October 24, 2022 - TC-MLAI Meeting Minutes

Chairs in attendance: Brecht De Man, Christian Uhle, Gordon Wichern

Other attendees:

- Flavio Everardo
- Jean-Marc Jot
- Andy Sarroff
- Angeliki Mourgela
- Christian Steinmetz
- Keith McElveen
- JT Colonel
- Jan Skoglund
- Shahan Nercessian
- David Prince
- Dylan Flesch
- Matthew Pitkin
- Mikus Salgravis
- Nyssim Lefford

+ Others (sorry I didn't record all names)

Agenda

- Welcome and Introduction
- Thank you to Andy (outgoing chair / co-founder)
- Review of past events
    - Workshop on Perceptual Loss Functions
    - AES/IEEE joint webinar
- Proposals for the coming year (bring your ideas)
    - Collect examples where ML system and listening test results diverge
    - Other proposals
- Open Discussion

Notes

- Andy Sarroff who helped organize the 2020 AES Virtual Symposium on Applications of Machine Learning in Audio, out of which this TC was born, and then co-founded the TC has decided to step aside to due to increased management responsibilities in his role at iZotope/Soundwide. The committee thanks Andy for his service.

- **Review of past events**
    - We had the first ever live event sponsored by our TC - “Teaching AI to hear like we do: psychoacoustics in machine learning” at the 153rd convention in NYC. The panelists were:
        - Gerald Schuller and Renato Profeta (TU Ilmenau) on perceptual loss functions including those inspired by audio coding
        - Gordon Wichern (MERL) on benefits and drawbacks of time and frequency domain loss functions.
        - Stefan Goetze and George Close (Univ. of Shefield) on MetricGan formulations for using non-differentiable metrics such as PESQ as differentiable loss function
        - Bernd Edler and Martin Strauss (Audiolabs Erlangen) on perceptual conditioning for flow-based models and the gap they observed between objective metrics and subjective listening tests.

      The workshop was well attended and we received positive feedback. Thanks to Gerald for organizing.

    - The AES/IEEE (IEEE CTSoc and SPS + AES) webinar on multichannel audio and psychoacoustics was successfully completed on September 15, 2022. This event ended up not being too focused on ML/AI, but Andy helped get the ball rolling on this and we discussed at previous TC meetings. TC-MLAI committee member Jean-Marc Jot represented the AES, while Sascha Spors represented IEEE SPS, and Ken Sugiyama represented IEEE CTSoc.

- **Proposals for the coming Year**
    - JT Colonel proposed an idea of a panel related to speech enhancement and processing using modern machine learning techniques. Related to this Brent Harshenberger mentioned work from his colleagues at Warner Bros using automatic dialog matching for things like voice characteristics and mentioned this is something they have published and may be able to discuss at an upcoming event.
    - Christian Uhle proposed an idea of a panel discussing new AI-based products. This has the advantage that we can repeat it at multiple conventions in a series-style format as new products are released.
    - Shahan Nercessian proposed a panel on ML for music – what is the technology that is considered mature and what is upcoming?
    - Shahan also proposed an idea for AI/ML track at an upcoming convention, that grouped together related proceedings papers and panels/workshops. No one is sure exactly what it would take to make this happen, but Brecht will follow up with the upcoming convention chairs. There was much content at the recently completed convention, focused on ML/AI, so this seems like a good possibility.
    - Brent Harshenberger and Christian Steinmetz proposed related ideas on a paper/workshop related to “The state of AI in audio engineering.” We should try to identify open problems and large research questions. Christian thinks it could

be in two parts, one part would be collecting the tasks, but our contribution could be around translating that into more solid research directions. This would be particularly useful for new graduate students and researchers in the field.
    - Keith McElveen proposed an idea related to audio deepfake detection, which is currently of great interest in the forensic audio community. Brent said this is also of great interest at Warner Bros./Discovery. Christian said he has seen at least 10 papers on this topic in the last month, and JT mentioned it's relation to watermarking.
    - Gordon mentioned an idea he has been thinking about related to collecting examples where ML systems and listening test results diverge. This could be useful to identify open research problems. Christian mentioned that for image synthesis (e.g., the recent Diffusion model craze) people no longer look at perceptual metrics. Is this also true for audio synthesis? If not, explaining why not could be interesting…
- **Action items**
    - Brecht to follow-up with upcoming convention organizers on the feasibility of an AI/ML track
    - Gordon will create a Google doc for members to sign-up to help organize proposals around some of the ideas we discussed