

Adaptive Multi-Channel Audio Processor

Abstract:

Extraordinary music can be produced, but however, without mixing and mastering the music can't be sold into the market. For new producer/artist or small band or even an individual artist this process will be rather difficult to be achieved. This paper will discuss concepts and approaches to achieve adaptive multi-channel audio processor technology from different perspectives, data, designs, modellings and algorithms in order to improve opportunities for new artists/producers and those who pursue career in this industry. In the end, the goal of the technology is to increase music production rate, reduce production cost, and provide healthy production methods. Whilst, gaps and questions are rises throughout the study and the discoveries of how the concept of psychoacoustic significantly impact this technology.

Introduction:

Music producers in the industry are required to have a musical background, and engineer sounds used/designed in each music tracks. In this modern era, mixing and mastering are essential process that must be done by a music producer to achieve balance frequencies across the spectrum to acquire pleasing mix (Daniel 2019). The term mixing is to combine and adjust sounds from multiple audio channels in a music and balance it with one to another. The output desired from this process is to produce music with pleasing and consistent sound frequency. Since mixing is a very complex process that deals with technical issues and requires producer or mix engineer to create creative decisions (De Man et al. 2015), difficulties to identify and create a decision to manipulate audio channels that become the problem source to the overall audio output will be met. Mastering, on the other hand, is a process to finalize all audio mixes to be louder, unique and ready for distribution. The challenges in this process is to manipulate the final mix to be consistent across different environment (e.g. cars, smart phones, stereos and etc.) (Birtchnell 2018). Those two processes are time-consuming, and exposed human ears to sounds in long duration. Long exposure period leads to ear fatigue with multiple symptoms such as inner ear pain, temporary loss of hearing sensitivity (listener needs more sound volume and worsen the situation). Furthermore, sounds become "blurry", thus inaccurate mixes. In the end, ear fatigue caused discomfort that leads to stress. Potentially, leads to tinnitus which is the "ringing" phenomenon in human ear. For new artists in the industry, the time spent to reach a success point is unknown. Thus, producers that are inexperience in mixing/mastering, required to hire an engineer, hence high music production cost in long term (\$22 - \$45 per hour) (PayScale 2020). This review will focus on data, designs, modellings, and algorithms from different perspectives related to an adaptive multi-channel audio processor to develop this technology. Thus, improving opportunities for new/inexperience producers, increase productivity rate, reduce production cost and allows for healthy production methods.

Literature Review:

Moffat & Sandler (2019) demonstrate studies to approach intelligent music production. The study introduced intelligent music production (IMP) as an approach to utilize and develop artificial intelligent (AI) in music production. Provided, there are various of processes to perform audio manipulation, analysis, and interaction by the intelligent system. The term IMP in this study is divided into two components. The first system is a collaborative music production or human-computer interaction (HCI). The second system introduced is fully autonomous program that has "black box" characteristics. The study claimed, in order to achieve IMP, it is integral to consider consumers interaction and its consistency with the system.

Furthermore, considering how the system will interact with audio inputs. Despite providing innovative and unique processes to audio signal manipulation/analysis, risk arises of how an individual misappropriate the technology. It is wise to consider the idea introduced by De Man et al. (2015) which involved frequency masking that compare one audio frequency to another that is loud or has the same frequency in order to eliminate any unnecessary audio (very low audio volume). Hence developing auditory spatial perception and intelligent clean mix algorithm.

Reiss (2011) conduct a study which describes multichannel audio signal processing that exploit inter-channel relationship to achieve multichannel content manipulations. Furthermore, the study provides descriptions of the required technologies and program architecture to develop this system. It is believed that to mix multiple audio channel it can be done efficiently and simpler by utilizing intelligent plugin/program. The tools would reduce workload of an audio engineers by automate any repetitive tasks and assist inexperienced producers/engineers to develop high-quality audio mixes. Furthermore, it is beneficial for those with economical constraint to hire an audio or a recording engineer. However, it is doubtful for a machine to conduct creative decisions that depend on artistic point of views. If the technology is achieved, it allows musician to focused on music creation and allows engineers to construct new innovative ideas. Psychoacoustic is human psychological perception towards sound and how it physiologically affects them. Different human has different opinions to sounds, that change the demand for the system to perform artistic decision for sound manipulation. Since, perceptions toward sound are complex and sound contains both psychological and physical components (Iakovides et al. 2004), psychoacoustic data still becoming the barrier for system engineers to develop this technology to its full potential. According to Reiss (2011), it is recommended to perform a listening test with different listener in order to determine of what kind of undesired artifacts can be found in audio production process. Hence, fulfilling consumers demand and develop this technology.

Mimilakis et al. (2013) proposed a study of an automated tonal balance enhancement for un-mastered audio due to its process which involves the selective enhancement on audio frequency. Music information and mode of audio material become the foundation of the computational process of this program. The work introducing an automated adaptive equalizer (EQ) that perform a fundamental frequency/pitch tracking to conduct audio mastering procedure where the system provides a powerful sample by sample computations. The performance of the system is conducted via Perceptual Evaluation of Audio Quality (PEAQ). Reiss (2011) recommend to perform listening test engaged with listeners, and in this study, Mimilakis et al. (2013) perform a listening test to collect data of which sound is more preferable. The study by Mimilakis et al. (2013) found that PEAQ measurements of this technology acquire PEAQ human preferences rate of 21.25% for original audio, and 78.75% for enhanced audio with tonal balance enhancement. However, the study claimed that the PEAQ results do not affect human perception since the output is highly-identical to the original audio but enhanced perceptually with tonal balance system's characteristics. Mimilakis et al. (2013) recommend to involved alternative equalization strategies, additional compression and spatial enhancement to develop the "brain" of the system. Hence, to minimize human involvement and enhance human-computer interaction (HCI) for audio mastering. As the machine learn and gather data from the analyzed un-mastered audio, human able to manipulate and improve those data to engineer/develop new optimum setups and algorithms to be implemented in the machine. Thus, smarter machine to master audio to its finest condition/quality.

Ramirez & Reiss (2017) perform a study/research about the relationship between intelligent audio mixing system with deep neural network. This research desire to explore of how deep neural network can be trained to perform audio as a content-based transformation by utilizing any of the standard mixing tools such as compressor, equalizers, limiters, etc. Ramirez, & Reiss (2017) investigate frequency of raw audio recording/input and target frequency of audio recording (stem) to train deep auto encoder (DAE). Through this method, the study would able to determine the capability of the system to learn audio transformation that utilize the same chain of audio effects. Overall, it is found that each DAE contains of 3 hidden layers of 1024 neurons and 1025 layers of input and output layers, where each DAE is independently trained to individual musical instrument group. DAE performances on analyzing bass and guitar are more robust compare to its performances to analyze vocals and keys. It logical for this phenomenon to occur as human has vocal frequencies and complexities that differ with other individuals. While keys have frequencies and roles that differ depending on music genres. The study recommended to perform more advance listening test to acquire more accurate system's performance measurement. In the end, the study believes that intelligent technology will ideally work as the user's assistant for sound manipulation/analysis, where the system allows user to learn from the technology. Followed by an interesting fact of how the concept of studies done by Fazekas (2012) intersect with the conclusion in this study where machine becomes a technical assistant and enhances user's performance in mixing process.

The study, performed by Fazekas (2012), investigates the use of Semantic Web technologies to integrate information management in music informatics and productions. The core contribution of the study is a semantic web technologies that describes recording studios, artefacts in music production technology, music production workflow data collection methodologies, facilitation of audio engineers in music production and finally its applications to develop an automated audio analysis. Fazekas (2012) claimed that the recent audio analysis and music understanding development are sufficient to significantly enhance music production workflow. Fazekas (2012) claimed that the technology will facilitate the creation of semantic audio software to be utilized as tutoring program for inexperience users/consumers. Thus, provide service and guidance for consumer to grow in music industry. On the other hand, it will improve the skill and knowledge of a professional engineer to innovate new idea for sound manipulation and analysis.

Research Gaps & Questions:

Gaps:

1. Information about economic benefits from those studies are vague and hasn't been clarified
2. The impact of the technology to listener physically is not being studied
3. Data of average time spent to mix and master hasn't been explored
4. Number of music genre tested are limited.

Questions:

1. Is it possible to integrate these technologies to develop a system that will adapt to any changes made to the sound input and filter all the unnecessary audio frequency?
2. How the current audio processor can be developed so that the adaptive analysis and manipulation can be computed during recording process?
3. How efficiency it would be to utilize frequency spectrum based on music genre as a preset to allow more freedom and flexibility to mix/master music based on preferences?
4. What features in these technologies can be developed for error identification systems?
5. How many percent of average normal time spent on mixing or mastering can be reduced by this technology?

References:

Birtchnell, T., 2018. Listening without ears: Artificial intelligence in audio mastering. *Big Data & Society*, 5(2), p.205395171880855.

Clason, D., 2017. *Hearing Loss And Listening Fatigue*. [online] Healthy Hearing. Available at: <<https://www.healthyhearing.com/report/52807-Hearing-loss-and-listening-fatigue>> [Accessed 8 August 2020].

De Man, B., Boerum, M., Leonard, B., King, R., Massenburg, G. and Reiss, J., 2015. Perceptual Evaluation of Music Mixing Practices. 138th Audio Engineering Society convention 2015, 1(9235), pp.129-136.

Dixon, D., 2019. *Psychoacoustics: How Perception Influences Music Production*. [online] iZotope. Available at: <<https://www.izotope.com/en/learn/psychoacoustics-how-perception-influences-music-production.html>> [Accessed 8 August 2020].

Fazekas, G., 2012. Semantic Audio Analysis Utilities and Applications, PhD thesis, Queen Mary University of London, London.

Hafezi, S. and Reiss, J., 2015. Autonomous Multitrack Equalization Based on Masking Reduction. *Journal of the Audio Engineering Society*, 63(5), pp.312-323.

Hcaudiology. 2020. *Tinnitus*. [online] Available at: <[http://www.hcaudiology.net.au/tinnitus/#:~:text=Tinnitus%20is%20a%20physical%20condition,or%20ringing%20like%20a%20bell'](http://www.hcaudiology.net.au/tinnitus/#:~:text=Tinnitus%20is%20a%20physical%20condition,or%20ringing%20like%20a%20bell'>)> [Accessed 7 August 2020].

H., J., 2020. *Ear Fatigue Symptoms When Mixing With Headphones & Monitors | LN*. [online] LedgerNote. Available at: <<https://ledgernote.com/columns/mixing-mastering/ear-fatigue-symptoms/>> [Accessed 7 August 2020].

Iakovides, S., Iliadou, V., Bizeli, V., Kaprinis, S., Fountoulakis, K. and Kaprinis, G., 2004. Psychophysiology and psychoacoustics of music: Perception of complex sound in normal subjects and psychiatric patients. *Annals of General Hospital Psychiatry*, 3(1), p.6.

Mastering the Mix. 2015. *Creating Space In Your Mix*. [online] Available at: <<https://www.masteringthemix.com/blogs/learn/48456965-creating-space-in-your-mix#:~:text=Try%20and%20play%20around%20with,to%20give%20your%20mix%20depth.&text=Panning%20is%20a%20useful%20tool,focus%20on%20high%20energy%20elements.&text=You%20can%20use%20reverb%20and,without%20affecting%20your%20mono%20signal.>>> [Accessed 7 August 2020].

Mimilakis, S., Drossos, K., Floros, A. and Katerelos, D., 2013. Automated Tonal Balance Enhancement for Audio Mastering Applications. *AES 134th Convention*,.

Moffat, D. and Sandler, M., 2019. Approaches in Intelligent Music Production. *Arts*, 8(4), p.125.

Payscale. 2020. *Sound Engineer Hourly Pay In Australia | Payscale*. [online] Available at: <https://www.payscale.com/research/AU/Job=Sound_Engineer/Salary> [Accessed 7 August 2020].

August 2020].

Ramirez, M., and Reiss, J., 2017. Deep Learning and Intelligent Audio Mixing. *Proceedings of the 3rd Workshop on Intelligent Music Production*

Reiss, J., 2011. Intelligent systems for mixing multichannel audio. *2011 17th International Conference on Digital Signal Processing (DSP)*,.

Rosenkranz, R., 2018. *Psychoacoustics*. [online] TU Dresden. Available at: <https://tu-dresden.de/ing/elektrotechnik/ias/aha/forschung/akustik/psychoakustik?set_language=en> [Accessed 8 August 2020].

Stamper, A., 2018. *In The Studio: Beware Of Ear Fatigue - Prosoundweb*. [online] ProSoundWeb. Available at: <<https://www.prosoundweb.com/in-the-studio-beware-of-ear-fatigue/>> [Accessed 9 August 2020].