DMRN+16: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2021

QUEEN MARY UNIVERSITY OF LONDON
TUE 21 DECEMBER 2021

A Benchmark Dataset to Study Microphone Mismatch Conditions for Piano Multipitch Estimation on Mobile Devices

Jakob Abeßer, Franca Bittner, Maike Richter, Marcel Gonzalez, Hanna Lukashevich

Semantic Music Technologies, Fraunhofer IDMT, Ilmenau, Germany, jakob.abesser@idmt.fraunhofer.de

***Abstract*— In this paper, we present the IDMT-PIANO-MM dataset, which allows to evaluate piano transcription algorithms under microphone mismatch conditions. In particular, we discuss specific constraints that these algorithms need to face when being used in music learning applications on mobile devices. Then, we describe the dataset w.r.t. recording locations and devices as well as the recorded music pieces. We intend this dataset to be a public benchmark to evaluate the robustness of AI-based MPE models within realistic microphone-mismatch conditions, which are to be expected with the large number of potential users of music learning applications.**

***Index Terms*—** Multipitch estimation, piano transcription, microphone mismatch, mobile devices

I. INTRODUCTION

In the field of Music Information Retrieval (MIR), the pitch detection of multiple simultaneous tones (multipitch estimation, MPE) is a challenging research task. MPE is commonly approached by recognizing characteristic patterns such as fundamental frequencies and their corresponding overtones in spectrogram representations. Traditional methods use decomposition techniques such as Non-Negative Matrix Factorization (NMF) whereas recent methods solely focus on deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [1].

A particularly interesting application scenario for multipitch estimation algorithms are music learning applications. Here, audio recordings of musical performances need to be transcribed and compared to a given reference notation with near real-time latency in order to assess the user's performance. Furthermore, music learning applications need to run on mobile devices, which have limited computational resources and microphones of very different quality. As a consequence, these constraints limit the complexity of the applied MPE algorithm. Finally, the applied MPE algorithm needs to be robust to acoustic parameters such as reflection times in the users' practice rooms. It has been widely observed that AI-based audio analysis algorithms exhibit a performance drop in domain shift scenarios, which are caused for instance by a microphone mismatch between the initial training data and the test data.

II. DATASET

Established evaluation datasets for piano transcription include audio files from YouTube videos[1], recordings of acoustic grand pianos such as the Yamaha Disklavier[2], or synthesized using professional sample libraries[3]. These datasets mostly lack the required variety of instrument models as well as metadata details about the spatial parameters of the recording locations. As the main contribution of this work, we present the IDMT-PIANO-MM dataset[4], which allows to study microphone mismatch conditions for piano multipitch estimation recorded with mobile phones. The dataset includes a total of 432 piano recordings (around four hours), which cover nine music pieces recorded in eight different rooms using six different recording devices. The pieces cover classical music (B. Bartók, W. A. Mozart, J. Pachelbel, and L. v. Beethoven) as well as jazz (S. Joplin as well as own compositions) and range from simple to medium difficulty. All music pieces are in the public domain. The recording locations range from small rooms to a large lecture hall. Information about the room geometries, piano position within the room, as well as wall materials are documented. The rooms include four different grand pianos, three upright pianos, and one stage-piano. At each location, audio recordings were made with three mobile phones (iPhone 6S Plus, Redmi Note 8, LG G6), two tablets (iPad Air 2, Amazon Fire tablet), and one stereo setup using two high-quality Oktava MK 012 microphones in an AB recording setup. In our presentation, we will show the results of an initial data inspection focusing on properties such as the dynamic range of the recordings. Also, we compare the different microphone characteristics using the spectrum correction method proposed in [2].

III. REFERENCES

[1] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic Music Transcription: An Overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2019.

[2] M. Kośmider, "Calibrating Neural Networks for Secondary Recording Devices," DCASE2019 Challenge, Tech. Rep., 2019.

[1] e. g., Giant-MIDI Piano, https://github.com/bytedance/GiantMIDI-Piano

[2] e. g., MAESTRO, https://magenta.tensorflow.org/datasets/maestro, MAPS https://hal.inria.fr/inria-00544155/en

[3] e. g., SMD-Synth https://zenodo.org/record/4637908

[4] https://www.idmt.fraunhofer.de/en/publications/datasets.html