Ten Years of Automatic Mixing

Conference Paper · September 2017

3 authors:

Brecht De Man
Birmingham City University
31 PUBLICATIONS **280** CITATIONS

Joshua Reiss
Queen Mary, University of London
245 PUBLICATIONS **2,178** CITATIONS

Ryan Stables
Birmingham City University
45 PUBLICATIONS **180** CITATIONS

TEN YEARS OF AUTOMATIC MIXING

Brecht De Man and Joshua D. Reiss
Centre for Digital Music
Queen Mary University of London
{b.deman, joshua.reiss}@qmul.ac.uk

Ryan Stables
Digital Media Technology Lab
Birmingham City University
ryan.stables@bcu.ac.uk

ABSTRACT

Reflecting on a decade of Automatic Mixing systems for multitrack music processing, this paper positions the topic in the wider field of Intelligent Music Production, and seeks to motivate the existing and continued work in this area. Tendencies such as the introduction of machine learning and the increasing complexity of automated systems become apparent from examining a short history of relevant work, and several categories of applications are identified. Based on this systematic review, we highlight some promising directions for future research for the next ten years of Automatic Mixing.

1. MOTIVATION

The democratisation of audio technology has enabled music production on limited budgets, putting high-quality results within reach of anyone who has access to a laptop, a microphone and the abundance of free software on the web. Similarly, musicians are able to share their own content at very little cost and effort, again due to high availability of cheap technology. Despite this, a skilled mix engineer is often still needed in order to deliver professional-standard material. Raw, recorded tracks almost always require a considerable amount of processing before being ready for distribution, such as balancing, panning, equalisation (EQ), dynamic range compression and artificial reverberation, to name a few. Furthermore, an amateur music producer will almost inevitably cause sonic problems while recording [1]. Uninformed microphone placement, an unsuitable recording environment, or simply a poor performance or instrument further increases the need for an expert mix engineer [2]. In live situations, especially in small venues, the mixing task is particularly demanding and crucial, due to problems such as acoustic feedback, room resonances and poor equipment. In informal amateur productions, having a competent operator at the desk is the exception rather than the rule. These observations indicate that there is a clear need for systems that take care of the mixing stage of music production for live and recording situations. By obtaining a mix quickly and autonomously, home recording becomes more affordable, smaller music venues are freed from the need for expert operators for their front of house and monitor systems, and musicians can increase their productivity and focus on the creative aspects of music production.

Meanwhile, professional audio engineers are often under pressure to produce high-quality content quickly and at low cost [3]. While they may be unlikely to relinquish control entirely to autonomous mix software, assistance with tedious, time-consuming tasks would be highly beneficial. This can be implemented via more powerful, intelligent, responsive, intuitive algorithms and interfaces [4].

Throughout the history of technology, innovation has traditionally been met with resistance and scepticism, in particular from professional users who fear seeing their roles disrupted or made obsolete. Music production technology may be especially susceptible to this kind of opposition, as it is characterised by a tendency towards nostalgia, skeuomorphisms and analogue workflows [1], and it is concerned with aesthetic value in addition to technical excellence and efficiency. However, the evolution of music is intrinsically linked to the development of new instruments and tools, and essentially utilitarian inventions such as automatic vocal riding, drum machines, electromechanical keyboards and digital pitch correction have been famously used and abused for creative effect. These advancements have changed the nature of the sound engineering profession from primarily technical to increasingly expressive. Generally, there is economic, technological and artistic merit in exploiting the immense computing power and flexibility that today's digital technology affords, to venture away from the rigid structure of the traditional music production toolset.

2. HISTORY

Coined by Dan Dugan, the term 'Automatic Mixing' (or 'Automatic Microphone Mixing') first referred to the application of microphone gain handling for speech [5, 6]. Almost exactly ten years ago, Enrique Perez Gonzalez gave new meaning to the term by publishing a method to automatically adjust not just level, but also stereo panning of multitrack audio [7]. Between 2007 and 2010, he went on to do more work on automating processes for music mixing, including level [8, 11], pan pots [15], EQ [12], unmasking [10] and delay correction [44]. To our knowledge, this was the inception of the field as it is known today.

Figure 1 shows a comprehensive but not exclusive overview of published systems or methods to automate mixing and mastering tasks. Some trends are immediately apparent from this timeline. For instance, machine learning

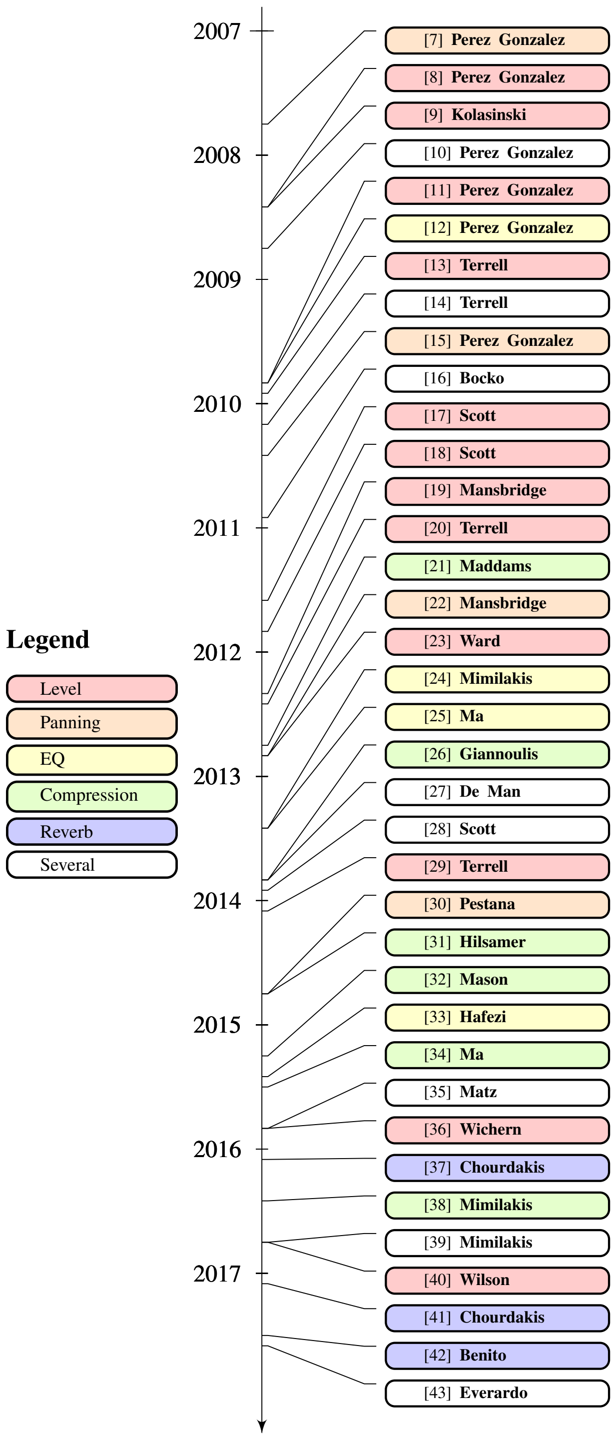

Figure 1: Timeline of prior work 2007–2017

methods seem to be gaining popularity [37–42]. 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 [21, 26, 31, 32, 34, 38] and reverb effects [37, 41, 42].

Table 1 further divides the same works into single track and multitrack, cross-adaptive systems, and indicates which have been evaluated objectively (deviation from a target metric) or subjectively (formal listening test).

Research on such systems has additionally inspired several works on furthering understanding of the complex mix process and its perception, formalising the knowledge on which this systems are based [45–48].

3. APPLICATIONS

Understanding and automating mix engineering processes has many immediate applications, of which some are explored here. They range from completely autonomous mixing systems, to more assistive, workflow-enhancing tools. The boundaries between these categories are vague, and most systems can be adapted for less or more user control.

3.1. Black box

In engineering terms, a 'black box' is a system which can only be judged based on its output signals, in relation to the supplied input. In other words, the user does not know what goes on inside, and cannot control it except by modifying the incoming signals. One or more mix tasks could be automated by such a device so that no sound engineer is required to adjust parameters on a live or studio mix. Many of the academic approaches are presented this way (e.g. [7, 23, 24, 34, 41]), given the appeal of presenting a fully automatic system, although they could be generalised or only partially implemented to give the user more control. The absence of a need—or option—for user interaction is a desired characteristic of complete automatic mixing solutions, for instance for a small concert venue without sound engineers, a band rehearsal, or a conference PA system. A recent surge in work on generative music and automatic composition has further increased the need for fully automated music production systems.

3.2. Assistants

When using Automatic Mixing systems, sound engineers of varying levels will typically want some degree of control, adjusting a small number of parameters of mostly automatic systems. This can range from very limited automation, such as the already common automation of ballistics, time constants and make-up gain in dynamic range compressors [26], to only exposing a handful of controls of a comprehensive mixing system [50]. Rather than corrective

Table 1: Overview of systems that automate music production processes

	Objective evaluation	**Subjective evaluation**	**No evaluation**
Single track	[8, 14, 24, 25, 31, 37]	[24, 26, 38, 41, 49]	[32, 42]
Multitrack	[7, 9–13, 17, 18, 20, 23, 29, 30, 33]	[15, 19, 21, 22, 27, 28, 30, 33–36]	[16, 39, 40, 43]

tools that help obtain a single, allegedly ideal mix [51], this results in creative tools offering countless possibilities and the user-friendly parameters to achieve them. Even within a single processor, extracting relevant features from the audio and adjusting the chosen preset accordingly would represent a dramatic leap over the static presets commonly found in music production software [9]. A more comprehensive mixing system can quickly provide a starting point for a mix, or reach an acceptable balance during a sound check, like a digital assistant engineer. On the multitrack editing side, [52] presents an Intelligent Audio Editor, which uses a MIDI score and music information retrieval methods to correct pitch and timing, and equalise the loudness of coincident notes.

3.3. Interfaces

Another class of intelligent music production tools, complementary to Automatic Mixing in the strict sense, comprises more or less traditional processors controlled in novel ways. For instance, a regular equaliser can be controlled with more semantic and perceptually motivated parameters, such as ‘warm’, ‘crisp’ and ‘full’ [53, 54], which increases accessibility towards novices and enhances the creative flow of some professionals. Deviating from the usual division of labour among signal processing units, the control of a single high-level percept can be achieved by a combination of EQ, dynamic range compression, harmonic distortion, reverberation, or spatial processing. An early example of a mixing GUI, where metaphorical stage positions determine parameters for spatial processing, is described in [55].

3.4. Metering and diagnostics

Finally, even when the traditional controls and processors are preserved entirely, intelligent technologies can play a role by providing additional alerts and visualisations related to high-level signal features. For instance, taking the ubiquitous level and loudness meters, goniometers and spectrograms a step further, the operator can be warned when the overall reverb level is high [56], an instrument is masked [57], or the spectral contour is too ‘boxy’. By defining these high-level attributes as a function of measurable quantities, mix diagnostics become more useful and accessible to both experts and laymen. Such applications also present opportunities for education, where aspiring mix engineers can be informed of which parameter settings are generally considered extreme. Once such perceptually informed issues have been identified, a feedback loop could adjust parameters until the problem is mitigated [48], for instance turning the reverberator level up or down until high-level attribute ‘reverb amount’ enters a predefined range.

4. FUTURE PERSPECTIVES

Despite the coverage of the most relevant processes, the different approaches taken and the available commercial applications, the authors believe the field of Automatic Mixing is still in its infancy.

A significant obstacle to the development of high quality systems, especially those based on machine learning methods, is the relative shortage of reliable data to inform or test assumptions about mix practices [9]. Recent efforts towards sharing datasets [58–61] and accommodating efficient capture of mix actions [62] may help produce the critical mass of data needed for truly intelligent mixing systems.

With analysis of high volumes of data it may also become possible to uncover the rules that govern not just mix engineering in general, but particular mixing styles [63]. From an application point of view, a target profile can thus be applied to source content to mimic the approach of a certain engineer [64], to fit a specific musical genre, or to achieve the most suitable properties for a given medium.

Almost all related work thus far only considers mixes with at most two channels. Expanding the current knowledge and implementations to surround sound, object-based audio and related formats, would allow Automatic Mixing applications in the increasingly important domain of AR and VR systems, as well as game and film audio.

Finally, as the perception of any one source is influenced by the sonic characteristics of other simultaneously playing sources and their processing, the problem of mixing is multidimensional. Consequently, the various types of processing on the individual elements cannot be studied in isolation only. Automatic Mixing of multitrack music remains an unsolved problem, with several established research directions but none of them exhausted. In the next decade, these and other challenges will have to be addressed, possibly revolutionising the music production workflow.

5. REFERENCES

[1] G. Bromham, “How can academic practice inform mix-craft?,” *Mixing Music*, Routledge, 2017.

[2] R. Toulson, “Can we fix it? – The consequences of ‘fixing it in the mix’ with common equalisation tech-

niques are scientifically evaluated," *J. Art of Record Production*, vol. 3, Nov 2008.

[3] A. Pras, C. Guastavino, and M. Lavoie, "The impact of technological advances on recording studio practices," *J. Assoc. Inf. Sci. Technol.*, vol. 64, Mar 2013.

[4] D. Reed, "A perceptual assistant to do sound equalization," *5th Int. Conf. on Intelligent User Interfaces*, Jan 2000.

[5] D. Dugan, "Automatic microphone mixing," *J. Audio Eng. Soc.*, vol. 23, June 1975.

[6] S. Julstrom and T. Tichy, "Direction-sensitive gating: a new approach to automatic mixing," *J. Audio Eng. Soc.*, vol. 32, Jul/Aug 1984.

[7] E. Perez Gonzalez and J. D. Reiss, "Automatic mixing: live downmixing stereo panner," *7th Int. Conf. on Digital Audio Effects (DAFx-07)*, Sep 2007.

[8] E. Perez Gonzalez and J. D. Reiss, "An automatic maximum gain normalization technique with applications to audio mixing," *Audio Engineering Society Conv. 124*, May 2008.

[9] B. Kolasinski, "A framework for automatic mixing using timbral similarity measures and genetic optimization," *Audio Engineering Society Conv. 124*, May 2008.

[10] E. Perez Gonzalez and J. D. Reiss, "Improved control for selective minimization of masking using interchannel dependancy effects," *11th Int. Conf. on Digital Audio Effects (DAFx-08)*, Sep 2008.

[11] E. Perez Gonzalez and J. D. Reiss, "Automatic gain and fader control for live mixing," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct 2009.

[12] E. Perez Gonzalez and J. D. Reiss, "Automatic equalization of multichannel audio using cross-adaptive methods," *Audio Engineering Society Conv. 127*, Oct 2009.

[13] M. J. Terrell and J. D. Reiss, "Automatic monitor mixing for live musical performance," *J. Audio Eng. Soc.*, vol. 57, Nov 2009.

[14] M. J. Terrell, J. D. Reiss, and M. Sandler, "Automatic noise gate settings for drum recordings containing bleed from secondary sources," *EURASIP J. Adv. Sig. Pr.*, Feb 2010.

[15] E. Perez Gonzalez and J. D. Reiss, "A real-time semi-autonomous audio panning system for music mixing," *EURASIP J. Adv. Sig. Pr.*, May 2010.

[16] G. Bocko et al., "Automatic music production system employing probabilistic expert systems," *Audio Engineering Society Conv. 129*, Nov 2010.

[17] J. Scott et al., "Automatic multi-track mixing using linear dynamical systems," *8th Sound and Music Computing Conf.*, July 2011.

[18] J. Scott and Y. E. Kim, "Analysis of acoustic features for automated multi-track mixing," *12th Int. Society for Music Information Retrieval Conf.*, Oct 2011.

[19] S. Mansbridge, S. Finn, and J. D. Reiss, "Implementation and evaluation of autonomous multi-track fader control," *Audio Engineering Society Conv. 132*, Apr 2012.

[20] M. J. Terrell and M. Sandler, "An offline, automatic mixing method for live music, incorporating multiple sources, loudspeakers, and room effects," *Computer Music Journal*, vol. 36, May 2012.

[21] J. A. Maddams, S. Finn, and J. D. Reiss, "An autonomous method for multi-track dynamic range compression," *15th Int. Conf. on Digital Audio Effects (DAFx-12)*, Sep 2012.

[22] S. Mansbridge, S. Finn, and J. D. Reiss, "An autonomous system for multitrack stereo pan positioning," *Audio Engineering Society Conv. 133*, Oct 2012.

[23] D. Ward, J. D. Reiss, and C. Athwal, "Multitrack mixing using a model of loudness and partial loudness," *Audio Engineering Society Conv. 133*, Oct 2012.

[24] S. I. Mimilakis et al., "Automated tonal balance enhancement for audio mastering applications," *Audio Engineering Society Conv. 134*, May 2013.

[25] Z. Ma, J. D. Reiss, and D. A. A. Black, "Implementation of an intelligent equalization tool using Yule-Walker for music mixing and mastering," *Audio Engineering Society Conv. 134*, May 2013.

[26] D. Giannoulis, M. Massberg, and J. D. Reiss, "Parameter automation in a dynamic range compressor," *J. Audio Eng. Soc.*, vol. 61, Oct 2013.

[27] B. De Man and J. D. Reiss, "A knowledge-engineered autonomous mixing system," *Audio Engineering Society Conv. 135*, Oct 2013.

[28] J. Scott and Y. E. Kim, "Instrument identification informed multi-track mixing," *14th Int. Society for Music Information Retrieval Conf.*, Nov 2013.

[29] M. J. Terrell, A. Simpson, and M. Sandler, "The mathematics of mixing," *J. Audio Eng. Soc.*, vol. 62, Jan/Feb 2014.

[30] P. D. Pestana and J. D. Reiss, "A cross-adaptive dynamic spectral panning technique," *17th Int. Conf. on Digital Audio Effects (DAFx-14)*, Sep 2014.

[31] M. Hilsamer and S. Herzog, "A statistical approach to automated offline dynamic processing in the audio mastering process.," *17th Int. Conf. on Digital Audio Effects (DAFx-14)*, Sep 2014.

[32] A. Mason et al., "Adaptive audio reproduction using personalized compression," *Audio Engineering Society 57th Int. Conf. (The Future of Audio Entertainment Technology)*, Mar 2015.

[33] S. Hafezi and J. D. Reiss, "Autonomous multitrack equalization based on masking reduction," *J. Audio Eng. Soc.*, vol. 63, May 2015.

[34] Z. Ma et al., "Intelligent multitrack dynamic range compression," *J. Audio Eng. Soc.*, vol. 63, June 2015.

[35] D. Matz, E. Cano, and J. Abeßer, "New sonorities for early jazz recordings using sound source separation

and automatic mixing tools," *16th Int. Society for Music Information Retrieval Conf.*, Oct 2015.

[36] G. Wichern et al., "Comparison of loudness features for automatic level adjustment in mixing," *Audio Engineering Society Conv. 139*, Oct 2015.

[37] E. T. Chourdakis and J. D. Reiss, "Automatic control of a digital reverberation effect using hybrid models," *Audio Engineering Society 60th Int. Conf. (DREAMS)*, Feb 2016.

[38] S. I. Mimilakis et al., "Deep neural networks for dynamic range compression in mastering applications," *Audio Engineering Society Conv. 140*, May 2016.

[39] S. I. Mimilakis et al., "New sonorities for jazz recordings: Separation and mixing using deep neural networks," *2nd Workshop on Intelligent Music Production*, Sep 2016.

[40] A. Wilson and B. Fazenda, "An evolutionary computation approach to intelligent music production, informed by experimentally gathered domain knowledge," *2nd Workshop on Intelligent Music Production*, Sep 2016.

[41] E. T. Chourdakis and J. D. Reiss, "A machine learning approach to application of intelligent artificial reverberation," *J. Audio Eng. Soc.*, vol. 65, Jan/Feb 2017.

[42] A. L. Benito and J. D. Reiss, "Intelligent multitrack reverberation based on hinge-loss markov random fields," *Audio Engineering Society Int. Conf. (Semantic Audio)*, June 2017.

[43] F. Everardo, "Towards an automated multitrack mixing tool using answer set programming," *14th Sound and Music Computing Conf.*, July 2017.

[44] E. Perez Gonzalez and J. D. Reiss, "Determination and correction of individual channel time offsets for signals involved in an audio mixture," *Audio Engineering Society Conv. 125*, Oct 2008.

[45] P. D. Pestana and J. D. Reiss, "Intelligent audio production strategies informed by best practices," *Audio Engineering Society 53rd Int. Conf. (Semantic Audio)*, Jan 2014.

[46] B. De Man et al., "An analysis and evaluation of audio features for multitrack music mixtures," *15th Int. Society for Music Information Retrieval Conf.*, Oct 2014.

[47] E. Deruty, F. Pachet, and P. Roy, "Human-made rock mixes feature tight relations between spectrum and loudness," *J. Audio Eng. Soc.*, vol. 62, Oct 2014.

[48] A. Wilson and B. Fazenda, "Variation in multitrack mixes: Analysis of low-level audio signal features," *J. Audio Eng. Soc.*, vol. 64, Jul/Aug 2016.

[49] B. De Man and J. D. Reiss, "Adaptive control of amplitude distortion effects," *Audio Engineering Society 53rd Int. Conf. (Semantic Audio)*, Jan 2014.

[50] A. Tsilfidis, C. Papadakos, and J. Mourjopoulos, "Hierarchical perceptual mixing," *Audio Engineering Society Conv. 126*, May 2009.

[51] E. Deruty, "Goal-oriented mixing," *2nd AES Workshop on Intelligent Music Production*, Sep 2016.

[52] R. B. Dannenberg, "An intelligent multi-track audio editor," *Int. Computer Music Conf.*, Aug 2007.

[53] S. Stasis, R. Stables, and J. Hockman, "A model for adaptive reduced-dimensionality equalisation," *18th Int. Conf. on Digital Audio Effects*, Dec 2015.

[54] R. Stables et al., "Semantic description of timbral transformations in music production," *ACM Multimedia*, Oct 2016.

[55] F. Pachet and O. Delerue, "On-the-fly multi-track mixing," *Audio Engineering Society Conv. 109*, Sep 2000.

[56] B. De Man, K. McNally, and J. D. Reiss, "Perceptual evaluation and analysis of reverberation in multitrack music production," *J. Audio Eng. Soc.*, vol. 65, Jan/Feb 2017.

[57] J. Ford, M. Cartwright, and B. Pardo, "MixViz: A tool to visualize masking in audio mixes," *Audio Engineering Society Conv. 139*, Oct 2015.

[58] B. De Man et al., "The Open Multitrack Testbed," *Audio Engineering Society Conv. 137*, Oct 2014.

[59] R. Bittner et al., "MedleyDB: A multitrack dataset for annotation-intensive MIR research," *15th Int. Society for Music Information Retrieval Conf. (ISMIR 2014)*, Oct 2014.

[60] R. Bittner et al., "MedleyDB 2.0: New data and a system for sustainable data collection," *17th Int. Society for Music Information Retrieval Conf. (ISMIR 2016)*, Aug 2016.

[61] B. De Man and J. D. Reiss, "The Mix Evaluation Dataset," *20th Int. Conf. on Digital Audio Effects (DAFx-17)*, Sep 2017.

[62] N. Jillings and R. Stables, "Investigating music production using a semantically powered digital audio workstation in the browser," *Audio Engineering Society Int. Conf. (Semantic Audio)*, Jun 2017.

[63] B. De Man, *Towards a better understanding of mix engineering*. PhD thesis, Queen Mary University of London, May 2017.

[64] H. Katayose, A. Yatsui, and M. Goto, "A mix-down assistant interface with reuse of examples," *Int. Conf. on Automated Production of Cross Media Content for Multi-Channel Distribution*, Nov 2005.