25-29 August 2024
inter.noise
NANTES FRANCE

Multimodal Hong Kong: A multisensory study of environmental perception of places through sight, sound, and smell

Qianxi Jiang[1]
SoundLab, School of Creative Media, City University of Hong Kong
Run Run Shaw Creative Media Centre, 18 Tat Hong Avenue, Kowloon Tong, Hong Kong

PerMagnus Lindborg [2]
SoundLab, School of Creative Media, City University of Hong Kong
Run Run Shaw Creative Media Centre, 18 Tat Hong Avenue, Kowloon Tong, Hong Kong

Lishan Wei[3]
SoundLab, School of Creative Media, City University of Hong Kong
Run Run Shaw Creative Media Centre, 18 Tat Hong Avenue, Kowloon Tong, Hong Kong

Jieling Xiao[4]
Birmingham City University
Birmingham School of Architecture and Design, The Parkside Building, 5 Cardigan Street, Birmingham, B4 7BD, United Kingdom

Ran Yue[5]
SoundLab, School of Creative Media, City University of Hong Kong
Run Run Shaw Creative Media Centre, 18 Tat Hong Avenue, Kowloon Tong, Hong Kong

ABSTRACT

Despite the importance of soundscape and smellscape in intangible cultural heritage, little is known about multisensory interactions. To explore the impact of multiple senses on environmental perception, the present study focused on six stalls at Shek Kip Mei wet market in Hong Kong. In the first part of the study, an online survey collected perceptual ratings of whole, auditory, and visual environmental quality, as well as free-form responses of imaginations of smell. Participants (N = 36) were presented with 12 audiovisual captures where audio was either matched with video or mismatched. ANOVA was used to examine the influence of three independent factors - scenes, modes, and a priori valence groups - on the ratings. The imagined smells were analysed through frequency counts of keywords. The results demonstrate participants' capacity to notice changes in audio modes and to perceive the environment through the recognition of scene context and multisensory interactions. In the second part of the study, not

[1] qjiang28-c@my.cityu.edu.hk
[2] pm.lindborg@cityu.edu.hk
[3] lishawei@cityu.edu.hk
[4] jieling.xiao@bcu.ac.uk
[5] ranyue6-c@my.cityu.edu.hk

discussed in the present paper, we will extend this experimental method to include smells in a laboratory setting. We discuss the importance of understanding multisensory interactions between sight, sound, and smell for environmental perceptions and Hong Kong's intangible cultural heritage and urban studies.

1. INTRODUCTION

The urban development and transformation have reshaped cityscapes and people's daily lives and traditions, leading to continuous changes in soundscape and smellscape [1-2]. To better connect people to the past and present and to understand the lived heritage of people and places, senses are widely used in heritage studies, and sensory heritage has become an emerging research topic [3-4]. Although there is increased emphasis on sound and smell as important components of cultural heritage, the research on sensory correspondences in this context remains insufficient compared to sight which traditionally dominates the realm of sensory heritage [1, 4]. Furthermore, many studies only focus on one individual sensory heritage component and still lack the perspective of multisensory environmental experience [4], particularly sensory interaction involving sound and smell.

Hong Kong, despite its unique cultural history, is facing the risk of neglecting essential threads of the city fabric in its rapid urban transformation [5]. Soundscape and smellscape are not the mainstream focus in narratives of urban research related to Hong Kong, which hinders the comprehensive and in-depth documentation and conservation of its intangible cultural heritage (ICH) represented by these senses. As significant media for perceiving the environment, it is crucial to gain a deeper understanding of the impacts that sounds (keynotes, sound signals, and sound marks) and smells (odours, scents, flavours, and smellmarks) on people's daily lives [6-7].

Sound, along with sight, is typically considered a sensory heritage component that has been relatively well-explored in research [4]. The International Organization for Standardization [8] defines soundscape as an "acoustic environment as perceived or experienced and/or understood by a person or people, in context." People interact with soundscapes through perception, thereby constructing meaning and understanding of their surroundings, which can be considered cultural construction [9-10]. Dumyahn and Pijanowski [9] noted that the social and cultural significance of soundscape is presented in terms of "a sense of place," cultural and historical heritage values, and interaction with landscape perception. For instance, Yelmi [1] recorded the perspectives of both locals and tourists on the urban soundscape of Istanbul, demonstrating that sound, as a component of everyday life and activities, holds significant value in perceiving urban spaces and constructing cultural identity.

Referring to the basic spatial vocabulary from Schafer's concept of the soundscape, Porteous [11] proposed the term smellscape, arguing that "smells may be spatially ordered or place-related." It refers to "the olfactory environment as perceived and understood, consisting of odours and scents from multiple smell sources" [12]. Defining the smellscape of heritage sites and exploring the changes in smells can lead to richer memories of places and events [3]. Skrede and Andersen [13] documented a putrid smell produced by digesters in Moss, Norway's factories. As industrial facilities undergo deindustrialisation, some workers desire to revive the local olfactory heritage, known as "the smell of Moss," by reproducing this smell to reconstruct their memory of the cityscape [13]. The documentation of smells is considered beneficial for the reproduction and creation of smell environments [6], however, current research on smell as a component of heritage remains relatively underdeveloped compared to that of sight and sound [4].

Multisensory interactions shape people's comprehensive responses to their living environments [14]. The experienced sights, sounds, smells and other senses can establish a connection between people and their individual and public pasts through memory [4]. This sensory connection contributes to creating immersive experiences that evoke emotion and

reminiscence [15]. Compared to visual-audio, there are relatively few studies on sensory heritage involving auditory-olfactory association. Lindborg and Liew [12] compared smell imaginations and memories evoked by audio-video in an online survey, and compared them to multimodal observations that had been made on-site at seven specific locations at a Singaporean heritage market. The results suggested that episodic memory and cultural stereotypes can influence the imagination of smell sources from audiovisual information [12]. Within the context of sensory heritage, there remains a notable gap in the overall research on multisensory environmental experience, particularly concerning the multisensory interactions perceived and experienced on-site [4].

Therefore, proposing a multimodal research approach incorporating sight, sound, and smell, the Multimodal Hong Kong Project (MMHK) has conducted an online survey to investigate people's interpretation and experience of multimodal environments in Hong Kong's wet market. The market environments were captured through 360-degree video and surround sound. We recorded and compared participants' multisensory experiences in the environments to examine the influence of different scenes, modality modes, and valence groupings on their environmental perceptions. We also explored whether participants can make meaningful smell imaginations within audiovisual environments. The objective of this study is to find boundary conditions, whether present in the stimuli, the perceiver, or both, that appear to have a positive or negative impact on people's impressions of the sites.

2. METHODS

2.1. Materials

Shek Kip Mei Market is located in Shek Kip Mei Estate in Sham Shui Po, Hong Kong. It is accessible for residents due to its proximity to the metro station and numerous bus stops. The market has a total of 88 stalls, each offering a diverse range of goods. The fish and prawns stall as well as the raw meat stalls are positioned close to Woh Chai Street, which has more servers and a ventilation system to drain water and strong smells. In the middle of the market, there are many dry food stalls, veggies stalls, and fruit stalls with narrower passageways. Although the smells in this area are lighter than meat stalls, each stall has its own unique smells bringing different sensations. On the outskirts of the market, there are a few flower stalls, convenience stalls, and religious goods stalls. The lighting here is slightly dimmer compared to the central market since the streets have natural light coming in. The presence of fresh flowers and their fragrant scent effectively captures people's attention.

2.2. Stimuli

This study selected six locations at Shek Kip Mei Market: the dry food stall, fish stall, fish and prawns stall, flowers stall, fruit stall, and veggies stall (see Figure 1). They were captured at specific sites in the market. In order to make the study manageable, we split the six locations into three groups, and created two matched and two mismatched conditions within each. Thereby, the total number of stimuli was 12, out of which six were matched and six were mismatched. In the first pairing of locations, we included:

1. Fish stall video + Fish stall audio (Matched)
2. Fish stall video + Fruit stall audio (Mismatched)
3. Fruit stall video + Fruit stall audio (Matched)
4. Fruit stall video + Fish stall audio (Mismatched)

In the second pairing, we included:

5. Fish and prawns stall video + Fish and prawns stall 2 audio (Matched)
6. Fish and prawns stall video + Dry food stall audio (Mismatched)
7. Dry food stall video + Dry food stall audio (Matched)
8. Dry food stall video + Fish and prawns stall audio (Mismatched)

In the third pairing, we included:

9. Veggies stall video + Veggies stall audio (Matched)
10. Veggies stall video + Flowers stall audio (Mismatched)
11. Flowers stall video + Flowers stall audio (Matched)
12. Flowers stall video + Veggies stall audio (Mismatched)

We made audiovisual recordings at the sites, and selected a 90-second clip from each. The ambisonic audio was synchronised and spatially adjusted to match the 360-degree video. When played back on YouTube, mouse movements correctly change the audio and video first-person perspective.

Figure 1: The six stalls at Shek Kip Mei Market chosen for the study.

2.3. Participants

In a preliminary test, we found that the median time required to complete the online test would be approximately 20 minutes, which seemed appropriate. We then reached out using snowball sampling to gather a sufficient number of responses, targeting around 40. As an exclusion criterion, we excluded cases in which the respondent had used less than 7 minutes on the questionnaire. This left a final sample of 36 responses, with completion times ranging from 7 minutes and 13 seconds to over 50 minutes. The long time reported by around ten participants can be justified by them taking one or more breaks underway, or taking extra time for translating the questions, such as from English to Chinese. The mean time spent was about 24 minutes, and the median was about 23 minutes.

Out of the 36 participants, 20 were female and 16 were male. Their ages were in a range between 21 and 53, with a median age of 26 years. Regarding their place of origin, 22 participants indicated being from Mainland China, 5 from Hong Kong, 4 from Malaysia, and the rest from several other countries including India, the UK, France, and the US. Most of the participants, 28, were Chinese (Mandarin) speakers.

2.4. Procedure

In an online survey designed using QuestionPro (www.questionpro.com), each participant was assigned and required to view all 12 stimuli (listed above) in random order. To maximise the clip-viewing experience, participants were asked to be in a calm and distraction-free environment and to use quality headphones and desktop's full-screen mode. They were allowed to rotate the view by dragging the mouse or tapping the "compass" icon on the clip.

For each stimulus, participants were asked to evaluate the environmental quality by answering three questions in random order, using a seven-point Likert scale ranging from -3 “Very unpleasant” to 3 “Very pleasant” with 0 “Neutral” in the middle:

- How pleasant do you feel that this environment is?
- Are the audio elements in this environment pleasing?
- Are the visual elements in this environment pleasing?

Then, they imagined what the smells might be like and described them in their own words. This was an open-ended question.

3. RESULTS

3.1. Environmental Quality

In the online survey, participants (N = 36) rated the whole, audio, and visual environmental quality of 12 stimuli in terms of pleasantness. Following the Stimuli section in which the six locations were divided into three pairings, so as to be able to make a priori comparison between relatively higher and lower pleasantness, within each pairing. The scenes were then grouped by relative valence, as low pleasant (Fish stall, Fish and prawns stall, and Veggies stall) and high pleasant (Fruit stall, Dry food stall, and Flowers stall), in which the first group (low) was assumed a priori to have lower pleasantness than the second (high). This prior assumption was borne out by the results. Therefore, Mode (matched and mismatched), Scene (six locations listed in the Stimuli section), and Group (low and high) were set as independent variables, while rated Quality (Environment as a whole, Audio, and Video) was the dependent variable.

We first analysed participant ratings of pleasantness across conditions, i.e. the questionnaire’s scales for Environment (as a whole), Audio elements, and Video elements, while controlling for each participant’s differing use of the seven-point Likert scale by applying within-participant z-score normalisation. This revealed a significant difference (p = 0.018*, η^2 = 0.007) between Video and Audio ratings, with Environment in between. See the error bar plot in Figure 2.

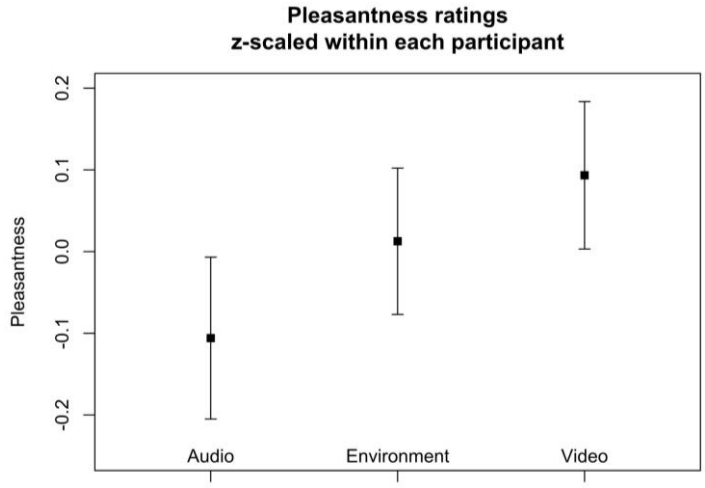

Figure 2: Error bar plot of rated Pleasantness across all twelve stimuli (N = 36) for Environment as a whole, Audio elements, and Video elements, with 95% confidence intervals around the means.

We then looked at the main effects of Mode (i.e. matched or mismatched audio) and Scene (six different stalls) across rating conditions using multivariate analysis of variance (MANOVA), showing that there were strongly significant impacts of Scene and the interaction between Mode and Scene on the environmental quality ratings [Scene: Pillai’s trace (5 df) = 0.10, p = 0.00006***; Mode*Scene: Pillai’s trace (5 df) = 0.071, p = 0.0081**], while Mode did not [p =

0.73]. For Mode and Group (i.e. a priori high valance versus low valence locations), the results of MANOVA revealed the significant effects of Group and the interaction between Mode and Group on the environmental quality ratings [Mode: Pillai's trace (1 df) = 0.025, p = 0.012*; Mode*Group: Pillai's trace (1 df) = 0.044, p = 0.00018***], while Mode, again, did not [p = 0.73]. This shows that our prior assumption of low versus high within the location pairings was supported. The results prompted us to continue with three univariate ANOVAs to further examine the effects of Mode and Group on the single scales of Quality: Environment as a whole, Audio, and Video. Eta-squared (η^2) was used to evaluate effect size. The results showed that Group had a significant impact on the quality rating of Environment as a whole and Video, while the interaction effect of Mode and Group notably affected Environment as a whole and Audio (see Table 1). Note that the interaction effect between Mode and Group on Audio pleasantness ratings, η^2 = 0.066, can be considered to be of medium size, while other effect sizes are small.

Table 1: Univariate ANOVA and eta-squared results for the effects of Mode and Group on each Quality (Audio, Environment as a whole, and Video).

	Audio				**Environment**				**Video**			
	SS	**F**	**p**	**η^2**	**SS**	**F**	**p**	**η^2**	**SS**	**F**	**p**	**η^2**
Mode	1.5	1.02	0.31	0.003	2.8	2.2	0.14	0.005	1.1	0.84	0.35	0.002
Group	2	1.35	0.25	0.003	16.3	13	0.0004***	0.031	17.4	13.4	0.00029***	0.032
*Mode*Group*	42.9	28.7	0.0000007***	0.066	12	9.6	0.0021**	0.023	4	3.05	0.081	0.007

Error bar plots were created for the respective univariate analysis comparison of two independent variables' conditions, regarding their influence on the quality ratings of Environment as a whole, Audio, and Video. In Figure 3 involving Scene with six locations, across all three environmental quality scales, the flowers stall presented the highest pleasantness degree, while the fish stall had the lowest, with the greatest difference between the two.

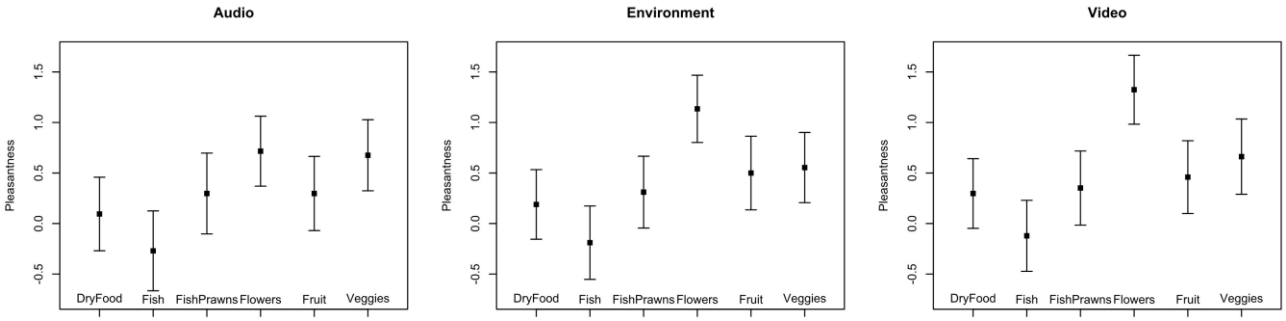

Figure 3: Error bar plots of rated Pleasantness (N = 36; Audio elements, Environment as a whole, and Video elements) with 95% confidence intervals around the means, for six locations (i.e. factor: Scene).

For Group with scenes divided into two groups according to a priori valence assumption, recall that Fruit, Dry food, and Flowers had been grouped as "high quality" locations, in relation to the (paired) Fish, Fish and prawns, and Veggies as "low quality" locations. Figure 4 illustrates the distinctions made between the two groups in the three rating conditions. The "high" group demonstrated a more positive impact on the pleasantness ratings of all scales, which supported the prior assumption of low versus high within the location pairings. While there were significant differences between the two groups in pleasantness ratings of Environment as a whole and Video, this was not the case for ratings of Audio pleasantness. This is an important finding that shows the effect of directing participants' attention towards audio elements compensated for the effect of the influence from video elements. Notably, in mismatched

stimuli, there would be video elements from a low-quality location presented together with audio elements from a high-quality location. This might explain the observation that the auditory modality compensated for, but did not overrule, the effect of the visual modality.

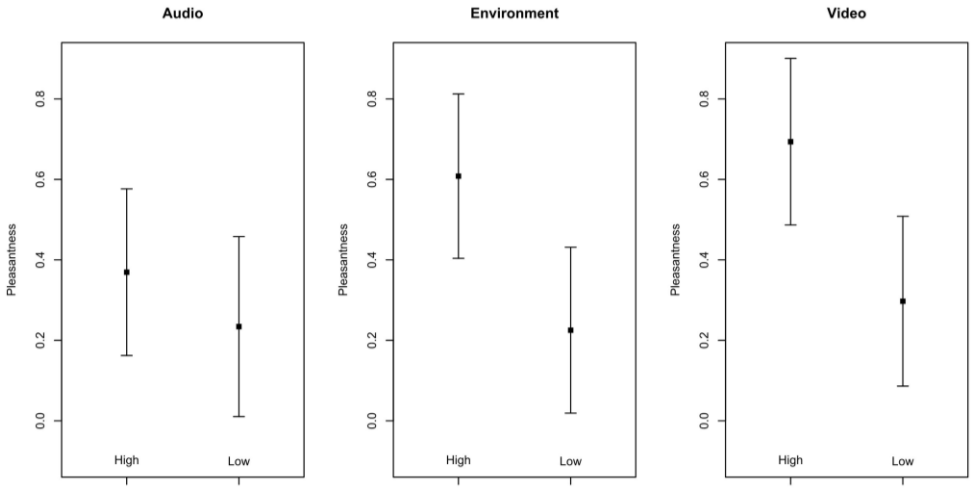

Figure 4: Error bar plots of rated Pleasantness (conditions as in Figure 3) for locations grouped a priori for high and low valence (i.e. factor: Group).

We then move on to the interaction effect, which (as shown in Table 1) was significant for the ratings of Environment as a whole and for Audio, but not for Video. Figure 5 shows interaction plots for the three conditions of pleasantness ratings. Firstly, the interaction effect in the Audio pleasantness ratings was the strongest, with η^2 almost three times as large for Audio as for Environment ratings. Secondly, for the ratings of the pleasantness of Audio elements, the wording of the rating scale guided participants to focus on what they heard, therefore leading to the most significantly differing ratings between high and low pleasantness groups. Thirdly and by contrast, for the ratings of Video elements, the questionnaire design made participants think about what they saw (possibly ignoring the audio), leading to a non-significant interaction effect. Note that there was a main effect of high versus low group for Video, but as the participants' attention was in this case drawn towards the video elements, ratings were not confounded by mismatched audio that would, in half of the cases, have higher a priori pleasantness. Finally, the ratings of Environment as a whole directed participants' attention not only to video, or to audio, but to both (to some degree). This led to a smaller yet significant interaction effect, with η^2 for Environment about a third as the one for Audio.

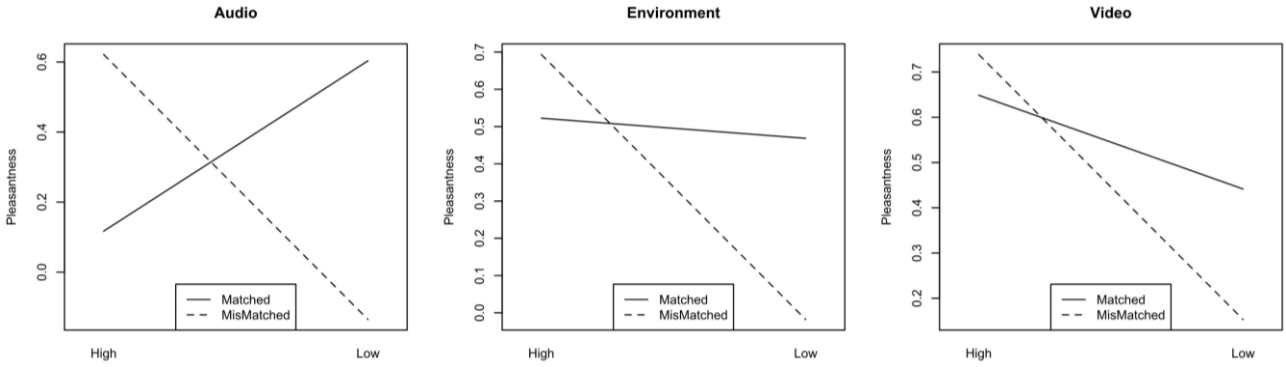

Figure 5: Interaction plot of the interaction effect between Mode and Group on each Quality.

3.2. Imagined Smells

Participants (N = 36) described through open-ended questions what the environments smelled like while viewing the audiovisual clips. There were 408 responses with a total of 2185 individual words. The word frequency was analysed through NVivo (lumivero.com/products/nvivo/) addressing the issue of synonyms. Stop words and a minimum character length of three for a word length were also set. Table 2 presents the top five frequency counts of imagined smells in different scenes and modes.

For the same scene but with matched or mismatched audio mode, despite slight differences in wording and frequency, the words used to image and describe smells were generally similar and directly related to the scene. Particularly, the first most common word was consistent across the scene's both modes. This preliminary finding suggested that participants' imagination of smells primarily relies on their judgement of the context of the scene, with minimal influence from the audio mode. Furthermore, when participants imagined and described the smells of a certain scene, they may also associate it with other scenes. For instance, the table showed that the association between the fruit stall and veggies stall was commonly observed, leading to the mixed usage of olfactory words related to both. This tentative result does not support the contention in our previous study [12], where participants were able to correctly imagine smell sources from the audio-only condition. However, that study used a different design and analysis. More research is needed to probe cross-modal association and imagination in complex environments.

Table 2: Most common words in imagined smells for the 12 stimuli from online survey participants (N = 36).

Scene	Mode	Smell
Fish stall	Matched	fish [+ fishy] (10.45%), meat [+ meats] (5.22%), damp [+ dampness, moist] (2.99%), mixed [+ combination] (2.24%), wet [+ moisture] (2.24%)
	Mismatched (Fruit stall audio)	fish [+ fishy] (12.38%), meat [+ meats] (6.67%), bad (1.90%), mixed [+ mixed] (1.90%), damp [+ moist] (1.90%)
Fruit stall	Matched	fruit [+ fruite, fruits, fruity] (9.46%), fresh [+ freshness, clean] (7.63%), vegetables [+ vegetable, veggies, vegs] (5.34%), floral [+ flowers] (3.05%), mixed [+ mix] (2.29%)
	Mismatched (Fish stall audio)	fruit [+ fruite, fruits, fruity] (11.72%), vegetables [+ vegetable, veggies, vegs] (5.60%), fresh [+ refreshing] (3.38%), air [+ atmosphere] (2.40%), floral [+ flowers] (2.40%)
Fish and prawns stall	Matched	fish [+ fishy] (9.10%), seafood (5.74%), food [+ foods] (4.13%), mixed [+ combination] (4.13%), old [+ previous] (2.48%), meat (1.65%)
	Mismatched (Dry food stall audio)	fish [+ fishy] (10.74%), seafood (5.41%), fresh (4.50%), food [+ foods] (3.60%), mixed [+ combination] (2.70%)
Dry food stall	Matched	dried [+ dry] (8.87%), food (4.84%), mixed [+ blend] (4.03%), seafood (4.03%), fishy (3.23%)
	Mismatched (Fish and prawns stall audio)	dried [+ dry] (9.45%), food (6.30%), mixed [+ combination, mixture] (3.94%), fish (2.36%), musty [+ moldy] (2.36%)
Veggies stall	Matched	vegetables [+ vegetable, veggies] (12.78%), fresh [+ invigorating] (10.53%), mixed [+ mixture] (3.01%), earthy (2.26%), damp (1.50%)
	Mismatched (Flowers stall audio)	vegetables [+ vegetable, veggies] (12.90%), fresh (7.26%), fruit [+ fruite, fruits] (5.65%), fragrance [+ aroma, perfume, scent, scents, sweet] (4.03%), mixed (3.32%)
Flowers stall	Matched	flower [+ floral, flowers] (20.54%), fragrant [+ fragrance] (7.14%), fresh (4.46%), sweet (2.68%), air (1.79%)
	Mismatched (Veggies stall audio)	flower [+ floral, flowers] (18.49%), fragrant [+ fragrance] (3.36%), fresh [+ clean] (3.36%), sweet (2.68%), air (2.52%)

4. DISCUSSION

4.1. Multisensory Perceptions of Environmental Quality

The usage of a multisensory approach allowed this study to explore the interaction among sight, sound, and smell, particularly within a complex servicescape environment. Compared to matched or mismatched modes of the audio-visual composites, the results of rating environmental quality and imagining smells indicated the greater impact of scenes and groups on participants' environmental perception. When participants focused on the whole environment, visual elements, and imagination of smells, they might pay little conscious attention to audio elements. Considering the overall research findings, the multisensory interaction in perceiving scenes might be the underlying factor. When viewing the audiovisual clips, the participants could imagine other sensory content they might perceive within their surroundings, rather than just audio or video content. Despite the matched or mismatched modes influencing the audio information to some extent, participants were still able to define a basic context through their judgement of scenes based on visual and audio cues, therefore further imagining potential smells [12]. In this context, we note that a recent study found that smell might have an impact on the evaluation of audio environments [7]. If this is true, the contradictory observation in our data, that participants were able to evaluate the environmental quality and imagine smells without being influenced by the audio modes, remains a topic for further research.

It is noted that the interaction effects of Mode and Group had different impacts on audio and video environments. The wording or phrasing of respective questions in the survey may direct participants' main attention towards what they heard or saw. Therefore, for the Audio pleasantness ratings, participants could perceive more clearly whether the audio was matched with the video. Conversely, for the Video ratings, in addition to the fact that vision is still the dominant sense through which we perceive the world, the visuals may make people neglect some changes in sounds and smells within the environment [7]. Therefore, the differing matched or mismatched audio made no difference to participants' perception of the Visual elements in the audio-visual composites.

In this study, the actually perceived sight and sound, and the imaged smell collectively construct the more comprehensive context of the scene. Its data and findings can further be applied not only to sensory cultural heritage, but also to other fields such as virtual tourism, museum experience design, and marketing. The emphasis on sound, smell, and multisensory interactions is relatively limited in these areas, but in fact, they influence how people construct scenes and cityscapes, and how they evaluate the pleasantness of these environments. Managers and designers can use this data to create multisensory scenes or transform existing cityscapes to maximise pleasant experiences. On the other hand, it also provides tourists and users with information on perceiving the degree of satisfaction of the destination. In addition, the study will also contribute to the improvement of some research methods, such as sensory walking. We used on-site capturing to document and understand people's responses to the audio, visual, smell, and multisensory stimuli in cityscapes. This may help reduce the impact of continuous movement on participants' experiences and responses in sense walking experiments [22, 25].

4.2. Intangible Cultural Heritage in Hong Kong

Hong Kong's traditional lifestyle has been greatly changed in the rapid urbanisation process, resulting in the destruction of intangible cultural heritage (ICH) [16]. Therefore, there is a call to protect ways of life rather than just physical buildings [17]. The urban wet market is not only a place for trade and consumption, but also has gradually been connected with issues of heritage and tourism, symbolising identity and memory [5, 12, 18]. In some Asian studies, the wet market is regarded as a tourist destination and a form of ICH (e.g. [19-20]).

With the long-standing history and integration into the daily lives of residents, we regarded the wet market as Hong Kong's ICH, representing a part of Hong Kong's traditional culture and identity. We combined the 360-degree video and 3D audio recordings to document the wet market, capturing the servicescape including market layout and hygiene conditions, as well as the social activities of Hong Kong residents within the wet market, such as social exchanges and interactions and language usage. In particular, we included the sense of smell, which is rarely explored in Hong Kong's cultural heritage research, and found that it can enhance people's perception and understanding of the urban market and Hong Kong's local culture. The findings of this study align with some previous studies on the wet market, affirming their cultural and social significance [12, 21].

In addition, the study also presented a viable multisensory approach including sight, sound, and smell to cultural heritage research, which can be applied to record other cityscapes in Hong Kong and the corresponding lifestyles and cultures, such as temples and food streets, thus further promoting the preservation and research of Hong Kong's ICH.

4.3. Urban Studies in Hong Kong

With the increasing research on non-visual senses, soundscape and smellscape are now gradually being applied in urban study and design [6, 22, 23]. Urban sounds encompass numerous "complex and ever-changing small scenes" that are intricately intertwined with residents' daily lives but are often ignored [23]. The Hong Kong wet market is an integral part of the residential infrastructure and also a reflection of Hong Kong's administrative and management system [18]. We documented the soundscape in a Hong Kong wet market as a social space where residents engage in various social activities. This allowed us to observe the everyday diversity within Hong Kong communities [21]. The smellscape in this study also provided insights into the categories of foods and issues related to health and food hygiene of Hong Kong's wet market [18].

5. LIMITATIONS AND FUTURE WORK

The findings are limited in terms of stimuli and survey design. On the one hand, the 90-second duration of an audiovisual clip stimulus was relatively long, affecting participants' willingness to complete and submit the questionnaire, as well as their engagement with clips. In addition, despite incorporating the smell in the study, there were no actual smell stimuli used in the survey, thus hindering a deeper exploration of the smellscape and its interaction with other senses. On the other hand, snowball sampling limited the inclusion of target participants from Hong Kong. Although many participants from Mainland China in the study work or study in Hong Kong, they may still have different understanding and identification with Hong Kong's culture. Furthermore, as mentioned in the Results and Discussion sections, the wording of some questions in evaluating the environmental quality may influence participants' attention to a specific sensory perception at the time of responding, thereby impacting the data analysis to some extent.

In the next step of the present research, we are conducting a controlled laboratory experiment with actual smell stimuli to reproduce the environments of Hong Kong's cityscapes in the laboratory through VR video, surround sound, and smell compounds. The grouping of scenes according to a priori valence assumption and the results of imagined smells will be further examined referring to the hedonic tone and descriptor of smells [24]. Also, the duration of each audiovisual clip stimulus will be shortened to 30 seconds, considering the time required to complete the survey in this study. In addition, for sampling, more participants who are from Hong Kong or speak Cantonese at home will be included, and a Chinese-language version of the online questionnaire will be used in parallel with the English. How participants will be asked about their environmental perceptions should also be considered to minimise the potential impact of question phrasing on their feelings and experiences.

6. CONCLUSION

The soundscape and smellscape play significant roles in perceiving cityscapes, preserving intangible cultural heritage, and fostering cultural identity. However, there is limited research on them specifically focused on Hong Kong. We captured and documented a Hong Kong wet market, Shek Kip Mei market, as a reflection of the residents' way of life and culture. By asking participants to view 12 captured audiovisual clips with matched or mismatched audio modes in an online survey, we investigated the impact of scenes from six market stalls, audio modes, and a priori valence groups on the perceptual ratings of whole, auditory, and visual environmental quality. The results show that the influence of scenes and a priori valence groups was more significant, while the effect of audio elements was minimal when participants focused on the whole environment, visual elements, and imagination of smells. Nevertheless, even in the mismatched audio-visual composites, participants were able to recognise contextual information about the scenes and evaluate the environmental quality through multisensory interactions. In the second part of the study, we will extend the multisensory approach and include real smell stimuli in a laboratory setting. The study contributes to exploring multisensory perceptions of environmental quality within the sensory cultural heritage research, and providing new insights into the intangible cultural heritage and urban studies in Hong Kong.

ACKNOWLEDGEMENTS

The work presented in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China to PerMagnus Lindborg (GRF #11605622 – CityU project #9043455).

REFERENCES

1. P. Yelmi. Protecting contemporary cultural soundscapes as intangible cultural heritage: Sounds of Istanbul. *International Journal of Heritage Studies,* **22(4)**, 302–311, 2016.
2. J. Skrede and B. Andersen. Remembering and reconfiguring industrial heritage: The case of the digester in Moss, Norway. *Landscape Research*, **46(3)**, 403–416, 2021.
3. L. Davis and L. Thys-Şenocak. Heritage and scent: Research and exhibition of Istanbul's changing smellscapes. *International Journal of Heritage Studies,* **23(8)**, 723–741, 2017.
4. M. Parker, D. H. R. Spennemann, and J. Bond. Sensory perception in cultural studies—A review of sensorial and multisensorial heritage. *The Senses and Society*, 1–31, 2023.
5. P. Lindborg, F. Aletta, K. M. Liew, Y. Matsuda, J. L. Xiao, R. Yue, L. H. Lam, and L. S. Wei. Multimodal Hong Kong: Understanding the intangible cultural heritage through sight, sound, and smell. Proceedings of Inter-noise, Nantes, France, 2024.
6. J. Xiao, F. Aletta, A. Radicchi, K. McLean, L. E. Shiner, and C. Verbeek. Recent advances in smellscape research for the built environment. *Frontiers in Psychology*, **12**, 1–7, 2021.
7. M. Ba, Z. Li, and J. Kang. The multisensory environmental evaluations of sound and odour in urban public open spaces. *Environment and Planning B: Urban Analytics and City Science*, **50(7)**, 1759–1774, 2023.
8. International Organization for Standardization. *ISO 12913-1:2014 Acoustics - Soundscape - Part 1: Definition and Conceptual Framework*. 2014.
9. S. L. Dumyahn and B. C. Pijanowski. Soundscape conservation. *Landscape Ecology*, **26(9)**, 1327–1344, 2011.
10. C. Bartalucci and S. Luzzi. The soundscape in cultural heritage. *IOP Conference Series: Materials Science and Engineering*, **949(1)**, 1–9, 2020.
11. J. D. Porteous. Smellscape. *Progress in Physical Geography,* **9(3)**, 356–378, 1985.
12. P. Lindborg and K. Liew. Real and imagined smellscapes. *Frontiers in Psychology*, **12**, 1–19, 2021.

13. J. Skrede and B. Andersen. Remembering and reconfiguring industrial heritage: The case of the digester in Moss, Norway. *Landscape Research*, **46(3)**, 403–416, 2021.
14. C. Spence. Using ambient scent to enhance well-being in the multisensory built environment. *Frontiers in Psychology,* **11**, 1–19, 2020.
15. S. Wang. Museum as a sensory space: A discussion of communication effect of multi-senses in Taizhou Museum. *Sustainability,* **12(7)**, 1–19, 2020.
16. K. E. Kuah and Z. Liu. Intangible cultural heritage in contemporary China and Hong Kong: An introductory overview. In *Intangible cultural heritage in contemporary China: The participation of local communities*, pages 1–10, Routledge, 2016.
17. S. Veg. Cultural heritage in Hong Kong, the rise of activism and the contradictions of identity. *China Perspectives*, **2007(2)**, 1–3, 2007.
18. V. Chan. Markets made modular: Constructing the modern "wet" market in Hong Kong's public housing estates, 1969–1975. *Urban History*, **50(4)**, 799–817, 2023.
19. S. M. Shariff, M. S. M. Zahari, M. H. Hanafiah, and N. Ishak. Traditional gastronomy knowledge transfers among Malay women: An exploratory study. *Journal of Foodservice Business Research*, **25(3)**, 277–301, 2022.
20. E. Park, K. Muangasame, and S. Kim. "We and our stories": Constructing food experiences in a UNESCO gastronomy city. *Tourism Geographies*, **25(2–3)**, 572–593, 2023.
21. C. Mele, M. Ng, and M. B. Chim. Urban markets as a "corrective" to advanced urbanism: The social space of wet markets in contemporary Singapore. *Urban Studies*, **52(1)**, 103–120, 2015.
22. N. Bruce, J. Condie, V. Henshaw, and S. R. Payne. Analysing olfactory and auditory sensescapes in English cities: Sensory expectation and urban environmental perception. *Ambiances*, 1–14, 2015.
23. R. Yue, Q. Meng, D. Yang, Y. Wu, F. Liu, and W. Yan. A visualized soundscape prediction model for design processes in urban parks. *Build. Simul.*, **16(3)**, 337–356, 2023.
24. A. Dravnieks, T. Masurat, and R. A. Lamm. Hedonics of odors and odor descriptors. *Journal of the Air Pollution Control Association*, **34(7)**, 752–755, 1984.
25. R. Yue, P. Lindborg, R. Ikeshiro, and R. Gui. Multimodal Hong Kong: Perceptions of a pedestrian footbridge in Mong Kok. Proceedings of Inter-noise, Nantes, France, 2024.