Fellowship ID：BR200203

独立行政法人日本学術振興会理事長　殿
To: President, Japan Society for the Promotion of Science

研究活動報告書

Research Report

１．受入研究者/ Host researcher

受入研究機関・部局・職
Name of Host Institution, Department and Title
国立情報学研究所
コンテンツ科学研究系　助教

受入研究者氏名
Host Researcher's Name
ユ　イ

２．外国人再招へい研究者/ Fellow

所属研究機関・部局・職
Name of Institution, Department and Title
CNRS & Sorbonne Université, LIP6, Research Director

外国人再招へい研究者氏名
Fellow's Name
Jean-Pierre BRIOT

３．採用期間/ Fellowship Period

2022 年　10 月　20 日　　　～　　　2022 年　11 月　7 日

４．研究課題/ Research Theme

Deep learning techniques for music generation, computational synthesis and creative systems

５．研究活動報告/ Research Report
（１）研究活動の概要/ Summary of Research Results

We would like to thank JSPS support for having made this mission possible.

Unfortunately, the organization of this mission turned out very difficult because of
1) the Covid effect onto travel and work requirements, and
2) changes and complications in the visa demand process.

The initial schedule, planned for February 2021 had to be changed and extended various times. The process for the visa also turned out very long and complicate. The final schedule was planned as from 30th of September 2022 to 29th of October 2022. The process to get the Certificate of Eligibility (CoE) started in May 2022 and ended up in late July with only some partial documents (Certificate for Completion of Registration to the ERFS system) delivered by the immigration service. Unfortunately, these documents were then refused in August by the Visa service at the Japan Embassy in Paris. We had to restart the demand for CoE, wait for them for more than a month a half, receiving them in Tokyo only on 7th of October. (While the deadline for starting the fellowship was 30th of September). We had to replan in a hurry the mission and visit, in order to rematch with the availabilities of visited researchers, resulting in a new schedule from 20th of October to 7th of November 2022.

（注）採用期間終了後３ヶ月以内に提出。
(Note) Submit the form within 3 months after the expiration of fellowship.

Jean-Pierre Briot visited three research institutions:
- the Speech and Audio Processing Group at Kyoto University in Kyoto, hosted by Professor Yoshii Kazuyoshi
- the STAIR (Software Technology and Artificial Intelligence Research) Lab at Chiba Institute of Technology in Chiba, hosted by Professors Akikazu Takeuchi and Akinori Yonezawa
- the Digital Content and Media Sciences Research Division at National Institute of Technology in Tokyo, hosted by Professor Yi Yu (the initial/main host)

In each of these three institutions, Jean-Pierre Briot has conducted a research seminar, titled: "Music creation with deep learning techniques: Achievements and challenges", with a large audience and following extensive discussions.

In each of these three institutions, Jean-Pierre Briot also had several research meetings and discussions about general and more specific topics about AI and machine learning-based music creation and co-creation. This included presentation of works by various researchers and students, with discussion and comments and suggestions provided by Jean-Pierre Briot.

We also had meetings and presentations with researchers and PhD students from other research groups, e.g., Xin Wang, from Professor Junichi Yamagishi group at NII. We also met with Professor Emmanuel Planas, in charge of international cooperation at NII. Last, we also had meetings with long time colleagues, such as Professor Shinichi Honiden, formerly from NII and now at Waseda University, and with Professor Mario Tokoro, formerly from Keio University and Sony Computer Science Laboratory, and also with Professor Philippe Codognet, Director of JFLI, the joint France/ Japan international research laboratory in information science.

In National Institute of Informatics, Professor Yi Yu and Jean-Pierre Briot have jointly organized a special NII-LIP6 research seminar on 1st of November about deep learning-based music generation, with the following presentations:

- Tokenization of symbolic music and its impact on generation quality, Nathan Fradet, LIP6 & Aubay
- Integrating energy consumption in the evaluation of neural audio synthesis models, Constance Douwes, IRCAM
- Melody Generation from Lyrics with Controllability of Musical Style, Zhe Zhang, NII
- Cross-Modal Generative Model for Music, Wei Duan, NII
- Co-Attention Model for Audio-Visual Cross-Modal Correlation Learning, Jiwei Zhang, NII
- Data Augmentation for Lyrics and Melody Generation, Louis Ngo, NII

We have found connections between the work of Nathan Fradet, PhD student advised at Sorbonne Université by Jean-Pierre Briot, working on token representations for deep-learning based music generation and the works of Zhe Zhang [YZD+2022], Wei Duan [DYZ+2022], and Jiwei Zhang, PhD students advised at NII by Yi Yu, about music melody generation from lyrics and vice-versa. Zhe Zhang has already investigated the use of MidiTok, a framework for music tokens by Nathan Fradet for melody generation [FBC+2021]. Another important issue for music generation is objective evaluation [BHP2019]. Standard objective metrics are about quality of prediction or reconstruction, but they are not adapted for original creation. Nathan Fradet is currently investigating new approaches for designing evaluation metrics and we plan to experiment such novel metrics on melody generation testbeds designed in the research team of Professor Yi Yu.

We have also found connections between the work of Constance Douwes, PhD student advised at Sorbonne Université by Philippe Esling and Jean-Pierre Briot, working on a methodology for evaluating the energy cost of deep learning-based audio generation and the balance with quality objective [DEB2021], and the work of Peter Bryzgalov, senior scientist at STAIR Lab in Chiba Institute of Technology, on estimation of training time for deep learning architectures [BMS2021]. Prior exchanges and discussion has already started. Jean-Pierre Briot has invited Peter Bryzgalov as a member of the PhD defense committee of Constance Douwes, at IRCAM, Paris, on 10th of March 2023.

（注）採用期間終了後 3 ヶ月以内に提出。
(Note) Submit the form within 3 months after the expiration of fellowship.

2

In summary, we found this mission quite successful and productive although the various organizational problems encountered. This mission allowed reconnexion of Jean-Pierre Briot with ancient universities/research colleagues in Japan and with more recent and new colleagues, with professor Yi Yu in the first place. As mentioned above, it also provided constructive ongoing prospects for the future. We would like to thank JSPS for their support which made this mission possible. We also would like to thank Professor Yi Yu and her team and NII for all their support and care for hosting us.

（２）研究キーワード/ Keywords

Information science, machine learning, deep learning, music, generation.

（３）主な研究発表（雑誌論文、学会、集会、知的財産権等/ Main Research Publications

[BHP2019] Jean-Pierre Briot, Gaëtan Hadjeres, and François-David Pachet. Deep Learning Techniques for Music Generation. Computational Synthesis and Creative Systems. Springer Verlag, 2019.

[BMS2021] Peter Bryzgalov, Toshiyuki Maeda, and Yutaro Shigeto. Predicting How CNN Training Time Changes on Various Mini-Batch Sizes by Considering Convolution Algorithms and Non-GPU Time. In Proceedings of the 2021 Workshop on Performance EngineeRing, Modelling, Analysis, and VisualizatiOn STrategy (PERMAVOST 2021).

[DEB2021] Constance Douwes, Philippe Esling, and Jean-Pierre Briot. Energy Consumption of Deep Generative Audio Models, ArXiv:2107.02621, October 2021.

[DYZ+2022] Wei Duan, Yi Yu, Xulong Zhang, Suhua Tang, Wei Li, and Keizo Oyama. Melody Generation from Lyrics with Local Interpretability. Accepted for publication in ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2022.

[FBC+2021] Nathan Fradet, Jean-Pierre Briot, Fabien Chhel, Amal El Fallah-Seghrouchni, and Nicolas Gutowski. MidiTok: A Python Package for MIDI File Tokenization. 22nd International Society for Music Information Retrieval Conference (ISMIR 2021), Late Breaking Demo, Online, November 2021.

[YZD+2022] Yi Yu, Zhe Zhang, Wei Duan, Abhishek Srivastava, Rajiv Shah, and Yi Ren. Conditional Hybrid GAN for Melody Generation from Lyrics. Neural Computing and Applications, 35:3191–3202, 2023.

[ZYT+2022] Jiwei Zhang, Yi Yu, Suhua Tang, Jianming Wu, and Wei Li. Variational Autoencoder with CCA for Audio-Visual Cross-Modal Retrieval. Accepted for publication in ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2022.

（４）その他/ Remarks

受入研究機関事務担当者記入/ Filled in by Host Institution

部署・氏名		電話番号	

※　部局名等の名称含め、内容に誤りが無いか必ずご確認ください。

（注）採用期間終了後 3 ヶ月以内に提出。
(Note) Submit the form within 3 months after the expiration of fellowship.