Audio Engineering Society

Conference Paper

Presented at the International Conference on Spatial & Immersive Audio,

2023 August 23-25, Huddersfield, UK

Exploring experience and expertise in the context of preference for spatial sound treatment in immersive environments

Angela McArthur[1], Andrew Johnston[2], Mark Sandler[3], Sam Ferguson[2], Rebecca Stewart[4]

[1]University College London
[2]University of Technology Sydney
[3]Queen Mary University of London
[4]Imperial College London

Correspondence should be addressed to Angela McArthur (a.mcarthur@ucl.ac.uk)

ABSTRACT

As experiences for immersive environments (IEs) mature, levels of audience experience are increasingly differentiated. This is complicated by their prior aesthetic experiences and expertise. When creating content for IEs, spatial sound practitioners may employ too much, or too little, complexity in their sound design. This potentially risks plausibility, immersion, or engagement. In this study, short clips of ecologically valid cinematic content - some containing sound-image distance displacements, some not - are examined for two kinds of IEs: VR, and a CAVE. Audience expertise and experience is considered, alongside aesthetic tastes, ideas and behaviors. A Repertory Grid Technique-derived process elicits free verbalizations in response to perceived similarities or differences in the sound between stimuli. The elicited verbal constructs are subsequently used to subjectively explore participants' preferences for treatment of sound in the IEs. From these preferences, a further interview identifies super-ordinate constructs which explore participant priorities for immersive experience, beyond the experimental conditions here. This approach focuses on participant-specific variables and subjective experience, to be of applied use to practitioners in the field. Three groups of experience and expertise levels are identified from self-reported measures. Participant priorities for spatial sound treatment are considered by experience and expertise, and grouped into themes. Themes were consistent across both IEs (VR and a CAVE). There may be a 'hierarchy' which indicates salience within these themes, based on the frequency with which they were identified by participants. This hierarchy differs somewhat for expert and non-expert audiences, though certain themes cut across groups (a desire for realism, a desire for engagement/ connection/ understanding, and a desire for balance between cognitive activity/ passivity). An overall trend in experience and expertise is found. Moderately experienced/ expert audiences have less definite preferences for spatial sound treatment in IEs. The least and most experienced audiences, express clearer preferences. These two groups often agree with each other about the priorities for the immersive experience. A potential explanation for these findings is speculatively explored, in which the 'inverted U-curve '(which has been used to explain liking and preference in related fields) is modified, and where sound-image distance displacements are considered to increase stimuli complexity.

1 Introduction

Aesthetic experience is dynamic, changing over larger or smaller temporal scales both within and across people. It resists attempts to be quantitatively deconstructed. Berlyne [1], [2] developed an idea of Wundt's (1872–1912), that there is an optimal level in the relationship between arousal and complexity, and that this optimal point related to optimal aesthetic pleasure. Berlyne proposed that this relationship varied systematically, and could be regarded as an inverted U-shaped curve, with arousal on one axis and 'hedonic value' on the other, as in Fig. 1. This simple idea offers an important concept in terms of subjective aesthetic experience (preference and liking) and stimulus (complexity). For Berlyne, the tension of opposites inherent in the curve derived from two adaptive mechanisms - a reward system activated by initial increases in arousal (so seeks them) and an avoidance system activated *beyond* a certain level of arousal.

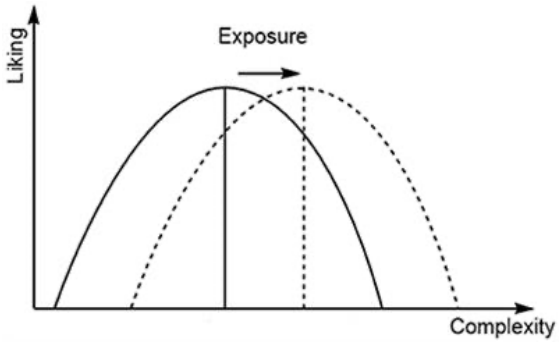

Figure 1. The Inverted U-Curve or Wundt Curve [3]

The field of empirical aesthetics has contributed much to aesthetics, though by focusing on the cognitive aspects of art appreciation it largely ignores (by necessity) phenomenological aspects or 'aesthetic as experience'. To create a participant-led design which does not narrow the range of experience captured for (novel) immersive media, this study draws from the experimental designs of empirical aesthetics by using a bespoke 'aesthetic questionnaire' including some items (some are adapted) from the Vienna Art Interest and Art Knowledge Questionnaire (VAIAK) [4]. This test, developed collaboratively by academics from psychology and art history [5] tests for art expertise and interest. The adaptation of the VAIAK extends its use to works which are not exclusively paintings (which it was developed for). The study also uses an extended form of the Repertory Grid Technique (RGT). The RGT was originally conceived of by Kelly, as part of personal construct theory [6] and the extension of it used here, the 'laddering interview', has been outlined elsewhere [7]. The RGT provides a structured means of eliciting free verbalizations, thus reducing researcher-led biases. It offers participants opportunities to reflect during trials. Participants can offer as many bipolar constructs as they wish. If they do not freely offer both poles of a construct, they are prompted to do so. They are encouraged to use their own vocabulary and reassured that there are no 'wrong' answers. This helps reduce performance pressure and increases ecological validity further.

A full range of demographic data is collected to establish experience and expertise. Aesthetic experience is compared across two IEs – 3DoF cinematic VR, and a 360º immersive facility at the University of Technology Sydney, the 'Data Arena' (DA) with a 4m-high, 360-degree high-resolution cylindrical screen (formed with 6 projectors), and 16 loudspeakers (14 + 2 sub-bass speakers) fitted at head-height behind a perforated screen (see Fig. 2, [8], [9] for more information, an animated visualisation and specifications).

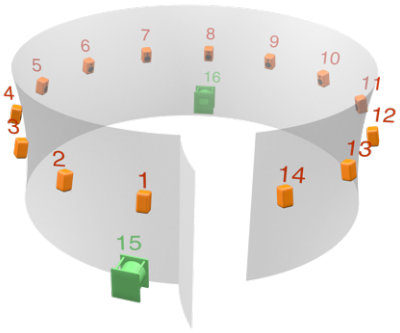

Figure 2. Data Arena Audio Channel Configuration [9]

The stimuli comprised clips taken from a cinematic VR work 'Rumpus' (original idea, production sound, production 360º video was produced by the BBC R&D North Team's Alia Sheikh and Tom Nixon/ post-production sound by the lead author) which was shown at Ars Electronica 2019 in Linz, Austria, in VR format. This validates it ecologically.

2 Method

Procedure: Participants complete two questionnaires prior to the study. The first screens for study suitability, the second gathers (self-reported) demographic data and information about

aesthetic tastes, viewing/ listening habits, and spatial sound training. During trials, the adapted VAIAK is administered using (mostly) a 5-point Likert scale.

Example items with minimal adaption (indicated in bold, such as changing the term ‘art' to ‘sound and image' or adding ‘music’) are:

- I enjoyed art **or music** classes at school
- I like to talk about **sound and/ or image** works with others

Items designed for this study included:

I understand ‘aesthetics in sound' to mean:

- Something that indicates good taste for most people vs Something that indicates good taste for specific people (like sub-cultures and genres)
- A subjective experience vs An objective fact
- A design / composition to give pleasure or be beautiful vs A design / composition that can ignore popular concepts of ‘the beautiful’
- A set of underlying stylistic principles that are evident to the creator but not necessarily anyone else vs A set of underlying stylistic principles that are evident to both creator and listener
- Something experienced mainly through the rational (thinking) part of me vs Something experienced mainly through the sensual (feeling) part of me
- Something that exists within the work, independently vs Something that exists between the listener and the work, as a relationship
- Something which needs to be consciously experienced vs Something that can be unconsciously experienced

After completing the adapted VAIAK, participants watch both the congruent and the incongruent versions of one clip in one IE, both congruent and incongruent versions of the second clip in the other IE (order of presentation is balanced to mitigate against effects arising from this). Immediately after each IE screening they are asked to provide verbal constructs to describe any similarities or differences relating to the sound (this constitutes the RGT bipolar construct elicitation phase).

Finally, the laddering interview is undertaken. Each participant is prompted to identify preferences from the bipolar constructs they elicited previously, given the kind of film clips they experienced in the study, and the media environments in which they experienced them. As participants articulate their preferences, super-ordinate constructs are elicited, and increasingly abstracted preferences are expressed. After the first question asking for a preference, the researcher concentrates on asking ‘why' one pole is preferable over the other. The process continues until no new constructs are offered. An example (for the construct ‘present ---- absent') of the conversation can be seen in Fig. 3.

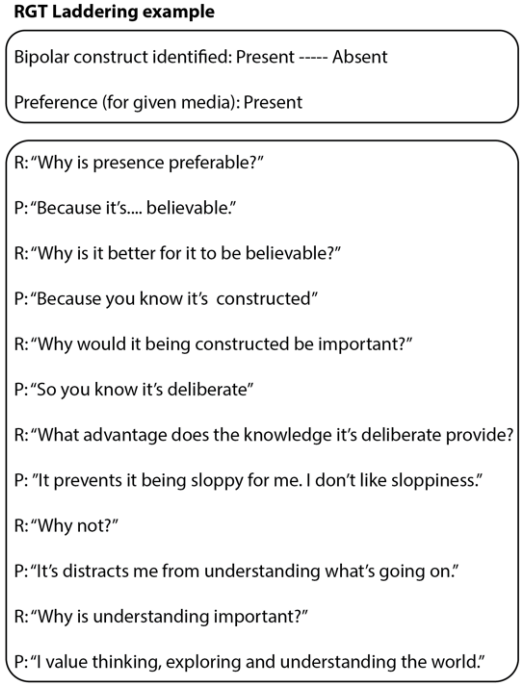

Figure 3. RGT Laddering Interview Process, R = researcher, P = participant.

Stimuli: Auditory distances broadly corresponding to two sets of image events (from the same two human actors) in two clips from ‘Rumpus’ are varied to be evidently either congruent (sound-image distances match) or incongruent (sound-image distances mismatch). The events are at varying distances, elevations and azimuths, according to the character positions in the clips. The varied sounds were recorded and rendered separately to production sound, though later combined. Production sound was up-mixed from double-mid-side to first-order Ambisonics (FOA) using the Harpex plugin [10]. Source distances were rendered with the ‘dearVR PRO' [11] plugin within Reaper [12] and Unity [13] (whether congruent or incongruent, for consistency). This plugin allows sound to be mixed within a 360º video environment using a headset, controllers, and headphones. This overcomes the need to manipulate a small, overhead element in the graphical user interface of an audio plugin within a DAW. It also allows the mix to be created in its target output environment. Sounds were rendered to be perceptually plausible/ congruent *or* implausible/ incongruent, but distances were not objectively verifiable. Sounds created with varied distance were:

Clip One
- Female protagonist footsteps (nearer in incongruous condition)
- Sound of lantern held by female protagonist (further in incongruous condition)
- Male protagonist breathing (nearer in incongruous condition)

Clip Two
- Female protagonist singing (further in incongruous condition)
- Sound of ukulele strummed by female protagonist (further in incongruous condition)
- Sound of mechanical fan (non-diegetic, nearer in incongruous condition)
- Male protagonist laughing (nearer in incongruous condition)

Clips were rendered as video files, in H.264 MPEG-4 format in a .mkv container, and 24-bit 48kHz FOA audio. Because production sound had been captured non-spatially, because distance was only varied for post-produced sound, and because the DA's reproduction system was on one horizontal plane, FOA rendering provided adequate resolution. Clips can be inspected at https://vimeo.com/838245694 (24") and https://vimeo.com/838248189 (28").

Sound design had as its primary purpose, the artistic rendering of the scene (its objects, actors and atmospheres) prioritising the emotional narratives distinct to each clip. Variables in the 8 experimental conditions are summarised:
- *Congruent vs Incongruent* -- In congruent versions of the clips, sound events are mapped to image events. In incongruent conditions, sound events are brought nearer or further than their image event counterparts (via subjectively designed distances).
- *Indoors vs Outdoors* -- 2 different locations for filming. One scene is outside in a street, one inside, in a curtain-lined tent.
- *Data Arena Vs VR* -- 2 different immersive environments for reproduction.

For VR trials a HTC Vive pro headset with Behringer BH770 closed-back headphones were used. The Vive Cinema app played back spherical video whilst decoding the audio (ambiX channel ordering) using the SADIE Binaural Measurement KU100 HRTF set [14]. For DA trials participants entered the space and could navigate it freely.

Participants

Twenty participants (N=11 female, N=9 male), ages 22 - 62 years, participated. Most (N=16) reported no eyesight difficulties. Four participants reported wearing prescription reading glasses. Where possible, participants' eyewear was worn under the VR headset during trials, however in some cases this was not feasible due to size, fogging, or discomfort. In comparison to the inherent issues of visual quality with headset image reproduction and 360° video content resolution, the participants' short sightedness was not a significant concern. Eyesight difficulties were not an issue for the DA.

3 Results

Participants were segmented into groups, based on study preparation data relating to experience and expertise. The following groups were defined:
- Sound experts / non-experts (those with formal training in music or sound, theory or practice Vs those without)
- Spatial sound experts / non-experts (those who had either studied spatial sound and/ or worked with it Vs those who had not)
- IE experts / non-experts (those who have worked with VR, AR, etc Vs those who hadn't)
- Spatial sound consumers Vs non-consumers (those who have experienced regular and/ or varied content using spatial formats)
- IE consumers / non-consumers (those who have experienced regular and/ or varied content in such environments)
- Attentive Vs inattentive (those who tended to watch **and** audition content singularly Vs those who tended to watch and/ or audition it whilst doing other things).

Participants were then split into three categories – (a) those with specific experience or expertise with sound, IEs, and attentiveness (EXP), (b) those without expertise or experience (IEXP), and (c) those without expertise who were 'particularly' inexperienced (PIEXP). Experience is not conflated with expertise in the data, but to be succinct here, we now combine the terms into simply 'experienced'. Fig. 4 highlights the 'count' of expertise or experience across measures for each participant. They were weighted (e.g. being a spatial sound expert is more important to expertise than being a spatial sound consumer).

Experience/ expertise x aesthetic interest: As we see in Fig. 5, the groups correlate well between measures of experience and of aesthetic interest, with only a few participants scoring higher or lower levels of aesthetic interest than expected based on experience. The correlation between these measures is clear, though we cannot infer a linear relationship between score values, and interest (the questions being subjective in nature). As a broad measure, the questionnaire serves to validate the previous

segmentation, and to highlight specific participants whose data may warrant future analysis.

sound expert	spatial sound expert	IE expert	spatial sound consumer	IE consumer	attentive
P1	P1	P1	P1	P1	P1
P2	P2	P2	P2	P2	P2
P3	P3	P3	P3	P3	P3
P4	P4	P4	P4	P4	P4
P5	P5	P5	P5	P5	P5
P6	P6	P6	P6	P6	P6
P7	P7	P7	P7	P7	P7
P8	P8	P8	P8	P8	P8
P9	P9	P9	P9	P9	P9
P10	P10	P10	P10	P10	P10
P11	P11	P11	P11	P11	P11
P12	P12	P12	P12	P12	P12
P13	P13	P13	P13	P13	P13
P14	P14	P14	P14	P14	P14
P15	P15	P15	P15	P15	P15
P16	P16	P16	P16	P16	P16
P17	P17	P17	P17	P17	P17
P18	P18	P18	P18	P18	P18
P19	P19	P19	P19	P19	P19
P20	P20	P20	P20	P20	P20

Figure 4. Participant segments - participants with expertise/ experience highlighted. Bright yellow = EXP, faded yellow = IEXP, no color = PIEXP

Number of constructs initially offered: The EXP group averages 6.57, with a range of 4 - 11 and a median of 7. The PIEXP group is distinct in their reduced number of initial constructs. The average number of constructs is 3.5, with a range of 2 - 6 and a median of 3, about half that of the EXP group. The IEXP group has an average of 5.66 initial constructs, a range of 3 - 8 and a median of 6. This group seems to be placed between the two 'extreme' groups.

Contextual preferences: Where a clear preference for a pole (of a given construct) was not provided, preference appeared to be contextual. The EXP group offered five such constructs out of 46 (about 1%). The (smallest in size) PIEXP group offered 4/14 such constructs, 29% of which half came from one participant (P17) who offered only 2 constructs in total. The IEXP group offered 21/51 such constructs - a huge 41% of all its constructs. Two participants contributed a lot here - P16 offered 6/ 9 such constructs, and all of P14's 3 constructs were contextually preferential.

Super-ordinate constructs: Super-ordinate constructs are more suitable for abstraction than initial constructs (which relate more specifically to the stimuli). The laddering interview thus progressed the initial constructs significantly, with as minimal interpretation as possible so as to preserve original meanings. Consequently, super-ordinate constructs were treated as separate and self-contained for such analyses. Example of super-ordinate constructs (for P12) are shown in Table 1, where we see how some of the constructs did not progress beyond even 1st order during the interview, how some repeat thematically (or literally) and how free verbalization leads to gaps in the data.

1st order super-ordinate	2nd order	3rd order	4th order	5th order
Engaging				
More complete	Engaging			
You feel it, you're not analysing it	Not 'everyday' experience	You're in the moment/ you're present	Pleasure, sensual	
Made me more curious	Allows for your own narrative/ imagination	Allows the brain to work	Discorda -nt	Pleasu -rable
More elements to the narrative	Stronger feeling of atmosphere			
More interesting	Subtle			
Feels more complete, more complex, fatter				
Sensual	My body feels it's participating, is there, isn't detached			

Table 1. Super-ordinate constructs by order for P12, EXP group (duplicate constructs = a parent construct led to multiple child constructs)

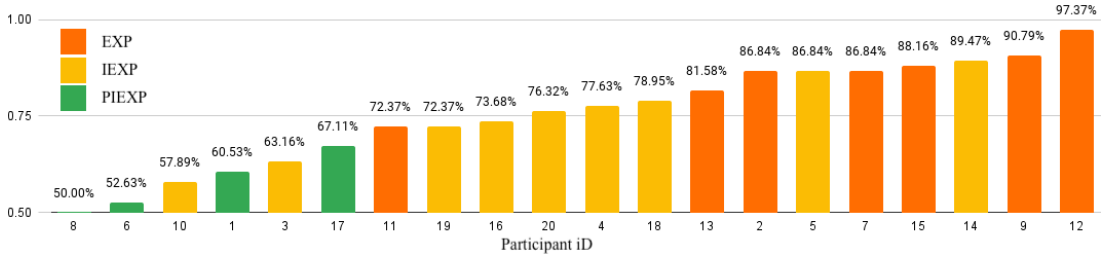

Figure 5. Aesthetic Questionnaire Scores (as percentages on y axis) x Experience/ expertise segmentation (by participant, on x axis)

Super-ordinate themes: Counting super-ordinate constructs thematically, we can infer individual participant themes. Themes are based on the subjective interpretation of the subjective reports of participants, so the following numerical data should not be treated objectively. It offers a guide to emergent patterns in free verbalizations, a bridging mechanism: not fully subjective nor objective, yet of practical use in creating work for IEs. By subsequently interpreting and combining participants' super-ordinate themes, we can consider themes across all participants (in some cases, individual themes did not correspond to an aggregate theme at all). This is done by interpreting and counting thematic occurrences. In the list below, count is provided in brackets for each theme. Where letters are appended to numbers in the list (e.g. 3a, 3b) the same count of occurrence is designated:

1. The idea of realism is generally preferred (24)
2. Engagement, connection, and understanding are important aspects of the experience (21)
3a. There is a 'sweet spot' in terms of the balance between cognitive activity and passivity, for an optimal experience (sensual and cognitive engagement may compete for resources (potentially attentional resources) (13)
3b. Expectation is an important component of the immersive experience (13)
4a. Immersion can heighten / enhance the sensual and/ or emotional experience (sound contributes greatly in this regard) (11)
4b. Spatial sound can contribute to narrative understanding, whether active (through participant exploration) or passive (through attentional cueing) (11)
5a. Sound-image displacements are preferred when the intention for the displacement is made clear (9)
5b. Spatial sound can contribute to a sense of space, atmosphere, and being placed 'in' that space (9)
6. Clear, close sound is generally preferred, whether realistic or not (8)
7. Participants want to be immersed (7)
8. Immersion can overwhelm participants (sensually, leading to an emotional response of them feeling unsafe) (6)
9. Participants want to experience well-conceived and realized work (4)
10. Immersion can transport participants to other worlds (3)

From counts, we can establish a sense of priority or 'hierarchy'. This is limited due to variations across participants (some are more verbose and use a term frequently, skewing results). However, accounting for all 20 participants, we somewhat mitigate against this impact. To underline the subjectivity of these analyses, we can consider theme 1 and 6. Theme 6 was counted 2/3 less than theme 1, but expresses a clear set of priorities for participants' needs. Without clear sound, the immersion is more threatened, the narrative less clear and intelligible, the experience is less engaging, and requires more cognitive effort. Yet clear sound is often not realistic, causing it to apparently conflict with theme 1 (the desire for realism).

Hierarchy of themes by experience: The EXP group, has a count of 51 (an average of 7.3 themes per participant). Interestingly, all 13 of the original themes are present, showing a level of variation in the aspects of the experience articulated. This group seemed more focused on the sensual and physical properties of the experiences, and expressed the least potential overwhelm. For our largest group (IEXP, 9 participants) there was a slight reduction in the variety of themes overall, with 11 of the original 13 represented. This group has a count of 68 (an average of 7.6 themes per participant). Themes 9 and 10 were not represented in this group. These themes were the lowest ranking in any case, but here they are completely absent. The PIEXP group has a count of just 20, an average of 5 themes per participant. This average is not particularly meaningful. One participant (P17) only contributed one individual theme, and two participants (P6 and P8) contributed individual themes that counted for only three of the overall themes. In such a small group, individual variations are a more appropriate measure than averages. 11 themes are represented, which is the same level of variety as the IEXP group (which is more than double the size). This shows the consistency and validity of the themes themselves. However, we should recall that participants were able to express as many constructs as they wished.

An overall pattern in the interaction between experience and priority for the themes is evident. In 7/13 themes, *both* the most and least experienced participants expressed a given theme was either a high *or* low priority for them. A multitude of patterns for this interaction could exist, so to have one of these expressed in about half the themes is striking. In 2/7 of the themes where this pattern existed, it was more mildly expressed. Our moderately experienced participants expressed these themes less frequently so (we infer) are less of a priority for them. The themes that did *not* conform to this pattern, show either consistently high priority across all groups, or are highest priority for *either* our most or least experienced groups:

Priority for all:

- The idea of realism is generally preferred
- Engagement, connection, and understanding are important aspects of the experience

Priority for most experienced participants:
- Spatial sound can contribute to a sense of space, atmosphere, and being placed 'in' that space

Priority for least experienced participants:
- There is a 'sweet spot' in terms of the balance between cognitive activity and passivity, for an optimal experience (sensual and cognitive engagement may compete for resources (potentially attentional resources)
- Participants want to be immersed
- Immersion can overwhelm participants (sensually, leading to an emotional response of them feeling unsafe)

Priority for most and least experienced participants:
- Immersion can heighten / enhance the sensual and/ or emotional experience (sound contributes
- greatly in this regard)
- Participants want to experience well-conceived and realised work
- Immersion can transport participants to other 'worlds'

Least priority for most and least experienced participants:
- Expectation is an important component of the immersive experience
- Spatial sound can contribute to narrative understanding, whether active (through participant exploration) or passive (through attentional cueing)
- Sound-image displacements are preferred when the intention for the displacement is made clear
- Clear, close sound is generally preferred, whether realistic or not

4 Discussion

Number of constructs initially offered: The most experienced group offered the highest number of constructs and the highest number and variety of themes. The least experience group offered the fewest of each. Experts may have a higher confidence, or overall ability to articulate themselves in an area with which they are familiar. Perhaps they felt less overwhelmed and were able to consciously process more of their experience. This would have helped them subsequently verbalise it. This does not mean they would have a lesser aesthetic response -

the most profound experiences are not ones that we have but rather those that have us, overwhelming the experiencer to the point where one cannot properly talk about appreciative knowledge of the experience [15, p. 88].

An inability to verbalise may indicate *more* impact, though we need to exercise caution about whether this correlates with a more intense *aesthetic* impact.

The moderately experienced group showed most ambivalence in their preferences. The most and least experienced groups were more definite about which pole of their constructs was preferable for spatial sound in IEs.

Hierarchy of themes by experience: We now consider a speculative application of the inverted U-curve. Usually, the curve is applied to one listener for various aesthetic works. Here, we apply it for one aesthetic work, and three varieties of listener with different experience levels (see Fig. 6). Incongruent sound-image relationships are more complex than congruent ones (our perceptual processes must work to overcome apparently conflicting information). So we see increasing sound-image displacements (in either number or intensity of displacement) represented on the x axis.

The theme of engagement was important for all participants and is represented on the vertical axis. Verifying that this is the correct term to use here would require further investigation. Presently, we can say is that something *like* engagement (motivated action to engage, whether out of necessity or choice) is represented on this axis. We see how the moderately experienced group has most engagement, due to an intermediate level of effort required on their part. Very inexperienced audiences may be overwhelmed by the effort of decoding aesthetics in IEs and consequently may disengage. Experienced participants may struggle to turn off the analytical cognition which mediates affective experience, and thus not be able to engage.

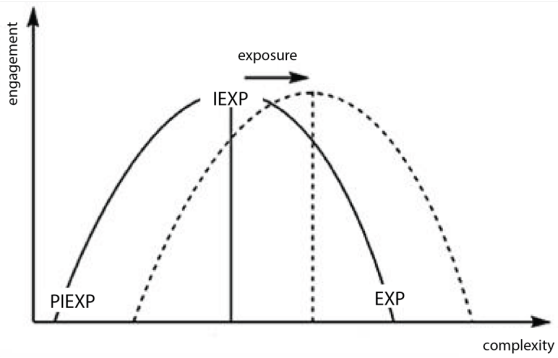

Figure 6. The inverted U-curve for a single aesthetic work and multiple audiences (PIEXP = particularly inexperienced, IEXP = inexperienced, EXP = experienced)

This overall trend in an 'optimal experience' level translates via the speculative inverted U-curve as a hypothesis for future investigation.

5 Conclusion

An overall trend in expertise is found. Moderately experienced participants show less definite preferences, whereas the 'extremes' of experience and expertise are unambiguous about their responses, and often express the same themes as priority.

The 'inverted U-curve' describes the relationship between complexity and liking, based on exposure. We assume sound-image incongruency to be more complex than congruency. We can also assume that exposure to IEs is limited at present. How quickly exposure effects take place for distance (or other) displacements in IEs is unknown, but they likely *do* occur. With experience, audiences will become more sensitive to the aesthetic elements of work [16], they will operate at higher levels of interpretation [17] and will be more clear about how to understand work. For now, designing more complexity into spatial sound treatment in IEs needs to be approached carefully. Establishing clear sound design principles could help establish audience expectations, and a period of in-experience exposure ('training') may help with this, if we are to engage audiences across a range of experience levels. Our sense of aesthetic experience is only hurt when violations in our expectations cannot be understood [18]. Practitioners can creatively violate expectations through the treatment of spatial sound in IEs, with this proviso in mind.

References

[1] D. E. Berlyne, *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts, 1971.

[2] D. E. Berlyne, "Conflict, arousal, and curiosity.," *Conflict, arousal, Curiosit.*, Oct. 1960.

[3] G. Madison and G. Schiölde, "Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music," *Front. Neurosci.*, vol. 11, no. MAR, p. 147, 2017.

[4] M. Pelowski, P. S. Markey, M. Forster, G. Gerger, and H. Leder, "Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates," *Phys. Life Rev.*, vol. 21, no. February, pp. 80–125, 2017.

[5] E. Specker *et al.*, "The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A Unified and Validated Measure of Art Interest and Art Knowledge," *Psychol. Aesthetics, Creat. Arts*, no. October, 2018.

[6] G. A. Kelly, *The psychology of personal constructs. Vol. 1. A theory of personality. Vol. 2. Clinical diagnosis and psychotherapy.* 1955.

[7] F. Fransella, R. Bell, and D. (Donald) Bannister, *A manual for repertory grid technique*, Second. West Sussex, UK: John Wiley {\&} Sons, 2004.

[8] "About — UTS Data Arena." [Online]. Available: https://dataarena.net/about. [Accessed: 16-Jun-2023].

[9] "Tech Specs — Wiki — UTS Data Arena." [Online]. Available: https://dataarena.net/dive-in/wiki/tech-specs. [Accessed: 16-Jun-2023].

[10] Harpex Ltd, "Harpex - Documentation." 2017.

[11] "dearVR PRO - Binaural, Ambisonics & Multi-Channel Spatializer - Dear Reality." [Online]. Available: https://www.dear-reality.com/products/dearvr-pro. [Accessed: 16-Jun-2023].

[12] "REAPER | Audio Production Without Limits." [Online]. Available: https://www.cockos.com/reaper/. [Accessed: 16-Jun-2023].

[13] "Get Your Unity Pro Subscription Today | Unity." [Online]. Available: https://unity.com/pages/unity-pro-buy-now. [Accessed: 16-Jun-2023].

[14] "SADIE | Spatial Audio For Domestic Interactive Entertainment." [Online]. Available: https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html. [Accessed: 27-May-2021].

[15] R. Shusterman and A. Tomlin, *Aesthetic experience*. Routledge, 2008.

[16] A. Goldman, "Evaluating Art," in *The Blackwell Guide to Aesthetics*, P. Kivy, Ed. Blackwell, 2004.

[17] C. F. Nodine, P. J. Locher, and E. A. Krupinski, "The Role of Formal Art Training on Perception and Aesthetic Judgment of Art Compositions," *Leonardo*, vol. 26, no. 3, p. 219, 1993.

[18] L. B. Meyer, *Emotion and meaning in music*. University of Chicago Press, 1956.