Analysis of Peer Reviews in Music Production

Brecht De Man and Joshua D. Reiss

Analysis of Peer Reviews in Music Production

Brecht De Man and Joshua D. Reiss

Centre for Digital Music, Queen Mary University of London

Abstract

The mix is an essential part of the music production process, which has an important but poorly understood impact on the perception of a record. Little is known about which aspects are the most important, and how to acquire such information. In this work we collect, annotate and analyse over 1400 reviews by trained listeners on 98 mixes. We assess which instruments, types of processing and mix properties are most apparent when comparing mixes, and explore which challenges arise when interpreting these comments. The benefits of using such unstructured data are discussed and a methodology for analysing it is proposed.

Keywords

mix engineering, perception, subjective evaluation, music production, peer review

Introduction

Mixing recorded music is a complex task, with creative and technical aspects (Moylan: 2006, p. xxxvii), that includes dynamically adjusting levels, stereo positions, filter coefficients, dynamic range processing parameters, and effect settings of multiple audio streams (Izhaki: 2008, p. 40). As there are many viable ways to mix a given song, it may not be possible to compile a single set of rules underpinning the mixing process (Case: 2011, p. 15). However, some mixes are clearly favoured over others, suggesting there are 'best practices' in music production, and uncovering them could further understanding of the mixing process and perception of mix decisions, and inform the design of intelligent audio processing tools.

Perceptual evaluation of mixes is essential when investigating these music production practices, as it indicates which processing corresponds with a generally favoured effect. In contrast, when mixes are studied in isolation, without comparison to alternative mixes or without feedback on the engineer's choices, it cannot be assumed that the work is representative of what the audience perceives to be a good mix. In this work, we study the comments of mixing engineers on each other's anonymised mixes of different songs. We analyse these comments in terms of focus on different instruments and processing, and the ratio between positive and negative comments. As such, we show what aspects of a mix are most apparent to trained listeners.

By analysing ratings or rankings of different mixes, we can sometimes indirectly infer which mix qualities or settings are likely detrimental or favourable (De Man et al.: 2015), but it requires a large amount of data to reliably discover a direct relation between preference and a complex objective feature. In contrast, a limited number of short reviews by experienced listeners can reveal a general dislike for an excessive reverb on the lead vocal, an overall lack of low frequency content, or a poorly controlled bass line. More generally, by collecting assessments for a number of mixes of a number of songs, it becomes possible to infer overall tendencies in perception of mixing engineering, and the relative influence of genre and instrumentation on these tendencies. Comments accompanying mix ratings can tell us what type of processing or instruments are likely to catch our attention in a mix we deem excellent or poor, help us find examples of a good or bad treatment of a particular instrument, and reveal differences in perception between listeners of varying degrees of expertise.

Conversely, having assessments of mixes in natural language - as opposed to numeric ratings of different aspects or the mix as a whole (De Man et al.: 2015) - can be instrumental in investigating the meaning of the subjective terms we use to describe sound ('punchy', 'thin', 'muddy', 'flat'), as previously investigated in (Cartwright and Pardo: 2013) and (Stables et al.: 2014). While such subjective terms do not allow

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

3

accurate communication about sound properties (Moylan: 2006, p. xxxvi), (Burgess: 2013, p. 74), they are prevalent among professionals and amateurs to effectively convey complex concepts. A selection of terms often used to describe spectral properties of sound and their definitions by various sound engineering handbooks is included with this work. Eventually, further analysis of the data we collect will allow a more accurate definition of such terms. Meanwhile, this table can be a useful reference for engineers and others seeking to translate unfamiliar terms to a general frequency range.

The double-blind reviews (listening test subjects are unaware of who produced which mix, and mix engineers don't know who wrote which comment) provide a unique type of feedback that is especially useful in an educational setting. Educators and students participating in the experiments described in this work have confirmed that this type of feedback is very insightful, and unlike any form of conventional evaluation where generally only a teacher comments on a student's mix. On the other hand, subjective evaluation participants experience an interesting critical listening exercise that is also useful in an educational setting. By making our tools and multitrack material available to the public, we enable other institutions to use this approach for evaluating recording, mixing and recording exercises as well as practising critical listening skills.

Beyond simply furthering the understanding of perception of the mix engineering process, analysis of this data also enables many important applications. Among them is a

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

4

more informed design of interfaces for hardware and software audio processors, where users manipulate low-level features of a sound by controlling a more abstract, high-level quality (the Waves OneKnob series is a well-known commercial example of allowing the user to adjust a range of low-level features by adjusting just one parameter which has a single, intuitive term). An increasing number of research groups and audio companies are investigating these relationships (Cartwright and Pardo: 2013), (Stables et al.:2014) and whether the definition of these terms holds across different regions and genres (Zacharakis et al.: 2015). Identifying salient terms in these mix reviews and determining how they relate to objective features is essential for the successful creation of such tools. Similarly, through revealing the relation between descriptive terms and the measurable audio features they correspond with, tools that allow searching for audio with a certain high-level property ("Find me a punchy snare drum sample, a fuzzy bass guitar recording, ...") or that indicate or visualise how well a given term describes the considered audio ("Attention: your vocal is too notchy.") are within reach.

A final, practical reason for allowing subjects to write comments when rating as many as ten different versions of the same song is that taking notes on shortcomings or strengths of the mixes helps keep track of which fragment is which and as such facilitates the complex task at hand.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

5

In this work, we explore how mix reviews can be used to further understanding of music production practices and perception thereof, and provide a methodology for aiding research based on this data. Specifically, we demonstrate how quantitative analysis of this data can lead to conclusions about these mixes, and we pave the way for further analysis of this data by discussing potential problems arising when collecting, interpreting and annotating these reviews. Furthermore, we share all subjective evaluation data and, where, possible, the raw audio, stereo mixes and DAW session files for further research, and document how these have been collected.

In the following sections, we describe the experiment through which we obtained mixes for each of the songs, the perceptual evaluation in which we collected comments on each of these mixes, and the analysis of these comments. We investigate which instruments, sonic qualities and processors trained test participants listen for when assessing different mixes of the same source material. We then discuss obstacles we encountered during the annotation process, and possible challenges - and proposed solutions - when interpreting these comments for further research. Based on our findings we outline future work.

Mix Experiment

We asked students and teachers of the MMus in Sound Recording at the Schulich School of Music, McGill University, to produce mixes of ten different songs (De Man et al.: 2014a, De Man et al.: 2015). For each song, one professional mix - often the original released version - and eight student mixes were produced. In all but two instances, an automatic mix based on research from (Mansbridge et al.: 2012a, Mansbridge et al.: 2012b, Giannoulis et al.: 2013, Pestana et al.: 2013, Hafezi and Reiss: 2015, Ma et al.: 2015) was produced as well.

Each student was allowed six hours to create their mix, using an industry standard digital audio workstation (DAW), with included plugins and an additional high quality reverberation plugin. As such, their toolset was very similar to their usual workflow, and constrained enough so that every session could be recalled and analysed in depth. Any form of editing the source material (other than processing with the plugin), rerecording, the use of samples, or otherwise adding new audio was prohibited, to ensure maximum comparability between the different mixes. As the participants were primarily trained to produce high quality mixes, they were not expected to deliver a mastered version for this experiment.

Table 1 lists the songs used in this experiment, along with the number of tracks #T (mono and stereo) and the group of students (eight each) that mixed each particular song[1]. The raw tracks and all mixes (audio and Pro Tools session) of six of these songs are available on the Open Multitrack Testbed (De Man et al.: 2014b) through a Creative Commons license (see 'CC' column in Table 1). The files can be accessed directly at `multitrack.eecs.qmul.ac.uk`.

Table 1: Songs used in the experiment

Artist	Song	#T	Class	Term	CC
The DoneFors	Lead Me	23	1st year	Fall 2013	✓
Joshua Bell, Kristin Chenoweth	My Funny Valentine	17	1st year	Fall 2013	
-[1]	-[1]	22	2nd year	Fall 2013	
Dawn Langstroth	No Prize	20	2nd year	Fall 2013	
Fredy V	Not Alone	24	1st year	Spring 2014	✓
Broken Crank	Red To Blue	39	1st year	Spring 2014	✓
-[1]	-[1]	24	2nd year	Spring 2014	
The DoneFors	Under A Covered Sky	28	2nd year	Spring 2014	✓
Fredy V	In The Meantime	24	1st year	Fall 2014	✓
The DoneFors	Pouring Room	19	2nd year	Fall 2014	✓

---

[1] For two songs permission to disclose artist and song name was not granted.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

8

Perceptual Evaluation

For each song, all mixes were evaluated in a blind perceptual evaluation experiment, where the subjects were all trained listeners - students and staff members at the same institution. Except for the two last songs, where the student mixing engineers assessed their own mix as well, every mix was evaluated by subjects who did not mix it and therefore were not familiar with the song.

Because of the need for a high quality, acoustically neutral listening environment (Moylan: 2006, p. xxxix), the test took place in the Critical Listening Lab at the programme's institution. The impulse responses of this listening environment can be found on `www.brechtdeman.com/research.html`.
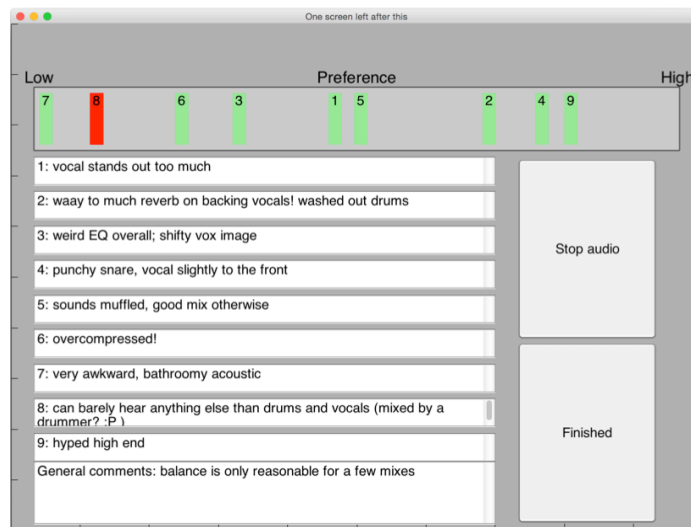

Figure 1: Perceptual evaluation interface

We used the freely available Audio Perceptual Evaluation toolbox from (De Man and Reiss: 2014). For each song, the subjects were presented an interface as shown in Figure 1. Movable markers on the top of the screen allowed ranking/rating of each of the mixes, on a scale from lowest to highest preference. Text boxes on the bottom allowed for comments of any length corresponding with the different mixes, or the song/experiment in general. We randomised the order and starting position of the numbered markers associated with each mix. Only a fragment of each song, corresponding with the second chorus and second verse, was selected for evaluation, to reduce strain on the test participants and to focus on a part during which most sources were active. When switching from one mix to another, the playhead skipped to the corresponding position in the next mix, to avoid excessive focus on the first few seconds of the fragment. A 'Stop audio' button allowed interrupting playback and rewinding to the beginning of the fragment.

In this work, we consider the comments associated with each of the mixes. For an analysis of the attributed ratings and rankings, see (De Man et al.: 2015).

Analysis

In total, we collected 1397 comments, from 1498 mix evaluations: nine to ten mixes of ten songs evaluated by between 13 and 22 trained listeners. Each of these comments is available on `www.brechtdeman.com/research.html`.

For four out of ten songs, the interface we used had one general comment field, where each numbered line corresponded with a mix (De Man and Reiss: 2014). In these experiments we received a comment for 82.1% of the assessed mixes. The average length of each comment was 60.0 characters. Upon providing a separate comment box for each mix (see Figure 1), we witnessed a 47% increase in comment length (88.3 characters on average), and the comment rate was as high as 96.5%. In this case, just two people accounted for the vast majority of the omitted comments.

We annotated each of these comments by treating them as a sequence of *statements*, where each statement referred to a single instrument (or the whole mix) and single processor or sonic attribute. Then, each statement was labelled as referring to a certain instrument or group of instruments (vocals, drums, bass, guitars, keyboards, or the mix as a whole) and a certain processor or feature (balance, space, spectrum, dynamics), as well as classified as 'positive', 'negative' or 'neutral'. We further split up the drums in

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

11

'kick drum', 'snare drum', 'cymbals', or the drums in general, and split up the space-related mix features in panning, reverb, and other.

For instance, the comment

"Drums a little distant. Vox a little hot. Lower midrange feels a little hollow, otherwise pretty good."

consists of the four separate statements "Drums a little distant.", "Vox a little hot.", "Lower midrange feels a little hollow" and "otherwise pretty good.". The first statement relates to the instrument group 'drums' and the sonic feature group 'space', and is a 'negative' comment or criticism. The second statement is labelled as 'vocals', 'level'[2], and 'negative'. The third pertains to the spectral properties of the mix in general (negative) and the fourth is a general, positive remark, again about the whole mix.

The 1397 comments thus resulted in a total of 3633 statements. On average, one comment consisted of 2.8±1.5 statements (median 2). The maximum number of statements within one comment was 10.

We found that 33% of the statements were about the mix in general (or an undefined subset of instruments), 31% regarded vocals (lead or backing) and 19% were related to

---

[2] 'Hot' means 'high in level', from electronic engineering jargon (Nielsen and Lund: 2003).

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

12

drums and percussion (see Figure 2). Within the drums and percussion category, 24%
referred specifically to the snare drum, 22% to the kick drum, and 4% to the hi-hat or
other cymbals. Note that to ensure some degree of comparability, all songs had vocals,
drums, keyboard, and bass guitar tracks, and all but one included guitars. From this, it
can be inferred that in the considered genres the quality of the vocals and their treatment
is crucial for the overall perception of the production, as trained listeners clearly listen
for it and comment on positive and negative aspects in roughly a third of all statements
made.
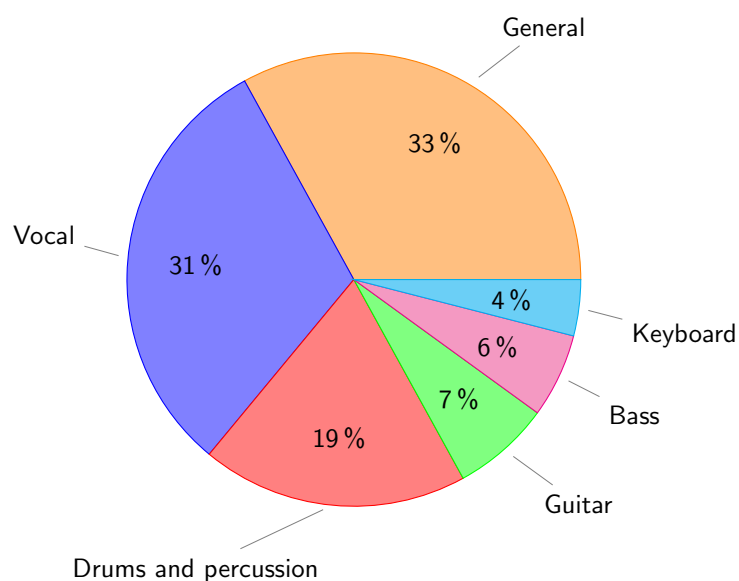
Figure 2: Representation of instrument groups across statements

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

Figure 3 shows the distribution of statements in terms of different processing or feature categories. 'Dynamic processing' includes automation, dynamic range compression and temporary level fixes. Within the category 'space', 58% of the statements were related to reverb, and 16% to panning.
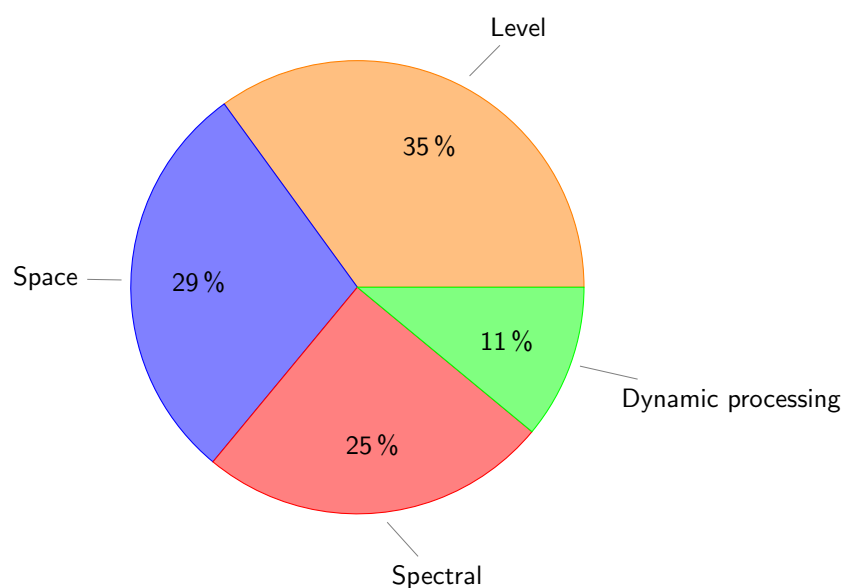
Figure 3: Representation of processors/features across statements

Three out of four statements were some form of criticism on the mix. Of the 24% positive statements, many were more general ('good balance', 'otherwise nice', 'vocals sound great'). In the remaining cases, the statement had no clear positive or negative implication. The difference between the number of positive and negative comments

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

showed a medium correlation (.23) with the numeric rating values from each subject from (De Man et al.: 2015).
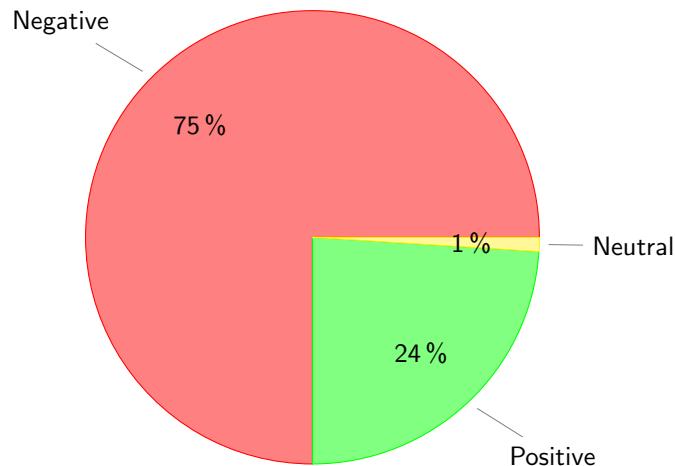
Figure 4: Proportion of negative, positive and neutral statements

Challenges

Which processor/sonic feature does this relate to?

The main challenge with interpreting the comments in this study is that it is often unclear to what processors or objective features the comment relates. Because of the multidimensionality of the complex mix problem, many perceived issues can be attributed to a variety of processors or properties of the perceived sound.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

This is further complicated by the subjects' use of semantic terms to describe sound or treatment which do not have an established relationship with sonic features or processor parameters - even if they are agreed upon by many subjects assessing the same mix and frequently used in a sound engineering context.

> *"Drums are flat and lifeless, no punch at all."*

> *"Snare is clear, but kick is lugubrious..."*

> *"Too much 'poof' in kick. Not enough 'crack' in snare."*

> *"Thinking if it had a bit more 'ooomf' in the lows it would be great."*

> *"Punchy drums."*

> *"I need some more meat from that snare."*

The term 'present' (which could relate to level, reverb, EQ, dynamic processing, and more (Burgess, 2013, p. 73)) is but one example of this.

> *"Electric guitar is a little too present."*

> *"Vox nice and present."*

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

16

"Hi-hat too present."

"Lead vocals sometimes not present enough, other times too present."

Some terms are related to a lack, presence or excess of energy in a certain frequency band in mixing handbooks, but even then this is not rigorously investigated, and the associated frequency range varies between and even within sources.

"Vocal a little thick."

"Piano a little muddy."

"Kick is a bit tubby sometimes."

"Drums sound a little thin to me."

"Very bright guitars."

"Vocal sounds dull."

"Guitars have no bite."

"Bass is dark."

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

"Nasal vocals."

"Guitars are woofy and too dark."

Table 2 lists terms used in such literature with their associated frequency ranges. Due to the subjective nature of these terms, their definitions vary - sometimes within the same book - and are meant as an approximation. Several sources also suggest that the frequency range to which such a term refers may depend on the instrument in question (Izhaki: 2008, p. 245), (Huber and Runstein: 2013, p. 485), (Gibson: 2005, p. 118). In some cases, it needs to be made clear that the term refers to a lack of energy in this band. For instance, 'dark' would refer to a lack of high frequencies. In other cases, the term is simply associated with a certain frequency range: more or less 'edge' depends on more or less energy in that frequency range. This table serves as a rough translation from subjective terms to excesses or deficiencies in particular frequency ranges. Conversely, the data from the research presented herein allows one to investigate whether these terms indeed correspond with a relatively higher or lower energy in this range.

Some statements are more generic and offer even less information on which of the mix's properties, instruments or processing the subject was listening for or (dis)pleased by.

"Nice balance."

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

18

"Best mix."

"Lead vocal sounds good."

"Nice vocal treatment."

"Bad balance."

"Guitars sound horrible."

"This is awful."

"Everything is completely wrong."

On the other hand, such a vague assessment of a complete mix, a certain aspect of the mix, or the treatment of a specific instrument can be useful when looking for examples of appropriate or poor processing.

Good or bad?

In other instances, it is not clear whether a statement is meant as positive (highlighting a strength) or negative (criticising a poor decision), or neither (neutral).

"Pretty dry."

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

19

"Lots of space."

"Round mix."

"Wide imaging."

"Big vocals."

"This mix kind of sounds like Steely Dan."

However, many of these can be better understood by considering other comments of the same person (if a similar statement, or its opposite, was made with regards to a different mix of the same song, and had a clear positive or negative connotation), other statements in the same comment (e.g. two statements separated by the word 'but' will mostly be a positive and a negative one), comments by other subjects on the same mix (who may express themselves more clearly and remark similar things about the mix), or the rating attributed to the corresponding mix by the subject (e.g. if the mix received one of the lowest ratings from the mix, the comment associated with it will most likely consist of mostly negative statements).

Another statement category consisted of mentions of allegedly bold decisions, that the subject condoned or approved of despite sounding unusual.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

20

"A lot of reverb but kind of pulling it off."

"Horns a bit hot, but I kind of like it except in the swells."

"Hated the vocal effect but in the end got used to it, nice one."

"Most reverb on the vocals thus far, but I like it."

This highlights a potential issue with comparative perceptual evaluation studies to reveal 'best practices' in mixing: there may be a tendency towards (or away from) more conventional sounding mixes and more mundane artistic decisions when several versions are judged simultaneously, whereas commercial music is typically available as one mix only, meaning that bold mix moves may not be questioned to the same extent by the listener.

Cryptic comments

It takes at least a basic background in sound engineering to interpret the following statements.

"Kick has no punch."

"Lots of drum spots."

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

21

"Vocals too wet."

Someone with sound engineering background will know to connect the use of the word 'punchy' to the dynamic features of the signal (Fenton et al.: 2014), that 'spots' refers to microphones at close distance (Faller and Erne: 2005), and that the term 'wet' is used here to denote an excessive amount of reverberation (Barthet and Sandler: 2010). Again, Table 2 and practical literature in general can help understand the more or less 'established' terms.

On the other hand, some comments are hard to understand even with years of sound engineering expertise, possibly because the subject forgot to complete his sentence, or because they are intended mainly to remind the subject which mix was which.

"Vocals."

"Reverb."

"Get the music."

Scaling these experiments to substantially higher numbers of evaluations could prompt automated processing of comments, using natural language processing (NLP) or similar, but because of the lack of constraints many comments are near impossible to interpret by a machine. It would be very challenging to automatically and reliably extract instrument,

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

22

process or feature, and whether the statement is meant as criticism or highlighting a strength, especially when humour is involved:

"Why is the singer in the bathroom?"

"Where are the drums? 1 800 drums? Long distance please come home in time for dinner..."

"Is this a drum solo album instead of a lead female group?"

"Do you hate high frequencies?"

"Lead vocal, bass, and drum room does not a mix make."

"No bass. No kick. No like."

"If that was not made by a robot, that person has no soul."

It would also take an advanced algorithm to understand these speculations about the mixing engineer's main instrument:

"Sounds like drummer mixed it..."

"Mixed by a drummer?"

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

"Guitar player's mix?"

or the following comic references (each from a different participant):

"Holy hi-hat!"

"Holy high end Batman!"

"Holy reverb, Batman!"

"Holy noise floor & drum compression!"

At this point, it seems a trade-off has to be made: do we want to process large amounts of machine-readable feedback, by imposing constraints on the feedback, or do we prefer a free form text field so as not to interrupt or bias the subject's train of thought? If we were to collect feedback with a limited vocabulary (for instance borrowing from the Audio Effects Ontology (Wilmering et al.: 2011), Music Ontology (Raimond et al.: 2007) and Studio Ontology (Fazekas and Sandler: 2011)), or via user interface elements such as checkboxes and sliders instead of text fields, we could almost effortlessly acquire unambiguous information on the processing of different sources in different mixes. This data could then readily be used by machine learning systems or other data processing tools. On the other hand, by studying free-form text feedback we learn how listeners naturally react to differences in music production, and even what

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

24

exactly these ill-defined terms and expressions mean and how they relate to different aspects of the mix. Which approach to choose therefore has to be informed by the research questions at hand.

Concluding remarks

We conducted a study where a large number of mixes of ten different songs were produced and perceptually evaluated. This led to a body of over 3600 statements describing different aspects of the mixes. We annotated each of those and discussed the distribution of references to instruments, processors, and sonic features. From this data, we quantified the proportion of attention paid to different instruments, types of processing, and categories of features, showing which are the most apparent and therefore arguably most important or challenging aspects of the mix, as perceived by trained listeners. Most of the statements were criticising aspects of the mix rather than praising them.

We provided motivation for creating this dataset and presenting a methodology for practical use in a research context. Some challenges in the interpretation of these statements were considered and, where possible, solutions were proposed. The main challenge when deriving meaningful information about the mix from its reviews, is to understand to which process or objective feature a statement relates. The wealth of

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

25

subjective terms used in the assessments of mixes is an important obstacle in this regard. A rough translation of terms relating to spectral properties of sound has been summarised in Table 2, as a proof of concept of how practical audio engineering literature can help interpret common but non-technical descriptive terms. Furthermore, reliably inferring whether a short review is meant as positive or negative is not always possible; however, considering numerical rating or ranking of the same mix as well as comments by others on the same mix, or by the subject on other mixes, often helps determine what the subject intended. Finally, due to the rich vocabulary and at times cryptic expressions used to describe various aspects of the mix in free-form text, the tedious annotation process could only be automated if feedback were more constrained. While these obstacles certainly increase the effort and time needed to annotate and analyse these comments, merits of free-text reviews as opposed to constrained feedback are highlighted as well. This discussion lays the foundation for in-depth analysis of such data.

We shared the raw tracks and mixes of six songs, the reviews on all mixes of all ten songs, the perceptual evaluation toolbox, and the listening room's impulse responses for reference and further research.

Future work

In this work, we analysed mixes from engineers at one particular institution, and investigated perception of those mixes by subjects at the same location. To remove the potential bias that comes with the fact that mixing engineers in this experiment were taught by the same faculty, live in the same city (Owsinski: 2006, p. 3), and have a similar level of experience, the experiment will be repeated at different schools. At the same time, this will allow us to investigate what influence geographic location and background has on mixing practices and perception thereof. Perceptual evaluation of the mixes will take place at other sites than where the mixes were produced. A larger body of comments on an expanded set of source material would improve the accuracy and the usefulness of the results of any analysis of this data.

The mix reviews collected in this and future studies warrant more advanced analysis, including but not limited to deriving preferred ranges of parameters and features (by identifying upper and lower bounds of these values outside of which subjects seem to find it too low or high), inferring the most damaging defects and rewarding features (by investigating which statements generally correlate strongly with rating or ranking), and how comments given relate to comments received (a subject who finds the level of the drums is too low on many occasions may be likely to mix drums loudly and receive feedback reflecting this). Furthermore, approximate definitions like those provided in

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

27

Table 2 could be empirically confirmed or improved by analysing the spectral properties of sound described with such a term, and comparing them to spectral properties of differently processed versions of the same source.

We highly welcome institutions to take part in further experiments, to contribute to this research and help their students receive feedback in a unique way.

Resources

Open Multitrack Testbed: `multitrack.eecs.qmul.ac.uk`

Listening test tool: `code.soundsoftware.ac.uk/projects/ape`

Browser-based listening test tool:

`code.soundsoftware.ac.uk/projects/webaudioevaluationtool`

Other data: `brechtdeman.com/research.html`

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

About the authors

Brecht De Man

b.deman@qmul.ac.uk

Joshua D. Reiss

joshua.reiss@qmul.ac.uk

Centre for Digital Music

School of Electronic Engineering and Computer Science

Queen Mary University of London

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

29

Table 2: Spectral descriptors in mixing engineering literature. The frequency ranges in this table are approximate, as they are often derived from figures, and originally meant as a rough estimate of the frequencies corresponding with the terms in their respective contexts.

Term	Range	Reference
air[3]	5-8 kHz	(Gibson: 2005, p. 119)
	10-20 kHz	(Coryat: 2008, p. 99)
	10-20 kHz	(Izhaki: 2008, p. 211)
	11-22.5 kHz	(Owsinski: 2006, p. 26)
	12-15 kHz	(White: 1999, p. 103)
	12-16 kHz	(Katz: 2002, p. 43)
	12-20 kHz	(Coryat: 2008, p. 25)
	12-20 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 108)
	12-20 kHz	(Waddell: 2013, p. 86)
anemic	lack of 20-110 Hz	(Izhaki: 2008, p. 211)
	lack of 40-200 Hz	(Gibson: 2005, p. 119)
articulate	800-5000 Hz	(Gibson: 2005, p. 119)
ballsy	40-200 Hz	(Gibson: 2005, p. 119)
barrelly	200-800 Hz	(Gibson: 2005, p. 119)
bathroomy	800-5000 Hz	(Gibson: 2005, p. 119)
beefy	40-200 Hz	(Gibson: 2005, p. 119)
big	40-250 Hz	(Owsinski: 2006, p. 25)
bite	2-6 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 106)
	2.5 kHz	(Huber and Runstein: 2013, p. 484)
body	100-500 Hz	(Coryat: 2008, p. 99)

---

[3] In some books, 'air' is also used to denote a part of the audible frequency range, exceeding 'highs' (Waddell: 2013, p. 86), (Cousins and Hepworth-Sawyer: 2013, p. 108).

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

30

	100-500 Hz	(Izhaki: 2008, p. 211)
	150-600 Hz	(Coryat: 2008, p. 24)
	200-800 Hz	(Gibson: 2005, p. 119)
	240 Hz	(Huber and Runstein: 2013, p. 484)
boom(y)	20-100 Hz	(Izhaki: 2008, p. 211)
	40-200 Hz	(Gibson: 2005, p. 119)
	60-250 Hz	(Owsinski: 2006, p. 25)
	62-125 Hz	(Katz: 2002, p. 43)
	90-175 Hz	(Owsinski: 2006, p. 26)
	200-240 kHz	(Huber and Runstein: 2013, p. 484)
bottom	40-100 Hz	(Gibson: 2005, p. 119)
	45-90 Hz	(Owsinski: 2006, p. 26)
	60-120 Hz	(Huber and Runstein: 2013, p. 484)
	62-300 Hz	(Katz: 2002, p. 43)[4]
boxy, boxiness	250-800 Hz	(Izhaki: 2008, p. 211)
	300-600 Hz	(Owsinski: 2006, p. 31)
	300-900 Hz	(Katz: 2002, p. 43)
	800-5000 Hz	(Gibson: 2005, p. 119)
bright	2-12 kHz	(Katz: 2002, p. 43)
	2-20 kHz	(Izhaki: 2008, p. 211)
	5-8 kHz	(Gibson: 2005, p. 119)
brilliant, brilliance	5-8 kHz	(Gibson: 2005, p. 119)
	5-11 kHz	(Izhaki: 2008, p. 211)
	5-20 kHz	(Huber and Runstein: 2013, p. 484)
	6-16 kHz	(Owsinski: 2006, p. 25)
brittle	5-20 kHz	(Huber and Runstein: 2013, p. 484)
	6-20 kHz	(Coryat: 2008, p. 25)
cheap	lack of 8-12 kHz	(Gibson: 2005, p. 119)
chunky	800-5000 Hz	(Gibson: 2005, p. 119)
clarity	2.5-4 kHz	(Waddell: 2013, p. 86)
	2.5-5 kHz	(Huber and Runstein: 2013, p. 484)
	3-12 kHz	(Izhaki: 2008, p. 211)
	4-16 kHz	(Owsinski: 2006, p. 26)
clear	5-8 kHz	(Gibson: 2005, p. 119)
close	2-4 kHz	(Huber and Runstein: 2013, p. 484)
	4-6 kHz	(Owsinski: 2006, p. 25)
colour	80-1000 Hz	(Izhaki: 2008, p. 211)

---

[4] More specifically, Katz calls this 'extended bottom'.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

31

covered	lack of 800-5000 Hz	(Gibson: 2005, p. 119)
crisp, crispness	3-12 kHz	(Izhaki: 2008, p. 211)
	5-10 kHz	(Huber and Runstein: 2013, p. 484)
	5-12 kHz	(Gibson: 2005, p. 119)
crunch	200-800 Hz	(Gibson: 2005, p. 119)
	1400-2800 Hz	(Owsinski: 2006, p. 26)
cutting	5-8 kHz	(Gibson: 2005, p. 119)
dark	lack of 5-8 kHz	(Gibson: 2005, p. 119)
dead	lack of 5-8 kHz	(Gibson: 2005, p. 119)
definition	2-6 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 106)
	2-7 kHz	(Izhaki: 2008, p. 211)
	6-12 kHz	(Owsinski: 2006, p. 26)
disembodied	200-800 Hz	(Gibson: 2005, p. 119)
distant	lack of 200-800 Hz	(Gibson: 2005, p. 119)
	lack of 700-20 000 Hz	(Izhaki: 2008, p. 211)
	lack of 4-6 kHz	(Owsinski: 2006, p. 25)
	lack of 5 kHz	(Huber and Runstein: 2013, p. 484)
dull	lack of 4-20 kHz	(Izhaki: 2008, p. 211)
	lack of 5-8 kHz	(Gibson: 2005, p. 119)
	lack of 6-16 kHz	(Katz: 2002, p. 43)
edge, edgy	800-5000 Hz	(Gibson: 2005, p. 119)
	1-8 kHz	(Izhaki: 2008, p. 211)
	3-6 kHz	(Owsinski: 2006, p. 26)
	4-8 kHz	(Katz: 2002, p. 43)
fat	50-250 Hz	(Izhaki: 2008, p. 211)
	60-250 Hz	(Owsinski: 2006, p. 25)
	62-125 Hz	(Katz: 2002, p. 43)
	200-800 Hz	(Gibson: 2005, p. 119)
	240 Hz	(Huber and Runstein: 2013, p. 484)
flat	lack of 8-12 kHz	(Gibson: 2005, p. 119)
forward	800-5000 Hz	(Gibson: 2005, p. 119)
full(ness)	40-200 Hz	(Gibson: 2005, p. 119)
	80-240 Hz	(Huber and Runstein: 2013, p. 484)
	100-500 Hz	(Coryat: 2008, p. 99)
	175-350 Hz	(Owsinski: 2006, p. 26)
	250-350 Hz	(Katz: 2002, p. 43)
glare	8-12 kHz	(Gibson: 2005, p. 119)
glassy	8-12 kHz	(Gibson: 2005, p. 119)

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

32

harsh	2-10 kHz	(Izhaki: 2008, p. 211)
	2-12 kHz	(Katz: 2002, p. 43)
	5-20 kHz	(Huber and Runstein: 2013, p. 484)
heavy	40-200 Hz	(Gibson: 2005, p. 119)
hollow	lack of 200-800 Hz	(Gibson: 2005, p. 119)
honk(y)	350-700 Hz	(Owsinski: 2006, p. 26)
	400-3000 Hz	(Izhaki: 2008, p. 211)
	600-1500 Hz	(Coryat: 2008, p. 24)
	800-5000 Hz	(Gibson: 2005, p. 119)
horn-like	500-1000 Hz	(Huber and Runstein: 2013, p. 484)
	500-1000 Hz	(Owsinski: 2006, p. 25)
	800-5000 Hz	(Gibson: 2005, p. 119)
impact	62-400 Hz	(Katz: 2002, p. 43)
intelligible	800-5000 Hz	(Gibson: 2005, p. 119)
	2-4 kHz	(Huber and Runstein: 2013, p. 484)
in-your-face	1.5-6 kHz	(Coryat: 2008, p. 24)
lisping	2-4 kHz	(Owsinski: 2006, p. 25)
live	5-8 kHz	(Gibson: 2005, p. 119)
loudness	2.5-6 kHz	(Izhaki: 2008, p. 211)
	5 kHz	(Huber and Runstein: 2013, p. 484)
metallic	5-8 kHz	(Gibson: 2005, p. 119)
mud(dy)	16-60 Hz	(Owsinski: 2006, p. 26)
	20-400 Hz	(Izhaki: 2008, p. 211)
	60-500 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 104)
	150-600 Hz	(Coryat: 2008, p. 24)
	175-350 Hz	(Owsinski: 2006, p. 26)
	200-400 Hz	(Katz: 2002, p. 43)
	200-800 Hz	(Gibson: 2005, p. 119)
muffled	lack of 800-5000 Hz	(Gibson: 2005, p. 119)
nasal	400-2500 Hz	(Izhaki: 2008, p. 211)
	500-1000 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 105)
	700-1200 Hz	(Katz: 2002, p. 43)
	800-5000 Hz	(Gibson: 2005, p. 119)
natural tone	80-400 Hz	(Izhaki: 2008, p. 211)
oomph	150-600 Hz	(Coryat: 2008, p. 24)
phonelike	800-5000 Hz	(Gibson: 2005, p. 119)
piercing	5-8 kHz	(Gibson: 2005, p. 119)

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

33

point	1-4 kHz	(Owsinski: 2006, p. 27)
power(ful)	16-60 Hz	(Owsinski: 2006, p. 26)
	40-200 Hz	(Gibson: 2005, p. 119)
	40-100 Hz	(Izhaki: 2008, p. 211)
presence	800-12k Hz	(Gibson: 2005, p. 119)
	1.5-6 kHz	(Coryat: 2008, p. 24)
	2-8 kHz	(Katz: 2002, p. 43)
	2-11 kHz	(Izhaki: 2008, p. 211)
	2.5-5 kHz	(Huber and Runstein: 2013, p. 484)
	4-6 kHz	(Owsinski: 2006, p. 25)
projected	800-5000 Hz	(Gibson: 2005, p. 119)
punch	40-200 Hz	(Gibson: 2005, p. 119)
	62-250 Hz	(Katz: 2002, p. 43)[5]
robustness	200-800 Hz	(Gibson: 2005, p. 119)
round	40-200 Hz	(Gibson: 2005, p. 119)
rumble	20-100 Hz	(Izhaki: 2008, p. 211)
	40-200 Hz	(Gibson: 2005, p. 119)
screamin'	5-12 kHz	(Gibson: 2005, p. 119)
searing	8-12 kHz	(Gibson: 2005, p. 119)
sharp	8-12 kHz	(Gibson: 2005, p. 119)
shimmer	7.5-12 kHz	(Huber and Runstein: 2013, p. 484)
shrill	5-7.5 kHz	(Huber and Runstein: 2013, p. 484)
	5-8 kHz	(Gibson: 2005, p. 119)
sibilant, sibilance	2-8 kHz	(Izhaki: 2008, p. 211)
	2-10 kHz	(Katz: 2002, p. 43)
	4 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 120)
	5-20 kHz	(Huber and Runstein: 2013, p. 484)
	6-12 kHz	(Owsinski: 2006, p. 26)
	6-16 kHz	(Owsinski: 2006, p. 25)
sizzle, sizzly	6-20 kHz	(Izhaki: 2008, p. 211)
	7-12 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 107)
	8-12 kHz	(Gibson: 2005, p. 119)
slam	62-200 Hz	(Katz: 2002, p. 43)
smooth	5-8 kHz	(Gibson: 2005, p. 119)
solid(ity)	35-200 Hz	(Izhaki: 2008, p. 211)

[5] More specifically, Katz calls this 'punchy bass'.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

34

	40-200 Hz	(Gibson: 2005, p. 119)
	62-250 Hz	(Katz: 2002, p. 43)
sparkle, sparkling	5-10 kHz	(Owsinski: 2006, p. 27)
	5-15 kHz	(Izhaki: 2008, p. 211)
	5-20 kHz	(Huber and Runstein: 2013, p. 484)
	8-12 kHz	(Gibson: 2005, p. 119)
steely	5-8 kHz	(Gibson: 2005, p. 119)
strident	5-8 kHz	(Gibson: 2005, p. 119)
sub-bass	16-60 Hz	(Owsinski: 2006, p. 25)
subsonic	0-20 Hz	(Izhaki: 2008, p. 209)
	0-25 Hz	(Waddell: 2013, p. 84)
	10-60 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 102)
sweet	250-400 Hz	(Katz: 2002, p. 43)
	250-2000 Hz	(Owsinski: 2006, p. 25)
thickness	20-500 Hz	(Izhaki: 2008, p. 211)
	40-200 Hz	(Gibson: 2005, p. 119)
	200-750 Hz	(Katz: 2002, p. 43)
thin	lack of 20-200 Hz	(Izhaki: 2008, p. 211)
	lack of 40-200 Hz	(Gibson: 2005, p. 119)
	lack of 60-250 Hz	(Owsinski: 2006, p. 25)
	lack of 62-600 Hz	(Katz: 2002, p. 43)
thump	40-200 Hz	(Gibson: 2005, p. 119)
	90-175 Hz	(Owsinski: 2006, p. 26)
tinny	1-2 kHz	(Huber and Runstein: 2013, p. 484)
	1-2 kHz	(Owsinski: 2006, p. 25)
	5-8 kHz	(Gibson: 2005, p. 119)
tone	500-1000 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 105)
transparent	lack of 4-6 kHz	(Owsinski: 2006, p. 25)
tubby	200-800 Hz	(Gibson: 2005, p. 119)
veiled	lack of 800-5000 Hz	(Gibson: 2005, p. 119)
warm, warmth	90-175 Hz	(Owsinski: 2006, p. 26)
	100-600 Hz	(Izhaki: 2008, p. 211)
	200 Hz	(Huber and Runstein: 2013, p. 484)
	200-800 Hz	(Gibson: 2005, p. 119)
	200-500 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 105)
	250-600 Hz	(Katz: 2002, p. 43)

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

35

whack	700-1400 Hz	(Owsinski: 2006, p. 26)
wimpy	lack of 40-200 Hz	(Gibson: 2005, p. 119)
	lack of 55-500 Hz	(Izhaki: 2008, p. 211)
woody	800-5000 Hz	(Gibson: 2005, p. 119)
woofy	800-5000 Hz	(Gibson: 2005, p. 119)
zing	4-10 kHz	(Coryat: 2008, p. 99)
	10-12 kHz	(Coryat: 2008, p. 24)
bass/low end/ lows	25-120 Hz	(Waddell: 2013, p. 84)
	20-150 Hz	(Coryat: 2008, p. 23)
	20-250 Hz	(Izhaki: 2008, p. 209)[6]
	20-250 Hz	(Katz: 2002, p. 43)
	20-250 Hz	(Mixerman: 2010, p. 72)
	40-200 Hz	(Gibson: 2005, p. 119)
	60-150 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 103)
	60-250 Hz	(Owsinski: 2006, p. 25)
low mids/ lower midrange	120-400 Hz	(Waddell: 2013, p. 85)
	150-600 Hz	(Coryat: 2008, p. 24)
	200-500 Hz	(Cousins and Hepworth-Sawyer: 2013, p. 104)
	250-500 Hz	(Katz: 2002, p. 43)
	200-800 Hz	(Gibson: 2005, p. 119)
	250-1000 Hz	(Mixerman: 2010, p. 73)
	250-2000 Hz	(Owsinski: 2006, p. 25)
	250-2000 Hz	(Izhaki: 2008, p. 209)
(high) mids/ upper midrange	250-6000 Hz	(Katz: 2002, p. 43)[7]
	350-8000 Hz	(Waddell: 2013, p. 85)[8]
	600-1500 Hz	(Coryat: 2008, p. 24)
	800-5000 Hz	(Gibson: 2005, p. 119)
	1-10 kHz	(Mixerman: 2010, p. 73)
	1.5-6 kHz	(Coryat: 2008, p. 24)

---

[6] Izhaki distinguishes between low bass (20-60 Hz), mid bass (60-120 Hz) and upper bass (120-250 Hz)

[7] Katz distinguishes between lower midrange (250-500 Hz), midrange (250-2000 Hz) and upper midrange (2-6 kHz).

[8] Waddell distinguishes between midrange (350-2000 Hz) and upper midrange (2-8 kHz).

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

36

	2-4 kHz	(Owsinski: 2006, p. 25)
	2-6 kHz	(Izhaki: 2008, p. 209)
	2-6 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 106)
highs/high end/ treble	5-12 kHz	(Gibson: 2005, p. 119)[9]
	6-20 kHz	(Izhaki: 2008, p. 209)
	6-20 kHz	(Coryat: 2008, p. 24)
	6-20 kHz	(Katz: 2002, p. 43)[10]
	7-12 kHz	(Cousins and Hepworth-Sawyer: 2013, p. 107)
	8-12 kHz	(Waddell: 2013, p. 86)
	10-20 kHz	(Mixerman: 2010, p. 74)

References

Barthet, M. and Sandler, M. (2010). On the effect of reverberation on musical instrument automatic recognition. In *128th Convention of the Audio Engineering Society*.

Burgess, R. J. (2013). *The Art of Music Production: The Theory and Practice*. OUP USA.

---

[9] Gibson distinguishes between highs (5-8 kHz) and super highs (8-12 kHz)
[10] Katz distinguishes between lower treble or highs (6-12 kHz) and extreme treble (12-20 kHz).

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

37

Cartwright, M. and Pardo, B. (2013). Social-EQ: Crowdsourcing an equalization descriptor map. *14th International Society for Music Information Retrieval Conference (ISMIR 2013)*.

Case, A. (2011). *Mix Smart: Professional Techniques for the Home Studio*. Focal Press.

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.

Cousins, M. and Hepworth-Sawyer, R. (2013). *Practical Mastering: A Guide to Mastering in the Modern Studio*. Taylor & Francis.

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

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)*.

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

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

38

De Man, B. and Reiss, J. D. (2014). APE: Audio Perceptual Evaluation toolbox for MATLAB. In *136th Convention of the Audio Engineering Society*.

Faller, C. and Erne, M. (2005). Modifying stereo recordings using acoustic information obtained with spot recordings. In *118th Convention of the Audio Engineering Society*.

Fazekas, G. and Sandler, M. B. (2011). The Studio Ontology framework. In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*.

Fenton, S., Lee, H., and Wakefield, J. (2014). Elicitation and objective grading of punch within produced music. In *136th Convention of the Audio Engineering Society*.

Giannoulis, D., Massberg, M., and Reiss, J. D. (2013). Parameter automation in a dynamic range compressor. *Journal of the Audio Engineering Society*, 61(10):716–726.

Gibson, D. (2005). *The Art Of Mixing: A Visual Guide To Recording, Engineering, And Production*. Thomson Course Technology.

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

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

39

Huber, D. and Runstein, R. (2013). *Modern Recording Techniques*. Taylor & Francis.

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

Katz, B. (2002). *Mastering Audio*. Focal Press.

Ma, Z., De Man, B., Pestana, P. D., Black, D. A. A., and Reiss, J. D. (2015). Intelligent multitrack dynamic range compression. *to appear in Journal of the Audio Engineering Society*.

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 *132nd Convention of the Audio Engineering Society*.

Mixerman (2010). *Zen and the Art of Mixing*. Hal Leonard Corporation.

Moylan, W. (2006). *Understanding and Crafting the Mix: The Art of Recording*. Focal Press, 2nd edition.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

40

Nielsen, S. H. and Lund, T. (2003). Overload in signal conversion. In *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*.

Owsinski, B. (2006). *The Mixing Engineer's Handbook*. Course Technology, 2nd edition.

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

Raimond, Y., Abdallah, S. A., Sandler, M. B., and Giasson, F. (2007). The Music Ontology. In *8th International Society for Music Information Retrieval Conference (ISMIR 2007)*, pages 417–422.

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

Waddell, G. (2013). *Complete Audio Mastering: Practical Techniques*. McGraw-Hill Education.

White, P. (1999). *Basic Mixers*. The Basic Series. Music Sales.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

41

Wilmering, T., Fazekas, G., and Sandler, M. (2011). Towards ontological representations of digital audio effects. In *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*.

Zacharakis, A., Pastiadis, K. and Reiss, J. D. (2015). An Interlanguage Unification of Musical Timbre: Bridging Semantic, Perceptual and Acoustic Dimensions. In *Music Perception*, 32, 4.

B. De Man and J. D. Reiss, "Analysis of Peer Reviews in Music Production," Journal on the Art of Record Production, Vol. 10, 2015.

42