Applications of Cross-Adaptive Audio Effects: Automatic Mixing, Live Performance and Everything in Between

Joshua D. Reiss[1]* and Øyvind Brandtsegg[2]

[1] *Centre for Digital Music, Queen Mary University of London, London, United Kingdom,* [2] *Department of Music, Norwegian University of Science and Technology, Trondheim, Norway*

This paper provides a systematic review of cross-adaptive audio effects and their applications. These effects extend the boundaries of traditional audio effects by potentially having many inputs and outputs, and deriving their behavior based on analysis of the signals. This mode of control allows the effects to adapt to different material, seemingly "being aware" of what they do to signals. By extension, cross-adaptive processes are designed to take into account features of, and relations between, several simultaneous signals. Thus a more global awareness and responsivity can be achieved in the processing system. When such a system is used in real-time for music performance, we observe cross-adaptive performative effects. When a musician uses the signals of other performers directly to inform the timbral character of her own instrument, it enables a radical expansion of the human-to-human interaction during music making. In order to give the signal interactions a sturdy frame of reference, we engage in a brief history of applications as well as a classification of effects types and clarifications in relation to earlier literature. With this background, the current paper defines the field, lays a formal framework, explores technical aspects and applications, and considers the future of this growing field.

Keywords: cross-adaptive, audio effects, music performance, music production, automatic mixing

1. INTRODUCTION

The technology that surrounds us in daily life is becoming increasingly "smart" and adapting to our habits and needs. This trend permeates all aspects of human endeavor, from the social to the medical, from the practical to the economic, from the local to the global and further on to space. Similarly we see this in creative aspects of human life, from the creative industries and on to the refined arts. The exploration of creativity in artificial intelligence has been gaining traction over the last decade. Our focus in this article is on recent advances in audio production techniques, where decisions on processing methods, or fine-tuning of their parameters, are done by an automated process. Some of these techniques fall under the heading of intelligent audio production tools. The term intelligent here refers in general to techniques that analyze the properties of the audio material to be processed and apply appropriate processing without human intervention. The basis for these automated decisions are often derived from best practices of professional audio engineers. The tools might be used to automatically correct for well known and frequently occurring problems in audio production. In other cases, the techniques are used for more creative explorations, allowing the automated processes to create tension and friction, as well as

new potential for musical expression in artistic, performative and perceptual dimensions. Analyzing the basic properties of a signal or a collection of audio material, extracting perceptual features, and creating new mappings, dependencies, correlations, dimensions of attraction and repulsion. The aim of this branch of techniques is to create new landscapes for human creativity to unfold. In the case of intelligent automation, it relieves the human operator from some repetitive and time consuming tasks, freeing resources for more creative parts of the production process. A similar but complementary effect can be seen in the applications on cross-adaptive performance, where habitual performative actions are challenged and sometimes interrupted. In all cases, the techniques raise questions that can engage in an active reappropriation of creativity in the age of artificial intelligence.

2. RECENT HISTORY OF APPLICATIONS

Even though we have seen an explosive growth in the utilization of adaptive processing and signal interaction in the last 15 years or so, we note a longer history of crossmodulation and signal interaction that has been active for the last 70 years or so. *Ring modulation* is the time domain multiplication of two signals. An early example of creative use can be seen in Stockhausen's "Mixtur" from 1964, while the guitar solo on Black Sabbath's "Paranoid" from 1970 serves as an example of usage in popular music. The *Vocoder* is based on a multi-band envelope follower on one sound, controlling levels of the same frequency bands on another signal. This has been in popular use, for example in Wendy Carlos' music for the film "A Clockwork Orange" from 1971. Vocoders were also used extensively by Laurie Anderson, for example in the 1981 " O Superman." Within electroacoustic art music, Philippe Manoury and Miller Puckette have explored adaptive techniques as a means to enable expressive performances where electronic material needs to be combined with a live acoustic performer. This can be heard in pieces like "Jupiter" (1987) and "Pluton" (1988). Another classic effect where signal analysis has been used to drive processing parameters is the *Auto-wah"*, where an envelope follower dynamically controls the cutoff frequency of a bandpass filter. A famous example is the trademark Clavinet sound of Stevie Wonder, for example on "Superstition" and "Higher Ground" (1972–73). More recently, we have seen adaptive processing permeating whole genres of popular music with the use of sidechain compression to create ducking and pumping effects. A clear example of this can be heard in Eric Prydz' "Call On Me" from 2004. Lately, we have seen adaptive control explorations by performer-instrument-designers like Stefano Fasciani and Hans Leeow. The cross-adaptive effects of Brandtsegg et al. (2018a) and Baalman et al. (2018) seems currently to be in the forefront of the field.

3. CLASSIFICATION OF AUDIO EFFECTS BY CONTROL

For understanding the architectures of complex or intelligent effects, we use another approach proposed in Verfaille et al.

(2006), which classifies digital audio effects in terms of the ways in which control parameters may be determined by input signals. At first, this may seem counterintuitive—why would the control parameters, as opposed to the perceptual properties or signal processing techniques, be an important aspect? However, certain forms of control are an essential enabling technology for intelligent audio effects.

Having control parameters depend on an audio signal is what defines an adaptive audio effect. Verfaille et al. (2006) further breaks down adaptive audio effects into several subcategories. Block diagrams of auto-adaptive, external-adaptive and cross-adaptive processing devices have been depicted in **Figure 1**, and these categories will be discussed below.

3.1. Non-adaptive

For non-adaptive audio effects, all parameters that influence the processing are either fixed or directly controlled by the user, as depicted in **Figure 2**. In general, features are not and do not need to be extracted from input signals. For instance, a graphic equalizer is non-adaptive since no aspect of the processing is dependent on any attributes of the input signal. The locations of the frequency bands are fixed and the gain for each band is controlled by the user. A multisource extension of this approach is to link the user interface, so that both channels of a stereo equalizer are controlled by just one set of controls. This provides exactly the same equalization for the left and right channel using a single user panel. Although the user interface is linked, the output signal processing is still independent of the signal content.

Figure 2 shows the standard implementation of a non-adaptive audio effect, where n is the discrete time index in samples, $x[n]$ is the input source and $y[n]$ is the output resulting from the signal processing.
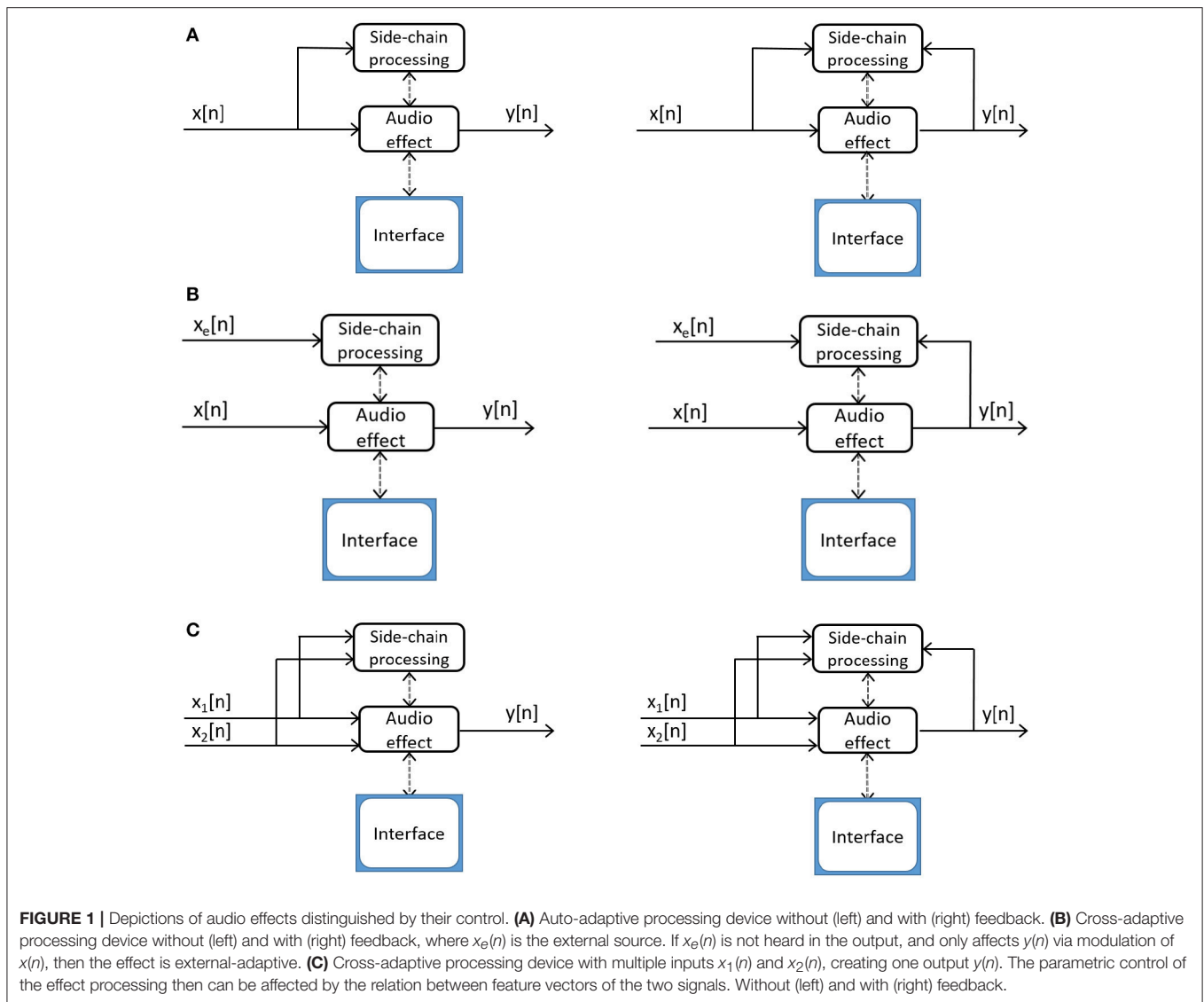
Adaptive digital audio effects, in contrast, use features extracted from the audio to control the signal processing. The dynamic range compressor is a good example of an adaptive effect. The gain applied by a compressor depends on the input signal level. Thus, the input signal needs to be analyzed and the level determined.

In some cases, the analysis of the input signal is done in a separate module. In other cases it is incorporated into the processing module, blurring the line between adaptive and non-adaptive. The classification is generally based on the manner of signal interaction, while the implementation details should not affect classification.

This distinction between non-adaptive and adaptive effects is not the same as between linear and nonlinear. Consider treating an audio effect as a black box, which is fed sine waves at different frequencies. For a linear, time-invariant effect, the same frequencies occur at the output, only with the amplitude and phase changed. One could imagine an effect that, regardless of input, always adds a 1 kHz tone at the output. Such a system would be non-adaptive, but it would also be nonlinear.

3.2. Auto-Adaptive

In an auto-adaptive audio effect, the control parameters are based on a feature extracted from the input source itself. Here "auto" means "self," and is not an abbreviation of "automatic," in contrast to auto-tuning or auto-mixing. In feedforward designs,

FIGURE 1 | Depictions of audio effects distinguished by their control. **(A)** Auto-adaptive processing device without (left) and with (right) feedback. **(B)** Cross-adaptive processing device without (left) and with (right) feedback, where $x_e(n)$ is the external source. If $x_e(n)$ is not heard in the output, and only affects $y(n)$ via modulation of $x(n)$, then the effect is external-adaptive. **(C)** Cross-adaptive processing device with multiple inputs $x_1(n)$ and $x_2(n)$, creating one output $y(n)$. The parametric control of the effect processing then can be affected by the relation between feature vectors of the two signals. Without (left) and with (right) feedback.

the analysis and feature extraction is performed only on the input signal. However, auto-adaptive audio effects can also use feedback from the output signal. An example is again the dynamic range compressor, where the gain applied to the signal is dependent on a measurement of either input or output signal level, for feedforward and feedback compressors respectively (Giannoulis et al., 2012).
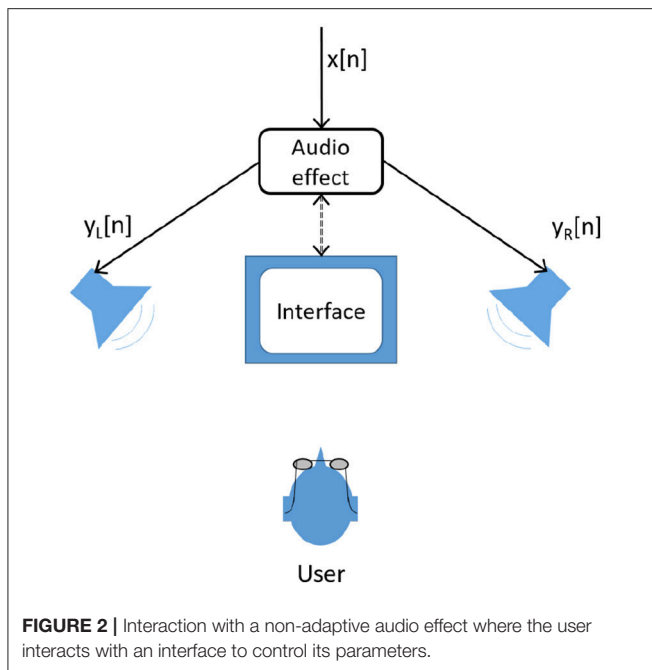
Some brief examples of auto-adaptive audio effects were given in Verfaille et al. (2006). such as noise gates, compressors and time-warping. Panning strategies were revisited in Morrell and Reiss (2009), which presented a "creative effect" type of implementation, where pan position could be time-varying and level or frequency dependent.

In 2000, James Moorer laid out a vision for audio production technology emerging over the next 20 years (Moorer, 2000). It described an Intelligent Assistant which would incorporate psychoacoustic models of loudness and audibility, to "take over the mundane aspects of music production, leaving the creative

side to the professionals, where it belongs." Details of this vision were further given in Dannenberg (2007), based on an auto-adaptive environment where tracks could be automatically tuned. It would maintain user-specified loudness relationships and autonomously time-align and pitch-correct performances, hence moving auto-adaptive effects into the realm of automatic editing.

3.3. External-Adaptive and Cross-Adaptive

In an external-adaptive audio effect, the system derives its control processing variables from analysis of a source other than the one to which it is applied. The usage of the terms *external adaptive* and *cross-adaptive* have been somewhat ambiguous in the literature. Even though external adaptive could be used as a generalization, we use the term exclusively for cases where the external signal itself is not heard, i.e., it is only active by means of modulation. The majority of cases of adaptive processing will then fall into the auto-adaptive and cross-adaptive categories.

being used, often taking into account the relative level of each microphone, or the relative level of each source.

A cross-adaptive effect is feed-forward if its control variable is derived only from the input signals, and it is feedback cross-adaptive if it also relies on analysis of the output signal. A feedback implementation may indirectly adapt to the input source since the output could depend on both input and an external source.

Verfaille et al. (2006) briefly described several auto-adaptive effects, but their only description of a cross-adaptive effect was dynamic time-warping. In Morrell and Reiss (2009) and Pestana and Reiss (2014a), auto-adaptive effects from Verfaille et al. (2006) were rephrased and extended as cross-adaptive applications.

In Brandtsegg (2015), Brandtsegg et al. (2018a), and Baalman et al. (2018), cross-adaptive methods were used for live performance, creating interconnected modulations between live instruments. Here, one can consider the performer part of the feedback and modulation network since a sensitive performer may adapt her playing to the sounding conditions of her instrument. Sarkar et al. (2017) considered a similar approach where cross-adaptive effects were used to allow performers to achieve synchronous effects that ordinarily would only be possible in the studio.

3.4. Cross-Adaptive Networks of Processing

Cross-adaptive processing networks are characterized by having control parameters which are determined by several input audio signals. They include cross-adaptive effects that analyze multiple input signals (and often the relationships between them) in order to produce at least one output track. Each node in such a network can conform to any of the previous adaptive topologies. It generalizes the single- or dual track adaptive processing approach and provides the greatest design flexibility. Cross-adaptive networks can be further enhanced using a feedback loop.

The cross-adaptive architecture includes many potential designs that can be used in full mixing systems. The following is a list of possible network nodes.

- Adaptive — features are extracted from at least one input signal. X_f and the feature values are used to drive processing on at least one input signal X_p, producing at least one output signal Y_p.
- Auto-adaptive — X_f and X_p are the same input signal.
- Cross-adaptive — X_f and X_p are not exactly the same signals.
- External-adaptive — Cross-adaptive, but X_f is truly external and is only perceived in Y_p by means of its modulation.
- Bilateral cross-adaptive — Two signals modulating each other, where both processed signals are heard in the output. (X_{p1} modulates X_{p2}, creating Y_{p2}) while also (X_{p2} modulates X_{p1} and creates Y_{p1}).
- Multi-input cross-adaptive — Cross-adaptive, but X_f is more than one signal.

FIGURE 2 | Interaction with a non-adaptive audio effect where the user interacts with an interface to control its parameters.

There is a perceptual aspect to the term, in that certain features of one sound are allowed to "cross over" to the other sound. For instance, cross-adaptive also relates to *cross synthesis* where characteristics of two sounds may be shared to create some merged output.

Consider ducking effects. Whenever the secondary audio input reaches a certain volume level, the audio source being processed "ducks" out of the way by reducing in volume. The technique can clearly be heard on the radio whenever a DJ begins to speak and the music automatically reduces in volume. The same effect has been used creatively in music production with sidechain pumping, where the bass drum envelope inversely control the level of e.g., synth pads. In some productions, the effect has been used in such a way that only the secondary sound is heard, e.g., the synth pads, rhythmically ducked are heard, but we can not hear the bass drum sound that activates the ducking. This would be a case where the term external adaptive is appropriate.

The Talk Box effect, popular since the 1970's, is an external-adaptive effect where the frequency content of the instrument's sound is filtered and modulated by the shape of the performer's mouth before amplification. The traditional Talk Box is interesting since it does not use analog or digital circuitry for the analysis and modulation, but rather the physical characteristics of how the mouth processes the human voice and the audio input signal. The Talk Box is often confused with the vocoder technique, which is an application of cross-synthesis where the spectral envelope of a signal modifies the spectral envelope of another signal.

One of the first examples of cross-adaptive audio effects is the automatic microphone mixer (Dugan, 1975). Automixers balance multiple sound sources based on each source's level. They reduce the strength of a microphone's audio signal when it is not

- Multi-output cross-adaptive — Cross-adaptive, but X_p (and hence Y_p) is more than one signal.
- Multi-input multi-output cross-adaptive — Cross-adaptive, but features are extracted from more than one input signal and more than one output signal.

4. INTELLIGENT AND ADAPTIVE DIGITAL AUDIO EFFECTS

Intelligent adaptive effects *listen* to the audio signal. They are imbued with an understanding of their intended use, and are often intended to control their own operation in much the same way as a sound engineer would control effects at a mixing desk. A knowledge engineering approach may be applied to gather and apply best practices in sound engineering, supplemented by listening tests to further establish preferences or machine learning from training data based on previous use. Thus, intelligent audio effects may provide parameter settings for dynamics processors (Ma et al., 2015), set appropriate equalization (Hafezi and Reiss, 2015), apply preferred reverberation (De Man et al., 2017a; Pestana et al., 2017), and implement stereo panning to more effectively distinguish the sources (Perez Gonzalez and Reiss, 2010).

Figure 3 gives a block diagram of an intelligent audio effect. Analysis is performed in a side-chain, which ensures that the audio signal flow is unaffected. This side-chain extracts and analyses features, resulting in controls which are used to modify the audio signal.

5. CROSS-ADAPTIVE PROCESSING

So far, the architecture for intelligent audio effects assumes that each effect processes a single audio track. At most, they may extract features from an additional input in the side-chain. But in multitrack mixing, the signal processing applied to a given source will depend on the properties of many other sources. So relationships between sources must be taken into account. This is relevant not only to produce a high quality mix by professional standards, but also for unconventional and creative sound design strategies. Any effects process that can be parameterized can be controlled in this manner. As an example, we could control the reverb time for the drums according to the level of rhythmic activity in the guitars, or control the flanger rate of the guitar in accordance with the noisiness of the vocals.

This can be conceptualized as an extension of the cross-adaptive audio effect, where the processing of an input source is the result of analysis of the features of each source and the relationships between all of those sources. Such an effect makes use of a multiple input multiple output (MIMO) architecture and may be considered inter-channel dependent.

This is depicted in **Figure 4** for one track within a multitrack, cross-adaptive audio effect. Assuming the cross-adaptive tool has the same number of inputs and outputs, inputs may be given as $x_m[n]$ and outputs as $y_m[n]$, where m has a valid range from 1 to M given that M is the maximum number of input tracks involved in the signal processing section of the tool.

A cross-adaptive multitrack audio effect is typically comprised of two main sections, the signal processing section and the side-chain processing section. The signal processing algorithm may be a standard audio effect and can include a user interface if the tool is meant to provide visual feedback or metering for its actions. The side-chain consists of a feature extraction block, and a cross-adaptive feature-processing block.

The feature extraction block will extract features from all input tracks. Consider feature vector $fv_m[n]$ obtained from the feature extraction section for source m. It may correspond to different features in each cross-adaptive audio effect. For example, for
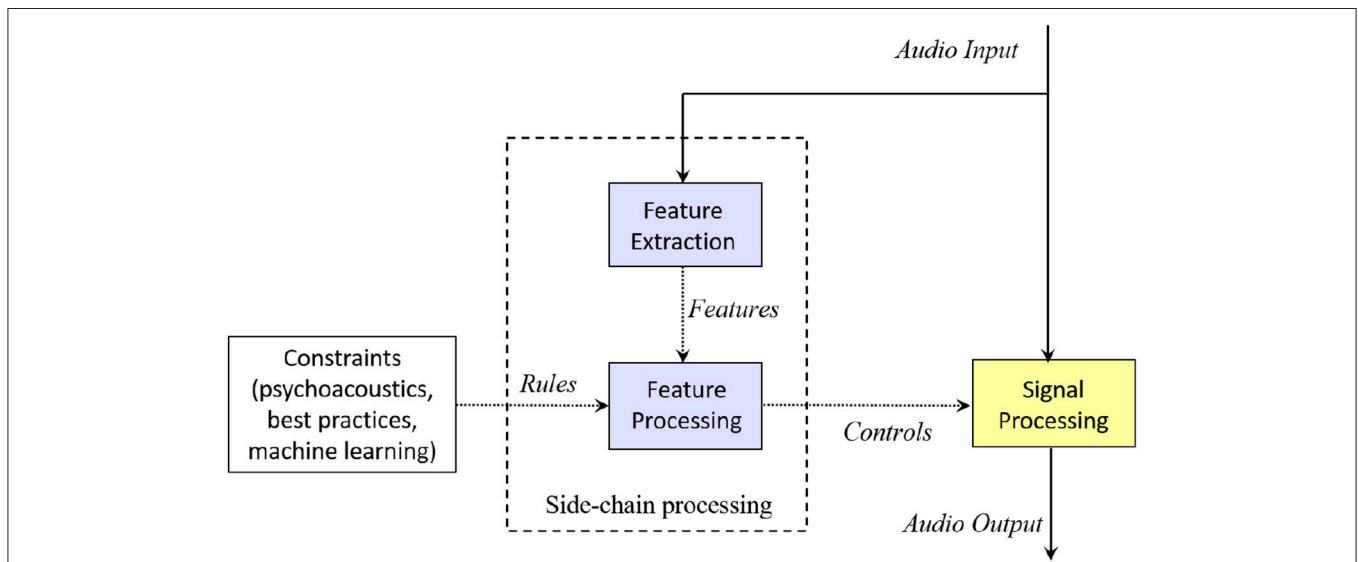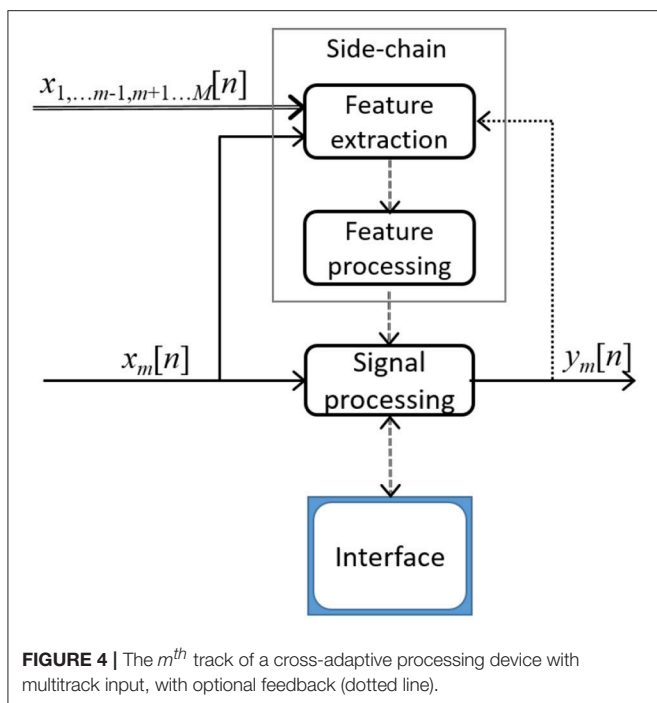

FIGURE 3 | An intelligent audio effect. Features of the input audio signal are extracted and analyzed to produce controls. The control values are often based on rules intended to mimic the behavior of a trained engineer.

FIGURE 4 | The mth track of a cross-adaptive processing device with multitrack input, with optional feedback (dotted line).

time alignment it may correspond to a time delay value but for spectral enhancement it may consist of a spectral decomposition classification feature.

These feature values are sent to a cross-adaptive feature processing section. Feature processing is achieved by mapping relationships between input features to control values applied in the signal processing section, as depicted in **Figure 5**. Typically, either the feature values or the obtained control values are smoothed before being sent to the next stage.

As with the feature vector, the control vector $c_m[n]$ will correspond to different parameters according to the aim of each multitrack audio effect. For example, suppose the objective is to achieve equal level for all channels, where $l_m[n]$ is the level per track, and the processing per track is given by $y_m[n] = c_m[n] \cdot x_m[n]$. A simple control could be given by $c_m[n] = \text{mean}(l[n])/l_m[n]$, which ensures equal level for all channels.

5.1. Cross-Adaptive Mixing in Practice

Some techniques used by mixing engineers may lend themselves to a cross-adaptive implementation. For instance, "corrective summing equalization" performs equalization to solve conflicts in between tracks, thus being cross-adaptive. This is in contrast to the auto-adaptive "Corrective track equalization", which corrects imperfections in an individual source or recording technique based on analysis of the audio signal (Bitzer and LeBoeuf, 2009; Wakefield and Dewey, 2015).

Dynamic spectral panning was proposed in Pestana and Reiss (2014a) as a means to reduce masking (an extension of auto-adaptive spectral panning Verfaille et al., 2006), and Pestana et al. (2015) suggested dynamic polarity switching in order to reduce the peak-to-RMS ratio, hence giving a higher overall loudness.

5.2. Cross-Adaptive Performance and Live Use

The activity of *cross-adaptive performance* is related to cross-adaptive processing, in that the latter is incorporated as part of the instrument used for performance. It also implies the action of human performers in the modulation loop. This means that perception and performative intention takes an integral role, and works together with the feature extraction and resulting parametric modulation. The "Cross adaptive processing as musical intervention"[1] project has explored the performative and musical implications of these techniques. A set of tools (Brandtsegg, 2015) have been developed to assist such explorations, in the form of open source VST plugins[2]. Since these can be used to control any parameter within a common DAW, their affordances can be incorporated into a more standardized workflow of music production. There are separate plugins for feature extraction and modulation mapping, and the mapping module allows control signal input from analysis of several sources. In this manner, analysis signals can be combined in several ways: mixing of control signals allows several sources to influence one control parameter; absolute differences between analysis signals allows the relationship between sources to be used; and conditional gates may be constructed so that the value of one analysis signal enables or disabled the mapping of another analysis signal.

A particular focus has been maintained on how this changes the affordances of the musical instrument, and how an ensemble approaches a shared instrumentality where each performer's action is mediated by whatever action other members of the ensemble may be executing simultaneously (Baalman et al., 2018). It can be described such that each musician no longer has her private instrument, but a collective instrument emerges out of their individual contributions and their interconnected mappings. With this domain follows a particular set of challenges to familiarize the performers with the expressive potential both of the electronic processing and the collective interaction. Different approaches from atomistic (bottom up) to holistic (top down) have been suggested, each with their own strengths and problems. More details on the working methods related to this kind of performance can be found in Brandtsegg et al. (2018a).

In addition to the analysis-based control signal structures outlined above, the cross-adaptive performance research has also used more direct forms of audio signal interaction. One such method is convolution, which has been used for creative sound design purposes since the mid 1980's (Dolson, 1985; Roads, 1993). To enable this technique to be used for live interaction between two audio signals, certain modifications to the implementation has been made (Brandtsegg et al., 2018b). With these modifications, the impulse response for the convolution can be recorded from a live audio signal and activated in the filter even before the recording has been completed. This allows two improvising musicians to play interactively with the convolution process together, where the sound from one is immediately convolved with the sound of the

---

[1]http://crossadaptive.hf.ntnu.no/
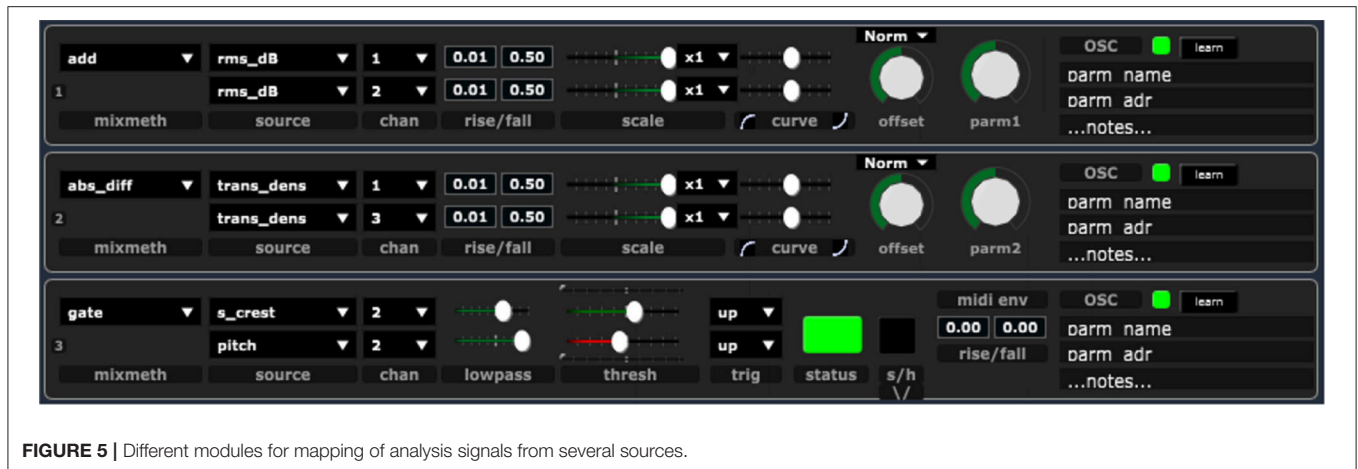[2]https://github.com/Oeyvind/featexmod

FIGURE 5 | Different modules for mapping of analysis signals from several sources.

other. This is also an example of how cross-adaptive performance extends the stricter technical notion of cross-adaptive processing as defined above.

6. MIXING SYSTEMS

Audio mixing is concerned with the editing, routing and combining of multiple audio tracks. It is a core challenge within the field of music production, though music production is much broader and includes instrument design, selection, microphone selection, placement and techniques, and overlaps other areas such as composition and arranging.

In Perez Gonzalez and Reiss (2011), Reiss (2011), and Reiss (2017), many cross-adaptive digital audio effects were described that aimed to replicate the mixing decisions of a skilled audio engineer with minimal or no human interaction. All of these effects produce mixes where the lth output channel is given by applying controls as shown in 1;

$$mix_l[n] = \sum_{m=1}^{M} \sum_{k=1}^{K} c_{k,m,l}[n] * x_m[n],\qquad(1)$$

where M is the number of input tracks and L is the number of channels in the output mix. K represents the length of control vector c and x is the multitrack input. Thus, the mixed signal at time n is the sum over all input channels of control vectors convolved with the input signal. The control vectors could also be expressed as transfer functions or difference equations, though this may be less compact.

Any cross-adaptive digital audio effect employing linear filters may be described in this manner. For automatic faders and source enhancement, the control vectors are simple scalars, and hence the convolution operation becomes multiplication. For polarity correction, a binary valued scalar is used (Perez Gonzalez and Reiss, 2008b). For automatic panning, two output channels are created, where panning is also determined with a scalar multiplication (typically, the sine-cosine panning law). For delay correction, the control vectors become a single delay operation. This applies even when different delay estimation

methods are used, or when there are multiple active sources (Clifford and Reiss, 2010, 2013). If multitrack convolutional reverb is applied, then c represents direct application of a finite room impulse response. Automatic equalization employs impulse responses for the control vectors based on transfer functions representing each equalization curve applied to each channel. And though dynamic range compression is a nonlinear effect, the application of feedforward compression is still a simple gain function. So multitrack dynamic range compression would be based on a time varying gain for each control vector.

In Uhle and Reiss (2010), a source separation technique was described where the control vectors are impulse responses that represent IIR unmixing filters for a convolutive mix. Thus, each of the resultant output signals mix_l in 1 represents a separated source, dependent on filtering of all input channels.

Terrell and Reiss (2009) attempted to deliver a live monitor mixing system that was as close as possible to a predefined target. It approximated the cause and effect between inputs to monitor loudspeakers and intelligibility of sources at performer locations. Stage sound was modeled as a linear multi-input multi-output system that simultaneously considered all performer requirements. The preferred mix was defined in terms of relative sound pressure levels (SPLs) for each source at each performer. So constrained optimization thus yielded scalar valued control vectors that resulted in optimal mixes for specified stage positions.

As described in Barchiesi and Reiss (2010), reverse engineering of an audio mix assumed 1 holds when presented with raw tracks and a final mix. It used least squares approximation to approximate control vectors as fixed length FIR filters. Justified assumptions then allow parameters for different effects to be derived. The gain in the filter represents fader values, initial zero coefficients represent delays, left and right output channel differences are based on sine-cosine panning, and anything remaining represents equalization. However, this would not be considered an intelligent audio effect since it aims to rederive settings, and it does not make use of best practices or preference.
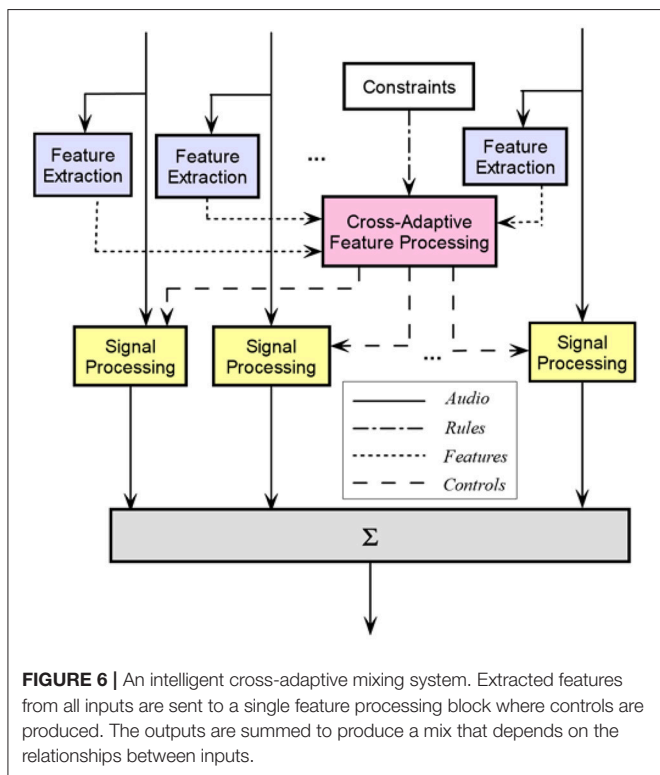
This approach also allows the mixing systems that incorporate several cross-adaptive effects to be implemented in a parallel or serial manner. The convolution operation may incorporate reverberation and equalization, for instance, or the result of applying cross-adaptive reverberation may then be passed through a second stage for cross-adaptive equalization before finally being summed to produce the output channels.

6.1. Automatic Mixing Systems

The design objective for automatic mixing systems is that the effects applied to a given source should depend on relationships between all involved sources. This can be met by the use of cross-adaptive processing. In this context, we can fully describe the architecture of an intelligent multitrack audio effect, shown in **Figure 6**. The side-chain consists of a feature extraction on each track and a single analysis section that processes the features extracted from all tracks. The cross-adaptive processing block outputs the control data based on consideration of relationships between input features. These controls provide the parameters in the signal processing of the multitrack content.

The side-chain processing section performs the analysis and decision making. It takes audio from one or more tracks together with optional external inputs, and outputs the derived control data. The control data drives the signal processing algorithm, thus establishing a cycle of feature extraction analysis and decision making. This aspect of the architecture is characteristic of adaptive effects, even when not intelligent or cross-adaptive.

The notion of adapting parameters in real-time can lead to results which go beyond the typical automation of mix

parameters (usually track level) and apply the classic processors in entirely new ways (e.g., dynamically carving out frequency content of a backing track in response to the lead vocal's features ...) (Pestana and Reiss, 2014b).

7. THE AUTOMATIC MIXING REVOLUTION

Automatic Mixing first referred to the application of microphone gain handling, and mainly for speech (Dugan, 1975). Between 2007 and 2010, Enrique Perez Gonzalez, an experienced sound engineer and music technology researcher, gave new meaning to the term by publishing methods to automatically adjust not just level (Perez Gonzalez and Reiss, 2009), but also stereo panning of multitrack audio (Perez Gonzalez and Reiss, 2007), equalization (Perez Gonzalez and Reiss, 2009), delay and polarity correction (Perez Gonzalez and Reiss, 2008b). He also automated complex mixing processes such as source enhancement within the mix (a multitrack form of mirror equalization) (Perez Gonzalez and Reiss, 2008) and acoustic feedback prevention (Perez Gonzalez
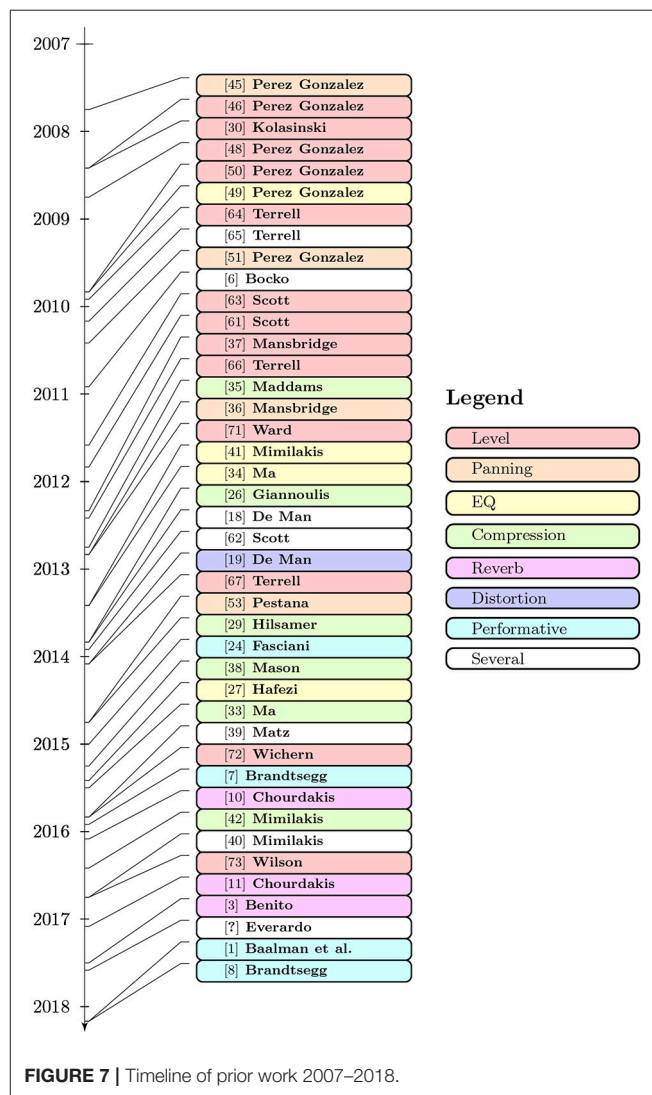
FIGURE 6 | An intelligent cross-adaptive mixing system. Extracted features from all inputs are sent to a single feature processing block where controls are produced. The outputs are summed to produce a mix that depends on the relationships between inputs.

FIGURE 7 | Timeline of prior work 2007–2018.

and Reiss, 2008a), and performed the first formal evaluation of an automatic mixing system (Perez Gonzalez and Reiss, 2010). To our knowledge, this was the inception of the field as it is known today.

This then ushered in a burst of sustained activity in the Automatic mixing field (De Man et al., 2017b). **Figure 7** shows a comprehensive but not exclusive overview of published systems or methods to automate production tasks in the last ten years. Some trends are immediately apparent. For instance, machine learning methods have been gaining popularity (Chourdakis and Reiss, 2016, 2017; Mimilakis et al., 2016a,b; Wilson and Fazenda, 2016a; Benito and Reiss, 2017). Whereas a majority of early automatic mixing systems were concerned with setting levels, recent years have also seen automation of increasingly "complex" processors such as dynamic range compressors (Maddams et al., 2012; Giannoulis et al., 2013; Hilsamer and Herzog, 2014; Ma et al., 2015; Mason et al., 2015; Mimilakis et al., 2016b) and reverb effects (Chourdakis and Reiss, 2016, 2017; Benito and Reiss, 2017). Research on such systems has also inspired several works on furthering understanding of the complex mix process and its perception (De Man et al., 2014a; Deruty and Tardieu, 2014; Pestana and Reiss, 2014b; Wilson and Fazenda, 2016b). Recently, we have also seen significant activity in exploring these techniques for live performance in a diversity of manners (Leeuw, 2009, 2012; Fasciani, 2014; Brandtsegg, 2015; Baalman et al., 2018; Brandtsegg et al., 2018a).

8. CONCLUSION

This paper gave a systematic review of the field of cross-adaptive audio effects. It defined the field and approaches, described the major advances and the broad range of applications, most notably in automatic mixing and as a new form of interaction in collaborative performance.

Beyond the scope of this paper, the cross-adaptive feature extraction architecture has also been used in the field of music information retrieval (Hargreaves et al., 2012), and new data sets have emerged for which cross-adaptive effects can be applied and evaluated (Bittner et al., 2014; De Man et al., 2014b). Such extensions of the field and its applications bode well for the future.

Cross-adaptive effects allow one to break free of the constraints of the traditional audio workstation plug-in architecture, and move beyond traditional approaches to performance interaction. Yet most existing work is still scratching the surface, often replicating existing approaches with new tools, e.g., intelligent systems that mimic the mixing decisions of sound engineers. The wide range of opportunities for new forms of interaction and new approaches to music production are yet to be explored.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Baalman, M., Emmerson, S., and Brandtsegg, Ø. (2018). "Instrumentality, perception and listening in crossadaptive performance," in *Proceedings of the 2018 Conference on Live Interfaces* (Porto).

Barchiesi, D., and Reiss, J. D. (2010). Reverse engineering the mix. *J. Audio. Eng. Soc.* 58, 563–576. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=15512

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

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 *ISMIR*, Vol 14, (Taipei), 155–160.

Bitzer, J., and LeBoeuf, J. (2009). "Automatic detection of salient frequencies," in *126th Audio Engineering Society Convention* (Munich).

Brandtsegg, Ø. (2015). "A toolkit for experimentation with signal interaction," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)* (Trondheim), 42–48.

Brandtsegg, Ø., Engum, T., and Wærstad, B. (2018a). "Working methods and instrument design for cross-adaptive sessions," in *Proceedings of the 2018 Conference on New Instruments for Musical Expression* (Blacksburg, VA).

Brandtsegg, Ø., Saue, S., and Lazzarini, V. (2018b). Live convolution with time-varying filters. *Appl. Sci.* 8:103. doi: 10.3390/app8010103

Chourdakis, E. T., and Reiss, J. D. (2016). "Automatic control of a digital reverberation effect using hybrid models," in *60th Audio Engineering Society Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)* (Leuven).

Chourdakis, E. T., and Reiss, J. D. (2017). A machine learning approach to application of intelligent artificial reverberation. *J. Audio Eng. Soc.* 65. doi: 10.17743/jaes.2016.0069

Clifford, A., and Reiss, J. D. (2010). "Calculating time delays of multiple active sources in live sound," in *129th Audio Engineering Society Convention* (San Francisco, CA).

Clifford, A., and Reiss, J. D. (2013). Using delay estimation to reduce comb filtering of arbitrary musical sources. *J. Audio Eng. Soc.* 61, 917–927. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=17072

Dannenberg, R. B. (2007). "An intelligent multi-track audio editor," in *Proceedings of the International Computer Music Conference* (Copenhagen).

De Man, B., Leonard, B., King, R., and Reiss, J. D. (2014a). "An analysis and evaluation of audio features for multitrack music mixtures," in *15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (Taipei).

De Man, B., McNally, K., and Reiss, J. D. (2017a). Perceptual evaluation and analysis of reverberation in multitrack music production. *J. Audio Eng. Soc.* 65, 108–116. doi: 10.17743/jaes.2016.0062

De Man, B., Mora-Mcginity, M., Fazekas, G., and Reiss, J. D. (2014b). "The open multitrack testbed," in *Audio Engineering Society Convention 137* (Los Angeles, CA).

De Man, B., Reiss, J. D., and Stables, R. (2017b). "Ten years of automatic mixing," in *3rd Workshop on Intelligent Music Production* (Salford, UK).

Deruty, E., and Tardieu, D. (2014). About dynamic processing in mainstream music. *J. Audio Eng. Soc.* 62, 42–55. doi: 10.17743/jaes.2014.0001

Dolson, M. (1985). "Recent advances in musique concrète at CARL," in *International Computer Music Conference (ICMC)* (Vancouver, BC).

Dugan, D. (1975). Automatic microphone mixing. *J. Audio Eng. Soc.* 23, 442–449.

Fasciani, S. (2014). *Voice-controlled Interface for Digital Musical Instruments.* PhD thesis, Singapore: National University of Singapore.

Giannoulis, D., Massberg, M., and Reiss, J. D. (2012). Digital dynamic range compressor design—a tutorial and analysis. *J. Audio Eng. Soc.* 60, 399–408. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=16354

Giannoulis, D., Massberg, M., and Reiss, J. D. (2013). Parameter automation in a dynamic range compressor. *J. Audio Eng. Soc.* 61, 716–726. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=16965

Hafezi, S., and Reiss, J. D. (2015). Autonomous multitrack equalization based on masking reduction. *J. Audio Eng. Soc.* 63, 312–323. doi: 10.17743/jaes.2015.0021

Hargreaves, S., Klapuri, A., and Sandler, M. (2012). Structural segmentation of multitrack audio. *IEEE Trans. Audio Speech Lang. Process.* 20, 2637–2647. doi: 10.1109/TASL.2012.2209419

Hilsamer, M., and Herzog, S. (2014). "A statistical approach to automated offline dynamic processing in the audio mastering process," in *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx-14)* (Erlangen).

Leeuw, H. (2009). "The electrumpet , a hybrid electro-acoustic instrument," in *NIME* (Pittsburgh, PA).

Leeuw, H. (2012). "The electrumpet, additions and revisions," in *NIME* (Ann Arbor, MI).

Ma, Z., De Man, B., Pestana, P. D., Black, D. A. A., and Reiss, J. D. (2015). Intelligent multitrack dynamic range compression. *J. Audio Eng. Soc.* 63, 412–426. doi: 10.17743/jaes.2015.0053

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)* (York, UK).

Mason, A., Jillings, N., Ma, Z., Reiss, J. D., and Melchior, F. (2015). "Adaptive audio reproduction using personalized compression," in *57th Audio Engineering Society Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet* (Hollywood, CA).

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* (London, UK).

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* (Paris).

Moorer, J. A. (2000). Audio in the new millennium. *J. Audio Eng. Soc.* 48, 490–498. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=12062

Morrell, M., and Reiss, J. D. (2009). "Dynamic panner: An adaptive digital audio effect for spatial audio," in *Audio Engineering Society Convention 127* (New York, NY).

Perez Gonzalez, E. and Reiss, J. D. (2007). "Automatic mixing: live downmixing stereo panner," in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx'07)* (Bordeaux), 63–68.

Perez Gonzalez, E., and Reiss, J. D. (2008a). "An automatic maximum gain normalization technique with applications to audio mixing," in *Audio Engineering Society Convention 124* (Amsterdam).

Perez Gonzalez, E., and Reiss, J. D. (2008b). "Determination and correction of individual channel time offsets for signals involved in an audio mixture," in *Audio Engineering Society Convention 125* (San Francisco, CA).

Perez Gonzalez, E., and Reiss, J. D. (2008). "Improved control for selective minimization of masking using inter-channel dependancy effects," in *11th Int. Conference on Digital Audio Effects (DAFx)* (York, UK), 12.

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

Perez Gonzalez, E., and Reiss, J. D. (2009). "Automatic gain and fader control for live mixing," in *2009. WASPAA'09. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY: IEEE), 1–4.

Perez Gonzalez, E., and Reiss, J. D. (2010). A real-time semiautonomous audio panning system for music mixing. *EURASIP J. Adv. Signal Process.* 2010:436895. doi: 10.1155/2010/436895

Perez Gonzalez, E., and Reiss, J. D. (2011). "chapter 13: Automatic mixing," in *DAFX: Digital Audio Effects, 2nd Edn.*, ed U. Zölzer (Chichester, UK: Wiley Online Library), 523–549

Pestana, P. D., and Reiss, J. D. (2014a). "A cross-adaptive dynamic spectral panning technique," in *DAFx* (Erlangen), 303–307.

Pestana, P. D., and Reiss, J. D. (2014b). "Intelligent audio production strategies informed by best practices," in *Proceedings of the Audio Engineering Society 53rd International Conference on Semantic Audio* (London).

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* (Warsaw).

Pestana, P. D., Reiss, J. D., and Barbosa, Á. (2017). User preference on artificial reverberation and delay time parameters. *J. Audio Eng. Soc.* 65, 100–107. doi: 10.17743/jaes.2016.0061

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

Reiss, J. D. (2017). "chapter 15: An intelligent systems approach to mixing multitrack music," in *Mixing Music*, eds R. Hepworth-Sawyer, and J. Hodgson (New York, NY: Routledge), 201–213.

Roads, C. (1993). "Musical sound transformation by convolution," in *Opening a New Horizon: Proceedings of the 1993 International Computer Music Conference, ICMC* (Tokyo).

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

Terrell, M. J., and Reiss, J. D. (2009). Automatic monitor mixing for live musical performance. *J. Audio Eng. Soc.* 57, 927–936. Available online at: http://www.aes.org/e-lib/browse.cfm?elib=15229

Uhle, C., and Reiss, J. D. (2010). "Determined source separation for microphone recordings using iir filters," in *129th Convention of the Audio Engineering Society* (San Francisco, CA).

Verfaille, V., Zolzer, U., and Arfib, D. (2006). Adaptive digital audio effects (a-dafx): A new class of sound transformations. *IEEE Trans. Audio Speech Lang. Proces.* 14, 1817–1831. doi: 10.1109/TSA.2005.858531

Wakefield, J., and Dewey, C. (2015). "Evaluation of an algorithm for the automatic detection of salient frequencies in individual tracks of multitrack musical recordings," in *Audio Engineering Society Convention 138* (Warsaw).

Wilson, A., and Fazenda, B. (2016a). "An evolutionary computation approach to intelligent music production, informed by experimentally gathered domain knowledge," in *2nd AES Workshop on Intelligent Music Production* (London, UK).

Wilson, A., and Fazenda, B. (2016b). Variation in multitrack mixes: analysis of low-level audio signal features. *J. Audio Eng. Soc.* 64, 466–473. doi: 10.17743/jaes.2016.0029