



Aalto University
School of Electrical
Engineering

November 14, 2023, Espoo, Finland

To the head of the doctoral school of Sorbonne Université

Review of the Doctoral Work of Mr. Antoine Lavault

“Generative Adversarial Networks for Synthesis and Control of Drum Sounds”

I have read and reviewed the doctoral thesis manuscript submitted by Mr. Antoine Lavault. This doctoral work focuses on the use of a deep learning technique in sound synthesis.

The manuscript of Mr. Lavault’s doctoral dissertation has 116 pages with additional 10 pages containing the table of contents, list of figures, list of tables, and the summary in English and French. The manuscript begins with an Introduction (Chapter 1), which first briefly describes the history of sound synthesis and the objectives and background of the work. Somewhat surprisingly, the Introduction then continues into the details of the physics and sound synthesis of drums, basics of deep learning, and perception and evaluation of sound. Finally, the Introduction lists the main contributions of the thesis.

Chapter 2 describes the current deep learning methods for synthesis as well as evaluation strategies. Chapter 3 discusses databases used for training the neural network models. Chapter 4 contains the main results of this work, mainly the StyleWaveGAN structure, methods to control it, and the objective and subjective evaluations. Chapter 5 concludes the thesis.

There are several nice photos and visualizations throughout the thesis, which I enjoy, including high-quality photos of various drums. The spectrograms are also exceptionally nice and clear. The thesis is written with practically flawless English and is easy to read.

The main result of the thesis is the StyleWaveGAN structure, which can be used for drum synthesis based on meaningful control using only three descriptors: brightness, depth, and warmth. It is compared with DrumGAN proposed by Nistal et al. and with NeuroDrum introduced by Ramires et al. and is shown to outperform them in terms of parametric control. In a perceptual evaluation, a version of the StyleWaveGAN also wins DrumGAN in the synthesis of cymbal and snare sounds.

When finalizing the doctoral thesis I suggest several minor improvements, which can be easily implemented under control of the supervisor. My main concerns are the following.

- The Summary on page XI, which I read as the Abstract of the thesis, should be improved. It now reads as a short version of the introduction, but the abstract should contain also the outcomes of the research in a nutshell. The summary now lacks the results and implications. The results are usually something tangible, like the StyleWaveGAN method, which was introduced in this work and has a nice name. In this case, results would also include the outcome of the quality evaluations and comparisons, such as the winning method and in what respect. The implication, on the other hand, would be how the results of the work are applicable, also possibly for future research. I suggest that the Summary could end with at least one sentence giving hints about applications or uses cases of deep-learning-based drum synthesis.

- The first few pages of the Introduction lack references. A reference should be given when a non-trivial subject is first mentioned, such as additive, subtractive, or FM synthesis. Also, this is not the first text listing limitations and goals of sound synthesis, and for this reason, some previous papers or books discussing these aspects should be mentioned on page 2. I find that the discussion here is very modern, so the list is at least partly different from historical references. Still, references are needed, not to give the false impression that this is the first time that someone understands to hope for good sound quality, real-time capability, and other useful features. Here's one example: J. O. Smith, "Viewpoints on the history of digital synthesis," in *Proc. Int. Computer Music Conf.*, Montreal, Canada, pp. 1–10, 1991. Extended version: <https://ccrma.stanford.edu/~jos/kna/>.
- The Introduction is rather long and could be divided into Chapters 1 and 2, so that Chapter 1 would be very short and Chapter 2 would contain all technical background material. Chapter 1 could still be called Introduction and would consist of the beginning of the current Introduction, with Sections 1.1 and 1.2 as well as Sec. 1.7., whereas the new Chapter 2 could be called Background or something like that and would consist of Sections 1.3-1.6.
- The description of the scientific novelties of this doctoral work, which are now presented in Sec. 1.7, could be improved. The StyleWaveGAN method should be explicitly mentioned also here, as it is the main novelty and has a name.
- In the Conclusion, it is claimed that the proposed method is "sufficiently fast" "to be capable of being integrated into real-time applications". However, was this formally shown somewhere? I could not find a description of a real-time test in the thesis. Unless this has been tested and documented, the wording must be changed so that it is true to what is known. An estimate of the number of operations per sample may still be easy to do.
- More references are needed also elsewhere in the manuscript, such as on previous attempts on drum sound synthesis and on subjective evaluation of synthetic sounds.

I have also mailed a list of detailed suggestions, such as some references and where they are needed, and all corrections (including very few typos), which I have noticed, to Mr. Lavault and Dr. Roebel, so that also these suggestions can be accounted for in the final version of the thesis. In total, all my comments correspond to a minor revision. Overall, I find that the results of the work are valuable for the audio research community and for future research in sound synthesis.

As the scientific quality and novelty of this thesis work are sufficient high, I agree that the defense of this thesis should proceed.

Your sincerely,



Dr. Vesa Välimäki, Professor of audio signal processing, Vice Dean for research (electrical engineering)
Aalto University
Department of Information and Communications Engineering / Acoustics Lab
P.O. Box 13000, FI-00076 AALTO, Espoo, Finland
E-mail: vesa.valimaki@aalto.fi