



## **Journal on the Art of Record Production**

ISSN: 1754-9892

Issue 8, Dec. 2013

### A Semantic Approach To Autonomous Mixing

Brecht De Man and Joshua D. Reiss

# A Semantic Approach To Autonomous Mixing

## Abstract

In this work, we give an overview of current automatic mixing approaches, and note the absence of using high-level, semantic knowledge to inform mixing decisions. We then propose a system where processing parameters for music production are inferred from instrument-specific rules rather than low-level features. To this end, we compile a body of rules from practical mixing engineering literature, and assess the potential of a system using just these rules. We compare the system to human mixing engineers and traditional automatic mixing systems that only use extracted, low-level signal features. Its performance is discussed and future directions are established.

## 1 Introduction

*“There’s no reason why a band recording using reasonably conventional instrumentation shouldn’t be EQ’d and balanced automatically by advanced DAW software.”*

— Paul White, Editor In Chief of Sound On Sound magazine

There is a clear need for systems that take care of the mixing stage of music production for live and recording situations. The democratisation of music technology has allowed musicians to produce music on limited budgets, where decent results are within reach for everyone with access to a laptop, a microphone under \$200, and access to the abundance of free software to be found on the internet (Tomaz De Carvalho: 2012). Similarly, at the distribution side, musicians can distribute their own content at very little cost and effort, also due to high availability of cheap technology

(compact discs, the internet) and, more recently, the emergence of online platforms like Soundcloud, Myspace, Youtube, and many more.

However, in order to deliver high quality material a skilled mixing engineer is still needed. Raw, recorded tracks almost always require a fair amount of processing (balancing, panning, equalising, compression, artificial reverberation and more) before being ready for distribution. Furthermore, despite the availability of reasonably high quality recording gear on a budget, an amateur musician or recording engineer will almost inevitably cause sonic problems while recording, due to less than perfect microphone placement, an unsuitable recording environment, or simply a poor performance or instrument. These artefacts are a challenge to fix post-recording, which only increases the need for an expert mixer (Toulson: 2008). In live situations, especially small venues, the mixing task is especially demanding and essential due to problems such as feedback, imbalance, room resonances and poor equipment. Unfortunately, having an experienced operator at the desk is the exception rather than rule.

Mixing multichannel audio comprises many expert but non-artistic tasks that, once accurately described, can be implemented in software or hardware (Reiss: 2011). By obtaining a high quality mix fast and autonomously, studio or home recording becomes more affordable for musicians, smaller music venues are freed of the need for expert operators for their front of house and monitor systems, and both audio engineers and musicians can increase their productivity and focus on the creative aspects of music production. In this way, one could argue the idea of automatic mixing is to music production what the autofocus function of a camera is to

photography (White: 2008). In both cases, the creativity is not left out as the user could have access to various degrees of control, from a completely autonomous system for laymen to an assisted but completely adjustable system for professional audio engineers.

Current automatic mixing systems already show adequate performance using basic extracted audio features (Perez-Gonzalez and Reiss: 2007; Perez-Gonzalez and Reiss: 2009a; Perez-Gonzalez and Reiss: 2009b; Perez-Gonzalez and Reiss: 2010; Reiss: 2011; Mansbridge et al.: 2012a; Mansbridge et al.: 2012b) or machine learning techniques (Scott and Kim: 2011; Scott et al.: 2011), and sometimes outperform amateur mixing engineers. Additionally, they are able to consistently outperform even professional mixing engineers in a fraction of the time when it comes to certain corrective tasks. These tasks seek to solve artefacts that are often the result of poor recording practice, such as compensating for time-varying delays (Clifford and Reiss: 2010), interference (Clifford and Reiss: 2011a), popping artefacts (Clifford and Reiss: 2011b), and comb filtering (Clifford and Reiss: 2011c).

However, few intelligent systems seem to take semantic, high-level information into account. The applied processing is dependent on low-level signal features, but no information is given (or extracted) about the instruments, recording conditions, playback conditions, genre, or target sound, to name a few. This information, which can be provided by an amateur end user at little cost, could significantly increase the performance of such semi-autonomous mixing system. Moreover, using feature extraction for instrument and even genre recognition, a fully autonomous system could be designed.

Having access to this type high-level knowledge, and choosing high-level parameters such as targeted genre or sound, shifts the potential of automatic mixing systems from corrective tools that help obtain a single, allegedly ideal mix, to providing the end user with countless possibilities and intuitive parameters to achieve them. For instance, an inexperienced user could then produce a mix that evokes a ‘classic rock’ sound, an ‘Andy Wallace’ sound, or a ‘1960’ sound. To maintain the photography analogy: in addition to an autofocus, end users would have a sonic equivalent of Instagram at their fingertips.

In order to get there, we need to identify the characteristics of the mixes we want to obtain, as well as the corresponding settings for different instruments. Many sources, among which numerous audio engineering books and websites, report standard settings for the signal processing of the various instruments a production consists of (White: 2000b; Katz: 2002; Gibson: 2005; Owsinski: 2006; Coryat: 2008; Izhaki: 2008; Case: 2011; Senior: 2012; Case: 2012). These settings depend on the engineers style and taste, the presence or absence of certain other instruments, and to some extent the characteristics of the signal. This involves preferential values for relative level, panning, equalising, dynamic range compression, and time-based effects such as reverb and delay.

The very books that provide this information, stress that mixing is a highly non-linear (Case: 2011), unpredictable (Senior: 2012) business, devoid of ‘hard and fast rules’ (Case: 2011), ‘magic settings’ (Case: 2012) or one-size-fits-all equaliser presets (Senior: 2012). This work does not seek to contradict this; rather, we investigate how

this knowledge can be used to infer or (when used in combination with low-level approaches) enhance autonomous mixing methods. As no two recorded tracks sound the same (even for identical instruments), it can be expected that spectral and dynamic features will have to be taken into account to consistently achieve good mixes. However, as a starting point it is valuable to test how effective a system relying on a body of practical knowledge is without accounting for low-level signal features.

In this work we first provide an overview of the state of the art, and identify the shortcomings of various approaches and schools of automatic mixing. We then propose a largely unexplored direction of automatic mixing research, i.e. using high-level, semantic information to infer mixing decisions. To this end, we assess the potential of ‘best practices’ derived from a broad selection of audio engineering literature and expert interviews, to constitute a set of rules that define to the greatest possible extent the actions and choices audio engineers make, given a song with certain characteristics. Rule-based processing is then applied to reference material (raw tracks) to validate the semantic approach. A formal comparison with state-of-the-art automatic mixing systems as well as human mixes as well as an unprocessed version is conducted, and future directions are identified.

## 2 Approaches in automatic mixing

### 2.1 Machine learning

In machine learning approaches to automatic mixing, the system can be trained on initial content (i.e. example mixes) to infer how to manipulate new content. An example of this is described in (Scott et al.: 2011), where a machine learning system was provided with both stems and final mixes for 48 songs. This was used as training

data, and it was then able to apply time varying gains on each track for new content. However, this approach is limited due to the rarity of available multitracks, and the related copyright issues. This is in stark contrast to music informatics research, usually concerned with the analysis of mixed stereo or monaural recordings, where there is a wealth of available content.

## 2.2 Grounded theory

Grounded theory and its methodology may be used to acquire basic knowledge which may subsequently be transferred to the intelligent system. For audio production, this suggests psychoacoustic studies to define mix attributes, and perceptual audio evaluation to determine listener preference for mix approaches. An important downside of this approach is that it is very resource intensive. Though there has been some initial work in this area (Perez-Gonzalez and Reiss: 2010), it is too limited to constitute a sufficient knowledge base for the implementation of an overall system.

## 2.3 Knowledge engineering

A traditional approach to intelligent systems design would exploit knowledge engineering. In contrast with grounded theory, knowledge engineering assumes the rules are already known and they should just be integrated in a system. This involves integrating established knowledge into the rules and constraints under which the system operates. However, best practices in record production are generally not known. That is, although there are numerous resources describing a particular record producers approach, there are few systemic studies that have defined standard practice (Pestana: 2013). Even established best practices may still not be well-defined (e.g., the mix should not sound ‘muddy’). Thus, the information, knowledge and

preferences are not easily acquired, and cannot be effectively transferred to the intelligent system.

In this paper, we follow a knowledge-engineered approach, acquiring a body of knowledge from practical literature. We do acknowledge the limitations of the knowledge engineered approach, especially as the knowledge we collect is often ill-defined (no exact figures provided), and scarce (no rules available for some instrument, for a certain processor). However, it serves as a good starting point for this experiment, where we attempt to explore the potential of high-level rules as opposed to the generally used extracted, low-level signal features. It can be appreciated that a higher number of high-level rules can be extracted, and validated, using machine learning or systemic studies.

### 3 Practical literature survey

We explore practical audio engineering literature for guidelines, rules of thumb and example settings, as a knowledge base for an automatic mixing system that infers its mixing decisions from a combination of user-inputted or extracted high-level semantic information (such as the instrument labels, the song's genre, the targeted mixing style) and the knowledge from such a rule base. Most of the information on settings of different processors found in practical literature occurs in the form of example parameters for a certain instrument. The standard processors in music production are dynamic range compression, equalisation and filtering, panning, fading, and time-based effects such as reverberation and echoes/delay. Other, more creative effects and processors can be used, but they are of less interest to automatic mixing research. It could be argued that they are entirely dependent on the artistic freedom of the

producer, artist and engineer, and not functional, subject to rules, or essential to the sonic validity of the mix.

### 3.1 Dynamic range compression

A simple version of a dynamic range compressor (usually referred to as ‘compressor’, but not to be confused with data compression) has parameters threshold, ratio, attack time, release time, and knee width. In the simplest of terms, the compressor applies a negative gain to the signal whenever it exceeds a threshold. This negative gain is proportional to by how much the signal level exceeds the threshold, with a factor determined by the ratio. The attack and release time control how fast the compressor decreases and increases its gain. The ‘knee’ is a smooth transition from the uncompressed to the compressed region. Quite often make-up gain is included (sometimes automatic) to compensate for the difference in loudness after a negative gain is applied to parts of the signal. However, we do not consider this in this work.

In practical literature, a significant number of compressor settings for different instruments and genres can be found (White: 1996; Robjohns: 1999; Rudolph: 1999; White: 2000a; Owsinski: 2006; Haas: 2008; Izhaki: 2008; Senior: 2009; Case: 2011; Senior: 2012). It should be noted that sometimes not all parameters are explicitly specified. An example could feature a range wherein a certain parameter should be chosen, or no value could be mentioned at all. In this case, we choose to default to a sensible standard value for this parameter, such as the ones featured in (White: 1996). Furthermore, testing a few different compressor models, we have found it should be taken into account that the definitions of attack and release time may vary. This

suggests that values found in text books and interviews with audio professionals may refer to compressor models with different definitions, too.

To fill the gaps, the recommended settings found in literature are complemented with the built-in compressor settings of the Platinum Compressor, shipped with Logic Pro 9.

### 3.2 Equalisation and filtering

A second essential compressor in audio engineering is the filter (of which equalisers technically are a subset). A static filter, meaning its parameters do not vary during playback, emphasises and attenuates certain parts of the frequency spectrum of the incoming signal. In this work, we use peaking filters (attenuating or emphasising a frequency region around a centre frequency, with a specified width), shelving filters (attenuating or emphasising all frequencies above a certain frequency by a fixed number of decibels) and high-pass filters (blocking everything below a certain cutoff point, as much as physically possible given the filter's complexity).

Again, practical audio literature features many examples of equaliser settings for different instruments and different effects (Gibson: 2005; Owsinski: 2006; Izhaki: 2008; Case: 2011; Senior: 2012). In this context, too, more often than not some parameters are not specified, which leaves room for interpretation. An example is the degree of attenuation or emphasis, which is usually not explicitly stated in decibel but rather in vague terms such as 'a little'. In some cases frequency ranges are not described in exact terms, but with semantic terms such as 'boom', 'air', and 'thump' (Gibson: 2005; Owsinski: 2006; White: 2006; Coryat: 2008; Izhaki: 2008). In this

case, however, the frequency parameter can be derived as most of the aforementioned sources do include diagrams or lists where these words are correlated with frequency ranges. For example, ‘air’ corresponds with 10-20 kHz (Owsinski: 2006; Izhaki: 2008; Coryat: 2008), 7-20 kHz (White: 2006) or 5-8 kHz (Gibson: 2005). Note that the definitions can differ substantially between sources. Another approach to defining these terms is to use crowdsourcing to learn which equaliser curve corresponds to a certain term (Cartwright and Pardo: 2013).

We also included equaliser settings from Logic Pro 9’s built-in channel equaliser for various instruments.

### 3.3 Panning

The panning value associated with each instrument track determines which fraction goes to the left and which to the right channel (assuming a stereo playback system). In case these fractions are equal, the source is said to be panned exactly ‘in the centre’. When panned all the way to the left (right), the source will only be present in the left (right) channel, speaker or earphone.

There is a fair amount of example positions for different instruments to be found in practical literature. Some sources list example panning positions for a hypothetical mix (Izhaki: 2008, p. 198), others give more general rules of thumb regarding panning low-frequency sources in the centre (White: 1998; White: 2000c; Izhaki: 2008; Case: 2011; Senior: 2012) and avoiding instruments to have the exact same position (Owsinski: 2006, p. 22).

### 3.4 Balance

Although balancing (changing the relative levels of the different instruments) is arguably the most obvious part of the mixing process, it proves very hard to find example settings, common practices or even vague rules of thumb describing the audio professional's approaches towards balancing in practical literature. An important reason for this is that the ideal levels very much depend on the loudness of the sources. Furthermore, they seem to depend heavily on the context, e.g. lead or background, presence of similar tracks, sparsity of the arrangement, intended target sound, ...

In this work, we first set the track levels so that the loudness of every track is the same, which is a bold assumption (Pestana: 2013). From this point onwards, we consider rules found in literature about which tracks should be louder (lead vocals, solo instruments) and which should be softer (background instruments, multiple instruments or microphones of the same functional section). In this case, too, few example settings include exact values in decibels, so a default level boost or drop has to be set (at e.g. 3 dB), with possibly a higher or lower value for those rules of thumb where either 'a little' or 'a lot' of boost (drop) is suggested.

### 3.5 Time-based effects

As mentioned, we are not investigating time-based effects such as echoes and reverberation, among many others. However, it is worth noting the scarcity of reverberation rules throughout literature. While some sources report recommend certain types of reverb for different sources (e.g. plate reverb for vocals (Izhaki: 2008, p. 413), spring reverb for electric guitar (Izhaki: 2008, p. 412)), very little information

is available on preferred settings of other reverb parameters such as decay time, wet/dry ratio, or early reflections vs. late reverberation ratio. At best, vague rules of thumb such as an inverse relationship between the tempo of a song and the reverb's decay time provide some help (Izhaki: 2008, p. 429).

## 4 Experiment

### 4.1 Setup

To investigate whether semantic information can be used successfully to infer mixing decisions, five songs are mixed using a body of rules derived from the aforementioned practical literature. No low-level features are extracted, except the track's peak and RMS levels (to be able to set the compressor's threshold to a meaningful value), as opposed to other autonomous systems. For every song, a number of other mixes are provided for comparison: two human mixing engineers' mixes, a fully autonomous mix solely based on low-level extracted features (Perez-Gonzalez and Reiss: 2007; Perez-Gonzalez and Reiss: 2009a; Perez-Gonzalez and Reiss: 2009b; Giannoulis et al.: 2012; Mansbridge et al.: 2012a; Mansbridge et al.: 2012b), and a monophonic sum of all (peak-normalised) sources without any processing whatsoever. The latter was meant as an anchor as it is to be expected this would result in a poor mix, although this could by chance result in a decent mix that some will prefer over a poor human or automatic mix.

The multitrack audio used for this experiment are high quality raw audio files that were publicly available through a website by Shaking Through (<http://www.shakingthrough.com>). They are unprocessed, except for occasional compressing or equalising applied during tracking. To avoid straining the participants' attention, and

because the parameters applied by the semantic mixing system are static (i.e. they do not vary with changing features or between song segments), we selected a fragment four bars out of each song. The resulting files are between 11 and 24 seconds long.

Both the human engineers and the automatic systems only use dynamic range compression, equalisation, (high pass) filtering, fading and panning. No editing or automation is allowed (all parameters static), although the parameters in the automatic system based on low-level feature extraction are continuously varying by design as they are implemented as real-time VST plug-ins.

The five different mixes of the five songs have been assessed through a formal listening test with fifteen participants. Following common practices, the order of the songs and the order the stimuli (different mixes) are presented per song is randomised. Per song, all stimuli are presented at once (multiple stimuli rather than pairwise evaluation) as it has been found in previous research that this yields results of similar accuracy while requiring less effort and time from participants, for this type of perceptual audio evaluation (De Man and Reiss: 2013). The interface used in this experiment is shown in Figure 1.

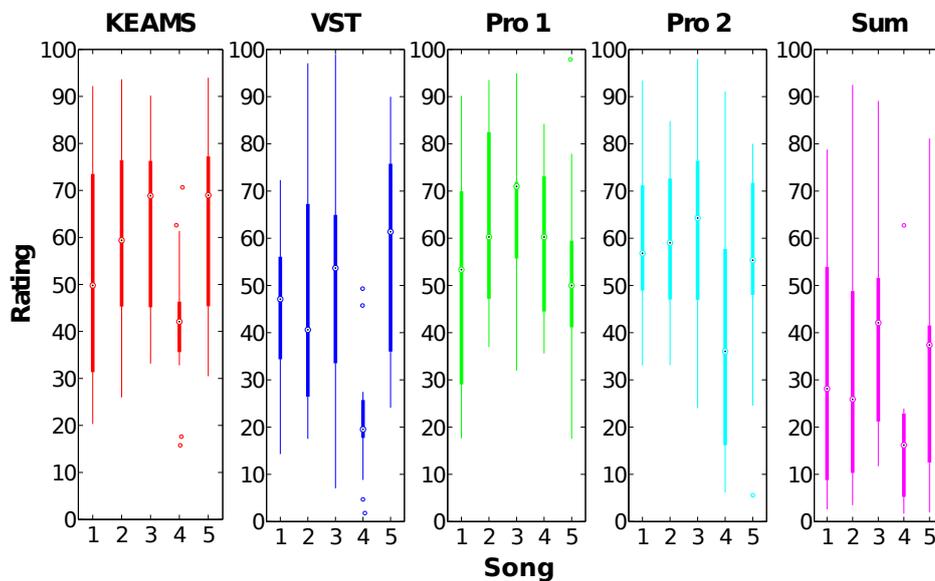


**Figure 1** Interface used during the listening experiment. Every marker corresponds to a different mix of the same song fragment. The highlighted marker represents the fragment that is currently playing. To avoid a bias on the first seconds of the fragments, the audio skips to the same position in a different fragment when another marker is clicked.

## 4.2 Results

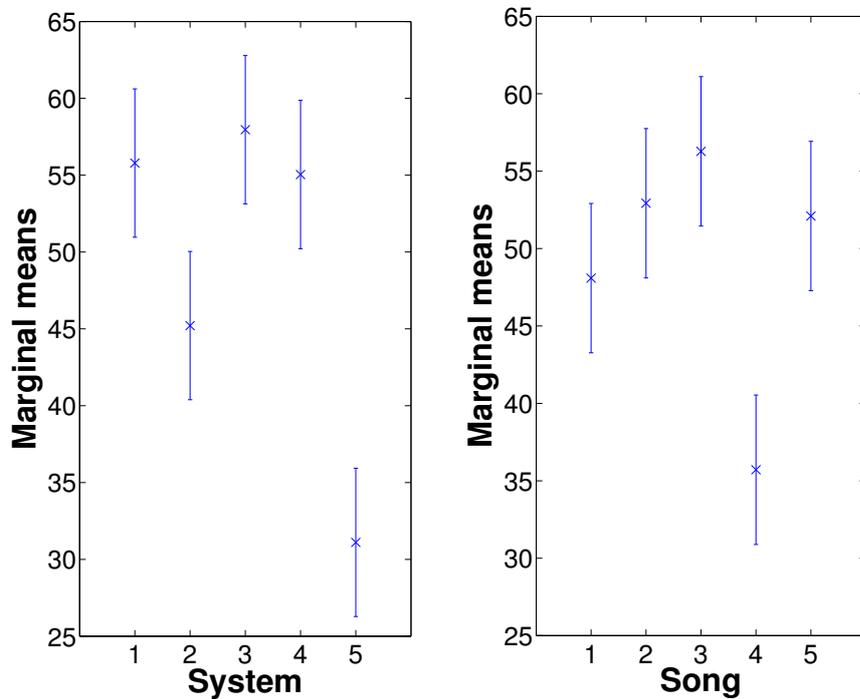
The results of this test are displayed in box-plot fashion in Figure 2. From these results, it is apparent that, as expected, the unprocessed sum of all files performs notably worse than the other mixes (although in some cases, a participant happens to like the raw, unprocessed mix more). Furthermore, there is a tendency to rate the fourth song poorly (a rather heavy alternative rock song, which the most experienced mixing engineer involved called ‘untreatable’, referring to the way it was recorded). To distinguish song effect from system effect, we perform an analysis of variance (ANOVA). The system effect size is  $R^2_{\text{system}} = 0.17$  (large effect (Cohen:

1988)), the song effect size is  $R^2_{\text{song}} = 0.09$  (medium effect). A multiple comparison of the population marginal means (a Bonferroni test with tolerance of 0.05) shows us the pairwise rather than familywise error rate (see Figure 3). This figure confirms a significantly worse rating of the plain sum, as well as of the fourth song. Furthermore, it clearly shows the automatic mix based on low-level features (where no high-level knowledge is supplied to the system) is rated lower than both human mixes and the automatic mix based on semantic information. The experiment reveals no significant difference between the human and semantic mixes.



**Figure 2** Box plot representation of the ratings per song and per system. System 1 only uses semantic knowledge (instrument labels) of the tracks; system 2 takes no user input and only extracts low-level features of the input signals; system 3 represents a mixing engineer with over 12 years of professional experience; system 4 represents a mixing engineer with 3 years of professional experience; and system 5 represents the monophonic sum of all tracks. Following the classic definition of a box plot, the dot represents the mean, the bottom and top of the ‘box’ represent the 25% and 75% percentile, the vertical lines extend from the

minimum to the maximum when smaller than 1.5 times the 25% and 75% percentiles, and the outliers are represented by open circles.



**Figure 3** Multiple comparison of population marginal means showing the effect of system and song.

In order to understand the differences in rating between the systems, and spot the shortcomings and advantages of each, we held a short, informal discussion with each of the test's participants immediately after they completed the evaluation. The subjects were asked what they thought of the difficulty of the test, in what way they thought the stimuli were different, and how it influenced their ratings of the different fragments. Although they weren't told they were listening to different mixes, nor which type of systems (human or automatic) were involved in the creation of the audio fragments, all 15 participants mentioned they rated the audio fragments at least to some extent based on the balance and audibility of the different sources. For example, at times the lead vocals were found to be too quiet, or the backing vocals

were overpowering. These shortcomings seemed to refer to the unprocessed, monophonic sum, where peak-normalising all sources doesn't guarantee an equal loudness, and the automatic mixing system, which didn't take high-level information into account (three backing vocals can easily overpower one lead vocal, as the system doesn't know which sources should be in the background and which on the foreground).

Many also reported the panning or 'location' of instruments had an influence on their liking of the different fragments, most often in a bad way when either some instruments were panned to the side when they were expected to be more central (e.g. lead vocal, snare drum) or when all instruments were panned to the centre. The former was found to refer to the low-level automatic system, whereas the latter undoubtedly corresponded with the unprocessed, monophonic sum.

In a few cases, the lack of reverb (and the corresponding 'blend') was also mentioned as a general shortcoming, confirming that in order to achieve high-quality mixes, time-based effects should be included in the system. In accordance with the practical literature (Owsinski: 2006), 10 out of 15 participants explicitly mentioned the level or position of vocals.

It should be noted that the used raw audio was recorded in a very professional fashion, and thus virtually free from artefacts or problems. It can be appreciated that a system that does not take signal features such as spectral or dynamic anomalies into account, is not able to fix sonic problems, which automatic systems have shown to be able to solve (Perez-Gonzalez and Reiss: 2009a).

## 5 Conclusions and future perspective

In this work, we assessed the potential of rules directly derived from practical engineering literature for use within a partly or fully autonomous mixing system. We have found such a knowledge-engineered system performs well, with no measured difference in subject preference from professional human mixes, and outperforming a system based on recent automatic mixing work using only low-level features to inform mixing decisions.

In particular, we have found that high-level information, such as distinction between background and foreground instruments, is essential to automatic mixing as systems acting on low-level features alone often fail to make proper mixing decisions as they do not know the instrument roles.

However, we do believe using both low- and high-level features, the latter either supplied by the user or extracted from the audio, could mean a substantial improvement for automatic systems. This would either mean including advanced measuring modules to extract features that can be referred to in the rule base of a semantic mixing system, or embedding semantic rules in an automatic system based on low-level features.

Future work will be concerned with expansion and validation of the rule base. Machine learning, grounded theory and knowledge engineering approaches can all be used to identify or confirm rules:

- Depending on availability, data mining for correlation between parameter settings and signal features (both low- and high-level) could reveal further underlying

rules mixing engineers tend to follow, and confirm or contradict rules already discovered.

- Discovery and validation of rules can happen through perceptual evaluation and psychoacoustic studies, as per grounded theory approach in automatic mixing research.
- Finally, further knowledge can be acquired through books or interviews, although it is advisable to confirm any rules coming from practical sources through careful perceptual evaluation, or supporting data from machine learning approaches. It has been evident that some rules may contradict each other, be specific to a certain genre or mixing style (although this may not be mentioned explicitly), or simply be inaccurate or too vague.

It is also important to provide a suitable format for the rules, to facilitate sharing, editing, adding, and using in a description logic context. Incorporating the Audio Effects Ontology (Wilmering et al.: 2011) will also be instrumental in this matter.

In order to provide a completely autonomous system, an automatic reverberator module should be included too, as this proves to be an essential element of the audio production process, next to dynamic range processing, equalisation, panning and balance.

The audio used during the listening test can be found on [www.brechtdeinan.com/research.html](http://www.brechtdeinan.com/research.html).

## 6 Acknowledgements

The authors wish to thank Pedro Duarte Pestana and Gauthier Grandgirard for providing the human mixes of the test audio, as well as everyone who participated in the listening test.

## Bibliography

- Cartwright, M. and Pardo, B. (2013). Social-EQ: Crowdsourcing an equalization descriptor map. *14th International Society for Music Information Retrieval*.
- Case, A. (2011). *Mix Smart: Professional Techniques for the Home Studio*. Focal Press. Taylor & Francis.
- Case, A. (2012). *Sound FX: Unlocking the Creative Potential of Recording Studio Effects*. Taylor & Francis.
- Clifford, A. and Reiss, J. D. (2010). Calculating time delays of multiple active sources in live sound. In *129th Convention of the Audio Engineering Society*.
- Clifford, A. and Reiss, J. D. (2011a). Microphone interference reduction in live sound. In *Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11)*.
- Clifford, A. and Reiss, J. D. (2011b). Proximity effect detection for directional microphones. In *131st Convention of the Audio Engineering Society*.
- Clifford, A. and Reiss, J. D. (2011c). Reducing comb filtering on different musical instruments using time delay estimation. *Journal of the Art of Record Production, Issue 5*.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, Second Edition*. Lawrence Erlbaum Associates Incorporated.
- Coryat, K. (2008). *Guerrilla Home Recording: How to Get Great Sound from Any Studio (no Matter how Weird Or Cheap Your Gear Is)*. MusicPro guides. Hal Leonard Corporation.

- De Man, B. and Reiss, J. D. (2013). A pairwise and multiple stimuli approach to perceptual evaluation of microphone types. In *134th Convention of the Audio Engineering Society*.
- Giannoulis, D., Massberg, M., and Reiss, J. D. (2012). Parameter automation in a dynamic range compressor. *Journal of the Audio Engineering Society*.
- Gibson, D. (2005). *The Art Of Mixing: A Visual Guide To Recording, Engineering, And Production*. Thomson Course Technology.
- Haas, W. (2008). The SOS guide to mix compression - hitting harder. *Sound On Sound*.
- Izhaki, R. (2008). *Mixing audio: concepts, practices and tools*. Focal Press.
- Katz, B. (2002). *Mastering Audio*. Focal Press.
- Mansbridge, S., Finn, S., and Reiss, J. D. (2012a). An autonomous system for multi-track stereo pan positioning. In *133rd Convention of the Audio Engineering Society*.
- Mansbridge, S., Finn, S., and Reiss, J. D. (2012b). Implementation and evaluation of autonomous multi-track fader control. In *132th Convention of the Audio Engineering Society*.
- Owsinski, B. (2006). *The Mixing Engineer's Handbook*. Course Technology, 2nd edition.
- Perez-Gonzalez, E. and Reiss, J. D. (2007). Automatic mixing: Live downmixing stereo panner. *Proc. of the 10th Int. Conference on Digital Audio Effects*.
- Perez-Gonzalez, E. and Reiss, J. D. (2009a). Automatic equalization of multi-channel audio using cross-adaptive methods. *127th Convention of the Audio Engineering Society*.

- Perez-Gonzalez, E. and Reiss, J. D. (2009b). Automatic gain and fader control for live mixing. *IEEE Workshop on applications of signal processing to audio and acoustics*.
- Perez-Gonzalez, E. and Reiss, J. D. (2010). A real-time semi-autonomous audio panning system for music mixing. *EURASIP Journal on Advances in Signal Processing*.
- Pestana, P. (2013). *Automatic Mixing Systems Using Adaptive Digital Audio Effects*. PhD thesis, Catholic University of Portugal.
- Reiss, J. D. (2011). Intelligent systems for mixing multichannel audio. In *2011 17th International Conference on Digital Signal Processing (DSP)*.
- Robjohns, H. (1999). How and when to use mix compression. *Sound On Sound*.
- Rudolph, B. (1999). Understanding audio compressors and audio compression. *Mix*.
- Scott, J. and Kim, Y. (2011). Analysis of acoustic features for automated multi-track mixing. *Proc. of International Society for Music Information Retrieval (ISMIR)*.
- Scott, J., Prockup, M., Schmidt, E., and Kim, Y. (2011). Automatic multi-track mixing using linear dynamical systems. In *Proceedings of the 8th Sound and Music Computing Conference, Padova, Italy*.
- Senior, M. (2009). Compression made easy - demystifying compressor controls and parameters. *Sound On Sound*.
- Senior, M. (2012). *Mixing Secrets*. Taylor & Francis.
- Tomaz De Carvalho, A. (2012). The Discourse Of Home Recording: Authority Of “Pros” And The Sovereignty Of The Big Studios. *Journal of the Art of Record Production*, Issue 7.

- Toulson, R. (2008). Can we fix it? – the consequences of ‘fixing it in the mix’ with common equalisation techniques are scientifically evaluated. *Journal of the Art of Record Production*, Issue 3.
- White, P. (1996). Compression and limiting - exploration. *Sound On Sound*.
- White, P. (1998). 20 tips on mixing. *Sound On Sound*.
- White, P. (2000a). *Basic Effects & Processors*. The Basic Series. Music Sales.
- White, P. (2000b). *Basic mixing techniques*. The Basic Series. Music Sales.
- White, P. (2000c). Improving your stereo mixing. *Sound On Sound*.
- White, P. (2006). Mixing essentials - the fundamentals of mixing. *Sound On Sound*.
- White, P. (2008). Automation for the people. *Sound On Sound*.
- Wilmering, T., Fazekas, G., and Sandler, M. (2011). Towards ontological representations of digital audio effects. In *Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11)*.