arts

MDPI

Article

Approaches in Intelligent Music Production

David Moffat * and **Mark B. Sandler**

School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK

* Correspondence: d.j.moffat@qmul.ac.uk; Tel.: +44-(0)20-7882-5555

Received: 18 July 2019; Accepted: 10 September 2019; Published: 25 September 2019

Abstract: Music production technology has made few advancements over the past few decades. State-of-the-art approaches are based on traditional studio paradigms with new developments primarily focusing on digital modelling of analog equipment. Intelligent music production (IMP) is the approach of introducing some level of artificial intelligence into the space of music production, which has the ability to change the field considerably. There are a multitude of methods that intelligent systems can employ to analyse, interact with, and modify audio. Some systems interact and collaborate with human mix engineers, while others are purely *black box* autonomous systems, which are uninterpretable and challenging to work with. This article outlines a number of key decisions that need to be considered while producing an intelligent music production system, and identifies some of the assumptions and constraints of each of the various approaches. One of the key aspects to consider in any IMP system is how an individual will interact with the system, and to what extent they can consistently use any IMP tools. The other key aspects are how the target or goal of the system is created and defined, and the manner in which the system directly interacts with audio. The potential for IMP systems to produce new and interesting approaches for analysing and manipulating audio, both for the intended application and creative misappropriation, is considerable.

Keywords: intelligent music production; automatic mixing; adaptive audio effects; audio processing; artificial intelligence; machine learning

1. Introduction

Intelligent music production (IMP) is the approach of bringing some artificial intelligence into the field of music production. In the context of this article, we consider music production to be the process of mixing and mastering a piece of music. As such, IMP has the capacity to work alongside mix engineers in a supportive and collaborative capacity, or to significantly change and influence their pre-existing workflow.

IMP is a developing field. It has the prospects to fundamentally change the manner in which engineers and consumers interact with music. The ability, not only to allow for the collaboration with an intelligent system, but also to explore and understand new dimensions and control approaches to sonic spaces can unleash potential new concepts and ideas within the space of music production. This will in turn challenge the current creative processes, change the immersion and interaction experienced by a consumer, and potentially change the way in which everyone interacts with music. It is clear that there is a need to further understand and develop future technologies that investigate the prospect of advancing music production. Fundamentally, the audio industry is already far behind technological advancements (Bromham 2016). Many multimedia fields, such as photography and film, have embraced new technology, with the likes of facial recognition, red eye removal and auto-stabilisation techniques (Ooi et al. 1990; Reiss 2018; Smolka et al. 2003). Very few similar advances in music production have emerged as industry standard tools in quite the same way. There is considerable scope for technological advancement within the field of music production.

Dugan (1975) presented seminal work in this area with a fully deterministic adaptive gain mixing system. IMP was drastically changed when Pachet and Delerue (2000) proposed a constraint optimisation approach to mixing multitrack audio, and thus implied there can be some computational solution to be optimised for. Since then, there have been a multitude of development in mixing of musical content (Gonzalez and Reiss 2007) and video game audio (Schmidt 2003). De Man et al. (2017) presented a review of IMP research, classifying research based by the audio effect that was automated. De Man et al. then go on to discuss the evaluation undertaken. However, there is no acknowledgement of any of the three aspects of IMP presented later in this paper and no processes for developing future technology.

Music production in general is a highly dimensional problem relying on combining multiple audio tracks, over time, with different audio processing being applied to each track. The way in which audio tracks are combined is highly dependent on all of other audio tracks present in the mix. The processing being applied to each of the audio tracks is highly subjective and highly creative. Due to the complexities involved in the mixing process, including the interdependencies of audio processing, the reliance of each individual aspect of a mix to tie together, and the high levels of subjectivity in the creative process, makes computational understanding of the mixing process very difficult. There have been a number of cases where a mix engineer will state that *mistakes* can often lead to the best mixing choices (Cascone 2000), often known as *happy accidents*. Some of these challenges in music production make IMP an interesting research area. There are many approaches which attempt to produce rules for mixing (De Man and Reiss 2013a; Pestana 2013), but fundamentally we do not yet fully understand the full mixing process.

The aim of this article is to take an overview of approaches taken in IMP. In every IMP tool, there are numerous decisions made, and we aim to identify these design decisions, whether they are made explicitly or implicitly. The understanding and acknowledgement of these design decisions can hopefully facilitate a better understanding of the intelligent approaches that are being taken, and result in more effective IMP systems.

A general view, as to how an IMP tool performs, is presented in Figure 1. This demonstrates all the key stages of interacting with audio. There is input and output audio, interaction with a human engineer, a decision-making process and the ability to perform an action. There are many examples where one or more of the parts of this process do not exist, or where parts are merged into a single component, but, in general, this outlines the expected requirements of an IMP system.

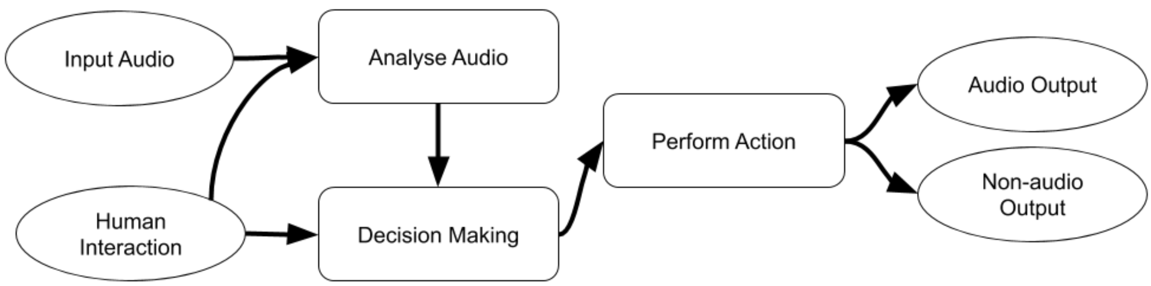

Figure 1. A generalised flow diagram of a intelligent music production tool.

In the field of AI, an intelligent agent is required to have three different aspects, the requirement to observe or perceive the environment, the requirement to act upon the environment, and the ability to make some decision surrounding the goals to be achieved, which includes the interpretation of the environment (Russell and Norvig 2016). This identifies three key aspects of an IMP system, where some approaches will try to solve one aspect, or combine multiple together, to solve both simultaneously. The three aspects of IMP are:

Levels of Control —The extent to which the human engineer will allow the IMP system to direct the audio processing. The restrictions places on the IMP system to perform a task based on the observations made (Palladini 2018).

Knowledge Representation —The approach taken to identify and parse the specific defined goals and make a decision. This aspect of the system is where some knowledge or data are represented, some analysis is performed and some decision-making is undertaken (De Man and Reiss 2013a).

Audio Manipulation —This is the ability to act upon an environment or perform an action. A change is enacted on the audio, either directly, through some mid-level medium, or where suggestions towards modification are made.

The rest of this article is presented as follows: approaches towards producing assistive versus automated mixing tools are discussed in Section 2. The decision-making process is discussed further in Section 3. Section 4 presents the approaches for manipulating and modifying audio content. A discussion as to the the benefits of different approaches in IMP and potential future directions in the future are presented in Section 5, detailing and contrasting different approaches in the field. Conclusion as to the use of IMP approaches are highlighted in Section 6.

2. Levels of Control

IMP is enacted in a number of different ways, depending on the situation. In other fields, there are discussions towards the levels of automation that are implemented (Rödel et al. 2014; Sheridan and Verplank 1978). However, this is typically relating to automating heavy industry or highly process driven tasks, and focuses on aspects from the artificial intelligence and robotics approach. In the context of IMP, the important aspect of automation compared to control, is the manner in which an engineer is able to interact with the intelligent system. This means there needs to be considerable focus on the interaction between the human and the intelligent system, where the individual is handing off control to the intelligence, rather than the automation taking control from the individual. Given this, there are a number of different levels to which intelligence can be incorporated into a music production system (Palladini 2018). These levels of control are: Insightive; Suggestive; Independent; and Automatic.

2.1. Insightive

An insightive approach provides the engineer with the greatest level of control, where some additional insight or advice is provided. The computer has the lowest level of control over the situation. This is the approach designed to increase informed decision by a mix engineer, and, as such, includes any enhanced informative approaches. In this instance, some supportive intelligence is able to present the user with some additional insight into the audio being produced or mixed. This takes the form of supportive textual information, visualisations, or analysis tools that are used to help inform the user of the current state of the audio mixture. The feedback is presented as visualisations (Ford et al. 2015) or haptics (Merchel et al. 2012). The principal is that this approach presents users with additional information or knowledge, allowing them to better understand the mixing that has been produced. These informative representations take a range of approaches including comparing masking levels (Wichern et al. 2015), identifying the impact reverberation on each track in a mix (De Man et al. 2017), or visualising the spatial and frequency components of a mix.[1] In this approach, users are provided with a greater level of understanding as to the current audio mixture, or a tool to aid identifying specific aspects of the mix. There are several examples of existing audio plugins that currently support this approach, such as a masking visualisation plugin (Ford et al. 2015), or a system that can arrange drum loops to allow searching for appropriate content (Bruford et al. 2019). This approach is beneficial to professionals mix engineers, as it allows the most direct control

[1] https://nugenaudio.com/visualizer/.

over the audio, and can simply provide some easier access to information to allow faster or better decisions to be made.

2.2. Suggestive

A suggestive mixing system is one which the user is able to ask a system to analyse and interpret the existing mix. From this, the engineer receives suggestions as to specific parameter settings or approaches for a mix to be developed. Such suggestions include making recommendations as to the audio processing chain to be applied (Stasis et al. 2017), analysing and recommending changes to spectral characteristics (Pestana et al. 2013), applying some initial sub-grouping tasks (Ronan et al. 2015), suggesting parameter settings to reduce perceived masking (Jillings and Stables 2017) or identifying occasions when a single effect has been overused (Vickers 2010). This approach includes the use of automated mix analysis tools (Wilson and Fazenda 2015), which allow for analysis, identifying potential issues or mistakes within a mix. Recommendations are then made to correct or improve the identified issues (Jun et al. 2015). The primary advantage of this approach is that it is able to understand aspects of the mix and present an argument or suggestion to the user, which requires the mix engineer to actively engage with the intelligent system. This ensures the mix engineer is in complete control of the system at all times, but has to actively accept suggestions from an automatic system, either in an exploratory manner, or in agreement with the intelligent system as a more efficient approach. It is important to note that control is always relinquished back to the engineer. The intelligent system can be useful for initial setups and getting the engineer closer to their target, but then will release control, so as not to interrupt the engineer while performing their primary task.

2.3. Independent

An independent system is considered the converse of a suggestive system. Within this approach, an intelligent system is allocated specific tasks to perform, and is able to go about them freely, with a mix engineer acting in a supervisory role. The engineer will always have some facility to overrule or change decisions made by an independent IMP system. This includes selecting control parameters towards a specific target (Gonzalez and Reiss 2008), the use of adaptive plugin audio effect presets, where analysis of the audio facilitates different settings, or selecting the gain parameters of audio tracks, as performed in the Izotope Neutron Plugin.[2] Other options for independent systems include setting up initial mix parameters based on pre-training (Moffat and Sandler 2019) or adaptive filtering for acoustic feedback removal (Van Waterschoot and Moonen 2011). At this stage, a mix engineer is giving over some control of the mix to some intelligent system. The key aspects that are important for this will be ensuring consistency in the approach, so that engineers may have the time to learn to trust the system (Muir 1994). One of the most challenging aspects of this approach is producing an intelligent system that an engineer can interact and collaborate with, without there becoming a battle or a fight with the system. The intelligent system must never be allowed to directly contradict the engineers approach, without good reason, but may develop additional ideas alongside it.

2.4. Automatic

A fully automatic mixing system is one where all control is passed over to the intelligent system, fully automating parameters to a predefined goal (Ronan et al. 2018), or performing a full mix (Terrell and Sandler 2012), given a predefined set or restrictions. Predefined targets or machine learned approaches are used, with the aim of mixing for a specific type of sound (Martínez Ramírez and Reiss 2017a; Mimilakis et al. 2016b). A fully automatic system has the potential to be a full mixing system, where a series of tracks are passed in, and an example mix is returned, without any human interaction,

---

[2] https://www.izotope.com/en/products/mix/neutron.html.

taking complete control of the manner in which audio is mixed and produced. This is advantageous for occasions where amateurs are wanting some system that will mix their tracks together without involvement, as a learning tool to compare to some student work, or in some cases where it is necessary for mixes to change in real time, where it is not possible for a mix engineer to be able to directly interact with the mix, such as in virtual reality and video game audio situations (Selfridge et al. 2018; Stevens and Raybould 2013).

2.5. Control Level Summary

It is clear that intelligent systems interact with music production in numerous different levels. The key differentiator is the level to which the engineer is in control of the system and how the interaction takes place, though these are not the only important factors of an IMP system to consider. There is a vital importance that the engineer using the system can set their own constraints and trust the system to interact with the audio in any predefined and specified manner. The extent to which suggestive and independent systems have been created, in comparison to fully automatic systems, is fairly small. The current field has focused more significantly on constructing fully automatic systems, which are often not the most useful or helpful systems for the end users and practical applications in the field. Given the importance of the control and interaction of the IMP system, it is also of vital importance how the IMP system will create rules and follow them, as will be discussed in the following section.

3. Knowledge Representation

The method in which an intelligent system understands and represents *knowledge* is an important factor, as it will greatly impact its ability to perform tasks, adapt to situations, and the methods in which users interacts or collaborates with it. The approaches to represent knowledge could either be a rule based structure, where a rule is manually defined, such as *all tracks need to have the same perceptual loudness*, or some system from learning these rules and decisions from data, in a machine learning approach. Fundamentally, the ability to understand the goals of the system, the decision constraints put upon the IMP system, and the understanding of intention are all represented in this category. The knowledge representation approach is highly important, as this is the way in which an intelligent system can understand the context of audio being analysed and mixed. Three knowledge representation approaches used in IMP, first identified by De Man and Reiss (2013a), are grounded theory; knowledge based system; and data driven approaches.

3.1. Grounded Theory

Grounded theory is a research method commonly used in the social sciences, where theories are systematically constructed through data gathering and analysis. This has been one of the most common approaches to IMP in recent years. Though the work described in this section does not strictly adhere to the original definition of grounded theory as propose by Glaser and Strauss (1967), the term grounded theory to describe this approach to IMP has been used in this way since De Man and Reiss (2013a). The principal aim of the grounded theory approach is to creating a formal understanding of the mix process and limits of perception (Bromham et al. 2018; De Man et al. 2016), and using this understanding to model the intention of mix engineers (Moffat and Sandler 2018). This data gathering is performed through ethnographic studies (Cohen 1993) and by close interview and analysis of mixing practice (Pestana 2013). Mixing practices are understood through the discussion with mix engineers (Ronan et al. 2017), though there are cases where the professional mix engineer will state that they always take one approach but consistently take an alternative approach (Pestana 2013). This demonstrates the general complexity and nature of the music production problem. This approach is vital to fully understanding the *human element* of music production and mixing. This approach is challenging, as there are regular cases where individuals will believe in one set of rules or approaches, however they may perform in another. There are examples of cases where the best aspects of a mix are created by a *happy accident*, and a large amount of creativity within the music production space is part of breaking the rules, rather than understanding and conforming to a rigid set of rules (Cascone 2000). Therefore, any grounded theory approach also needs to take a rebellious approach, and be able to break rules within a set of larger fixed constraints.

3.2. Knowledge Based Systems

A knowledge based expert system is an approach with some ability to interpret knowledge and reason on the results. The knowledge based system will define or formulate a set of rules or knowledge about the environment, and then allow for some inference engine to apply the rules. These rules take many forms, such as *if-then-else* rules. More complex rules do exist, which require some optimisation approach to solve them. De Man and Reiss (2013b) performed a review of knowledge based systems in IMP and identified that knowledge based systems take the form of either a constraint optimisation problem (Terrell and Sandler 2012; Terrell et al. 2014), or as a set of defined rules, which are solved together through some inference engine (Benito and Reiss 2017; Deruty 2016; Moffat et al. 2018). This is an active area of research, and many approaches identify different aspects of a mix to optimise, such as masking (Jillings and Stables 2017; Ronan et al. 2018), musical score and timbral features (Bocko et al. 2010; Kolasinski 2008), specific loudness targets (Fenton 2018; Wichern et al. 2015), or mixing to a target reference track (Barchiesi and Reiss 2009). One of the key advantages of these systems is the ability to consider multiple objectives and find an optimal solution which fulfils the largest number of targets. As a result, this system is able to consider when and how to break certain sets of *rules* or allocated targets, and interpret different rules with different levels of priority in parallel. There are computational approaches which may never give a single *optimal* solution, but are able to propose a series of comparable *scored* different answers. Similarly, in music production, there is no one single ground truth optimal mix of a music track, but a series of different options that are all considered appropriate of *a good mix* in some respects (Jillings and Stables 2017). This is due to the subjective nature of music production. The primary challenges with a knowledge based system approach is designing systems capable of capturing the rules of a mixing system and represent them in a machine interpretable manner. Many constraint optimisation systems will never be in a position to operate in real-time, due to the design and constraints of the artificial intelligence approach. This will significantly impact the ability to interact with this system, and react to changing aspects of a mix. Instead, some knowledge based systems require a period of time to calculate the requested parameters. This considerably impacts the traditional music production studio workflow. The ability to design and compare mixing rules, which are formally evaluated, will provide a more concrete understanding of exactly when specific rules can be applied in a highly quantitive approach.

3.3. Data Driven

Data driven approaches to IMP, such as machine learning, has developed considerably in recent years. These systems produce some mapping between audio features (Moffat et al. 2015; Scott and Kim 2011) and mixing decisions (Martínez Ramírez et al. 2019; Martínez Ramírez and Reiss 2017b). They are used in a number of different ways, such as the automation of a single audio effect (Chourdakis and Reiss 2016 2017), to perform some track mastering (Mimilakis et al. 2016b), the ability to learn and apply a large amount of nonlinear processing (Martínez Ramírez and Reiss 2019), or even to simply perform an entire mix in a single *black box* system (Martínez Ramírez and Reiss 2017a). The primary aspects of a data driven system is that all understanding of the process taking place is learnt from data. This data could be an individual providing a series of training examples for the system to learn from (Chourdakis and Reiss 2017), learning some particular aspects from a specific dataset (Mimilakis et al. 2013), or learning from large scale curated datasets, such as the Open Multitrack Testbed (De Man et al. 2014) or MedleyDB (Bittner et al. 2014). The entire machine learning field has been growing significantly over the past few decades, and there is ample opportunity for this growth to be further encouraged and utilised in IMP. The largest limiting factor on any data driven approach is the quantity of data that can be collected (Witten et al. 2016). Despite this limitation, data driven approaches are shown to produce highly impressive results when provided with considerable datasets (Arel et al. 2010).

3.4. Knowledge Representation Summary

The method in which some information or knowledge is represented in an IMP system is of vital importance, as each different approach will present restrictions and limitations to what the IMP can achieve. A knowledge based system allows an individual to interact directly with rules, which they prioritise, or remove, whereas a machine learning approach is driven by the presented data. This then leads to problems with data collection and manipulation. Machine learning is more restrictive in the options it presents, when compared to some grounded theory or knowledge based approaches, as only human mixing processes, derived from data, are to be applied. Knowledge based systems have the advantage of finding new approaches based on key aspects of a mix. This also allows for misappropriation and the potential for creative uses of IMP tools, in unforeseen ways. In music production, the processing and mixing applied to each track will be highly dependent on every other track in the mix. There is a high level of interdependence of each and every track. As such, this will considerably grow the scale of the problem, and most of these issues need to be controlled for in the knowledge representation aspect of the IMP system. Once the knowledge and decision-making process has been represented, the IMP system will then perform some interaction with the audio, as discussed in the following section.

4. Audio Manipulation

One of the fundamental objectives for IMP systems is to interact and modify audio. This necessitates an approach for audio manipulation. The approach with which an intelligent system interacts with, and manipulates, audio will place considerably limitations on the structure of the intelligent system. The most common approaches to audio manipulation is the use of adaptive audio effect; however, there is some work that performs direct audio transformation.

4.1. Adaptive Audio Effects

Reiss (2011) proposed adaptive audio effects for IMP. Figure 2 shows the typical flow diagram of an adaptive audio effect. The audio effect is the part of the system that directly modifies audio, with a set of predefined parameters. These parameters are modified by analysis of audio through the feature mapping. The feature mapping aspect performs some mapping or relation between the audio analysis aspects and the audio effect. The audio analysis part performs some feature analysis of an audio signal and typically represents the signal with some audio features (Verfaille et al. 2011). Verfaille et al. (2006) states that adaptive audio effects fall into four different categories:

Auto-adaptive —Analysis of the input audio to be modified will impact the control parameters.
External-adaptive —Alternative audio tracks are used in combination with the input signal, to adjust control parameters.
Feedback-adaptive —Analysis of the output audio to change the control parameters.
Cross-adaptive —Alternative audio tracks are used to effect control parameters, and the input track in turn effects control parameters on external tracks.

These adaptive effects are used within IMP systems, where some parameter automation on each of the pre-existing parameters is undertaken (Pestana 2013). This has been a common approach for some time in IMP, where traditional audio effects are made into auto-adaptive or cross-adaptive audio effects, based on a some grounded theory approaches (Reiss and Brandtsegg 2018). The advantage of this approach is that it allows for easy transition to traditional music production paradigms, with the ability for parameter recommendation and adaptive plugin presets. However, the restrictions in the ways that audio can be manipulated, and the specific complex search space of the plugin limits what an IMP system achieves.

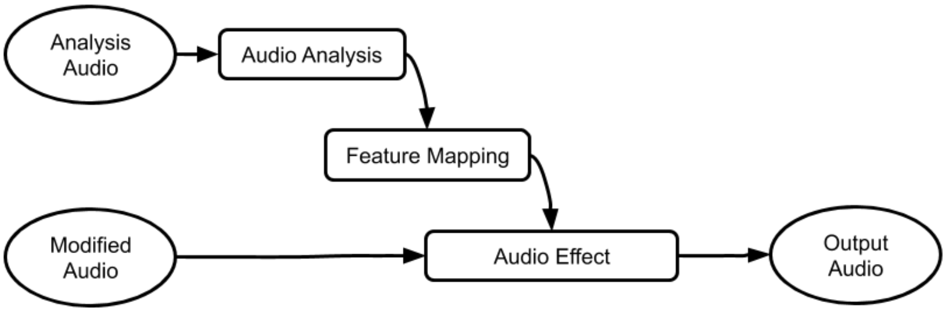

Figure 2. A flow diagram of a typical adaptive audio effect.

4.2. Direct Transformation

The alternative approach is to modify a piece of audio directly. In this approach, instead of using some mid-level audio manipulation which an IMP system needs to learn and understand, the system has the ability to learn the best audio manipulation and enact the audio change directly (Martínez Ramírez and Reiss 2017a; Mimilakis et al. 2016b). This approach is commonly undertaken with neural networks, either in audio mixing (Mimilakis et al. 2016b) or audio style transfer (Verma and Smith 2018), but could as easily be performed with any other approach that is able to learn some signal processing transform. The resulting IMP system is not limited to the human methods and way of interacting with audio, and, not limited in approaches, it can take to produce the desired effect. It also means that there are not complex mapping layers to understand what an audio effect can achieve, and no reliance on a specific DSP (Digital Signal Processing) technology. This has the potential to facilitate the creation of completely new audio effects, which could be more interesting and insightful methods for manipulating audio, in ways that we had never intended, but allowing a greater freedom and flexibility to manipulate audio. This has considerable beneficial impact on traditional music production approaches, as the full IMP system does not necessarily need to be used as in intelligent agent, but instead the approach is used to identify and create new approaches to interact with and modify sound, with a set of exposed parameters that could then facilitate an engineer to use this as a new form of audio effect. The direct transform approach has had little investigation in the field of IMP; almost all intelligent systems work on the basis of modelling and automating traditional audio effects.

4.3. Audio Manipulation Summary

The approach of modifying audio through an adaptive audio effect is an intuitive and natural approach to take, as it agrees with pre-existing notions of music production and thus suggests the possibility to automatically produce a *mix like a human*. However, there are a large number of limitations with this process; where some complex multichannel signal modification is required, it is difficult to achieve this without a large number of different plugs in very specific setups to achieve something that the system would be able to learn in a much more interpretable way. This direct modification is then implemented directly into an audio plugin, and integrated into pre-existing work-flows very easily. The opportunity for new and interesting audio processors to be developed, through understanding how an IMP system may produce an automated mix, is highly advantageous to expert and amateur mix engineers alike.

5. Discussion

There are a number of ways in which IMP tools can be constructed. They may focus on supporting engineers through recommendation of audio effect parameter settings based on some expert knowledge from professional engineers, or they can be data driven fully automated mixing systems. The key aspect of any system is to understand what level of control is expected, and thus selecting different approaches to best achieve the results. There are opportunities to merge or combine different categories together in different ways, which will fundamentally shape the ability and affordances of any IMP tool. For example, it is possible to use a machine learning system to predict sets of audio effect parameters, based on learning data from users; however, this relies on a greater understanding as to the intention of the mix engineer when they performed in initial change. The understanding as to what mixing decisions are made, and why, cannot be underestimated. In another case, a fully automated data-driven approach could be taken, where large quantities of data can be collected, and then the mixing processes extrapolated from the data are applied, with no human interaction. This would act very much as a black-box system, and rely heavily on the quantity and quality of gathered data. It is expected that a automatic data-driven IMP system would produce a range of consistently average results. Although not particularly useful to professional engineers, this is advantageous as an education tool, ideal for an amateur who is looking to get a first recording of their band at minimal cost, and would improve the baseline audio quality of music production being posted online and to streaming services.

Machine learning systems also have the potential to grow quickly, but this relies on solving the issue of accessing and gathering quantities of data, and interpreting mixing decisions made in each case. As the intelligent mixing problem is somewhat different from many other data science based problems, it is required to undertake a number of different approaches to simplify or re-frame the problem, such as data fusion (Hargreaves 2014), which allows modelling of the interdependence of the audio tracks.

The focus on traditional audio effect parameter automation is a limiting factor of current IMP systems, as they are severely restricted to a single, predefined, set of DSP tools. It could instead be the case that transforms or effects are learned with specific intentions, such as an effect for noise removal, or effect for spectral and spatial balancing. It should not require the use of multiple effects to produce a single intention within an IMP system; instead, a single intention can be produced from a single action. The interdependent nature of processing on each individual track, along with recent studies (Ronan et al. 2015), suggests that a hierarchical approach to mixing audio tracks would also be highly advantageous.

Misappropriation is the use of a piece of technology, other than as intended. The misuse and misappropriation of technology has been prevalent in music production for years (Bartók and Suchoff 1993; Prior 2012). Any new technological advancement has the capacity to change the limits and boundaries of the creative process, and thus shape and change potential creative outcomes (Eno 2004). As discussed in Section 4.2, the use of intelligent systems to directly transform and enact a process on the audio signal will allow for the creation of a whole new set of audio effects, where these audio effects are not limited by existing signal processing techniques. These will inevitably allow for a different set of tools and different approach to music production (Bromham 2016; King 2015).

There are numerous promising approaches that have recently been realised in IMP. There have been recent indications that the use of advanced signal processing techniques could have considerable benefits to the field of IMP (Pardo et al. 2018). In particular, the use of source separation and music information retrieval technologies have facilitated better mixing tools and the ability for a system to better classify and understand the audio it is working with. The recent realisation that using a hierarchical mixing structure can improve the resulting mix (Ronan et al. 2018) has considerable implications for future work in IMP.

The growth of internet technologies and the remix culture (Xambó et al. 2019) has a range of opportunities for IMP systems. The use of interactive dynamic music mixing systems can engage audiences (Paterson et al. 2016), and many of these tools rely on some form of IMP system. Dynamic music mixing, similar to video game audio (Stevens and Raybould 2013), requires some automated system which allows levels of interaction and control, the ability for a human engineer to create constraints to an IMP system and for agreed interaction. As such, the understanding of IMP approaches is vital, in the context of any dynamic processes, whether mixing or content creation.

The focus on real time vs. offline mixing processes will both improve the results of IMP systems in general, but also allow for more deterministic interpretable systems that can be used in real time. While it is clear that IMP as a field has considerable challenges to overcome, the fundamental approaches and developments in the field are being made, meaning there are considerable opportunities available in the field of IMP for future research and innovation.

6. Conclusions

The use of IMP tools can have considerable advantages for the music production field. Three key aspects of an IMP system have been identified, which lead to design decisions being considered when producing an intelligent system to interact with audio. The benefits and downfalls of each decision are identified. Current research avenues have explored the use of grounded theory approaches and their application to automate existing audio effects; however, the future of music production involves manipulating audio in ways designed specifically for the required task, and allowing engineers the choice as to how they wish to interact with their IMP tools. There is considerable scope for new research to be performed, to develop different tools for interacting with audio and to facilitate a better understanding of the mixing process. The potential for creating new effects and methods for manipulating audio in new ways can have considerable impact for mixing and misappropriation of the effects. The way in which a mix engineer might interact with any IMP system is of vital importance to their ability to understand, effectively use, trust, and interact with IMP systems.

Author Contributions: Conceptualization, D.M.; methodology, D.M.; writing—original draft preparation, review and editing, D.M.; supervision, M.B.S.; funding acquisition, M.B.S.

Funding: This paper is supported by EPSRC Grant EP/ L019981/1, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption. Mark B. Sandler acknowledges the support of the Royal Society as a recipient of a Wolfson Research Merit Award.

Acknowledgments: The authors would like to thank Gary Bromham for insightful discussion and comments on this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

Arel, Itamar, Derek C. Rose, and Thomas P. Karnowski. 2010. Deep machine learning—A new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine* 5: 13–18. [CrossRef]

Barchiesi, Daniele, and Josh Reiss. 2009. Automatic target mixing using least-squares optimization of gains and equalization settings. Paper presented at the 12th Conference on Digital Audio Effects (DAFx-09), Como, Italy, September 1–4.

Bartók, Béla, and Benjamin Suchoff. 1993. *Béla Bartók Essays*. Lincoln: University of Nebraska Press.

Benito, Adán L., and Joshua D Reiss. 2017. Intelligent multitrack reverberation based on hinge-loss markov random fields. Paper presented at the Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio, Erlangen, Germany, June 22–24.

Bittner, Rachel M., Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam, and Juan Pablo Bello. 2014. Medleydb: A multitrack dataset for annotation-intensive MIR research. Paper presented at the International Society for Music Information Retrieval (ISMIR) 2014, Taipei, Taiwan, October 27–31, vol. 14, pp. 155–60.

Bocko, Gregory, Mark F. Bocko, Dave Headlam, Justin Lundberg, and Gang Ren. 2010. Automatic music production system employing probabilistic expert systems. In *Audio Engineering Society Convention 129*. New York: Audio Engineering Society.

Bromham, Gary. 2016. How can academic practice inform mix-craft? In *Mixing Music*. Perspective on Music Production. Edited by Russ Hepworth-Sawyer and Jay Hodgson. Oxon: Taylor & Francis, chp. 16, pp. 245–56.

Bromham, Gary, David Moffat, György Fazekas, Mathieu Barthet, and Mark B. Sandler. 2018. The impact of compressor ballistics on the perceived style of music. In *Audio Engineering Society Convention 145*. New York: Audio Engineering Society.

Bruford, Fred, Mathieu Barthet, SKoT McDonald, and Mark B. Sandler. 2019. Groove explorer: An intelligent visual interface for drum loop library navigation. Paper presented at the 24th ACM Intelligent User Interfaces Conference (IUI), Marina del Ray, CA, USA, March 17–20.

Cascone, Kim. 2000. The aesthetics of failure: "Post-digital" tendencies in contemporary computer music. *Computer Music Journal* 24: 12–18. [CrossRef]

Chourdakis, Emmanouil Theofanis, and Joshua D. Reiss. 2016. Automatic control of a digital reverberation effect using hybrid models. In *Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. Leuven: Audio Engineering Society.

Chourdakis, Emmanouil T., and Joshua D. Reiss. 2017. A machine-learning approach to application of intelligent artificial reverberation. *Journal of the Audio Engineering Society* 65: 56–65. [CrossRef]

Cohen, Sara. 1993. Ethnography and popular music studies. *Popular Music* 12: 123–38. [CrossRef]

De Man, Brecht, Nicholas Jillings, David Moffat, Joshua D. Reiss, and Ryan Stables. 2016. Subjective comparison of music production practices using the Web Audio Evaluation Tool. Paper presented at the 2nd AES Workshop on Intelligent Music Production, London, UK, September 13.

De Man, Brecht, Kirk McNally, and Joshua D. Reiss. 2017. Perceptual evaluation and analysis of reverberation in multitrack music production. *Journal of the Audio Engineering Society* 65: 108–16. [CrossRef]

De Man, Brecht, Mariano Mora-Mcginity, György Fazekas, and Joshua D. Reiss. 2014. The open multitrack testbed. Paper presented at the 137th Convention of the Audio Engineering Society, Los Angeles, CA, USA, October 9–12.

De Man, Brecht, and Joshua D. Reiss. 2013a. A knowledge-engineered autonomous mixing system. In *Audio Engineering Society Convention 135*. New York: Audio Engineering Society.

De Man, Brecht, and Joshua D. Reiss. 2013b. A semantic approach to autonomous mixing. Paper presented at the 2013 Art of Record Production Conference (JARP) 8, Université Laval, Quebec, QC, Canada, December 1.

De Man, Brecht, Joshua D. Reiss, and Ryan Stables. 2017. Ten years of automatic mixing. Paper presented at the 3rd AES Workshop on Intelligent Music Production, Salford, UK, September 15.

Deruty, Emmanuel. 2016. Goal-oriented mixing. Paper presented at the 2nd AES Workshop on Intelligent Music Production, London, UK, September 13.

Dugan, Dan. 1975. Automatic microphone mixing. *Journal of the Audio Engineering Society* 23: 442–49.

Eno, Brian. 2004. The studio as compositional tool. In *Audio Culture: Readings in Modern Music*. Edited by Christoph Cox and Daniel Warner. Londres: Continum, chp. 22, pp. 127–30.

Fenton, Steven. 2018. Automatic mixing of multitrack material using modified loudness models. In *Audio Engineering Society Convention 145*. New York: Audio Engineering Society.

Ford, Jon, Mark Cartwright, and Bryan Pardo. 2015. Mixviz: A tool to visualize masking in audio mixes. In *Audio Engineering Society Convention 139*. New York: Audio Engineering Society.

Glaser, Barney, and Anselm Strauss. 1967. Grounded theory: The discovery of grounded theory. *Sociology the Journal of the British Sociological Association 12*: 27–49.

Gonzalez, Enrique Perez, and Joshua D. Reiss. 2007. Automatic mixing: Live downmixing stereo panner. Paper presented at the 10th International Conference on Digital Audio Effects (DAFx'07), Bordeaux, France, September 10-15, pp. 63–68.

Gonzalez, Enrique Perez, and Joshua D. Reiss. 2008. Improved control for selective minimization of masking using interchannel dependancy effects. Paper presented at the 11th International Conference on Digital Audio Effects (DAFx'08), Espoo, Finland, September 1–4, p. 12.

Hargreaves, Steven. 2014. Music Metadata Capture in the Studio from Audio and Symbolic Data. Ph.D. thesis, Queen Mary University of London, London, UK.

Jillings, Nicholas, and Ryan Stables. 2017. Automatic masking reduction in balance mixes using evolutionary computing. In *Audio Engineering Society Convention 143*. New York: Audio Engineering Society.

Jun, Sanghoon, Daehoon Kim, Mina Jeon, Seungmin Rho, and Eenjun Hwang. 2015. Social mix: Automatic music recommendation and mixing scheme based on social network analysis. *The Journal of Supercomputing* 71: 1933–54. [CrossRef]

King, Andrew. 2015. Technology as a vehicle (tool and practice) for developing diverse creativities. In *Activating Diverse Musical Creativities: Teaching and Learning in Higher Music Education*. Edited by Pamela Burnard and Elizabeth Haddon. London: Bloomsbury Publishing, chp. 11, pp. 203–22.

Kolasinski, Bennett. 2008. A framework for automatic mixing using timbral similarity measures and genetic optimization. In *Audio Engineering Society Convention 124*. Amsterdam. Audio Engineering Society.

Martínez Ramírez, Marco A., Emmanouil Benetos, and Joshua D. Reiss. 2019. A general-purpose deep learning approach to model time-varying audio effects. Paper presented at the International Conference on Digital Audio Effects (DAFx19), Birmingham, UK, September 2–6.

Martínez Ramírez, Marco A., and Joshua D. Reiss. 2017a. Deep learning and intelligent audio mixing. Paper presented at the 3rd Workshop on Intelligent Music Production, Salford, UK, September 15.

Martínez Ramírez, Marco A., and Joshua D. Reiss. 2017b. Stem audio mixing as a content-based transformation of audio features. Paper presented at the 19th IEEE Workshop on Multimedia Signal Processing (MMSP), Luton, UK, October 16–18, pp. 1–6.

Martínez Ramírez, Marco A., and Joshua D. Reiss. 2019. Modeling nonlinear audio effects with end-to-end deep neural networks. Paper presented at the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, May 12–17, pp. 171–75.

Merchel, Sebastian, M. Ercan Altinsoy, and Maik Stamm. 2012. Touch the sound: Audio-driven tactile feedback for audio mixing applications. *Journal of the Audio Engineering Society* 60: 47–53.

Mimilakis, Stylianos Ioannis, Estefanıa Cano, Jakob Abeßer, and Gerald Schuller. 2016b. New sonorities for jazz recordings: Separation and mixing using deep neural networks. Paper presented at the 2nd AES Workshop on Intelligent Music Production (WIMP), London, UK, September 13. Volume 2.

Mimilakis, Stylianos Ioannis, Konstantinos Drossos, Andreas Floros, and Dionysios Katerelos. 2013. Automated tonal balance enhancement for audio mastering applications. In *Audio Engineering Society Convention 134*. Rome, Italy. Audio Engineering Society.

Moffat, David, David Ronan, and Joshusa D. Reiss. 2015. An evaluation of audio feature extraction toolboxes. Paper presented at the 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, November 30–December 3.

Moffat, David, and Mark B. Sandler. 2018. Adaptive ballistics control of dynamic range compression for percussive tracks. In *Audio Engineering Society Convention 145*. New York: Audio Engineering Society.

Moffat, David, and Mark B. Sandler. 2019. Automatic mixing level balancing enhanced through source interference identification. In *Audio Engineering Society Convention 146*, Dublin: Audio Engineering Society.

Moffat, David, Florian Thalmann, and Mark B. Sandler. 2018. Towards a semantic web representation and application of audio mixing rules. Paper presented at the 4th Workshop on Intelligent Music Production (WIMP), Huddersfield, UK, September 14.

Muir, Bonnie M. 1994. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics 37*: 1905–22.

Ooi, Kazushige, Keiji Izumi, Mitsuyuki Nozaki, and Ikuya Takeda. 1990. An advanced autofocus system for video camera using quasi condition reasoning. *IEEE Transactions on Consumer Electronics 36*: 526–30.

Pachet, François, and Olivier Delerue. 2000. On-the-fly multi-track mixing. In *Audio Engineering Society Convention 109*. Los Angeles: Audio Engineering Society.

Palladini, Alessandro. 2018. Intelligent audio machines. Paper presented at the Keynote Talk at 4th Workshop on Intelligent Music Production (WIMP-18), Huddersfield, UK, September 14.

Pardo, Bryan, Zafar Rafii, and Zhiyao Duan. 2018. Audio source separation in a musical context. In *Springer Handbook of Systematic Musicology*. Berlin: Springer, pp. 285–98.

Paterson, Justin, Rob Toulson, Sebastian Lexer, Tim Webster, Steve Massey, and Jonas Ritter. 2016. Interactive digital music: Enhancing listener engagement with commercial music. *KES Transaction on Innovation In Music 2*: 193–209.

Pestana, Pedro Duarte, Zheng Ma, Joshua D. Reiss, Alvaro Barbosa, and Dawn A. A. Black. 2013. Spectral characteristics of popular commercial recordings 1950-2010. In *Audio Engineering Society Convention 135*. New York: Audio Engineering Society.

Pestana, Pedro Duarte Leal Gomes. 2013. Automatic Mixing Systems Using Adaptive Digital Audio Effects. Ph.D. Thesis, Universidade Católica Portuguesa, Porto, Portugal.

Prior, Nick. 2012. Digital formations of popular music. *Réseaux* 2: 66–90. [CrossRef]

Reiss, Joshua, and Øyvind Brandtsegg. 2018. Applications of cross-adaptive audio effects: Automatic mixing, live performance and everything in between. *Frontiers in Digital Humanities* 5: 17. [CrossRef]

Reiss, Joshua D. 2011. Intelligent systems for mixing multichannel audio. Paper presented at the17th International Conference on Digital Signal Processing (DSP), Corfu, Greece, July 6–8, pp. 1–6.

Reiss, Joshua D. 2018. *Do You Hear What I Hear? The Science of Everyday Sounds*. Inaugural Lecture. London: Queen Mary University.

Rödel, Christina, Susanne Stadler, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. Paper presented at the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, September 17–19, New York: ACM, pp. 1–8.

Ronan, David, Brecht De Man, Hatice Gunes, and Joshua D. Reiss. 2015. The impact of subgrouping practices on the perception of multitrack music mixes. In *Audio Engineering Society Convention 139*. New York: Audio Engineering Society.

Ronan, David, Hatice Gunes, and Joshua D. Reiss. 2017. Analysis of the subgrouping practices of professional mix engineers. In *Audio Engineering Society Convention 142*. New York: Audio Engineering Society.

Ronan, David, Zheng Ma, Paul Mc Namara, Hatice. Gunes, and Joshua D. Reiss. 2018. Automatic minimisation of masking in multitrack audio using subgroups. *arXiv* arXiv:1803.09960.

Ronan, David, David Moffat, Hatice Gunes, and Joshua D. Reiss. 2015. Automatic subgrouping of multitrack audio. Paper presented at the 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, November 30–December 3.

Russell, Stuart J., and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*, 3rd ed. Malaysia: Pearson Education Limited.

Schmidt, Brian. 2003. Interactive mixing of game audio. In *Audio Engineering Society Convention 115*. New York: Audio Engineering Society.

Scott, Jeffrey J., and Youngmoo E. Kim. 2011. Analysis of acoustic features for automated multi-track mixing. Paper presented at the 12th International Society for Music Information Retrieval Conference (ISMIR), Miami, FL, USA, October 24–28, pp. 621–26.

Selfridge, Rod, David Moffat, Eldad J. Avital, and Joshua D. Reiss. 2018. Creating real-time aeroacoustic sound effects using physically informed models. *Journal of the Audio Engineering Society* 66: 594–607. [CrossRef]

Sheridan, Thomas B., and William L. Verplank. 1978. *Human and Computer Control of Undersea Teleoperators*. Technical Report. Cambridge: Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Smolka, Bogdan, K. Czubin, Jon Yngve Hardeberg, Kostas N. Plataniotis, Marek Szczepanski, and Konrad Wojciechowski. 2003. Towards automatic redeye effect removal. *Pattern Recognition Letters 24*: 1767–85.

Stasis, Spyridon, Nicholas Jillings, Sean Enderby, and Ryan Stables. 2017. Audio processing chain recommendation. Paper presented at the 20th International Conference on Digital Audio Effects, Edinburgh, UK, September 5–9.

Stevens, Richard, and Dave Raybould. 2013. *The Game Audio Tutorial: A Practical Guide to Creating and Implementing Sound and Music for Interactive Games*. London: Routledge.

Terrell, Michael, and Mark Sandler. 2012. An offline, automatic mixing method for live music, incorporating multiple sources, loudspeakers, and room effects. *Computer Music Journal 36*: 37–54.

Terrell, Michael, Andrew Simpson, and Mark Sandler. 2014. The mathematics of mixing. *Journal of the Audio Engineering Society* 62: 4–13. [CrossRef]

Van Waterschoot, Toon, and Marc Moonen. 2011. Fifty years of acoustic feedback control: State of the art and future challenges. *Proceedings of the IEEE 99*: 288–327.

Verfaille, Vincent, Daniel Arfib, Florian Keiler, Adrian von dem Knesebeck, and Udo Zölter. 2011. Adaptive digital audio effects. In *DAFX—Digital Audio Effects*. Edited by Udo Zölter. Hoboken: John Wiley & Sons, Inc., chp. 9, pp. 321–91.

Verfaille, Vincent, Udo Zolzer, and Daniel Arfib. 2006. Adaptive digital audio effects (A-DAFx): A new class of sound transformations. *IEEE Transactions on Audio, Speech, and Language Processing* 14: 1817–31. [CrossRef]

Verma, Prateek, and Julius O. Smith. 2018. Neural style transfer for audio spectograms. Paper presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017), Workshop for Machine Learning for Creativity and Design, Long Beach, CA, USA, December 4–9.

Vickers, Earl. 2010. The loudness war: Background, speculation, and recommendations. In *Audio Engineering Society Convention 129*. San Francisco: Audio Engineering Society.

Wichern, Gordon, Aaron Wishnick, Alexey Lukin, and Hannah Robertson. 2015. Comparison of loudness features for automatic level adjustment in mixing. In *Audio Engineering Society Convention 139*. New York: Audio Engineering Society.

Wilson, Alex, and Bruno Fazenda. 2015. 101 mixes: A statistical analysis of mix-variation in a dataset of multi-track music mixes. In *Audio Engineering Society Convention 139*. New York: Audio Engineering Society.

Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. 2016. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann.

Xambó, Anna, Frederic Font, György Fazekas, and Mathieu Barthet. 2019. Leveraging online audio commons content for media production. In *Foundations in Sound Design for Linear Media*. London: Routledge, pp. 248–82.