

## SONIFICATION LISTENING: AN EMPIRICAL EMBODIED APPROACH

*Stephen Roddy*

Dept. Of Electronics and Engineering,  
Printing House,  
Trinity College, Dublin 2, Ireland  
roddyst@tcd.ie

*Dermot Furlong*

Music and Media Technologies,  
Dept. Of Electronics and Engineering,  
Printing House, Trinity College, Dublin 2, Ireland  
dfurlong@tcd.ie

### ABSTRACT

This paper presents a sonification listening model built from models of embodied cognitive meaning-making faculties. The aim of such a model is to aid in understanding how meaning is applied to auditory stimuli at the cognitive level. This in turn can aid auditory display designers in creating more effective auditory displays. The concept of ‘scale’ in sonification is considered in relation to the faculties described in the model. An experiment that explores how embodied auditory cognition, as described by the model, understands and interprets sonifications is then presented. This examining two specific kinds of ‘scale models’ listeners employ to interpret a sonification. The results obtained from this experiment are particularly convincing showing that a listener’s knowledge of the data-set being sonified will determine how they interpret changes in the auditory stimuli in a sonification. The existence of these scale models, the impact of a listener’s knowledge on their perception of a sonification and the implications imposed by the embodied nature of auditory cognition suggest a new avenue for auditory display researchers interested in developing meaningful sonifications that exploit the embodied nature of auditory cognition.

### 1. INTRODUCTION

There is a strong precedent in cognitive musicology and perceptual studies towards embodied cognitive models for auditory and musical cognition [1], [2], [3], [4]. These approaches treat auditory cognition as an embodied phenomenon that relies heavily on functions and processes determined by emergent parameters of lived bodily experience. The embodied approach suggested here differs from the standard approaches to auditory cognition in a number of important ways. The idea that auditory cognition is about the interception and computation of sonic symbols is flatly rejected. The listener is not a computer; the listener an embodied and environmentally bound organism. Auditory cognition is about making the chaos of raw auditory perception meaningful enough to enable the listener to operate effectively in their environment [5],[6],[7],[8],[9]. That environment is richer than the one presented in ecological psychacoustics as it encompasses social, cultural and political levels beyond the merely physical habitat. Auditory stimuli are not subjective impressions of an objective mind independent reality, nor are they purely mental abstractions divorced from extra-mental causal factors. They emerge in the dynamic bodily-mediated relationship between

subjective impression and objective motivation. That bodily mediation is the key factor to understanding auditory cognition as it provides the experiential grounding for mental content that is required for an auditory stimuli to have any meaning to a listener at all [1],[2],[3],[4],[5],[6],[7],[8],[9],[10],[11]. A designer cannot hope to create a truly effective Auditory Display without a thorough understanding of how auditory cognition assigns meaning to stimuli. Embodied cognition does not refer to some special case of cognitive processing that becomes active when a human engages in physical activity. This is a misuse of the term that betrays a serious underestimation of the topic at hand. There is much more to embodied cognition than novel movement based interfaces for human computer interaction. Embodied cognition is defined by how it answers the age old mind-body problem, the question of how the human mind relates to physical matter. Embodied cognition says that the body, as mediator between mind and world, grounds all mental content in perceptual and sensorimotor experience. That means all thoughts, words, and imaginative acts are structured, constrained, motivated, and made both intelligible and meaningful only in relation to perceptual and sensorimotor experience. Embodied cognition bridges the Cartesian divide and offers a truly cohesive account of auditory meaning-making. Though there are many different theories on auditory meaning making, it is becoming increasingly hard to compete against the results of the 30 years of *empirically grounded* neuroscientific research behind the embodied approach [6]. The real value of Embodied Cognition to AD is that it describes how auditory stimuli become meaningful for a listener, a crucial point for the AD designer. This paper aims to start taking those results seriously by exploring the contributions that an embodied model of sonification listening can make to the field.

### 2. EMBODIED COGNITION

Before presenting the Embodied Sonification Listening Model some of the embodied cognitive faculties upon which it is built must first be presented and discussed. The cognitive faculties described here are embodied at both the neural and phenomenal level. They have evolved out of earlier sensorimotor and perceptual faculties and make use of those sensorimotor and perceptual faculties. They are embodied meaning-making faculties in that they make perceptual, conceptual and imaginative domains of experience meaningful through the addition of structure and content derived from perceptual and sensorimotor experience [6],[7],[8],[9].

#### 2.1. Embodied Schemata

Embodied schemata, sometimes referred to as image schemata, are *the* basic elements of human cognition that



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

provide the crucial bridge between embodied experience, and cognitive meaning-making. They are pre-linguistic models of the common patterns encountered across the multiple modes of perceptual and sensorimotor experience that emerge from having a human body constrained by physical and biological necessity. They are cross-culturally universal and develop at an early age to structure both auditory perception and auditory imagery. They provide the basic rules that govern logic, reasoning and meaning-making and are imaginatively extended to ground abstract domains of experience in familiar domains, further facilitating understanding, reasoning and meaning-making. All of this is carried out using the same neural circuitry that drives sensorimotor and perceptual activity [6][7][8][9].

They often reflect the simple logical structure that describes the interactions between landmarks and trajectories. This generality means they can apply to a broad range of experiences.

For example, the PATH schema consists of three components. There are two landmarks (LMs) - the “source” and the “goal”. The source is where the path begins and the goal is where it ends. The third component is the path between source and goal. The behaviour of a trajectory (TR) is constrained by specific logical entailments of the LMs.

The SCALE schema is an extension of the PATH schema that includes cumulativeness, normativity and fixed directionality in the trajectory component. A TR traveling the path from source to goal in the SCALE schema is increasing in value. Progress along the path will have a valence being either a positive or a negative thing. The path between source and goal is set in place and the further a TR moves along the scale the greater the intensity [7]. Embodied schemata exhibit a bi-polar structure where one pole tends to have a positive, and its opposite a negative valence. This valence is transferred onto any structure to which the image schema is mapped.

## 2.2. Domains

Domain is a term used loosely across EC research to describe, not a cognitive faculty, but any coherent area of human experience sensorimotor, conceptual or otherwise. In essence they are collections of embodied schemata, concepts, metaphors and blends relating to a specific topic. It is a basic structure of knowledge. Sensorimotor domains are those domains of embodied experience uncovered through perceptual and bodily action. For example when one uses the word ‘running’ to reference their experiences of running, they are referencing a sensorimotor domain. Conceptual domains are abstract. They revolve around a specific topic of knowledge and contain concepts and relationship roles relevant to the understanding of that topic. For example when one says ‘running’ to refer to the abstract idea of bi-pedal locomotion, and all that goes with it, they are referencing a conceptual domain. All conceptual domains, including the fantastical, are ultimately grounded in sensorimotor domains. Domains are sometimes described in terms of Fillmore’s (1976) Frame Theory [8]. For example, the *fishing* domain would include elements like, fish, fishing-rod, river, casting out and reeling in and also the relationships between these elements. Conceptualising a domain as a frame encourages a focus on the relationships between elements of the domain.

## 2.3. Cross Domain Mapping

Cross-domain mapping is one of the simplest and yet most important cognitive faculties discussed here. It also forms an integral part of the other two faculties, metaphor and blending, discussed shortly. Cross-domain mapping maps elements from a source domain onto elements of a target domain to add an additional level of meaning to the target domain [8]. For example when one says ‘The TV is too loud’ they are mapping the concept ‘loud’ from the ‘sound’ domain into the ‘TV’ domain. This is a cross-domain mapping at the conceptual level, but this faculty spreads far beyond conceptual and linguistic domains. Cross-domain mapping structures and binds one’s basic perception of reality by weaving together the chaotic strands of embodied multi-modal perception and integrating them with conceptual reality to create a cohesive and meaningful world.

When one perceives the sound of a roaring engine as emanating from under the hood of a car to which they are attending visually a cross-domain mapping from the auditory to the visual domain is taking place. When one perceives the smell of baking bread, and sound of a roaring fire as emanating from bakers oven, to which they are attending visually, a cross-domain mapping is taking place between olfactory, auditory and visual domains. The critical importance of cross-domain mapping to both perceptual and conceptual experience cannot be stressed enough. The process is also integral to the perception and interpretation of sonification. For example, when auditory icons are presented alongside visual messages in a human computer interaction context the listener unconsciously performs a cross domain mapping that relates the auditory stimulus to the visual message.

## 2.4. Conceptual Metaphor

Conceptual metaphor is a specific kind of cross-domain mapping where elements of a source domain are used to reframe a target domain [7],[8]. They can make an abstract and unfamiliar conceptual domain meaningful by presenting it in terms of a more familiar source domain. The classic example is highlighted in the phrase ‘Love is a Journey’ [8]. In that phrase the abstract conceptual domain of love receives a structure that is mapped from the familiar domain of journeys. One can then reason, make inferences, about ‘love’ as though it were a journey. One might infer, for example, that love has a beginning, middle and end and is typified by forward motion along that arc. The embodied schema that lends a journey its structure, the SOURCE-PATH-GOAL schema, is extended in this way to structure one’s concept of love. Like standard cross-domain mappings metaphorical mappings also structure one’s perceptual experiences framing one perceptual or sensorimotor domain in terms of another perceptual, sensorimotor or conceptual domain. For example the desktop metaphor in human computer interaction (HCI) frames what would otherwise be an unfamiliar and abstract virtual space in terms of an office desk space. This structures how a user understands, reasons about and interacts with the virtual space. Conceptual metaphor is one of the cognitive faculties underlying sonification. For example, in a Parameter Mapping Sonification (PMSon) of stock market data, the *conceptual* domain of whatever ‘stocks’ is being used, is mapped onto the *perceptual* domain of the sound stimuli [8]. This results in a listener interpreting the sound as a representation of the data and reasoning about the sound on the basis of any knowledge they may have about the data.

Although these kinds of mappings are automatically happening at a cognitive level on the side of the listener, it is imperative that sonification designers take them into account if they wish to understand how auditory stimuli in a sonification become meaningful. The sonification mapping strategy then should support the innate cross-domain mapping strategies that listeners employ if they are to be meaningful.

### 2.5. Mental Space

Like domains, mental spaces are not cognitive faculties but structures of knowledge. They are not static but come into existence as one thinks and talks, by grouping together numerous conceptual domains, and elements of local context, and framing them under a common embodied schemata. As such they are larger and more general than the individual conceptual domains that they contain. They are general mental models that act to aid in thinking and reasoning and disband once the task for which they were assembled is complete [9]. At the neural level they are activated sets of neurons. The concept is key to the idea of a conceptual integration network.

### 2.6. Conceptual Blending

A conceptual integration network (CIN) is a dynamic mental model that structures thought and reason and generates meaning and understanding. Conceptual integration is a process of cross-domain mapping between input mental spaces resulting in the emergence of new conceptual content and relations in a blended space [9]. Where cross-domain mappings alone bind domains together, and conceptual metaphors frame one domain in terms of another, CINs integrate multiple input spaces through processes referred to as composition, completion and elaboration. Composition juxtaposes elements of the input spaces in relation to one another. Completion rounds out this composition to frame it in terms of a familiar embodied schemata while elaboration runs the blended space through an imaginative mimetic simulation on the basis of the principles and logic of the blended space. CINs also contain a generic space, which unites the content of the input spaces. Meaning arises in the blend as a function of the entire network as a whole. Like the other cognitive faculties described here it is an imaginative process that extends concepts, domains embodied schemata to create meaning. CINs are thought to account for auditory meaning-making in electroacoustic music [10] [4], and in harmonic music [1] [11]. The minotaur, a mythological creature of ancient greek descent, provides a simple example of a blend. In the concept of the minotaur the input space of man and bull are mapped into the blend space of a CIN to give rise to a new element the minotaur. The sonification model described in this paper is a further example of a conceptual integration network.

## 3. SONIFICATION LISTENING

The perception of sonification is like a strange mixture of disjointed elements from Chion's "Casual Listening" [12] and Smalley's "Source Bonding" [13]. One listens to a sonification to gain insight into a data-set rather than a sound source, and rely on innate parameters of auditory cognition and perception rather than semantic abstraction. Sonification listening is a categorically different type of listening experience, and the cognitive processes that underlie it reflect

this difference. Sonification listening can be modelled and considered as a CIN. There are no less than four mental spaces in a CIN where a cross domain mapping from a 'data space' to a 'sound space' motivates an emergent blended space to facilitate interpretation and understanding.

### 3.1. Sound Space

The sound space contains our actual perception of sonic parameters like pitch, timbre and tempo. The wider sonic space contains broader associations which might be made between sonic parameters and extra-auditory parameters in the form of concepts, domains, schemata and metaphors.

For example, if a violin timbre is used it may evoke associations to the concept of a musical instrument, the domain of orchestral string instruments, common schematic transformations (patterns of physical activity, harmonic/melodic structures) familiar to violin-based music of a certain era, and metaphors for movement (e.g. tremolo passages reflecting speedy movement).

### 3.2. Data Space

The Data space does not contain the actual data itself, rather it contains what is known about the data and any broader associations that might be made between that knowledge and other relevant background knowledge in the shape of concepts, domains, other mental spaces, category mappings, schemata and metaphors.

For example, when dealing with unemployment data, and associated spaces of economics and social issues, the domains of work and money, and metaphors such as loss may be involved. Any previous knowledge or expectations we have about the unemployment rate during the years in question will also come into play. It is important to remember that the actual objective data record itself does not come into play.

Critical to the data spaces are data concepts. The data space can contain at least one of at least two categories of data domain. These are the Amount data domain, and the Attribute data domain. Amount data is data where each data point is a measure of some amount of objects. Attribute data is data where each point is a measure of the value of some property of either (a) a single object, or (b) a single commonly shared property across multiple objects.

### 3.3. Sonification Listening Metaphor

The Sonification Listening Metaphor (SLM) is what sets sonification apart from other aural practices like music or conversational listening. Embodied metaphors make sense of abstract domains by framing them in terms of familiar domains. The SLM frames the sound space in terms of the data space. As a result, the listener interprets changes in the sonic stimuli *as* changes in the data. For the listener, at both the cognitive and phenomenal level, the sonic stimulus is understood *as if* it were the data itself. The knowledge that defines the listener's understanding of changes in the data now defines the listener's interpretation of changes in the sonic stimulus. The sonification does not become meaningful until attended to in this way.

### 3.4. The Sonification Blend

When a listener listens to a sonification in order to understand the data they are reframing the sonic stimuli in

terms of the data space. This creates a new level of meaning for the listener in the form of a scalar model. Changes in the sonic stimuli are understood as changes in the data changes along a specific scale. Scale emerges as a result of the structure that the SLM imposes on the listener's sonification experience. This is the point where all of the spaces have integrated into a seamless whole in the listener's experience, and the sonification is a meaningful spectrum of data changes expressed along a sonic scale.

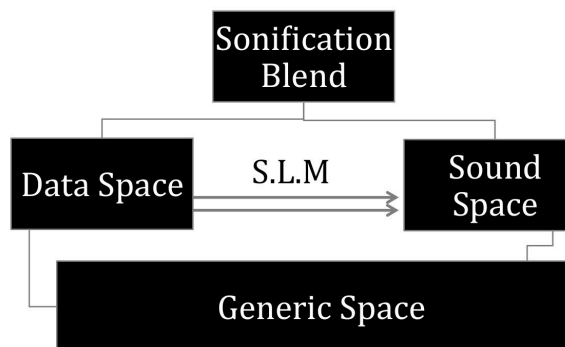


Figure 1: The Embodied Sonification Listening Model.

## 4. SONIFICATION SCALES

### 4.1. Scale in Sonification Listening

The projection of a SCALE schema [7] onto auditory parameters is crucial to sonification. Changes in auditory parameters can only be meaningfully interpreted as changes in the data because of the emergence of a SCALE schema in the blended sonification space. For example, in standard pitch mapping the SCALE schema structures the mapping of the pitch parameter so that changes in pitch are interpreted as changes in the data. Pitch becomes a scale along which data is expressed.

### 4.2. Types of Cognitive Scale

Two specific types of scale, “Amount” and “Attribute”, have been identified as objects of interest for this paper. Amount scales are used to comprehend collections of discrete entities, and Attribute scales to comprehend the values of properties within a single entity. An Amount scale occurs when the SCALE schema [7] is projected onto a COLLECTION schema [14] so that elements of the SCALE can be quantified (counted) in terms of the SCALE. Amount scales are used to make sense of data that deals with collections of discrete entities e.g. population data. An Attribute scale occurs when the SCALE schema is projected onto an ATTRIBUTE schema [15] so that degrees of an ATTRIBUTE can be quantified (valuated) in terms of the SCALE. Attribute scales make sense of data that deals with the proprietary attributes native to one specific entity e.g. measurements of thickness.

### 4.3. Amount Scales in Sonification.

The Amount scale is often designed into PMSon mapping strategies by way of vertical pitch mapping where pitch is treated as a scale, with low pitches representing low values

and high pitches reprinting high values. This type of mapping is grounded in subjects' physical experiences of MORE and LESS.

When more objects are put into some container the overall physical level of the object(s) within the container rises [15]. This provides the basis for multiple metaphorical mappings where AMOUNT is interpreted in terms of VERTICALITY and MORE is interpreted in terms of HEIGHT [16]. In PMSon, this mental processing is applied to the pitch domain to create a basic Amount scale.

The Amount scale can also be designed into a mapping strategy by mapping an increase in rhythmic repetition (tempo) of some note to an increase in the data. This can be explained as a very direct cross-domain mapping that interprets notes in terms of amount so that more notes means more numbers in accordance with the MORE-LESS schema [7].

### 4.4. Attribute Scales in Sonification.

When a sonification is supposed to represent attribute data, i.e. a qualitatively different kind of data, these scales can be confusing to a listener. Increases in attribute on a level observable in human experience do not necessarily relate to increases in height, or in the amount of objects presented in a certain time frame. Increases in Attribute values are often questions of increasing intensity and degree within some property of an individual object, rather than increases in amount across a group of objects. Additionally, increases in attribute properties often manifest differently in the spatial domain. For example, in the concept of WEIGHT, MORE weight causes an object to move DOWN in the spatial domain and LESS weight causes an object to move UP. For WEIGHT, DOWN maps to MORE and LESS maps to UP. This is based in experiential knowledge of heavier objects exerting more downwards force than lighter objects. So while an increase in weight maps to an increase in value [16] it also maps to a decrease in verticality. This is the exact opposite of the Amount scale where an increase in amount maps to an increase in verticality.

## 5. EXPERIMENTAL INVESTIGATION

The Embodied Sonification Listening Model predicts that a listener's interpretation of a sonic stimuli will change based on their understanding of the data represented by represented in the sonification. It also suggests that there are at least two kinds of scale employed by a listener in the interpretation of a sonification. Amount scales are applied to interpret sonic stimuli when the listener believes the data values represented are values of amounts. Attribute scales are applied to interpret sonic stimuli when the listener believes the data represents values of specific attributes of objects. Amount scales map increases in data value to increases along auditory dimensions, e.g. pitch or tempo, while Attribute scales are more likely to map increases in data value to decreases along those dimensions. The following experiment is intended to test that hypothesis.

### 5.1. Participants

Participants were recruited from a large international pool via online crowdsourcing platform Crowdfunder.com. Experimental materials were hosted on Survey Gizmo:

<http://www.surveymogzmo.com/>. Participants were compensated for their participation. There were a total of 112 participants, 83 male and 29 female. The age ranges break down as follows:  
Under 18 = 2; 18-24 = 2; 25-34 = 38; 35-54 = 38; 55+ = 4.

## 5.2. Experimental Procedure

Participants take part in a short training exercise to familiarize themselves with the procedure before beginning. They are instructed to listen to the sounds as many times as needed to before answering a question. Listeners are presented with a stimulus and told that it represents a specific data-set e.g. depth of a body of water, population of a town. After listening they are asked to determine whether the values in the data are increasing or decreasing over the course of the clip. Each stimulus is presented twice and listeners are told that the stimulus represents a different type of data on each presentation.

Listeners are said to have categorised a stimulus using an amount scale when they identified an increase in pitch or tempo as an increase in data value or a decrease in pitch or tempo as a decrease data value.

Listeners are said to have categorised a stimulus using an attribute scale when they identified an increase in pitch or tempo as a decrease in data value or a decrease in pitch or tempo as an increase in data value.

## 5.3. Audio Specifications

Twelve individual stimuli of 30 seconds length each are used in the experiment. 6 stimuli show linear variation in the frequency of a sine tone over time and the other 6 show linear variations in the tempo of a sine tone of 440hz frequency over time.

3 of the pitch stimuli begin at 440hz and rise to 1760Hz over the course of the clip. The other 3 begin at 1760hz and fall to 440hz. Tempo is modulated by an amplitude envelope.

3 of the tempo stimuli begin at 120bpm and rise to 600bpm. The other 3 begin at 600bpm and fall to 120bpm.

## 6. RESULTS

These results were analysed with a McNemar test on the twelve pairs of stimuli (6 rhythm and 6 pitch), comparing perceptions of positive and negative valence for indications of Amount type data vs. indications of Attribute type data.

### 6.1. Dissimilarity Testing

For pitch stimuli B, quite a significant effect was recorded. ( $X^2(1, N = 112) = 19.1, p = 0.001, \phi = 1.80$ , the odds ratio is 3.75). This strongly suggests that acceleration is interpreted using the Amount scale, and although thickness is interpreted primarily using the Amount scale, a statistically significant amount of listeners interpret it using the Attribute scale.

For pitch stimuli C, a significant effect was recorded. ( $X^2(1, N = 112) = 7.4, p = 0.01, \phi = .70$ , the odds ratio is 2.54). This strongly suggests that temperature is interpreted using the Amount scale, and although volume is interpreted primarily using the Amount scale, a statistically significant amount of listeners interpret it using the Attribute scale.

Stimuli	Data-Type	Amount scale	Attribute scale
<i>Pitch</i>			
A	Amount of Students	75.9%	24.1%
	Surface Area	68.8%	31.2%
B	Thickness	51.8%	48.2%
	Acceleration	81.3%	18.7%
C	Volume	67.9%	32.1%
	Temperature	83%	17%
D	Empty Space	60.7%	39.3%
	Amount of Ants	82.1%	17.9%
E	Concentration	83%	17%
	RPM	87.5%	12.5%
F	Mass	65.2%	34.8%
	Population	79.5%	20.5%
<i>Tempo</i>			
A	Strength	65.2%	34.8%
	Stiffness	61.6%	38.4%
B	Depth	49.1%	50.9%
	Absorbency	83%	17%
C	Production-Rate	59.8%	40.2%
	Angle	63.4%	36.6%
D	Hardness	64.3%	35.7%
	Stock Count	82.1%	17.9%
E	Yield	58.9%	41.1%
	Tension	81.3%	18.7%
F	Size	70.5%	29.5%
	Speed	71.4%	28.6%

Table 1: Experimental Results.

For pitch stimuli D a significant effect was recorded ( $X^2(1, N = 112) = 16.94, p = 0.001, \phi = 1.60$ , the odds ratio is .17).

This suggests that amount of ants is strongly interpreted using the Amount scale, and empty-space although is interpreted primarily using the Amount scale, a statistically significant number of listeners interpret it using the Attribute scale.

For pitch stimuli F, a significant effect was recorded. ( $X^2(1, N = 112) = 7.11, p = 0.01, \phi = 0.67$ , the odds ratio is .38). This suggests that amount is population interpreted using the Amount scale, and although mass is interpreted primarily using the Amount scale, a statistically significant number of listeners interpret it using the Attribute scale.

For tempo stimulus B, quite a significant effect was recorded. ( $X^2(1, N = 112) = 26.74, p = 0.001, \phi = 2.52$ , the odds ratio is 5.75). This strongly suggests that depth is interpreted using an Attribute scale, while absorbency is interpreted using the Amount scale.

For tempo stimuli D, a significant effect was recorded ( $X^2(1, N = 112) = 9.52, p = 0.01, \phi = .89$ , the odds ratio is .35). This strongly suggests that hardness is interpreted using an Attribute scale, while stock is interpreted using the Amount

scale. This suggests that stock is strongly interpreted using the Amount scale, and although hardness is interpreted primarily using the Amount scale, a statistically significant number of listeners interpret it using the Attribute scale.

For tempo stimuli E, a significant effect was recorded ( $X^2(1, N = 112) = 15.24, p = 0.001, \phi = 1.44$ , the odds ratio is 4.12). This suggests that tension is strongly interpreted using the Amount scale, and although yield is interpreted primarily using the Amount scale, a statistically significant number of listeners interpret it using the Attribute scale.

## 6.2. Similarity Testing

Pitch stimuli A and E, and rhythm stimuli A, C and F showed no significant effect. This means that the results for Attribute and Amount tests were quite similar.

For Pitch stimuli A, testing compared amount with surface area to examine their similarity ( $X^2(1, N = 112) = 1.6, p = 0.2, \phi = .15$ , the odds ratio is .66). This indicates that both stimuli were being interpreted in the same manner. This would imply that surface-area tends to be interpreted with a Amount scale and not an Attribute scale.

The results for Pitch stimuli E showed no significant difference between Concentration and RPM and rated towards the criteria that suggest a Amount scale ( $X^2(1, N = 112) = 1.31, p = 0.25, \phi = .12$ , the odds ratio is 1.71). This indicates that listeners rate Concentration using the same kind of scale with which they rate RPM, a vertical Amount scale where high means more and low means less.

Rhythm stimuli A compared two attribute values to one another - Strength to Stiffness. It showed a no significant difference in interpretation of with 39 of 112 interpreting Strength with an Attribute scale, and 43 interpreting Stiffness with an Attribute scale. ( $X^2(1, N = 112) = .38, p = 0.25, \phi = .03$ , the odds ratio is .82). This indicates that listeners rate Strength using the same kind of scale with which they rate Stiffness, when interpreting the rhythm dimension. This is primarily a vertical Amount scale where high means more, and low means less, but also a vertical Attribute scale where high means less and low means more.

The results for Rhythm stimuli C also showed no significant difference between Production Rate and Angle and weighted towards the criteria that suggest an Attribute scale ( $X^2(1, N = 112) = 1.31, p = 0.25, \phi = .12$ , the odds ratio is 1.71). This indicates that listeners rate Production Rate using the same kind of scale with which they rate Angle, primarily a vertical Amount scale where high means more, and low means less, but also a vertical Attribute scale where high means less and low means more.

The results for Rhythm stimuli F also showed no significant difference between Size and Speed and weighted towards the criteria that suggest an Attribute scale. ( $X^2(1, N = 112) = 3, p = 0.05, \phi = .28$ , the odds ratio is .33). This indicates that listeners rate Size using the same kind of Rhythm scale with which they rate Speed, when interpreting the rhythm dimension. This is primarily a vertical Amount scale where high means more, and low means less, but also a vertical Attribute scale where high means less and low means more.

## 7. DISCUSSION

The results show that whether people think they are listening to an Attribute value of some entity, or an Amount value of a number of entities will impact their perception of a sonification. This happens in a predictable manner. An

Amount value usually has a direct vertical mapping where increase in parameter is interpreted as an increase in data value. An Attribute value tends to be more ambiguous, with a significant tendency towards an inverted mapping where an increase in parameter is interpreted as a decrease in data value.

These differences in interpretation between Amount values and Attribute values speak to the value of taking the embodied cognitive aspects of sonification design (and specifically the SCALE schema) into account when designing mapping strategies. The study highlights the need for congruence between cognitive mappings and data-to-sound mappings and also lays out some key advice for the mapping of Amount and Attribute data in a sonification.

The variances in interpretation across the two types of scale discussed above suggest that the processes of embodied cognition described in this paper are active in sonification listening. Metaphorical cross-domain mappings from the data space to the sound space are framed in terms of a sonification listening strategy sufficiently conditioning the emergence of the blended space which the listener experiences as the sonification. A scalar model emerges in this blended sonification space as a result of cognitive processes of composition, completion and elaboration. It is this scalar model, which makes a sonification what it is - a scalar auditory representation of some data set or other.

The experiment offers some interesting insights into the cognitive mapping strategies employed in sonification listening. For example, one might presume that a significant amount of people would interpret SIZE using an Attribute scale because SIZE is listed as a basic Attribute schema by both Lakoff [17] and Johnson [7]. 29.5% of listeners interpreted size on the attribute scale. This suggests that there is a tendency for listeners towards the attribute scale for size. However, this result is not as strong as one would expect based on both the literature and the other results presented here, which tended to be higher than 29.5%. This implies that the scale for speed (the concept against which size was contrasted) also has a tendency towards the Attribute scale with 26.5% of listeners interpreting it that way. 24.1% interpreted the number of students using an Attribute scale. Though this was not a strong result, it was also an unexpected one as the number of students satisfies the description of a Amount type data concept. Concentration and Angle also showed a very strong tendency towards the Amount type, which was also unexpected as both satisfy the conditions of an Attribute data type. Strength and Stiffness were both expected to show a significant trend towards the Attribute scale. They did, with 34.8% and 38.4% respectively choosing the answer that co-related to an Attribute scale. The strongest attribute results came for data concepts like Thickness: 48.2%, Production Rate: 40.2%, Depth: 50.9% Crop Yield: 41.1%, Free Space: 60.7%. Most of these data concepts share a spatial component. This may suggest that the attribute scale works best with spatial concepts or that there is a spatial component to attribute schema.

## 8. CONCLUSIONS

This paper opens with a discussion of the need for a serious account of Embodied Cognition as it relates to sonification listening. It then discusses the embodied cognitive faculties that are employed by a listener in auditory cognition. An Embodied Sonification Listening Model which explains sonification listening in terms of these cognitive faculties is then presented. This is followed by a discussion on the

importance of scale in sonification and an account of the two models of scale, the amount scale and attribute, suggested by the sonification model to account for a listener's experience of scalar data changes in a sonification. Some persuasive experimental evidence for the existence of these scales is then presented before a discussion of their relevance. The two scalar models are the amount scale and the attribute scale. Listeners apply an amount scale to interpret data values that represent amounts of objects. This scale interprets increases in auditory parameters, pitch and tempo, as increases in data values and decreases in those parameters as decreases in the data. Listeners apply an attribute scale to interpret data values that represent the values of specific attributes of an object(s). The attribute scale interprets increases in auditory parameters, pitch and tempo, as decreases in data values and decreases in those parameters as increases in the data.

These results are of critical importance to the design of effective sonification mapping strategies as they highlight an underlying bias that exists in a listener, as a result of the embodied nature of their auditory cognitive faculties, towards interpreting and reasoning about a sonification differently in terms of their knowledge and beliefs about that sonification. Furthermore the results show that this process of filtering a sonification stimulus through previous embodied knowledge is not random but highly systematic and can be predicted by the application of the Embodied Sonification Listening Model. This opens a promising new avenue for AD practitioners who wish to develop richly communicative sonifications that are designed to fit with a listener's innate meaning-making faculties.

## 9. REFERENCES

- [1] L.M. Zbikowski, *Conceptualizing music: Cognitive Structure, Theory, and Analysis*. New York, US: Oxford University Press, 2005.
- [2] M. Johnson. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago, IL: University of Chicago Press, 2008.
- [3] C. Brower, "A Cognitive Theory of Musical Meaning." Annual Meeting of the Society for Music Theory, Phoenix, 1996.
- [4] A. Cox, "The Mimetic Hypothesis and Embodied Musical Meaning". *Musicae Scientiae*, 5(2), 195-212, 2001.
- [5] S. Roddy, D. Furlong, "Embodied Aesthetics in Auditory Display." *Organised Sound*, 19(01), 70-77, 2014.
- [6] G. Lakoff, "Explaining Embodied Cognition Results." *Topics in Cognitive Science*. 4 (4), p773–785, 2012.
- [7] M. Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press, 1987.
- [8] G. Lakoff, M. Johnson, *Metaphors We Live By*. Chicago, IL: Univ. of Chicago Press. p8-170, 1980.
- [9] G. Fauconnier, M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Arizona, US: Basic Books, 2002.
- [10] G.S. Kendall, "The Feeling Blend: Feeling and Emotion in Electroacoustic Art". *Organised Sound*, 19(02), 192-202, 2014.
- [11] S. Barrass, P. Vickers, *Sonification Design and Aesthetics*. In Hermann, T., Hunt, A., Neuhoff, J. G., editors, *The Sonification Handbook*, pp145–71. Berlin, Germany: Logos Publishing House, 2011.
- [12] M. Chion, M., *Audio-vision: sound on screen*. West Sussex, UK: Columbia University Press, 1994.
- [13] D. Smalley, "Spectromorphology: Explaining Sound-Shapes". *Organised sound*, 2(02), 107-126, 1997.
- [14] G. Lakoff, R.E. Núñez, *Where Mathematics Comes From: How the embodied mind brings mathematics into being*. New York, US: Basic books, 2000
- [15] G. Lakoff, M. Johnson, *Philosophy in the Flesh: The Embodied Mind and its Challenges to Western Thought*. New York, US: Basic Books. pp3-602, 1999.
- [16] G. Lakoff, J. Espenson, and A. Schwartz, *Master Metaphor List*. Berkeley. CA: *Cognitive Linguistics Group*, 1991.
- [17] G. Lakoff, *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press, 1987.