



Voice activity detection in a multitask approach.

Subject proposed by Nathan SOUVIRAA-LABASTIE (Nahimic / Villeneuve d'Ascq)

Master internship, Lille (France)

Expected start : September - October 2019

Advisors

- Nathan SOUVIRAA-LABASTIE, R&D engineer (Nahimic) nathan.souviraa-labastie@nahimic.com
- Damien GRANGER, R&D engineer (Nahimic) damien.granger@nahimic.com
- Raphaël GREFF, R&D Director (Nahimic) raphael.greff@nahimic.com

Company description

Nahimic (a.k.a. A-Volute) is a company based in Villeneuve d'Ascq (France) that publishes audio enhancement software for the gaming industry, in particular the Nahimic software on MSI laptop. Nahimic has developed a solution for digital and real-time 3D sound. The suite of audio effects proposed by Nahimic includes effects to improve multimedia content (music or movie), as well as microphone effects for communication (noise reduction).

Conundrum

Voice activity detection, more commonly known as VAD, is a speech processing technique used to detect the presence or absence of human speech in an audio signal.

The main applications of VADs are in speech coding and recognition, but it can also be used to disable some processes during the non-voice part of an audio session [1]. This feature would thus reduce the CPU load of our algorithms.

In addition, voice activity can be jointly estimated together with other tasks such as noise reduction or other classification. Examples of multi-tasking approaches can be found in [4, 3].

Subject

The first part of the internship will focus on benchmarking several state-of-the-art techniques (classical signal processing [2], deep learning [5]) and the adaptation of one of these techniques to the needs of Nahimic, the latency and computation cost properties being more important than accuracy for our application. Thus, knowledges in hardware or embedded software would be a plus.

In a second part and if the student has a particular aspiration for machine learning, it will be possible to work on a multi-task approach based on internal work that for the moment focuses on multi-task treatment of music.

Skills

Who are we looking for ? Preparing an engineering degree or master's degree, or even a PhD (3 month visit), you preferably have knowledge in the development and implementation of advanced algorithms for digital audio signal processing.

In addition, advanced notions in the following various fields would be highly appreciated : - Audio, acoustics and psychoacoustics - Audio effects in general : compression, equalization, etc. - Machine learning and artificial neural networks. - Statistics, probabilist approaches, optimization. - Programming language : Matlab, Python.

And experiences in the following areas would be a plus : - Sound spatialization effects : binaural synthesis, ambisonics, artificial reverberation. - Voice recognition, voice command. - Voice processing effects : noise reduction, echo cancellation, antenna processing. - Virtual, augmented and mixed reality. - Computer programming and development : Max/MSP, C/C++/C#. - Video game engines : Unity, Unreal Engine, Wwise, FMod, etc. - Audio editing software : Audacity, Adobe Audition, etc. - Scientific publications and patent applications. - Fluent in English and French. - Demonstrate intellectual curiosity.

Other offers

<https://nahimic.welcomekit.co/>

<https://www.welcometothejungle.co/companies/nahimic/jobs>

Références

- [1] T.-S. CHAN et al. « Vocal activity informed singing voice separation with the iKala dataset ». In : *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, p. 718-722.
- [2] J. RAMIREZ et al. « Efficient voice activity detection algorithms using long-term speech information ». In : *Speech communication* 42.3-4 (2004), p. 271-287.
- [3] A. SØGAARD et Y. GOLDBERG. « Deep multi-task learning with low level tasks supervised at lower layers ». In : jan. 2016, p. 231-235.
- [4] D. STOLLER, S. EWERT et S. DIXON. « Jointly Detecting and Separating Singing Voice : A Multi-Task Approach ». In : *CoRR* abs/1804.01650 (2018). arXiv : 1804.01650.
- [5] X.-L. ZHANG et J. WU. « Deep belief networks based voice activity detection ». In : *IEEE Transactions on Audio, Speech, and Language Processing* 21.4 (2012), p. 697-710.