9-12 AUGUST 2015
inter.noise 2015
SAN FRANCISCO, CALIFORNIA USA

Cognition of soundscapes and other complex acoustic scenes

William J. Davies[a)]
Acoustics Research Centre, University of Salford
Newton Building, The Crescent, Salford M5 4WT, UK

The fields of soundscapes, music cognition and audio quality have many similarities. Researchers in these areas are all dealing with perception and cognition of complex acoustic scenes. To date, there has been little cross-fertilisation between them. This paper examines some key concepts and results from soundscapes, music and audio. It is shown that perceptual dimensions, categories, and figure/ground play important roles in all three areas. The concept of the scale of the cognitive structure is introduced, building on results in music cognition. Scale refers to the way we can attend to the soundscape as a whole, or zoom in to a sound within it, or further in to a component or feature of that sound. It is suggested that scale links the models of perceptual dimensions, categories and auditory objects. The idea of scale is further used to suggest why the attention mechanism is so important in complex scenes, and why the concept of listening modes may be more simply explained as the consequences of attention. The paper concludes by speculating on the inherent cognitive apparatus applied to all complex scenes and suggests fruitful avenues for future research.

1 INTRODUCTION

As I write this paper, I'm sitting in the University library staring out of the window at the park, which is full of trees. This is satisfying partly because I can choose to look at it at different scales. I can look globally at the whole thing – how pleasant the different shades of green are! Or I can focus on a specific tree – what a nice explosion of blossom that one has! Or (putting my glasses on) I can zoom further in to a specific leaf – one on the end of a branch catches my attention as it bobs in the breeze. If I were close enough, I could examine the veins in the leaf. That's four different scales in one scene. All but the simplest visual scenes have this feature of

[a)] email: w.davies@salford.ac.uk

multiple scales and we are very used to zooming in and out of them as we navigate our surroundings (or stare idly out of the window). Does the same thing happen with sound and auditory scenes? A moment's reflection shows that it does. Imagine you are listening to some music. You could pay attention to the shape or mood of the whole piece, focus on the melody or the words, listen to a specific instrument, or look out for a particular catchy phrase or rhythm. If you're a musician, you might zoom in further to examine the way the bass player managed that tricky note. As with the visual scene of the park, three or four levels of scale exist when listening to music. It seems clear that this kind of structural scale is an important feature of how we listen to music. But it's not just music. Scale is also very apparent in the way we process our everyday soundscapes and in the specific field of audio quality. These three research fields – music cognition, soundscapes and audio quality – are all studying specific examples of the same thing: how we listen to complex acoustic scenes.

The main idea of this paper is to show how scale is an underlying concept that ties together and explains several aspects of the way we listen to complex acoustic scenes.

The specific contributions of this paper are to:

- Show how the three research areas of soundscapes, music cognition and audio quality are closely related. Results and concepts from each can be used to help solve problems in the others. Findings on structural scale from music have implications for soundscape research (section 3.1). Object-based representations in audio reproduction could be very useful for soundscape work (section 3.5).
- Show how structural scale underlies the familiar concepts of soundscape dimensions and categories (section 3.2).
- Show how scale is related to attention (section 3.3). I suggest that attention and comparison are the underlying mechanisms used in processing all complex scenes (section 3.6).
- Show that the concept of listening mode may be explained more simply as the work of attention across different scales (section 3.4).

2 SOUNDSCAPES AND COMPLEX ACOUSTIC SCENES

Soundscape research often features locations like urban squares and parks. These have complex acoustic scenes and the acoustic scenes have several levels of structural scale, like music. For example, consider the soundscape of an urban square. In the urban square we might listen to the overall scene and gain a global impression – perhaps it is calming, or vibrant, etc. Or we might notice that the sound made by people is dominant in the soundscape. We might zoom in a little further to listen to the hubbub of speech (perhaps excluding the footsteps and other people sounds). We could zoom further in to isolate a specific conversation. And further still to an attribute of one speaker; her regional accent, perhaps. All these sounds and features have previously been identified as components of an urban soundscape in the research literature. But soundscape research tends to focus on one level of scale, usually on the top level, the overall impression of the whole soundscape. In our everyday life it would be unusual to keep our attention at one scale all the time. Instead we tend to zoom in and out of the different scales. Depending on our activity, this can be done rapidly. For example, when we listen to a new soundscape for the first time, then zooming in or out once a second does not feel uncomfortably fast. This is one problem that this paper addresses: how does scale work in soundscapes?

A second problem I want to address is really a solution in disguise. This is that essentially the same topic is being studied by several fields, but communication between the fields seems to be limited. Soundscape research studies the topic of how people process the complex acoustic scenes that we find in our everyday environments. The main characteristics of this topic include multiple complex sources, variation with time and space, complex environmental filtering effects, listener expectation, emotional state and general context. So we have a complex listener listening to a complex acoustic scene. But this description could also be applied to research in music cognition, a field that seeks to understand what happens in our heads when we listen to music. Yet a third field also has the same general problem. Researchers in audio quality are trying to understand how to improve the perceived quality of reproduced sound. This field has recently moved away from assessments of low-level objective metrics like signal distortion towards a comprehensive quality model that includes a complex signal being altered by a complex spatial reproduction system and evaluated by a listener with a context, emotional state and so on. For example, imagine that you are listening to a stereo recording of music over loudspeakers in a quiet room. How should the quality of the audio be assessed? You could focus on a specific detail – the quality of the bass drum sound, for example. Or you could zoom out a bit to evaluate the spatial separation – are the foreground sources rendered convincingly in front of you? Or you could zoom all the way out to evaluate the overall quality of the audio system, perhaps by comparison with another reproduction system. It seems that structural scale and zooming are also features of audio quality assessment.

In this paper I want to argue that soundscapes, music cognition and audio quality assessment are all particular cases of the same general problem: that of human cognition of complex acoustic scenes. When we look at the problem in this more general way, it is easier to see what the important common features are. Concepts familiar to soundscape researchers, like perceptual dimensions and category systems are two examples of useful common features that have meaning across the general problem. But the idea of structural scale underlies both dimensions and categories and is underexplored.

3 FEATURES OF COMPLEX SCENES

3.1 Scale

Perceptual scale in a visual scene seems at first sight to be strongly related to physical distance and size. In the visual scene of the park, we moved from looking at the whole park to the veins on one leaf, across four scales A simpler example was studied by the Gestalt psychologists: the human figure standing in a landscape.[1] Our attention is drawn to the figure in the foreground and away from the landscape in the background. The Gestalt psychologists understood that this is not simply a matter of the figure being closer to the observer. The person standing in front of us is salient. It catches our attention because it has the right size and shape to do so. We pay less attention to the bush that is just as close to us; the bush is part of the landscape. So foreground and background have a dual meaning. They can refer simply to spatial distance from the observer. Or they can refer to importance or salience. The cognitive representation of scale is not a simple plotting of distance and size. Objects are grouped and assigned to a scale level partly on their importance to our understanding of the scene. This grouping is not fixed and can adapt to the content. If we put a single human figure in front of the visual scene of the park, then the cognitive representation may collapse into just two levels of scale. These are the figure and the background. The figure/ground duality is the simplest system of scale, with two levels of scale and often just one component at each level (one figure in one landscape).

What is the experimental evidence that scale is a feature in soundscapes and other complex acoustic scenes? Schafer and his acoustic ecology colleagues described a soundscape in a similar way to the Gestalt figure/ground model, with the terms keynote (background), sound signals (foreground) and soundmark (unique identifier of the soundscape).[2] In a review of over 500 soundscape papers, Payne et al. found that a two-scale model was the most common, consisting of the elements of sound and soundscape. They note that the parsing into foreground and background may vary with the individual perceiver.[3] The positive soundscape project found that people soundwalking in Manchester and London described what they could hear mainly at the scale of individual sounds, but they also talked about the larger scale of the soundscape, and the smaller scale of features or components of an individual sound source.[4]

Scale is also seen in studies of audio quality. Several researchers have explicitly addressed the question of how the scale of the reproduced audio scene should be divided up. The most common model has two scales of foreground sources and the background or whole scene.[5] An influential paper by Rumsey[6] however, proposed a comprehensive framework for describing auditory scenes, in which he used a three-scale model consisting of individual scene elements, ensembles of elements, and attributes related to the environment. An ensemble is a set of individual sources that are grouped together cognitively by a listener (i.e. orchestra, band, string section, etc.). Rumsey used a series of elicitation and listening tests to derive a set of attributes that listeners could apply to elements at each of the three scales.

However it is in the field of music cognition that we find the most developed model of scale. Several research groups have addressed the question of how the structure of a piece of music is represented internally by the listener and useful reviews are provided by Krumhansl[7] on the cognitive elements and by Purwins et al.[8] on how they can be modelled. These reviews show that music perception is built up from perceptions of the basic elements of rhythm and pitch, in a rough hierarchy of increasing complexity and scale. The elements for which neurological evidence exists are shown in Table 1. Krumhansl notes that the top level of the hierarchy continues, with more complex and longer patterns of rhythms and melodies, sometimes featuring nested repetition. Listeners can organise complex compositions into many levels of hierarchical structure by identifying the repetition and variation of simpler phrases. One fascinating finding is that the cognitive structure of music that is revealed by this research resembles established music notation, though it's unclear which way causation works in this respect.

Table 1 – Elements of music perception (after Purwins et al.[8]).

Pitch-based		Time-based
Melody categorisation	↑ Increasing scale	Meter (complex)
Harmony		Meter (simple)
Contours		Pulse
Intervals		Grouping
Chroma		
Tones (complex)		
Pitch (f0)		

It is interesting that music has such a broad range of scale in its hierarchy, from pitch to the shape of a whole symphony, while soundscape research has tended to concentrate at two scales: sounds and soundscapes. This might be because musical listening is special (as Gaver suggests[9]) but it might be simply because we haven't looked. The thought experiment posed earlier of conversation in a city square, suggests that it would be worth specifically researching scale in soundscapes.

3.2 Dimensions, categories and scale

People use categorisation to make sense of their environment by grouping perceptions in an efficient way. It is a strategy that is found in all sensory modalities and widely reported in many environmental contexts. Many papers have reported categorisation of both everyday sounds and soundscapes. The most common categories found for individual sounds are natural, human and technological.[3] These vary somewhat in different environments, showing that our category system is adaptive. For example, in a city street the category technological might be replaced by transport, or inside a house a category of household might be added. Researchers have also asked people to categorise sets of complete soundscapes. This seems to produce more varied results across the literature, perhaps because grouping whole soundscapes is not something most of us do very often. Maffiolo et al. found soundscapes grouped into two categories of event sequences (distinguishable sounds present) and amorphous sequences.[10] Raimbault and Dubois reported that people divided soundscapes into two top-level categories: 'transportation and works soundscapes' and people presence soundscapes'.[11] Below these, they suggested a hierarchical structure of sub-categories, down to specific objects in the environment.

It is clear that human categorisation is adaptive and can be applied at least at two scales (individual sounds and whole soundscapes). Music researchers have demonstrated categories at work across a wider range of scales, from the low level (notes, articulations) through mid-level (chords, phrases) to high level (symphonies, genres). Indeed, when considering music, Purwins et al. make the point that categories function as just-noticeable differences for higher-level organisation.[8] That is, they show us what can be usefully distinguished at that level, just as the jnd does for signal-level perception. The evidence on categories suggests that scale is an important underlying feature. People can probably apply their categorisation mechanism at any scale: categorising soundscapes, sounds, or something lower-level produces a different view of the same complex acoustic environment. Scale is thus potentially useful in understanding how our cognition of soundscapes is organised.

Another way to understand the structure of a set of perceptions is with dimensional analysis. This usually results from either a set of paired comparisons of individual elements, or from a large set of semantic differential ratings of each individual element. In both methods, a high-dimensional dataset results, and a multi-dimensional model is then fitted to the results to try to find a low number of dimensions (principal components) that explain most of the variance in the data. Soundscape researchers have produced a few such dimensional results, almost all at the scale of the whole soundscape. The results from different experiments are fairly similar, so that it has been suggested[12] that the first two dimensions of calmness/pleasantness and activity/eventfulness could be regarded as the 'standard model' of soundscape dimensions. There have been surprisingly few dimensional analyses at the scale of individual everyday sounds. An exception is due to Gygi et al. who ran paired comparisons for similarity on fifty everyday sounds and produced a three-dimensional model.[13] Interestingly, they were able to relate the perceptual dimensions of the sounds to objective acoustic features. This has not generally been the case for dimensions of whole soundscapes.

In music, dimensional analysis was successfully applied at the level of individual musical notes. Grey[14] ran a pair comparison experiment on the similarity of notes from different instruments and fitted three significant dimensions to the results. The dimensions were related to acoustic features of the notes: spectral energy distribution, whether the amplitude envelopes of the partials were coincident, and noise at the start of the sound.

It seems clear that scale underlies dimensional models as well. Successful dimensional models can be obtained by focusing on a particular scale. A more complete picture should be obtained if dimensions are identified at all scales involved in everyday perception. Using the idea of scale to organise dimensional models in a hierarchy raises an interesting question. How are the dimensions at each scale related? For example, in the soundscape of a city square, do the positions of the individual sounds on Gygi's dimensions predict the position of the whole soundscape on the calmness/pleasantness and activity/eventfulness graph? The answer to this question requires more research.

3.3 Attention and scale

Attention is the mechanism by which we focus (zoom in) up or down to a particular scale, or move across from one element to another within a scale. When gazing at the park, I can choose to look at the whole scene, a particular tree, a particular leaf, a vein in the leaf, thus traversing scales. Or I can stay at the scale of the whole tree and look from one tree to another. When listening in the city square, I can take in the whole soundscape, the voices, a particular conversation, or the accent of a speaker. Or I can stay at the level of the sound object and try to listen in turn to each of the conversations around me. These all are examples of top-down (TD) attention, where we consciously select one element to attend to. At the same time, there is a competing bottom-up (BU) attention process, whereby a particular element in the scene can capture our attention because it is salient. While I sit in the city square directing my top-down attention from one conversation to another, you can walk up behind me and clap your hands to suddenly capture my attention. Salience can be modelled for individual sounds,[15] while TD auditory attention is harder to model because we can't directly observe it happening. Simple models of TD/BU attention working with two competing sounds produce plausible results, however.[16]

We don't usually have to give much thought to how our attention mechanism works with everyday sound but it does seem possible to observe it happening within ourselves (a kind of meta-attention). One can walk down the street and observe how the attention spotlight selects one element after another. Selection can be quite rapid: changing focus more than once a second in a complex scene does not seem to be fatiguing.

Many writers have pointed out how important attention is in selecting between elements at a particular scale: listening to each sound source in a soundscape, for example. I suggest that attention actually allows us to move our focus in two ways: between elements at one scale, and across different scales. In fact, it is the attention mechanism that allows us to use scale to structure our understanding of complex scenes.

3.4 Attention and listening mode

Several authors have proposed different models to explain how we sometimes attend to one element and sometimes another. Particularly in soundscape research, the concept of the listening mode is popular. Truax[17] proposed that there are three listening modes: listening in search (analytical, focused on sounds related to one's activity), listening in readiness (intermediate, focusing on some sounds, but also alert for information), and background listening (distracted, tuned out, focus on something else, such as reading). Gaver[9] proposed two modes: musical listening (focus on detailed attributes like loudness and timbre) and everyday listening (gathering information about the environment as a whole, more gestalt). Gaver's model is perhaps the one most widely adopted in soundscape research, though the modes are often called analytic listening and holistic listening. Raimbault[18] proposed two modes: holistic hearing (processing the soundscape as a whole) and descriptive listening (focusing on the meaning of the sound). Stockfelt[19] introduced the term 'dishearkening' to refer to the processing of disregarding aspects of the sonic environment. It is described as an active process, as the individual constantly alters which aspects of the soundscape are ignored over others.

I suggest that all these listening modes might be explained as the consequence of attention, both across scales and within one scale. Truax's listening in search is TD in action, sweeping across the scene. Listening in readiness describes the competition between TD and BU attention. Background listening has TD attention directed away from the external scene altogether. Gaver's musical or analytic listening is TD attention at the scale of the sounds, moving from one sound to another, or (at a lower scale) within one sound to examine its attributes. Everyday or holistic listening happens when the attention is zoomed out to the scale of the whole scene. In Raimbault's model, holistic hearing is again attention zoomed out to the whole scene, while descriptive listening presumably involves using TD attention to select each element in turn for semantic processing. (There is conflicting evidence and some debate on whether selection by attention comes before or after semantic processing in the neural sequence of events.[20]) Finally, Stockfelt's dishearkening clearly describes the way TD attention suppresses BU cues and selects just one element of the scene at a time.

The advantages of using attention to explain listening mode are that one mechanism can unify the different proposals for listening modes, and that computational models exist for attention.

3.5 Objects

Auditory scene analysis has uncovered the perceptual rules by which simple auditory scenes are parsed into auditory objects.[21] For example, the rule of common fate says that harmonics which start and stop together will tend to be integrated into one auditory object by the listener. Describing simple scenes as a collection of objects is now widely accepted in many fields.[22, 23] At the same time, spatial audio developers have standardised methods for describing an audio recording as a set of low-level objects.[24] At the moment, object-based audio implementations deal mainly with signal-level concerns such as programming strategies and rendering (see[25] for an example). However, object-based audio offers potentially great flexibility over channel-based systems like stereo or 5.1 for recording, broadcast and storage. Research has therefore started to determine how to make a listener-centred object representation of complex recordings like sports matches and radio dramas.[26]

This should be interesting to soundscape researchers because we might think of objects as sounds, sources or musical instruments. In such a schema, we see that objects could be defined simply as the things that are grouped in categories or located along perceptual dimensions. This in turn suggests that objects can exist at different scales, perhaps in a hierarchy. For example, in an orchestra, each violin might be an object, but some of these objects are members of the larger object called second violins, and the second violins object is itself a member of the larger object called strings. The ontology of objects in the scene will depend on the schema applied. Soundscape researchers might apply a different classification scheme (voices, vehicles, busker, etc.) to recording engineers (foreground dialogue, background, FX, etc). But it is likely that audio researchers will develop tools to support parsing or decomposing a complex scene into an object representation, and these may open up new possibilities for experimentation and understanding in soundscapes.

3.6 Inherent apparatus

Models of listening to complex acoustic scenes applied to outdoor soundscapes, music and spatial audio have many features in common. Dimensions, categories and listening modes are found in all three fields. It is tempting to ask if there are more fundamental cognitive mechanisms that underlie these representations. Dimensions and categories are essentially distance measures. They are ways of representing the size of the difference between two entities, or objects. Similarity and preference are the two most common distance measures in the literature. Distance measures can be usefully represented by categories and/or dimensions. Distance measures generate the attributes that we measure: we define an attribute when it seems to capture some quality of the difference between two objects, such as loudness, calmness or complexity. It should be emphasised that distance measures can be obtained at all scales; for example, we can compares two soundscapes, or two sounds, or two sound features (like pitch). The cognitive mechanism that performs comparison thus seems to be a fundamental one. The second fundamental mechanism would seem to be attention. This is the mechanism that selects objects to be compared or further processed. It too works across scale, allowing us to zoom out to listen to the scene as a whole or to zoom in to an object or sub-object within the scene. I suggest that attention and distance measurement are the two fundamental mechanisms that underlie human cognition of complex acoustic scenes.

4 CONCLUSIONS AND FURTHER WORK

I have attempted to show that the problem studied by researchers in soundscapes, spatial audio quality and music cognition is essentially the same: how humans represent complex acoustic scenes. Each field has produced similar models of categories, dimensions and listening modes. I have proposed that scale is an important feature of how we process complex acoustic scenes. When listening to a complex scene we use our attention to focus in turn on different objects at one scale, but we can also use attention to zoom in and out of different scales. Attention across scale can be used to explain how different listening modes work.

The arguments developed in this paper suggest several possibilities for future work. Explicit evidence of soundscape listening at different scales might be achieved by adapting methods from music cognition. The relationships between soundscape representations at different scales could be sought. For example, whether the dimensions of sounds and the dimensions of soundscapes can be linked might be approached by having listeners manipulate synthesised soundscapes in an interactive experiment. Direct comparisons of broadcast audio scenes and soundscape recordings could be used to determine whether object schemas applied by listeners are consistent between broadcast audio and naturalistic soundscapes. This might lead to an evidence-based taxonomy for the elements of complex scenes. Finally, it might be possible to extend existing models of attention to reproduce the features of different listening modes.

5 REFERENCES

1. G. Kanizsa, *Organization in vision: Essays on Gestalt perception*, Praeger Publishers, (1979).

2. R.M. Schafer, *The tuning of the world*, Knopf, (1977).

3. S.R. Payne, W.J. Davies, and M.D. Adams Research into the Practical and Policy Applications of Soundscape Concepts and Techniques in Urban Areas. Department for Environment, Food and Rural Affairs: London, 2009

4. W.J. Davies, M.D. Adams, N.S. Bruce, M. Marselle, R. Cain, P. Jennings, J. Poxon, A. Carlyle, P. Cusack, D.A. Hall, A. Irwin, K.I. Hume, and C.J. Plack, "Perception of soundscapes: An interdisciplinary approach", *Applied Acoustics* **74**, 224-231 (2013).

5. P. Jackson, M. Dewhirst, and R. Conetta, "QESTRAL (Part 3): System and metrics for spatial quality prediction", *AES 125th Conv.* (San Francisco, 2008).

6. F. Rumsey, "Spatial quality evaluation for reproduced sound: Terminology, meaning , and a scene-based paradigm", *J. Audio Eng. Soc.* **50**, 651–666 (2002).

7. C.L. Krumhansl, "Rhythm and pitch in music cognition", *Psychological Bulletin* **126**, 159-179 (2000).

8. H. Purwins, P. Herrera, M. Grachten, A. Hazan, R. Marxer, and X. Serra, "Computational models of music perception and cognition I: The perceptual and cognitive processing chain", *Physics of Life Reviews* **5**, 151–168 (2008).

9. W.W. Gaver, "What in the world do we hear? An ecological approach to auditory scene perception", *Ecological Psychology* **5**, 1-29 (1993).

10. A.V. Maffiolo, M. Castellengo, and D. Dubois, "Qualitative judgements of urban soundscapes.", *Inter-Noise 99* (Fort Lauderdale, Florida, USA, 6–8 December 1999).

11. M. Raimbault and D. Dubois, "Urban soundscapes: Experiences and knowledge", *Cities* **22**, 339-350 (2005).

12. W.J. Davies and J.E. Murphy, "Reproducibility of soundscape dimensions", *Internoise 2012* (New York, 19-22 August 2012).

13. B. Gygi, G.P.K. Kidd, and C.S. Watson, "Similarity and categorization of environmental sounds", *Perception & Psychophysics* **69**, 839-855 (2007).

14. J.M. Grey, "Multidimensional perceptual scaling of musical timbres", *Journal of the Acoustical Society of America* **61**, 1270-1277 (1977).

15. C. Kayser, C. Petkov, M. Lippert, and N. Logothetis, "Mechanisms for allocating auditory attention: an auditory saliency map", *Curr. Biol.* **15**, 1943-7 (2005).

16. B. De Coensel and D. Botteldooren, "A model of saliency-based auditory attention to environmental sound", *Proc. 20th International Congress on Acoustics (ICA 2010)* (Sydney, Australia, 23–27 August 2010 2010).

17. B. Truax, *Acoustic Communication*, Ablex Publishing, (2001).

18. M. Raimbault, "Qualitative Judgements of Urban Soundscapes: Questionning Questionnaires and Semantic Scales", *Acta Acustica united with Acustica* **92**, 929-937 (2006).

19. O. Stockfelt, *Cars, buildings and soundscapes*, in *Soundscapes: Essays on Vroom and Moo*, H. Jarviluma, Editor. 1994, Tampere University Press: Tampere.

20. C. Spence and V. Santangelo, *Auditory attention*, in *Oxford Handbook of Auditory Science: Hearing*, C.J. Plack, Editor. 2010, OUP: Oxford. p. 249-270.

21. A.S. Bregman, *Auditory scene analysis: the perceptual organization of sound*, MIT Press, (1994).

22. T.D. Griffiths and J.D. Warren, "What is an auditory object?", *Nature Reviews Neuroscience* **5**, 887-892 (2004).

23. J.K. Bizley and Y.E. Cohen, "The what, where and how of auditory-object perception", *Nature Reviews Neuroscience* **14**, 693-707 (2013).

24. J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "Mpeg-h audio: The new standard for universal spatial/3d audio coding", *Journal of the Audio Engineering Society* **62**, 821-830 (2015).

25. M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer: A unified spatial audio reproduction framework for arbitrary rendering methods", *124th Convention of the Audio Engineering Society* May 2008).

26. Engineering and Physical Sciences Research Council, S3A: Future Spatial Audio for an Immersive Listener Experience at Home. URL: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L000539/1