OPEN ACCESS Freely available online
M. N. Lefford, G. Bromham, G. Fazekas, and D. Moffat, "Context-Aware Intelligent Mixing Systems"
*J. Audio Eng. Soc.*, vol. 69, no. 3, pp. 128–141, (2021 March).
DOI: https://doi.org/10.17743/jaes.2020.0043



# Context-Aware Intelligent Mixing Systems*

**M. NYSSIM LEFFORD,**[1] *AES Associate*, **GARY BROMHAM,**[2] *AES Student Member*,
(nyssim.lefford@ltu.se) (g.bromham@qmul.ac.uk)

**GYÖRGY FAZEKAS,**[2] **AND DAVID MOFFAT,**[3] *AES Member*
(g.fazekas@qmul.ac.uk ) (david.moffat@plymouth.ac.uk)

[1]*Luleå University of Technology, Sweden*
[2]*Queen Mary University of London, London, United Kingdom*
[3]*University of Plymouth, Plymouth, United Kingdom*

Intelligent Mixing Systems (IMS) are rapidly becoming integrated into music mixing and production workflows. The intelligences of a human mixer and IMS can be distinguished by their abilities to comprehend, assess, and appreciate context. Humans will factor context into decisions, particularly concerning the use and application of technologies. The utility of an IMS depends on both its affordances and the situation in which it is to be used. The appropriate use for conventional purposes, or its utility for misappropriation, is determined by the context. This study considers how context impacts mixing decisions and the use of technology, focusing on how the mixer's understanding of context can inform the use of IMS, and how the use of IMS can aid in informing a mixer of different contexts.

## 0 INTRODUCTION

Intelligent mixing systems (IMS) and technologies are rapidly growing in sophistication and *intelligence*, and thus they are increasingly useful for facilitating and assisting mix engineers in or automating the mixing process. However machine intelligence, computational models, and algorithmic representations of the mixing process are failing to capture some vital aspects of human mixing knowledge. In particular, there has yet to be created an IMS that actively incorporates or imitates into its functionality anything like the human ability to adapt actions and decisions to a specific context. This ability enables a mix engineer to appropriately and precisely align mixing decisions to the musical content, client needs and expectations, audience expectations, and the time and place in which the mix engineer works. Responding to context is also one way that a mix engineer imparts their unique sonic signature, which is typically a goal of accomplished engineers [1–3]. Mix engineers adapt to specific situations and constraints axiomatically. They might be aided in doing so by context-aware mixing technologies.

Mixing may be described as "a process in which multitrack material . . . is balanced, treated and combined into a multichannel format, most commonly two-channel stereo. But a less technical definition . . . is that a mix is a sonic presentation of emotions, creative ideas and performance" [4]. The role of the mix engineer is to process, combine, and balance instruments and tracks with the aim of producing a mix that is of high audio quality as well as expressive. A mix engineer works with 'aesthetic criteria' [5] as much as technical ones to achieve this goal. Throughout this article, we shall refer to a mixer as an individual that takes up the task of mixing.

The following analysis deconstructs the influences of context on mixing decisions with the aim of making a context-aware IMS that appears *smarter* and offers users new utility. The study's focus is on pop music mixing though many findings are applicable to other forms of mixing. From the impacts identified, it is possible to see how prior research from semantic audio (SA), music information retrieval (MIR), the semantic web, and context-aware computing (CAC) hold relevance for developing context awareness in IMS. In the future, IMS may be driven by information gleaned from a user's sensory context, and/or systems may detect context-significant information directly from the user's actions and choice history. Furthermore, IMS may encourage the mixer's own context awareness and thereby offer a new kind of support for decision making.

*This paper is supported by EPSRC Grant EP/S026991/1.

There are various approaches to designing IMS with different intended uses in mind. Commercial products include *smart* plugin effects processors such as Soothe from Oaksound[1] or VocalAlign from SynchroArts[2] as well as creative environments such as Neutron from Izotope[3] or Faders from Semantic Audio Labs.[4] The latter offer intelligent assistance functionality. For a survey of design approaches see [6]. To date, two approaches appear to have dominated IMS design in academic research and commercial products. Both approaches involve deriving mixing *rules*.

In one approach, mixing decisions are explained by experts and experts' actions are observed. The music production literature and other relevant ethnographic and musicological literature about mixing practices and praxis are also consulted. This practice-oriented literature provides generalizations about mixing norms and conventions [5], technical standards, and common artistic considerations. Rules are extrapolated then proceduralized for systems explicitly using an ontology or implicitly in machine learning models. Elsewhere rules are reverse engineered through data-driven analysis of existing musical mixes [7]. This requires training systems on corpora of prototypical examples. These two approaches can sometimes be combined [6].

These systems, no matter how sophisticated and powerful, are limited in some ways. Firstly systems have a finite set of rules, and that set is constrained by the observations of practice made or the corpus analyzed. Unlike a human mixer, at present, IMS do not typically generate or learn new rules. Though machine learning may do this implicitly, it is not yet common practice. And IMS do not adapt and modify rules for specific content, circumstances, or context. However, "the mix is dependent on the music, and mixing is not just a set of technical challenges" [4, p. 4]. Human mixers may have to disentangle attributes rooted in technical concerns from those that pertain to genre or musical aesthetics [4]; for example particular genres may warrant compromises in audio quality because it can suggest authenticity. Mixers balance technical and aesthetic concerns [8]. Information such as genre helps mixers to determine which choices are appropriate, given the context. IMS cannot yet make these distinctions. Systems may prioritize attributes or rules but as compared to a mixer they work with far fewer criteria for determining relative importance. They do not have access to contextualizing criteria.

Genre conventions are particularly important in IMS as they can be represented as rules. They are often key considerations in system design and mixes rendered by IMS tend toward the conventional. Human mixers, however, are expected to introduce elements of novelty and take decisions that align with their personal aesthetics. As Zak puts it, "While musicians leave the traces of their emotions, experiences, and the sounds of their musical expression on tape, the composite sound image that we recognize as the musical work is fashioned by recording engineers and producers—*performers* in their own right. They are the musicians' artistic collaborators, and their actions and aesthetic choices, too, are represented in the form of the finished work" [9, p. 17]. When and how much artistic personality to interject depends on context.

Mixes are not merely stylized but highly idiosyncratic. Mixers use minute adjustments that are carefully matched to performances and content. Idiosyncrasies may also emerge spontaneously simply due to circumstances, such as the physical setting of the studio and the combination of technologies used. Nevertheless, they impart unique and valued qualities to the mix [3, 10].

Many of these minute adjustments are made to shape the perceived emotional characteristics of a mix. "As a mixing engineer, one of our prime functions, which is actually our responsibility, is to help deliver the emotional context of a musical piece. From the general mix plan to the smallest reverb nuances, the tools we use—and the way we use them—can all sharpen or sometimes even create power, aggression, softness, melancholy, psychedelia and many other moods" [4, p. 4].

Critical listeners, mixers, clients, and discerning audiences are able to detect these variations mix to mix [3, 11, 12]. They hear subtlest modulation in a timbre or difference in balance and relative to other attributes of a mix. So far these are patterns that not even today's most advanced deep learning can identify. Moreover, human listeners are fickle about their preferences for these affectations. What is unacceptable in one mix is fantastic in another. Liking depends on the context. IMS is able to produce variation, but it cannot yet reproduce the idiosyncrasies produced by a mixer responding to context.

### 0.1 Context

There are numerous ways to define any given context and select relevant attributes. Some of the most pragmatic approaches come from the CAC literature and are geared toward designing useful and assistive computational tools. The application of any technology is contextualized by the setting surrounding its application [13]. Dey claims, "If a piece of information can be used to characterize the situation of a participant in an interaction, then that information is context" [14, p. 2]. A setting may be defined by features such as time of use, concurrent activities, the user's geographic location, the proximity of objects, and people in the user's surroundings as well as by information about the task for which the technology is applied [14]. In pop music mixing, tasks are contextualized, for example, by genre. Certain tasks, effects, attributes, and conventions may be performed only within the context of particular genres. To function with context awareness, an IMS needs to perform mixing tasks that are appropriate given these kinds of considerations.

When humans assess a context to determine what is relevant and to source relevant information, they can direct their attention dynamically, deliberately, and according to their values [15]. They survey the contextualizing elements

[1] https://oeksound.com/plugins/soothe/
[2] https://www.synchroarts.com/products/vocalign-pro/overview
[3] https://www.izotope.com/en/products/neutron.html
[4] https://www.faders.io/

within the frame of their attention at any given moment. For technology like IMS to provide context-aware assistance, system functionality needs mixing rules and also abilities to make multi-criteria decisions about options from among those that are available, likely to be effective, and desirable, based on the musical content.

## 1 A CASE FOR CONTEXT

The working context for each mixing project affords certain degrees of freedom for action; it imposes constraints. Magnusson separates subjectively imposed and objective constraints. “Objective constraints represent the physical limitations of the environment or physical material and the designed constraints” that result from technology or system configurations and functionality [16, p. 54]. The physical environment of the studio, including the acoustic space, equipment, furnishings, and anything that can physically impact the mixing of sounds, is potentially pertinent to the mixer’s working context because it influences mixing actions. Subjective self-imposed constraints are not directly accessible for observation, and therefore it is hard to represent them as system rules, to model them, and to incorporate them into a system’s context awareness. Nevertheless, as they are part of a mixer’s working context, a context-aware IMS can be aware that subjective impressions are a factor in mixing.

Some objective constraints are similarly difficult to generalize for a system because they emerge under specific physical conditions such as a studio location. In the ecological sense, the mixer’s physical environment mediates all perceptual information within. The environment surrounds the mixer, and in addition to biology, physical position and activity within the environment determines what the mixer perceives. This combination of factors is referred to as a *vantage*. Vantage is the sum total of a perceiver’s position relative to objects that might influence perceptions, perceptual acuity, and what information holds relevance for the perceiver [17]. All collaborators in a production space work not only within but also with what the environment makes perceptible [18]; for example, by adapting to it for musical listening [19].

The environment *situates* [20] the mixing work. Mixing is situated in a social and socio-technical environment as much as a physical one. A mixer’s actions are potentially influenced by, for example, a client’s presence at the mixing session. Clients bring their expectations and priorities, knowledge of musical genres, technical standards, conventions, practices, biases, and preferences. This perspective can encroach on the mixer’s actions and decisions.

In a socio-technical activity like mixing “interaction between people and machines necessarily implies mutual intelligibility, or shared understanding” [20, p. 7] of the work and the use of technology in that work. For example, in pop music, it is understood that quantization is usually applied to rhythmically align musical parts. It might be employed for other purposes, but actions are purposeful and intelligible only within the circumstances in which they occur [20]. When people work together anywhere to produce anything “the coherence of situated action is tied in essential ways not to a priori prescriptions, but to the action’s particular circumstances” [20, p. 21], or context. Technology plays a crucial role in a music production working context and the role of intelligent machines is growing [21]. And intelligent tools remain largely indifferent to the user’s circumstances.

### 1.1 Experience, Appropriateness, and Appreciating the Context

Mixers are not all equally sensitive to the social, technical, and physical conditions of their environment, nor even aware that each mixing context presents unique possibilities and constraints. Knowledge of mixing praxis, common practices, and mixing experience teach a mixer to direct attention toward significant attributes and issues and to interrogate technological and sonic affordances. Experience informs a mixer’s ability to foresee the utility and appropriateness of a particular tool for a particular application in a particular context.

Experienced mixers know which problems are likely to be encountered and have abilities to identify reliable solutions to such challenges. They are able to detect which problems are most relevant given a particular mix and then discriminate carefully among options given those conditions. The deeper the mixers’ general understanding of the mixing space and problem space, technology, acoustics, musical considerations, and mixing aesthetics, the more subtle the realm of possibilities appears to be in the context of a particular mix because they have experienced a greater number of variants and solutions. Also, the more experienced mixer can envision more plausible approaches for meeting expectations. They perceive flexibility in available ways to work with and around conventional practices and attributes since experience gives them the confidence to measure decisions against the context for appropriateness.

For example, a mixer with a thorough appreciation of a particular effects processor and its conventional application can make informed decisions about adjusting its parameters in atypical ways. They might even misappropriate the tool to achieve a novel effect, if the context makes that choice appropriate. By contrast, less experienced mixers may know in broad terms what they want to do with the processor, but they are less certain about how exactly to realize it in a mix or the appropriate margins for variation [22]. Since they are still learning the rules and procedures, they are more likely to adhere to strict interpretations of conventional axioms and practices rather than adapt them given what a particular context makes possible.

Experience is not the only means through which mixers can build knowledge of mixing technologies and traditional paradigms for their use. This knowledge is also accrued by reading manuals or about mixing practices and by watching video tutorials produced by professionals and/or hobbyists. Learning resources are abundant, and as a result, inexperienced, amateur engineers are increasingly enabled in their attempts to mix somewhat like experienced engineers.

Hoare et al. [23] observes that between the amateur and professional mixer is a growing base of professional amateur mixers (pro-ams) [24]. The Fusing Audio and Semantic Technology for Intelligent Music Production and Consumption (FAST-IMPACT Project) also acknowledges the growing significance of pro-ams in the production to consumption chain. FAST-IMPACT also draws attention to the developing relationship between pro-ams and intelligent mixing technologies [25]. It seems that many pro-ams are looking to attain professional-sounding results without much concern for how this goal is achieved [22]. As compared to the experienced engineer, pro-ams are usually less sensitive to influences within their circumstances that could impact the intelligibility of certain actions. They know less about adapting rules and margins for variation. In these matters, context-aware IMS could provide them with assistance.

Ronan et al. show that professionals and amateurs react differently to different mixes of the same musical content [26]. In particular, they make different associations between mix attributes and emotional cues. Professionals, who are highly critical listeners, demonstrated higher levels of valence and arousal in response to music production quality than their amateur, non-critical counterparts [27]. As Izhaki asserted, emotional impact is a vital consideration in the context of music mixing [4]. Taking decisions with the intention of influencing listeners' valence and arousal levels requires context sensitivity [27].

The pro-am in mixing may differ from pro-ams in other creative fields. Leadbeater and Miller, who observed the working practices of pro-ams in fields other than music production, suggest the pro-am's work "involves the deployment of publicly accredited knowledge and skills, often built up over a long career" [28, p. 20]. It appears to be otherwise in music production. Hoare et al. [23] and McGrath et al. [22] followed mixer pro-ams through different studio setups and found that amateurs, pro-ams, and professionals have different technical needs. It appears that pro-ams are able to function seemingly at a professional level in the absence of experience and deep knowledge of mixing practices because they have access to and the abilities to wield sophisticated and increasingly intelligent tools. The more intelligent the tool, the more professional sounding the pro-am's product. The more context aware the IMS, the more a pro-am's mixing would appear to be a response to specific material and circumstances. But these reactions are difficult to replicate with IMS. Judgments about appropriateness and context are largely tacit knowledge and difficult to represent as rules for an IMS. Mixing, in practice, combines tacit mixing knowledge, broad knowledge of aesthetics, and expressive communication and rules.

## 2 CONVENTIONS, TRADITIONS, AND SEMANTICS

Mixing also requires knowledge of genre conventions, technical standards, and best practices. Some constants do hold mix to mix, context to context. In such cases, rule-based features and recommendations made by an IMS are very useful and appropriate. Technologies that have been designed around conventions or common rules can have great utility for both experienced and inexperienced mixers.

Mixers of all levels attempt to some extent to recreate sounds from other mixes. Experienced and less experienced mixers both have awareness of mixing codes and canon. For example, in popular music, ranges of parameter settings on processors are associated with particular genres. Some processing techniques are so prevalent that they are considered traditional paradigms for technology use. For example, the excessive pitch correction employed in contemporary R&B and hip-hop is a kind of *musical code* applied by the mixer to communicate something specific about the genre and artist [29]. It is important to distinguish here between kinds and instances. Experienced mixers create sounds that belong to categories and that are also context appropriate. The quality of being emblematic does not guarantee appropriateness.

Similarly there are standards that regulate the technical composition of a mix regardless of context and enable professionals to deliver work that is technically appropriate for mastering, distribution, and broadcast. Professional audio technologies offer minimal and maximal levels of bit depth and sampling rates because these ranges correspond to levels of audio quality that have been formally standardized. There are many standards that mixers are required to meet, for example to publish a track on any online platform. The product needs to match certain standardized levels of audio quality. Each online distributor defines their own standards, and as such different platforms require different standards, and this is acknowledged in stated requirements (for example, by Apple[5]). These are implicit norms in the context of professional music production. By offering functions that enable conformity, technology designs acknowledge that standards and conventions are inherently part of the mixing context.

Technologies for all kinds of artistic production are routinely and intentionally designed to normalize to conventions [30]. IMS should not be viewed differently in these regards. Some conventions or norms are so fundamental that they are taken for granted, such as tonal harmony in western music, which is the norm in western pop music. Western tonal harmony has implications for tuning and intonation and thus for any pitch-related signal processing. There are also western conventions regarding rhythmic organization [31]. The implicit expectation that mixers working in a pop production context will adhere to these norms is evident in the functionality of every digital audio workstation (DAW), every pitch correction effect, every drum machine, quantizer, etc.

Assistance in matching conventions is in many respects desirable, but simultaneously it can interfere with a user's abilities to innovate [32]. Also, it limits a technology's applicability across different cultures, musical traditions, and contexts that fall outside the norm [33]. A system design assumes a particular kind of context, whether the user is aware of these assumptions or not. Both experienced and less experienced mix engineers have reasons to recreate some conventions, norms, and standards. This is expected and along these lines, their thinking aligns well with rule-based IMS.

[5]https://www.apple.com/itunes/mastered-for-itunes/

All these common practices, genre-related conventions, and traditional paradigms for effects use together give rise to mixing semantics and ontologies. These core concepts and their associated descriptors are an implicit part of the mixer's working context, just as linguistic semantics and ontological positioning are implicit in the context of language usage. Every conversation is structured by grammatical rules that govern what constitutes a well-formulated sentence. Analogously, there are implicit rules that structure mixes, and these rules may help to make IMS more context aware.

SA research studies the mixing semantics of practitioners and formalizes mixing lexicons and ontologies of mixing techniques, types of sounds and timbres, balances, etc. [34–36]. As a result of this research, many descriptors for common types of sounds, effects, signal processing, and processors have been codified [5, 37, 38]. Tool designers are increasingly trying to leverage these findings to make informed choices about how to label parameters or presets on mixing technologies. Since these semantics are part of the context in which a mixer works, embedding them in the technologies themselves makes it even easier for users to comply with expected norms and standards.

This has obvious utility unless it inadvertently discourages or prevents mixers from exercising a degree of creative freedom or breaking with convention to innovate as appropriate. At times, to create novel and also aesthetically interesting mixes, mixers break with traditional paradigms and even misappropriate technology to create effects that appear novel and particularly expressive given the musical context. An example might be applying a plugin that simulates a guitar amplifier, such as Guitar Amp Designer in Logic Pro X, to a voice or drum. This could be considered subversive or breaking with a norm, and such decisions are not captured by any standard ontology.

Professional mixers deliver on three fronts. They match norms, such as genre conventions and technical standards. They defy some expectations by working around norms in creative, innovative, or unexpected ways. Finally, mixers exude an expressive, personal aesthetic style. The mixer's ability to find confluence among these three deliverables is context dependent. The context determines which norms are to be met and which are appropriate to work around, whether novelty appears novel or not, and the significance and coherence of aesthetic choices. Technology, especially semantic or intelligent technology, can potentially assist the mixer in delivering in all three areas because a mixer's available technology draws attention to information that is presumably relevant to the mixing context.

## 3 TECHNOLOGY AND DECISION MAKING

System functionality and the way information is exposed to the user influences users' decisions and is therefore important in defining a context. Decision makers respond to how decisions are "framed" [39], and interface designs frame choices by activating perceptual and cognitive information processing [40]. Users act on information that is clear to them. The presentation of information about parameter settings and processing options can make some choices more clear than others.

Decision makers are generally more willing to take risky choices but not ambiguous ones [41]. In mixing technology, semantics may help to clarify the relevance of certain choices in a certain context, but mixers appreciate descriptors differently. Appreciation, level of experience, and context awareness are interconnected. Take as an example two equalization effects, one that uses as a parameter label the semantic description *warmth* and another that uses frequencies and ranges displayed as numbers. An engineer who knows how to scan regions for frequencies that are appropriate is more likely to find the numbers to be clear. While scanning, that mixer may even take opportunities to explore unconventional, aesthetically motivated options. A warmth label, in this situation, may appear ambiguous. It may not seem to offer the user the desired amount of precision and control. To a less experienced engineer or pro-am the warmth label is likely to be more clear than numbers. A pro-am is less certain about how exactly to set an EQ to boost warmth, and a potential problem here is that the setting may give the instrument a conventionally or objectively warmer sound without producing a timbre that is well suited to the particular mix or context.

Without the experience, knowledge, and appreciation of context appropriateness, mixers are understandably more likely to rely on technical functionality, like IMS, that promises to deliver sounds that fit within a generic category, whether or not they are intelligible in their exact mixing situation. Presets are another type of functionality that promises to deliver sounds that fit within generic categories and without awareness. IMS can take some cues from how presets are used in practice.

### 3.1 Presets and Bounded Exploration

Presets are ubiquitous in audio technology. When they came into prominence in the late 1970s in synthesizers, the manufacturers of these instruments were assuming that many users would be unlikely to program their own sounds [42]. Théberge argues that the Yamaha DX-7's oversimplistic interface design actively discouraged the creation of new sounds. This conclusion was based on the observations of repair engineers who noticed that when keyboards were returned to the factory for service the presets had not been edited or modified in any way [42]. In other words, they were never adapted to a specific context.

The primary purpose of a preset in music production software is to provide the user with the means to easily navigate the control parameter space. Technology designers may include presets to mitigate issues arising from a large number of varied controls exposed to the user or complex inter-dependencies among controls. They may also serve as aids to new users, allowing them to experience the full extent and capabilities of an individual effect. Presets are designed by the technology manufacturer to have utility in many user contexts. That potential might exist because they are conventional parameter settings or otherwise typical of a commonly used effect. Some mixers develop their abilities to *interrogate* presets for their usefulness. They are then able to find sounds they like with few parameter adjustments. Other engineers take a preset to be merely a suggestion for how it might manipulate the character of an instrument. It is a tool for ideation rather than a solution.

Paterson observes that different levels of experience correlate to different behaviors around the use of presets. The amateur is "likely to wish to use named presets as a starting point in their productions, possibly not even adjusting them to any great effect," meaning that exploration of the potential or experimentation "is potentially negated" [43]. A key difference between an amateur and an experienced mixer is that an experienced engineer knows that a generic preset, for instance, for compressing a bass is not appropriate for all basses or contexts. Using presets as a suggestion is a way to find potential in the context of a particular mix. The experienced mixer knows how to interrogate the options available in pursuit of appropriate options and how to adjust parameters on a preset to make it appropriate.

Paterson observes that the use of presets has "long been scorned by professional producers" [43], who possess the wherewithal to render the sounds they intend and imagine without assistance. Furthermore, because of their comparatively broader experience, professionals can imagine applying preset effects in numerous and varied situations and therefore may even be inspired to appropriate that preset for non-traditional application [44].

Experienced mixers are empowered to misappropriate because they know how to evaluate appropriateness in a specific context and generally perceive more degrees of freedom and options. They have deeper appreciation of constraints. "Constraints map out a territory of structural possibilities which can then be explored" [45, p. 95]. Context steers the experienced mixer's workflow; it narrows the plausible choices. Within those limits, a constraint may be "transformed to give another one" [45, p. 95]. When experienced mixers hit the boundaries, they turn to exploring their problem space in divergent and sometimes creative ways. Their experience has taught them how to find potential and constraint workarounds, and in this way limitations inspire exploration [44]. Explorations are more likely to be fruitful when the context can be used to clearly identify the limits of appropriateness. Otherwise, too many options are available to the mixer, without any clear strategies for evaluating them.

It is ironic that the preset feature originally emerged to serve the needs of amateurs [42] who do not have highly developed skills and the experience to make context-specific distinctions. In some respects, by conforming to rules, IMS is following suit. Since the amateur may not make the most informed decisions about the appropriateness of presets, an intelligent system could actively help by reducing the user's choices to a set of options that are likely to be appropriate given the context, the other presets in use, and other information about the mix (e.g., spectral balance, dynamics, etc.). For mixers who are highly aware of the importance of the context, context-appropriate presets foster creative exploration of the mixing problem space. They are a tool for rapidly simulating subtly varied appropriate possibilities.

## 4 CONTEXT AND STYLE

Appropriate choices are often those that have appeared previously in mixes created within similar contexts. To achieve these ends, mixers' workflow decisions are at least partially guided by established methods developed by others and by paradigms established over their own practices. Utilizing these methods requires at the very least familiarity with common as well as contemporary tools, and knowledge of their conventional applications inform best practices [46]. However, in many contexts, traditional methods are also important and may be growing in importance. Some mixers use terms such as *retro* or *vintage*, and distinctive *nostalgic* sounds are very desirable in many genres [47]. Therefore, in many contexts, both contemporary and traditional paradigms of tool use may be appropriate choices. Mixers not only need to have the skills to create retro attributes, they also need to determine when *nostalgia* is appropriate. These determinations may be informed by experience or an IMS.

### 4.1 Nostalgia

Nostalgia is an example of an attribute that is desirable in certain contexts, or nostalgia may be considered context in its own right. A mixing language has developed to describe the attributes of nostalgia. It is typically associated with semantic descriptors such as warmth. These labels or parameters can be found in audio effects. Similarly there are parameters in effects that can be mapped to emotions that are also suggestive of retro, vintage, or nostalgia [47]. Nostalgia narrows a mixer's set of appropriate choices. Thus IMS that can render retro-sounding attributes, recreate traditional paradigms of application, and be aware of contemporary mixing contexts could be a powerful tool particularly for mixers with less knowledge of traditional methods.

Awareness in the nostalgic idiom requires an appreciation of the deficiencies of vintage studio equipment. Limitations like the sounds associated with tape, tubes, and transformers are not considered technical flaws but part of their aura and inherent appeal [48, 49]. Outside the context of nostalgia, similar characteristics are avoided. Nostalgia-capable IMS would need to discriminate among desirable vintage and undesirable distortions. Moreover, a nostalgia-capable IMS may have to adapt to a range of nostalgic possibilities, spanning from a recreated vintage mix to a contemporary mix with retro elements. Also, as nostalgic sounds are frequently desirable, processors are sometimes chosen not for their traditional paradigm of use but to produce these artifacts. A compressor, for example, might be chosen for the harmonic distortion properties exhibited by the hardware unit or plugin rather than its ability to control dynamic range [50].

Pestana [5] highlights that there is more than one way to achieve a desired mix result. Both compression and distortion, for example, are able to alter the shape and envelope of a sound and thereby change its perceptual attributes [51]. Either might be utilized effectively depending on the context of desired qualities, of the signal(s) being processed, the other sounds in the mix, the affordances of the particular processors available to create the desired effect, etc. To provide useful and context-aware assistance, a nostalgia-capable IMS may need to suggest to the user very different modes of applications. Where there are numerous ways to achieve seemingly appropriate effects, an IMS will need to differentiate between the appropriateness of multiple types of audio processing to select the one most suited to the context of the program material.

Professionals, pro-ams, and amateurs do not approach imitating retro and vintage sounds in the same way and require different kinds of assistance. The experienced mixer listens very critically when appraising these qualities. The less experienced mixer may have only ever heard emulations of vintage sounds. Each is therefore likely to have different subjective impressions of the color, saturation, warmth, and brightness. So, particularly in contexts where the mixer tried to create vintage sound in the absence of the original analog devices, a nostalgia-capable IMS may need to assist the mixer in discriminating among choices by directing the user's attention toward particular attributes.

Mixers refer to a retro digital aesthetic or *digital warmth* [47]. Within many genres, retro and *Lo-Fi* aesthetics are embraced as a particular form of artistic expression [29]. Given that these terms are so widely used in varied genres and contexts, a pertinent question for designing nostalgia-capable IMS is 'What do terms such as retro, vintage, and nostalgia sound like?'

There appears to be some agreement about the semantic meaning and timbral attributes of warmth (associated with nostalgia), brightness, air, punch, and space. Pearce et al. identified a number of commonly used audio semantic terms and using statistical analyses created models of what they represent as plotted along a scale (e.g., not bright to very bright) [52]. However, there are many terms that still lack models or are concepts that require relative discernment. For example, in digital effects, the glitchy sound that could be achieved with early versions of Antares Autotune remains highly desirable [29]. That effect, though not an application intended by the system designers, is prevalent among artists such as Kanye West or T-Pain. Future IMS may be able to similarly exploit the limitations of mixing technologies to create effects that sound both surprising and appropriate if IMS become capable of determining what is appropriate given the context.

### 4.2 Misappropriation

Mixers tend to deliberately choose to work around traditional paradigms of use and avoid traditional sounds. They *misappropriate* tools to create new sounds and effects [29]. The context determines what sort of application is appropriate and also if an application appears non-traditional or novel. Misappropriation for the purposes of this discussion can be loosely defined as using a tool for an *unintended* purpose. When a mixer misappropriates, they are questioning the notion of appropriateness [50]. For example, an Opto style device, such as a Universal Audio LA2A, might be the more traditional choice for compressing a bass, but a VCA style dynamic range compressor such as a DBX 160 VU may suit the particular context better [53]. Another example is using a preset that had been intended for processing guitars on a drum loop instead. The creative impact of making such decisions can have far-reaching consequences and implications for the direction of the entire mix. In other words, it changes the context for further work on the mix.

'Creep,' a song by Radiohead, is a good example of working around traditional paradigms. The song contains a guitar part in the chorus that is likely to be considered by most mixers to be far too loud, as compared to conventional mixes of a similar style. It subverts the idea of what a balance should be like. Also, the guitar part heavily masks the lead vocals at points in the song, breaking another convention. However, given the musical material, genre, and artist, these choices are considered not just appropriate but highly affective. Similarly, in the appropriate contexts, the misappropriation of a preset can make a mix more interesting [54].

These kinds of misappropriations and creative and novel solutions may emerge, as Boden [45] explains, from how a problem space is explored. Human mixers do not explore their problem spaces randomly. Their explorations are bounded by knowledge of the context and what is potentially appropriate in it. A system capable of breaking mixing rules appropriately needs rules for breaking rules or to be trained on more than conventional mixes.

Yet even without this level of sophistication, a system that could merely determine what is conventionally appropriate or not in a given a context could still provide a lot of intelligent assistance to the user. It may function as a motivator, encouraging users to take interest in attributes or elements they may have overlooked. Consequently, the interactions between a computational system and human might be compared to that of two mixers or producers creatively bouncing ideas off each other. Additionally, IMS could have a role to play in cautioning against misappropriation where it is not likely to be perceived as appropriate. These kinds of intelligence are not trivial to implement. However, research in related fields hint at potential context-oriented solutions for IMS.

## 5 CONTEXT IN SEMANTIC AUDIO AND MUSIC INFORMATION RETRIEVAL

Some areas of SA and MIR research have investigated the semantics of context and how to utilize information about the user's context or manage the lack of it. For some time, SA and MIR along with Semantic web research have been wrestling with the problem of managing the differences between what humans know, understand, and expect and what machines are capable of representing and figuring out autonomously. This issue is referred to as the *semantic* *gap*. As of yet, there is no consensus around an exact definition of the semantic gap. However, Hein's definition, "the difference in meaning between constructs formed within different representation systems" [55], is highly relevant to this discussion. In the domain of Music Informatics, Schedl, Gómez, and Urbano have attempted to quantify the gap and refer to it as a "mismatch between machine-extractable music features and semantic descriptors that are meaningful to human music perception" [56, p. 6].

One key issue for those working on this problem is how to represent human knowledge in a machine interpretable way. There is a lack of understanding around knowledge and mental processing and this impedes the formulation of robust representations. Gaps between human and machine understanding are widest when they involve very abstract concepts such as context. Although it is possible to describe the process of adapting to context in general terms, it is very difficult to generate consistent axioms and rules that cover many or all eventualities.

Casey et al. suggest that many MIR studies lack "high-level intuitive information about music embod[ying] the types of knowledge that a sophisticated listener would have about a piece of music," such as abilities to appraise musical surface, timbre, or emotion. These features, they contend, are derivable from extractable features with appropriate models of cognition, but these have not yet been sufficiently developed [57, p. 671]. Wiggins argues that the semantic gap is not even visible from the *auditory domain* of extractable or perceptible features; and complex musical knowledge, for example, about interpretation and the structure of musical phrases are not easily translated into machine readable axioms. The gap is further exacerbated, says Wiggins, by musical experience, memory, culture, linguistics, and word-grounding, all of which may be associated with context and all of which are inseparable from musical inference, musical exchange, and communication and the making of meaning in music [58].

The modeling of mixing intelligence in IMS is not yet confronting the semantic gap. Ontologies have been engineered for SA and MIR technologies that could be co-opted for building up the context awareness of IMS. For example, Raimond et al. propose *the Music Ontology*, "a formal framework for dealing with music-related information on the Semantic Web. It includes editorial, cultural and acoustic information" [59, p. 1]. It assumes a workflow that spans the production of a musical recording to the recording's release and incorporates concepts pertaining to "music production, music consumption, music recommendation, etc." [59, p. 1]. It builds on a generic knowledge representation and knowledge sharing technology, the W3C's Web Ontology Language (OWL),[6] which is a logic-based formal semantic language for representing *things* or entities in any domain, and the relationships between the designated entities. Logic is used as grounding for OWL to support automated verification and inference from instance data expressed using the language and entailment rules associated with its constructs.

[6]https://www.w3.org/TR/owl2-overview/

The Music Ontology also inherits concepts from relevant domain ontologies such as the Timeline[7] and Event[8] ontologies, which provide concepts that are not music specific but highly important in the description of music [60]. The concept of a *Timeline* is a "coherent backbone for addressing temporal information." Other temporal concepts include *Intervals* and *Instances*, which may be used to describe *Events* and chain them in a workflow or provenance description. Music production concepts include, for example, *Work*, *Person*, *Group*, *Record*, *Track*, *Stream*, *Performance*, *Mixing* (limited to DJ mixing), and *Sampling*. These concepts may have a place in a mixing ontology. They may help to establish the context, but they are many steps removed from the knowledge necessary to take a mixing decision.

Fazekas and Sandler propose a music production ontology that captures more studio-based concepts [34] in the context of a proposed intelligent DAW and associated Multitrack ontology.[9] It too inherits well-established non-(music) domain specific concepts and standardized ontologies and uses OWL. It captures information about multitracking and distinguishes between *Clips* and *Tracks* and even *Multitrack sessions* as well as track types in a DAW, *Mediatrack*, *Videotrack*, *Audiotrack*, and so on. The granularity extends to *Automationtrack*, *Automationevent*, and *Automationparameter* and details like *Microphone* and the *Location* of that microphone. This is more directly relevant to the mixer's domain but still barely covers what is *meaningful* to a mixer in an actual mixing scenario. The aforementioned ontologies have been incorporated into a broader framework that aims to provide an application and situation independent conceptualization of the entire recording studio domain [38, 61].

The Studio Ontology (SO) Framework[10] extends the Music Ontology, adding details about the music production workflow. It covers events and actions that fall between capturing an audio signal and releasing a track or album. The core of the SO is a parallel event and signal flow model. This model can be used to describe, for instance, the precise placement of a microphone, the mixing of the produced signal with other signals in the environment, or a sequence of audio signal transformations such as the application of filters, audio effects, and elements of final mastering. The signal flow is associated with the event flow, which helps to capture the precise signal routing in a studio setup. Each event in the model may be associated with agents. For example, it can describe who was responsible for a decision such as the choice and settings of a compressor, or it can describe individual mixing decisions like moving a fader.

Besides the core model, the SO includes a number of foundational elements, consisting of small domain ontologies in support of its model. The Device Ontology and Connectivity Ontology, for example, define entities for describing low-level detail such as (studio) device terminals,

[7]http://purl.org/NET/c4dm/timeline.owl

[8]http://purl.org/NET/c4dm/event.owl

[9]https://sourceforge.net/p/motools/code/HEAD/tree/studio/multitrack.owl

[10]http://isophonics.net/content/studio-ontology

protocols, and audio connector types. High-level extensions include domain ontologies for microphones, mixing, basic audio effects, and audio editing. The framework also provides hooks for domain-specific extensions such as detailed models of audio effects described in the Audio Effects Ontology [62] using a conceptual layering system originally proposed for intellectual works [63].

The aim of the SO Framework and associated ontologies is to capture the technical workflow rather than the intent of the mixer. Therefore it falls short of providing sufficient means for mapping mixing decisions to goals, intent, context, and all other human considerations involved in mixing decisions. There have been some IMS approaches that use these ontologies, defined rules, and constraint optimization methods for musical mixing [36]. These ontologies have also been used to represent music production data captured from a DAW in order to learn associations between semantic labels, mixing decisions, and acoustic features [35].

### 5.1 Semantics in Intelligent Mixing Tools

Some IMS use semantic descriptions to connect user input and expectations with machine functionality. Semantic descriptors are used to label parameters in audio effects, implying that adjustments help the engineer to achieve these sounds (e.g., making a sound boomy or thin). The *Semantic Audio Feature Extraction* (SAFE) project [64] investigated the relationships between semantic descriptors and the resulting processed signal, specifically in terms of objectively measurable acoustic features. As part of this research, the SAFE project released four VST plugins: a parametric equalizer, dynamic range compressor, distortion, and reverb effect. Each required the user to enter a sound descriptor associated with the applied processing as well as information about the context in which the transformation is taking place. This contextual information includes genre, location, and demographic data of the user.

SAFE identified a set of semantic descriptors that correlate with clusters of equalization or compression parameters. The research also showed that compression and equalization share a similar vocabulary, but reverb and distortion have a dissimilar description schema [65]. It was informally observed that the requirement to provide contextualizing information increases the variability of semantic descriptors, suggesting that different contexts evoke different associations. For instance, an R&B producer in New York and Grime producer in London may use different terms to describe the same transformation. Hence, it can be inferred that descriptions depend on the genre and locality of the production [66].

Semantic descriptors for emotional content are also being utilized in intelligent mixing tools. Ronan et al. have shown mixing decisions are guided by emotions perceived to be expressed by the musical content or that are induced by listening to it [27]. There are various models of music and emotion that provide a basis for connecting emotion perception to decision making. Appraisal theories such as the component process appraisal model proposes five functional components of emotion: cognitive, peripheral efference, motivational, motor expression, and subjective feeling [67]. A number of studies discussed in [68] show the relevance of this model to the perception of sound more generally.

Emotion as well as other contextualizing information such as genre, style, and semantic descriptors have also been shown to assist professional and amateur users of music libraries, who expect this information to map to particular acoustic features [69]. An IMS with the necessary capabilities to detect emotional content and emotional context could factor both into numerous mixing decisions, such as those that impact timbre or balance. It could also assist less experienced engineers who may be less sensitive to emotional cues by directing their attention to pertinent considerations. Few systems are capable of inferring context from an audio signal to assist mixing decisions, e.g., the configuration of audio effects [70]. These approaches however only represent baby steps toward context-aware IMS.

## 6 CONTEXT AWARENESS IN CAC

The field of CAC offers additional insight for IMS. CAC seeks solutions that make computational assistance and services “adaptable, responsive, personalized, dynamic, and anticipatory” [71, p. 5]. In broad terms, context-aware applications are those that are able to use information about context. They ordinarily fall into three categories: “the presentation of information and services to the user” that are relevant to the user’s context, “automatic execution of a service” and “contextual augmentation” or “tagging of contextual information” to enhance interactions [72, p. 59].

As with IMS design, intelligent systems engineering of any sort requires a positivistic approach that treats context as “a form of information. . . delineable. . . stable. . . context and activity are separable” [13, p. 25]. Given this definition, it is not immediately obvious how CAC and IMS can be compatible with the exploratory, artistic, and novelty-seeking modes of mixing described above. In mixing it is often hard to predict the extent to which a given aspect of the context will be relevant to actions and decisions. As with CAC, in IMS, it appears that “the kind of thing that can be modelled. . . is not the kind of thing that context is” [13, pp. 26–27].

There are *operational* solutions to address the challenges of acquiring contextual data [72, p. 59], for example querying users directly about the contexts in which they work, as in the SAFE project. However, experience from CAC suggests that this does not work well for very abstract concepts and axioms. This is particularly true when a given context may only be relevant some of the time and consequently only partially or occasionally useful. Dey and Abowd argue that deriving contextual information from user input is problematic because the notion of a user input suggests one user and one application, but context implies varied extra-application contingencies, many of which are other technologies.

A mixer taking a decision about a parameter setting on a single reverb does so in the context of many sonic elements. Although users interact with individual applications and take decisions about those interactions, those interactions occur within a context that typically includes multiple applications [73]. Mixers usually use one DAW, many plugins or processors, and a variety of other technologies, synthesizers, drum machines, triggering devices, work, or control surfaces, etc. Each technology is designed as an individual system. Each user interface is designed for that particular system. But these systems are often technically, musically, or conceptually interconnected. Peripheral technologies can impact a mixer's interactions with the DAW and plugins within the DAW. To contextualize interactions with any one of these applications, that application requires knowledge representations with additional layers of *abstraction* [73]. Context-aware IMS require awareness of inter-operations and other inter-relations among technical components.

The distributed nature of a production might also frame the context. Tasks or any form of user cognition may be distributed among clients and collaborators. Context-aware IMS may need to keep track of who contributes what and differences among contributors. The SAFE project [64] worked with and around some of these issues. Users were asked to describe effects and the system adjusted to those descriptions. SAFE queried users directly about location and other demographic data. Users also specified genre.

In SAFE the relevance of the user's location was not always clear. Mixers' semantics appeared to be influenced by their geography. Geography probably does have an impact on style [10]. In SAFE, users tended to set the demographic information only once, even when they worked on multiple mixes. In those cases, location information could not be reliably used as a criterion for mixing decisions. On the other hand, music production is a global business. It is not unusual for professional mixers to have clients from other countries and the target audience may be a global audience. IMS may be able to utilize information about the mixer, client, and audience in mixing decisions and might assist mixers by compensating for localizing factors when they are not desirable.

Finally, context awareness could arise from distributed processing [73] such as information gathered from the world wide web or a centralized server. For example, mixers frequently use reference recordings. These tracks could be accessed and analyzed by an IMS to refine and contextualize the assistance it provides. Also, systems may use crowd sourcing for decisions or tap into web-accessible knowledge repositories. As of yet, there do not appear to be many mixing technologies with these capabilities, but such functionality could contribute to the sophistication of IMS in the future.

To manage the complexity of context, Abowd and Dey suggest categorizing contexts and thereby reducing definitions to only the most relevant information given a particular task or scenario [74]. However, Dourish cautions that "contextuality. . . is not. . . made a priori. It is an emergent feature of the interaction, determined in the moment and in the doing" and by what has "general bearing" [13, pp. 29, 28] on a task. It may be possible to shift among categories if the IMS has the information and procedures necessary to make those shifts intelligently.

A mixer's perspective on context shifts from low (perceptual) and high (abstract) aspects, as they seek relationships that connect each detail with impressions of the mix as whole. There is much that might be *sensed* by a context-aware system [75] and from that abstractly defined features of the context could be inferred. In Gray and Salber's definition of "sensed context," "properties of real world phenomena" are extracted using sensor technology, but only those properties that are used by system functionality and make it "more effective or usable" are considered relevant to the sensed context [75, p. 318]. "Not only does high-level context depend on low-level sensed context but also different kinds of low-level parameters are interrelated" [71, p. 12].

Generating information about context this way is somewhat analogous to the MIR technique of extracting signal features and using them to infer musically relevant information. MIR utilizes features such as *temporal shape* as in attack time; *energy*-related descriptors, *global energy*, *noise energy*, etc.; *spectral shape* including the *centroid*, *spread*, short-time fourier transforms, etc.; *harmonic features*; and various *perceptual features* including *loudness*, *sharpness*, etc. [76]. With low-level features like these, it is possible to determine things like beat, and hence tempo. Then patterns detected across low-level features may be correlated and matched to complex concepts as they have been predefined.

All these approaches could go a long way in realizing the potential of context-aware IMS. Still, thorough definitions of mixing contexts may be forever out of reach. All "situated action emphasizes the improvisational aspects of human behavior and deemphasizes a priori plans that the person simply executes" [77, p. 52]. Context-aware systems will always be *imperfect* and must always contend with what is *unknown* (undetectable), *ambiguous*, *imprecise*, or *erroneous* [71, p. 9]. Chalmers suggests "allowing applications to choose what aspects they interpret" as a way "to cope with temporary loss of some aspects of context." They may make substitutions for "missing data" [78, p. 69]. Yet even a minimally context-aware system with the ability to detect that something is ambiguous, erroneous, non-normative, or conventional may provide utility to a mixer.

For example, if tracks have primarily been mixed consistently and following a traditional paradigm, an IMS could identify where and how outlier tracks fail to meet the conventions met by the rest of the mix. That information alone is enough to provoke context awareness in mixers and thereby inspire and enable them to better situate their work and creative choices within the context of their circumstances. Dourish notes, "*being ordinary* is something that people work at, by acting in ways that they understand to be the normal activities of the groups of which they are members, and making no issue out of them in the course of their interactions" [13, p. 31]. By even minimally assisting in situating the mixer's work, by constructively challenging the intelligibility of actions and choices, a context-aware IMS could encourage epistemic reflection, creative exploration, novelty-seeking approaches, and more context-aware decisions.

## 7 NEXT STEPS AND CONCLUSION

This article highlights the importance of context for music mixing and IMS. Deepening our understanding of the influences of context on mixing enables us to develop an IMS that is more flexible and better aids mixers in identifying relevant aspects of context given the situation of the mix. Both these features could facilitate greater creative freedom. To this end, much can be learned from research on mixing practices as well as from SA, MIR, and CAC, all of which are dealing with related challenges.

It is clear from prior work that there are a number of different types of context, and they can be represented in a range of different ways. One of the most commonly considered contextualizing factors in mixing, and thus particularly applicable to IMS, is musical style and genre. Though genre remains difficult to define, even a basic awareness as to the intended style of a mix impacts the mixer, client, and audience's expectations, drastically changing the ways mix decisions are perceived and appropriate mixing conventions are selected.

Mixing is also influenced by mixers' depth of knowledge regarding common practices, mixers' semantics, and the words they use to describe particular musical or sonic attributes. These factors also help to define their working context. There may be geographically determined local practices, too. Mixes are produced in a particular time period. The latest releases may further influence what mixing decisions are expected or appropriate. Similarly, recent technological advances also add to expectations or inspire and make appropriate explorations into new creative territory. Any of the factors can contribute to a mixing context.

The context of a mix is not fixed at the onset or indeed at any time as the work progresses. Mixers' attention shifts naturally over time to different attributes. Thus varied instances of appraisal and perception are contextualized differently with some factors being relevant for some decisions and less relevant or irrelevant for others. Mixing is a feed forward process. Once decisions and actions have been taken, they potentially contextualize future decisions and actions.

Given these observable impacts on mixing, several avenues for making IMS more context aware are immediately apparent. Intelligent mixing tools are already routinely being trained on corpora of examples to identify genre-associated patterns, for example spectral balances. This means they can and do utilize these patterns to adjust mixes or recommend adjustments to mixes in ways that bring mixes in line with the norms. To make this even more context aware, rather than relying solely on corpora compiled and analyzed prior to mixing, an IMS could gather data on the latest musical releases and relevant reference recordings. With this additional material, systems could identify additional, more refined sub-sets of patterns and mix attributes. Also, the more IMS can analyze the choices a mixer makes while mixing, the more those attributes can be factored into the system's context awareness.

A key consideration for making this kind of assistance truly useful for the mixer is how functionality and options are presented to the user. The interface design can draw the user's attention toward different types of decisions, solutions, and problems. It can encourage the user to explore available degrees of freedom and reflect on decisions. For example, an assistive IMS interface might make it simple for a mixer to quickly prototype different genre-appropriate spectral balances. That would enable a mixer to identify which kind of balances can work best given the musical content. The key here is to not merely offer suggestions but step the user through a prototyping process. Such assistance can serve as a starting place and even draw the mixer's attention toward attributes with creative potential.

Finally, IMS is an excellent way to help an engineer explore context itself. IMS can make visible and audible different aspects of context, those that originate internally or externally to the mix itself. Any mix created, by human and/or machine, will rely heavily on and be viewed within a context. As such, the understanding and appreciation for context is a key attribute for any IMS designer to consider.

## 8 REFERENCES

[1] S. S. Horning, "Engineering the Performance: Recording Engineers, Tacit Knowledge and the Art of Controlling Sound," *Soc. Stud. Sci.*, vol. 34, no. 5, pp. 703–731 (2004).

[2] B. Anthony, "Mixing as a Performance: Creative Approaches to the Popular Music Mix Process," *J. Art Rec. Prod.*, vol. 11 (2017).

[3] T. Seay, "Capturing That Philadelphia Sound: A Technical Exploration of Sigma Sound Studios," *J. Art Rec. Prod.*, vol. 6 (2012).

[4] R. Izhaki, *Mixing Audio: Concepts, Practices, and Tools* (Routledge, Abingdon, UK, 2017). https://doi.org/10.4324/9781315716947.

[5] P. D. L. G. Pestana, *Automatic Mixing Systems Using Adaptive Digital Audio Effects*, Ph.D. thesis, Universidade Católica Portuguesa, Porto, Portugal (2013 Feb.).

[6] D. Moffat and M. B. Sandler, "Approaches in Intelligent Music Production," *Arts*, vol. 8, no. 4, p. 125 (2019).

[7] D. Moffat and M. Sandler, "Machine Learning Multitrack Gain Mixing of Drums," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 527.

[8] D. Beer, "The Precarious Double Life of the Recording Engineer," *J. Cult. Res.*, vol. 18, no. 3, pp. 189–202 (2014).

[9] A. J. Zak, *The Poetics of Rock: Cutting Tracks, Making Records* (University of California Press, Berkeley, CA, 2001).

[10] S. Zagorski-Thomas, "The US vs the UK Sound: Meaning in Music Production in the 1970s," in *The Art of Record Production*, pp. 79–97 (Routledge, Abingdon, UK, 2016).

[11] J. R. Covach and A. Flory, *What's That Sound?: An Introduction to Rock and its History* (WW Norton & Company, New York, NY, 2006).

[12] W. Moylan, *Understanding and Crafting the Mix: The Art of Recording* (Routledge, Abingdon, UK, 2014).

[13] P. Dourish, "What We Talk About When We Talk About Context," *Pers. Ubiq. Comput.*, vol. 8, no. 1, pp. 19–30 (2004).

[14] A. K. Dey, "Understanding and Using Context," *Pers. Ubiq. Comput.*, vol. 5, no. 1 (2001).

[15] A. Johnson and R. W. Proctor, *Attention: Theory and Practice* (Sage, Thousand Oaks, CA, 2004).

[16] T. Magnusson, "Designing Constraints: Composing and Performing With Digital Musical Systems," *Comput. Music J.*, vol. 34, no. 4, pp. 62–73 (2010).

[17] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition* (Psychology Press, London, UK, 2014).

[18] M. N. Lefford, "Information, (Inter)action and Collaboration in Record Production Environments," in *The Bloomsbury Handbook of Music Production*, pp. 145–160 (Bloomsbury, London, UK, 2020).

[19] E. F. Clarke, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning* (Oxford University Press, Oxford, UK, 2005).

[20] L. A. Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication* (Cambridge University Press, Cambridge, UK, 1987).

[21] T. Wilmering, D. Moffat, A. Milo, and M. B. Sandler, "A History of Audio Effects," *Appl. Sci.*, vol. 10, no. 3 (2020). https://doi.org/10.3390/app10030791.

[22] S. McGrath, A. Chamberlain, and S. Benford, "Making Music Together: An Exploration of Amateur and Pro-Am Grime Music Production," in *Proceedings of the Audio Mostly Conference*, pp. 186–193 (Norrköping, Sweden) (2016 Oct.).

[23] M. Hoare, S. Benford, R. Jones, and N. Milic-Frayling, "Coming in From the Margins: Amateur Musicians in the Online Age," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1295–1304 (Toronto, Ontario) (2014 Apr.).

[24] P. Nick, "The Rise of the New Amateurs: Popular Music, Digital Technology, and the Fate of Cultural Production/Nick Prior," in J. R. Hall, L. Grindstaff, M. -C. Lo (Eds.), *Handbook of Cultural Sociology*, pp. 398–407 (Routledge, Abingdon, UK, 2010).

[25] M. Sandler, D. De Roure, S. Benford, and K. Page, "Semantic Web Technology for New Experiences Throughout the Music Production-Consumption Chain," in *Proceedings of the 2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*, pp. 49–55 (Milano, Italy) (2019 Jan.).

[26] D. Ronan, D. Moffat, H. Gunes, and J. D. Reiss, "Automatic Subgrouping of Multitrack Audio," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)* (2015).

[27] D. Ronan, J. D. Reiss, and H. Gunes, "An Empirical Approach to the Relationship Between Emotion and Music Production Quality," *arXiv e-prints arXiv:1803.11154* (2018).

[28] C. Leadbeater and P. Miller, *The Pro-Am Revolution: How Enthusiasts Are Changing Our Society and Economy* (Demos, London, UK, 2004).

[29] R. Brøvig-Hanssen and A. Danielsen, *Digital Signatures: The Impact of Digitization on Popular Music Sound* (MIT Press, Cambridge, MA, 2016).

[30] H. S. Becker, "Art as Collective Action," *Am. Soc. Rev.*, vol. 39, no. 6, pp. 767–776 (1974).

[31] E. Quinton, C. Harte, and M. Sandler, "Extraction of Metrical Structure From Music Recordings," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx)* (2015).

[32] S. L. Star and J. R. Griesemer, "Institutional Ecology, Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39," *Social Studies Sci.*, vol. 19, no. 3, pp. 387–420 (1989).

[33] N. Bryan-Kinns, Z. Li, and X. Sun, "On Digital Platforms and AI for Music in the UK and China," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)* (2020).

[34] G. Fazekas and M. Sandler, "Ontology-Based Information Management in Music Production," presented at the *126th Convention of the Audio Engineering Society* (2009 May), paper 7665.

[35] S. Enderby, T. Wilmering, R. Stables, and G. Fazekas, "An RDF Representation for Semantic Music Production Data," in *Proceedings of the 2nd AES Workshop on Intelligent Music Production*, pp. 24–28 (London, UK) (2016).

[36] D. Moffat, F. Thalmann, and M. B. Sandler, "Towards a Semantic Web Representation and Application of Audio Mixing Rules," in *Proceedings of the 4th Workshop on Intelligent Music Production (WIMP)* (2018).

[37] B. De Man and J. D. Reiss, "A Knowledge-Engineered Autonomous Mixing System," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), paper 8961.

[38] G. Fazekas and M. B. Sandler, "The Studio Ontology Framework," in *Proceedings of the 12th International Society for Music Information Retrieval Conference* (2011).

[39] A. Tversky and D. Kahneman, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, vol. 47, no. 2, pp. 263–291 (1979).

[40] E. Giaccardi, "Metadesign as an Emergent Design Culture," *Leonardo*, vol. 38, no. 4, pp. 342–349 (2005).

[41] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, and C. F. Camerer, "Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making," *Science*, vol. 310, no. 5754, pp. 1680–1683 (2005).

[42] P. Théberge, *Any Sound You Can Imagine: Making Music/Consuming Technology* (Wesleyan University Press, Middletown, CT, 1997).

[43] J. Paterson, "The Preset Is Dead; Long Live the Preset," presented at the *130th Convention of the Audio Engineering Society* (2011 May), paper 11.

[44] V. Zappi and A. McPherson, "Hackable Instruments: Supporting Appropriation and Modification in Digital Musical Interaction," *Frontiers ICT*, vol. 5, p. 26 (2018). https://doi.org/10.3389/fict.2018.00026.

[45] M. A. Boden, *The Creative Mind: Myths and Mechanisms* (Routledge, Abingdon, UK, 2004).

[46] P. Pestana and J. Reiss, "Intelligent Audio Production Strategies Informed by Best Practices," in *Proceedings of the AES 53rd International Conference: Semantic Audio* (2014 Jan.), paper S2-2.

[47] G. Bromham, D. Moffat, M. Barthet, A. Danielsen, and G. Fazekas, "The Impact of Audio Effects Processing on the Perception of Brightness and Warmth," in *Proceedings of the ACM Audio Mostly Conference* (2019).

[48] A. Williams, "Technostalgia and the Cry of the Lonely Recordist," *J. Art Rec. Prod.*, vol. 9 (2015).

[49] S. Bennett, "Endless Analogue: Situating Vintage Technologies in the Contemporary Recording Workplace," *J. Art Rec. Prod.*, vol. 7 (2012).

[50] A. Moore, R. Till, and J. Wakefield, "An Investigation Into the Sonic Signature of Three Classic Dynamic Range Compressors," presented at the *140th Convention of the Audio Engineering Society* (2016 May), paper 9572.

[51] T. Wilmering, G. Fazekas, and M. B. Sandler, "Audio Effect Classification Based on Auditory Perceptual Attributes," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), paper 9009.

[52] A. Pearce, T. Brookes, and R. Mason, "Hierarchical Ontology of Timbral Semantic Descriptors," Tech. rep., Audio Commons project deliverable D5.1 (2016 Aug.), https://www.audiocommons.org/materials/.

[53] G. Bromham, D. Moffat, M. Barthet, and G. Fazekas, "The Impact of Compressor Ballistics on the Perceived Style of Music," presented at the *145th Convention of the Audio Engineering Society* (2018 Oct.), paper 10080.

[54] G. Bromham, "How Can Academic Practice Inform Mix-Craft?" in R. Hepworth-Sawyer, J. Hodgson (Eds.), *Mixing Music, Perspectives on Music Production*, pp. 245–256 (Taylor & Francis, Oxfordshire, UK, 2016).

[55] A. M. Hein, "Identification and Bridging of Semantic Gaps in the Context of Multi-Domain Engineering," in *Proceedings of the Forum on Philosophy, Engineering & Technology (fPET)* (2010).

[56] M. Schedl, E. Gómez Gutiérrez, and J. Urbano, "Music Information Retrieval: Recent Developments and Applications," *Found. Trends® Inf. Retr.*, vol. 8, no. 2-3, pp. 127–261 (2014). https://doi.org/10.1561/1500000042.

[57] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-Based Music Information Retrieval: Current Directions and Future Challenges," *Proc. IEEE*, vol. 96, no. 4, pp. 668–696 (2008).

[58] G. A. Wiggins, "Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music," in *Proceedings of the 2009 11th IEEE International Symposium on Multimedia*, pp. 477–482 (2009 Dec.).

[59] Y. Raimond, S. A. Abdallah, M. B. Sandler, and F. Giasson, "The Music Ontology," in *Proceedings of the 8th International Society for Music Information Retrieval Conference* (2007).

[60] G. Fazekas and M. B. Sandler, "Knowledge Representation Issues in Audio-Related Metadata Model Design," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), paper 8765.

[61] G. Fazekas, *Semantic Audio Analysis: Utilities and Applications*, Ph.D. thesis, Queen Mary University of London, London, UK (2012).

[62] T. Wilmering, G. Fazekas, and M. B. Sandler, "AUFX-O: Novel Methods for the Representation of Audio Processing Workflows," in *Proceedings of the 15th International Semantic Web Conference (ISWC)*, vol. 9982 of *Lecture Notes in Computer Science*, pp. 229–237 (2016). https://doi.org/10.1007/978-3-319-46547-0_24.

[63] M. -F. Plassard (Ed.), *Functional Requirements For Bibliographic Records: Final Report / IFLA Study Group on the Functional Requirements for Bibliographic Records*, vol. 19 (K.G. Saur, Munich, Germany, 1998).

[64] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. D. Reiss, "SAFE: A System for Extraction and Retrieval of Semantic Audio Descriptors," in *Proceedings of the 15th International Society for Music Information Retrieval Conference* (2014).

[65] R. Stables, B. De Man, S. Enderby, J. Reiss, T. Wilmering, and G. Fazekas, "Semantic Description of Timbral Transformations in Music Production," in *Proceedings of the 24th ACM International Conference on Multimedia*, pp. 337–341 (Amsterdam, the Netherlands) (2016 Oct.). https://doi.org/10.1145/2964284.2967238.

[66] S. McGrath, A. Chamberlain, and S. Benford, "The Grime Scene: Social Media, Music, Creation and Consumption," in *Proceedings of the Audio Mostly Conference*, pp. 245–250 (Norrköping, Sweden) (2016 Oct.).

[67] K. R. Scherer, A. Schorr, and T. Johnstone, *Appraisal Processes in Emotion: Theory, Methods, Research* (Oxford University Press, Oxford, UK, 2001).

[68] M. Barthet, G. Fazekas, and M. Sandler, "Music Emotion Recognition: From Content- to Context-Based Models," in *Proceedings of the International Symposium on Computer Music Modeling and Retrieval*, pp. 228–252 (London, UK) (2012 Jun.).

[69] P. Saari, G. Fazekas, T. Eerola, M. Barthet, O. Lartillot, and M. Sandler, "Genre-Adaptive Semantic Computing Enhances Audio-Based Music Mood Prediction," *IEEE Trans. Affect. Comput. (TAC)*, vol. 7, no. 2, pp. 122–135 (2016). https://doi.org/10.1109/TAFFC.2015.2462841.

[70] D. Sheng and G. Fazekas, "A Feature Learning Siamese Model for Intelligent Control of the Dynamic Range Compressor," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)* (2019).

[71] L. Feng, *Context-Aware Computing, Volume 3 of Advances in Computer Science* (De Gruyter, Berlin, Germany, 2018). https://doi.org/10.1515/9783110556674.

[72] U. Alegre, J. C. Augusto, and T. Clark, "Engineering Context-Aware Systems and Applications: A Survey," *J. Syst. Software*, vol. 117, pp. 55–83 (2016).

[73] A. K. Dey and G. D. Abowd, "The Context Toolkit: Aiding the Development of Context-

Aware Applications," in *Proceedings of the Workshop on Software Engineering for Wearable and Pervasive Computer*, pp. 431–441 (Limerick, Ireland) (2000 Jun.).

[74] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a Better Understanding of Context and Context-Awareness," in *Proceedings of the International Symposium on Handheld and Ubiquitous Computing*, pp. 304–307 (Karlsruhe, Germany) (1999 Sep.).

[75] P. Gray and D. Salber, "Modelling and Using Sensed Context Information in the Design of Interactive Applications," in *Proceedings of the IFIP International Conference on Engineering for Human-Computer Interaction*, pp. 317–335 (Toronto, Canada) (2001 May).

[76] G. Peeters, "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project," Tech. rep., IRCAM (2004).

[77] G. D. Abowd, E. D. Mynatt, and T. Rodden, "The Human Experience of Ubiquitous Computing," *IEEE Pervasive Comput.*, vol. 1, no. 1, pp. 48–57 (2002).

[78] D. Chalmers, *Sensing and Systems in Pervasive Computing: Engineering Context Aware Systems* (Springer, New York, NY, 2011).

## THE AUTHORS


M. Nyssim Lefford


Gary Bromham


György Fazekas


David Moffat

M. Nyssim Lefford is a cognitive scientist researching creative processes in record production contexts. She is an associate professor at Luleå University of Technology, where she teaches a wide range of audio, music, sound design, and interactive media related topics. She studied music production and engineering and film scoring at Berklee College of Music and from the Massachusetts Institute of Technology received her MS for work on network music collaboration and PhD for investigations into the perceptions of music creators as they work in situ. She is a member of the AES.

•

Gary Bromham is a part-time PhD researcher at Queen Mary University of London, researching the role that traditional studio paradigms and retro aesthetics play in intelligent music production systems (2016–). He has several publications in this field and has contributed a chapter to the recent Routledge publication 'Perspectives on Music Production: Mixing Music' (2017). He was also a research assistant on the EPSRC funded project called FAST (Fusing Audio and Semantic Technologies), where he is employed as an industry advisor (2017–2020). In addition to his research interests, Gary is a practicing music producer, songwriter, and audio engineer, with over 30 years' experience (1989–2020). He has worked with artists as diverse as Bjork, Wham, Blur, and U2, during a period that has witnessed several technological changes. Gary is well versed in most popular music making software and has extensive knowledge of using analog hardware, acting as a product designer and specialist for the renowned mixing desk company Solid State Logic. He is also a frequent guest lecturer and external advisor at several universities in the UK, Norway, and Sweden, speaking on songwriting, music production aesthetics, and audio engineering and bringing some of his extensive knowledge and experience to both undergraduate and Master's degree level programs.

•

George Fazekas is a Senior Lecturer (Associate Prof.) in Digital Media at the Centre for Digital Music, Queen Mary, University of London (QMUL). He holds a BSc, MSc, and PhD degree in Electrical Engineering. He is an investigator of UKRI's 6.5M Centre for Doctoral Training in Artificial Intelligence and Music (AIM CDT). He published over 140 academic papers in the fields of Music Information Retrieval, Semantic Web, Ontologies, Deep Learning, and Semantic Audio, including an award-winning paper on transfer learning. He was QMUL's Principal Investigator of the H2020 Audio Commons project (grant no. 688382, EUR 2.9M, 2016–2019), receiving best score from European Commission reviewers, and PI/Co-I of additional research projects worth over £230K. He was general chair of ACM's Audio Mostly 2017 and papers co-chair and organizing committee leader of the AES 53rd International Conference on Semantic Audio. He is regular reviewer for IEEE Transactions, Journal of the AES, Journal of New Music Research, and others. He is a member of IEEE, ACM, BCS, and AES and received the Citation Award of the AES for his work on the Semantic Audio Analysis Technical Committee.

•

Dr. David Moffat is a Lecturer in Sound and Music Computing at the University of Plymouth. He received his PhD and MSc from Queen Mary University. His research focuses on intelligent and assistive mixing and audio production tools through the implementation of semantic tools and machine learning. David has been a member of the Audio Engineering Society since 2014, is a member of the AES UK committee and Vice Chair of the Technical Committee on Semantic Audio Analysis.