Intelligent Audio Production Strategies Informed by Best Practices

Pestana, PD; Reiss, JD; AES 53rd International Conference on Semantic Audio

Intelligent Audio Production Strategies Informed by Best Practices

Pedro Pestana[1], and Joshua D. Reiss[2]

[1] *School of the Arts, Catholic University, Oporto, Portugal*

[2] *Queen Mary University of London, London, UK*

Correspondence should be addressed to Pedro Pestana (ppestana@porto.ucp.pt)

ABSTRACT

The main focus of this article is to explore and investigate the fundamental constraints that should be at the basis of algorithm development in intelligent audio production systems. Through mix analysis and grounded theory strategies, a best-practices framework on the craft of mixing is sought out. Findings, while not to be taken as dogmatic, give a clear indication of preferred implementation strategies, and show what still needs to be done to fully understand the technical choices that audio mixing has incorporated throughout its history.

1. CONTEXT

The last five years have witnessed blooming of research in the field of automatic mixing [1], powered by cross-adaptive digital audio algorithms [2]. Most of the developed strategies, while showing promising results, have mainly relied on the author's experience, or on literature review where literature is exiguous. We argue that a more thorough exploration on what the premises are is essential for more effective mapping and design strategies in intelligent audio production tools.

We have approached this by doing extensive work on the best practices of mixing, which culminated in the first author's PhD thesis [3]. The text herein is a short summary of selected findings, highlighting those for which conclusions were strongly drawn. We are especially interested in conclusions that go against the stabilized assumptions found in previous research. This work resorts to approaches based on knowledge engineering (KE) [4], grounded theory (GT) [5] and machine learning (ML) [6].

KE seeks to integrate expert knowledge into computer systems for task solving that usually requires a high level of human expertise. For intelligent audio production we know the knowledge lies in the hands of top practitioners, but extracting it is not always trivial as practical sound engineering has moved away from a technical to an artistic field in the last half a century, and practitioners are often inclined to believe there is no knowledge implied. GT is a discipline that strives to systematically generate theory from data stemming from empirical research . For our case it means looking at complex psychoacoustic evaluation studies and extracting meaningful data out of listener preference. Finally, ML relies on the construction of systems that can learn from data. After a training step on a learning data set, the algorithm should be able to perform accurately on new examples.

In the original work [3] we have first relied on literature review and an extensive interview process to crystalize upon 88 potential assumptions for how technical decisions in mixing are performed. Figure 1 highlights some assumptions of this work. These suppositions were examined and eventually validated by one of seven different strategies:

1. Measuring parameters from mixing sessions of successful songs.
2. Having successful sound engineers perform specifically tailored mixing exercises.
3. Measuring features from completed successful mixes.
4. Performing subjective listening tests on experienced subjects.
5. Analyzing through quantitative surveys the habits of successful mixing engineers.
6. Performing exploratory interviews with successful mixing engineers.

#	Title	Proven	Origin	Tested
01	All signals should be presented with equal loudness.	False	PI	SE; Q
02	The main element should be up by an understandable amount of loudness units.	True	INT	EX; MM; SE; Q
03	Vocals should be ridden above the backing track	True	INT; LIT	EX; Q
04	No element should be able to mask any of the frequency content of the vocals.	True	INT; PI	Q
05	Track panning affects partial loudness	True	LIT	EX; SE
06	Dynamic Range Compression affects relative loudness choices.	False	INT	SE
07	Low-end frequencies should be centrally panned.	True	LIT; INT; PI	MM; SE
08	The main track is always panned centrally.	True	LIT; INT; PI	MM
09	Remaining tracks are panned out of the center.	True	LIT; INT	EX; MM; Q
10	The higher the frequency content the more a track can be panned sideways.	False	LIT; PI	MM
11	Frequency balance should be kept between left and right.	True	LIT; INT; PI	MM; Q
12	Hard panning should be avoided.	False	LIT; PI	SE ; Q
13	Sources recorded with close (mono) and far (stereo) techniques simultaneously should have the mono source panned to the same perceived position featured in the stereo source.	True	INT	Q
14	Monophonic compatibility should be kept.	True	LIT, INT	MM; Q
15	Panning is mostly done audience-perspective.	False	LIT	Q
16	It is customary to apply temporal cues to panning.	False	PI	Q
17	Equalization is frequently done to avoid inter-track masking effects.	True	LIT; INT; PI	EX; Q
18	Salient resonant frequencies should be subdued.	True	INT	Q
19	High-pass filters should be used in all tracks with no significant low-frequency content.	False	LIT; PI	SE; Q
20	There is a specific low-mid region that can be attenuated to improve clarity.	False	LIT	SE ; Q
21	Expert mixers tend to cut more than boost.	False	LIT	Q
22	High Q-factors should be used when cutting and low Q-factors when boosting.	True	LIT; INT	Q
23	Equalization use should always be minimized.	False	LIT	Q
24	Every song is unique in its spectral/timbral contour.	True	INT	MM; Q
25	Reverb time is strongly dependent on song tempo.	False	INT	SE ; Q
26	Reverb time is strongly dependent to an autocorrelation measure.	True	-	SE
27	Delay times are typically locked to song tempo.	True	LIT; INT	SE ; Q
28	The pre-delay is timed as a multiple of the subdivided song tempo.	True	LIT; INT	SE ; Q
29	The level of the reverb returns is on average set to a specific amount of loudness lower than the direct sound.	True	-	SE
30	Low-end frequencies are less tolerant of reverb and delay.	True	LIT; INT	EX; Q
31	Transients are less tolerant of reverb and delay.	True	LIT; INT	EX; Q
32	The sends into the reverbs should be equalized.	True	INT	Q
33	Reverbs can be carefully substituted by delays to lessen masking effects.	True	INT	SE; Q
34	Compression takes place whenever a source track varies too much in loudness.	True	LIT; INT	EX; SE; Q
35	Compression takes place whenever headroom is at stake, and the low-end is usually more critical.	True	INT	MM; EX; SE; Q
36	Gentle bus/mix compression helps blend things better.	True	LIT; INT	SE; Q
37	There is an optimal amount of compression in terms of dB and it depends on sound source features.	True	LIT	EX; Q
38	Compression should not be overused and there are maximum values for it.	False	LIT	EX; Q
39	? Compressor attack is set up so that only the transient goes through.	False	LIT	EX; Q
40	Compressor release is set up so that it is over when the next note is about to start.	False	LIT	EX; Q
41	It is acceptable to judiciously lop off some micro-burst transients to gain peak-to-RMS space.	True	-	SE ; Q
42	In deciding a tracks dynamic profile, an expert engineer will shift the focus of the listener by enhancing different tracks over time, with volume changes that may some times be quite big.	True	INT	EX; Q

Fig. 1: Selected assumption overview. The origin of the assumption can either be literature review (LIT), the interview process with professionals (INT), or the assumption made on previous implementations (PI). The method of testing is either through mixing exercises by professionals (EX), measuring number one hit singles for features (MM), subjective evaluation with a listening panel (SE) or a questionnaire sent to professionals (Q).

7. Using literature review.

This is a purposely ordered list, as each element yields more robust conclusions than those that succeed it. Options 1 and 2 grant us objective, quantifiable access to the workings of the mind of successful engineers performing successful mixes. Option 3 is equally robust, but may be tainted by mastering and conversion practices, and is limited in the scope of assumptions it can prove. Option 4 is not as objective in nature, but if performed on a large enough scale with experienced subjects, can give a good estimation of best practices. Option 5 and 6 introduce the problems of bias and status that arise from the sharing of private methodologies, 5 having the advantage of being quantifiable. Finally 7 is considered the less revealing because technical literature in mixing is scarce and written by authors that are not as successful in the craft of mixing as those in options 5 and 6.

In our research [3] we were able to have contributions from nearly 60 successful professional sound engineers (the criteria were having mixed number one albums or singles or having won a prestigious award for sound engineering), to build a panel of up to 70 listeners for subjective evaluation purposes, and to extract information from a dataset of over 900 songs that were number one singles in either the UK or the US. We performed over 20 subjective evaluation tests, had over 100 interviews, and examined a 49-question survey directed at the almost 60 experts.

We shall now explore significant conclusions on topics of loudness (Section 2), panning (Section 3), equalization (Section 4), temporal processing (Section 5) and dynamic range control (Section 6). We then move on to a broader overview of our conclusions, looking at potential areas for further work.

2. LOUDNESS

In terms of loudness balance it became clear that a perfect mix is not one where there is equal-loudness among tracks and yet every element should be perceived, contrary to previous belief [7, 8]. This is clearly depicted in the answers to the questionnaire illustrated in Figure 2. Instead, there is an order of importance, and we managed to quantify some aspects that are close to universal in deciding upon it. We later found that this should change through time, and that mixing engineers use the fact that listeners cannot process everything simultaneously to chose which signal is presented at the forefront at each instance in time, another indication that equal-loudness is not a concern, at least not a local one. The idea that the loudness choice of a signal was affected by its loudness range was also proven to be wrong. More stable signals were expected to endure a softer presentation, but that was not observed in subjective testing.

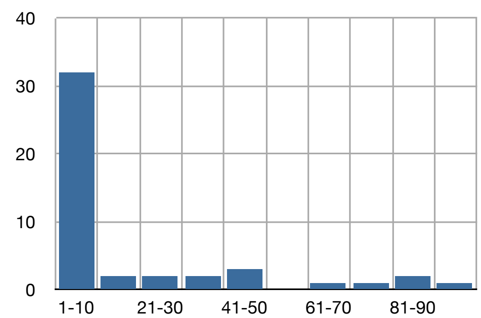

Fig. 2: Histogram showing accumulated answers to the question: "Over the course of your last 100 individual mixes, how often did all sound sources except for the vocals end up with nearly-equal loudness? It was posed to professional sound engineers who had at least one a grammy award or produced a number one album. It was carefully explained that it was not level but loudness that was to be judged. The results are clear and contrary to previous assumptions.

Whenever vocals are present, they seem to feature as the main element, and require a louder presentation. We have concluded that on average the vocals are equally loud to the sum of all other tracks, which means that individually they will be from 3 to 8 Loudness Units (LU) up from another element, depending on the number of total tracks [3]. Whichever value is chosen, it was quite evident from our findings that it is kept at a constant loudness differential to the backing elements, in what is usually termed *vocal riding*. This means that it is very important to create *a priori* a loudness profile that will be used as a blueprint for vocal placement. A system that does not implement a long-range loudness profile is bound to have vocal riding that acts as a dynamic range compressor and not an actual rider. The question of loudness of the main track is so crucial that we felt compelled to include another assumption on it, which was also considered to be true, stating that additional steps should be undertaken to ensure no frequency range of the vocal was masked. It was seen that this might not emerge from other assumptions, and should be catered to separately.

An idea that was all-pervasive throughout this topic was that of partial loudness [9], or how much the level of one track must be brought up in the presence of another to appear to be as loud as in its absence. Partial loudness relates to masking and is therefore stereophonic-position-dependent. Still, for all practical terms, it seems this concern's solution will emerge as a result of panning, equalization and loudness rules.

3. PANNING

Panning practices are possibly the most well-defined. It seems universal that low-frequency content is centered, along with the main track [10]. This typically results in a group of four elements, the back-bone, that are precisely centered. These are the kick drum and bass guitar (as a result of the low spectral centroid), the vocals (for their importance) and the snare drum (for reasons that will be seen below). In all songs where the track count is large enough, all other elements should be kept out of the center. This is probably related to best practices in terms of release from masking, even if sound engineers have reported masking problems as only the third most important reason to pan [3]. Track count is something that might be important to take into consideration, as atypical, sparse mixes will sometimes make the other rules discardable. One practice that seems universal, is that of keeping a balance of energy and energy per frequency band between left and right channels. The vast majority of commercially successful music of the last 60 years has strongly abided by this principle, and overall balance of left-to-right energy is shown in Figure 3.

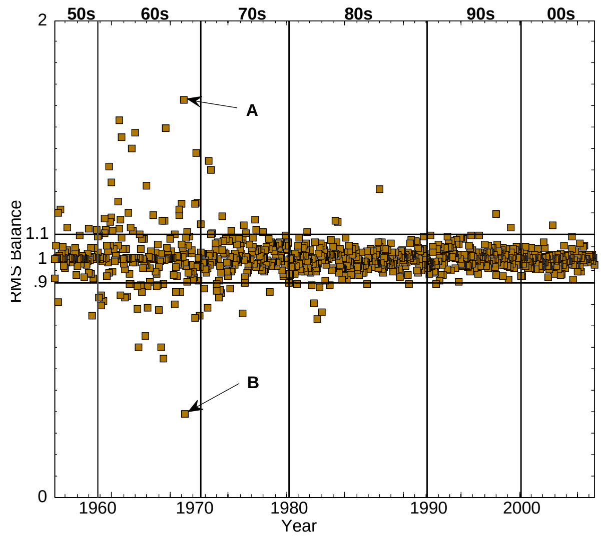

Fig. 3: Ratio of the left to the right channel's RMS value for the 928 commercial number ones in our dataset. There is a strong clustering of values around equal-energy per channel. Songs are sorted by date and decade division shown as vertical black lines. Other than the early mono years, and the following two experimental decades, difference between left and right lies consistently within .8 of a dB (the horizontal black lines).

Whenever there are point elements that are featured in overall microphone pairs (the typical situation being the individual drum element in both spot microphones and overheads), the monaural elements are placed to match the positions in the stereophonic general track. This is the reason for the snare drum being the fourth universally centered element. It speaks of a principle of verisimilitude that is more important for more natural styles. In jazz, for example, it is typical that the pan positions follows the placement of elements in a live stage, allowing for the disrespect of the centering of low-content or main element. One last thing we found to be true in terms of panning is that monaural compatibility should be checked.

Contrary to what has been believed in previous automixing implementations [11], we have established that wide panning is perfectly acceptable. Not only is it allowed, but it represents around one third of all panning decisions. Another assumption that was proven false was that the higher the frequency content the wider the pan position might be, a strategy that had been followed in [12]. It seems that other than low-frequency content (below 200 Hz) everything is open to wide panning. An aspect we have found to be irrelevant is the panning perspective of elements that have a clear left and right side. We proposed that audience perspective is more frequent, following most examples in literature (e.g. [13]), but it seems to be an almost perfectly split decision. The use of temporal panning strategies also constitutes a minority of cases. One topic we touched lightly because of its smaller user base is that of surround panning. We have found that, contrary to stereophonic situations, the best practices are very ill-defined, both in terms of perspective, center channel content and surround content.

4. EQUALIZATION

As for equalization and filtering, we seem to have started with most of the assumptions phrased wrong. Our misinformation came from literature and lore surrounding the known practices of the most revered sound engineers. Since we have surveyed some of these, we can only infer that such lore is in fact a myth. The practice of using high-pass filters on all tracks that have no bottom end content, which is very frequent in live-sound engineering for good reasons, is seldom used in studio mixing and bears no justification according to subjective testing. The same goes for any blind approach to the improvement of clarity through low-mid region cuts, as this practice depends on source content. Expert sound engineers do not tend to use subtractive equalization more than they do additive. The assumption that expert mixers are minimal in their use of equalization, particularly when boosting also does not find echo among our interviewees. An interesting area within equalization is that of whether there is a target frequency contour for a mix. Contrary to popular belief, we have found [14] there is, and it is similar to the shape of pink noise, with some additional roll-off in both top and bottom end. This proved to be subconscious, since most practicing mixing engineers will swear that they will mix towards a different contour every time. But when we talk of integrated spectral response there are clear boundaries, where a lack of one *dB* in the high-end can lead to a mix that is perceived as dull and unsuccessful.

On the other hand, we postulated that subtractive equalization should have higher quality factors than additive equalization, and that appears to be true.

We have found that the two major corrective purposes of equalization are the avoidance of spurious resonances and the unmasking of sound sources (see Figure 4). The former is an isolated concern, while the latter is the cross-adaptive ideal that must guide computer-assisted implementations. Its importance is so ubiquitous here that we suggest that most of the other assumptions in equalization may simply be its reduced version. The aforementioned low-cuts and low-mid cuts may well arise because of masking concerns. The same can be said of relating track count to the broadness of frequency range for each track. It is apparent that masking worsens with number of tracks, so one must start cutting frequency regions to make way for new elements. Dealing with kick drum and bass guitar separation in a very sensitive frequency region, is also a consequence of the need for *release from masking*, and it may be that, while true, one needs not impose any special constraint to enforce it.

5. TEMPORAL PROCESSING

Moving out of the relative simplicity of one or two parameter axes into the complexity of temporal and dynamic processing, our views in terms of assumptions must be limited by what are the fundamental matters, especially those useful for upcoming research. Linking reverberation time to song tempo failed to unravel a tight coupling. However, throughout our validation attempts it became clear that maybe tempo is not the real explaining factor, and we have actually seen that the signal's auto-correlation can be a stronger indication of the parametric choice, much like had been proposed by Ando [15], when exploring hall acoustics. The same does not happen in the case of delay time, where all data points towards the idea that coupling it to tempo is far more recommendable than failing to do so. The third tempo-related aspect is the reverb's pre-delay, which often appears in literature as a crucial parameter. Our results indicate that experts do not consciously time to tempo, though they might still do it subconsciously. Subjective testing suggests that there is a strong benefit in having the pre-delay value exceed $30-40$ *ms* and it might well be that timing to tempo immediately beyond that is the best option. However, as our test was discrete in terms of time differences, we could not tell if there was no time-uncoupled lower value that yielded better results.

We also found that the loudness of the reverb relative to the original source is related to tempo and sparseness, hovering around a value of 9 *LU* down from the dry track. This was confirmed with subjective testing and analysis of songs mixed by expert engineers, giving us what are curiously stable results, in the light of the fact that we could not extract useful information from the interview process. See Figure 5 for global test results.

We tried to understand how reverb style depended upon low-level or semantic features, but failed to come up with a strong causal link. There are some trivial remarks that can be made, such as the prevalence of plate reverb use on vocals, but none that is clearly quantitative. What we could prove is that low-frequency content simply does not get sent into artificial reverberation. It appears that percussive elements are less tolerant of reverberation level and time, but we have not found more than a weak indication of this. The content that does get reverberated is equalized going into or coming out of the unit. This is usually done with high-passing (at around 200 *Hz*) and low-passing (at around 5 *kHz*) and there is no consensus on whether it should be done on sends

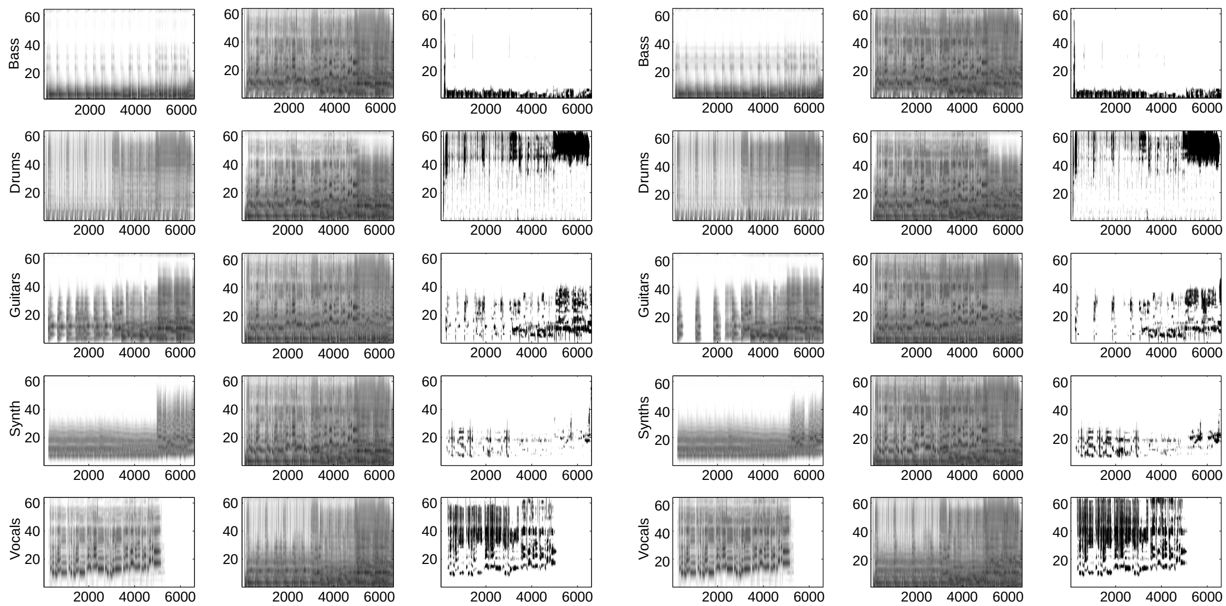

Fig. 4: Masking analysis of the equalization choices in a mixing exercise (one engineer's result shown, yet all illustrated the same trend), the left three columns are pre-EQ elements, and the remaining are post-EQ. Gammatonegrams of each element are shown full left, followed by the same depiction of the remaining instruments. Unmasked regions are shown in black. There is a clear increase in unmasked regions after EQ has been applied (see [3] for details on masking calculations).

or returns, even if it yields different results. Reverbs and delays are interchangeable in applications for 50% of our questioned experts, and subjective testing shows no preference of reverb over a well-crafted delay as a space-enhancing strategy.

6. DYNAMICS

We tried to ascertain what is the most important reason to compress dynamic range, which would lead us to the important mapping parameter to be considered when deciding compression amount. We have seen that the two main technical reasons to compress are erratic loudness ranges and control of low-frequency content. Even though experts point to the former as a more typical reason to compress, the latter seems to yield better results in subjective testing, and also seems to resonate better with the decision made by experts on actual compression exercises. This must mean that either we could not identify the true meaning of 'erratic', or the engineers have a misconception about their practice. We have suggested that automatic compression due to fluctuating loudness should only be performed in extreme cases, and that the best mapping parameter is related to frequency content. This is contrary to the approach taken in [16].

It was apparent that overall mix compression is a best practice, typically near the 3 dB mark[1], but there is no indication that there is a maximum value for the amount of compression professionals may use. As far as differences in compression amount being due to specific features, we have found that it probably relates more to instrument type that to features *per se*, except for a frequency-dependency that can be embodied by a modified version of the spectral centroid [3]. We have suggested that the most robust strategy is to perform instrument identification and rely on a fixed compression table.

We have tried our best to understand compressor attack and release times, but there seems to be no expert consensus on how these should be set, and how to relate them to sound features. Through analysis we have come up with some peculiar proposals (e.g. a relationship to the 4^{th}

[1] The amount of compression is not a parameter in itself but the result of several parameter interactions. It is common choice to think only of the end result, though.

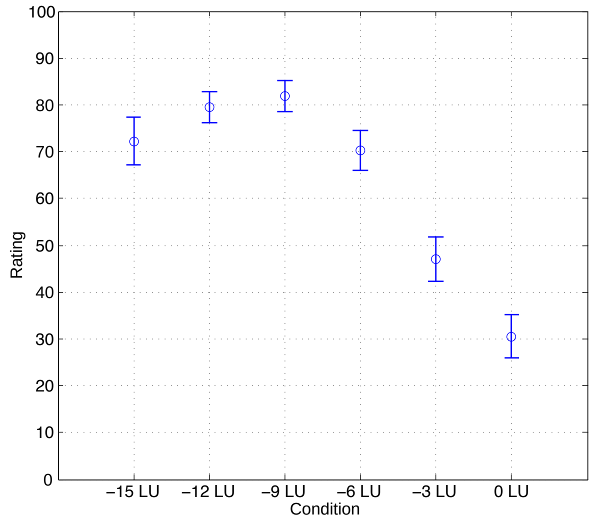

Fig. 5: Close to thirty evaluators in a panel took a multi-stimulus test where they were given six versions of six songs to rate. The versions differed in the loudness ratio of the reverberation returns to the original sources, shown as the x-axis here. The averaged evaluations show there is a peak around $-9LU$ down, and that subjects truly dislike excessive reverberation in a popular music context.

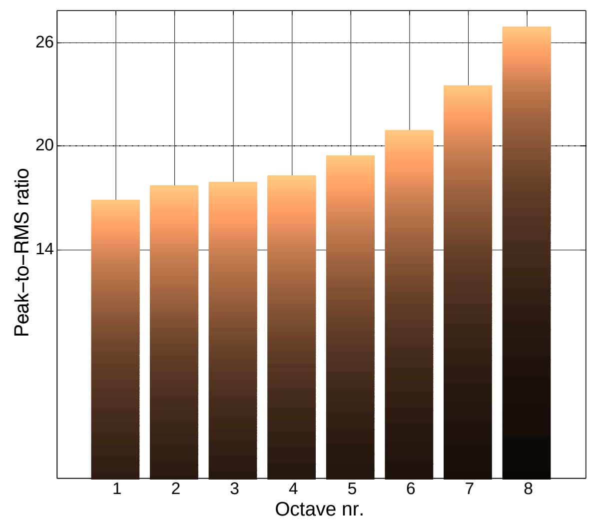

Fig. 6: Peak-to-RMS ratio per octave band for the 928 commercial number ones in our dataset. We analyzed the result against non-mixed sound sources, and it indicates that the trend of having lower dynamic range on the lower octaves is a consequence of mixing. Professionals therefore seem to apply more compression to low-frequency sources.

MFCC) that certainly will elicit more research. Allowing for peak-limiting seems to be the norm and we have seen that generally 6 dB cuts are unnoticeable to knowledgeable listeners. We suggest a more thorough test that relates this value to the duration of a transient, even though we cannot see implementation strategies that will depend on this level of detail. Multi-band compression was found to be residual in use, but mix bus compression is the norm.

We have dealt with the question of parameter change through time, in what we propose calling *bird's-eye dynamics*[2]. It was generally found very important to have dynamic richness and to feature relevant elements at relevant times. We were able to understand that level automation was by far the most frequent form of long-term temporal manipulation, and it can be easily argued that panning and equalization automation are bound by artistic criteria. On the other hand, allowing for rich temporal change of these parameters, constrained by unmasking concerns, can yield results that go well beyond the possibilities of human mixing, and is an area ripe for research. What we could not find, and is ultimately very relevant, are the time constants for permissible change. It is clear that changes that are too rapid will produce artifacts, distractions or even physical unrest. It is also clear that too slow a change will miss the opportunity of unmasking to its full potential. The acceptable value is very elusive, and we could not gather any clues from our validation strategies. We propose that this can only be verified with subjective testing of already implemented full-systems.

7. CONCLUSION

There are two overall concepts that stand out. The first one is that mixing is inevitably linked to the problem of masking, which was been pervasive through almost a quarter of the assumptions we examined. This might even be more important in computer-assisted systems, where the thoroughness with which one can deal with masking is virtually infinite. The second highlighted concept is that all integrated automatic mixing systems

[2] The idea of micro- and macro-dynamics are already in standard use for the ITU/EBU loudness measures [17, 18]. The definition of macro-dynamics for those standards is much shorter than what we need to describe the overall structural changes (e.g. chorus is louder than the verse).

must operate with hindsight. This means that while there is a set order of events to make decisions upon, those decisions should not be implemented until the aggregate result is known, and redundant calculations should not be allowed, something that separates computer-approaches from the human, iterative approach.

It seems the strongest and most novel conclusions among all the specific assumptions are:

- It appears that there is a long-term target equalization contour for a mix.
- There is a quantifiable way of implementing loudness balance, including reverb return loudness.
- Compression choices are strongly frequency-dependent.
- Mix compression and peak limiting are widespread practices, even among conservative sound engineers.
- Some assumptions that have been considered obvious so far have been proven false. These include the use of high pass filters and the idea that the higher the frequency content, the wider the panning, among others.

Though we have focused on unmasking, there may be an antonymic approach. The traditional suggestion is that when the time-frequency content starts to get cluttered, one tries to pull things out of the way of each other, either with equalization or panning. The reverse approach also has its proponents. They claim that when things get crowded, one should start doubling parts to get them thicker, pan them together and build more massive construction blocks. This is the Motown/Phil Spector wall-of-sound school of thought and fights masking through texture. While this was proposed as an alternative in the masking assumptions, we still could not gather enough evidence to place it in isolation, but note that this could be a future direction of research.

Amidst our discussions arose many hints for future work. As for broad areas of interest, automatic reverberation is clearly the most unexplored field. Automatic dynamic range compression may be close on the horizon as a viable solution, but only as far as a cornerstone goes. And questions related to integration of a multi-function system are also a region of much potential work.

We have paved the way for future informed implementations, and hopefully opened a discussion on a practice that has so far been too hermetic, and lacking in research. The numerous expedients we have relied upon to give us understanding on the validity of our assumptions have made this a very thorough and rewarding work, and we hope to have struck a balance between completeness, and the opening of new windows of opportunity for curiosity and exploration.

8. ACKNOWLEDGMENT

The author P.D. Pestana was sponsored by national funds through the Fundação para a Ciência e a Tecnologia, Portugal, grant number SFRH/BD/65306/2009, and projects:
"PEst-OE/EAT/UI0622/2011"
"PEst-OE/MAT/UI2006/2011"
Support was also provided by EPSRC grant EP/K007491/1, 'Multisource audio-visual production from user-generated content".

9. REFERENCES

[1] J. D. Reiss, "Intelligent Systems for Mixing Multichannel Audio," in *17th Intl Conf on Digital Signal Processing (DSP)*, 2011.

[2] V. Verfaille and U. Zölzer, "Adaptive Digital Audio Effects (A-DAFx): A New Class of Sound Transformations.," *IEEE Transactions on Audio, Speech, and Language.*, vol. 14, no. 5, pp. 1817 – 1831, 2006.

[3] P. D. Pestana, *Automatic Mixing Systems Using Adaptive Digital Audio Effects*. Phd, Universidade Católica Portuguesa, 2013.

[4] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods Title," *Data and knowledge engineering*, vol. 25, no. 1, pp. 161–197, 1998.

[5] S. Looso, R. Borner, and M. Goeken, "Using grounded theory for method engineering," in *Fifth International Conference on Research Challenges in Information Science (RCIS)*, IEEE, 2011.

[6] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. New York: Springer, first ed., 2006.

[7] E. P. Gonzalez and J. D. Reiss, "Automatic Gain and Fader Control For Live Mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009.

[8] S. Mansbridge, S. Finn, and J. D. Reiss, "Implementation and Evaluation of Autonomous Multi-track Fader Control," in *Proceedings of the 132nd AES Convention*, (Budapest), Audio Engineering Society, 2012.

[9] B. R. Glasberg and B. C. Moore, "A Model of Loudness Applicable to Time-Varying Sounds," *Journal of the Audio Engineering Society*, vol. 50, no. 5, pp. 331–342, 2002.

[10] R. Izhaki, *Mixing Audio — Concepts, Practices and Tools*. Oxford: Elsevier Science & Technology, first ed., 2008.

[11] E. P. Gonzalez and J. D. Reiss, "A Real-Time Semi-autonomous Audio Panning System for Music Mixing," *EURASIP Journal on Advances in Signal Processing*, 2010.

[12] S. Mansbridge, S. Finn, and J. D. Reiss, "An Autonomous System for Multi-track Stereo Pan Positioning," in *Proc of the 133rd AES Convention*, 2012.

[13] B. Owsinski, *The Mixing Engineer's Handbook*. Vallejo, CA: Mix Books, first ed., 1999.

[14] P. D. Pestana, Z. Ma, J. D. Reiss, D. Black, and A. Barbosa, "Spectral Characteristics of Popular Commercial Recordings 1950-2010," in *Proceedings of the 135th Audio Engineering Society Convention*, vol. submitted, Audio Engineering Society, 2013.

[15] Y. Ando, *Concert Hall Acoustics*. Berlin: Springer Verlag, first ed., 1990.

[16] J. A. Maddams, S. Finn, and J. D. Reiss, "An Autonomous Method for Multi-track Dynamic Range Compression," in *Proceedings of the 15th International Conference on Digital Audio Effects*, (York), DAFx, 2012.

[17] ITU, "ITU-R-BS.1770: Algorithms to Measure Audio Programme Loudness and True-peak Audio Level BS Series," tech. rep., International Telecommunications Union, Geneva, 2006.

[18] EBU, "Tech Doc 3342 "Loudness Range: A Measure to Supplement Loudness Normalisation in Accordance with EBU R 128"," Tech. Rep. August, European Broadcast Union, Geneva, 2010.