Queen Mary
University of London

User Preference on Artificial Reverberation and Delay Time Parameters

REISS, J

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/19480

User Preference on Artificial Reverberation and Delay Time Parameters*

PEDRO D. PESTANA, *AES Member*   **JOSHUA D. REISS,** *AES Member*   **AND**   **ÁLVARO BARBOSA**

(ppestana@porto.ucp.pt)   (joshua.reiss@qmul.ac.uk)   (abarbosa@usj.edu.mo)

Universidade Católica Portuguesa, CITAR   C4DM, Queen Mary University of London   U. Saint Joseph, Macau

It is a common belief that settings of artificial reverb and delay time in music production are strongly linked to musical tempo and related factors. But this relationship, if in existence, is not yet understood. We present the results of two subjective tests that evaluate user preference of young adults with formal training in audio engineering on artificial reverb and delay time, while trying to relate choice to tempo and other low-level explaining factors. Results show there is a conclusive relationship between musical tempo and delay time preference as described by users. Reverb time setting preference, however, cannot be explained in such a strong manner. In this latter aspect the present work has nevertheless uncovered some ideas on how to proceed in order to quantify the phenomenon.

0 INTRODUCTION

The current work consists of the analysis of two subjective tests, performed with knowledgable practitioners, which strive to explain the relationship between the choice for the time parameter in artificial temporal processing units, and the underlying musical content. Specifically, we hypothesise, following technical literature [1, 2], that there is a relationship between a song's musical tempo and the definition of artificial reverb and delay times.

A delay, or echo, consists of a discrete repetition of the signal after a given period of time. This repetition can be individual or can have sequels, which are frequently (but not necessarily) evenly spaced in time. Below a delay time of about 30 *ms*, the human ear does not perceive a repetition, and it integrates both dry and delayed sounds, which means we will consider that the processing we call 'delay' to consist of times that are greater than this interval.

Artificial reverberation is a process that strives to emulate and complement the real phenomena of room reverberation. The physical manifestation of this effect depends upon the numerous reflections that spring from the room's boundaries creating a series of differently timed echoes that blend into a tail that will prolong the sound. It is typical to distinguish between early reflections (sparse and colouring) and reverberant sound (dense and statistically uncorrelated). It is reverberation that offers the sonic footprint that enables one to identify the sound of a room. One crucial parameter is the Reverberation Time (RT_{60}), which for historical reasons is given as the time it takes for the tail to decay 60 *dB* after the original sound has ceased to exist.

In the following Section we will contextualize the current work, looking also into the reasons and possible applications, while highlighting its differences to previous approaches. In Section 2 we discuss the subjective test methodology and statistical approach which was common to both tests. In Section 3 we present and interpret the results and further comments by test subjects, leading to some post-hoc analysis in Section 4. Some tentative conclusions are drawn in Section 5, along indications for future work.

1 PRIOR WORK AND MOTIVATION

In [3], 60 successful practicing sound engineers were interviewed and no conscious method for regulating artificial delay and reverb parameters was suggested, other than the idea that slower tempos lead to longer reverbs, and that there is a stylistic aspect to choice. In the current work we change the approach from producer to listener, looking for patterns in subjective preference of time-constant setting in delay and reverb. User preference should be content-, context- and epoch- dependent, and the test does not propose to seek rigid correlations, merely trends that can lead to further investigation.

We acknowledge that extensive work has been done on listener preference regarding acoustic spaces [4, 5, 6] (a particularly good overview in [7]), but highlight the fact

---

*To whom correspondence should be addressed: P. D. Pestana, CITAR - Catholic University of Portugal, Rua Diogo Botelho, 1327, 4169-005 Oporto. Phone: +351226196200. Mail: ppestana@porto.ucp.pt

that setting artificial time constants in production is a related but very different problem. On the one hand the acoustic approach always starts from space - given a specific space, how does the music fit, while the production approach always starts with music - given a specific song, how to set the parameters. There is also a strong difference between perception of spatial cues in three-dimensional spaces such as live halls, and through a stereophonic setup where all cues collapse to two points in space [8].

A note as to the terminology in this work: we are looking for relationships between musical tempo and processing time constants as set by users. While in terms of delay processing the idea of a link is immediately as the repetition falls within the sense of musical beat, in terms of reverb it is not, as the vague idea of "decaying 60 dB" does not imply that a decay time equaling the quarter note would mean the listener would have ceased to hear the reverberating field by the time the next beat would settle in. This does not preclude from searching for a relationship, as some author's have indicated there may be one [2, 3]. We have chosen the term "coupled" to refer to a relationship where the time constant would imply a subdivision of the beat for brevity sake, knowing there might not be a consensual term here.

Trend knowledge from the subjective tests presented herein can be useful in automatic settings of reverb time by assisting mixing systems; adaptive parametrisation of presets in artificial reverberation plug-ins; or any adaptive/automatic system that may have access to acoustic cues of the listening environment in order to remix audio content in real-time.

2 METHODOLOGY

The tests were performed by a pool of experienced listeners from three different training centres: the School of Arts of the Catholic University of Oporto, the Communication Science department at the Lusíada University in Lisbon, and the professional audio school Restart, also in Lisbon. Both students and faculty members volunteered to collaborate, and an isolated room in each institution was chosen for the testing apparatus. Due to logistic concerns there was no formal pre-screening session, but the first runs of each test included redundancy testing that was used as a post-pre-screening method, in that examples were repeated and user consistency checked, leading to rejection of subjects that did not perform well in repeated tests. With this process around 10% of the participating subjects were rejected and an unofficial listening panel began to emerge. An identified problem is the fact that ours is a convenience sample, something that usually afflicts audio testing, and we can only suggest future replications of the test in different settings, so that a meta-analysis can be used to synthesise results.

Test design and procedures followed closely the recommendations in [9]: individual duration of a test was targeted at under 20 minutes; the subjects were well informed on the test procedure, and exploratory interviews were made at the end. Song excerpts were kept under 30 seconds, the

listener being able to listen as often as needed, and was allowed to answer at any time.

Tests were performed with professional grade circumaural earphones (Sennheiser HD650), previously calibrated with a dummy head. The signal chain was consistent and the listening level stable across the procedures. Works such as [10] have confirmed that relative level setting, for example, is different over headphones and loudspeakers, with no consistent tendency as to sign or magnitude of difference. Our choice was mainly related to the necessity of running test at three facilities and maintaining consistency, and repeatability in further tests.

The choice of presentation level is essential not only for repeatable results, but to keep the influence of level out of some perceptual attributes, as defined in [11]. Most recommendations oscillate between levels of 83 and 85 $dB(A)$ SPL [12], and we opted for the lower figure for all tests that did not deal with loudness control themselves.

As a music production problem, we strived to reduce ambiguity to the minimum: all songs presented were recorded on acoustically dead rooms ($RT_{60} < 0.15s$), so the amount of temporal processing artificially introduced is much larger the the recorded acoustic footprint. This is similar to the majority of contemporary recording practice approaches [1, 2, 13].

The two critical tests presented here followed a multi-stimulus approach, where subject j is allowed to classify condition k of song i on a quality preference scale ranging from 0 to 100. The different conditions typically vary exclusively in one parameter, and subjects are simply asked to rate based on either quality or clarity, with no indication about the aspect they are differentiating, except where noted.

All tests were double blind, with completely randomised full-factorial design, where song order and condition order varied arbitrarily between subjects (who were informed of this). To minimize end point effect and subjective use of scaling, it was suggested that respondents should screen all conditions for the best one on a first pass, and rate it at 100, so that all others could be rated relatively to it, never exceeding the scale. Test instructions were pre-screened with colleagues and presented verbally before the test and a very brief sentence describing the task was featured on the test interface. The instructions were always straightforward and the procedure was well discussed with the subjects. Whenever doubts remained, a training run was performed. A mandatory training procedure was not part of test design as it was seen in pilot tests that accuracy was stable from the beginning, and would have declined after a period, due to listening fatigue, should the procedure have been made longer. We find that this is peculiar to the type of comparison proposed.

The test interfaces were custom-built in the software Max/MSP, and an example is shown in Figure 1. The question was posed as *"Please evaluate according to the quality of the mix"* on a scale of 0 to 100.
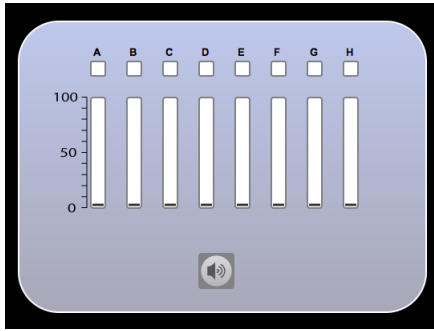
Fig. 1. Multi-stimulus test interface for the second test. Each test has a label box (removed) reminding of the proposed task.

The evaluation of condition k of song i by subject j[1] typically leads us to a multi-variate matrix that can be analysed on a per-song basis, but also on a per-subject base, or, if song differences prove to be irrelevant, our matrix collapses to two dimensions, by bundling every observation of the same condition together.

An overview of characteristics for each test is presented in Table 1. The three first rows are self-explanatory. Six songs enabled us to keep the total test durations under the estimated 20 minutes for most subjects.

Test	Reverb	Delay
Number of conditions	8	6
Number of subjects	20	20
Number of songs	6	6
$\bar{A}_k$ (condition)	9.63	10.38
$\bar{A}_j$ (subject)	13.67	18.89
$\bar{A}_i$ (song)	8.67	10.38
Subjects non-randomness (Friedman)	86%	100%
Mean Spearman ρ	0.33	0.41
Mean answer time (s)	18:29	17:11
Mean answer time per song (s)	03:05	02:52
Mean subject age (yrs)	26.9	25.3
Mean subject experience (yrs)	5.1	5.4
Mean identification difficulty	49.2	1.8
Mean judgement difficulty	78.6	12.4

Table 1. Overall characteristics of the two multi-stimulus tests.

The three values A_k, A_j and A_i are mean range indicators. Let the range of the confidence interval for condition k, integrating all songs and subjects be given by:

$$A_k = 2 \times t_{0.975} \times s_k / \sqrt{IJ}, \tag{1}$$

with the standard deviation s_k. The mean range per condition is then simply:

$$\bar{A}_k = \frac{1}{K} \sum_{k=0}^{K} A_k, \tag{2}$$

something that can be easily extended to i and j and give us some idea on how wide the confidence intervals typically are within each independent variable.

$\bar{A}_k$, $\bar{A}_j$ and $\bar{A}_i$ show how consistently conditions were evaluated by subject/song pairs, how closely subjects agreed between themselves for song/condition pairs and how close songs were to each other in subject/condition pairs, respectively. We had designed our sampling strategy for an A_k of around ± 5, given 25 subjects, which is roughly achieved.

To test for non-randomness, we used the Friedman test [14]. This is a rank test, where we first order each subjects ranking of each condition on a per song basis, turning $\{x_{jk}\}_{J \times K} \to \{r_{jk}\}_{J \times K}$. We then calculate the rank sum for each evaluator: $R_k = \sum_{j=1}^{J} r_{kj}, k = 1, 2, ..., K$. For cases where there are no equal ranks (which happened to account for all our cases, not by design but by serendipity), we can simplify the test statistic as:

$$Q = \frac{12}{JK(K+1)} \sum_{k=1}^{K} R_k^2 - 3J(K+1). \tag{3}$$

As our values for K and J are large enough, this can be approximated by a chi-squared distribution with p-value given by $P\left(\chi_{K-1}^2 \geq Q\right)$. For p-values below the typical significance level of 0.05 we reject the null hypothesis, H_0: Subject's judgements are arbitrarily attributed.

We have also followed the suggestion in [15] to check the correlation of each subject to the average subject and red flag low scores. This has been done with Spearman's Rho, which, in the case of no rank ties, is given by:

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_j^2}{n(n^2 - 1)}, \tag{4}$$

where d_j is the rank difference between the jth subject and what we call the 'Typical independent (TI) subject'[2]

In Table 1, subject's non-randomness and average Spearman's ρ follow the discussions above, and are indicators of how reliable were the subject's judgements. For each test we performed $I + 1$ Friedman tests, one for the judgement of each song, and one for the overall mean judgements across songs. The percentage indicated here is how many tests rejected the null hypothesis that evaluations were random. The mean Spearman ρ is the average of correlations between each subject and the TI-subject.

The last two rows of Table 1 pertain to subjects being asked how easy it was to identify the differentiation parameter in the test, and how easy was it to judge those differences. The mean score for those questions is presented, 0 being easiest and 100 hardest.

---

[1] We shall consider K conditions, I songs, and J subjects in total.

[2] The typical independent subject is calculated by averaging the observations across subjects, excluding the subject that is being analysed prior to ranking. This allows preservation of independence of the two sets of ranks being compared.

3 TEST CONDITIONS AND RESULTS

3.1 Delay Time Preferences

For the setting of delay times it is customary to lock precisely to tempo, to the extent of many mix engineers using delay charts or calculators [1, 2]. It is interesting to confirm this connection with blind subjective evaluation, and for this we used six songs, which, ordered by tempo are at a speed of 67, 84, 105, 120, 143 and 180 bpm. We compared the following conditions:

- Condition α: Completely dry, unprocessed mix of all tracks, performed by a mixing engineer.
- Condition β: As in α, but the vocal is sent through a delay unit, set to a quarter-note, with 33% feedback, 80/20 dry/wet level. These two parameters are kept constant through all remaining examples.
- Condition γ: As in α, but the vocal is sent through a delay unit, set to an eighth-note.
- Condition δ: As in α, but the vocal is sent through a delay unit, set to a half-note.
- Condition π: The vocal delay time is now uncoupled to tempo and set to a quarter-note multiplied by $\pi/2$.
- Condition ε: The vocal delay time is now uncoupled to tempo and set to four fifths of a quarter-note.

The values are detailed in Table 2, which includes additional values that were used for the second test below. It is also important to notice that songs were of different genres, song 1 a slow pop ballad, song 2 groove rock from the 70s, song 3 is smooth jazz, song 4 funk/rock crossover, song 5 is uptempo classic rock, and song 6 electronic synthpop.

Song #	1	2	3	4	5	6
bpm	67	84	105	120	143	180
beat (β)	895.5	714.3	571.4	500	419.6	571.4
1/2 (γ)	447.8	357.2	285.7	250	209.8	166.7
dbl (δ)	1791	1428.6	1142.8	1000	839.2	666.6
ϕ	1448.9	1155.7	924.5	809	678.9	539.3
4/5 (ε)	716.4	571.44	457.12	400	335.68	266.6
π	1406.6	1122	897.6	785.4	659.1	523.5

Table 2. Characteristics of each song for both tests presented herein. The tempo influences the several subdivisions durations.

An overview of the overall results is given in Figure 2, where conditions are ordered by increasing delay time. There is a clear separation between times that are coupled and times that are not, and the former are clearly preferred. It seems that no delay on the vocals is subjectively better than a delay that is off-subdivision. Considering only beat-coupled settings, there is also an evident preference for faster subdivisions. Table 2 can be consulted for the absolute meaning of the preferred eight-note subdivision: it lies somewhere between 167 and 448 *ms*, slower than typical slap-back echoes, suggesting we could have even tried for faster subdivisions.

How relevant is song content in this scenario? Figure 3 shows us the inter-song relationships for each condition. The slower song scores highly in all conditions where delay is present, but it is considered the worse to be left dry.
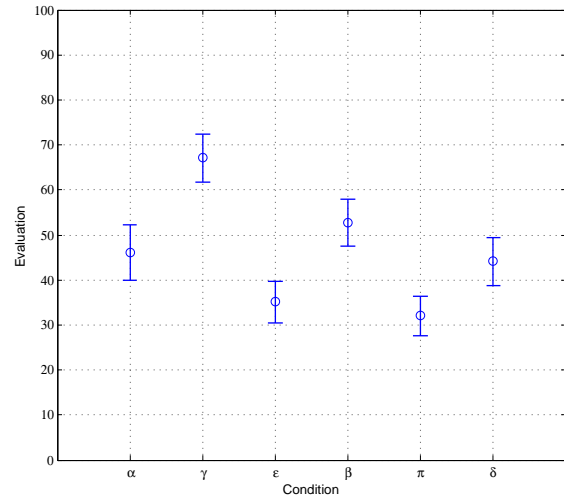

Fig. 2. Mean and confidence intervals for the evaluation of each condition, considering inter-song differences to be irrelevant. Condition order is changed so that left-to-right corresponds to decreasing tempi.

Condition γ is very homogenous between songs, and it is always rated the best condition, except for song one, where timing to the beat (β) is rated higher.
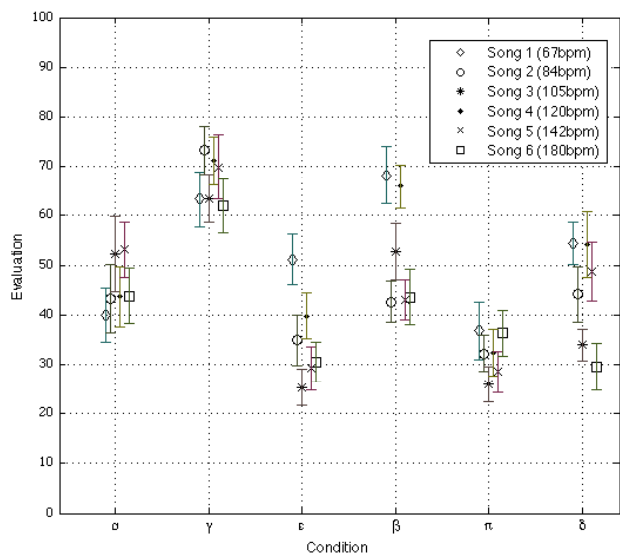

Fig. 3. Mean and confidence intervals for the evaluation of each condition, divided by songs.

Subjects are fairly balanced in their replies, as can be seen in Figure 4. There are no clusters of subjects, and an analysis of the correlation between subjects and the average subject, shows that only 7 and 19 deviate from the norm, and this is simply because subject 7 prefers double (δ) to eighth (γ) and 19 prefers quarter (β).

The results on the whole seem rather clear and robust, the Friedman test scores given in Table 3 indicate that evaluations are by far not random.

Most subjects found the test simple and the differentiation parameter readily understandable. Two subjects went so far as to state they were aware they had chosen first the
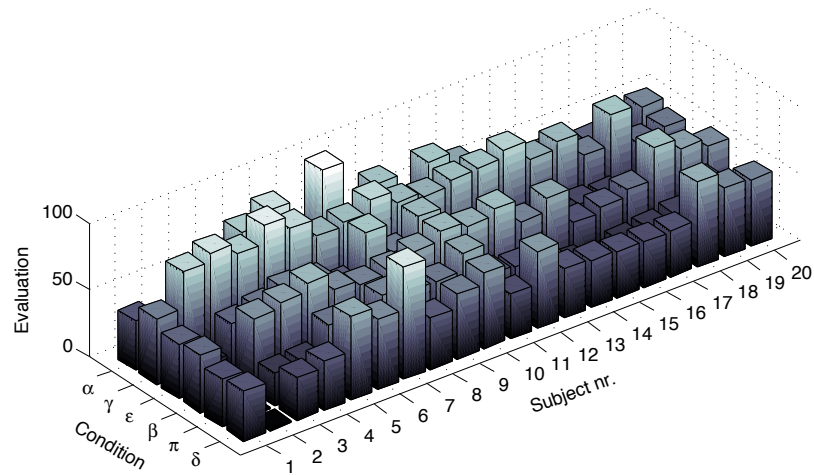
Fig. 4. Bar charts of the mean evaluation of each condition by each subject, averaged over the six songs.

	Reverb		Delay	
	Test Stat	*p*-value	Test Stat	*p*-value
Song 1	1.85	0.96	16.22	0.006
Song 2	40.91	< 0.001	29.9	< 0.001
Song 3	18.89	0.009	32.37	< 0.001
Song 4	34.59	< 0.001	21.85	< 0.001
Song 5	20.07	0.005	18.72	0.002
Song 6	47.1	< 0.001	17.8	0.003
All data	39.2	< 0.001	48.83	< 0.001

Table 3. Results of the Friedman test for randomness of evaluations.

short delay, followed by the dry version, followed by the beat version. Subject 6 mentioned feeling that it depends on genre — for rock the choice would be on short delays, for the jazzy theme, a quarter-note delay sounded well. Subjects 3 and 4, both the most experienced subjects, also reported that it was crucial how the delay time would fit with the phrasing of the melodic line, but for the evaluated pieces, their judgements were pretty standard and consistent between songs. Subject 9 states that echoes are annoying in being reminiscent of dated production values — this type of observation could result in a marked bi-polarisation of opinions, but it did not, and subject 9 is the only case where this approach is reflected on the results.

3.2 Reverberation Time Preferences

Again, we presented subjects with eight different conditions of timing reverb decay to tempo, over a range of the same six songs of different tempos.

Reverb decay time cannot be disassociated from reverb level, as the ability to hear the tail very much depends on the loudness of the reverberant field in short musical gaps. We ran pilot studies to have a rough idea of preference so that we could lock one parameter at a comfortable level while varying the other. The reverb loudness was thus set so that it was 9 *LoudnessUnits* (LU) lower than the direct sound, and it was applied equally across all elements except for kick drum, bass guitar and overheads. The reverb unit was a TC Electronic 4000 with a hall algorithm,

no pre-delay, and all other settings left as in preset 1. One should refer back to Table 2, which indicates song characteristics and decay times for each of the conditions and each of the songs. The conditions used in this test instance are:

- Condition α: all tracks dry.
- Condition β: specified tracks sent to the unit with the decay time set to the beat (quarter-note) of the musical tempo.
- Condition γ: specified tracks sent to the unit with the decay time set to half the beat (8th-note) of the musical tempo.
- Condition δ: specified tracks sent to the unit with the decay time set to double the beat (half-note) of the musical tempo.
- Condition ϕ: specified tracks sent to the unit with the decay time set to a quarter-note multiplied by the golden ratio. This was suggested by a renowned engineer (in [3]) as being a personal approach, and was one of the most quantitative responses we had, so we wanted to test for it. We are, however, unsure of whether we should consider this to be a coupling to tempo or not.
- Condition ε: decay time uncoupled with musical tempo, by subdividing each beat in five different parts and choosing the time it takes to complete four of these.
- Condition π: decay time uncoupled with musical tempo, by multiplying the beat by $\pi/2$, clearly an irrational measure. This is very close to ϕ in practice.
- Condition ζ: decay time set to 2 seconds. This is typical of the best music halls in the world for classical reproduction [4]: the Amsterdam Concertgebouw, Vienna Großer Musikvereinssaal and Boston Symphony Hall. We are aware it is considered good for orchestral music, but not for the pop/rock styles we are evaluating.

Overall results are shown in Figure 5 illustrating that other than subjects clearly disliking the dry option (α), and moderately disliking the long, two-second option (ζ), no

substantial differences emerge. Condition π neutralizes a very soft bell-shape curve, which could indicate that there is an optimal tempo:decay relationship unattached to beat-coupling. Furthermore, options β, γ and δ, the coupled ones, show no clear advantage over the uncoupled ε and π.

The figure shows a smooth arching trend, which is only marred by condition π. Even considering that the the homogeneity of the mean and confidence intervals is too high to make bold statements, it is tempting to suggest that the optimal decay time lies between the quarter and the half-note, but the π condition's lower status means some degree of coupling to the beat is preferred.
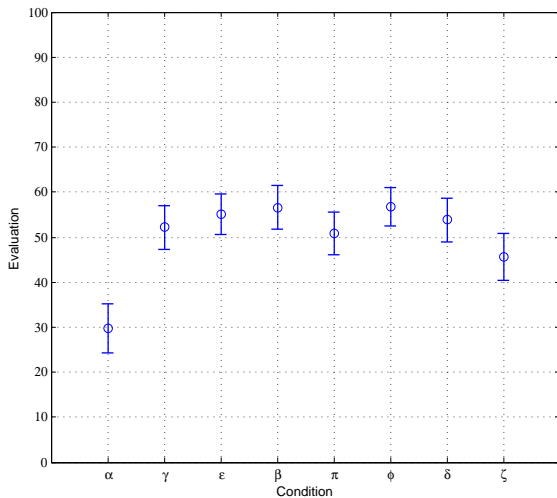

Fig. 5. Mean and confidence intervals for the evaluation of each condition, assuming no inter-song difference. The conditions are ordered by increasing decay time.

It is tempting to imagine that inter-song and inter-subject differences could result in this blurred overall picture, but our analysis shows it is not so: both subjects and songs show the same indistinct behavior. One interesting conclusion that emerges from Figure 6 is that the evaluation of fixed condition ζ decreases monotonically with increasing tempo. This is a clear reinforcement of the initial assertion that slow songs allow for longer reverbs. Another interesting observation from the figure is that the evaluation of the dry version's quality is markedly higher for songs 1, 3 and 5, than for songs 2, 4 and 6. This is perhaps surprisingly not related to tempo, but looking at Table 2, a pattern emerges — the songs that work well dry are those with more syncopation, as opposed to the three straight, strong-pulsed songs.

The Friedman test report on Table 3 also raises a curious question: why is song 1 the only one rated arbitrarily, particularly considering how low the remaining p-values were?

Looking at the raw subject data, we confirm that there is a clear disagreement between subjects on how to evaluate the different conditions in the case of song one. For example, subject 1 only seems to care for the quarter-note decay time, whereas subject 20 has a strong preference for an uncoupled decay and subject 7 for a half-note decay. These
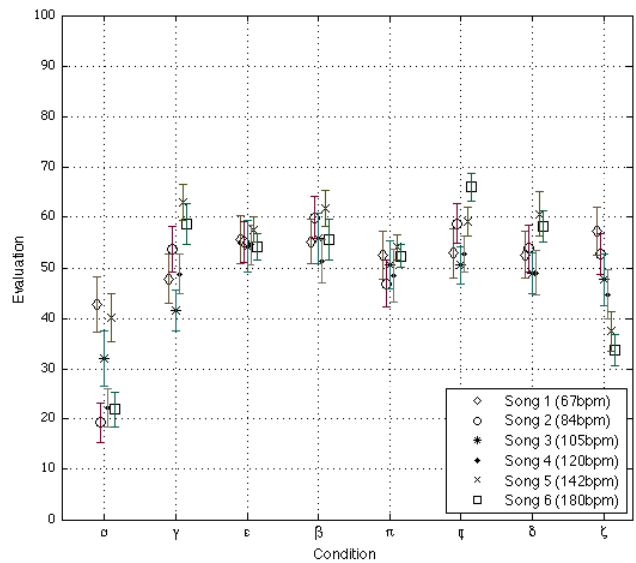

Fig. 6. Mean and confidence intervals for the evaluation of each condition, separated by song.

are polarised cases, but most subjects are more blurred in their judgements. This difference is emphasised on Table 4 showing the Spearman's rank correlation between each subject's judgements and the TI-subject for song 1. This dramatic variation in preference is not seen on the other songs.

#	ρ	#	ρ	#	ρ	#	ρ
1	0.109	6	-0.539	11	-0.263	16	0.359
2	-0.481	7	0.356	12	-0.196	17	-0.204
3	0.024	8	0.06	13	-0.325	18	-0.192
4	0.738	9	-0.069	14	0.738	19	0.738
5	-0.412	10	0.667	15	-0.096	20	0.364

Table 4. Correlation between each subject's evaluation and the average subject's evaluation of the conditions for song 1.

The open interview at the end of the test revealed the confusion on the parameter that was being differentiated. Even though 15 subjects were right in that it was reverberation, they were not confident on which aspect of reverberation it was. The remaining 5 subjects thought timbre and equalisation, reinforcing the idea that single-parameter testing with complex material can lead to much confounding. Most subjects stated that there was one condition that was radically different from the rest, and from the results, we would say that it was the dry condition α.

4 POST HOC ANALYSIS

For the reverberation case, there might be better criteria than tempo, and we followed the tests with an exploration on the correlation of our results with a large amount of audio descriptors. We found interesting links to two features: the signal's autocorrelation function (ACF), and spectral flux (feature calculations and definitions in [3, p. 107]). The first feature's relevance may relate to Ando's work [6], where the author finds that in good concert halls, the preferred delay time of a single reflection could be esti-

mated with the ACF of the signal, the delay being determined from $|r_{xx}^*(\Delta t_1)| = 0.1 \cdot r_{xx}^*(0)^3$. Preferred reverberation time is then: $RT_{60} = 23 \cdot \Delta t_1$.

Our test design did not plan for this sort of conclusion, but we can measure *a posteriori* both features and see if user preference is better justified from them. We are assuming thus that decay time preference lies in the value that was maximum for each song, regardless of whether it was a significant or an overlapping maximum. These values are shown in the second column of Table 5.

#	Pref. RT_{60}	$T_{sub} = 23 \times (\tau_e)_{min}$	Ξ_{flux} ($\times 10^{-6}$)	BPM
Song 1	2000	10,189	3.24	67
Song 2	714	2,495	5.93	84
Song 3	571	2,817	5.66	105
Song 4	400	1,407	4.69	120
Song 5	209	2,789	7.01	143
Song 6	539	1,909	4.42	180

Table 5. Other potential explaining factors for decay time preference. T_{sub} is the prediction according to [6], Ξ_{flux} the spectral flux, and BPM our original explanation proposal, the beats per minute.

The next column shows Ando's prediction of preferred reverberation time. As can be seen this is completely exaggerated by calculating auto-correlation in wet signals. The third column shows the spectral flux and the fourth shows our original explaining factor. If we calculate the correlation between subjective preference and each of these explanations, we get $r = 0.951$ for the auto-correlation, $r = -0.755$ for the spectral flux, and $r = -0.667$ for the BPM. A logarithmic transformation on the flux or BPM improves r (-0.812 and -0.765 respectively), but still it looks like auto-correlation could be the best explanation for this sort of choice.

5 CONCLUSIONS

Two similar tests were presented, evaluating user preference for reverb and delay time parameters and its relation to song tempo. In terms of reverb time, the proposed relationship does not hold true. Decay time is still related to tempo, as both the interview process and the subjective test showed a negative rank correlation between tempo and RT_{60}, as the homogeneity in results throughout all conditions reinforces the fact that conditions were themselves a ratio of tempo to bpm.

Unlike reverb time, the user test on delay time showed stronger results, and there are two quite definite conclusions we can draw from it: coupled delay times work better perceptually than uncoupled ones when attention is drawn to them (as in placing them on a vocal) and faster delays are preferred over longer ones, given the same conditions.

Post-test interviews with subjects helped understand that the setting of reverb time is seen as too multi-dimensional

to be correlated to a single factor (namely song tempo), as it was hinted that stylistic concerns and song genre have a bearing on user's choice. However, results still seem to indicate that even if song tempo is not the main correlate, there may be other low-level factors that strongly explain this variable. A new test design is needed to bring those to light, particularly because the decision to offer songs of different genres, production values and instrumentation may be sensible in terms of mimicking real-world situations, but was seen to bring too much confusion into the test.

Further work is also needed in analysing the way several parameters interact, especially in what relates reverb time to reverb level. Here a more interactive method-of-adjustment test might prove more adequate in explaining the underlying factors. While aspects related to delay time showed to be conclusive, much more work is needed in terms of reverb time prior to establishing a definitive model.

We have provided enough information to ground explorative approaches to automating the time parameter of temporal process in an intelligent audio mixing context, and an alternate and foreseeable route for further work would be to try an initial implementation where the mapping [16, 17] follows the broad rules hinted to herein, and evaluate its success.

6 ACKNOWLEDGMENT

7 REFERENCES

[1] B. Owsinski, *The Mixing Engineer's Handbook*, 1st ed. Vallejo, CA: Mix Books, 1999.

[2] R. Izhaki, *Mixing Audio — Concepts, Practices and Tools*, 1st ed. Oxford: Elsevier Science & Technology, 2008.

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

[4] L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. New York: Springer, 2004.

[5] M. Long, *Architectural Acoustics*, 1st ed. Oxford: Elsevier Science & Technology, 2006.

[6] Y. Ando, *Concert Hall Acoustics*, 1st ed. Berlin: Springer Verlag, 1990.

[7] ——, "Concert Hall Acoustics Based on Subjective Preference Theory," in *Springer Handbook of Acoustics*, T. D. Rossing, Ed. New York: Springer, 2007, ch. 10, pp. 351–386.

[8] D. Griesinger, "The Theory and Practice of Perceptual Modeling — How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity,"

---

[3] The time it takes for the envelope of the normalised auto-correlation function to decay to one tenth of its value at zero. The actual value used is the minimum of the running ACF, $2T = 2\ s$, with an interval of $100\ ms$.

in *Proceedings of the Tonmeister Conference*, Hannover, 2000.

[9] S. Bech and N. Zacharov, *Perceptual Audio Evaluation — Theory, Method and Application*. Chichester: John Wiley & Sons, 2006.

[10] R. King, B. Leonard, and G. Sikora, "The Effects of Monitoring Systems on Balance Preference: A Comparative Study of Mixing on Headphones Versus Loudspeakers," in *Proc. of the 131st AES Convention*, New York, 2011.

[11] A. Gabrielsson and H. Sjögren, "Perceived Sound Quality of Sound-reproduction Systems," *JASA*, vol. 65, no. 4, pp. 1019–1033, 1979.

[12] SMPTE, "RP 200:2012 "Relative and Absolute Sound Pressure Levels for Motion-Picture Multichannel Sound Systems — Applicable for Analog Photographic Film Audio, Digital Photographic Film Audio and D-Cinema"," Tech. Rep., 2012.

[13] P. Newell, *Recording Studio Design*, 2nd ed. Oxford: Focal Press, 2008.

[14] F. Mosteller and R. E. Rourke, *Sturdy Statistics: Nonparametrics and Order Statistics*, 1st ed. Boston: Addison Wesley, 1973.

[15] T. Sporer, J. Liebetrau, and S. Schneider, "Statistics of MUSHRA revisited," in *Proc. of the 127th AES Convention*, New York, 2009.

[16] J. D. Reiss and E. P. Gonzalez, "Automatic Mixing," in *DAFx*, 2nd ed., U. Zölzer, Ed. Chichester: John Wiley & Sons, 2011, pp. 523–552.

[17] A. T. Sabin, Z. Rafii, and B. Pardo, "Weighted-function-based rapid mapping of descriptors to audio processing parameters," *Journal of the AES*, vol. 59, no. 6, pp. 419–430, 2011.

THE AUTHORS

Pedro D. Pestana   Dr. Josh Reiss   Álvaro Barbosa

Pedro Duarte Pestana is currently teaching at the School of Arts at the Catholic University of Portugal (UCP) where he is also the director of the Research Center for Science and Technology of the Arts (CITAR). He received his Ph.D. in computer music at UCP, specializing in adaptive digital audio effects for automatic mixing. He has investigated topics pertaining to perception and cognition in audio, machine learning systems, acoustics, interactive sound design, and digital audio effects. He has been an active member of the AES for over a decade and won the best paper award at the 134th AES Convention in Rome.

●

Dr. Josh Reiss is a Senior Lecturer with the Centre for Digital Music at Queen Mary University of London. He received his Ph.D. in physics from Georgia Tech, specializing in analysis of nonlinear systems. Dr. Reiss has published over 150 scientific papers and serves on several steering and technical committees and is co-founder of the company MixGenius. He has investigated music retrieval systems, time scaling and pitch shifting techniques, loudspeaker design, automatic mixing, and digital audio effects, among others. His primary focus of research is on the use of state-of-the-art signal processing techniques for professional sound engineering.

●

Álvaro Barbosa is an Associate Professor and Dean of Faculty of Creative Industries at University of Saint Joseph (USJ) in Macau SAR, China. Holding a Ph.D. degree in Computer Science and Digital Communication from UPF in Spain, his academic activity is mainly focused on the field of Design for Audio and Music Technology. His recent R&D work, on experimental Network Music and Interactive Sound-Design Systems, was largely fostered in 2010 during a Post-Doctoral Research Position at Stanford University in the Center for Computer Research in Music and Acoustics (CCRMA). His current projects have special emphasis in Sound and Music Design pieces, Design Thinking and Systematic Creativity.