ANDREW MITCHELL PHD THESIS CORRECTIONS

PREDICTIVE MODELLING OF COMPLEX URBAN SOUNDSCAPES: ENABLING AN ENGINEERING APPROACH TO SOUNDSCAPE DESIGN

The following are all of the corrections made in response to the examiners' comments given in the joint report following the viva. Here, I have only specifically addressed comments from the joint report although the corrected thesis also includes changes made in response to some of the comments in the individual reports and from the viva. My response is given in red, with italic text showing what has been added to the thesis. These corrections have also been highlighted in orange in the marked changes PDF and marked 'proof'.

I'd like to thank the examiners for the insights and their helpful comments both in the viva and in the report.

JOINT REPORT MINOR CORRECTIONS

1. Separate the layers of information according to their importance and better highlight the main results of your work.
   As suggested in the viva, to better highlight the main results in a clear and concise way, I have added the following key takeaway sections to the empirical study chapters:

   Chapter 5:

   The key takeaways demonstrated in this chapter are

   - *The modelling achieved reasonable performance and allowed me to investigate soundscape perception in a unique scenario.*
   - *The model was able to capture the expected changes in perception within a wide variety of location types. The results of the model were consistent with what would typically be expected for the more obvious soundscape changes (i.e. Monumento Garibaldi became much more chaotic with the introduction of lawn works).*
   - *Location-context effects are very important for pleasantness but not for eventfulness.*
   - *A reduction in sound level does not always lead to an increase in pleasantness; the change in sound source composition and sonic character is important.*

   Chapter 7:

   The main conclusions of this study are:

   - *When considering short recordings of single-source environmental sounds, no significant differences in noise annoyance were observed as a function of demographic factors, such as gender and self-reported area of residence (i.e. from very quiet to very noisy).*
   - *The multi-level linear regression model derived from this case study achieved an overall $R^2=0.64$, using sharpness as a fixed effect (the first level), and impulsiveness, roughness, and tonality as random effects allowed to vary according to the type of sound (the second level) as predictors for perceived noise annoyance.*
   - *By incorporating sound source information along with the psychoacoustic metrics, the model can better reflect how listeners will respond to different sounds. As such, a general model should*

strive to automatically recognise sound sources and combine this information with the psychoacoustic analyses.

Chapter 8:

The key takeaways of this chapter are:

- *Psychological well-being is positively associated with pleasantness and negatively associated with eventfulness in males and individuals that did not report their occupations.*
- *Occupational status, in particular retirement as a proxy of age and gender, was related to the perceptions of pleasantness and eventfulness.*
- *In total, personal factors were shown to account for 1.4% of the variance for pleasantness and 3.9% of the variance for eventfulness.*
- *The overall models, including the location-context, achieved* $R^2_{ISOPl}=0.35$ *and* $R^2_{ISOEv}=0.18$*, indicating that the location-context accounted for the majority of the variance.*
- *Several potential methods of incorporating personal factors into a general model are proposed in order to account for differences in demographic patterns, however based on the variance explained by these factors, they do not appear to be crucial to include at this stage.*

To make it more clear which layers of information were deemed to be important for a general predictive model, I have added the following to Section 10.1 "Key Findings and Contribution to Knowledge":

In all, the results of the empirical studies have helped to determine the most important information to consider in a general predictive soundscape model, at this stage of the state-of-the-art for these models:

1. ***Sonic Character*** *- unsurprisingly, quantifying the sonic character of a soundscape through acoustic, psychoacoustic, spectrogram, and/or bioacoustic analysis is crucial to predicting the soundscape perception. While the models used in this thesis focussed on acoustic/psychoacoustic metrics, future work will involve expanding this suite of metrics to capture of aspects of the sonic character.*
2. ***Sound Source Identification*** *- in conjunction with the sonic character analysis, sound source identification should be performed. As demonstrated in Chapter 7, including sound source labels introduces specific context to the psychoacoustic features, greatly improving accuracy and usefulness of the model. Future work will require implementing this sound source identification in an automated way, using AI environmental source recognition models.*
3. ***Location-context Information*** *- The models in Chapter 5 and 8 clearly demonstrated the importance of considering the context of the location in the predictive model, especially for predicting pleasantness. The first step of automating and generalising this effect will be to incorporate visual analysis, making use of the 360°videos and photos, to characterise the location.*

Beyond this, while ***personal and demographic*** *information was shown to have an identifiable effect on soundscape perception. However, in the context of developing a practical predictive model, the relatively low amount of variance explained by these factors and the difficulties in* *obtaining the necessary input data in a design application, indicates that personal factors should likely be considered optional factors.*

2. Better highlight the data that was used for modelling.
   As suggested in the viva, I have reproduced versions of Fig. 3.3 showing specifically which portions of the database were used for the two modelling chapters which use the ISD, Chapter 5 and Chapter 8. These are now Figs. 5.2 and 8.1.

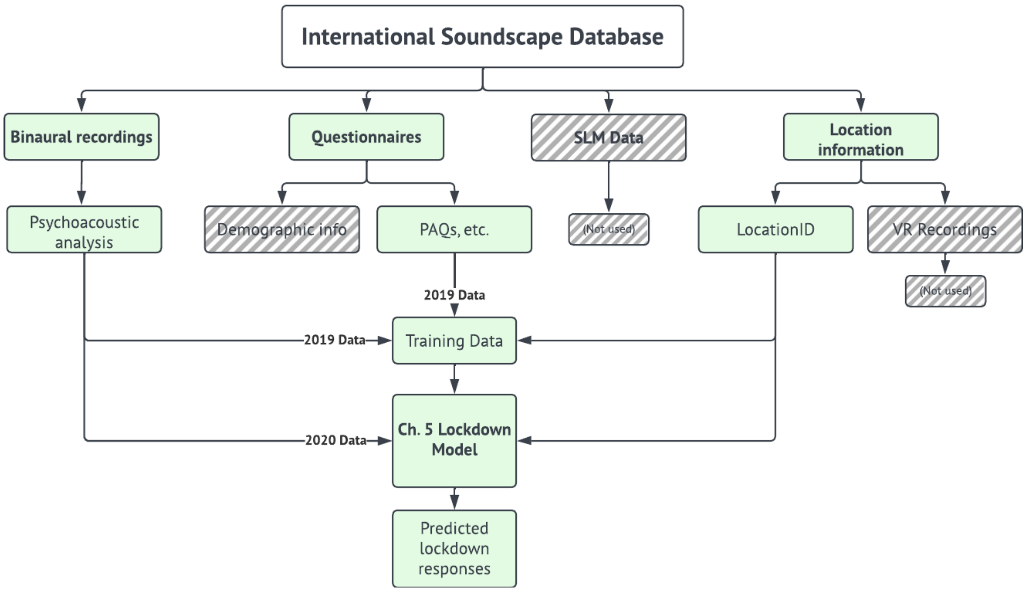

Fig 5.2: Version of Fig. 3.3 showing which subsets of the data from the ISD are used in the study. Those portions used are highlighted in green while the portions not used in this study are crossed out.

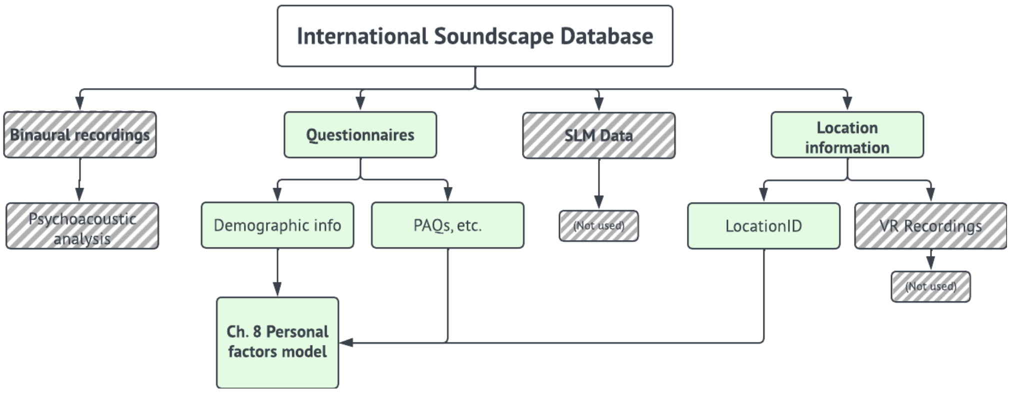

Fig 8.1: Version of Fig. 3.3 showing which subsets of the data from the ISD are used in the study. Those portions used are highlighted in green while the portions not used in this study are crossed out.

3. Mention your hypothesis for why the online survey did not work well.
   Based on your comments, **I have conducted additional analysis** to compare the online perceptual data responding to the 2019 recordings and the 2019 in situ data. This had two relevant outcomes: (1) there was a significant difference between the online and in situ responses, confirming that we could not directly use the online data to confirm the model predictions; (2) the fact that the online pleasantness

responses were much more different than the eventfulness responses, may additionally confirm the importance of context for pleasantness, but not for eventfulness.

I have added the following section to Chapter 5. In addition, Table 5.3 has been added showing the differences between online responses to 2019 recordings and the in-situ 2019 surveys.

Online perceptual data

However, after the initial data collection, questions were raised as to how the playback loudness impacts ecological validity as it relates to the perceived affective quality of the soundscape. The results of the listening study indicated that the PAQ responses collected in the online study to the 2019 recordings were not consistent with those collected on site. Table 5.3 shows the range, mean, and standard deviation of the responses collected for each location. By looking at the difference between the mean responses for online vs. in-situ we can see multiple large differences between the two. The largest difference is for the RegentsParkJapan pleasantness - while this was rated highly pleasant (0.62) on site, the online study respondents rated it as neutral (0.05). In general, the online listeners rated the soundscapes as much more neutral than the in-situ listeners. This difference was less pronounced for eventfulness, where the biggest difference between the in-situ and online measurements was 0.22, for St Pauls Row.

This difference in the online study could have a few possible explanations. The first is the lack of control over the playback level. Without a reference level for the participants to calibrate to, it was not possible to verify the actual sound level they were exposed to. The second is a lack of contextual information; the participants were exposed only to the binaural recording, with no additional information about the visual environment, the architectural typology, or the general setting of the space. The particular online listening setup we designed appears to be insufficient for collecting perceptual responses, however since the COVID-19 lockdowns, new developments in online listening studies have been made, which may enable this in the future (Peng et al., 2022; Tan, Hasegawa, & Lau, 2022).

Table 5.3:

Table 5.3.: Comparison in perceptual responses between the online listening study and *in situ* survey data. Both sets include only 2019 data; responses to 2019 recordings for the online study and survey data collected in 2019 for the *in situ* data. The range, mean, and standard deviation of the responses are reported for each location. The difference between the online and *in situ* responses is also reported.

		Online Study			*In situ*			Difference (Online minus *in situ*)		
	LocationID	**Range**	**Mean**	**Std. Dev.**	**Range**	**Mean**	**Std. Dev.**	**Range**	**Mean**	**Std. Dev.**
Pleasant	CamdenTown	1.41	-0.32	0.27	1.38	-0.10	0.29	0.03	-0.22	-0.01
	EustonTap	1.25	-0.11	0.30	1.21	-0.20	0.28	0.04	0.09	0.02
	MarchmontGarden	1.53	-0.03	0.35	1.85	0.27	0.42	-0.32	-0.30	0.07
	MonumentoGaribaldi	1.06	0.06	0.30	0.93	0.42	0.27	0.13	-0.36	0.03
	PancrasLock	1.38	-0.19	0.29	1.68	0.25	0.39	-0.29	-0.43	0.09
	RegentsParkFields	1.18	0.34	0.27	2.00	0.50	0.39	-0.82	-0.16	0.12
	RegentsParkJapan	1.75	0.05	0.49	1.78	0.62	0.33	-0.03	-0.58	0.16
	RussellSq	1.28	-0.12	0.35	1.32	0.47	0.30	-0.04	-0.59	0.06
	SanMarco	1.19	-0.10	0.31	1.43	0.21	0.32	-0.23	-0.32	0.01
	StPaulsCross	1.03	-0.05	0.27	1.56	0.36	0.33	-0.53	-0.41	0.07
	StPaulsRow	1.28	-0.22	0.32	1.60	0.22	0.34	-0.32	-0.44	0.02
	TateModern	1.31	-0.12	0.33	1.43	0.37	0.29	-0.12	-0.50	0.04
	TorringtonSq	1.41	0.10	0.30	1.85	0.08	0.31	-0.44	0.02	0.01
Eventful	CamdenTown	1.50	0.26	0.38	1.75	0.36	0.35	-0.25	-0.10	0.03
	EustonTap	1.10	0.06	0.32	1.32	0.18	0.30	-0.22	-0.12	0.02
	MarchmontGarden	1.34	0.04	0.36	1.21	-0.04	0.26	0.13	0.08	0.10
	MonumentoGaribaldi	1.78	-0.10	0.37	0.88	-0.01	0.24	0.90	-0.09	0.12
	PancrasLock	1.38	0.08	0.33	1.31	0.07	0.28	0.07	0.00	0.05
	RegentsParkFields	1.18	0.01	0.26	1.35	-0.05	0.27	-0.18	0.06	0.00
	RegentsParkJapan	1.15	-0.07	0.25	1.27	-0.01	0.25	-0.12	-0.06	0.01
	RussellSq	1.21	0.13	0.29	1.04	0.05	0.25	0.16	0.08	0.04
	SanMarco	1.31	0.44	0.33	1.28	0.37	0.29	0.03	0.07	0.05
	StPaulsCross	1.46	0.07	0.29	1.38	0.14	0.29	0.07	-0.07	0.00
	StPaulsRow	1.21	0.34	0.27	1.51	0.11	0.27	-0.31	0.22	0.00
	TateModern	1.72	0.30	0.36	1.68	0.21	0.31	0.04	0.09	0.05
	TorringtonSq	1.25	0.02	0.31	1.60	0.18	0.32	-0.35	-0.16	0.01

This led to a small addition to the discussion section:

> *[…] my results indicate that these types of contextual factors are not significant for ISOEventful, and do not affect the relationship between the acoustic features of the sound and the perception. It is possible that this result also explains the difference in the online and in-situ responses highlighted in Section 5.2.3. I observed a much greater discrepancy in pleasantness ratings between the two modalities than for eventfulness ratings. This may be the result of the lack of contextual information for the online study, greatly impacting pleasantness, while the eventfulness rating was not greatly affected.*

4. Better highlight the near-term future prospects for validating and improving the proposed model.

   My main proposals for near-term improvements have been detailed in the discussion sections of Chapters 7 and 8 (Sections 7.4 and 8.4.1) and most importantly in the second half of Chapter 9 (Sections 9.4 'Making use of the soundscape circumplex', and 9.5 'Probabilistic predictions'). I have added the following to Section 10.2 'Limitations and future work' to highlight this:

   > *In particular, the immediate future work which I will be undertaking to create an improved model consists of incorporating sound source information (see Section 7.4) and generating predicted distributions of responses (see Section 9.5).*

   Regarding the future prospects for validating the model, I have added the following section on longitudinal data collection and validation, as discussed in the viva:

   > *10.2.2 Longitudinal validation*
   >
   > *Within a machine learning context, the model presented in Chapter 5 was tested against an unseen set of data by creating the 80%/20% training/testing set. However, as the*

testing data was randomly selected from across the pooled dataset, the testing data itself is not entirely indepedent. The data collected for the ISD often took place over several sessions/days, but in all cases the goal was to select days which maintained consistent environmental and soundscape conditions and formed one coherent dataset from which both the training and testing sets were drawn. To truly test both the consistency of soundscape perception of a space (see the following section for more on this) and the predictive performance of the model, we should return to the same spaces over and repeat the data collection protocol. By returning now, nearly 4 years later, we could assess whether the soundscape of these spaces has fundamentally changed, confirming or raising doubts about the consistency of how the soundscapes are perceived. More importantly for the model validation, this new set of data collected in the same locations under different conditions would provide a truly independent testing set with which to validate the model presented in Chapter 5. As the SSID project progresses and comes to an end, one of the remaining data collection goals should be carrying out these return visits to the same spaces to reassess their soundscapes and validate the success of the predictive mode.

5. Mention how your proposed protocol could be soberer, using the minimum number of items needed for a similar/useful soundscapes collection.
   Two levels of simplification with minimum items required have been added to the protocol chapter as Section 3.6.4:

   > *The SSID Protocol was initially developed as the overarching data collection procedure for the Soundscape Indices project, which incorporates many avenues for investigation and different research agendas. As part of an exploratory research task, this set of data was intended to characterise the conditions in the locations as fully as possible, allowing us to investigate research questions we may not have considered at the start. However, this full characterisation is often not necessary when performing a soundscape assessment. Instead, the researcher may only wish to capture the most important features of the soundscape in a way which is compatible with the ISD. To represent the potential use cases and access to equipment that researchers may have, I will present three levels of data collection complexity.*
   >
   > *The full SSID Protocol presented above is the most comprehensive, comprising both Recording and Questionnaire stages, including the binaural recordings, stationary SLM, environmental data, 360°video, and ambisonic recording.*
   >
   > *The simplified protocol is intended to fully capture the auditory and visual environment, along with the soundscape perception assessed through questionnaires. This can be achieved by conducting only the Questionnaire Stage and including the following:*
   >
   > - *Stationary SLM*
   > - *Binaural recordings*
   > - *360°photos*
   > - *Questionnaires*
   >
   > *This would enable a much simpler equipment setup and would allow the researchers carrying out the data collection to focus on gathering as many questionnaire responses as possible. Since the goal is only to assess the in situ environment, this method would not require the ambisonic*

recording or 360°video intended for VR playback.

In some cases, even this simplified protocol may prove inaccessible to some researchers, especially as it still requires access to a 360°camera and binaural headset. In these situations, a minimal protocol would comprise only the following:

- *Questionnaire*
- *SLM*

I would suggest that ideally two SLMs should be used - one as the standard stationary SLM running throughout and one to replace the function of the binaural headset, taking short term measurements with audio recording while each participant is completing the survey.

In both the simplified and minimal protocols, the Questionnaire stage should be carried out with the same steps as the full protocol, using the same (or an expanded) survey questionnaire. In all cases, the researcher's notes and additional documentation of the space, including detailed photographs, are crucial to provide further information to anyone analysing the results of the soundscape assessment.