CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The design and formalization of an embodied soundscape sonification framework
Authors	Roddy, Stephen
Publication date	2024-11-22
Original Citation	Roddy, S. (2024) 'The design and formalization of an embodied soundscape sonification framework', International Journal of Human-Computer Interaction, 41(16), pp.9893-9907. https://doi.org/10.1080/10447318.2024.2429757
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1080/10447318.2024.2429757
Rights	© 2024, Informa UK Limited, trading as Taylor & Francis Group. This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Human-Computer Interaction on 22 November 2024, available online: https://doi.org/10.1080/10447318.2024.2429757
Download date	2026-11-03 07:53:52
Item downloaded from	https://hdl.handle.net/10468/16700

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The Design and Formalization of an Embodied Soundscape Sonification Framework

Dr. Stephen Roddy
Radical Humanities Laboratory,
Department of Digital Humanities,
University College Cork, Cork.
sroddy@ucc.ie

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Abstract
This paper explores the design, evaluation, and formalisation of a framework for the creation of soundscape sonifications: data-to-sound mappings which make use of sounds recorded from real-world sources to communicate information to a listener. This approach is informed by design principles from the field of embodied cognitive science. It opens with a consideration of the relationship between soundscapes, sonification, and embodied cognition design principles.
Four approaches to soundscape sonification are presented and evaluated. Evaluation results are presented and analysed before the paper closes with a discussion of the findings and a formalisation of the most effective candidate framework. The article shows how embodied cognition principles can be applied to the design of communicatively effective soundscape sonifications with the takeaway point for researchers being the formalisation of an 'Embodied Soundscape Sonification Framework' as a guide to the design of effective soundscape sonifications.

Keywords: Sonification, Embodied Cognition, Soundscape, Framework.

1. Introduction

Environmental soundscapes, the environmental variant of the soundscapes defined by Schafer (1969), explored in the work of Truax (1996; 2001; 2012) and Westerkamp (1999; 2007) and extended to sonification by Walker and Kramer (2004), provide a communicatively effective domain of sonic complexes for sonification. Communicatively effective in this instance meaning that the intended information the designer wishes to convey about the data is communicated to the listener. This work demonstrates how soundscape sonifications can be organised and presented to a listener in ways that exploit their embodied cognitive meaning-making faculties.

Schafer (1969) originally defined the soundscape as any perceptible ‘sonic environment’, geographical, musical, or otherwise which is distinguished from the wider acoustic environment which contains further non-perceptible acoustic components. The Handbook of Acoustic Ecology (Truax, 1978) further defined the soundscape as follows:

> An environment of sound (sonic environment) with emphasis on the way it is perceived and understood by the individual, or by a society. It thus depends on the relationship between the individual and any such environment. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an artificial environment.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Schafer's definition of the soundscape as the experience of the perceptible signals in an acoustic environment carries the subtle but important distinction that the soundscape results from the filtering and interpretation of the acoustic environment by the perceptual apparatus of the perceiving agent. The second definition presented here carries this further by explicitly highlighting the role of the listeners understanding as crucial to the definition of a soundscape. Truax (1996; 2012) argues that the relationships between human organisms and their environments are primarily mediated by soundscapes through dynamic processes of embodied cognition. A similar argument about the relationship between environment and organism is put forward by Maturana and Varela (1987) and Varela *et al.* (1991). They argue that the organism and environment share a close relationship of mutual specification or structural coupling. This is a closed-loop relationship in which the environment defines and shapes the organism's perceptual and cognitive functions, through the conduit of the physical body, and those functions also define the organism's experience of the environment at the levels of perception, cognition and interaction, through the conduit of the physical body, in equal measure. This differs from Gibson's ecological view of perception in its focus on the physical body as the conduit between mind and world. The environment's ability to shape the cognitive faculties of the organism is constrained by and mediated in terms of the physical body and the organism's ability to shape the environment is constrained by and mediated in terms of the physical body. Sound localisation by interaural time and interaural level difference provides a relevant example of how the physical body's constraint and mediation of the environment defines the organism's cognitive faculties in terms of the listener's ability to determine the location of a sound source.

Soundscapes are embodied to the degree that their experience results from the constraint and mediation of the acoustic environment by the listener's physical body. This suggests that commonly studied embodied cognitive capacities might be at play in the listener's perceptual organisation and conceptual understanding of the soundscape. Kendall (2010; 2014) has argued that the embodied cognitive faculties of embodied schemata and conceptual blending are crucial to the emergence of meaning in soundscape listening in the context of electroacoustic music. Antle *et al.* (2008) and Droumeva *et. al.* (2009) have demonstrated how the experience of a soundscape and its conceptual interpretation can be structured on the basis of conceptual metaphors and embodied schemata in an auditory display context. For a sound to be embodied it need not represent some aspect of the human body and it need not sound as though it originated with the human body either. Auditory cognition itself is embodied and as a result, all sounds are understood and made meaningful in terms of a mind that has been defined and shaped by the physical human body. The embodied approach advocated here involves using and controlling sounds in ways that exploit those aspects of cognition, which have been shaped by the body; it is the nature of cognition, and not the sounds themselves, that is considered embodied. The evaluations presented below examine organisational strategies by which soundscapes can be configured to best exploit the embodied meaning making capacities of the listener in a sonification context.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

1.1 Soundscape and Sonification in HCI Research

Recognising 'soundscapes' as an increasingly active topic in HCI research, Johansen *et al.*, (2022) undertook a meta-analysis of 235 HCI papers on soundscapes, identifying two common conceptualisations of the term across the field. The first is the soundscape as an 'acoustic environment' and the second is the soundscape as a 'composition'. They also acknowledge a hybrid approach accommodating both definitions and rooted in Schafer, Westerkamp's work. The term soundscape sonification is used here to refer to the use of recordings of real-world locations, the use of soundscapes composed from recordings of real-world environmental elements, and the use of soundscapes which have been composed from synthesised elements, in the sonification of data. An early example of the application of frameworks associated with the soundscapes, to auditory displays in a HCI context is Gaver's (1988) Auditory Icons. Building on Bly's (1982) earlier use of sound to convey information in HCI contexts. Gaver's Auditory Icons map computing events to everyday sound-producing events. They exploit the listeners' familiarity with the everyday sounds of the soundscape to communicate information. This idea was further explored in the Sonic Finder system Gaver developed at Apple (Gaver, 1989) and expanded upon again with his synthesized Auditory Icons (Gaver, 1993). Non-speech audio was becoming an important channel for displaying information in this period (Buxton, 1989) culminating in the establishment of the International Community for Auditory Display in 1994. Not long after Macaulay *et al.*, (1998) made the argument that enough knowledge and understanding of sound and soundscapes now existed for the topic to be taken seriously in HCI and interface design. Following directly from Gaver's work, Kobyashi and Schmant (1997) and Fernström and Brazil, (2001) focus on the design of soundscapes to support browsing applications while Brewster (2000) approaches auditory display for wearable computing applications from a soundscape perspective. Blattner (1996) begins to introduce Schafer's (1969) ideas on the soundscape into HCI. The integration of soundscape and sonification techniques for process monitoring tasks is pursued by both Kilander and Lönnqvist (2002), and Mauney and Walker, (2004). Walker and Kramer's (2004) application of Neuhoff's (2004) Ecological Psychoacoustics to Auditory Display, provided an ecologically grounded re-appraisal of Gaver's (1989) original work and a novel ecological framework for sonification design. Vickers (2005, 2011, 2012) and Poli (2005, 2012) draw more explicit links between sonification and the soundscape as it is characterized in the seminal work of Schafer and later Chion (1994), more tightly integrating soundscape studies into sonification research. Interaction with the soundscape (Yücetürk *et al.*, 2017) and interactive soundscapes (Eggen, 2016) became increasingly popular research topics in HCI as Sonic Interaction Design (Franinovic & Serafin, 2013) and Sonic Information Design (Jeon *et al.*, 2018) foregrounded questions about interaction, embodied cognition, and the soundscape in auditory display research (Diniz *et al.*, 2012; Roddy & Bridges, 2018). This period also saw a focus on soundscape sonifications in which different elements are mapped to unique data streams, sometimes

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

called data-driven soundscapes (Wolf, Gliner and Fiebrink, 2015; Hermann, *et al.*, 2015). This has been particularly successful in the context of in-vehicle auditory displays (Nadri et al, 2021). The work presented below shares a similar ethos toward, and interest in, the integration of soundscape and sonification techniques seen in projects by Lenzi *et al.* (2024) and Seiça *et al.* (2023) while also returning to address an earlier line of inquiry introduced by Walker and Kramer (2004) and extend it with reference to key concepts from embodied cognition as they are applied in HCI and sonification research (see Hurtienne, 2017; Antle *et al.*, 2011; Imaz & Benyon, 2007). Walker and Kramer (2004) recognise the importance of environmental sound to sonification. They argue that traditional psychoacoustics does not offer a comprehensive description of how listeners hear and think about real-world sound and suggest that an ecological psychoacoustics that considers the relationship between organism and environment is required for the design of effective sonifications. As such, a deeper understanding of the role of the body as the mediating and constraining conduit between the organism's cognitive faculties and environment is required to better understand meaning making in the context of soundscape sonification. An embodied soundscape sonification is here defined as a sonification in which environmental soundscape elements are organised, mapped to data, and presented in a manner that exploits the embodied nature of auditory cognition as described by the Embodied Sonification Listening Model (ESLM) presented elsewhere (Roddy and Bridges 2020). The work presented here assembles a design framework for creating embodied soundscape sonifications.

1.2 Candidate Frameworks for Embodied Soundscape Sonification

Hurtienne (2017) provides an overview of the application of concepts from the field of embodied cognition in HCI, highlighting the impact of conceptual metaphors conceptual blending and embodied schemata to design practices in the field and briefly demonstrating how they might inform sonification design tasks. The four frameworks presented and evaluated below extend these design practices to soundscape sonification. These frameworks utilise two methodological approaches to relating an ecological sonic source to a data set and two approaches to the formal organisation of those sonic sources to make them meaningful for the listener's embodied cognitive meaning making faculties.

The first way of relating a sound source to a data source is here termed the metaphorical technique. It exploits the listener's faculty for conceptual metaphorical mapping. The second, here termed the blending technique, exploits the listener's faculty for conceptual blending. In the metaphorical approach, illustrated in Figure 1, designers make use of ecological sonic materials that are metaphorically related to the source of a data set. An example of the metaphorical method is the use of a soundscape of stormy weather to represent hard times because bad weather is often used as an embodied conceptual metaphor for hardship (Żołnowska, 2011). This approach differs from standard parameter mapping sonification (PMSon), in which data is directly mapped to sonic parameters. In this approach the data to sound mapping is mediated by an embodied

conceptual metaphor, as introduced by Lakoff and Johnson (1980) Conceptual metaphors have been successfully applied in auditory display contexts in the past (Kantan *et al.*, 2021; Goßman, 2010). A conceptual metaphor is the cognitive process by which embodied schemas and relations derived from sensorimotor activity in familiar domains of thought and experience are leveraged to make sense of novel or abstract domains of thought and experience through a process of reframing. A classic example is the metaphor 'argument is war', here the source domain, 'argument', is reframed in terms of the target domain, 'war'. Relations in the source domain can be framed and reasoned about in terms of analogues in the target domain. For example, one might defend their position or employ a strategy of attack against an opponent. The approach outlined here applies conceptual metaphors theory to guide sonification design. Sonic materials, components of the soundscape, are chosen and designed to frame changes in the dataset in terms of a familiar domain of embodied human experience. Thus, the final sonification operates as a sonic metaphor for the original data. Returning again to the sonification technique, what this means at a design level is that the sound must bear some metaphorical relation to the data being represented so that changes in the data can be framed by, and interpreted through the changes in the sound.

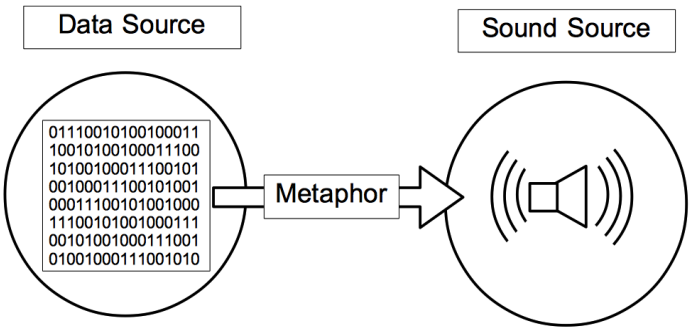


Figure 1. The Metaphorical Technique

In the blended approach, illustrated in Figure 2, designers create a soundscape that represents the imagined shared environment in which the data source is revealed by its causal effect on an independent sound source. An example of the blended method is the use of increasingly agitated birdsong to represent the increasing magnitude of an earthquake. In this blended soundscape the sound source, the birds, and the data source, the earthquake, co-exist in the same space. The magnitude of the earthquake, the data dimension, determines the magnitude of agitation of the birds, the sonic dimension. This differs from the metaphorical approach because it is the imagined causal relationship between the data source and the sonic environment that the designer leverages to link data to sound. Essentially, the data has a causal effect on the sounds within a soundscape environment where the sounds change in reaction to changes in the data. The changes in

the sounds happen in a manner that is logically consistent with the source producing the sounds thus revealing changes in the data. This approach differs from standard PMSon in that the data-to-sound mapping is designed to represent a conceptual blend, which is thought to work by the processes discussed by Fauconnier and Turner (2002). It is an embodied approach because conceptual blends are said to derive from previous embodied experience and are used to lend structure to acts of creative thought. Conceptual blending is a cognitive process involving the integration of two or more familiar concepts to create a novel concept which contains some properties that were not present in either of the two concepts in isolation. For example, the mythological concept of the Pegasus is a blend of the concepts of 'horse' and 'bird' and likewise the Centaur is a blend of ‘man’ and ‘horse’ (Martinez *et al.* 2012). Coming back to consider our sonification technique from the design perspective again, this means that the data will be mapped to some sonic element of a larger soundscape environment. Imaz and Benyon (2007) provide a detailed guide for applying conceptual metaphors to interface design, and for analysing and understanding existing interfaces in terms of conceptual blends.

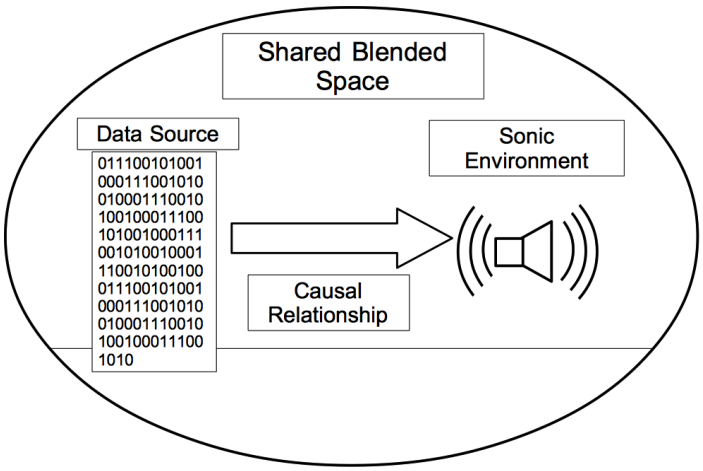


Figure 2. The Blended Technique

The research described previously suggests that the formal embodied structure of a soundscape sonification can be arranged in at least one of two ways to help support the meaningful communication of a data set to a listener. There are also two methods for applying formal embodied structure to the sonification. These are the balance method and the twin-pan balance method. The first method of applying formal embodied structure is based on an embodied schema called the balance schema proposed by Johnson (1987). Embodied schemata (or image schemata) are described as building blocks of thought that are derived from frequently encountered structures within sensorimotor experience. According to this theory, we draw upon embodied schemata to lend structure and organization to acts of reasoning, and imagination and also to structure our perceptual experiences. This is achieved by extending these schemata through conceptual metaphors

and conceptual blending. According to Johnson the balance schema is a formal pattern that humans derive from their first-hand experiences of learning to balance their bodies within their physical environments. A soundscape can be conceptualised as a set of unique individual sonic elements that bear a relationship to one another along common auditory dimensions e.g., amplitude and timbre, spatial dimensions e.g. proximity and orientation, and embodied dimensions, e.g. weight and roughness. The balance method, illustrated in Figure 3, operates by modulating the relationships between the internal sonic elements of a soundscape in accordance with the data. This represents changes in the data source to a listener by altering their overall perception of balance within that soundscape.
In a practical sense this involves designing a basic background soundscape and a set of sound elements, each mapped to a unique data source that are placed into the soundscape. As the data values across these elements converge on 0, they disappear from the soundscape and the background becomes increasingly audible. As the data values increase, they begin to dominate the soundscape. However, given that the data sources are unique and so the amplitudes of the sounds will rise and fall with unique profiles, the internal balance of elements within the soundscape will change and evolve over time with the data. Amplitude presents a critically useful and direct parameter for data mapping in this regard as it allows the overall balance of elements within the soundscape to be shaped by the data.

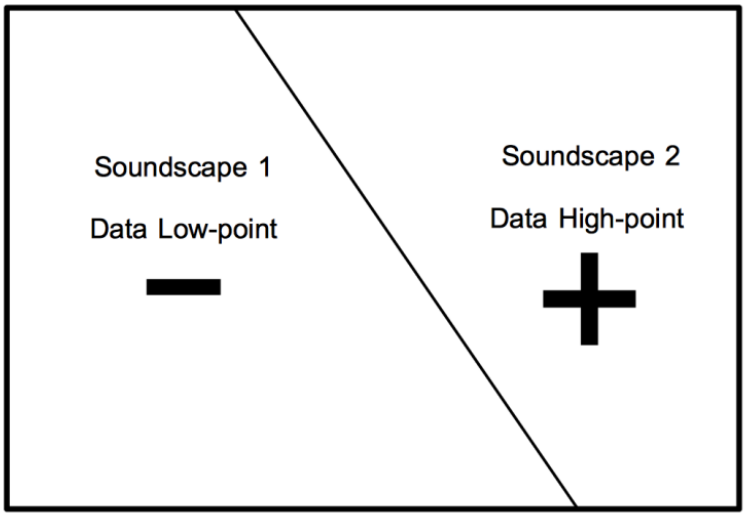


Figure 3. Balance Schema

The second method of applying formal embodied structure is based on an embodied schema called the twin-pan balance schema proposed by Johnson (1987) as a special case of the balance schema previously applied to auditory display in a HCI context by Antle *et al.*,(2011).

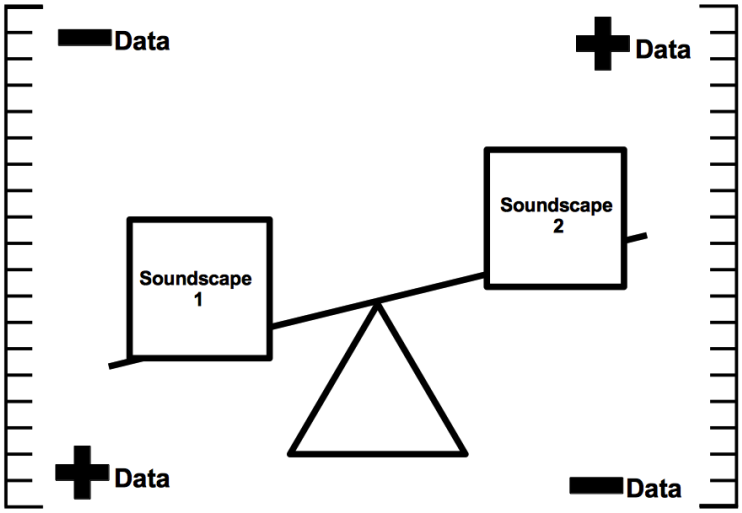


Figure 4. Twin-Pan Balance

According to Johnson, the twin-pan balance schema is a formal pattern that humans derive from their first-hand experiences of multiple counteracting forces finding balance around a fulcrum. The twin-pan balance method, illustrated in Figure 4, operates by selecting two distinct soundscapes, one that represents the high point of the data, and one that represents the low point and mixing between those soundscapes at a rate determined by the data set. This creates two distinct soundscapes for a listener at either extreme of a multivariate data set and a spectrum of transformations for all data points in between. In order to achieve the twin-pan balance effect, the data must be redundantly mapped to both soundscapes with an inverse relationship. As with the previous approach, amplitude provides a useful and direct parameter to which to map the data. As the amplitude of one soundscape increases with changes in the data, the amplitude of the opposing soundscape falls.

Table 1 below presents an overview of each candidate framework. The metaphorical balance framework involves a single soundscape that is metaphorically related to the data source and within which changes in the data values are mapped to changes in the amplitudes of individual sounds. The metaphorical twin-pan framework involves two opposing soundscapes which metaphorically relate to the highest and lowest values of the data. The overall amplitudes of both soundscapes are inversely mapped to changes in the data. The blended balance framework involves one soundscape that represents a blend of the data source and the soundscape. In this framework, the data is directly mapped to control the amplitudes of the sounds that represent reactions to changes in the data. The blended twin-pan framework involves two soundscapes which each represent a blend of the data source and a reactive soundscape impacted by the data source. The overall amplitudes of these soundscapes are inversely mapped to the data.

	Metaphorical	**Blended**

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Balance	• Single soundscape metaphorically related to the data. • Amplitudes of sounds within soundscape directly mapped to data.	• Single soundscape representing a blend of data source and a soundscape for data source to impact. • Amplitudes of sounds within soundscape directly mapped to data.
Twin-pan	• Two opposing soundscapes metaphorically related to the data. • Overall amplitudes of both soundscapes inversely mapped to the data.	• Two opposing soundscapes each representing a blend of data source and a reactive soundscape for data source to impact. • Overall amplitudes of both soundscapes inversely mapped to the data.

Table 1. Overview of the Candidate Frameworks.

2. Materials and Methods

Both approaches to handling sonic materials discussed here rely heavily on the communicative effectiveness of the sonic material used. When applying the framework, it is crucial to use rich sonic material capable of effectively communicating the information intended by the designer(s) when sonifying the data. Experimentation with prototyping platforms suggested that real-world recordings provide better sonic materials for soundscape sonification than synthesised sounds. It is also essential that each individual soundscape represents one single environment in order for the formal organisation methods to work effectively.

There are a number of approaches to applying conceptual metaphors as design tools in HCI contexts. Wilkie *et al.* (2013) and Antle *et al.* (2011) both describe a method for designing auditory displays on the basis of conceptual metaphors, and Roddy & Bridges (2018) lay out a detailed approach to designing sonifications and auditory displays around conceptual metaphors. Whatever approach a designer adopts, sonic materials of this type should always bear a metaphorical relationship to the data source being represented. Blended sonic materials require that the data source has a causal effect on the sound source with the audible result representing the effect of the data source on the sound source within a shared location. The soundscape then represents this environment-specific interaction.

Both approaches to organising the form of the soundscape sonification can be described by simple equations. The balance method operates by modulating the balance

within a single soundscape. It is applied by establishing a base soundscape (Sl) that corresponds to the low point of the data and an extra set of sonic elements (Se) that, when fully integrated with the base soundscape create a transformed soundscape (Sh) that corresponds to the high point of the data. The transformation from the base soundscape (Sl) to the transformed soundscape (Sh) is driven and controlled by the data series (D). Both soundscapes (Sl and Sh) can be related to the data via the conceptual metaphor or conceptual blending techniques. This method is represented in Equation 1 where R1 and R2 represent the low and high points of the data-series' range and F(t) represents the state of the sonification at any given time, t. In the balance method, the base soundscape (Sl) and transformed soundscape (Sh) must represent the same sonic environment, albeit somewhat transformed as the data increases.

$$F(t) = Sl + Sh(D/R2) + R1$$

Equation 1. Data-driven Soundscape Transformation

The twin-pan balance method operates by modulating the balance between two individual soundscapes. It requires the establishment of one soundscape that corresponds to the low point of the data (Sl) and another one that corresponds to the high point of the data (Sh). The high point soundscape (Sh) is modulated by the data (D) and the low soundscape (Sl) is inversely modulated by the data (-D). This method is represented in Equation 2 below where R1 and R2 represent the low and high points of the data-series' range. F(t) represents the state of the sonification at any given time, t.

$$F(t) = Sl(-D/R2) + Sh(D/R2) + R1$$

Equation 2. Twin-Pan Data-Driven Soundscape Transformation

The design for each framework is quite simple and can be expressed in any number of audio processing languages and applications. The examples presented here were realised using Logic Pro X and Csound.

2.1 Experimental Listener Evaluation: Selecting a Potential Candidate Framework

Two methods for selecting or designing sonic materials and two methods of formal organisation of a soundscape sonification are laid out above. Drawing from these there are four possible frameworks for embodied soundscape sonification proposed. A listener evaluation of all four frameworks was undertaken in order to determine which one is the most communicatively effective. The most communicatively effective framework will be considered to be the one that expresses:

(A) The dynamic data curve of an entire sonification.

(B) The value of a static data level presented in a sonification.

For the purposes of this article, the term data curve is defined as a curve that links a series of successive data levels. Data curves change and evolve over time and are analogous to the curves used in visual trend graphs *etc.* The term data level is here used to refer to one of three levels of measurement, high, medium, or low, represented in a visual graph or sonification. A static data level is one that does not change or evolve over the course of a sonification.

A listener evaluation was designed to test the effectiveness of each technique proposed in the soundscape sonification framework. Listeners were recruited from a large international pool of 40 countries via online crowdsourcing platform Crowdflower.com. This was done to ensure that the results were not specific to a particular culture and could be generalised across a large number of people. Experimental materials were hosted on Survey Gizmo. Participants were compensated for their participation. 128 people took part in the evaluations. Of that number 37 were female and 91 were male. 19 participants were between 18 and 24. 69 participants were between the ages of 25 and 34. 32 participants were between the ages of 35 and 54 and 6 participants were over 55. 2 participants chose not to provide an age. Precautionary measures were taken to ensure that participants were using proper equipment. Listeners who could not pass a simple evaluation proving that they were using a 2-channel stereo system were barred from partaking in the evaluation. Listeners completed a short training trial to familiarise themselves with the procedure at the outset of the evaluation. The experiment took approximately 25 minutes to complete and aimed to answer questions A and B above.

2.2 Stimuli and Procedure

Listeners were presented with seven individual evaluations termed A to G. Six of these, A to F, were soundscape evaluations and G was a baseline evaluation intended to generate a reference point against which the results of A to F could be compared. This evaluation made use of pitch-mapped sine tone stimulus to represent static and dynamic data values. The listeners completed two kinds of task during each evaluation, a matching task and an identification task. The matching task is intended to determine which candidate framework can communicate the dynamic data curve of an entire sonification most effectively, as reflected in (A). It asked listeners to match visual representations of data curves with sonic representations of data curves created using the candidate embodied soundscape sonification frameworks. These dynamic stimuli fluctuate across three data levels, small, medium, and large, over the course of a single presentation. They were 30 seconds long and formatted on the basis of each technique described in the embodied soundscape sonification framework. If listeners can correctly match a visual stimulus with the correct sonic stimulus they were said to have correctly identified the data curve from the sonic stimulus. Asking the listeners to draw the data curve may also have produced valid results. However, this solution proved difficult to implement as the evaluations were carried out online.

The identification task is intended to determine which candidate framework is most effective at communicating the individual data levels presented in a sonification as reflected in (B). The identification task asked listeners to identify the values of static data levels presented using the candidate embodied soundscape frameworks. These static stimuli remain at a single data level small, medium, or large, over the course of a stimulus. They were 6 seconds long and formatted on the basis of each candidate technique considered for the embodied soundscape sonification framework. Listeners were told that the stimuli presented in each of the evaluations represent one of the following socioeconomic data sets: unemployment rates, emigration rates, crime rates, GNP rates, suicide rates or poverty rates. The stimuli are not mapped to actual data but are instead designed to represent the data types listed.

The stimuli used in this evaluation were purely auditory. They were not mapped to any actual data. They were designed to undergo specific transformations and occupy specific states in order to determine how listeners relate those transformations and states to an imagined data source. Accordingly, the results obtained from the evaluation are not representative of any structure inherent in, or particular to, a specific data set, but rather describe how listeners map such sounds to imagined data values.

3. Results

Method	**Structure**	**Evaluation**	**Observed number**				**% Correct**
Metaphorical							
			1	**2**	**3**	**4**	
	Twin-pan						
		A	**81**	20	11	16	63.28
		E	52	**33**	15	28	25.78
	Balance						
		B	31	36	**36**	25	28.13
		C	34	18	**63**	13	49.22
		D	42	32	10	**44**	34.38
		F	**57**	30	21	20	44.53
Blended							
	Twin-pan						
		A	39	17	24	**48**	37.50
		D	**59**	29	14	26	46.09
		E	31	34	**31**	32	24.22
		F	40	29	40	**19**	14.84

	Balance						
		B	64	34	28	**13**	10.16
		C	**57**	14	13	44	44.53
Sine Tone							
	Pitch Mapping						
		G	28	23	**62**	15	48.44

Table 2. *Responses for all matching task stimuli. Correct choices in bold.*

The matching task was intended to determine how well listeners could determine the overall shape of the data curve from a soundscape sonification in accordance with (A). The results for the matching task are presented in Table 2. Results can be interpreted against the baseline results generated for the pitch-mapped sine tone. It should be noted that only 48% of listeners identified the pitch-mapped sine tone correctly. This can be explained by the profile of the listeners tested in this evaluation. This work examines embodied cognition as a method for sonification design in the context of an increasingly networked sonification listenership. To reflect this, participants in each evaluation represent a relatively large international pool of networked online listeners.

The strongest individual result of 81 correct identifications was recorded for the metaphorical twin-pan balance approach of evaluation A. This is 15% more than the 62 listeners who identified the pitch-mapped sine tone correctly. The weakest individual result of 13 correct identifications was recorded for the blended balance approach of evaluation B. This is 38% less than the 62 listeners who identified the pitch-mapped sine tone correctly. On average 57 listeners identified the metaphorical twin-pan balance stimuli correctly, 55 listeners identified the metaphorical balance stimuli correctly, 39 listeners identified the blended twin-pan balance stimuli correctly and 35 listeners identified the blended balance stimuli correctly.

Demographic	**Number**	**Mean % of Total Correct Answers**	**SD the Mean %**
Female	37	44.44	15.71
Male	90	43.82	15.52
Other	1	57.69	49.89
18-24	19	40.69	20.37
25-34	69	44.01	15.92
35-54	32	46.94	17.82
55+	6	42.31	19.93
Age Unreported	2	40.38	32.90

South America	21	43.04	16.97
North America	8	44.95	20.90
Europe	77	45.75	15.79
Asia	19	38.97	19.89
Africa	3	39.74	30.99

Table 3. *Percentage of Total Correct Responses by Demographic.*

The effect of demographics on the overall accuracy of participants, as presented in Table 3, was limited. For the most part, results seem to be quite stable across age, gender, and geographical location. Female (N=37) and male (N=90) respondents choose the correct answer across all evaluations roughly 44% of the time. The sample size for other genders was limited to N=1. There is a very slight trend of accuracy increasing with age, but this is weakened by broad standard deviation spreads and the low sample sizes for some age groups (i.e. N=6; SD=19.9 for 55+ age group and N=19; SD=20.37 for the 18-24 group). Similarly, accuracy remains stable across geographic locations for the most part, again accounting for N and SD with a possible dip in Asia.

Chi-square goodness-of-fit tests (see Lancaster, 1969) were performed on the observed results of each matching task as represented in Table 2. These were intended to determine whether the results were statistically or practically significant.

Evaluation G examined the pitch-mapped sine tone stimulus intended as a baseline against which the other stimuli could be compared. As discussed previously only 62 listeners correctly matched this stimulus to its visual counterpart X^2 (3, N = 128)=45.44, p<.001, V=.34. The significant p-value and large effect size suggest reliable result with a large degree of practical significance.

Evaluations A and E examined metaphorical twin-pan balance stimuli. 81 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical twin-pan balance stimulus A X^2 (3, N = 128)=101.313, p<.001, V=.51. The significant p-value and large effect size represented by V suggest that a clear majority of listeners were easily able to identify the data-contour for this sonification by matching it to a visual representation. The results obtained for this evaluation are stronger than those recorded for the pitch-mapped sine tone baseline. This suggests that the soundscape sonification stimulus examined in this evaluation might be better at expressing a data-contour than the standard pitch-mapping approach under the correct conditions. The significant p-value and large effect size also suggest that the results are reliable and of a large degree of practical significance. 33 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical twin-pan balance stimulus E X^2 (3, N = 128)=22.06, p<.001, V=.24. A clear majority of 52 listeners associated the sonification with the wrong visually graphed data-contour. This suggests that listeners were confused by the stimulus. The significant p-value and medium effect size indicate a reliable result with a medium degree of practical significance.

Evaluations B, C, D and F examined metaphorical balance stimuli. The results for the metaphorical balance stimulus of evaluation B clustered around the 25% mark for

each visual data-contour X^2 (3, N = 128)=2.56, p>.05, V=.08. This indicated that listeners were confused by the task. The non-significant p-value and small effect size further support this. 63 listeners, roughly 50%, associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation C X^2 (3, N = 128)=47.56, p<.001, V=.35. Because there were four possible matching options, the chance a listener would randomly select the right one is only 25% This suggests a strong result. The significant p-value and large effect size suggest that the results are reliable and of a large level of practical significance. 44 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation D X^2 (3, N = 128)=22.75, p<.001, V=.24. This was not a significant overall majority. 42 listeners associated the stimulus with a different graph suggesting a measure of confusion for the listeners between these two options. The significant p-value and large effect size suggest that this result is reliable and has a large degree of practical significance. 57 listeners associated the correct visual data-contours with sonified data-contours for the metaphorical balance stimulus of evaluation F X^2 (3, N = 128)=27.94, p<.001, V=.27. This result is slightly lower than the result of 62 recorded for the pitch-mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch-mapped sine tone. The significant p-value and medium effect size suggest that this is a reliable result of medium practical significance.

Evaluations A, D, E and F examined blended twin-pan balance stimuli. 48 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation A X^2 (3, N = 128)=18.56, p<.001, V=.22. The recorded figure is significantly lower than the result for the pitch-mapped sine tone baseline. The significant p-value and medium effect size suggest that this is a robust result of medium practical significance. 59 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation D X^2 (3, N = 128)=34.31, p<.001, V=.3. This result is slightly below the result for the pitch-mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch-mapped sine tone. The significant p-value and large effect size suggest that it is a reliable result of medium practical significance. The results are distributed almost evenly across all of the possible answers for evaluation E's blended twin-pan balance stimulus X^2 (3, N = 128)=.19, p>.05, V=.02. This suggests that they were due to chance and that listeners were confused by the stimulus. This is further supported by the non-significant p-value and small effect size produced by the analysis. Only 19 listeners associated the correct visual data-contours with sonified data-contours for the blended twin-pan balance stimulus of evaluation F X^2 (3, N = 128)=9.56, p<.001, V=.16. The significant p-value but small effect size produced by the analysis suggest that while these results are of practical significance a clear majority of listeners misunderstood the sonification.

Evaluations B and C examined blended balance stimuli. Only 13 listeners matched the soundscape sonification stimulus to the visually graphed data-contour in evaluation B and a clear majority matched the sonic stimulus to the wrong visual graph X^2 (3, N = 128)=25.68, p<.001, V=.26. The results suggest that listeners were confused by the

stimulus and the significant *p*-value and medium effect size indicate that the result is reliable and of a medium level of practical significance. 57 listeners matched the sonic stimulus to the correct visual graph for the blended balance stimulus of evaluation C X^2 (3, N = 128)=45.44, $p<.001$, V=.34. This is only slightly below the result for the pitch-mapped sine tone baseline suggesting that the stimulus is slightly less effective than a pitch-mapped sine tone. The significant *p*-value and large effect size of the analysis suggest that the result is reliable and has a large level of practical significance.

Method	**Evaluation**	**Observation**			**% Correct**
		Small	**Medium**	**Large**	
Metaphorical Twin-pan					
	A 1	**52**	35	41	40.63
	A 2	13	27	**88**	68.75
	A 3	9	**64**	55	50.00
	E 1	**41**	67	20	32.03
	E 2	26	**82**	20	64.06
	E 3	33	53	**42**	32.81
Metaphorical Balance					
	B 1	7	30	**91**	71.09
	B 2	**49**	58	21	38.28
	B 3	17	**71**	40	55.47
	C 1	**55**	53	20	42.97
	C 2	41	**71**	16	55.47
	C 3	12	49	**67**	52.34
	D 1	28	51	**49**	38.28
	D 2	**28**	65	35	21.88
	D 3	24	**61**	43	47.66
	F 1	13	42	**73**	57.03
	F 2	25	**83**	20	64.84
	F 3	**63**	49	16	49.22

Table 4. *Identification Tasks with Metaphorical stimuli. Correct choices in bold*

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Method	Evaluation	Observation			% Correct
		Small	**Medium**	**Large**	
Blended Twin-pan					
	A 1	28	**80**	20	62.50
	A 2	29	55	**44**	34.38
	A 3	**31**	59	38	24.22
	D 1	**40**	67	21	31.25
	D 2	17	65	**46**	35.94
	D 3	18	**90**	20	70.31
	E 1	32	41	**55**	42.97
	E 2	11	**62**	55	48.44
	E 3	**25**	42	61	19.53
	F 1	18	**68**	42	53.13
	F 2	13	56	**59**	46.09
	F 3	**42**	61	25	32.81
Blended Balance					
	B 1	35	61	**32**	25.00
	B 2	28	**72**	28	56.25
	B 3	**59**	39	30	46.09
	C 1	**40**	59	29	31.25
	C 2	13	**76**	29	59.38
	C 3	12	26	**90**	70.31

Table 5. *Identification Tasks with blended stimuli. Correct choices in bold*

Method	Evaluation	Observation			% Correct
		Small	**Medium**	**Large**	
Sine Tone Pitch Map					
	G 1	16	42	**70**	54.6875
	G 2	10	**78**	40	60.9375
	G 3	**44**	57	27	34.375

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Table 6. *Identification Tasks with pitch-mapped sine baseline. Correct choices in bold*

The identification task (B) was intended to determine how well listeners could determine the individual static data levels represented in a soundscape sonification in accordance with (B). The results for the identification task are presented in Tables 4, 5, and 6. 55% of listeners identified the pitch-mapped sine tone correctly. The strongest individual result of 91 correct identifications was recorded for a large stimulus created using the metaphorical balance approach in evaluation B. This is 16% more than the 70 listeners who identified the large stimulus of the pitch-mapped sine tone correctly. The weakest individual result of 25 correct identifications was recorded for the small stimulus of the blended twin-pan balance approach in evaluation E. This is 14% less than the 70 listeners who correctly identified the corresponding small stimulus of the pitch-mapped sine tone. This is 38% less than the 62 listeners who identified the pitch-mapped sine tone correctly. On average 62 listeners identified the metaphorical twin-pan balance stimuli correctly, 63 listeners identified the metaphorical balance stimuli correctly, 53 listeners identified the blended twin-pan balance stimuli correctly and 61 listeners identified blended balance stimuli correctly.

A number of chi-square goodness of fit tests were carried out in order to further analyse the data. Evaluation G examined the pitch-mapped sine tone baseline stimuli. The results for G provide the baseline against which other results can be interpreted. For evaluation G, 70 listeners identified the large value representation correctly X^2 (2, N = 128)=34.18, p<.001, V=.73. The significant p-value and large effect size suggest a reliable result with a large level of practical significance. 78 listeners identified the medium stimulus correctly X^2 (2, N = 128)=54.44, p<.001, V=.92 the significant p-value and large effect size suggest a reliable result with a large level of practical significance. Only 44 listeners identified the small value stimulus correctly X^2 (2, N = 128)=10.6, p<.01, V=.4. A clear majority of listeners incorrectly identified the stimulus suggesting that listeners were confused by it. The significant p-value and large effect size suggest a reliable result with a large degree of practical significance.

Evaluations A and E examined metaphorical twin-pan balance Stimuli. For evaluation A, 52 listeners identified the small value stimulus correctly X^2 (2, N = 128)=3.48, p>.05, V=.12. The non-significant p-value indicated that this result was unreliable. 64 listeners matched the medium X^2 (2, N = 128)=74.55, p<.001, V=.54 the significant p-value and large effect size suggest a reliable result with a large degree of practical significance. 88 listeners identified the large value stimulus correctly X^2 (2, N = 128)=40.8, p<.001, V=.4. This result is larger than the result obtained for the baseline of evaluation G, and the significant p-value and large effect size suggest that this is a reliable result with a large degree of practical significance. For evaluation E only 41 listeners identified the small value stimulus correctly X^2 (2, N = 128)=25.98, p<.001, V=.32. A clear majority identified stimulus incorrectly. This suggests that listeners were confused by the stimulus. The significant p-value and medium effect size suggest a reliable result with a medium degree

of practical significance. 82 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=54.81, p<.001, V=.46. This was a substantially larger result than that obtained for the pitch-mapped baseline and the significant p-value and large effect size suggest a reliable result with a large practical significance. For the large value stimulus, the results cluster around 33%, the number of observations expected from chance. X^2 (2, N = 128)=4.7, p>.05, V=.14. The analysis reveals that the results obtained are not statistically significant. This suggests that listeners were confused by the stimulus.

Evaluations B, C, D, and F examined metaphorical balance stimuli. 91 listeners identified the large value stimulus correctly for evaluation B X^2 (2, N = 128)=88.3, p<.001, V=.59. This is a substantially higher result than that recorded for the sine baseline. The significant p-value and large effect size indicate a reliable result with a large degree of practical significance. Only 49 listeners identified the small value stimulus correctly for B X^2 (2, N = 128)=17.45, p<.001, V=.26. A clear majority identified it incorrectly. The significant p-value and medium effect size suggest that this is a reliable result with a medium degree of practical significance. This suggests that listeners were confused by this stimulus. 71 listeners identified the medium level stimulus correctly for evaluation B X^2 (2, N = 128)=33.42, p<.001, V=.37. This is close to but below the level established by the baseline suggesting that the stimulus is slightly less effective than a pitch-mapped sine tone. The significant p-value and medium effect size suggest a reliable result with a medium degree of practical significance. Only 55 listeners identified the small value stimulus correctly for evaluation C X^2 (2, N = 128)=18.11, p<.001, V=.27. 53 listeners incorrectly categorised the stimulus as medium, suggesting that listeners were confused by the stimulus. This was supported by the significant p-value and medium effect size, which indicates a reliable result with a medium level of practical significance. 71 listeners identified the medium value stimulus correctly for C X^2 (2, N = 128)=35.55, p<.001, V=.37. This is close to but below the level established by the baseline suggesting that the stimulus is slightly less effective than a pitch-mapped sine tone. The significant p-value and medium effect size suggest a reliable result with a medium level of practical significance. 67 listeners identified the large level stimulus correctly for evaluation C X^2 (2, N = 128)=36.86, p<.001, V=.38. This is only slightly below the level established by the baseline suggesting that the stimulus is only slightly less effective than a pitch-mapped sine tone. The significant p-value and large effect size suggest a reliable result with a large level of practical significance.

Only 49 listeners identified the large value stimulus correctly for evaluation D X^2 (2, N = 128)=7.61, p<.05, V=.17. 51 listeners incorrectly categorised the stimulus as medium, suggesting that listeners were confused by the stimulus. The significant p-value and small effect size indicate a reliable result with a small degree of practical significance. Only 28 listeners identified the small value stimulus correctly for D X^2 (2, N = 128)=18.11, p<.001, V=.27. A clear majority identified the stimulus incorrectly. The significant p-value and medium effect size suggest that this is a reliable result with a medium level of practical significance. 61 listeners identified the medium level stimulus correctly for evaluation D X^2 (2, N = 128)=36.86, p<.001, V=.38. This is below the level established by the baseline

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

suggesting that the stimulus is less effective than a pitch-mapped sine tone. The significant *p*-value and large effect size suggest a reliable result with a large level of practical significance.

73 listeners identified the large value stimulus correctly for evaluation F X^2 (2, N = 128)=42.20, $p<.001$, V=.41. This is slightly above the baseline result and the significant *p*-value and large effect size indicate a reliable result with a large level of practical significance. 83 listeners identified the medium value stimulus correctly for evaluation F X^2 (2, N = 128)=57.48, $p<.001$, V=.47. This is above the baseline result and the significant *p*-value and large effect size suggest that it is a reliable result with a large level of practical significance. 63 listeners identified the medium value stimulus correctly for evaluation F X^2 (2, N = 128)=27.3, $p<.001$, V=.33. This is above the baseline result and the significant *p*-value and medium effect size suggest that it is a reliable result with a medium level of practical significance.

Evaluations A, D, E and F examined blended twin-pan balance stimuli. For evaluation A, 80 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=49.75, $p<.001$, V=.44. This result is substantially higher than the corresponding baseline result. The significant *p*-value and large effect size indicate a reliable result of large practical significance. Only 44 listeners identified the large stimulus for evaluation A correctly X^2 (2, 128)=7.98, $p<.05$, V=.18. A clear majority incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p*-value and small effect size indicate that the result is reliable but has a small level of practical significance. Only 31 listeners identified the small stimulus correctly for evaluation A X^2 (2, 128)=9.95, $p<.001$, V=.2 and a clear majority of listeners incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p*-value and roughly medium effect size indicate a reliable result of medium practical significance.

For evaluation D, 40 listeners identified the small value stimulus correctly X^2 (2, N = 128)=25.05, $p<.001$, V=.31. This result is slightly lower than the corresponding baseline result. The significant *p*-value and medium effect size indicate a reliable result of medium practical significance. Only 46 listeners identified the large stimulus for evaluation D correctly X^2 (2, 128)=27.39, $p<.001$, V=.33. A clear majority incorrectly identified the stimulus. This indicates that listeners were confused by the stimulus. The significant *p*-value and medium effect size indicate that the result is reliable and of medium practical significance. 90 listeners identified the medium stimulus correctly for evaluation D X^2 (2, 128)=78.81, $p<.001$, V=.55 The significant *p*-value and large effect size indicate a reliable result with a high level of practical significance.

For evaluation E, 55 listeners identified the large value stimulus correctly X^2 (2, N = 128)=6.30, $p<.05$, V=.16. This result is lower than the corresponding baseline result. The significant *p*-value and small effect size indicate a reliable result of limited practical significance. 62 listeners identified the medium stimulus for evaluation E correctly X^2 (2, 128)=35.83, $p<.001$, V=.37. This result is lower than the corresponding baseline result. The significant *p*-value and large effect size indicate that the result is reliable and of large practical significance. Only 25 listeners identified the small stimulus correctly for

evaluation E X^2 (2, 128)=15.20, p<.001, V=.24. A clear majority of listeners misidentified the stimulus, suggesting that they were confused by the stimulus. The significant p-value and medium effect size indicate a reliable result with a medium level of practical significance.

For evaluation F, 68 listeners identified the medium value stimulus correctly X^2 (2, N = 128)=29.31, p<.001, V=.34. This result is lower than the corresponding baseline result. The significant p-value and roughly large effect size indicate a reliable result of medium to high practical significance. 59 listeners identified the large stimulus for evaluation E correctly X^2 (2, 128)=31.05, p<.001, V=.35. 56 listeners categorised the stimulus as medium in size. This suggests that the stimulus was confusing. The significant p-value and large effect size indicate that the result is reliable and of large practical significance. Only 42 listeners identified the small stimulus correctly for evaluation E X^2 (2, 128)=15.20, p<.001, V=.24. A clear majority of listeners misidentified the stimulus, suggesting that they were confused by the stimulus. The significant p-value and medium effect size indicate a reliable result with a medium level of practical significance.

Evaluations B and C examined blended balance stimuli. For evaluation B, only 32 listeners identified the large value stimulus correctly X^2 (2, N = 128)=11.92, p<.01, V=.22. A clear majority of listeners identified the stimulus incorrectly. This suggests that the listeners were confused about the stimulus. The significant p-value and medium effect size indicate a reliable result of medium practical significance. 72 listeners identified the medium stimulus for evaluation B correctly X^2 (2, 128)=30.25, p<.001, V=.34. This was lower than the corresponding baseline result. The significant p-value and roughly large effect size indicate that the result is reliable and of medium to large practical significance. 59 listeners identified the small stimulus correctly for evaluation B X^2 (2, 128)=10.33, p<.01, V=.20. The significant p-value and roughly medium effect size indicate a reliable result with a small to medium level of practical significance.

For evaluation C, only 40 listeners identified the small value stimulus correctly X^2 (2, N = 128)=10.80, p<.01, V=.21. A clear majority of listeners identified the stimulus incorrectly. This suggests that the listeners were confused about the stimulus. The significant p-value and medium effect size indicate a reliable result of medium practical significance. 76 listeners identified the medium stimulus for evaluation C correctly X^2 (2, 128)=51.05, p<.001, V=.45. This was slightly lower than the corresponding baseline result. The significant p-value and large effect size indicate that the result is reliable and of large practical significance. 90 listeners identified the large stimulus correctly for evaluation C X^2 (2, 128)=81.06, p<.001, V=.56. This is substantially larger than the figure recorded for the corresponding baseline. The significant p-value and large effect size indicate a reliable result with a large level of practical significance.

4. Discussion of Results

Overall, the results of these evaluations suggest that on average twin-pan balance soundscapes are more effective than balance soundscapes for communicating the shape of a data curve and are of roughly equal effectiveness for communicating individual discrete

data levels to a listener. The results also suggest that the metaphorical soundscape stimuli were more communicatively effective than the blended soundscape stimuli for representing both a dynamically changing data curve and a static data level. This seems to be consistent with earlier findings on the applicability of conceptual metaphors and embodied schemata to auditory display and HCI design tasks (Antle *et al.*, 2008; Droumeva *et al.*, 2009), and with the application of the twin-pan balance schema to auditory display design (Antle *et al.*, 2011). More generally this corresponds with Hurtienne's (2017) analysis of the kinds of HCI design tasks for which conceptual metaphors and embodied schemata prove useful, namely intuitive interface design.

On average, none of the soundscape sonification techniques proved more effective than the pitch-mapped sine tone baseline for communicating a data curve. However, the individual metaphorical twin-pan balance stimulus of evaluation A did prove more effective. 81 listeners identified this stimulus correctly and this result shows a high level of practical significance for sonification design as evidenced in the large effect size generated. This suggests that in general, the metaphorical twin-pan balance approach holds the capacity to be more effective than the pitch-mapped sine tone approach. However, further research is needed to determine why this stimulus produced such a strong result in comparison to the other metaphorical twin-pan balance stimulus that was correctly identified by only 33 listeners.

The results obtained in this evaluation are varied. Certain results were weak where a stronger result would be expected and other results were strong where a weak result would be expected. For example, all of the four blended twin-pan balance stimuli with the exception of E produced statistically and practically significant results. Similarly, the metaphorical twin-pan balance stimulus of evaluation A produced a much stronger result than that of evaluation E. Both stimuli use similar sonic content and embodied schematic organisational patterns as their counterparts. This suggests that while the content of the sounds used is obviously important, it is how those sounds are mapped to data and arranged in a soundscape at a formal level that determines the overall efficacy of the sonification. In this particular case, the sounds of evaluation A that were metaphorically related to the data sources and mapped to data values in accordance with the twin-pan balance schema were found to be most effective.

The results obtained for the metaphorical balance stimulus of matching task B and the blended twin-pan balance stimulus of evaluation E suggested a particular lack of communicative effectiveness. This may in part be due to the design of the stimuli. Amplitude modelling was used to control the balance between soundscapes for all the stimuli in this evaluation. However, each of the soundscapes has its own internal ebb and flow in terms of the amplitudes of individual sonic elements. As such the internal amplitude fluctuations of the stimuli may have interfered with the amplitude modelling applied to the soundscapes in the stimulus. This in turn would make it hard for listeners to match the stimulus to a corresponding visually graphed data curve. This highlights the difficulty of working with the complexities of real-world soundscape recordings. This

further underlines the importance of developing good design practices for working with soundscape materials in a sonification context.

In the identification task, listeners incorrectly identified the small stimulus of the pitch-mapped sine baseline. They showed difficulties identifying all of the small, blended twin-pan balance stimuli and one of two small blended balance stimuli. Listeners also showed difficulty identifying one of the two small metaphorical twin-pan balance stimuli and three of the four small metaphorical balance stimuli. These results suggest that listeners find it difficult to determine low data levels for both pitch-mapped sine tone stimuli and embodied soundscape sonification stimuli. On average soundscape sonification methods were less effective than the pitch-mapping method. However, as discussed in the previous section a few stimuli proved to be more effective than the small pitch-mapped sine tone baseline. This suggests that in certain circumstances embodied soundscape sonifications can be more effective than pitch-mapped sine methods. It may be possible that listeners consider the presence of any sound whatsoever to represent the presence of a positive data value. It may be better to use a mapping strategy where the lowest point in the data is represented by silence and lower data values are mapped to quieter soundscape amplitudes. Listeners were more successful at identifying medium and high value soundscape sonifications then low value soundscape stimuli. On average listeners were more successful at identifying data levels from the pitch-mapped sine tone baseline than the soundscape sonification stimuli. However, some of the individual stimuli, discussed in the previous section, presented notable exceptions to this trend. This once again suggests that embodied soundscape sonification approaches have the capacity to be more communicatively effective than pitch-mapped sonifications for expressing medium and high-level data values. Further research is required to better understand listeners' interpretation of value levels from soundscape sonifications and to determine exactly when and why they are effective and how this effectiveness can be controlled.

While the results generated by this empirical listener evaluation are interesting, they are preliminary in nature. Many of the results generated require further research that might more precisely pinpoint how embodied soundscape sonification can be used to maximise communicative effectiveness. Nevertheless, these results are of practical significance in the context of sonification design as evidenced in the strong effect sizes discussed in the previous section.

The results suggest that the metaphorical twin-pan balance method for creating embodied soundscape sonifications is more effective at communicating data curves and discrete data points to a listener. While the metaphorical balance approach provided some higher individual results for the identification task, the metaphorical twin-pan balance approach proved more robust overall. As noted at the outset of this section, this is consistent with previous research in the field (Antle *et al.*, 2008; Droumeva *et al.*, 2009; Antle *et al.*, 2011; Hurtienne, 2017). The results also suggest that this approach has the capacity to be more effective than pitch-mapping sonifications under certain circumstances. While the exact nature of these circumstances requires further research, the limitations of mapping data to pitch have long been understood (Flowers, 2005). Several

researchers have discussed a distinction between primary metaphors and complex metaphors (Grady, 1997; Lakoff and Johnson, 1999), with Hurtienne (2017) underlining their importance in HCI contexts,. Primary metaphors directly connect subjective or abstract domains to embodied experiences, for example, the association people draw between the experience of warmth and the subjective impressions of affection. Complex metaphors are constructed from these primary metaphors through the process of conceptual blending, in similar fashion to how molecules are constructed from atoms through a process of chemical bonding. Primary metaphors are more immediately meaningful to a person because they are directly connected to embodied experience, and complex metaphors are less immediately meaningful because their connection to embodied experience is mediated by multiple primary metaphors the relationships between which must be grasped in order to understand these metaphors. The metaphorical soundscapes explored here are similar to primary metaphors because they relate directly to concrete embodied experiences. The blended soundscapes are analogous to complex metaphors because they reference multiple domains of embodied experience the relationships between which need to be unpacked and understood through a reversal of the conceptual blending process in order for the soundscape to become meaningful. This would add additional cognitive load to the listener and might explain why the metaphorical soundscapes performed better than the blended soundscapes. It should also be noted that the findings with regards to blended soundscape sonifications are consistent with Hurtienne's (2014) argument that conceptual blending, while a useful tool for analysing a design, is not a useful tool for planning and configuring a design.

It is likely that the twin-pan balance method creates soundscapes that can be more easily interpreted by the listener. In the balance soundscapes, listeners need to pay close attention to the individual elements in the soundscape, as well as the overall state of the soundscape in order to get a sense of changes in the data. In the twin-pan balance soundscapes, listeners only need to pay attention to which soundscape is more prominent in the mix and the degree to which they judge it to be prominent. This might explain why the twin-pan balance soundscapes performed better than the balance soundscapes.

The metaphorical method for relating data to sonic materials and the twin-pan balance method for organising and mapping data to those sonic materials represent an effective technique for the design of embodied soundscape sonifications. This indicates that these methods represent a viable framework for embodied soundscape sonification that might prove useful for sonification design and provide a focal point for further research. This in turn represents an iteration over previous approaches to soundscape research in HCI which, as demonstrated by Johansen *et al.*, (2022), clusters around one of two core themes: the soundscape as acoustic environment, and/or the soundscape as composition. The work described here asks HCI researchers engaged with auditory display to address the embodied dimensions of the soundscape in their sonification designs, in much the same way that Walker and Kramer (2004) did in their application of ecological psychoacoustics to auditory display. The key contribution of this research is that it links two important themes in HCI research of sonification, the soundscape, and the application of embodied cognition principles to the design process. The key takeaway point here for HCI researchers is that an approach to soundscape sonification that

integrates conceptual metaphor and the twin-pan balance schema is more effective at communicating data curves and discrete data points to a listener than the alternatives explored here. More specifically, this process involves the design of two distinct soundscapes, one corresponding to the maximum value of the data set, and another corresponding to the minimum. The sounds chosen for these soundscapes should bear a metaphorical relationship to the data being sonified. The data should be mapped directly to the amplitude of the soundscape representing the minimum value. As data values increase the soundscape's amplitude should also increase and vice versa. Data should also be mapped with an inverse polarity to the soundscape representing the maximum value. As the data increases in value, the amplitude of this soundscape should decrease and vice-versa. This approach is formalized as the Embodied Soundscape Sonification Framework in the following section.

4.1 A General Formalisation for an Embodied Soundscape Sonification Framework

The embodied soundscape framework is a framework for creating soundscape sonifications that exploits the embodied nature of human cognition to effectively communicate data to a listener.

In this framework, two soundscapes are used to represent a single data set. The first bears a metaphorical relation to the high point of the data and the second bears a metaphorical relation to the low point of the data. The original data set is mapped to modulate the perceptual salience of each soundscape in an inversely proportional ratio.

The entire framework can be described by the following transfer functions where msLo is the metaphorically related soundscape that represents the low point of the data msHi is the metaphorically related soundscape that represents the high point of the data, D represents the Data, R1 and R2 represent the low and high points of the data series' range respectively and F(t) is the state of sonification at a given time, t.

$$F(t) = msL(-D/R2) + msH(D/R2) + R1$$

Equation 3. The Embodied soundscape Mapping Transform

4.2 Limitations of the Framework & Future Work

One limitation of the embodied soundscape framework is that it has only been empirically verified to work for dynamic data curves and static sonifications across three levels of data: low, medium and high. As such, the framework in its current state is best suited to conveying an overall topological gist of a data set. Further research is required to determine how effectively the framework performs for high-resolution data sets with multitudes of data levels. The framework's limited capacity for representing higher-resolution data must be offset against the results obtained when the framework was compared to the pitch-mapping of a sine tone. These results suggest that in the correct circumstances the framework can potentially produce more communicative data sonifications than the pitch-mapping approach across three data levels.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Another limitation of this approach is that the four candidate frameworks utilised only two of the many possible embodied schematic approaches to structuring the form of a sonification: the balance and twin-pan balance schemata. An exploration of different kinds of embodied schematic structuring strategies may make for even more effective sonifications.

Alongside the limitations of the framework there are were also some shortcomings of the evaluation that must be addressed when considering the results. One such shortcoming was the decision to have participants choose the correct data curve from four alternatives provided to them during the matching tasks. This approach could not capture data about listeners who were able to correctly identify partial sections of the curve, nor could it record more specific details, like whether listeners were more likely to correctly identify rising, falling, or static sections of the overall curve across the different stimuli. Similarly, some participants may have chosen the correct curve while only recognising a portion of the curve from a given stimulus. A more robust evaluation method would be to allow the listeners to draw the curve that they heard and perform a match between their drawings and the actual data curve. This would help to overcome the problems described previously while also producing a richer representation of the listeners' subjective experience of the stimuli. While, technical limitations prevented the exploration of this approach for this study, future work will explore similarly robust methods for capturing user data.

Finally, there are subjective dimensions to sonification in that each user has their own unique and personal experience of a given soundscape and thus a given soundscape sonification. As such, understanding the experience of the individual user is important for the design of effective sonifications. A study by Nadri *et al.*, (2021) focusing the subjective experience of auditory displays in highly automated vehicles used. Subjective evaluation criteria to find that sonification of the surrounding environment was preferred by users over more abstract sounds. A similar evaluation of the subjective dimensions of the user experience of soundscape sonification may help to further refine this approach.

5. Conclusions

This article opened with a consideration of embodied soundscape sonification before discussing four possible frameworks for embodied soundscape sonification were then introduced and an example of one of the frameworks was presented. These frameworks were submitted to user testing the results of which determined the most effective framework for embodied soundscape sonification to be that in which sonic materials are metaphorically related to the data source and the soundscape is organised around the structure of Johnson's (1987) twin-pan balance schema. The key contribution of this research is that it shows how embodied cognition principles can be applied to the design of communicatively effective soundscape sonifications. Following from this, the main takeaway point for HCI researchers is the Embodied Soundscape Sonification Framework, formalized here in section 4.1, as a. guide for soundscape sonification design.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Declaration of Interest Statement

I hereby declare that the disclosed information is correct and that no other situation of real, potential or apparent conflict of interest is known to me. I undertake to inform you of any change in these circumstances, including if an issue arises during the course of the meeting or work itself.

References

Antle, A. N., Corness, G., & Bevans, A. (2011). Springboard: Designing image schema based embodied interaction for an abstract domain. *Whole body interaction*, 7-18.

Antle, A. N., Droumeva, M., & Corness, G. (2008, June). Playing with the sound maker: do embodied metaphors help children learn?. In *Proceedings of the 7th international conference on Interaction design and children* (pp. 178-185).

Blattner, M. (1996). Multimedia Interfaces: Designing for Diversity. In *Multimedia Systems and Techniques* (pp. 87-122). Boston, MA: Springer US.

Bly, S. (1982) Presenting information in sound. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems (CHI '82)* (pp. 371–375). ACM: New York, NY, USA.

Brewster, S., & Walker, V. A. (2000, January). Non-visual interfaces for wearable computers. In *Colloquium Digest-IEE* (pp. 6-6). IEE; 1999.

Buxton, W. (1989). Introduction to this special issue on nonspeech audio. *Human Computer Interaction*, *4*, 1-9.

Chion, M. (2019). *Audio-vision: sound on screen*. Columbia University Press.

Diniz, N., Coussement, P., Deweppe, A., Demey, M., & Leman, M. (2012). An embodied music cognition approach to multilevel interactive sonification. *Journal on Multimodal User Interfaces*, *5*, 211-219.

Droumeva, M., Antle, A.N., Corness, G. & Bevans, A. (2009). Springboard: Exploring embodied metaphor in the design of sound feedback for physical responsive environments. In *Proceedings of the 15th International Conference on Auditory Display, Cophenhagen, Denmark, May 18-22, 2009.*

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Eggen, B. (2016). Interactive soundscapes of the future everyday life. *Peripheral interaction: challenges and opportunities for HCI in the periphery of attention*, 239-251.

Fauconnier, G. & Turner, M. (2002). *The way we think: conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Fernström, M., & Brazil, E. (2001). Sonic Browsing: an auditory tool for multimedia asset management. Hiipakka, J, Zacharov, N., & Takala, T., (Eds). In *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, *Espoo, Finland, July 29-August 1, 2001.*

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. *In Proceedings of the First International Symposium on Auditory Graphs (AGS2005).* (pp. 1–5).

Franinovic, K., & Serafin, S. (Eds.). (2013). *Sonic interaction design*. Mit Press.

Gaver, W. W. (1988). *Everyday listening and auditory icons* (Doctoral dissertation, University of California, San Diego, Department of Cognitive Science and Psychology).

Gaver, W. W. (1989). The SonicFinder: An interface that uses auditory icons. *Human–Computer Interaction*, *4*(1), 67-94.

Gaver, W. W. (1993). Synthesizing auditory icons. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems* (pp. 228-235). ACM.

Grady, J. (1997). Theories are buildings revisited. *Cognitive Linguistics*, 8 (4), 267–290.

Grond, F. & Berger, J. (2011). Parameter mapping sonification. In Hermann, T., Hunt, A., Neuhoff, J. G., (Eds) *The Sonification Handbook,* (pp. 363–397). Berlin, Germany: Logos Publishing House.

Goßman, J. (2010). From metaphor to medium: Sonification as extension of our body. *International Community for Auditory Display. Washington, DC, USA: International Community for Auditory Display*.

Hermann, T., Hildebrandt, T., Langeslag, P. & Rinderle-Ma, S. (2015). Optimizing aesthetics and precision in sonification for peripheral process-monitoring. In *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015). Graz, Austria, July 8-10.*

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Hurtienne, J. (2014). Non-Linguistic Applications of Cognitive Linguistics on the Usefulness of Image-Schematic Metaphor in User Interface Design. *The Bloomsbury companion to cognitive linguistics*, 301-24.

Imaz, M., & Benyon, D. (2007). *Designing with blends: Conceptual foundations of human-computer interaction and software engineering methods*. MIT Press, Cambridge, MA.

Jeon, M., Walker, B. N., & Barrass, S. (2018). Introduction to the special issue on sonic information design: theory, methods, and practice, part 1. *Ergonomics in Design*, *26*(4), 3-3.

Johnson, M. (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Kantan, P. R., Spaich, E. G., & Dahl, S. (2021, June). A metaphor-based technical framework for musical sonification in movement rehabilitation. In *The 26th International Conference on Auditory Display (ICAD 2021)*.

Kendall, G. S. (2010). Meaning in electroacoustic music and the everyday mind. *Organised Sound, 15*(01), 63-74.

Kendall, G. S. (2014). The Feeling Blend: Feeling and emotion in electroacoustic art. *Organised Sound, 19*(02), 192-202.

Kilander, F., & Lonnqvist, P. (2002). A whisper in the woods—an ambient soundscape for peripheral awareness of remote processes. In *Proceedings of the 8th International Conference on Auditory Display (ICAD 2002), Kyoto, Japan, July 2-5, 2002*.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic books.

Lancaster, H. O. (1969). *Chi-square distribution.* John Wiley & Sons, Inc..

Lenzi, S., Terenghi, G., Meacci, D., Moreno Fernandez-de-Leceta, A., & Ciuccarelli, P. (2024). The design of Datascapes: toward a design framework for sonification for

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

anomaly detection in AI-supported networked environments. *Frontiers in Computer Science*, *5*, 1254678.

Macaulay, C., Benyon, D., & Crerar, A. (1998). Voices in the forest: Sounds, soundscapes and interface design. *Exploring Navigation: Towards a Framework for Design and Evaluation in Electronic Spaces, SICS Technical Report*, *98*.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding.* New Science Library/Shambhala Publications.

Mauney, B. S., & Walker, B. N. (2004). Creating functional and livable soundscapes for peripheral monitoring of dynamic data. In *Proceedings of the 10th International Conference on Auditory Display (ICAD-04),* Sydney, Australia, 2004.

Nadri, C., Ko, S., Diggs, C., Winters, M., Sreehari, V. K., & Jeon, M. (2021, September). Novel auditory displays in highly automated vehicles: sonification improves driver situation awareness, perceived workload, and overall experience. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 586-590). Sage CA: Los Angeles, CA: SAGE Publications.

Neuhoff, J. (2004). *Ecological psychoacoustics*. Brill.

Polli, A. (2005). Atmospherics/weather works: A spatialized meteorological data sonification project. *Leonardo*, *38*(1), 31-36.

Polli, A. (2012). Soundscape, sonification, and sound activism. *AI & Society, 27*(2), 257-268.

Roddy, S., & Bridges, B. (2018a). Sound, ecological affordances and embodied mappings in auditory display. *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*, 231-258.

Roddy, S., & Bridges, B. (2018b). Addressing the mapping problem in sonic information design through embodied image schemata, conceptual metaphors, and conceptual blending. *Journal of Sonic Studies*, (17).

Roddy, S., & Bridges, B. (2020). Mapping for meaning: the embodied sonification listening model and its implications for the mapping problem in sonic information design. *Journal on Multimodal User Interfaces*, *14*(2), 143-151.

Schafer, R. M. (1969). *The new soundscape: a handbook for the modern music teacher.* Canada: BMI.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Seiça, M., Roque, L., Martins, P., & Cardoso, F. A. (2023). An interdisciplinary journey towards an aesthetics of sonification experience. *Journal on Multimodal User Interfaces*, *17*(4), 263-284.

Truax, B. (1978). *Handbook for Acoustic Ecology,* Originally published by the World Soundscape Project, Simon Fraser University.

Truax, B. (1984). *Acoustic communication.* Norwood, NJ: Ablex Publishing Corporation.

Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review, 15*(1-2), 49-65.

Truax, B. (2012). Sound, listening and place: The aesthetic dilemma. *Organised Sound, 17*(03), 193-201.

Varela, F.J., Thompson, E.T. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: The MIT Press.

Vickers, P. (2005). Ars Informatica -- Ars Electronica: Improving Sonification Aesthetics. In: *Understanding and Designing for Aesthetic Experience* (workshop at HCI 2005: The 19th British HCI Group Annual Conference), 5-9 September 2005, Edinburgh.

Vickers, P. (2011). Sonification for process monitoring. In Hermann, T., Hunt, A., & Neuhoff, J. G., (Eds). *The Sonification Handbook,* (pp. 455–491). Berlin, Germany: Logos Publishing House.

Vickers, P. (2012). Ways of listening and modes of being: electroacoustic auditory display. *Journal of Sonic Studies, 2*(1), 1-23

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. *Ecological Psychoacoustics,* 150-175.

Westerkamp, H. (1999). Soundscape composition: Linking inner and outer worlds. *Soundscape Newsletter,* Amsterdam, Holland.

Westerkamp, H. (2007). The world soundscape project. *World,* 12, 41.

Wilkie, K., Holland, S., & Mulholland, P. (2013). Towards a participatory approach for interaction design based on conceptual metaphor theory: a case study from music interaction. *Music and Human-Computer Interaction*, 259-270.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757

Wolf, K. E., Gliner, G., & Fiebrink, R. (2015, March). A model for data-driven sonification using soundscapes. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion* (pp. 97-100). ACM.

Yücetürk, S., Obaid, M., & Yantaç, A. E. (2016, October). Probing human-soundscape interaction using observational user experience methods. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (pp. 1-4).

Żołnowska, I. (2011). Weather as the source domain for metaphorical expressions. *AVANT.*Pismo Awangardy Filozoficzno-Naukowej, (1), 165-179.

Biography

Stephen Roddy is a Musician, Researcher, and Lecturer working at the intersection of Engineering and Creative Arts at the Radical Humanities Lab and the Department of Digital Humanities at University College Cork. He has a background in Sound and Music Computing and holds a Ph.D. in Sonification.

Roddy, S. (2024). The Design and Formalization of an Embodied Soundscape Sonification Framework. International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2024.2429757