Towards Statistical Unsupervised Online Learning for Music Listening with Hearing Devices
Purwins, Hendrik; Marchini, Marco; Marxer, Richard

Published in:
First International Music & CI Symposium

Publication date:
2017

Citation for published version (APA):
Towards Statistical Unsupervised Online Learning for Music Listening with Hearing Devices

Hendrik Purwins, Marco Marchini, and Ricard Marxer

The same piece of music maybe be experienced differently by different subjects, depending on various factors: The perceptual processing of the piece (e.g. 3500 inner hair cells vs. a few electrodes in a cochlear implant) may allow for a representation with various degrees of frequency and temporal resolution. The listener’s cognitive capabilities, e.g. memory, voluntary attention direction, abstraction and inference (normal listener vs. a person with dementia) will cause different cognitive processing and musical experience. The familiarity with a particular musical style, instrument, piece may affect the emotional impact and intellectual understanding of a piece.

It could be desirable for a hearing device (hearing aid or cochlear implant) to perform musical signal processing to transfer the musical experience of a piece for a normal listener to a similar musical experience of the same piece in a listener with severe hearing limitations. Such a transformation would require a formal representation of the music piece that is scalable in complexity. Such a scalable representation is provided in [1], where the number of different sounds (e.g. considering open and closed hi-hats as one or two distinct sound categories) as well as the temporal context horizon (e.g. storing up to 2–note sequences or up to 10-note sequences) is adaptable. The framework in [1] is based on two cognitively plausible principles: unsupervised learning and statistical learning. Opposed to supervised learning in primary school children, where the school teacher points at a written letter and articulates its phonetic pronunciation, infants perceptually organise the phonemes they are exposed to into groups, based on the phoneme’s similarity and context (unsupervised learning). 8-month-old infants are able to learn statistical relationships between neighboring speech sounds [3] (statistical learning). In [1], Figure 1, unsupervised learning is implemented as agglomerative clustering, informed by the Gestalt principle of regularity. The model [1] performs statistical learning, applying variable length Markov chains. In [2], grouping of sounds into phonetic/instrument categories and learning of instrument event sequences is performed jointly using a Hierarchical Dirichlet Process Hidden Markov Model.

 Whereas machines often learn by processing a large data base and subsequently updating parameters of the algorithm, humans learn instantaneously, i.e. the mental representation is continuosly changed after every exposure to small batches of sound events (online learning). In [2] online learning is implemented via the interplay of Cobweb clustering and a hierarchical n-gram instantaneously updating the number of timbre groups and their respective transition counts. We propose to use online learning for the co-evolution of both CI user and machine in (re-)learning musical language.

REFERENCES