Chapter 13: Artificial Intelligence Music Mixing Systems

David Moffat
Interdisciplinary Centre for Computer Music Research,
University of Plymouth

13.1 Introduction

Mixing music, or music production, is the process of combining a series of different musical tracks together, while applying a range of audio processing to *blend* the tracks together in a pleasance and aesthetically pleasing way. Music mixing practices require the pleasant combination of all aspects of a musical piece, using a set of engineering tools to do so. Owsinski (2013) identified five key components to consider while making a mix, and they are

- Balance;
- Spectral;
- Spatial;
- Depth; and
- Dynamics

The balance is related to ensuring that all instruments can be heard within a mix, and that none are monopolising the mix, in terms of volume. Spectral is related to the frequency content on the mix, to ensure that this is balanced, and there is not too much weight on a particular frequency component, and that the frequency components of each sound source are distinct and clear. The spatial balance is to represent a good image of the sound between two ears, allowing for better differentiation between all the different sources, without one side sounding louder than the other. The depth, is to ensure that each sonic element has enough interest and complexity on its own, and that the richness of each sound can be clearly heard, while combining together to produce a pleasant rich timbral sound. The dynamics are related to the sudden transient nature of the sound, to ensure there is ample and adequate change in the overall, and individual, volumes of the tracks, by creating quieter parts of the music, the sounder parts can stand out more, creating interest and an evolving sonic signature over time.

These five dimensions of a mix are typically controlled through a range of different audio effects. The way in which a piece of music is mixed, can heavily influence the way in which is is perceived, in terms of preference (De Man et al., 2015), perceived quality (Wilson and Fazenda, 2013), and in the evoked emotion (Ronan et al., 2018b; Scherer et al., 2001). Music mixing is a highly complex, multi-dimensional problem, where a number of different complex sounds are combined in a multitude of different ways. The processing and modification of each and every track depends on all other tracks within the musical mixture, and often requires different processing and effects in different sections of the song. Some equaliser setting applied in the chorus maybe very different to the equaliser setting required in the verse. This results in a highly complex non-linear search space, relying heavily on human perception and preference of music, along with the limitations of human hearing and emotional responses to music.

The integration of an artificially intelligent music mixing system, or intelligent mixing system (IMS), has the capacity to change music production workflows and approaches (Wilmering et al., 2020). The use of an IMS can change the way in which a mix engineer can explore through

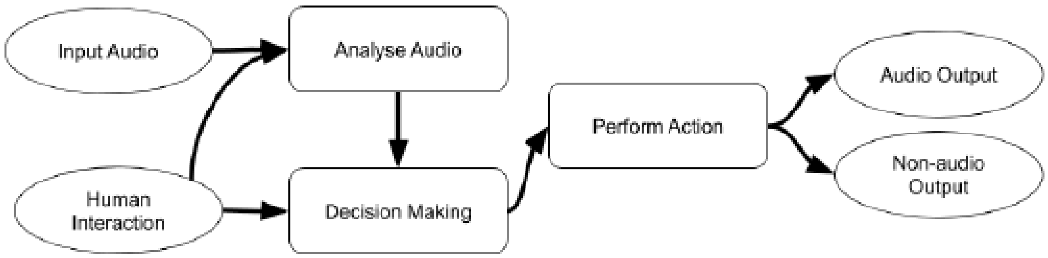


Figure 13.1: The typical structure of an IMS

the vast array of mixing options available to them, they could use this as an opportunity to reduce the dimensionality of the music mixing problem, controlling their path through the mixing environment, and could even use an IMS as a tool for collaboration, where they are both enacting control over the musical mix as a whole. The use of an IMS will even inform and influence professional practice (Bromham, 2016).

Historically, music mixing has grown and developed over time, constantly using new technology and practice to create new and interesting music (Burgess, 2014), and some musical trends and music styles are as a result of technology, rather than just cultural evolutions (Mauch et al., 2015). Some genres of music, such as *techno* and *acid-house*, were created as a result of technological innovations (Bull, 1997). A large proportion of these technological innovations are as a result of borrowing, using, or misappropriating technology from other fields, and applying them to music mixing processes (Wilmering et al., 2020). The culture of misusing and misappropriating technology, within music, has been prevalent throughout the history of music (Prior, 2012), as a way to be creative, explore new approaches and uncover the opportunities within different musical spaces. The use of IMS, in music mixing, brings a number of opportunities, not just for the new technology to be used as intended, but also for practitioners to take this new technology, use and misuse it, and explore the expressive opportunities it affords. New technologies have the advantages and opportunity to lead to new approaches for music production (King, 2015). This could either be intentionally, through understanding of how a tool works in one domain, and applying it to a music mixing context, or this could be accidental. It is well reported in music production, that often trying new things and exploring, even making mistakes can result in *happy accidents* which has resulted in many of the mixing practices that are common place today (Cascone, 2000). This was best summed up by Bromham, who said"Some of the most creative moments in recording have come from accidents where devices have been plugged up incorrectly and technology misappropriated" (Bromham, 2016).

There are numerous approaches for developing an IMS. In principal there are a few key aspects of an IMS that are necessary to consider. Figure 13.1 shows a typical IMS structure. There must be some system for parsing audio tracks, and potentially human input, then some interpretation and responce to the audio tracks input, which should directly modify and manipulate the audio, or present some representation of the audio, which can benefit the mix engineer. In the field of theoretical artificial intelligence, an *intelligent agent* is one which has three key components, the ability to to *observe* or perceive the environment, the ability to *act* upon the environment, and *decision making* capacity to achieve a desired goal (Russell and Norvig, 2016). This identifies three key aspects of an IMS, which will be discussed within this chapter. The three aspects of an IMS are:

Decision Making Process — The process of the IMS analysing the inputs, and using this to make some mixing decision. This process includes representing all the musical knowledge and concepts, creative and technical decisions, along with understanding why the decision is made.

Audio Manipulation — The way in which an IMS will act upon the world, how it can interact with the world and the tools it is provided with to have an impact.

Human Computer Interaction — The observations made of the wider environment, the way in which an IMS will have utility to any user, and how the tools can be used.

13.2 Decision Making Process

The decision making process, is arguably one of the most challenging components of an IMS. The ability to capture the concept of a creative decision that a mix engineer may make, or to understand the reasoning behind a single decision being made, is a challenging approach, and to embed this concept into an IMS can be even more challenging. The idea of modelling some *knowledge* of musical mixing, and using that to perform actions later on, is one of the key aspects of any intelligent system. In the case of music mixing, this could be through some domain knowledge, learning or defining a rule, such as "We want the vocals in this track to be really clear", or these rules could be implicitly learned, through the analysis of data collections. De Man and Reiss (2013b) identified three different approaches for modelling the decision making process in IMS; knowledge encoding; expert systems; and data driven approaches.

13.2.1 Knowledge Encoding

In most IMS literature, this approach is referred to as *Grounded Theory* (De Man and Reiss, 2013a). Grounded theory approach is a formalised approach taken within social sciences, where theories are create though systematic methodological collection and analysis (Glaser and Strauss, 1967). Many IMS approaches take an informal grounded theory inspired approach, without following the systematic practices of a formal grounded theory approach.

The knowledge encoding approach is to formalise the understanding of the mixing process. There are a number of different approaches taken to gather *knowledge* of mixing practices. Ethnographic studies can be conducted, as a formal framework analyse and understand the practice of mixing engineers (Cohen, 1993; McGrath and Love, 2017). Interviews and surveys can be conducted (Ronan et al., 2017; Pestana and Reiss, 2014), which can provide insight into how mix engineers state that they approach mixing problems. Often this can be verified though the analysis of mixing practices (Pestana, 2013; De Man et al., 2016) Published literature by respected practitioners (Izhaki, 2008; Senior, 2012; Owsinski, 2013) can often be a useful way to gain a better understanding of mixing processes.

Often practitioners experience, coupled with rules derived from literature, can be used to automate specific audio effects independently. There are a large number of studies looking into perceptual attributes of mixing production practice. Hermes (2019) performed an overview of mixing approaches, and focused on understanding spectral clarity for automatic mixing. Bromham et al. (2018) conducted a study to understand which compressor settings would be deemed appropriate for a given piece of programme material. Bromham et al. (2019) looked to understand how different audio effects would influence the perception of timbral attributes of a piece of music, including brightness and warmth. Weaver et al. (2019, 2018) investigated the impact of reverberation on how musicians perform together, which was further analysed by De Man et al. (2017). Fenton and Lee (2015) investigated the perceptual attribute punch, within a music mixing context, where as Moore (2020) investigated how aggressive a distortion effect can be. Both Wilson and Fazenda (2015) and Colonel and Reiss (2019) performed statistical analyses of a large numbers of musical mixes.

These inferred rules can then be applied into IMS. Perez Gonzalez and Reiss (2009b) proposed setting the gains of all audio tracks to the same perceptual loudness within a mix, and Moffat and Sandler (2019c) proposed including a source separation evaluation metric to compensate for crosstalk between different microphones in a live situation. Perez Gonzalez and Reiss (2010) sets the pan of some audio tracks to reduce spectral interference of different tracks, while maintaining the low frequency content as close to the centre as possible, and Ward et al. (2012) extends this to use a perceptual model of masking to place each track within the stereo field. Perez Gonzalez and Reiss (2009a) equalised tracks to reduce the spectral overlap of audio tracks, where Hafezi and Reiss

(2015) used a perceptual masking model to define the equalisation values. Maddams et al. (2012) automated the parameters of a dynamic range compressor, based on signal analysis, to consistently set the dynamic range of audio tracks. Moffat and Sandler (2018) identified that mix engineers will often use dynamic range compressors for a number of different uses, and developed a parameter setting for emphasising transients of drums. Moffat and Sandler (2019b) automated reverberation parameters, where the reverb time is controlled by the tempo of the audio track (Weaver et al. 2018).

It is clear here, that most approaches undertaken will only automate one type of audio effect, and will typically restrict themselves to a simple set of rules. Any more complex rule structures require more complex management of multiple conflicting rules, such as that described in Section 13.2.2. Throughout this approach, there is a necessity to consistently update the collection of mixing approaches, and to evaluate the approaches taken to implement them. Especially as it has been demonstrated that professional mix engineers may identify one approach whilst actually using an alternative approach (Pestana, 2013). This could easily lead to cases where an IMS approach is well intentioned, but never able to produce effective results.

The knowledge encoding approach is critical to understand the *human* approach to mixing. This could be ideal as a training system, where simplified use-cases could be given to an individual, to demonstrated isolated concepts or approaches. However, the ability to combine all these approaches together, create a much larger set of problems, where approaches will contradict each other, and there will be differences of opinion in the mixing approach. There are also numerous examples of *happy accidents*, where something is done accidentally, which results in producing a preferable mixing result, typically through breaking the rules, rather than confirming to existing rules (Cascone, 2000). This is both acknowledged and embraced by many practitioners, and some mix engineers embrace this approach (Eno and Schmidt, 1975).

13.2.2 Expert Systems

Expert system is the approaching where a human expert decision-making process is modelled by a computer system. The computer model is often more generally called a knowledge based system. Expert systems are designed to approach problems, through understanding the problem, and then representing a typical expert approach using a series of *if-then-else* rules. Expert systems are broken up into two different components sections: the knowledge based, and the inference engine. The knowledge base, is where a series of facts and rules can be stored, and the inference engine will then utilise these rules to make deductions and suggestions.

In the case of IMS, there have been a number of different expert system based approaches. An expert system can either explicitly state a set of rules, and use these to then perform some inference, or optimise towards a given result define rules and perform inference (Moffat et al., 2018; Benito and Reiss, 2017), or mixing can be set up as a constraint optimisation problem, where a series of goals are defined, in a mathematical form, and the system must perform a search for the mathematically optimal solution (Terrell and Sandler, 2012; Terrell et al., 2014).

The rule/inference approach can be used as an effective way to build on the grounded theory approaches outlined in Section 13.2.1. Specific rules are developed and coded, such as discussed by De Man and Reiss (2013a), which can then be applied to a given problem. Pachet and Delerue were the first to identify that musical mixing could be defined as an inference or optimisation problem. Pachet and Delerue (2000) constructed a full mixing system, based on sound source spatialisation and mixing, though placing individual musical sources in a sonic space, defining a series of rules, and allowing the inference engine to perform the mixing task. Deruty (2016) developed a range of high level mixing goals which should be achieved during the mixing process. Benito and Reiss (2017) constructed a probabilistic soft logic engine to apply reverb to a musical mix. Rules were collected from grounded theory approaches, coded into a logical inference engine, and applied to different musical tracks. The author notes the challenges in translating grounded theory rules into probability weighting. Moffat et al. (2018) created a generalised framework for constructing musical mixing rules to be applied to an inference engine, and suggests that there is potential to learn mixing rules from data, utilising the semantic web (Berners-Lee et al., 2001).

Mathematical optimisation approaches have also been effectively demonstrated in the music production field. Barchiesi and Reiss (2009) proposed setting the gain and equalisation parameters to mix towards a given reference track. Kolasinski (2008) performed an optimisation approach to mixing a series of tracks to match a given same timbre of some selected reference track. The timbre is defined using a spectral histogram, and only gains of different tracks could be adjusted to match the reference track. Gang et al. (2010) used timbre and a range of musical score based features to optimise a number of audio effect parameters towards a given reference track. A range of mixing targets have been used, such as mixing to a specific targets loudness (Fenton, 2018; Wichern et al., 2015), using a perceptual model of masking, to minimise the inter-track masking (Ronan et al., 2018a; Jillings and Stables, 2017b), as this is often considered a negative effect of track interference, or optimising to reduce a number of different objective measures (Terrell et al., 2010). Terrell and Sandler (2012); Terrell et al. (2014) investigated music mixing in a live music context, optimising the layout of different sources and speaker to counteract for room effects. Pestana et al. (2015) optimised the phase offset of each instrument track, to minimise the comb filtering effects of phase cancellation. Wilson and Fazenda (2016) proposed a *human-in-the-loop* mixing approach, where a human is able to state a preference over a set of mixes, which is used, in turn to generate more mixes, in the hope that a "personal global optima" (Wilson and Fazenda, 2016) is found.

There is also a variety of different optimisation approaches that have been taken, linear approaches, such as least squares (Barchiesi and Reiss, 2009; Terrell et al., 2014) or genetic inspired approaches such as genetic algorithms (Jillings and Stables, 2017b; Kolasinski, 2008), or particle swarm algorithms (Ronan et al., 2018a). Wilson et al. (2017) discusses the use of genetic algorithms, compared to other expert system approaches, in creating intelligent music systems.

Expert systems benefit from their ability to model highly complex rule structures, that are ever growing, with multiple target objectives, and aims to always find a solution that has the ability to fulfil as many of the targets as is possible. These systems are able to consider each and every rule, in turn, and identify when certain rules need to be broken or ignored, in order to produce the best overall system. The ability to create hard constraints and soft constraints, such that an IMS can navigate a mix-space, following an individuals intention.

The rule based mixing approaches present considerable power, as the ability to produce a formalised approach to construct, compare and evaluate formal mixing rules, in a simple structure could prove to be very powerful. As there are many cases where a mix engineer may give a rule that they follow, there are examples where mix engineers will say one thing, but do another (Reiss, 2018). This could be because an engineer does not objectively understand exactly what their mixing process is or that they feel a need to justify their approach. The formal and consistent evaluation of a range of mixing rules, through a quantitive approach, would be highly insightful into both a better understanding of mixing practice, and assist greatly in developing state of the art intelligent mixing systems.

The key encapsulating factor of an optimisation approach derived IMS, is contained within the *fitness function*. This is the component of the optimisation that defines what to prioritise, and how it should be evaluated. These fitness functions have been used, in optimisation approaches, to reveal greater understanding of the auditory system (McDermott and Simoncelli, 2011), perceptual similarity measures (Moffat and Reiss, 2018), adjusting synthesiser parameters (Garcia, 2001) and for musical composition (Miranda and Al Biles, 2007). The fitness function is required to encapsulate all the understanding and knowledge that the experts have, and how it can be applied to the mixing problem at hand.

A review of expert systems, and how they were applied to IMS, was performed by De Man and Reiss (2013b), where the challenges in defining rules for IMS are identified. The inherent complexity of music mixing means that, there is no certain *optimal solution*, but a number of different *appropriate mixes* given a set of contexts (Jillings and Stables, 2017b). There are a number of different mixes that are preferred by different individuals, in different moods, at different times. Mixing has the ability to change and transform a piece of music (Ronan et al., 2018b), and so any set of constraints defined would need to acknowledge this and take this into consideration while defining the rules to be applied (Lefford et al., 2020a).

Inherent to how optimisation approaches work, it is not possible for most of them to operate in real-time, and as such, they need to be seen as tasks where an entire track is given to an IMS, and the mix is produced at a later date. This can severally limit the ability for an individual to interact with the music mixing system, as this would not integrate well with traditional music production studio workflows. Expert based system relies on the assumption that experts will make consistent, agreeable decisions. This implies that experts should be considered to be time invariant - that a engineer who applies a given equaliser setting today, would apply the same equaliser setting tomorrow, or next year. There have been a number of cases where expert systems have been demonstrated to be highly effective AI approaches (Nelson et al., 1982; Rasmussen, 1990), however, there are few cases where these approaches have been demonstrated to creative approaches with great effect.

13.2.3 Data Driven

Data driven IMS approaches have been developing in recent years, particularly with the growth in machine learning and neural network techniques. These approaches rely on analysis on datasets or lots of example mixes and use this to extrapolate some set of mixing parameters. This is commonly done by selecting a set of relevant audio features, or audio descriptors (Moffat et al., 2015), typically designed to represent some semantic or perceptual attributes (Stables et al., 2014), and discover how these can be related to a specific mixing decision (Martínez Ramírez and Reiss, 2017b; Martínez Ramírez et al., 2019). Reed (2000) first proposed a data driven mixing approach, where data was analysed as to how the frequency band energy can influence the timbral attributes of brightness, darkness and smoothness, and this was used to automatically equalise a given audio track to an identified semantic term, using a nearest neighbour algorithm. Since then, machine learning approaches have grown considerably, and there many approaches for using a data driven approach to construct an IMS.

Deep learning approaches have become very relevant recently, since it was demonstrated that a neural network has the ability to parse 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). Moffat and Sandler (2019d) extracts gain parameters from a series of audio mixes, using a *reverse engineering the mix* approach, developed by Barchiesi and Reiss (2010), and then uses this to predict gain parameters and extrapolate to larger datasets, using a random forest approach. Pestana et al. (2013) analysed 60 years of the UK and USA pop chart music, and then Ma et al. (2013) used this to predict an *ideal* equalisation curve, which can be applied to different tracks. Martínez Ramírez et al. (2020a) and Sheng and Fazekas (2019) both generated a set audio samples modified with the use of a dynamic range compressor, and then learned the transformations applied by that compressor. Hestermann and Deffner (2020) took on the task of manually annotating a large dataset of audio tracks, to develop and intelligent de-esser. Chourdakis and Reiss (2016, 2017) developed an approach for learning reverberation parameters from a specific user input, which then extrapolates the selected reverb parameters to other tracks, though they comment on the challenges of finding appropriate quantities of data. Mimilakis et al. (2016a) constructed a neural network to learn the mastering process of jazz music, taken from the Jazzomat dataset (Jazzomat, 2018). Martínez Ramírez et al. (2020b) recently demonstrated a full end-to-end IMS, using drums. This system learns the full music production process of drum mixing and demonstrates that the intelligently produced mix is indistinguishable from the professional engineer generated mix.

Clearly, one of the largest restricting factors within a data driven approach for intelligent mixing, is the data gathering. Other than taking on large scale manual annotation approaches, there are numerous approaches that have been taken, such as using a mix parameter reverse engineer approach (Barchiesi and Reiss, 2010), as used by Moffat and Sandler (2019d). There are also a number of curated multitrack datasets, including the Cambridge MT Multitrack Dataset (Senior, 2011), the Open Multitrack Dataset (De Man et al., 2014), MedleyDB (Bittner et al., 2014). There are also instrument specific instrument dataset, such as the ENST Drum dataset (Gillet and Richard, 2006). Despite limitations, data driven approaches are highly effective results, once suitable datasets are curated (Arel et al., 2010).

The considerable growth in data science and machine learning approaches over the past decade (Witten et al., 2016), has resulted in their being ample directions for further work in data driven IMS. It has been demonstrated that data driven approaches are highly effective, and extendable, leaving considerable opportunities for future work in this space. The input track ordering is a present challenge, where any machine learning approach should be able to mix tracks together in a way that is invariant to the input track order. As the number of musical tracks grows, the problem search space becomes exponentially more and more complex, which can lead to real challenges. And networks need to be able to deal with missing instruments, eg. some tracks will have a brass section, or a violin track, but many will not, as Bittner et al. (2014) identify that within their dataset only two of the sixteen instruments exist in more than 50% of the multi-tracks.

13.2.4 Decision Making Summary

The decision making processes used by an IMS are of critical importance to both the inputs needed to perceive the system, and to how an IMS will operate. The decision making process encodes all the knowledge of a given system, and will enact some decision to the action component of the IMS. The input system is greatly influenced, as some approaches require a single audio track input, where as some require a full multitrack mixing context. Furthermore, if any additional metadata is required, or human interaction to involve in the decision making process. Knowledge encoding approaches rely heavily on professional mix engineer approaches and understanding, and approaches to try and represent this domain knowledge in a simple direct way, however they are highly restricted to often controlling single audio effect directly, and so do not model the interaction between different audio effects or processing chains. Expert systems attempt to quantify the uses more formally, performing some inference or optimisation based on these rules, which allow for considerably more complex rules. Conversely, data driven systems have recently shown that they are able to mix as effectively as a professional engineer(Martínez Ramírez et al., 2020b). Though this is in a simplified mixing task, mixing only drums, these results are highly promising for future research. The future of IMS can also lie heavily in expert systems. Through combining knowledge approaches, and learning defined rules from data, this approach could both be used to gain insight into music mixing practices and approaches taken, but also develop state of the art IMS. This will only be possible once data collection challenges within the data driven approaches have been addressed. The decision making process will then provide a decision, and an action will be taken to some change to the audio, as discussed in the following section.

13.3 Audio Manipulation

The manner in which an IMS will perform an action upon a piece of audio, will greatly influence the limitations and restrictions of the IMS, and will also emphasise the opportunities of the IMS. How a piece of audio is modified will greatly define and limit the IMS. There are two approaches to modify audio with an IMS, and they are either to use adaptive audio effects, or to perform a direct transformation on the audio.

13.3.1 Adaptive Audio Effects

The use of audio effects, to construct an IMS is the most common approach taken. Audio effects are the process used to manipulate and change a sound in an intentional manner. Audio effects have been around for as long as we have had documented music (Wilmering et al., 2020). In principal, an audio effect can be any sounds modification, from as simple as a loudness control, to a dereverberation algorithm (Naylor and Gaubitch, 2010) or noise removal algorithm (Mat et al., 2010). These audio effects are used as part of the music making, mixing and production, to shape and control the musical sounds. The principal aspect of this is to ensure that control can be harnessed over a pre-existing sound. An audio effect is the approach of taking a sound and modifying it, in some consistent predictable, and usually controllable, manner. An adaptive audio

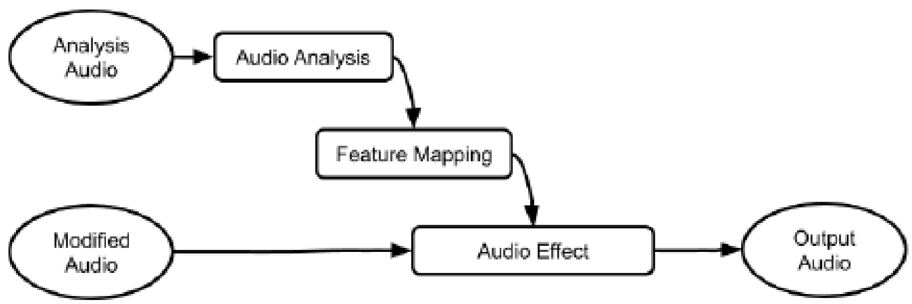


Figure 13.2: A flow diagram of a typical adaptive audio effect.

effect (AFx) is one in which, control parameters are changed over time, based on either analysis of audio, or an external sensor input, such as a gestural control.

Figure 13.2 shows the general structure of an AFx. Verfaille et al. (2011) developed a classification for AFx, which identifies the following AFx categories:

Auto-adaptive An effect where the audio analysis is performed on the input signal that is also being modified, as shown in Figure 13.3a. The AFx adapts directly to the audio signal being used. An example of this could be a dynamic range compressor.

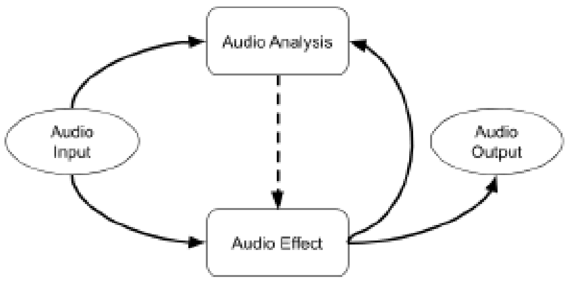
External adaptive An effect where the control analysis input is presented from something external, such as an alternative audio stream, or gestural input, as shown in Figure 13.3b, for example a side-chain compressor.

Cross-adaptive Cross-adaptive AFx, is where two different audio tracks are used to modified each other directly, where both the audio samples, and the AFx interact, potentially conflict with each other, and typically reach some equilibrium state, which changes as the audio channels progress, as shown in Figure 13.3c.

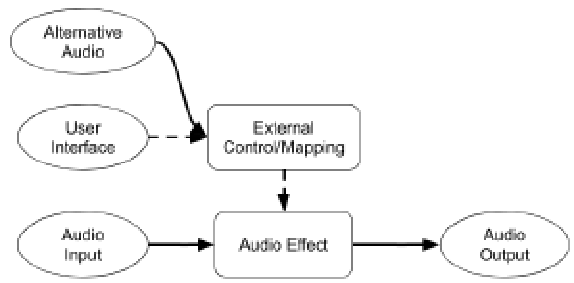
The use of adaptive audio effect for IMS was formalised by Reiss (2011). The audio effect is some signal processing block which modifies audio, and is adaptive in some way. The feature mapping or parameter automation is performed, so that the IMS can directly control a parameter, much in the same way a human engineer would. Adaptive effect implementation is performed in a number of different ways, depending on the type of audio effect being used (Verfaille and Arfib, 2002), and this model can be applied either directly to perceptual attributes of a piece of music (Holfelt et al., 2017), or for individual performed to be able to interact with each other in more musical ways (Sarkar et al., 2017).

AFx are used within IMS to automate pre-existing audio effect parameters (Reiss and Brandtsegg, 2018). This has been a common approach for some time, and intuitively it makes sense to maintain as much of the processing chain as constant when developing IMS. There are many auto-adaptive audio effects, such as dynamic range compressors, that are not considered to be intelligent, but do rely on some analysis of an audio signal to automate some internal parameters. Even cross-adaptive effects, such as a side-chain compressor, are considered *advanced mixing techniques* but not intelligent. The AFx approach to IMS allows for an easy and intuitive transition from IMS to traditional music production paradigms, and there are clear opportunities for how to interact with, or expose the parameters automated to a user, such as creating an *adaptive preset* (Paterson, 2011), through parameter setting recommendation, whether some or all of the settings, or by operating a fully automatic plug-in, which fits with typical mixing workflows.

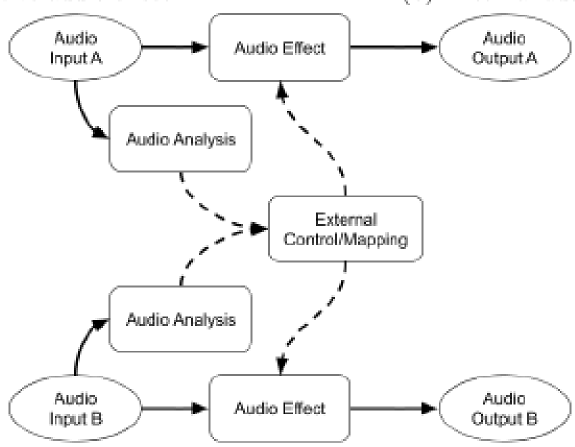
The use of adaptive audio effects within an IMS introduces a number of challenges. Primarily, that there are a large number of audio effects that can be applied in any order, to achieve a number of different goals. To this end, there are a number of approaches to analyse and propose an audio effect chain. McGarry et al. (2017) performed an ethnographic study into the music mixing processes that are undertaken in studios. Sauer et al. (2013) used natural language parsing and a range of semantic terms to define a target, which is analysed to propose a suggestion as to



(a) Auto-adaptive audio effect



(b) External adaptive audio effect



(c) Cross-adaptive audio effect

Figure 13.3: Flow diagrams of the three primary different types of adaptive audio effects.

the types audio processing that should be applied to a given audio track. Stasis et al. (2017a) conducted a study, evaluating the use and ordering of audio effects on a range of different audio tracks, in an attempt to understand the types of audio processing chains that are commonly used. This work was developed further, and related to semantic descriptors (Stasis et al., 2017b). Moffat and Reiss (2020) presented a review of semantic approaches to music production, with a focus on audio effect manipulation of semantic attributes.

There are, a large number of restriction of using AFx in IMS. The limitations on how the audio can be manipulated and changed is highly limiting, there may be a number of cases where a specific target is wanted, but the IMS is not able to understand how to achieve the desired outcome. This is often a problem with student engineers, who may know what they want to achieve, but do not know how to achieve it (Katz and Katz, 2003). Furthermore, constructing an independent IMS for a single audio effect will greatly limit the opportunity for that IMS to understand the complexity of the impact it may have on the signal, and a later IMS in the same chain may then be attempting to undo whatever that IMS is performing. As such, constructing a global IMS, with an understanding as to the overall musical context will have considerably greater power.

13.3.2 Direct Transformation

Alternative to using AFx, there have recently been a number of approaches which have demonstrated some sort of direct audio transformation. Instead of using some mid-level audio interaction algorithm, which are typically based on some electronic circuit or mechanical system (Wilmering et al., 2020), the audio can be modified directly. This can be performed either though direct audio sample modification, as performed in end-to-end learning (Martínez Ramírez et al., 2020b), or modifying the audio in some reversible domain, such as the short-time Fourier transform (Mimilakis et al., 2016a; Martínez Ramírez and Reiss, 2017a). The ability to directly transform audio is common in the field of neural networks, which have demonstrated the ability to directly modify audio for a range of different tasks, including intelligent mixing (Martínez Ramírez et al., 2020b), audio style transfer (Grinstein et al., 2018; Verma and Smith, 2018), audio effect modelling (Martínez Ramírez et al., 2020a; Hawley et al., 2019), signal denoising (Nercessian and Lukin, 2019) and even in sound synthesis (Gatys et al., 2015). Fundamentally, all of these approaches, whether called denoising, style transfer or timbre modification, are each audio effect, in one way or another.

An IMS developed with a direct transformation capacity would not be limited to traditional human approaches of modifying audio, instead able to learn to produce the desired effect, regardless of how possible or easy that would have been through modelling in electronic circuitry (Wilmering et al., 2020). There is no reliance on some complex non-linear mapping between audio attributes, perceptual attributes and audio effect parameters, and instead the important transformation, and ordering, are implicitly understood.

There are clear opportunities for a neural network approach to learn some signal processing transform, based on a given dataset, which could be used to create or realise new audio effect approaches or ways for interacting with audio, both from a creative and an engineering approach. This has the potential to allow for some new, interesting and highly meaningful audio effects to be created. Instead of an engineer having trouble with masking, and thus selectively using a panner, an equaliser and a compressor, they could instead load up their *demasker*, which could provide some composite tool consisting of aspects of all the individual audio effects that are relevant, and allow dynamic effect ordering where appropriate with a simple high level user interface to control. This system would not need to have any intelligent control systems, but would be an IMS, as the use of the AI technology to develop the audio effect would be the intelligence within the system.

The impact to both the traditional studio paradigm and the opportunities present in constructing an IMS are considerable, both easing the ability to shape audio as intended, but also for the creation of new audio effects, from both a technical and creative perspective. The direct transformation paradigm has yet to be fully explored, and new opportunities are being regularly coming to light, but none the less, the opportunities for the creation of new audio effects, and the recent outcomes suggest that the direct transform method has considerable opportunity for exploration and innovation.

13.3.3 Audio Manipulation Summary

The audio modification approach selected, while constructing an IMS, will influence a number of factors of the ISM. An AFx approach will modify existing known audio effect parameters in an intuitive *human-like* manner. This fits directly into the understanding of music mixing, and it can be believe that this approach would swiftly be taken up by practitioners. There are, however, a number of limitations of this approach. Music is a complex system, and the large multichannel signal modification required, with several different mapping layers and understanding the intercorrelation between different audio effect parameters, audio effect ordering in the signal chain and processing approaches makes this a highly complex and non-linear search space. It will take a human mixer years of experience to gain an intuition as to what effect ordering, and parameter modification will a given change to the audio, both on a single audio track and relative to the overall musical mix.

Conversely, when a direct modification IMS is implemented, the system would be able to learn exactly the transform required for each stage of the process, in a much more interpretable way, and directly implemented. It would even be possible for this direct transformation to be developed and framed in the more traditional audio effect domain, this allowing for easy integration into preexisting music mixing workflows. The ability for a direct transform IMS to create novel, interesting and creative audio effects processors can produce insights into existing mixing practices, and is highly advantageous to expert and amateur mix engineers alike.

13.4 Human Computer Interaction

There are a number of reasons for constructing an IMS. It could be that one is looking to completely automate the music mixing process, either due to budget constraints, or there is no way that a human can possibly be in the position to mix the piece of music, such as in a video game (Schmidt 2003). Alternatively, the aim could be to use an IMS to develop some form of technology that can help a mix engineer in a live conditions, such as microphone bleed removal (Clifford and Reiss 2013; Moffat and Reiss, 2016). The aim could be to provide a mix engineer with more insight into the mix, through some form of visualisation (Ford et al., 2015), or to be used as an educational tool (Lefford et al., 2020b). Based on the aim of developing a given IMS, it is vital to design and build an IMS, in acknowledgement of how it can interact with a human mix engineer.

In fields outside of music production, the introduction of IMS to provide task automation has been prevalent for decades (Sheridan and Verplank, 1978; Lindsay et al., 1993), however, the vast majority of these approaches relate to automating heavy industry, where working conditions are slow and dangerous, or on production lines, where repetitive tasks are performed in a highly repetitive manner, under the supervision of a human operator (Furman and Seamans, 2019; Bolton et al., 2018). In these cases, there are clear advantages to using an AI system, either to speed up the process, reduce the risk to workers or to maintain a 24 hour production cycle and vastly increasing the production outputs.

However, in music mixing, the purpose of an IMS would only ever be to act as a tool, for a practitioner to use, to allow them to produce their music. Whether they are amateur or professional, they will all require different types of tools (Sandler et al., 2019), but IMS can be used to construct useful tools that can provide some advantages to them each. The purpose of the tool, and the manner is which the tool is used, will define the interaction between the human and the computer, where it is acknowledged what capacities are given to the AI system, rather than kept within the human mixing domain. Palladini (2018) proposed a number of different levels of AI approaches, and how these can be used to construct an IMS. This approach is derived from the field of self-driving cars (Rödel et al., 2014), where there is a constant interaction between a human and an AI, in order to both build a trust of the AI system. The analogy of a self-driving car, and an IMS is a very effective one, as in both cases, there is a strict human-computer interaction, in multiple different ways, from the addition of an automatic breaking system, to automatic gear selection, to fully automatic driving systems. In both cases, the AI system is being used as a tool for a human to use, in a way they see fit, one in a very practical deliberate way, and the other in a very creative way.

Palladini (2018) identified that there are different levels to which a human can interact with an IMS. These levels of interaction can be described are: Automatic; Independent; Recommendation; and Discovery.

13.4.1 Automatic

An automatic IMS is where a series of audio tracks are provided to the IMS and a full mix of the audio tracks is expected in return. This approach will automate all aspects of the mix, with no human interaction, other than perhaps a few high level control parameters, where an individual may wish to select a style or genre to be mixed. This approach could take the form of defining a set of requirements or constraints (Terrell and Sandler, 2012), by identifying a target track to mix in the same way as (Barchiesi and Reiss, 2009), through the definition of some predefined mixing goal (Ronan et al., 2018a), or even through mixing examples to learn the style of a specific mix engineer (Martínez Ramírez et al., 2020b). In the self driving car analogy, this approach would be the fully autonomous driving car.

A fully automated mixing system does not require any external interventions, other than the input of some audio tracks, and thus could be advantageous for an amateur, who is not experienced with music mixing, but instead requires the highest quality produced audio content, with minimal effort (Lefford et al., 2020a). These systems could also be used as benchmark mixes, which could be analysed and compared to ones own mix, to reflect on what issues or challenges are being faced within a mix. There are cases where it is not possible for a mix engineer to produce a mix, such as in a video game, where objects and components of the mix are constantly changing and need to be dynamically mixed (Stevens and Raybould, 2013; Selfridge et al., 2018). There are a number of bespoke approaches for this challenge, which include implementing some level of audio detail (Schwarz et al., 2011; Durr et al., 2015; Tsilfidis et al., 2009), where sonic elements are only included when there is suitable *space* in the mix, for each sonic element. Other approaches include the dynamic generation of sonic elements within a video game (Farnell, 2007; Mengual et al., 2016), or generative music approaches, only creating and mixing voices as and when required (Dawson, 2013).

13.4.2 Independent

An independent IMS, is one where a series of tasks can be allocated to the IMS, which it can manage and perform, while a mix engineer acts as a supervisor to the system. Overall, the mix engineer has control over the system, with the ability to change or overrule a decision made by an IMS. The ISM would have to react to the dynamic changes of a mix engineer, who will be modifying and manipulating other aspects of the audio. This could be achieved through the automation of a single type of audio effect, such as gain across all tracks (Perez Gonzalez and Reiss, 2008), through the automation of an entire music mixing process, such as automating the final mastering process (Toulson, 2016; Mimilakis et al., 2016b), through enhancing the audio signal quality for the mix engineer, in an adaptive manner, by reducing microphone bleed (Van Waterschoot and Moonen, 2011; Moffat and Reiss, 2016), reducing the comb filtering effects of phase interference between sound sources (Clifford and Reiss, 2013), or by mixing a set audio effect to a given task, such as providing an equaliser to an IMS, and asking for it to maintain the relative masking below a given threshold (Hafezi and Reiss, 2015). One of the best examples of an independent IMS, is the "automatic microphone mixing" system, developed by Dugan (1975), where the gain control for a series of microphone channels is presented to the IMS for automating, but all other components of the mix are controlled by the mix engineer. This approach has been integrated into the Yamaha CL series sound desks.

Another approach that is taken in building an independent IMS, is to allow the IMS to create a *rough mix*, where a set of initial parameters are set up for a mix engineer (Cartwright et al., 2014). This rough mix could be based on direct microphone analysis (Moffat and Sandler, 2019c), through an understanding of the physical geometry of the room (Terrell and Reiss, 2009), or through some initialisation process – more commonly known as a *sound-check* (Ewert and Sandler, 2016). This is analogous to a car which is able to provide some basic self driving capabilities, such as automatic parking (Hsu et al., 2008), or monitoring to ensure that the steering is staying within a specific lane on a motorway. The small segmented tasks are highly useful, individually, but do not remove the overall system control from the end user.

An independent IMS provides the mix engineer with knowledge as to what the system is automating, and how, with an active acknowledgement as to when the IMS will relinquish control to the mix engineer. One of the most important aspects of an independent IMS, is that the mix engineer is able to trust the system (Muir, 1994). The requirement for a consistent, predictable outcome, that the mix engineer can rely upon, without being betrayed (Baier, 1986), will greatly influence the utility of the IMS. If a mix engineer is in a position where they need to battle with the IMS to achieve the desired outcome, if the IMS contradicts the mix engineer, or if the IMS introduces some challenges that frustrate or interfere with the mixing process, then the IMS has no purpose in that mixing context, and the mix engineer will quickly use their supervisory role to remove the IMS from the music mixing process. However, developing an independent IMS that can mix audio content within the constraints provided, and understand the greater context of the changes it makes, provides considerable benefit to pro-am (Sandler et al., 2019), and amateur mix engineers, who could be in a position to focus down on smaller simpler tasks, or not have to worry about the negative impacts of their mixing exploration, and ensuring that a good quality mix is presented at all points during the mixing process, rather than necessitating a destructive process before a new and improved mix can be found.

13.4.3 Recommendation

A recommendation IMS, sometimes called a suggestive mixing system (Moffat and Sandler, 2019a), is one where the IMS has the capacity to analyse and interpret the current mixing process, gaining an understanding of the current mixing context (Lefford et al., 2020a), and use this to provide the mix engineer with some recommendations or suggestions. These recommendations could be the automatic labelling of instrument tracks (Pauwels et al., 2017; Sandler et al., 2019), adaptive audio effect parameter setting (Paterson, 2011). Recommendations could take the format of suggesting an audio processing workflow or chain, either through suggestion of the audio effect chain directly (Stasis et al., 2017a), or through the hierarchical sub-grouping or sub-mixing of stems (Ronan et al., 2015b).

Pestana et al. (2013) developed an IMS which is able to analyse a set of audio tracks and recommend changes to spectral characteristics of the tracks. Stables et al. (2014) developed an approach where audio effect parameters can be suggested, based on semantic descriptors, and a mix engineer can search through lists of descriptors to find the most appropriate for their use case. Zheng et al. (2016) crowdsourced a range of semantic terms associated with different mixing audio effects, which Seetharaman and Pardo (2016) developed into an IMS. Jillings and Stables (2017b) suggested gain mixing parameter settings to reduce perceived masking of a set of audio tracks. Cartwright and Pardo (2014) developed an advisory approach to synthesis voicing, where given a midi score, it could make recommendations as to what instrument voice would be most appropriate for that track. Vickers (2010) identified occasions when a single effect, namely dynamic range compression, has been overused, and negatively impacts the musical mix. IMS can also utilise mix statistical analysis approaches (Wilson and Fazenda, 2015; Colonel and Reiss, 2019), querying attributes of a mix to identify any potential issues, and may be the cause of those issues. Suggestions can then be made to correct these problems (Jun et al., 2015), which a mix engineer would have the opportunity to engage with if they so choose. Extending the autonomous car analogy, a recommendation system would be comparable to a system to suggest which gear the car should be in, or making suggestions to slow down as the speed limit changes, as performed by modern day sat-nav systems.

The real benefit of a recommendation system, is the ability for the IMS to become interpretable and adaptable. It will make a suggestions to users, who can accept or reject the suggestions, and this in turn can be used to search for more appropriate answers to the problem, similar to the approach proposed by Wilson and Fazenda (2016). The engineer maintains control over the mix, and the IMS at all times, and has the opportunity to actively engage with the IMS, or to pursue their own mixing approach.

A recommendation IMS could be beneficial to amateur mix engineers, while learning or attempting to hone their mixing skills, or used by professional mix engineers when in a situation that they are unsure as to what to do, or what approach to take. This approach could even be considered the intelligent, adaptive and context dependent equivalent to *Oblique strategies*, developed by Eno and Schmidt (1975). Oblique strategies is a set of cards, which all have different, general comments to consider while mixing, such as "Honour thy error as a hidden intention" or "Only one element of each kind". The cards were developed to assist with challenging creative decisions, or situations where the engineer is not sure what to you. The key aspect, is the control lies with the mix engineer at all stages, which gives them the power to make any decision during the mixing process, with the possibility of querying the IMS if deemed relevant. Furthermore, the development of interpretable AI systems (Baehrens et al., 2010), which would provide insight into the mixing decisions taken, and justification for recommendations made, could be a highly insightful one, both from understanding the benefits and applicability of the IMS, but also to better understand of existing music production workflows.

13.4.4 Discovery

An IMS constructed for the purposes of discovery, is designed to provide the mix engineer with some additional insight into the mixing process being undertaken. At this stage, the IMS will have no ability to enact control over the audio signal, instead producing representations which a mix engineer can use to inform themselves and aid in their own decision making process. The IMS is designed to support the mix engineer with their current goals and targets in some way. This could be through some mix visualisations, comparisons to existing *target* mix approaches or a textual or numerical response to the current mix.

Hargreaves et al. (2012) developed an approach for structural analysis of multitrack audio, and used this to automatically identify the chorus and verse components of the musical track being mixed. Virtanen et al. (2015) present a review of different approaches for structural sound analysis. Wichern et al. (2015) developed an approach for analysing and visualising the level of inter-track masking within a mix, which Izotope developed into an audio plugin, which is presented as part of Neutron (Izotope, 2020). Ford et al. (2015) took a similar approach, to visualise the perceptual masking between sets of audio tracks in a mix. De Man et al. (2017) analysed and identifying the reverberation level and impact it had on every track within a mix. Sauer et al. (2013) made a number of mix processing recommendations based on intended target semantic descriptors. Cannam et al. (2010) developed an approach for advanced visualisation of an audio track, and allowed for multiple versions, or mixes, of the same piece of music to be compared to one another, allowing for an effective comparison and analysis between multiple mixes of the same raw audio input. Where as Gelineck and Overholt (2015) and Merchel et al. (2012) both developed approaches for providing haptic feedback of music mixing. Bruford et al. (2019) developed an approach for searching for appropriate drum loops, given the rest of the audio content. Moffat et al. (2017) created a hierarchical structure to sound, based on unsupervised learning and perceptual attributes, which is designed to assist with searching for audio samples and loops.

The principal value of this approach is to provide a greater level of understanding as to the current audio mixture. Following the autonomous car analogy, this would be the development of parking sensors or a sat-nav technology that can give a view of the traffic around the next corner. This approach is beneficial to amateur mix engineers, as it can bring insight to parts of the musical mix, where their hearing ability or experience may not allow them to be aware of otherwise. It is often the case that an amateur will know they have a particular issue, such as a muddy mix, but not know how they can fix it. This approach can provide simply, easy to interpret information which can lead to a faster and more effective decision making process.

13.4.5 Control Level Summary

The development of an IMS is heavily constrained by the way in which mix engineers are intended to interact with it. Engineers can remain in complete control, but allow an IMS to provide some insight or discovery of the music they are mixing, or they can hand-off the entire mixing process, allowing the IMS complete autonomous control, with little but the most high level controls over the result. They way in which this interaction takes place will completely change the dynamic of the situation, and will directly impact the usability and attitude of the engineer. It is vital that, at all stages, there is an agreement between the IMS and the engineer as to what the expectation is, and the IMS should never step outside of this boundary, without clear signposting. There is current research demonstrating all four approaches to developing an IMS, however the challenges of an automatic IMS are only just being overcome. There is a significant need for further investigation into how individuals can interact with an IMS, and how the IMS can learn from this approach to refine the mixing protocol implemented. Due to the critical importance of the interaction between the IMS and the mix engineer, it is also necessary to understand the approach the IMS will take to create rules, how they will be represented and what feedback can presented back to the user, at all stages.

13.5 Further Design Considerations

A number of decisions, that need to be made while creating an IMS have been been outlined. However, there are a number of music specific considerations, that can highly influence the effectiveness and capacity of an IMS. In this section we will discuss and outlines these approaches.

13.5.1 Mixing by Sub-grouping

Within mixing the process of sub-grouping or using *buses*, is one where groups of similar tracks are all mixed together, independent of the rest of the mix, and processed as a smaller group, or *stem*, which is then integrated into the main mix. This is most commonly done with vocals and drums (Ronan et al., 2015a), but there are many types of mix buses used in professional mixing approaches.

McGarry et al. (2017) performed an ethnographic study, where they discuss the importance of subgrouping in music production. Ronan et al. (2017) surveyed a number of professional mix engineers to ask about their subgrouping practices, and concluded that almost all mix engineers will perform some form of subgrouping and apply audio effect processing to the group. Ronan et al. (2015a) identified that the use of subgrouping will have a positive impact on the final mix produced. Ronan et al. (2015b) developed an unsupervised approach for automatic subgrouping of sets of musical stems, which was shown to improve the result of a mix when used in combination with an IMS (Ronan et al., 2018a). Wilmering et al. (2016) describes a formal structure of the mixing process and audio effect processing workflow, which Jillings and Stables (2017a) uses to make proposals for channel routing of audio tracks.

It is clear that the approach of grouping together audio tracks, and reduce the complexity of the music mixing problem, is highly advantageous, though the subject still demands further research. As such, there is an opportunity to develop a bespoke IMS for a given sub-group. It can be considered that where a given IMS would be useful and effective at mixing a drum stem, it may be less effective at mixing the vocal stem, and then there would be a stem-mixing IMS, which only needs to process a smaller number of preprocessed music stems. Current results demonstrate that this approach has the potential to be highly effective in the field of IMS.

13.5.2 Intelligent Mixing Systems in Context

Music production is consistently driven by context. The shape of the music industry now, vs 50 years ago, mean that they types of musical performances, the expectation on how the music will be consumed and the expectation of a piece of music have all changed. The music industry and social contexts of a piece of music shape the way that professional engineers mix (Pras and Guastavino, 2011), and the way that that music is consumed. It only makes sense that the expectation of an IMS is that this context can, and should, be considered when constructing an IMS. Pras et al. (2018) identified that cultural and geographic differences between groups of individuals will influence how different mixes of the same track are perceived, making it clear that cultural context, along with educational and semantic contexts will heavily influence how a piece of music is perceived, and thus this will influence the types of IMS required for these given contexts. Lefford et al. (2020a) present an in-depth discussion as to how mixing in context can be performed, and the necessity of this approach, whilst Pardo et al. (2018) discusses the use of Music Information Retrieval (MIR) tools, such as source separation and noise removal within musical contexts.

For example, Ma et al. (2013) developed an IMS, which analysed 60 years of UK and US pop music, and it follows that, there an IMS could be constructed, following this approach, that continually updates the given IMS parameters, identifying suitable genre and cultural contexts to select data from, and use this to apply an equalisation curve most appropriate to a chosen piece of music, based on cultural, genre and current societal contexts and trends.

This concept could surely be extended further, to draw inspiration and concepts from latest releases, or larger sets of audio tracks that fall into a similar cluster, based on the relevant contexts in a given situation.

13.6 Discussion

IMS can have a multitude of different aims and purposes. They can be designed to suggest parameters for a pre-existing digital audio effect, as an educational tool for training students, or be designed as *black-box* systems which will take a number of pre-recorded audio tracks and produce a mix, as high quality as is possible, subject to the quality of the input material. Regardless of the manner in which an IMS is used, it is designed as a tool to be used by an audio engineer or operator, and as such, the way the tool is interacted with and used is of critical importance. We can never forget about the human in the mixing process, as there will always be some creative intention or decision to be realised, and the form that this interaction takes, will heavily influence the usefulness of the tool for mixing music, whether that be to remove some noise, as a tool to produce a full mix, or as creative inspiration.

Sandler et al. (2019) identified that there are three different levels of experience of individuals who work in studio production: amateur; pro-am; and professional. The use of IMS will be considerably different, dependant on the individuals experience and knowledge of the music production field, and their experience of using the tools of the trade. A professional may use an IMS as an exploratory tool, to allow them to quickly prototype a number of ideas that they have, which gives them a wide range of rough mixes. This will then allow them to decide which creative direction to take the mix, which they can then pursue in a more traditional manner. An amateur, on the other hand, may allow the IMS to direct a specific creative vision, and they can gently direct the IMS within a small range on some specific details, that they are focused on. A pro-am may well use the IMS as a collaboration, where the IMS and the individual will work together at different points, bouncing ideas off each other, moving forward and constantly changing different elements of the mix until an agreement is met between the pro-am and the IMS.

Music mixing and production is inherently an interconnected process, where processing applied to each and every individual track is highly dependent on both that individual track, but on every other audio track that goes into the mix. This means that there are often highly iterative processes, where changes are made to several different tracks in turn, constantly making small changes to each track, until a desired effect is reached. Alternatively, a mix engineer may speculate as to how they can make a particular track sound, and then adjust a number of other tracks to fit with their imagined track. Regardless of which process is taken, it is clear that there are a high level of interdependencies between all the audio tracks within a mix. This is highly relevant to IMS, as firstly, there is a requirement to model the interdependencies between audio tracks. In a machine learning context, this is not entirely trivial. This concept, of modelling interdependence between variables is often called *data fusion*, and this concept has been applied to music mixing (Hargreaves, 2014), however there is no consensus as to which approach is most appropriate for an IMS. A review of data fusion approaches is presented by Castanedo (2013).

In considering the music mixing process, it could be viewed as a specific type of sound design, where different sonic elements are gathered together, and crafted into an aesthetically pleasing format. Similarly in the wider context of sound design, the purpose it to take a number of relevant sonic elements, select which sounds are most appropriate to include and exclude, and blend them together in a pleasant and believable manner. The creative design nature of music mixing is one which lends itself to a combination of both highly technical and highly creative approaches. Similarly, design approaches may include both design and technical approaches. Lefford et al. (2020b) proposed learning from computer assisted design approaches, when design an IMS. The growth of AI technology in both these fields draw some strong parallels, and there are certainly intersections between the concepts of these works.

The development of a fully automatic IMS is a contentious subject, as many aspects of music mixing are considered a *dark art*, and a highly creative practice, and AI systems will never be able to produce a mix in the same way as a professional mix engineer (Birtchnell and Elliott, 2018), however, there are counter arguments that suggest that the integration of AI tools into the music mixing workflow can provide powerful new insights into music mixing practices, facilitate new ways to explore musical content and even influence the creative capacity of practitioners (Bromham, 2016). Furthermore, the facility for a mix engineer to misappropriate these intelligent tools, to use them to create new types of sounds, interesting mixes and to guide and influence the creative music production space have considerable impact on the entire field of music production. The potential for this new technology to drive potential creative outcomes is well known (Eno, 2004), and there are many changes to the future of music production, with the inclusion of AI mixing technologies (Prior, 2010; Pras et al., 2013).

13.7 The Future of Intelligent Mixing Systems

It is clear that there are significant opportunities for new and interesting developments within music production, through the development and use of IMS. The opportunities to include some signal processing approaches, and computational musicology into the system, can provide the system with a better understanding as to the fundamental of music. Fields such as audio source separation and music information retrieval afford considerably better musical and sonic understanding of the music medium, and often of the psychoacoustic and perceptual attributes, as music is perceived by people in general.

In the process to developing highly effective IMS system, the opportunity to develop a simplified *stem mixing* approaches, which group together a small number of tracks and produce an effective submit, has been shown to be a highly promising approach, though clearly further exploration is needed within this space.

The development of assistive technologies, for the creation and production of music, also present a realm of exciting opportunities. Using audio analysis tools to provide recommendation as to music samples and stems that will work well together, building the structure of a piece of music, may help reduce laborious tasks of finding the appropriate sounds in libraries of millions of different sounds. Where as a recommendation for voicing or re-voicing different melodies allow a much greater control over the shape of the timbre. A method to analyse a melody, either midi or musical, and recommend alternative instruments to play that piece, that would work well within the content of the current musical composition, could greatly aid the creation, and mixing of any musical piece.

Video games, virtual reality and augmented reality all lend themselves highly to dynamic intelligent mixing systems. In these cases, it would never be possible to predict the exact detail of how an individual will interact with the environment, and there are clear opportunities for developing approaches that can mix both diegetic and non-diegetic audio content seamlessly, within these environments.

It is clear that there are considerable challenges to overcome within the field of IMS. A combination of fundamental research, coupled with creative development and integration of technologies have the potential to have a considerable impact on the music industry and field of music production.

References

Arel, I., Rose, D. C., and Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research. *IEEE computational intelligence magazine*, 5(4):13–18.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., and Müller, K.-R. (2010). How to explain individual classification decisions. *Journal of Machine Learning Research*, 11(Jun):1803–1831.

Baier, A. (1986). Trust and antitrust. *ethics*, 96(2):231–260.

Barchiesi, D. and Reiss, J. (2009). Automatic target mixing using least-squares optimization of gains and equalization settings. In *Proceedings of the 12th Conference on Digital Audio Effects (DAFx-09), Como, Italy*.

Barchiesi, D. and Reiss, J. (2010). Reverse engineering of a mix. *Journal of the Audio Engineering Society*, 58(7/8):563–576.

Benito, A. L. and Reiss, J. D. (2017). Intelligent multitrack reverberation based on hinge-loss markov random fields. In *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*.

Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, 284(5):28–37.

Birtchnell, T. and Elliott, A. (2018). Automating the black art: Creative places for artificial intelligence in audio mastering. *Geoforum*, 96:77–86.

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and Bello, J. P. (2014). MedleyDB: A multitrack dataset for annotation-intensive MIR research. In *15th International Society for Music Information Retrieval Conference (ISMIR 2014)*.

Bolton, C., Machová, V., Kovacova, M., and Valaskova, K. (2018). The power of human–machine collaboration: Artificial intelligence, business automation, and the smart economy. *Economics, Management, and Financial Markets*, 13(4):51–56.

Bromham, G. (2016). How can academic practice inform mix-craft? In Hepworth-Sawyer, R. and Hodgson, J., editors, *Mixing music*, Perspective on Music Production., chapter 16, pages 245–256. Taylor & Francis.

Bromham, G., Moffat, D., Barthet, M., Danielsen, A., and Fazekas, G. (2019). The impact of audio effects processing on the perception of brightness and warmth. In *ACM Audio Mostly Conference*, Nottingham, UK. ACM.

Bromham, G., Moffat, D., Fazekas, G., Barthet, M., and Sandler, M. B. (2018). The impact of compressor ballistics on the perceived style of music. In *Audio Engineering Society Convention 145*, New York, USA.

Bruford, F., Barthet, M., McDonald, S., and Sandler, M. B. (2019). Groove explorer: An intelligent visual interface for drum loop library navigation. In *24th ACM Intelligent User Interfaces Conference (IUI)*.

Bull, R. (1997). The aesthetics of acid.

Burgess, R. J. (2014). *The history of music production*. Oxford University Press.

Cannam, C., Landone, C., and Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the ACM Multimedia 2010 International Conference*, pages 1467–1468, Firenze, Italy.

Cartwright, M. and Pardo, B. (2014). Synthassist: an audio synthesizer programmed with vocal imitation. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 741–742.

Cartwright, M., Pardo, B., and Reiss, J. D. (2014). Mixploration: Rethinking the audio mixer interface. In *International Conference on Intelligent User Interfaces*.

Cascone, K. (2000). The aesthetics of failure: "post-digital" tendencies in contemporary computer music. *Computer Music Journal*, 24(4):12–18.

Castanedo, F. (2013). A review of data fusion techniques. *The Scientific World Journal*, 2013.

Choudakis, E. T. and Reiss, J. D. (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)*. Audio Engineering Society.

Chourdakis, E. T. and Reiss, J. D. (2017). A machine-learning approach to application of intelligent artificial reverberation. *Journal of the Audio Engineering Society*, 65(1/2):56–65.

Clifford, A. and Reiss, J. D. (2013). Using delay estimation to reduce comb filtering of arbitrary musical sources. *Journal of the Audio Engineering Society*, 61(11):917–927.

Cohen, S. (1993). Ethnography and popular music studies. *Popular music*, 12(2):123–138.

Colonel, J. and Reiss, J. D. (2019). Exploring preference for multitrack mixes using statistical analysis of mir and textual features. In *Audio Engineering Society Convention 147*. Audio Engineering Society.

Dawson, R. (2013). Cell: A generative music game. Master's thesis, University of Sussex.

De Man, B., Boerum, M., Leonard, B., Massenburg, G., King, R., and Reiss, J. D. (2015). Perceptual evaluation of music mixing practices. In *Audio Engineering Society Convention 138*.

De Man, B., Jillings, N., Moffat, D., Reiss, J. D., and Stables, R. (2016). Subjective comparison of music production practices using the Web Audio Evaluation Tool. In *2nd AES Workshop on Intelligent Music Production*.

De Man, B., McNally, K., and Reiss, J. D. (2017). Perceptual evaluation and analysis of reverberation in multitrack music production. *Journal of the Audio Engineering Society*, 65(1/2):108–116.

De Man, B., Mora-Mcginity, M., Fazekas, G., and Reiss, J. D. (2014). The Open Multitrack Testbed. In *Audio Engineering Society Convention 137*.

De Man, B. and Reiss, J. D. (2013a). A knowledge-engineered autonomous mixing system. In *Audio Engineering Society Convention 135*. Audio Engineering Society.

De Man, B. and Reiss, J. D. (2013b). A semantic approach to autonomous mixing. *Journal on the Art of Record Production (JARP)*, 8.

Deruty, E. (2016). Goal-oriented mixing. In *Proceedings of the 2nd AES Workshop on Intelligent Music Production*.

Dugan, D. (1975). Automatic microphone mixing. *Journal of the Audio Engineering Society*, 23(6):442–449.

Durr, G., Peixoto, L., Souza, M., Tanoue, R., and Reiss, J. D. (2015). Implementation and evaluation of dynamic level of audio detail. In *Audio Engineering Society Conference: 56th International Conference: Audio for Games*, London, UK.

Eno, B. (2004). The studio as compositional tool. In Cox, C. and Warner, D., editors, *Audio culture: Readings in modern music*, chapter 22, pages 127–130. Londres: Continuum.

Eno, B. and Schmidt, P. (1975). Oblique strategies. *Opal.(Limited edition, boxed set of cards.)[rMAB]*.

Ewert, S. and Sandler, M. (2016). Piano transcription in the studio using an extensible alternating directions framework. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11):1983–1997.

Farnell, A. (2007). An introduction to procedural audio and its application in computer games. In *Audio Mostly Conference*, pages 1–31, Ilmenau, Germany. ACM.

Fenton, S. (2018). Automatic mixing of multitrack material using modified loudness models. In *Audio Engineering Society Convention 145*. Audio Engineering Society.

Fenton, S. and Lee, H. (2015). Towards a perceptual model of "punch" in musical signals. In *Audio Engineering Society Convention 139*. Audio Engineering Society.

Ford, J., Cartwright, M., and Pardo, B. (2015). Mixviz: A tool to visualize masking in audio mixes. In *Audio Engineering Society Convention 139*. Audio Engineering Society.

Furman, J. and Seamans, R. (2019). AI and the economy. *Innovation Policy and the Economy*, 19(1):161–191.

Gang, R., Bocko, G., Bocko, M. F., Headlam, D., Lundberg, J., and Ren, G. (2010). Automatic music production system employing probabilistic expert systems. In *Audio Engineering Society Convention 129*. Audio Engineering Society.

Garcia, R. A. (2001). Automating the design of sound synthesis techniques using evolutionary methods. In *Proceeding 4th International Conference on Digital Audio Effects (DAFx)*, Limerick, Ireland.

Gatys, L., Ecker, A. S., and Bethge, M. (2015). Texture synthesis using convolutional neural networks. In *Advances in neural information processing systems*, pages 262–270.

Gelineck, S. and Overholt, D. (2015). Haptic and visual feedback in 3d audio mixing interfaces. In *Proceedings of Audio Mostly Conference on Interaction With Sound*, page 14. ACM.

Gillet, O. and Richard, G. (2006). ENST-drums: an extensive audio-visual database for drum signals processing. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR-06)*, pages 156–159.

Glaser, B. and Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the journal of the British sociological association*, 12(1):27–49.

Grinstein, E., Duong, N. Q., Ozerov, A., and Pérez, P. (2018). Audio style transfer. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 586–590. IEEE.

Hafezi, S. and Reiss, J. D. (2015). Autonomous multitrack equalization based on masking reduction. *Journal of the Audio Engineering Society*, 63(5):312–323.

Hargreaves, S. (2014). *Music Metadata Capture in the Studio from Audio and Symbolic Data*. PhD thesis, Queen Mary University of London.

Hargreaves, S., Klapuri, A., and Sandler, M. B. (2012). Structural segmentation of multitrack audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(10):2637–2647.

Hawley, S. H., Colburn, B., and Mimilakis, S. I. (2019). Signaltrain: Profiling audio compressors with deep neural networks. *arXiv preprint arXiv:1905.11928*.

Hermes, K. (2019). Enhancing creativity through automatic mixing research: testing spectral clarity predictors in the mixprocess. *The Art of Record Production*.

Hestermann, S. and Deffner, N. (2020). Enhanced de-essing via neural networks. In *Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part II*, pages 537–542. Springer.

Holfelt, J., Csapo, G., Andersson, N., Zabetian, S., Castenieto, M., Dahl, S., Overholt, D., and Erkut, C. (2017). Extraction, mapping, and evaluation of expressive acoustic features for adaptive digital audio effects. In *Proceedings of the Sound & Music Computing Conference*.

Hsu, T.-h., Liu, J.-F., Yu, P.-N., Lee, W.-S., and Hsu, J.-S. (2008). Development of an automatic parking system for vehicle. In *2008 IEEE Vehicle Power and Propulsion Conference*, pages 1–6. IEEE.

Izhaki, R. (2008). *Mixing Audio: Concepts, Practices and Tools*. Focal Press.

Izotope (2020). Neutron plugin. Website. https://www.izotope.com/en/learn/unmasking-your-mix-with-neutron.html.

Jazzomat (2018). Jazzomat dataset, the jazzomat research project. http://jazzomat.hfm-weimar.de/dbformat/dbcontent.html. Accessed 26 May 2020.

Jillings, N. and Stables, R. (2017a). Automatic channel routing using musical instrument linked data. In *3rd AES Workshop on Intelligent Music Production*.

Jillings, N. and Stables, R. (2017b). Automatic masking reduction in balance mixes using evolutionary computing. In *Audio Engineering Society Convention 143*. Audio Engineering Society.

Jun, S., Kim, D., Jeon, M., Rho, S., and Hwang, E. (2015). Social mix: automatic music recommendation and mixing scheme based on social network analysis. *The Journal of Supercomputing*, 71(6):1933–1954.

Katz, B. and Katz, R. A. (2003). *Mastering audio: the art and the science*. Butterworth-Heinemann.

King, A. (2015). Technology as a vehicle (tool and practice) for developing diverse creativities. In Burnard, P. and Haddon, E., editors, *Activating Diverse Musical Creativities: Teaching and Learning in Higher Music Education*, chapter 11, pages 203–222. Bloomsbury Publishing.

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

Lefford, M. N., Bromham, G., Fazekas, G., and Moffat, D. (2020a). Context aware intelligent mixing systems. *Journal of the Audio Engineering Society*. In Press.

Lefford, M. N., Bromham, G., and Moffat, D. (2020b). Mixing with intelligent mixing systems: Evolving practices and lessons from computer assisted design. In *Audio Engineering Society Convention 148*, Vienna, Austria.

Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., and Lederberg, J. (1993). DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence*, 61(2):209–261.

Ma, Z., Reiss, J. D., and Black, D. A. (2013). Implementation of an intelligent equalization tool using yule-walker for music mixing and mastering. In *Audio Engineering Society Convention 134*. Audio Engineering Society.

Maddams, J. A., Finn, S., and Reiss, J. D. (2012). An autonomous method for multi-track dynamic range compression. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, pages 1–8.

Martínez Ramírez, M. A., Benetos, E., and Reiss, J. D. (2019). A general-purpose deep learning approach to model time-varying audio effects. In *International Conference on Digital Audio Effects*.

Martínez Ramírez, M. A., Benetos, E., and Reiss, J. D. (2020a). Deep learning for black-box modeling of audio effects. *Applied Sciences*, 10(2):638.

Martínez Ramírez, M. A. and Reiss, J. D. (2017a). Deep learning and intelligent audio mixing. In *Proceedings of the 3rd Workshop on Intelligent Music Production*.

Martínez Ramírez, M. A. and Reiss, J. D. (2017b). Stem audio mixing as a content-based transformation of audio features. In *19th IEEE Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE.

Martínez Ramírez, M. A. and Reiss, J. D. (2019). Modeling nonlinear audio effects with end-to-end deep neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 171–175. IEEE.

Martínez Ramírez, M. A., Stoller, D., and Moffat, D. (2020b). A deep learning approach to intelligent drum mixing with the wave-u-net. *Journal of the Audio Engineering Society*. In Press.

Mat, M., Yassin, I. M., Taib, M. N., Zabidi, A., Hassan, H. A., and Tahir, N. M. (2010). Adaptive filter based on narx model for recorded audio noise removal. In *2010 IEEE Control and System Graduate Research Colloquium (ICSGRC 2010)*, pages 26–32. IEEE.

Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M. (2015). The evolution of popular music: Usa 1960–2010. *Royal Society open science*, 2(5):150081.

McDermott, J. H. and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, 71(5):926–940.

McGarry, G., Tolmie, P., Benford, S., Greenhalgh, C., and Chamberlain, A. (2017). " they're all going out to something weird" workflow, legacy and metadata in the music production process. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 995–1008.

McGrath, S. and Love, S. (2017). The user experience of mobile music making: An ethnographic exploration of music production and performance in practice. *Computers in Human Behavior*, 72:233–245.

Mengual, L., Moffat, D., and Reiss, J. D. (2016). Modal synthesis of weapon sounds. In *Proceeding Audio Engineering Society Conference: 61st International Conference: Audio for Games*, London. Audio Engineering Society.

Merchel, S., Altinsoy, M. E., and Stamm, M. (2012). Touch the sound: audio-driven tactile feedback for audio mixing applications. *Journal of the Audio Engineering Society*, 60(1/2):47–53.

Mimilakis, S. I., Cano, E., Abeßer, J., and Schuller, G. (2016a). New sonorities for jazz recordings: Separation and mixing using deep neural networks. In *2nd AES Workshop on Intelligent Music Production (WIMP)*, volume 13.

Mimilakis, S. I., Drossos, K., Virtanen, T., and Schuller, G. (2016b). Deep neural networks for dynamic range compression in mastering applications. In *Audio Engineering Society Convention 140*.

Miranda, E. R. and Al Biles, J. (2007). *Evolutionary computer music*. Springer.

Moffat, D. and Reiss, J. D. (2016). Dereverberation and its application to the blind source separation problem. In *Proceeding Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. Audio Engineering Society.

Moffat, D. and Reiss, J. D. (2018). Objective evaluations of synthesised environmental sounds. In *Proceeding 21th International Conference on Digital Audio Effects (DAFx-17)*, Aveiro, Portugal.

Moffat, D. and Reiss, J. D. (2020). Semantic music production. *Journal of the Audio Engineering Society*. In Press.

Moffat, D., Ronan, D., and Reiss, J. D. (2015). An evaluation of audio feature extraction toolboxes. In *Proceeding 18th International Conference on Digital Audio Effects (DAFx-15)*, Trondheim, Norway.

Moffat, D., Ronan, D., and Reiss, J. D. (2017). Unsupervised taxonomy of sound effects. In *Proceeding 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, UK.

Moffat, D. and Sandler, M. B. (2018). Adaptive ballistics control of dynamic range compression for percussive tracks. In *Audio Engineering Society Convention 145*, New York, USA.

Moffat, D. and Sandler, M. B. (2019a). Approaches in intelligent music production. *Arts*, 8(4):125.

Moffat, D. and Sandler, M. B. (2019b). An automated approach to the application of reverberation. In *Audio Engineering Society Convention 147*, New York, USA.

Moffat, D. and Sandler, M. B. (2019c). Automatic mixing level balancing enhanced through source interference identification. In *Audio Engineering Society Convention 146*, Dublin, Ireland.

Moffat, D. and Sandler, M. B. (2019d). Machine learning multitrack gain mixing of drums. In *Audio Engineering Society Convention 147*, New York, USA.

Moffat, D., Thalmann, F., and Sandler, M. B. (2018). Towards a semantic web representation and application of audio mixing rules. In *Proceedings of the 4th Workshop on Intelligent Music Production (WIMP)*, Huddersfield, Uk.

Moore, A. (2020). Dynamic range compression and the semantic descriptor aggressive. *Applied Sciences*, 10(7):2350.

Muir, B. M. (1994). Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922.

Naylor, P. A. and Gaubitch, N. D. (2010). *Speech dereverberation*. Springer Science & Business Media.

Nelson, W. R. et al. (1982). Reactor: An expert system for diagnosis and treatment of nuclear reactor accidents. In *AAAI*, pages 296–301. Pittsburgh, PA.

Nercessian, S. and Lukin, A. (2019). Speech dereverberation using recurrent neural networks. In *Proceeding 23rd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK.

Owsinski, B. (2013). *The Mixing Engineer's Handbook*. Nelson Education.

Pachet, F. and Delerue, O. (2000). On-the-fly multi-track mixing. In *Audio Engineering Society Convention 109*. Audio Engineering Society.

Palladini, A. (2018). Intelligent audio machines. Keynote Talk at 4th Workshop on Intelligent Music Production (WIMP-18).

Pardo, B., Rafii, Z., and Duan, Z. (2018). Audio source separation in a musical context. In *Springer Handbook of Systematic Musicology*, pages 285–298. Springer.

Paterson, J. (2011). The preset is dead; long live the preset. In *Audio Engineering Society Convention 130*. Audio Engineering Society.

Pauwels, J., O'Hanlon, K., Fazekas, G., and Sandler, M. B. (2017). Confidence measures and their applications in music labelling systems based on hidden markov models. In *ISMIR*, pages 279–285.

Perez Gonzalez, E. and Reiss, J. (2009a). Automatic equalization of multichannel audio using cross-adaptive methods. In *Audio Engineering Society Convention 127*. Audio Engineering Society.

Perez Gonzalez, E. and Reiss, J. (2009b). Automatic gain and fader control for live mixing. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'09)*, pages 1–4.

Perez Gonzalez, E. and Reiss, J. D. (2008). Improved control for selective minimization of masking using interchannel dependancy effects. In *11th International Conference on Digital Audio Effects (DAFx)*, page 12.

Perez Gonzalez, E. and Reiss, J. D. (2010). A real-time semiautonomous audio panning system for music mixing. *EURASIP Journal on Advances in Signal Processing*, 2010(1):436895.

Pestana, P. and Reiss, J. (2014). Intelligent audio production strategies informed by best practices. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. Audio Engineering Society.

Pestana, P. D., Ma, Z., Reiss, J. D., Barbosa, A., and Black, D. A. (2013). Spectral characteristics of popular commercial recordings 1950–2010. In *Audio Engineering Society Convention 135*. Audio Engineering Society.

Pestana, P. D., Reiss, J. D., and Barbosa, A. (2015). Cross-adaptive polarity switching strategies for optimization of audio mixes. In *Audio Engineering Society Convention 138*.

Pestana, P. D. L. G. (2013). *Automatic mixing systems using adaptive digital audio effects*. PhD thesis, Universidade Católica Portuguesa.

Pras, A., De Man, B., and Reiss, J. D. (2018). A case study of cultural influences on mixing practices. In *Audio Engineering Society Convention 144*.

Pras, A. and Guastavino, C. (2011). The role of music producers and sound engineers in the current recording context, as perceived by young professionals. *Musicae Scientiae*, 15(1):73–95.

Pras, A., Guastavino, C., and Lavoie, M. (2013). The impact of technological advances on recording studio practices. *Journal of the American Society for Information Science and Technology*, 64(3):612–626.

Prior, N. (2010). The rise of the new amateurs: Popular music, digital technology, and the fate of cultural production. In Hall, J. R., Grindstaff, L., and Lo, M.-C., editors, *Routledge Handbook of Cultural Sociology*, pages 398–407. Routledge.

Prior, N. (2012). Digital formations of popular music. *Réseaux*, 2:66–90.

Rasmussen, A. N. (1990). The inco expert system project: Clips in shuttle mission control. *First CLIPSConference*, page 305.

Reed, D. (2000). A perceptual assistant to do sound equalization. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 212–218. ACM.

Reiss, J. and Brandtsegg, Ø. (2018). Applications of cross-adaptive audio effects: automatic mixing, live performance and everything in between. *Frontiers in Digital Humanities*, 5:17.

Reiss, J. D. (2011). Intelligent systems for mixing multichannel audio. In *Proceeding 17th International Conference on Digital Signal Processing (DSP)*, pages 1–6. IEEE.

Reiss, J. D. (2018). Do you hear what i hear? the science of everyday sounds. Professorial: Inaugural Lecture Queen Mary University.

Rödel, C., Stadler, S., Meschtscherjakov, A., and Tscheligi, M. (2014). Towards autonomous cars: the effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 1–8. ACM.

Ronan, D., De Man, B., Gunes, H., and Reiss, J. D. (2015a). The impact of subgrouping practices on the perception of multitrack music mixes. In *Audio Engineering Society Convention 139*.

Ronan, D., Gunes, H., and Reiss, J. D. (2017). Analysis of the subgrouping practices of professional mix engineers. In *Audio Engineering Society Convention 142*.

Ronan, D., Ma, Z., Mc Namara, P., Gunes, H., and Reiss, J. D. (2018a). Automatic minimisation of masking in multitrack audio using subgroups. *ArXiv e-prints*.

Ronan, D., Moffat, D., Gunes, H., and Reiss, J. D. (2015b). Automatic subgrouping of multitrack audio. In *Proceeding 18th International Conference on Digital Audio Effects (DAFx-15)*, Trondheim, Norway.

Ronan, D., Reiss, J. D., and Gunes, H. (2018b). An empirical approach to the relationship between emotion and music production quality. *arXiv preprint arXiv:1803.11154*.

Russell, S. J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson Education Limited, third edition.

Sandler, M., De Roure, D., Benford, S., and Page, K. (2019). Semantic web technology for new experiences throughout the music production-consumption chain. In *2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*, pages 49–55. IEEE.

Sarkar, S., Reiss, J. D., and Brandtsegg, Ø. (2017). Investigation of a drum controlled cross-adaptive audio effect for live performance. In *Proceeding 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, UK.

Sauer, C., Roth-Berghofer, T., Auricchio, N., and Proctor, S. (2013). Recommending audio mixing workflows. In *International Conference on Case-Based Reasoning*, pages 299–313. Springer.

Scherer, K. R., Zentner, M. R., et al. (2001). Emotional effects of music: Production rules. *Music and emotion: Theory and research*, 361(2001):392.

Schmidt, B. (2003). Interactive mixing of game audio. In *Audio Engineering Society Convention 115*. Audio Engineering Society.

Schwarz, D., Cahen, R., Jacquemin, C., and Ding, H. (2011). Sound level of detail in interactive audiographic 3d scenes. In *International Computer Music Conference (ICMC)*, Huddersfield, UK.

Seetharaman, P. and Pardo, B. (2016). Audealize: Crowdsourced audio production tools. *Journal of the Audio Engineering Society*, 64(9):683–695.

Selfridge, R., Moffat, D., Avital, E. J., and Reiss, J. D. (2018). Creating real-time aeroacoustic sound effects using physically informed models. *Journal of the Audio Engineering Society*, 66(7/8):594–607.

Senior, M. (2011). Mixing secrets for the small studio: Additional resources.

Senior, M. (2012). *Mixing Secrets*. Focal Press.

Sheng, D. and Fazekas, G. (2019). A feature learning siamese model for intelligent control of the dynamic range compressor. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Sheridan, T. B. and Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Technical report, Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Stables, R., Enderby, S., De Man, B., Fazekas, G., and Reiss, J. D. (2014). SAFE: A system for the extraction and retrieval of semantic audio descriptors. In *15th International Society for Music Information Retrieval Conference (ISMIR 2014)*.

Stasis, S., Jillings, N., Enderby, S., and Stables, R. (2017a). Audio processing chain recommendation. In *Proceedings of the 20th International Conference on Digital Audio Effects*, Edinburgh, UK.

Stasis, S., Jillings, N., Enderby, S., and Stables, R. (2017b). Audio processing chain recommendation using semantic cues. In *3rd Workshop on Intelligent Music Production*.

Stevens, R. and Raybould, D. (2013). *The game audio tutorial: A practical guide to creating and implementing sound and music for interactive games*. Routledge.

Terrell, M. and Reiss, J. D. (2009). Automatic monitor mixing for live musical performance. *Journal of the Audio Engineering Society*, 57(11):927–936.

Terrell, M., Reiss, J. D., and Sandler, M. (2010). Automatic noise gate settings for drum recordings containing bleed from secondary sources. *EURASIP Journal on Advances in Signal Processing*, 2010:10.

Terrell, M. and Sandler, M. (2012). An offline, automatic mixing method for live music, incorporating multiple sources, loudspeakers, and room effects. *Computer Music Journal*, 36(2):37–54.

Terrell, M., Simpson, A., and Sandler, M. (2014). The mathematics of mixing. *Journal of the audio engineering society*, 62(1/2):4–13.

Toulson, R. (2016). The dreaded mix sign-off: Handing over to mastering. In *Mixing Music*, pages 257–269. Routledge.

Tsilfidis, A., Papadakos, C., and Mourjopoulos, J. (2009). Hierarchical perceptual mixing. In *Audio Engineering Society Convention 126*.

Van Waterschoot, T. and Moonen, M. (2011). Fifty years of acoustic feedback control: State of the art and future challenges. *Proceedings of the IEEE*, 99(2):288–327. mixing.

Verfaille, V. and Arfib, D. (2002). Implementation strategies for adaptive digital audio effects. In *Proceeding 5th International Conference on Digital Audio Effects (DAFx-02)*, pages 21–26, Hamburg.

Verfaille, V., Arfib, D., Keiler, F., Knesebeck, A. v. d., and Zölter, U. (2011). Adaptive digital audio effects. In Zölter, U., editor, *DAFX - Digital Audio Effects*, chapter 9, pages 321–391. John Wiley & Sons, Inc.

Verma, P. and Smith, J. O. (2018). Neural style transfer for audio spectograms. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Workshop for Machine Learning for Creativity and Design, Long Beach, CA, USA.

Vickers, E. (2010). The loudness war: Background, speculation, and recommendations. In *Audio Engineering Society Convention 129*. Audio Engineering Society.

Virtanen, T., Gemmeke, J. F., Raj, B., and Smaragdis, P. (2015). Compositional models for audio processing: Uncovering the structure of sound mixtures. *IEEE Signal Processing Magazine*, 32(2):125–144.

Ward, D., Reiss, J. D., and Athwal, C. (2012). Multitrack mixing using a model of loudness and partial loudness. In *Audio Engineering Society Convention 133*.

Weaver, J., Barthet, M., and Chew, E. (2018). Analysis of piano duo tempo changes in varying convolution reverberation conditions. In *Audio Engineering Society Convention 145*. Audio Engineering Society.

Weaver, J., Barthet, M., and Chew, E. (2019). Filling the space: The impact of convolution reverberation time on note duration and velocity in duet performance. In *Audio Engineering Society Convention 147*. Audio Engineering Society.

Wichern, G., Wishnick, A., Lukin, A., and Robertson, H. (2015). Comparison of loudness features for automatic level adjustment in mixing. In *Audio Engineering Society Convention 139*.

Wilmering, T., Fazekas, G., and Sandler, M. B. (2016). AUFX-O: Novel methods for the representation of audio processing workflows. In *International Semantic Web Conference*.

Wilmering, T., Moffat, D., Milo, A., and Sandler, M. B. (2020). A history of audio effects. *Applied Sciences*, 10(3):791.

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

Wilson, A. and Fazenda, B. M. (2013). Perception & evaluation of audio quality in music production. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*.

Wilson, A. and Fazenda, B. M. (2016). An evolutionary computation approach to intelligent music production informed by experimentally gathered domain knowledge. In *Proceedings of the 2nd Audio Engineering Society Workshop on Intelligent Music Production (WIMP)*, London, UK.

Wilson, A., Loughran, R., and Fazenda, B. M. (2017). On the suitability of evolutionary computing to developing tools for intelligent music production. In *3rd Workshop on Intelligent Music Production (WIMP)*, Salford, UK.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Zheng, T., Seetharaman, P., and Pardo, B. (2016). SocialFX: Studying a crowdsourced folksonomy of audio effects terms. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 182–186.