods such as diffusion and VQ-VAEs, as well as the usage of tremendous corpora of captioned images online to construct cross-modal representation extractors such as CLIP, have allowed these models to become highly skilled in various tasks: Prompt-conditioned image generation, image translation, image inpainting, representation learning... The potential such models represent for the music-making process is attractive but is held back by the amount of captioned musical audio data for training available online, prompting the exploration of cross-modal retrieval. Diffusion models and VQ-VAEs have made their way to the music space [5, 4] - and work has already been done on prompt-conditioned non-musical audio generation [6] and music translation [8] - but the results are comparatively less impressive than image generation, and the corpus of GAN-based generation models remains larger than that of newer methods. Diffusion models have even been adapted towards source separation and AMT recently. Though cross-modal music representation learning has recently been the subject of a growing number of studies [7, 12], a reference model for music such as CLIP for images has yet to be trained. By building upon these studies and cross-modal representations, one can imagine many applications that could be of use for musicians: generating production-ready samples and loops (percussive, melodic, or harmonic) conditioned on prompt, style, audio, or context. Time-domain, frequency domain, or symbolic domain melody or harmony infilling. Melody-conditioned Harmony generation. Zero-shot or few-shot source separation. Instrument conditioned/agnostic automatic transcription. These are all projects that would highly motivate me.

The singing voice: Though I am very much interested in other instruments, as a vocalist the capacities of the singing voice have always astounded me. Though the solo singing voice itself has been rather well explored for automatic transcription, generation, and analysis, I believe there is still work to be done. The generation of flawless singing voices is still a ways away and could be improved with previously mentioned generative methods. studies of vocal ensembles for transcription and source separation exist [3]. Still, they are mainly constrained to choirs or solo voices rather than pop-style backing vocals, which remain scarcely explored to the best of my knowledge. These studies are of interest to other ensemble transcription tasks such as brass and string ensembles. Another area that has seldom been studied is of singing voice cloning. Though applications exist [9], the obtained results are somewhat unnatural, and these studies are mostly lyric-less. The implications are attractive: sing a phrase and have Ariana Grande sing it back to you - better. Singing voice synthesis and analysis can be a novel tool for music producers and artists alike - and a powerful tool to better understand the most used instrument in the world

I believe my experiences and skills make me a perfect fit for the doctoral program at NYU Steinhardt. I have pursued three research internships during my studies, one exploring acoustic diode effects at Ecole Centrale de Lyon, one researching active control of the first vibration modes of a cello at IRCAM, and my master's thesis internship over the course of which I delved into music tagging towards a better recommender system at Groover. These experiences have allowed me to develop both my academic writing skills and my research skills. My studies in acoustics engineering have given me insight into fields closely related to music, AI, and research: Machine Learning, Digital signal processing, Acoustics, Musical acoustics, and statistics. Though pursuing my classes in Australia from France during the pandemic made keeping up a challenge, I obtained the golden key award for academic excellence (top 10% of class GPAs). To broaden my knowledge of ML and find my way toward AI in music, I pursued extra university ML courses. In one project-based course, I proposed and completed a project to classify vocalists using melgrams and computer vision. Through this entire experience, I have become aware of my capacity for discipline and resilience. Three data science internships have allowed me to develop and maintain a set of technical skills essential for music AI, including but not limited to: Python, TensorFlow, PyTorch, Librosa, C++, and Git. My passion for music and the knowledge domain that comes with it is also something I will bring to the table for this program: I believe my love of music will provide a solid basis for my research and my diverse musical experiences in the past years will help me adapt to any research that falls within the intersection of my interests and my supervisors'. It will also fuel the scientific curiosity which is core to any research program.

Throughout my studies, I've understood that music and research are two areas that inspire me and motivate me to expand my knowledge and go above and beyond expectations. I now find myself in the thrilling position of knowing - thanks to recent experiences - exactly how those puzzle pieces should fit together to fulfill me academically, professionally, and personally. Research - like music - is an arduous road filled with challenges and creative endeavors but I believe that the skills I have acquired in my past experiences, my passion for the field, and the resilience, and rigor I have displayed in my academic curricula have equipped me for the task at hand. I look forward to a path at NYU Steinhardt and the challenges it will bring. I am certain that I will discover new passions, curiosities, and questions as I prepare for a career in Music AI research.

References

[1] R. M. Bittner, B. McFee, J. Salamon, P. Li, and J. P. Bello. Deep salience representations for f0 estimation in polyphonic music. In *ISMIR*, pages 63–70, 2017.

[2] M. Buisson, B. Mcfee, S. Essid, and H. C. Crayencour. Learning Multi-Level Representations for Hierarchical Music Structure Analysis. In *International Society for Music Information Retrieval (ISMIR)*, Bengaluru, India, Dec. 2022.

[3] H. Cuesta, B. McFee, and E. Gómez. Multiple f0 estimation in vocal ensembles using convolutional neural networks. *arXiv preprint arXiv:2009.04172*, 2020.

[4] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever. Jukebox: A generative model for music, 2020.

[5] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. 2020.

[6] F. Kreuk, G. Synnaeve, A. Polyak, U. Singer, A. Défossez, J. Copet, D. Parikh, Y. Taigman, and Y. Adi. Audiogen: Textually guided audio generation, 2022.

[7] I. Manco, E. Benetos, E. Quinton, and G. Fazekas. Contrastive audio-language learning for music. 2022.

[8] N. Mor, L. Wolf, A. Polyak, and Y. Taigman. A universal music translation network, 2018.

[9] A. Polyak, L. Wolf, Y. Adi, and Y. Taigman. Unsupervised cross-domain singing voice conversion. *arXiv preprint arXiv:2008.02830*, 2020.

[10] Y. Wang, J. Salamon, M. Cartwright, N. J. Bryan, and J. P. Bello. Few-shot drum transcription in polyphonic music. *arXiv preprint arXiv:2008.02791*, 2020.

[11] H.-H. Wu, C.-C. Kao, Q. Tang, M. Sun, B. McFee, J. P. Bello, and C. Wang. Multi-task self-supervised pre-training for music classification. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.

[12] H.-H. Wu, P. Seetharaman, K. Kumar, and J. P. Bello. Wav2clip: Learning robust audio representations from clip. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022.