

The Nature of Modality and Learning Task: Unsupervised Learning of Auditory  
Categories

A dissertation presented to  
the faculty of  
the College of Arts and Sciences of Ohio University

In partial fulfillment  
of the requirements for the degree  
Doctor of Philosophy

Phillip A. Halsey

August 2015

© 2015 Phillip A. Halsey. All Rights Reserved.

This dissertation titled  
The Nature of Modality and Learning Task: Unsupervised Learning of Auditory  
Categories

by  
PHILLIP A. HALSEY

has been approved for  
the Department of Psychology  
and the College of Arts and Sciences by

Ronaldo Vigo  
Associate Professor of Psychology

Robert Frank  
Dean, College of Arts and Sciences

## **Abstract**

HALSEY, PHILLIP A., Ph.D., August 2015, Psychology

The Nature of Modality and Learning Task: Unsupervised Learning of Auditory

Categories

Director of Dissertation: Ronaldo Vigo

Categorization and concept-learning has a long-standing influence on the field of psychology because the notions of concept-learning are key to how individuals learn. Central to this idea is; how do we categorize stimuli that vary according to different dimensions? How do we categorize stimuli under different conditions? How do we store these categorizes as mental representations? And does the modality of the stimuli affect our construction of a mental concept, and to what extent does this affect categorization behavior? To partially answer this last question, it has been determined that the modality of a stimulus does influence categorization behavior but the extent of this is unknown. The current dissertation explores the manner in which stimulus modality, relationships between stimulus dimensions, and learning method affects categorization behavior. Two experiments are conducted in order to examine the auditory dimensions individuals attend to when making comparisons, and how individuals spontaneously categorize auditory stimuli based on the attended dimensions. Participant's data was then examined according to three models of unsupervised learning: the simplicity model, SUSTAIN, and GISTM.

## Dedication

*To my parents for their support*

*And to my cats for telling me when it was dinner time*

### **Acknowledgments**

First, I would like to thank my advisor Ronaldo Vigo for – essentially - accepting me to the program and helping me to further my education in mathematical modeling and cognitive science. Without his support of my research and his compassion for helping me with any obstacles, I certainly would not have made it to this point. I also most certainly would not have ended up where I am without his knowledge and interest in seeing students succeed.

Second, I would like to thank all the members of my committee – Steve Evans, Keith Markman, Robert Briscoe, and Mark Phillips – for reading this sizable document and sitting through, not just one but two, lengthy presentations. Your input most certainly made this dissertation better.

Finally, I'd like to thank Emmanuel Pothos for providing me with a copy of the simplicity model code (as well as great documentation on how to use it) and Todd Gureckis for his help with the SUSTAIN code.

## Table of Contents

	Page
Abstract.....	3
Dedication.....	4
Acknowledgments.....	5
List of Tables .....	8
List of Figures .....	9
Chapter 1: Introduction.....	11
Chapter 2: Theories of Concept Learning.....	16
Rule-Based Theories.....	16
Representational Theories.....	18
Prototype theory.....	18
Exemplar theory.....	20
Complexity Reduction Theories .....	24
Ideotype Theory.....	26
Chapter 3: Overview of the Dissertation .....	29
Chapter 4: Categorization .....	33
Supervised Learning .....	33
Unsupervised Learning.....	39
Chapter 5: Auditory Stimuli.....	44
Auditory Dimensions .....	44
Frequency and pitch.....	44
Amplitude and loudness.....	49
Amplitude envelope.....	52
Attack .....	52
Decay.....	53
Sustain .....	53
Release.....	53
Timbre.....	55
Chapter 6: Integral Dimensions .....	61

	7
Chapter 7: Auditory Categorization.....	67
Chapter 8: Current Research.....	75
Chapter 9: Experiment 1 .....	79
Method .....	79
Participants.....	79
Stimuli.....	79
Procdedure .....	80
Results.....	83
Chapter 10: Experiment 2 .....	89
Method .....	89
Participants.....	89
Stimuli.....	89
Procedure .....	89
Model Analysis .....	92
Simplicity model.....	92
SUSTAIN.....	98
Generalized structural invariance theory .....	99
Results.....	105
SUSTAIN.....	109
Simplicity model.....	112
Simplicity model vs. SUSTAIN.....	117
Generalized structural invariance theory .....	119
Chapter 11: Discussion .....	128
General Discussion .....	128
Limitations and Future Directions .....	130
References.....	167

## List of Tables

	Page
Table 1. Auditory Values.....	136
Table 2. Auditory Stimulus Boolean Values .....	137
Table 3. Auditory Stimuli According to Boolean Values.....	138
Table 4. Stress per NMDS Model.....	139
Table 5. Participant Ratings for Each Dimension.....	139
Table 6. Pairwise Tests of Audio Dimensions.....	139
Table 7. Participant Groupings per Category Type .....	140
Table 8. SUSTAIN Categorization Predictions .....	141
Table 9. Simplicity Model Predictions .....	142
Table 10. Simplicity Model Computations.....	143
Table 11. Structural Manifolds for Each Type as Computed by GISTM.....	144
Table 12. Difficulty and Invariance Values per Category .....	145



## List of Figures

	Page
Figure 1. The 3[4] SHJ category family .....	146
Figure 2. Unsupervised learning task.....	147
Figure 3. Pothos and Chater (2002) stars.....	148
Figure 4: Wave fundamentals .....	149
Figure 5: Wave changes according to frequency .....	150
Figure 6: Amplitude envelope (ADSR).....	151
Figure 7: Audio in the time-domain.....	152
Figure 8: Audio in the frequency-domain (spectral envelope) .....	153
Figure 9: Experiment 1 stimuli presentation.....	154
Figure 10: Experiment 2 stimuli presentation.....	155
Figure 11: Shepard plot of one-dimensional NMDS .....	156
Figure 12: Shepard plot of two-dimensional NMDS .....	157
Figure 13: Shepard plot of three-dimensional NMDS .....	158
Figure 14: Shepard plot of four-dimensional NMDS .....	159
Figure 15: Shepard plot of five-dimensional NMDS.....	160
Figure 16: Scree plot of metric Shepard $R^2$ .....	161
Figure 17: Scree plot of non-metric Shepard $R^2$ .....	162
Figure 18: Graph of 3 dimensional NMDS solution.....	163
Figure 19: Computer screen during learning task.....	164
Figure 20: SUSTAIN predictions against the human data.....	165

Figure 21: Simplicity model predictions against the human data.....	166
---	-----

## Chapter 1: Introduction

During one's lifetime, it is often advantageous to make categorical determinations about particular stimuli— be it an object or sound – encountered within a particular environmental context. The process of categorization allows an individual to determine whether that stimulus is similar enough to a group of objects categorized from prior knowledge, or whether that stimulus represents something new or unique that might represent an entirely new group or category. To illustrate, imagine your friend has invited you to their residence for the first time. After arriving, you enter the home only to find that rather than the standard couches and chairs you're accustomed to seeing and using, on the floor lay several large amorphous bags filled with undetermined material. Assuming no prior knowledge about these objects (although the reader would easily identify them as beanbag chairs), a key question arises; namely, would these objects (beanbags) be considered chairs or would they form their own group? Perhaps you observed your friend sitting down on one of the objects; certainly given this event you could determine that these objects serve a similar function to an idealized chair— namely sitting – but would you categorize these objects as chairs if asked later to formulate a list of all the chairs you could recall? Would you consider the physical features of these objects as variations of what constitutes a “chair”? In an extreme example, what if your friend had only small pillows on the floor surrounding a table rather than chairs or beanbags?

In both examples, the limits of what an individual might consider “chairness” – possible and allowable dimensional variations of an object considered as a chair, such as

the beanbag chair and the pillows – are tested. In other words, how does an individual form a concept containing possible defining characteristics or dimensions of an object such as a chair? Furthermore, how does a concept defined by such characteristics influence an individual's future classifications in determining whether a particular object is or is not a chair?

Researchers in the field of categorization behavior are interested in exactly the above questions. How does a concept - or mental representation of a category – develop and become defined over time? And when presented with a unique set of stimuli, how do we learn to classify and organize these stimuli into new concepts? In studying these phenomena, researchers are interested in examining two key aspects of categorization and concept learning. First, researchers examine the learnability of a category – typically assessed via proportion of errors – and determine how the overall complexity or amount of variation within a category affects learnability. Then based on these results, they aspire to predict individual future performance on a particular category solely on the inherent complexity within that category (Feldman, 2000, 2006; Vigo, 2006, 2009). Such prediction of categorization and concept learning behavior is typically assessed via supervised learning tasks in which individuals categorize objects into experimentally defined categories according to relevant object features (a more detailed discussion of supervised versus unsupervised learning will follow in Chapter 4).

The second aspect of research examines various category structures that are relevant to the conditions influencing or directing an individual's categorization behavior. Specifically, when allowing an individual to freely categorize objects, will contextual or

experimental manipulations result in a different classification sorting? To use the example above, what conditions may result in an individual classifying or not classifying the beanbag as belonging to the category of “chair”? Researchers examining experimental or contextual manipulations employ an unsupervised learning task - also referred to as a free sorting task (Ashby, Queller, & Berretty, 1999) – in which the individual is able to freely categorize the stimuli based on one of the more relevant features.

Although tasks such as supervised and unsupervised learning are not necessarily modality specific, research in categorization has favored the examination of visual information representation (Murphy, 2002). Despite humans relying primarily on visual learning (Shiffrar & Pinto, 2002), it has been determined that detection and processing using other modalities such as audition does assist in the visual processing of information (Thomas & Shiffrar, 2010). An unseen animal emitting a sound or vocalization may inform us whether or not we are in danger. We determine this by categorizing the stimulus we just heard based on previous stimulus categories, and then generalize our behavior to that sound accordingly. A “woof” might incline us to categorize a sound a certain way and behave - possibly more relaxed; a shrieking sound in a castle may incline us to determine we’ve heard a banshee and run.

This is but one simple example and there are certainly a myriad of circumstances in which we may wish to categorize auditory stimuli. For example, imagine attending an orchestral performance and listening to the following instrumentation: a string section (i.e. violin, viola, cello, and bass), the woodwinds (i.e., clarinet, flute, oboe), a brass

section (i.e., French horn, tuba), and finally the percussion instruments (i.e., timpani, chimes, bass drum). As you close your eyes and listen to the music, how might you categorize the instruments you are hearing? Would you do so based on how the instruments sound, the notes played, or play style of the notes? Or, alternatively, would you categorize the sounds according to several of these criteria? It may be that you group certain string instruments together because their sound is similar in some capacity. But suppose the violin and brass sections play a portion of the orchestral piece in which the musical dynamics of the piece are identical? How does categorization occur in this instance? As the remainder of the orchestra, the horns and violins, continue to play, the variance of pitch and dynamics will also influence the ability to categorize. In this situation would you categorize the violins together based on what they're playing (e.g., pitch), or continue to categorize the violins with the other string instruments?

This scenario provides an example of the importance of how context and/or conditions influence categorization behavior. The condition of “similar strings” may lead to one unique classification whereas “similar violins and brass” may lead to another. Thus, given certain conditions, how do individuals freely categorize auditory stimuli? Namely, when specific salient dimensions define the auditory experience, on what basis will individuals categorize the sounds? As an example, would an individual categorize tones of a similar sound together (their timbre) regardless of the pitch being played by the two instruments? Or would an individual consider these stimuli as belonging to separate categories (both timbre and pitch used for categorization)? In what ways are humans purported to develop internal representations of categories? In the following research, I

investigated these questions, and how they affect categorization and concept learning behavior. But first, a brief overview of the current theories exploring categorization and concept learning is necessary.

## **Chapter 2: Theories of Concept Learning**

One of the primary goals of concept learning is to determine how individuals categorize stimuli and how they later retrieve these categorical relationships via a mental representation. When we make contact with a particular stimulus – for example, a whale – how do we form a conceptual, mental representation of the category “whale”? Furthermore, what structure does this mental representation take and how will the hypothesized conceptual structure influence categorization? For example, under what conditions might we categorize a whale as a mammal rather than a fish? Can researchers explain errors in categorization such as categorizing a whale as a fish rather than a mammal? Below are some of the primary types of theories that researcher developed in order to account for categorization behavior and theorize the nature of a concept’s structure.

### **Rule-based Theories**

In the history of concept learning, describing concepts as rules has been one of the earliest known representational paradigms. The theory behind concepts-as-rules is that individuals define concepts according to their features or attributes. As such, these definitions need to be both necessary and sufficient for the inclusion of only that concept and the exclusion of all others. Such definitions can be translated into rules which adhere to the notions of logic, specifically logical connectives such as “AND”, “OR”, and “IF”. Accordingly, because of the adherence to logic, definitions are determined to be either “TRUE” or “FALSE”. By complying with these basic elements of logic, a particular concept is either included or excluded from a category; the degree to which a concept



occupies a category cannot be ambiguous, as determined by the dichotomy of the truth-values. Thus, in this respect, the distinction between belonging to category A or category B is defined by the truth of the definitional statements made about the concept. If the presence of an object's attributes is determined to be true according to the definitions then it is included in the category; otherwise, a participant may consider to the object to belong in another category.

For example, an individual may develop the concept of a "tree" according the definitions of "has leaves" and "has a woody trunk". However, such a description would not include elements of the concept that are known to be trees but do not have leaves, such as evergreen trees. Thus, it would be necessary to increase the specificity of the definition to include this aspect about trees (e.g., "has a woody trunk", and "has leaves" or "has nettles"). By adding more definitions according to the logical conjunctives, more conditions are necessary to include any members of the concept. The revised definition of trees would include deciduous and evergreen trees that have a woody trunk, but such a definition would exclude young trees that have not yet developed such a trunk. It would be necessary to further revise the definition such that young trees are included. However, what of deciduous trees that that shed their leaves during winter? According to the current revised definition, trees during winter would not satisfy the definition of being a "tree". The individual would need to revise the concept definition repeatedly to satisfy the conditions of tree-ness under any contingency. As a result, the definition of the concept would be exceedingly complex with no common attribute between the trees that unities them all under the concept of "tree".

Early offshoots of rule-based theories - such as semantic networks (Collins & Quillian, 1969; Quillian, 1967) - had some success despite examining only semantic relationships. Similar to the example above, semantic networks begin with a general set of features and iteratively decompose the feature set until a group of specific features described only one object<sup>1</sup>. The field of concept learning has largely progressed beyond these theories due to their inherent inadequacies (Murphy, 2002; Rumelhart & Ortony, 1976) and many theories have arisen that directly attempt to address how conceptual frameworks develop from categorization behavior.

### **Representational Theories**

Beyond observing and explaining how individuals categorize or organize different objects, researchers developed representational theories in order to create a plausible explanation for how individuals store a category as a mental representation, or concept, and how individuals access and use these concepts are accessed and used during the categorization process. Two theories exemplify the representational perspective of concept learning behavior; Rosch's (1975) Prototype Theory and the Exemplar Theory (Medin & Schaffer, 1978; Nosofsky, 1984).

**Prototype theory.** Rosch (1975) conceived the representational paradigm of prototypes in response to the inadequacies of the rule-based paradigm. Prototypes, according to Rosch, were summaries of all the features of that concept; such summary

---

<sup>1</sup> For example, through this iterative process of increasing specificity, an individual can describe a word or object such as "penguin" using adjectives and verbs such as "black", "bird", "swims", "and "flightless". Depending on the other words within the network, this may be sufficient enough to categorize and describe a penguin.

representations described the features that were most prominent (or most frequently occurring) in the members within that concept. Features occurring more frequently have higher weights; therefore our concept learning process may assign higher weights to features that are more prominent in a particular concept. For trees, the attribute of “woody” might rate highly as the majority of trees have woody trunks (a more typical attribute); atypical attributes (such as “non-woody”) rate lower. Thus, due to the use of weighting on these summary representation features, contradictions are permitted (e.g., both “woody” and “non-woody” are connected to the concept of tree) and, more importantly, concepts can occupy a continuum.

One final important feature of the prototype paradigm, and not present in the rule-based paradigm, is intransitivity. To put simply, the prototype theory allows for violations of the hierarchical nature present in the definitional theory; the transitive property of  $A < B < C$  need not hold – C does not necessarily need to be “greater” than A or B (where the values are features).

Certain issues do exist with the prototype representation. First, prototype in Rosch’s exact definition of what constitutes a prototype is ambiguous and unclear and as a result, there has been little consensus as to what exactly qualifies a prototype. This ambiguity has led researchers to interpret the definition of a prototype in such a manner that facilitates their own research endeavors rather than creating one absolute definition. Second, the prototype model is unable to account for attributes or features that are defined continuously; for example, when precise measurements are used (e.g., length, volume) where values can take a nearly infinite amount of values, it is necessary to

partition these measurements into different categories (e.g., small, medium, and large). Naturally, this partitioning leads to information loss and precision in defining concepts (in some respects, this issue is analogous to those of the rule-based concept learning). Finally, according to the prototype paradigm, an individual does not necessarily learn features in isolation; rather they learn feature combination in order to describe a concept. The number of possible feature combinations can lead to a combinatorial explosion; that is, the number of feature combinations that can define a particular concept becomes intractable for a human to cognitively manage. This very obviously presents a significant problem with the prototype paradigm; if the computational costs of learning a concept via a few features of a prototype easily becomes extensive, humans will be unable to learn concepts with more than a couple of features. These issues and deficiencies contributed to the rise of other representational models that attempted to correct these issues.

**Exemplar theory.** Exemplars provided an alternative mental representation to that of prototypes, beginning with the writing of Medin and Schaffer (1978) and their proposed model of exemplar-based concept learning, the context model. An exemplar refers to instance of a particular stimulus that is stored in memory; thus, instances that share certain similar features form a concept. For example, light fixtures may vary in size, placement, and color, but the relative similarity of these instances of light fixtures would differentiate them from coat racks. When encountering a unique instance of an exemplar, the individual compares the similarity between the features of the unique exemplar and the features of exemplars stored in memory. The saliency of these features determines classification; specifically, if an individual may categorize stimuli together if

the salient features between the unique exemplar and the exemplars in memory are similar.

Thus, exemplar theory differs from prototype theory in a very meaningful way. Rather than compare a test stimulus to a single internally represented average stimulus that defines the category structure, an individual compares the test stimulus to a known set of instances belonging to that particular category. For example, when an individual is attempting to determine whether an object that lacks legs but has a back is a chair, the individual retrieves a group of chair representations from memory that tend to exemplify the dimensions of “chair-ness.” To the extent that the observed stimulus demonstrates similarity to stored exemplars, the individual may categorize the stimulus as an instance of the concept chair.

Medin and Shaffer’s (1978) context model examined categorization behavior with the underlying assumption that individuals learn objects in a given category set as exemplars. For instance, the exemplars in Medin and Schaffer’s experiment comprised visual stimuli that each occupied some point in psychological space. In order to retrieve an exemplar from memory, a test stimulus presented to an individual would act as a cue for exemplar. If knowledge of a particular exemplar was incomplete, then an individual may categorize the test stimulus in a manner similar to the exemplar it most closely resembles. For example, an individual presented with a test stimulus of 1100 (where each value of 1 or 0 in the 4-value sequence represents a single binary dimension of the stimulus, e.g., the first value in the sequence refers to stimulus color where 0 = black and 1 = white), might categorize the test stimulus as the exemplar “11?0” (where the ‘?’

represents incomplete information) rather than “0?11” because the latter is more dimensionally divergent from the test stimulus whereas the former is more dimensionally congruent. Medin and Shaffer argued that a multiplicative rule in conjunction with dimensional saliency parameters best describes how to determine the probability of categorizing an object in a particular manner.

Nosofsky proposed an extension of Medin and Shaffer’s (1978) context model in order to interpret categorization through the lens of choice and similarity. The resulting model, the generalized context model (GCM; Nosofsky, 1984) united three distinct components from psychological science and information science; the aforementioned context model, Shepard’s psychological distance metric (1957, 1958a, 1958b), and Luce’s choice axiom (1963).

Nosofsky formulated that, mathematically, for such a multiplicative rule to predict classification behavior, the underlying measure for determining stimulus similarity must be an exponential decay function such that the distance metric between the stimuli in multidimensional scaling space was based on the city-block. The foundations of this are traceable to Shepard’s work on distance metrics in multidimensional scaling (Shepard, 1957; 1958a) and more recent works (Shepard, 1987). The incorporation of both Shepard’s exponential distance metric (and subsequently the Minkowski-r metric) and parameterization of the context model allowed for the GCM to make computations concerning the perceived similarity of a test stimulus to an exemplar in terms of psychological distance in multidimensional space between the stimulus and the exemplar.

In order to make predictions about categorization behavior, the final component incorporated into the GCM is Luce's choice axiom (1963; 1977), a measure also incorporated into the context model (Medin & Shaffer, 1978), as a description of the probability of a response given a particular test stimulus given a set of exemplars. Luce's choice axiom pertains to the probability that a chosen stimulus was proportional to the strength the stimulus exerted over the behavior. Put simply, it is the percentage an individual will correctly classify a test stimulus as belonging to category A (or alternatively category B) given the combined set of both category A and B.

Taken together, Nosofsky's GCM is composed of three distinct elements; Medin and Schaffer's context model, Shepard's psychological distance metric, and Luce's choice axiom. Together, they demonstrate that the probability that a test stimulus will be classified as belonging to category A (or B) given the test stimulus' similarity to exemplars in psychological space is a function of both the exponential decay of similarity between the compared stimuli and exemplar and the Minkowski-r distance metric provided the individual attends to the appropriate features of the stimuli for correct classification. Nosofsky elegantly described the relationship between these three components as:

$$APC = 100 \frac{[\sum_{x \in X} P(X|x, D, w) + \sum_{y \in Y} P(Y|y, D, w)]}{D}$$

where the percentage of categorizing an object in the correct category is a function of the sum weighted psychological distance between a test stimulus and the stimuli in a particular category (equation 2) with respect to the psychological distance of the test stimulus to all stimuli in both categories (A and B). As such, the predicted percentage of

correct categorization can be determined for each test stimulus with respect to the categories. The accuracy of the GCM in predicting certain categories and categorization difficulty orderings result in its use as the standard to compare against other concept learning models.

### **Complexity Reduction Theories**

How individuals minimize or reduce objects into their most important – or diagnostic – features is the concept behind complexity reduction theories. Such theories predominately use rule reduction as the complexity reduction mechanism. In order to reduce a category's complexity, these theories state that individuals examine a concept using a general set of rules and iteratively reduce these rules while also preserving category's original elements. For example, given a Boolean expression of  $xy' + xy$ , where each literal is an object feature ( $x$  and  $x'$  are black and white,  $y$  and  $y'$  are triangle and circle, respectively), the expression could be minimized to  $x(y' + y)$  and still actually describe the objects. Stated differently, rather than describe the two objects as “a black circle and a black triangle”, the objects could be denoted as “a black circle and triangle.” In this case, the definition of the category remains the same, but stylistically the expression is smaller.

As another example, let's say that I work at a grooming facility where I exclusively wash cats. As such, I may want to expedite the process so that I finish with easy-to-wash cats first in order to attend to more difficult cases. Cats are extremely multi-dimensional; there's fat ones, small ones, ones with an excessive number of toes, ones with a normal amount of toes, ones with black ones, multi-colored ones. My goal



would be to reduce all this complexity and variation into an easy to work with concept to facilitate the process I want to accomplish - separate the cats so I can find the easy ones. I may learn that cats with polydactyl toes, regardless of color, are quite docile when I dunk them in the tub and scrub them down. Already I've reduced my complexity down to "polydactyl, and black or multi-colored" for the "easy" category. In terms of accomplishing my washing task, I'm making progress. But will size matter? It certainly may. Given what I already know, I might learn that fat ones with polydactyl also have tendencies towards violence against my person when I dunk them in water, whereas the small ones are perfectly easy to deal with. I can now, for the most part, successfully categorize the cats and may assign a label to each category as being either "Nermals" (small and polydactyl, and black or multi-colored) or "Garfields" (everyone else).

According to minimization theory, individuals achieve their categorization objective when they attempt to minimize the features of a data set to its most basic diagnostic dimensions in order to increase the learnability of the set (Feldman, 2000, 2003, 2006; Vigo, 2006). More specifically, Feldman (2000) suggested that the underlying complexity of a given set of objects (i.e., category) corresponds to the shortest Boolean expression length of that set. Thus, individuals presented with a set of objects associated with a highly minimized Boolean expression will be able to quickly and accurately categorize those objects, because the underlying expression may be simple (such as the example above). Conversely, an expression that is incompressible or lacking in minimization manipulations, such as  $x'y'z + xyz'$ , would be more difficult for individuals to categorize, due to the lack of minimization "shortcuts." Although Vigo

(2006) found that the Boolean factorization used by Feldman demonstrated a computation bias towards proving Feldman's argument, models of complexity reduction still exist in the field such as QMV factorization (Vigo, 2006) and the simplicity model (Pothos & Chater, 2002).

### **Ideotype Theory**

The Generalized Invariance Structural Theory (GIST; Vigo, 2013, 2014), an extension of Categorical Invariance Theory (CIT; Vigo, 2009), is a more successful notion of how individuals store and retrieve concepts. Unlike other formal theories that rely on probabilistic notions of concept learning, the GIST is a general, deterministic framework that thus far, based on historical and recent data, has proven to more accurately predict concept-learning difficulty than the leading theories, such as Feldman's (2000) above-mentioned Boolean minimization Nosofsky's (1984) GCM, and Goodwin and Johnson-Laird's (2011) mental models.

The primary notion behind the GIST - and its predecessor, CIT - is that individuals are pattern detectors. Namely, when given a category, individuals will try to detect patterns - referred to as invariants - across the available stimuli. The usage of invariants as a measure of performance directly echoes other sciences, such as physics, where certain attributes of an object remain invariant regardless of transformations or alterations. However, Vigo (2008, 2009, 2011, 2013, 2014) introduced a new notion of invariance which he named "categorical invariance". Vigo's notion is more general than previous notions and not limited by spatial intuitions.

As an example, suppose there is a set of three objects defined dimensionally according to the Boolean rules explained under “Complexity Reduction Theories.” Each object varies according to color ( $x = \text{black}$ ,  $x' = \text{white}$ ), size ( $y = \text{small}$ ,  $y' = \text{large}$ ), and shape ( $z = \text{circle}$ ,  $z' = \text{triangle}$ ). The resulting three objects are; a black, small circle ( $xyz$ , or 000), a white, small circle ( $x'yz$ , or 100), and a black, large circle ( $xy'z$ , or 010). To establish invariance, a single dimension is perturbed, or transformed along a single dimension. Objects present in both the original and perturbed set are invariants. As a brief example, perturbing the color dimension using our example stimulus category would result in the following set of objects:

$$\{100, 000, 110\}$$

Compared against the original set  $\{000, 100, 010\}$ , it is apparent that two of the objects - 000 and 100 - are preserved or are invariant under a color perturbation.

Dimensional perturbations occur across all dimensions until exhausted.

Continuing to the next dimension would result in a perturbed set of:

$$\{010, 110, 000\}$$

Despite the perturbation of the dimension of size, two of the object-stimuli within the original category remain in the set.

These invariances present across stimuli lay the foundation for what Vigo (2013) refers to as the dimensional binding process. These bound dimensions regulate the process of pattern detection. The dimensional binding process makes intuitive sense; if the same attribute exists between several objects, it's more relevant to disregard this attribute so that the individual can focus their attention on other attributes that may

facilitate concept learning. This goal-oriented behavior frees attentional resources to focus, and subsequently encode, the relevant category structure.

The proportion of detected invariants for a given stimulus dimension relative to the number of stimuli within a given category set comprises what Vigo refers to as the structural kernel (Vigo, 2013). As the name implies, a structural kernel is an encoding of the structural information for a given stimuli dimension. Structural kernels encoded for each dimension carry the relevant structural information for the given dimension of the category stimuli. The structural kernels are stored as a memory trace, or ideotype (Vigo, 2013). Ideotypes retain the structural information of the structural kernels for a particular category and, as such, provide information about the relative difficulty of category learnability and allow for an informed formation of rules. Thus, ideotype theory posits a process of pattern perception, invariance detection, redundancy reduction (through dimensional binding), and structural encoding.

Of the previously explained theories, current research in categorization and concept-learning behavior has focused on three theories: exemplar theory (most notably Nosofsky's GCM), simplicity reduction, and ideotype theory. Reflecting this, the current experiment will use these theory's relevant mathematical and computational models to examine concept-learning behavior.

### Chapter 3: Overview of the Dissertation

Both concept learning and auditory perception are expansive fields of study. The intersection between these two fields is still growing; relative to visual categorization, less research exists for auditory categorization (a quick scan of Murphy, 2002 demonstrates just how prevalent visual categorization is within the field). However, as it is important to understand the context and conditions under which individuals will form a concept based on visually perceived objects, it is equally important to learn these conditions when perceiving auditory stimuli. Stimuli are not necessarily always visually perceived first (e.g., you may hear the Wampa before you see him, as Luke did in *The Empire Strikes Back*), and thus auditory categorization and concept learning requires the same examination as visual learning. The following chapters provide an overview of the primary experimental paradigms used in categorization and concept learning study; specifically, supervised and unsupervised learning of categories. These two paradigms differ in terms of how researchers conduct the experiment and, more importantly, in their experimental aims.

Supervised learning seeks to predict categorization and concept learning behavior by using an individual's number or proportion of errors in categorization when categorizing two labeled categories (e.g., category A and B). In contrast, the purpose of unsupervised learning is not necessarily to predict or anticipate an individual's categorization performance; instead, researchers use unsupervised learning to determine how the set of conditions and environmental context affects categorization and concept learning when an individual freely categorizes unlabeled objects. Experimental data on

human (and artificial intelligence) visual categorization and concept learning under these learning tasks is readily available as it has been extensively documented and studied for years (Murphy, 2002). In comparison, human auditory categorization and concept learning has been documented to a lesser degree compared to visual categorization, with a research focus on artificial intelligence auditory categorization (e.g., Blumensath & Davies, 2004; Cai, Lu, & Hanjalic, 2005; Park & Glass, 2008; Van Segbroeck & Van hamme, 2009; Wulfing & Riedmiller, 2002). The following chapters will provide a more complete examination of these paradigms, including the categorization of visual objects and how to define these stimuli dimensionally, with specific intent to extend those notions to an unsupervised learning task of auditory stimuli.

The perception of an auditory stimulus, generated through either computer music and synthesis or acoustic instrumentation, results according to changes to dozens of parameter values or, in the case of acoustic instruments, physical changes or variations (e.g., Grey, 1977). Given this vast variability, research has attempted to find the contributing factors to multi-faceted auditory dimensions such as timbre; as a primer, the major dimensions of auditory stimuli are discussed with respect to their importance and usage in the current literature (e.g., Bonebright, 2001; Bulgarella & Archer, 1962; Clarkson & Pentland, 1999; Gao, Lee, & Zhu, 2004; Goudbeek, Swingley, & Smits, 2009; Guastavino & Katz, 2004). In addition, I will discuss auditory categorization with respect to these dimensions with a focus on non-speech auditory categorization. Though the importance of examining speech-related categorization cannot be understated, it will be evident that less research has examined categorization and concept learning of non-

speech sound. Of the categorization research conducted in non-speech sound, particular inquiries analogous to those in visual categorization and concept learning have arisen (Goudbeek, Swingley, & Smits, 2009); for example, under what conditions do we freely categorize sounds that we hear? When given a set of sounds that conform to well-defined dimension (audio with dimensions defined according to Boolean rules, e.g., amplitude as either 0 = soft or 1 = loud), how will individuals categorize these sounds? Will individuals attempt to categorize with an easily learned rule that lacks precision - such as categorizing according to a single dimension, e.g., instrument type - or will individuals use multiple dimensions during the categorization process in order to create more categories that more accurately reflect the myriad of varying dimensional attributes between different sounds? It is the intention of the current dissertation to examine these lines of inquiry and begin the groundwork for future, similar lines of inquiry concerning the nature of unsupervised auditory learning.

In the final chapters of the current dissertation proposal, I will discuss the experimental procedures for both experiments with emphasis on the interconnected nature between the experiments. The analyses to be conducted for each experiment of the dissertation will be discussed, including a final comparison between three mathematical and computational models' performance in relation to the participants' data during the categorization task. These three models will include; Pothos and Chater's simplicity model (2002), Love, Medin, and Gureckis' SUSTAIN (2004) algorithm, and the generalized structural invariance model (GISTM; Vigo, 2013, 2014). Pothos and colleagues (2005) previously established the performance comparison between the

simplicity model and SUSTAIN on an unsupervised learning task. These two models use the theories of complexity reduction and specifically analyze unsupervised learning tasks. The GISTM, a more recently developed model based on the notion of ideotypes and the principles of invariance is notably accurate on supervised learning tasks, but it is as of yet untested according to an unsupervised learning task.



## Chapter 4: Categorization

Research in the field of categorization and concept learning has a rich and detailed history, including those investigations on the phenomena pre-dating the Cognitive Revolution of the 1950's (Hull, 1920; Smoke, 1933). After the Cognitive Revolution and the unification of cognitive psychology with a highly interdisciplinary, interest in the way in which individuals - and algorithms, a term used here for simplicity to refer to neural networks, support vector machines, and other programs that learn data - categorize and learn object classifications increased. Because concept learning and categorization represents a high level behavior that encompasses low-level cognitive phenomena such as perception, attention, and others, understanding how individuals engage in concept learning behavior is key in understanding human behavior in general (Murphy, 2002). As such, researchers have historically investigated human concept learning behavior according to supervised and unsupervised learning research designs.

### **Supervised Learning**

In the machine learning literature, a supervised learning task was one in which labels were assigned to specific objects in order to denote category membership (Kotsiantis, Zaharakis, & Pintelas, 2007). Thus, if given two categories of objects, A and B, an algorithm's performance relates to the proportion of correct object classifications. Prior exposure to the stimuli and their labels would initiate a learning mechanism within the algorithm. For example, a neural network would adjust "neural" weights to accommodate the incoming information in preparation for the supervised learning task. Categorization accuracy indicates algorithm performance when prompted to categorize

unlabeled stimuli. To the extent that an algorithm demonstrated a high degree of accuracy on the categorization task, one may arrive at the belief that the algorithm is successful. Often, if the research pursuit is to develop an algorithm that demonstrates some key capacities of human category learning - such as the aforementioned SUSTAIN (Love et al., 2004) - the algorithm's performance will be contrasted on human performance on several influential studies within the concept-learning field.

Highly influential studies in cognitive psychology, including those by Bourne, and Shepard, Hovland, and Jenkins (Bourne, 1963; Bourne & Guy, 1968; Shepard, 1991; Shepard, Hovland, & Jenkins, 1961), act as these performance benchmarks, not only for that of algorithms, but also for mathematical models of human categorization behavior. In addition, these studies helped to establish the role of task and environmental context as a governing factor in concept learning behavior (Billman & Knutson, 1996; Wattenmaker, Dewey, Murphy, & Medin, 1986; Wisniewski & Medin, 1994). In these seminal categorization experiments, researchers presented visual stimuli to participants according to a specific procedure and asked participants to categorize the stimuli according to the stimulus features or dimensions. Variation of stimulus features adhered to a binary scale; for example, the dimension of color included only black or white values with no permissible gradations. Shape was either one of two values, such as a triangle or circle (Shepard, Hovland, & Jenkins, 1961). The number of stimulus features under examination ultimately determined the number of stimuli within the overall category set. The total number of objects in a category is:

$$p = 2^D$$

where  $D$  is the number of binary-varying dimensions and  $p$  is the total number of objects in the category (Feldman, 2000; Vigo, 2009). Thus, a category with four varying dimensions consists of a total of 16 category objects. Researchers then partition the total category in any way such that some objects are in the “positive” category and some objects are in the “negative” category, or alternatively category A and B (e.g., Shepard, Hovland, & Jenkins, 1961; Vigo, 2009, 2013).

Particular interest has been given to the category of four positive objects defined by three dimensions, generally referred to as the  $3[4]$  category<sup>2</sup> (Feldman, 2000; Goodwin & Johnson-Laird, 2011; Kruschke, 1992; Nosofsky, 1984, 1986; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard, Hovland, & Jenkins, 1961; Vigo, 2009, 2011, 2013). This category allows for the creation of six unique groups – or types – each consisting of eight objects. In an influential study, Shepard, Hovland, and Jenkins (1961) examined the learning difficulty of these six types. A participant’s categorization of each object within a category type was successful if they used a particular “rule” for each type. Type I objects could be categorized by using only a single dimension; if the difference between groups was shape, participants only needed to attend to the object’s shape (e.g., triangle or circle) in order to correctly categorize every object in the set. Thus, objects within the Type I category share one similar quality (e.g., shape). Type II objects required attention to two of the objects’ dimensions and used an exclusive or

---

<sup>2</sup> Note that Vigo (2013) extended this notation such that it can now denote gradations. For example, the notation  $3_2[4]$  would be equivalent to the previous notation of  $3[4]$ ; in both examples, four objects vary according to three binary dimensions. A set of features could take on three possible values rather than two binary values, for example  $3_3[4]$ . The subscript specifies the possible values or states of the features.

(XOR) rule to determine membership. As a concrete example, objects in one particular category had to be either black and triangular *or* white and circular. Without satisfying both those conditions, the participant should not categorize the object as a member of that particular group. Three of the four objects in types III, IV, and V categories adhered to a single rule - triangular, for example - while the fourth a combination of rules to categorization. Individuals could not categorize type VI objects by the simple rules such as those above; in fact, no particular rule allows object categorization without engaging memory (Shepard, et al., 1961). Figure 1 presents a possible visual construction of the 3(4) category using Boolean rules.

The results from the study indicated that participants experienced increasing difficulty in categorizing objects correctly as the number of dimensions necessary to attend to increased. That is, as the “rule” necessary for categorization became more complex, the proportion of correct responses decreased. Resulting from this, the following difficulty ordering (proportion of errors) was established;  $I < II < III, IV, V < VI$  (Shepard et al., 1961). Extensive replication demonstrates a distinctive difficulty ordering of the six types occurs when individual engage in a supervised learning task (Nosofsky, et al., 1994; Shepard, Hovland, & Jenkins, 1961). Due to the robust ordering that is produced, the 3[4] category has been used as a performance benchmark for mathematical modeling and prediction (Feldman, 2000; Kruschke, 1992; Nosofsky, 1984; Nosofsky, et al., 1994; Vigo, 2009, 2011, 2013).

It is important to note that while the difficulty order is a robust effect, the proportion of errors between each type (or learnability) is distinctly affected by the

experimental presentation (Vigo, 2011, 2013). For example, in the Shepard, Hovland, and Jenkins (1961) experiment, researchers showed participants a series of images each containing a single object from one of two possible category types. Participants responded categorizing the objects and received corrective feedback. The feedback should allow the participant to categorize the objects with greater accuracy (e.g., less errors) when presented with the same category structure and type in the future. More specifically, participants learn to categorize the set of objects based on the relevant features of that category. For example, a category may vary on one dimension – let’s say color – and therefore an individual need only categorize the objects according to that dimension. That is, if the participant learns that objects of shape “circle” belong in category A and objects of shape “triangle” belong in category B, only the dimension of object “shape” is necessary for categorization.

This method of category presentation involving corrective feedback has been studied extensively in order to validate mathematical models (e.g., Nosofsky, 1984; Vigo, 2009, 2011, 2013, 2014), theories of concept learning behavior, such as prototype theory and exemplar theory (Medin & Schaffer, 1978; Nosofsky, 1984; Rosch & Mervis, 1975), and to present experimental replications of the study to verify results (Nosofsky, et al., 1994). Vigo referred to this as a serio-informative task (Vigo, 2013, 2014).

The 3[4] category has also been tested in a procedure that differs from the serio-informative task and – as a result – demonstrates the same difficulty ordering but with different proportions of errors on each category type (Feldman, 2000; Vigo, 2009, 2013). In this modification of the concept-learning task, participants can view the category or

categories of interest for some pre-defined duration. That is, the participant has access to perceive the category – consisting of the positive instances or both the positive and negative instances partitioned into the respective categories on-screen– prior to making responses to individual objects regarding their category of origin. Vigo referred to whole category presentation as the para-informative task (Vigo, 2013). Responses during the presentation of single objects in the para-informative task are identical to that of the serio-informative task; if the object belonged to the positive category, participants make a response to categorize that object as belonging to the positive category. Once made, the participant then views the next object in the category; the participants receives no feedback following each response. Thus, the serio-informative and para-informative tasks differ both in the presence of feedback and in the presentation of the category. Recent articles have examined how this method of presentation affects category learnability (Feldman, 2000; Vigo, 2009, 2013). As indicated previously, they have both found – despite some procedural inconsistencies in the Feldman experiment – that the learnability of the six 3[4] categories does decrease in terms of proportion of errors, but the overall difficulty ordering is preserved. Vigo found similar results in additional experiments (2009, 2013).

One of the fundamental goals of examining the learnability of category membership in a supervised learning experiment pertains to the ability to predict human (and nonhuman animal) performance given a particular category (Feldman, 2000; Goodwin & Johnson-Laird, 2011; Kruschke, 1992; Love, Medin, & Gurekis, 2004; Medin & Schaffer, 1978; Nosofsky, 1984; Vigo, 2009). If examining supervised learning

involving a particular category structure - for example the linearly separable 3[4] Type I category (e.g., groups easily separated by using only one dimension) - researchers can anticipate the individual's performance on that category relative to the specific presented task. It would further be anticipated that an individual presented with this category structure using a serio-informative task (single item presentation with corrective feedback) would demonstrate slightly different performance than an individual presented with the same category on a para-informative task (whole category presentation with no feedback) (Vigo, 2009; Vigo, 2012). However, because the outcomes (the associated labels) are known prior to the experiment, supervised learning can explicitly aim to predict and determine changes in human performance behavior between the specific conditions associated with the serio-informative and para-informative tasks (Pothos & Chater, 2002; Vigo, 2009, 2011).

### **Unsupervised Learning**

Unsupervised concept learning, in contrast to supervised concept learning, is used to examine the conditions under which an individual categorizes a group of stimuli given variations in the stimulus dimensions (Goudbeek, Swingley, & Smits, 2009; Pothos & Chater, 2002). In the machine learning literature, an algorithm is presented with a collection of stimuli or objects that it must then spontaneously sort into groups (Bengio, Courville, & Vincent, 2012; Ghahramani, 2004). Generally, a successful algorithm would ideally sort these objects in such a way that groups have a high degree of within group similarity while maintaining a low degree of similarity between different groups (although the exact mechanism would depend on the type of implemented algorithm). In

other words, the algorithm is assessing the underlying structure of the object collection in order to facilitate a partition or grouping. When presented with the 3[4]-1 objects that have thus far been the ever-present example in this dissertation, the algorithm should group similar shapes together. In doing so, each of the two groups - triangular and circular - would have a high degree of within-group similarity while, when comparing groups against each other, would demonstrate a low degree of similarity. Using categories that are not divided as easily as 3[4]-1 may result in more unique groupings in order to satisfy the within-group and between-group conditions (Fleiss and Zubin, 1969; Pothos & Chater, 2002, 2005; Zubin, 1938, etc...).

Research on human participants is similar (Colreavy & Lewandowsky, 2008; Pothos & Chater, 2002; Love, 2002). Individuals observe a collection of stimuli, and categorize the stimuli by assigning a category label to each one of the stimuli. Much like the algorithms described above, an individual should ideally be making these category assessments and groupings by evaluating the degree of similarity or coherence between the presented stimuli (Pothos & Chater, 2002). Unlike the serio-informative and para-informative (Shepard, Hovland, & Jenkins; Vigo, 2009, 2013) supervised learning tasks, unsupervised learning provides no corrective feedback or category cues. Instead the participant spontaneously constructs groupings of objects in order to increase within-group similarities and between-group differences. Once participants group all objects in such a way that they feel the groupings are subjectively coherent, the participant begins again with a new collection of objects. In experiments with visual objects, participants group similar objects by drawing lines to separate dissimilar objects or circling around



similar objects (Compton & Logan, 1993, 1999; Pothos & Chater, 2002). Figure 2 presents a simple example of this process; the participant would separate the objects by using lines or circles in order to establish groups of similar objects.

Ashby, Queller, and Berretty (1999) distinguished between two types of unsupervised learning tasks, what they referred to as the “unsupervised learning task”, in which individuals are not provided corrective feedback but are told the number of categories, and the “free sorting task” in which individuals are provided neither the corrective feedback nor the category number. Henceforth, we will discuss the free sorting task and refer to it as unsupervised learning since it is primarily used in unsupervised learning research (Edwards, Perlman, & Reed, 2012; Goudbeek, Swingley, & Smits, 2009; Pothos & Chater, 2001; Pothos & Chater, 2002).

In addition, the purpose of unsupervised learning research differs from supervised learning in that, rather than attempt to predict how individuals classify stimuli, unsupervised learning allows for the examination of the conditions under which an individual will classify stimuli a particular way (Pothos & Chater, 2002; Ashby, Queller, & Berretty, 1999). In other words, in what manner do experimental manipulations - whether through prompting or instructions, or through the choice of stimulus modality and dimensions – affect the free and spontaneous categorization of stimuli.

Over a series of studies, Pothos and Chater (2002) examined just this question; participants categorized visual objects according to different criteria across several experiments. In their first experiment, participants viewed a rectangular space containing markers placed in a coordinate plane, similar to that presented in Figure 2. Because

location was the factor that influenced categorization, each object varied according to two dimensions; the X-coordinate and the Y-coordinate. Participants categorized stimulus sets of increasing response variability. That is to say, early tasks in the experiment had a clear solution for grouping despite not being explicitly stated (e.g., two large groups clustered on the opposite ends of the coordinate space). When the task required partitioning the space into three clusters - response variability increased as compared to the two-cluster condition. Specifically, participants developed more “distinct solutions”, as Pothos and Chater (2002) refer to it, to the problem of dividing the space into three clusters. When comparing this result against the two-cluster condition, it appears that a uni-dimensional solution (categorizing solely by the X or Y coordinate in this case) will result in less response variability.

As mentioned, when reassigning dimensions used to manipulate stimuli to different attributes without a change in the inherent categorization requirements, responses to the unsupervised learning task change. Pothos and Chater (2002) reassigned the previously described dimensional variation from an X/Y coordinate system to the inner and outer diameter of a star. Specifically, changes along the Y axis in the previous experiment corresponded to the size of the inner star diameter, whereas changes in the X axis corresponded to the size of the outer star diameter (Figure 3). The dimensional reassignment resulted in more response variability or distinct solutions than the stimuli presented on the X/Y coordinate system despite the same categorization requirements being present (e.g., separating two large clusters). This refers back to the central goal of unsupervised learning; how do the conditions, context or stimuli affect the categorization

process? In the Pothos and Chater (2002) experiments, a simple dimensional reassignment fundamentally changes the unsupervised learning process by increasing categorization response variability to the prescribed conditions.

The above studies by Pothos and Chater (2002) provided an example of the delicate relationship the categorization process maintains to the available stimuli, and how changes in one subsequently affect the other. While these and other studies (Ashby, Queller, & Berretty, 1999; Love, Medin, & Gureckis, 2004) have focused exclusively on unsupervised learning of visual category stimuli, only a few have examined how individuals perform when given an unsupervised learning task involving auditory stimuli (Goudbeek, et al., 2009; Gygi, et al., 2007). Thus to further expand the knowledge of how conditions and stimulus modality affects categorization behavior, in the current experiment I will examine the behavioral response of individuals in an unsupervised learning task involving auditory concepts varying across multiple dimensions. The current experiment will expand on the usage of audio stimuli from previous research will be discussed in detail in Chapter 7.

## Chapter 5: Auditory Stimuli

Prior to discussing auditory categorization, it may be useful to provide a brief primer on auditory stimulus dimensions. Auditory stimuli vary according to three broad dimensions: frequency, amplitude, and timbre (Melara & Marks, 1990; Roads, 1996, 2002). However, these dimensions decompose further into sub-features, particularly timbre which when analyzed in the frequency domain is a composite of many individual features (Grey, 1977).

### Auditory Dimensions

Auditory stimuli contain a robust amount of information as evidenced by the number of dimensions in which they can vary (Bonebright, 2001; Bulgarella & Archer, 1962; Clarkson & Pentland, 1999; Gao, Lee, & Zhu, 2004; Goudbeek, Swingley, & Smits, 2009; Guastavino & Katz, 2004). The attributes used to define the auditory stimuli in the current experiments – namely, pitch, amplitude, and timbre - will be briefly described and discussed according to their definition and, where applicable, their perceptual definitions. Researchers have found the following features of auditory stimuli to be significant or necessary to the human perception of sound (Caclin, McAdams, Smith, & Winsberg, 2005; Grey, 1977; Lockhead & Byrd, 1981).

**Frequency and pitch.** Frequency describes the behavior of periodic vibrations - or waves - that may be perceivable by humans according to the auditory modality (Farnell, 2010; Roads, 1996; 2004) and described in units of hertz (Hz). The range of frequencies extends from 0 Hz to extreme ultra-frequencies, although the range of human auditory perception is limited to only frequencies between 20 to 20000 Hz. These values

represent a human with no auditory deficits (e.g., tinnitus) at a younger age. As individuals age, the upper bound of human hearing (20,000 Hz) tends to decline progressively (Robinson & Sutton, 1979).

Auditory waves behave in cycles – the rise and fall of the wave from the resting point (the horizontal line in Fig. 4). The amount of time it takes for a single cycle to complete is the period,  $T$ . Frequency can then be measured as an inverse relation of the period:

$$f = 1/T$$

Hertz (Hz) are the standard unit for the resulting frequency ( $f$ ). To provide an example, a wave with a frequency of 150 Hz would have a periodicity of  $1/150$ , or occurring every 0.006 seconds. Thus, according to this formula, cycles with shorter periods result in higher frequencies (see Figure 5 for a comparison of two frequencies/waves).

Frequency in more subjective and psychophysical terms is a sound's pitch. As such, this allows for individuals to articulate relationships between frequencies. A 12 note scale describes the relationship of frequency according to octaves. An octave provides a standard unit of measurement that describes a sound as twelve notes higher than the root pitch. An octave higher than that (+24) will be similar to the root pitch, and so on. Thus, pitch describes a highly linear relationship; increasing a root pitch by a certain number of notes will have the same relationship between the notes regardless of the originating octave (e.g., playing seven notes up from C3 results in the same note as playing seven notes up from C4). In terms of frequency, however, the relationship is

non-linear. The change in frequency between notes C2 and C3 is less than the change in frequency between notes C3 and C4.

Pitch is a highly identifiable aspect of musical and auditory content under different conditions. In individuals with normal hearing, they are able to assess a just noticeable difference between two complex tones differing by as little as 1 Hz (Kollmeier, Brand, & Meyer, 2008). Therefore, individuals assess two tones differing by only 1 Hz as being different along a dimension of pitch. Differences in behavioral history also affect pitch perception. Tervaniemi, et al., (2005) found that the behavioral history of the individual contributes to pitch identification and discrimination; the most striking and fundamental difference in this case is whether the individual has a history of playing and performing music or not. Thus, performance on an auditory measure can be highly dependent on the individuals' history. Musicians are more likely to discriminate and accurately respond to a change in pitch than non-musicians. The ability to determine these relations, described as relative pitch (RP) perception, is a behavioral characteristic of trained musicians in that they are able to identify pitch when given a contextual cue, such as an additional, named note like 'B' (Levitin & Rogers, 2005). Thus, because of their musical training, identifying previous musical training present in an individual's behavioral history is critical for avoiding any potentially confounded or biased results.

Differences in the timbre content also demonstrate an effect on an individual's ability to accurately report perceived pitch. Lockhead and Byrd (1981) reported that when a sound contained timbre, or overtone information in the form of a piano tone, participants identified absolute pitch with 99% accuracy. In contrast, when researchers

removed overtone information by using a pure sine tone – where no partial frequencies are present other than the fundamental frequency - absolute pitch identification dropped to 58% accuracy. Balzano (1986) found a somewhat conflicting result; while investigating absolute pitch identification of pure sine tones, he found a median accuracy of 84.3% identification. Despite this apparent contradiction, these results establish the interdependent relationship between pitch and timbre, (an additional discussion focusing on this interdependent relationship will continue in the timbre section of this paper.) Moore Glasberg, and Peters (1984) examined the contribution and dominance of each harmonic partial, or overtone, in an individuals' assessment of a complex tones' perceived pitch. Moore and colleagues (1984) found that the distribution of the dominant harmonic partials in a particular complex tone played at a particular pitch, varied between tones. However, when researchers detuned the dominant harmonic partials greater than +/- 2 to 3%, there was an overall change in judgment of the tone's pitch content.

One of the primary measures of pitch discrimination and accuracy is pitch matching. In such paradigms, individuals may be asked to do one of the following: vary a tone's pitch until it matches, or approximates, the pitch of the original tone (Plack & Oxenham, 2005), match a pitch by entering the note and octave value (Balzano, 1986), or determine the relative increases or decreases in pitch by making corresponding matches (Zatorre, Evans, & Meyer, 1994). By making individuals perform these comparisons, researchers are able to gauge the effect other auditory dimensions may have on the perception of pitch. It follows that individuals will tend to match or categorize similar pitches into groups. Kohler (1987) found that changes in the fundamental frequency

appeared to result in different classifications of the tones. More specifically, when increasing or decreasing the fundamental frequency, individuals classify manipulated tones as being different from each other. Thus, based on this result, I anticipate that changes in pitch across the stimuli will affect the resulting grouping structure in an unsupervised classification task.

The exact nature of pitch perception is continually under investigation, but it does appear that some group differences require documentation. As previously mentioned, musicians tend to perceive pitch more accurately than non-musicians (Tervaniemi, et al., 2005). However, there have been some disparate results with regard to age. Most notably, a child's ability to discriminate and perceive pitch accurately has seen some conflicting results (Bundy, Colombo, & Singer, 1982; Clarkson & Clifton, 1985; Speer & Meeks, 1985), but such results may be an effect of the natural contingencies in their developmental history, such as having a parent with musical training (Levitin & Roger, 2005). Thus, the connecting thread behind this body of research is musical training; either having the requisite training or growing up in an environment with strong musical ties appears to affect pitch detection and categorization behavior. Therefore, in assessing categorization behavior of individuals using the auditory dimension of pitch, participant screening should take place based on musical training or at least documenting the extent of training that they possess. Because the focus of the experiment does not relate to musical training, I will document and screen participants as the situation warrants.



**Amplitude and loudness.** The dimension of amplitude refers to the pressure a wave propagated through a medium, such as air. Amplitude has both positive and negative components associated with sound. Each are relative to the resting point of the sound which generates no waves and is the absence of sound. When generating a positive increase in the amplitude, the medium through which it travels is compressed. Negative decreases in amplitude result in rarefactions of the medium. A change in the compression and rarefaction through the medium corresponds with a change in amplitude; increases result in higher amplitude, decreases result in lower amplitude. The absolute amplitude describes the difference between the peak displacement and the zero (resting) point. In terms of computational descriptions of amplitude, this value varies between zero (no amplitude) and one (full amplitude). Changes in amplitude can further be described in units of decibels (dB), the logarithmic unit of the relative intensity of two auditory stimuli, or the RMS (root mean squared) amplitude. Of these, dB amplitude has been reported frequently as an analysis measure of amplitude (Glasberg & Moore, 2002; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Ohl, Scheich, & Freeman, 2001; Svirsky, 2000; Wong, Skoe, Russo, Dees, & Kraus, 2007).

Loudness – the psychophysical description of loudness - is the perceived, subjective determination of a stimulus' intensity relative to other factors, such as frequency or, subjectively pitch. To measure the subjective loudness of a sound, Stevens and Volkman (1940; Stevens, 1960, 1970) created the unit of sones. A sone is a relative ratio measurement, where one sone is equal to a 1 kHz sine wave at 40 dB. An increase in volume to 50 dB, a change of 10 dB, the wave in sones would be twice as loud. A

further increase – such as an increase to 60 dB – would result in the wave four times as the original 40 dB wave. Despite being the psychophysical analog of amplitude, current research infrequently uses sones as a measure of loudness. More typically, dB or RMS amplitude is used as the measurement of loudness in perceptual experiments because these units can easily be converted (e.g., Jesteadt, Luce, & Green, 1977; Khalfa, et al., 2004; Neuhoff, McBeath, & Wanzie, 1999; Puckette, 2006).

Generally, individuals are able to establish a just noticeable difference between the loudness of two sounds by only 1 dB (Howard & Angus, 2012). However, the changes in loudness can also affect the perception of other auditory dimensions. The apparent loudness of a stimulus is dependent upon frequency changes of the auditory stimulus. Neuhoff, McBeath, and Wanzie (1999) examined square waves varying in frequency across time. The frequency increased, decreased, or remained constant across the duration of the auditory stimulus. Additionally, each sound also varied in loudness under several conditions; loudness increased, decreased, or remained constant. Neuhoff et al. found that increases in the frequency across time results in the perception that the perceived loudness is also increasing. Similarly, decreases in the frequency across time resulted in a lower perceived loudness of the square wave. Other researchers have experienced similar results indicating that loudness and frequency are interrelated dimensions that affect one another (Canevet & Scharf, 1990).

Subjective, perceptual determinations of loudness differ across groups of individuals. For example, Khalfa and colleagues (2004) found that, relative to developmentally normal children, children with autism - in addition to having a smaller

auditory dynamic range - tend to have increased perception of auditory stimuli loudness. In other words, children with autism, perceive sounds as louder compared to loudness perception of developmentally normal children. Other groups of individuals perceive loudness differently. Similar to the subjective perceptual differences of pitch found between musicians and non- musicians, the subjective loudness of an auditory stimulus has also been found to differ depending on whether an individual plays or has played a musical instrument (Hoover & Cullari, 1992). In contrast, some groups of individuals, such as the blind, do not differ from individuals with normal vision (Yates, Johnson, & Starz, 1972). As with pitch, due to variation in loudness perception between groups, it is necessary to take note of the relevant physiological or behavioral history of the individuals in the experiment.

Perceived loudness of an auditory stimulus is highly dependent upon other features or dimensions of sound such as frequency (Neuhoff, McBeath, & Wanzie, 1999), and also highly dependent upon previous behavioral history and physical or psychological afflictions (Hoover & Cullari, 1992; Khalfa, et al., 2004; Yates, Johnson, & Starz, 1972). Additionally, the manner in which amplitude varies across time, known as the amplitude envelope, also has a profound influence on both pitch perception and perception of timbre (Berger, 1964; Houtsma, 1997). The contour of the amplitude envelope determines the duration of a sound; thus, the shape and attributes of the amplitude envelope determine the duration of an auditory stimulus. The amplitude envelope is traditionally – in the field of both computer music and synthesis (Roads, 1996) – described according to four components: attack, decay, sustain, and release.

**Amplitude envelope.** The amplitude envelope is a description of the flow of amplitude across time. For synthetic and physically modeled instruments, these are the parameters that govern the amplitude of the instrument across time, often referred to as the voltage controlled amplifier. For computer or synthesizer based instruments, the amplitude envelope is typically defined by the attack, decay, sustain, and release components (in short referred to as the ADSR; Figure 6), although variations of differing complexity such as ADR (attack, decay, release), AR (attack, release), and ADBDR (attack, decay, breakpoint, decay, release) exist.

**Attack.** The attack of the ADSR is the duration of time that an instrument requires to go from zero to maximum amplitude. This corresponds to the onset of a sound; sounds with a longer onset sonically appear to swell at a gradual rate while sounds with shorter onsets appear more percussive. Sounds with a very short attack,  $\sim 10$  ms, produce an audible click, although physical instruments rarely exhibit an attack of such short duration. Many instruments are physically capable of producing a continuum of attack times. For example, striking a violin string suddenly produces a very short attack – e.g. short rise time – with the result being that the maximum amplitude occurs in a very short period. In contrast, lightly dragging the bow across the strings and gradually increasing the pressure and intensity of the bow to the string creates a slow attack – or long rise time. The differing rates of attack also influence the perception of the instrument's timbre (Caclin, et al., 2005; Grey, 1997). Specifically, the same instrument played with different attack speeds generates introduces the listener to the sound's overtones at different rates. Thus, while the attack is a function of varying the amplitude

envelope over time, it also results in effects on the perceived timbre. I will address timbre more thoroughly in a later section.

***Decay.*** Auditory decay occurs immediately after the attack, and is the period where the force initially applied during the attack (e.g., blowing on a horn or striking the strings with a bow) decreases from the friction present on the instrument. As a result, the energy contributed during the attack portion exits the system and there is a subsequent decrease in amplitude. After dispersing excess energy from the initial attack, the amplitude decreases until it reaches an equilibrium referred to as sustain.

***Sustain.*** Sustain is an equilibrium state of energy input and output within the instrument. A constant flow of energy input is equal to a constant flow of energy output. To give a more specific example, the energy input by bowing a string on an instrument is equal to the output. Energy input continues into the system; the bow is still moving across the strings. Sustain continues until energy is no longer supplied to the system. Once energy input ceases, the release phase begins.

***Release.*** The dispersal of residual energy after no further energy input into the system results in the release. During the release, the instrument continues to produce some sound over time, but the amplitude of the tone decreases. The decrease in amplitude is consistent with the duration of the dispersal of energy. In other words, the duration from the moment equilibrium is disrupted (sustain) until the amplitude returns to a stable resting state, defines the release. As a practical example, once the bow stops making contact with the string the amount of time the instrument takes to go silent refers to the release period. The shape of the waveform's amplitude through time, as controlled

by the amplitude envelope, plays an important role in the both the perceived pitch of an instrument or synthesized tone. Changes in the amplitude envelope can have a dramatic effect on the perception of pitch. Hartmann (1978) found that when individuals are presented with tones, those tones that have an exponential decay of the amplitude envelope rather than a gated envelope are estimated as being higher in pitch (Hartmann, 1978). Neuhoff and colleagues (1999) found similar results; not only do changes in loudness affect pitch perception, but also changes in the amplitude envelope affect the release stage (Hartmann, 1978). The amplitude envelope also plays a critical role in the perception of speech, including perception of speech by both children and adults with dyslexia (Goswami, Gerson, & Astruc, 2010; Goswami, Thomson, Richardson, Stainthorp, Hughes, Rosen, & Scott, 2002; Corriveau, Pasquini, & Goswami, 2007), the ability for individuals to speech-read or lip-read (Grant, Ardell, Kuhl, & Sparks, 1985; Grant, Braid, & Renn, 1991; Grant, Braid, & Renn, 1994), and recognition of languages such as Chinese (Fu, Zeng, Shannon, & Soli, 1998; Luo & Fu, 2004). In fact, the amplitude envelope plays a very fundamental role in language perception and comprehension; recognition of elements such as consonant-vowel ratio (Freyman, Nerbonne, & Cote, 1991) and the perception of consonants and vowels (Shinn & Blumstein, 1984) are both highly dependent upon the shape of the amplitude envelope. As such, the amplitude envelope represents a critical feature in both perception of speech and musical tone.

**Timbre.** Acoustically, a precise and unanimous technical description of timbre has remained elusive in auditory research due to its fundamental nature. It is also difficult to define and specify the exact physical dimensions as well as other contributing dimensions to the overall timbre (Pitt, 1994). The longstanding definition, as put forth by the American National Standards Institute defines timbre as:

The attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. (American Standards Association, 1960)

In other words, timbre is the defining characteristic that would result in an individual's ability to determine: a) that an individual perceives the sound resulting from a piano played at several different octaves is the same instrument, and b) the tone from a piano and an oboe are from different instruments when played at the same pitch and same loudness. While such a definition may be adequate from a perceptual standpoint, many researchers have found this definition to be imprecise and incomplete from an acoustic standpoint (McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995; Pitt, 1994; Smalley, 1994) and, as a result, have made attempts at determining the relevant – and minimum – number of dimensions that contribute to the perception of an instrument's sound. From a more theoretical perspective, many have debated how timbre arises and how we are able to perceive the same timbre at two different pitches or two degrees of loudness as being the same or similar.

Often, researcher can extract the features of timbre via a Fourier transform – typically Fast Fourier Transform (or FFT) – which converts the audio signal from the

time domain to the frequency domain so that spectral analysis can be conducted on the features (Grey, 1977; Roads, 2004). The time domain consists of an audio signal as represented continuously through time (Fig. 7), wherein the shape of the amplitude envelope is easily discernable. Although information about a particular pitch is available in the time domain based on the analysis of single wave cycles at a microsound level, information concerning the frequency spectrum is unavailable (Roads, 2004). The frequency domain allows for analysis of the entire frequency spectrum of a particular sound in one instant, rather than over a lengthy period of time. The duration of the instant depends on the window size of the FFT; window sizes are based on the number of samples present within the window (e.g., 1024 is equivalent to two cycles of a waveform consisting of 512 samples). Researchers can perform any number of analyses once the audio is in the frequency domain, including analysis of the harmonic spectrum, power spectrum, spectral centroid, and others (Grey, 1977; Roads, 2004). FFT has been the standard method of analyzing the timbre of a sound.

However, there are some criticisms of the FFT process and inferring perceptual relationships based solely on the analysis of FFT. Balzano (1986) argued that because there exist variations in timbre, not only due to changes in pitch and loudness, but also due to natural variation between instruments, such a process will not necessarily result in an auditory stimulus that approximates the timbre of the original instrument at the recorded pitch and loudness. His solution was to examine the timbre for possible cues, or invariances, that provide information about the underlying structure in relation to the signal dynamics (Balzano, 1986). In this sense, spectral analysis is analogous to



descriptive statistics; descriptive information about a sound or wave results, but the relationships and underlying dynamics that connect the attributes of timbre perception are not directly considered.

Despite issues concerning the apparent superficiality of the FFT process, spectral analysis has served an important purpose in the analysis of timbre. Previous research has focused on the dimensional reduction of timbre such that only the most salient or diagnostic dimensions are preserved. Much of the research has found that timbre couldn't be isolated to a single, unifying dimension; it appears that timbre is composed of multi-dimensional physical features of sound (Grey, 1977; Houtsma, 1997; de Bruijn, 1978). Of these features, it appears that the spectral envelope represents an important dimension of timbre (de Bruijn, 1978; Ter Keurs, Festen, & Plomp, 1992; Warren, Jennings, Griffiths, 2005). The spectral envelope is the frequency-domain (FFT) representation of an auditory stimulus given a particular frame of time; the shape of the spectral envelope describes differences in amplitude intensity at each frequency. Figure 8 provides an example of two differing spectral envelopes; as the amplitude energy at each frequency changes, so does the perceivable timbre. When more overtones – or harmonics - are present at higher frequencies, individuals tend to label such sounds as being “brighter”, specifically in regards to what some describe as “tone” color (von Bismarck, 1974; Clarkson, Clifton, & Perris, 1988; Krumhansl & Iverson, 1992). Such subjective labels are characteristic of timbre research. A more precise measure of differences in auditory spectra may be the spectral centroid (McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995; Samson, Zatorre, & Ramsay, 1997). The spectral

centroid is the center, or average, of the spectrum; thus, sounds with higher averages (through frequency distributions weight more highly or more positively skewed towards higher frequencies) typically contain more high frequency overtones and result in descriptions such as “bright” (Grey & Gordon, 1978).

In addition to the spectral centroid, researchers have identified other dimensions as contributing to the characteristics that distinguish different auditory timbres. McAdams and colleagues (1995) investigated the number of features that distinguish timbre by using a latent class model. The result demonstrated three significant correlations associated between timbre dimensions and Krimphoff’s (1993) proposed acoustic correlates. The three dimensions are: log-attack time, spectral centroid, and spectral flux. Log-attack time represents the logarithm of duration of time from the onset of a sound (set via a threshold of 2% of the maximum amplitude) until the maximum amplitude is reached. Spectral centroid, as mentioned, is the average overtone or harmonic content of an audio source and spectral flux refers to the variance associated with the spectral envelope over time. Krimphoff (1993) measured spectral flux as the average of the correlations between the amplitude spectra in subsequent time windows. As further described by McAdams, the correlation would be highest when variation across each window is minimized. Thus, spectral flux in timbre is low when variability between windows is also low.

However, from a practical, psychological standpoint, Caclin and colleagues (2005) found that spectral flux demonstrated little effect on an individual’s assessment of dissimilarities as compared to other influencing dimensions such as log-attack time. In a

sequence of studies, Caclin and colleagues (2005) found that participants were generally unable to distinguish variations in spectral flux when it used as a varying dimension of timbre. Only when the spectral variability was at its highest did individuals use spectral flux to assess dissimilarity, and even given these conditions spectral flux acted as a supplement to other dimensions of timbre. Caclin and colleagues found that log-attack time and spectral centroid were the most salient attributes of timbre along with the spectral structure of the sound.

The result from Caclin's studies (2005) that three dimensions for timbre recognition is consistent with previous research (Krumhansl, 1989; McAdams & Cunibile, 1992; Miller & Carterette, 1975). Given this research, perhaps three features of timbre – specifically spectral centroid, spectral structure, and log-attack time – will provide a more objective starting point for assessment of dimensional importance in determining differences and similarities in auditory stimuli, via multi-dimensional scaling as possible subjective “definitions” of timbre by the participants.

Regardless of the varied definitions of timbre, the identification and discrimination of timbre has played a key role in auditory perception. Of particular interest has been timbre discrimination in infants and young children with the aim of answering key questions; namely, “Is timbre discrimination something that is learned?” and, “At what age can an individual effectively discriminate between different timbres?” (Clarkson, Clifton, & Perris, 1988; Trainor, Wu, & Tsang, 2004; Trehub, Endman, & Thorpe, 1990). In favor of the argument that timbre discrimination is something that is learned, musical training has been shown to influence timbre identification - similar to

the research of musician pitch identification – as the specific neural correlates of timbre perception between musicians and non-musicians has been studied (Chartrand & Belin, 2006; Crummer, Walton, Wayman, Hantz, & Frisina, 1994; Prior & Troup, 1988; Vurma, Raju, & Kuuda, 2010). Generally, musical training increases an individual's sensitivity to timbre changes when compared to non-musicians (McAdams, Winsberg, Donnadiu, De Soete, & Krimphoff, 1995). Therefore, as with pitch, due to known differences in performance between musicians and non-musicians, it will be necessary to document such behavioral history prior to the experiment.

## Chapter 6: Integral Dimensions

Not only can the dimensional variation of stimuli affect learning, but the degree to which there is interplay between those dimensions can also have effects. So far, the majority of the discussed studies have used separable dimensions – where an individual can isolate the contribution of each stimulus dimension – but some dimensions and stimuli display an interconnectedness that doesn't promote isolated processing. When dimensions behave as such, they possess integral dimensions. Learning processes differ between separable and integral dimensions (Garner, 1974; Garner & Felfoldy, 1970; Shepard, 1991).

When a stimulus' dimensions are separable, an individual can easily distinguish between the values or intensity of the dimensions (Garner, 1974; Pothos & Chater, 2002). For example, a blue colored circle possess two identifiable and distinct features. More formally, the separation between dimensions - such as dimension of color (“blue”) and the dimension of shape (“circle”) – refers to orthogonality. The example stimuli from Pothos and Chater (2002) in Figure 2 present a concrete example of separable dimensions. An individual can easily distinguish that there are two attributes that define each stimulus in the rectangular space; the X coordinate and the Y-coordinate. Nosofsky and colleagues (2013) used the apt term “highly analyzable” to describe such separable dimensions. Indeed, an individual can readily decompose a separable-dimension stimulus into the constituent components, in the case the X-coordinate and the Y-coordinate.

Little and colleagues (2013) found evidence that individuals perform serial or parallel processing when making categorization decisions of stimuli based on their separable dimensions. Based on a series of isolated dimensional assessments (e.g., taking into account only one dimension at a time), individuals successively determined how to categorize a given stimulus (Little, Nosofsky, & Denton, 2011; Little, et al., 2013). Some categorization errors may occur during this process; an individual examining an object with separable dimensions may encounter interference in stimulus categorization based on the attended orthogonal dimensions. Each of the independent dimensions may compete for attention resources and, as a result, influence the response made by the individual (Garner & Felfoldy, 1970). Conversely, when an individual categorizes separable-dimension stimuli, they may be able to filter out stimulus features that are not indicative of a category's structure. The reduction/filtering of attended stimulus dimensions facilitates the categorization process and, in such situations, an individual may categorize based on a single dimension (uni-dimensionally) and disregard - or filter out - all other dimensions (Gottwald & Garner, 1975).

Much of the traditional categorization literature has examined separable dimensions, including the seminal study by Shepard, Hovland, and Jenkins (1961). Current research in categorization and concept learning has also focused on categorization using separable dimensions, such as the amoebas used by Feldman (2000) and the flasks used by Vigo (2009, 2011, 2013; Vigo & Basawaraj, 2013; Vigo, Zeigler, & Halsey, 2013). Furthermore, researchers examined separable dimensions in eye-tracking experiments on categorization. Rehder and Hoffman (2005), much like Shepard

and colleagues (1961), conducted research in which they separated each dimension spatially during presentation to the participants, allowing stimuli or objects composed of separable dimensions to be examined while filtering out other features of the stimuli. Overall, consensus across studies indicates that individuals process separable-dimension stimuli by an individual through serial or parallel processing (Garner & Felfoldy, 1970; Little, Nosofsky, & Denton, 2011; Little et al., 2013).

In contrast to separable dimensions, integral dimensions lack the orthogonality of dimensions. When a stimulus' dimensions are integral, each feature is inseparable from the higher-order stimulus to which it contributes (Garner & Felfoldy, 1970; Nelson, 1993). Early research by Garner and Felfoldy (1970) found examples of this inseparability in several modalities. In vision, dimensions such as color, shape, and size are easily separable and psychologically distinct; color is independent of shape and individuals may attend to color accordingly. However, visual dimensions such as hue, color and brightness of a stimulus are integral (Garner & Felfoldy, 1970). An individual presented with such a stimulus would be unable to psychologically separate these dimensions and would, accordingly, process the whole stimulus rather than individual dimensions (Cheng & Pachella, 1984; Nosofsky, et al., 2013; Nosofsky & Palmeri, 1996; Patchella, Somers, & Hardzinski, 1981; Shepard, 1991).

Because of the inherent relationship between the stimuli dimensions, individuals are able to efficiently perform categorize integral-dimension stimuli via speed-sorting. In fact, this relationship between the stimulus features may under certain circumstances represent a redundancy (Garner & Felfoldy, 1970). While separate dimensional

processing is not possible, holistic relationships between dimensions facilitate an increase in discrimination and categorization response time by allowing the individual to selectively filter their attention from dimensional commonalities between stimuli (Garner & Felfoldy, 1970).

Research indicates that while separable-dimensional stimuli are processed serial or in parallel, integral-dimension stimuli are examined according to a coactive process (Little, Nosofsky, & Denton, 2011; Little, et al., 2013). That is, when categorizing integral-dimension stimuli, assessments of each dimension are pooled together to form a categorization decision. Compared to separable-dimension stimuli where each dimensional is assessed individually, a holistic processing of the collective dimensions informs the categorization process (Nosofsky, et al., 2013).

Auditory stimuli are dimensionally integral; each dimension has an effect upon every other auditory dimension (Melara & Marks, 1990). Because of the integrality of the dimensions, individuals often perceive the auditory stimulus as a whole rather than based on individual features (Nelson, 1993). Pitch and loudness are integral dimensions and, in addition, the dimensional value of each affects the other dimensions (Melara & Marks, 1990; Nelson, 1993; Pitt, 1994). Research by Grau and Nelson (1988) has supported this interpretation; because of the integrality of dimensions, the interference in a classification task is redundant between the dimensions – in this case, pitch and loudness. In other words, Grau and Nelson found that individuals perceive pitch and as a single unit rather than separate features. This is an intuitive result; it has been found that when pitch and loudness are correlated (i.e., high pitch matched with high loudness and



low pitch matched with low loudness), fewer errors are made in identification tasks (Neuhoff, McBeath, & Wanzie, 1999). That is, when auditory dimensions are both congruent, the audio tone is more easily perceived as a single, integral unit. Grau and Nelson also examined the integrality of pitch and timbre, and they found greater interference, or errors, in the tested task (Grau & Nelson, 1988). Thus, while pitch and loudness are integral dimensions, pitch and timbre may be somewhat more separable; specifically, the pitch of a note and the type of instrument are distinguishable as somewhat different dimensions.

However, this is not to say that pitch and timbre are as separable as, say, a shape and a color. There exists a level of interaction between these dimensions even though it may be possible to distinguish between an instrument's sound and the pitch (Melara & Marks, 1990; Pitt, 1994). When playing an instrument – for example a French horn – timbre changes as a function of the pitch. FFT analysis of the horn playing at two different fundamental frequencies reveals the presence of different overtones, or harmonics, in the spectrum (Roads, 2004). When holding timbre constant in the form of both notes emanating from a French horn, the fundamental frequency still affects timbre. However, in the case of timbre and pitch, both these dimensions typically co-vary; this is particularly apparent in the analysis of the spectral centroid (an increase in pitch results in an increase in the spectral centroid; Melera & Marks, 1990).

The relationship between timbre and loudness also demonstrates an interacting effect. When timbre is positively correlated with loudness (e.g., loudness increases as timbre increases), the information supplied by both features appears to be redundant. In

other words, the dimensions interact such that individuals perceive the stimulus as a single unit rather than two separate dimensions. There are some inherent interacting effects between each of the auditory dimensions (Melara & Marks, 1990; Nelson, 1993; Pitt, 1994). As a result of this interaction between features, auditory stimuli are not as readily separable according to dimension. Even analysis of timbre, such as FFT, still provides a result that is dependent upon both the loudness and pitch of a particular stimulus. This is due in part to interaction and natural variability (Balzano, 1986).

In summary, auditory stimuli are less perceivable by their dimensions than some visual stimuli. For example, when imaging the pitch of a sound, it will always be associated with a particular loudness. Psychologically it would be difficult if not impossible to image pitch isolation from loudness (e.g., pitch without any loudness). A level of interconnectedness between dimensions will always exist; pitch can't exist without loudness, timbre can't exist without pitch (or loudness again, for that matter!). The effects of this integrality might manifest in the form of how individuals allocate auditory attention when perceiving and attending to auditory stimuli. Much like perceiving the dimensions of color – hue, saturation, and brightness – individuals' may lack the cognitive ability to pointed and specifically allocate attention to a specific dimension when dealing with integral dimensions (Little, Nosofsky, & Denton, 2011; Little, et al., 2013). Thus, I will assume equal and distributed attention across all dimensions in the current experiment.

## Chapter 7: Auditory Categorization

Audio categorization research, while not as extensively studied as visual categorization, exists for both supervised and unsupervised learning tasks (Goudbeek, Swingley, & Kluender, 2007; Goudbeek, Swingley, & Smits, 2009; Howard & Silverman, 1976; Vandermosten, Boets, Luts, Poelmans, Wouters, & Ghesquiere, 2011; Vigo, Barcus, Zhang, & Doan, 2013). In particular, unsupervised learning of audio categories has been extensively researched by computer scientists and engineers in an attempt to create an algorithm that accurately and efficiently categorizes audio signals (Blumensath, Davies, 2004; Goa, Lee, & Zhu, 2004; Park & Glass, 2008; Van Segbroeck, & Van hamme, 2009). However, fewer studies exist in which humans freely categorize auditory stimuli. Despite this, results have found significant categorization changes in unsupervised learning tasks when the type of stimulus and the auditory dimensions are experimentally manipulated (Goudbeek, Cutler, & Smits, 2007; Goudbeek Swingley, & Smits, 2009; Vallabha, McClelland, Pons, Werker, & Amono, 2007).

Multidimensional scaling (MDS) solutions for auditory stimulus experiments – used to examine the relevant number of auditory dimensions that participants perceive and attend to during a similarity or dissimilarity task - result in a consistent solution across studies. The overall trend of discovery within the literature demonstrates that individuals tend to assess similarity based on three dimensions of the given set of dimensions examined (Alrich, Hellier, & Edworthy, 2008; Bonebright, 2001; Howard & Silverman, 1976; McAdams, et al., 1995; Samson, Zatorre, & Ramsay, 1997). For example, Howard and Silverman (1976) examined the number and type of relevant

dimensions used by individuals when making similarity assessments of non-speech sounds. Researchers defined each non-speech sound by four dimensions: fundamental frequency, waveform shape, formant frequency, and number of formants. Howard and Silverman found that using multidimensional scaling for individual differences, three dimensions were determined to be the most relevant based on model  $R^2$  and dimensional interpretability. Of the dimensions examined, fundamental frequency and waveform shaped played a substantial role in similarity judgments with the third remaining dimensions being a combination of the two formant dimensions. Thus, only three dimensions in this early experiment on auditory dimension assessment were sufficient for describing participant behavior (Howard & Silverman, 1976).

In another example, Gygi, Kidd, and Watson (2007) also examined the number of auditory dimensions participants used when assessing similarity and what specific auditory dimensions or categorization strategies these dimensions correlated. That is, Gygi and colleagues were interested in the dimensional criteria individuals used for similarity. Stimuli for this experiment were natural sounds - such as airplanes flying, the sound of bowling, a person coughing, glass breaking, toilet flushing, falling ice, and more – rather than dimensionally defined stimuli. Multi-dimensional scaling on these stimuli demonstrated that individuals assessed similarity based on three – then undetermined – dimensions. Visually, three distinct clusters of audio formed. On the basis of the qualitative features shared by the sounds contained within these clusters, Gygi and colleagues defined these dimensions as: harmonic (sounds with some perceivable pitch or pitch change, such as instrument sounds, babies crying, etc...), discrete impact

(percussive sounds with a distinct termination, such as a gunshot, glass breaking, etc...), and continuous (sounds with a longer duration that change across time in some way, such as a toilet flushing, a person coughing, etc...). Thus, even with the wealth of auditory information present in naturalistic sounds, it appears that based on MDS participants are reducing the dimensional space down to three dimensions in order to assess similarity (Gygi, Kidd, & Watson, 2007).

Similarity assessment is widely believed to play an integral role in the formation of categories and largely results in the development of an individual's conceptualization of that category (Medin & Schaffer, 1978; Nosofsky, 1984; Pothos, et al., 2002; Rosch & Mervis, 1975). Nosofsky's generalized context model (1984) uses similarity assessment between dimensions to determine categorization predictions within a given set (context) of objects. The similarity assessment of these models, and previous notions of similarity (Tversky & Gati, 1978, 1982), essentially constitutes multiple pairwise comparisons between stimulus features to determine the presences or absence of shared features. Recent theories have traded these notions of pairwise comparisons for theories of underlying complexity and invariance of which by far the most successful is Vigo's theories of categorical invariance (Vigo, 2009, 2013, 2014)<sup>3</sup>. The success of this model implies that individuals perceive and encode the underlying category structure rather than engage in a seemingly combinatorial number of pairwise comparisons. Therefore, when individuals are engaging in similarity judgments, it appears likely that similarity

---

<sup>3</sup> Although, in his book, Vigo (2014) defines a particular principle of GIST referred to as invariance-similarity equivalence that translates invariance into a pairwise comparison process that involves both dimensional binding and partial similarity assessment.

responses are the result of an underlying pattern detection process rather than pairwise comparisons between stimulus dimensions.

Of course, the extent to which individuals are able to make these comparisons is dependent upon their ability to perceive and attend to the relevant stimulus dimensions. The saliency of particular stimulus dimensions will influence an individual's ability to attend to and use those dimensions for similarity assessment (Fritz, et al., 2007). For stimuli such as audio, it may be that the integrality of the dimensions impedes individuals from attending to specific dimensions without interference from additional dimensions (Melara & Marks, 1990; Nelson, 1993; Pitt, 1994). In auditory similarity assessment, individuals may be attending to the dimensions that are relatively easiest to extract from the stimuli and use those to form their similarity judgments. And given the consistency of MDS solutions in previous audio research to find that three dimensions sufficiently describe human performance, three dimensions may represent an approximate upper bound in the amount of auditory information an average individual can attend to and processes (Chen & Cowan, 2005; Cowan, Chen, & Rouder, 2004; Sauls & Cowan, 2009; Tulving & Patkau, 1962).

Researchers have then used such MDS results to examine and/or predict by which dimensions individuals will categorize auditory stimuli. In a continuation of their MDS analysis, Gygi and colleagues (2007) studied the way individuals would freely categorize these natural auditory stimuli; specifically, would the dimensions implied by the MDS solution also be used in categorization, or would individuals establish different categorization criteria based on the vast number of auditory dimensions present in the

natural stimuli. As might be anticipated given the type of stimuli, individuals did not strictly group sounds together according to these similarity judgments; instead participants established their own categorization criteria based on the saliency of auditory dimensions, specifically audio source (machine vs. non-machine, or animal vs. human) and context (outdoors vs. indoors). As might be gleaned from these category divisions, participants tended to separate auditory stimuli uni-dimensionally, such that stimuli either belonged to one group or belonged to the other, and participants separated by sound source. Thus, rather than construct complex rules or categorization criteria, participants tended to define and categorize auditory stimuli based on a subset of salient stimulus features and use them in a straightforward manner (Gygi et al., 2007).

Goudbeek, Swingley and Smits (2009) found a somewhat similar pattern of uni-dimensional categorization behavior in both supervised and unsupervised learning tasks. In their experiment, Goudbeek and colleagues pre-defined the stimulus dimensions to vary according to duration and formant frequency rather than determine the relevant number and type of dimensions. Participants categorized stimuli either based on a single dimension (either duration or frequency) or according to multiple dimensions (duration and frequency). In supervised learning, these instructions would correspond to learning a rule of uni-dimensional categorization such that either frequency or duration was relevant to successfully categorize the stimuli. In such conditions, participants readily learned the categorization rule and demonstrated a high level of categorization performance in terms of percentage correct (although, there was a slight increase in performance when categorizing by duration rather than frequency). When the categorization rule required

participants to attend to both dimensions (similar to the 3[4]-2 category type previously discussed in Chapter 4), participant performance dropped such that correct categorization was only slightly better than chance.

Unsupervised learning of the auditory stimuli resulted in similar behavior by the participants. Because unsupervised learning does not entail the correct/incorrect performance metric of supervised learning, Goudbeek and colleagues examined the number of categorization solutions performed by participants when a particular dimension was sufficient for categorizing the stimuli. For example, if only duration varied between the two stimuli – taking on two possible values – then participants should optimally construct categories based on only that dimension. When tested, they found that participants frequently used the single relevant dimension to establish their category boundary. If duration was the relevant dimension, participants constructed two categories based on the two durations; if formant frequency was the relevant dimension, participants constructed two categories based on the two formant frequencies. However, when both dimensions were relevant towards creating possible categories, 7 of the 12 participants continued to use only a single dimension to create each category (Goudbeek et al, 2009).

Much like the study by Gygi and colleagues, (2007), this demonstrated persistence in using a single dimension rather than multiple dimensions in order to categorize auditory stimuli implies an important distinction between how we categorize different modalities and stimuli. When presented with visual stimuli, participants can readily learn and extract the relevant patterns for successful categorization. For instance,



category types of the 3[4] family, as previously mentioned, require differing amounts of attention allocation to the dimensions to categorize successfully (Nosofsky, 1984; Shepard, Hovland, & Jenkins, 1961). Somewhat remarkably, while participant performance indicates an obvious preference and ability to categorize uni-dimensionally, performance rates of category types requiring attention towards multiple dimensions continues to remain relatively high (e.g., 3(4)-2; Nosofsky, 1984; Nosofsky, et al., 1994). Although it might be reasonable to expect individuals to use the additional auditory information to assist in categorization much as they do with visual stimuli, this does not appear to be the case (Goudbeek, et al., 2009). Instead, individuals might be developing concepts based on single dimensions and their concept formation is apparent in their categorization strategy.

However, there is an obvious gap in knowledge between these and many other studies examining dimensional relevance and categorization. Namely, no study to date examines both similarity assessment – using multi-dimensional scaling to establish the number and type of auditory dimensions used by individuals – and how individuals engage in an unsupervised, free-sorting task using well-defined category structures. The study by Gygi and colleagues (2007) did examine auditory similarity assessment and free-sorting categorization but used naturalistic sounds. Such stimuli tend to be ill-defined with respect to category structure, such as those categories defined by Boolean logical rules (e.g., Feldman, 2000; Vigo, 2009; Vigo, 2013). Even when tightly controlling auditory stimulus dimensions through digital signal processing or meticulous field recording, dimensions can take on to a continuum of possible values and interact in

unique ways that inhibit a researcher's control of dimensional values. Understanding how individuals determine similarity and subsequently engage in free-sorting categorization using tightly controlled and defined auditory stimuli can go a long way towards understanding the conditions under which we spontaneously develop concepts of non-speech auditory stimuli.

Additionally, few studies have applied the current mathematical and algorithmic models of categorization and concept learning to auditory stimuli (only Vigo, Barcus, & Yu, 2015 appears to apply these types of models to audio). These models are proposed accounts of human concept learning and categorization behavior but have only examined visual stimuli (Love, et al., 2004; Pothos & Chater, 2002). Therefore, the extent to which these models successful account for human concept learning of other modalities – namely audio – needs to be addressed to assess their robustness not just dealing with different category types, but also with modeling different modalities. Should the models be successful, this can provide some insight into the underlying cognitive processes individuals engage in when developing and learning categories and concepts.

Perhaps by using well-defined auditory stimuli, inferences about human similarity assessment and the underlying categorization and concept learning mechanisms may become apparent. In the current experiments, we explore how individuals engage in an unsupervised free-sorting task in order to determine the manner in which individuals are processing these stimuli and developing concepts.

## Chapter 8: Current Research

The intention of this dissertation proposal is three-fold. The first objective of the current research will be to determine the number of auditory dimensions needed to create unsupervised learning categories as it relates to human attention of auditory dimensions, assuming the current methods are sensitive enough to establish this determination. As is typical in this research, participants will assess similarity between pairs of audio stimuli that vary according to several dimensions (Caclin et al., 2005; Grey, 1977; Krumshansl & Iverson, 1992; Mark & Melara, 1990). To determine the number of auditory dimensions, I will conduct multi-dimensional scaling (MDS) on similarity comparisons prior to the unsupervised learning task. The auditory stimuli will vary according to dimensions recognized in the literature as having a significant contribution on the tonal qualities of an auditory stimulus; namely, frequency/pitch, amplitude/loudness, timbre (spectral structure, spectral centroid), and the timbre-amplitude interaction (log-attack time). I expect that the MDS solution will correspond to limits in the participants' auditory attention systems. Specifically, because the limits of the average human auditory attention system hit an upper bound of approximately three to four dimensions (Chen & Cowan, 2005; Cowan, Chen, & Rouder, 2004; Sauls & Cowan, 2009; Tulving & Patkau, 1962), I anticipate that the best MDS solution will be approximately three dimensions. Such a finding would add to the current body of research also confirming similar results (Gygi, et al., 2009).

The second objective is to examine the manner in which individuals classify well-defined integral-dimension auditory stimuli using an unsupervised free-sort learning task.

Previous research (Gygi, et al., 2007) has examined unsupervised learning using categories of stimuli that are dimensionally more robust than well-defined categories (e.g., naturalistic sounds) and, as a result, may have contained excessive information that interferes with categorization behavior (Garner & Felfoldy, 1970). Previous research that has used dimensionally defined auditory stimuli has generated stimulus dimensions that were possibly arbitrary and not explicitly determined to be of importance to human categorization and concept learning behavior (Goudbeek, et al., 2009). By using well-defined stimuli developed based off MDS results, the current unsupervised learning experiment bridges the gap between the studies of Gygi and colleagues (2007) and Goudbeek and colleagues (2009). In the Goudbeek study, researchers examined unsupervised learning using well-defined stimuli but the dimensions of these stimuli were arbitrary defined. That is, the choice of dimensions on which stimuli varied was not necessarily based on how and what individuals dimensionally perceive and attend to in a given auditory stimulus. In contrast, Gygi (2007) did use MDS to determine dimensions which individuals may attend and use as the basis of categorization, but they used naturalistic sounds and the MDS analysis didn't inform the development of the auditory stimuli used in unsupervised learning.

Thus, in the current unsupervised learning experiment, using well-defined stimuli allows increased research control over auditory dimensions. The use of auditory stimuli developed in this manner allows researchers to make informed decisions not only about the unsupervised learning process but the unsupervised learning process in the context of how individuals initially reduce auditory dimensional space through attention,

discrimination, and similarity assessment of auditory dimensions. I anticipate that during the free-sorting task, individuals may exhibit a preference towards uni-dimensional categorization as demonstrated in previous experiments, even though the dimensional space of the stimuli in the current experiment is more tractable than that of naturalistic stimuli (Gygi, et al., 2009). That is, participants will prefer to categorize auditory stimuli using a single subjective diagnostic dimension (e.g., categorizing by pitch, or by instrument type) rather than categorizing based on multi-dimensionality. Overall, I believe that the data will tend to show a linearly separable trend, such that individuals categorize stimuli as belonging to one of two categories.

The third and final objective of the current experiments is to examine three models of unsupervised learning with respect to the categorization data: the simplicity model (Pothos & Chater, 2002), SUSTAIN (Love, Medin, & Gureckis, 2004) and GIST (Vigo, 2013, 2014). While all three of these models examine both supervised and unsupervised learning of visual stimuli, to date researchers have not examined the robustness of these models to account auditory categorization behavior<sup>4</sup>. Examining model performance in this capacity is dual-purpose; it allows researchers to make assessments about model performance and appropriateness with respect to the data and it may allow for some insight into the cognitive processes of individuals when behaving and engaging in auditory categorization tasks.

---

<sup>4</sup> The exception to this is a study conducted by Vigo and Barcus (2015) in which they used the pre-cursor to the GISTM, the categorical invariance model (Vigo, 2009), to examine auditory stimuli. However, the GISTM itself remains untested as of yet on such stimuli.

Of these models, it is hypothesized that the GIST (Vigo, 2013, 2014) and the SUSTAIN (Love, Medin, & Gureckis, 2004) model will demonstrate unsupervised learning more similar to that of the participants than the simplicity model. Previous research (Pothos, et al., 2008) using visual stimuli demonstrated that the lack of parameterization on the simplicity model a similar trend due to the high level of parameterization within the SUSTAIN model. Additionally, the simplicity model has difficulty with unsupervised categorization tasks that include multiple, smaller categories (Pothos, et al., 2011). Essentially, if the distribution of stimuli results in more than two groups, particularly if these groups contain an unequal number of stimuli, the simplicity model falls apart. Given these issues, it is anticipated that SUSTAIN will outperform and the simplicity model using auditory stimuli, especially if the stimulus groupings are complex. With respect to the GIST, although it is untested in regards to unsupervised learning experiments, previous research in supervised learning (Vigo, 2009, 2011, 2013, 2014) and information judgments (Vigo, 2011b; Vigo & Basawaraj, 2013) has established that the mathematical framework is both flexible enough to accommodate interpretations of different psychological phenomena and robust enough to make accurate predictions of human performance relative to other models of concept-learning behavior (Vigo, 2009, 2011, 2013, 2014). The GIST should, if prior research is any indication, provide an accurate description of human concept-learning behavior in the unsupervised learning task and should provide some key indicators of an underlying cognitive process.

## Chapter 9: Experiment 1

The purpose of the first experiment is to determine the number and type (e.g., pitch, loudness, etc.) of psychologically relevant auditory dimensions needed to create the auditory stimuli for unsupervised learning. To accomplish this, the number of relevant dimensions will be confirmed via multi-dimensional scaling (MDS) and using participants' rating of each of auditory dimension with respect to their allocation of attention will determine which dimensions are used in the stimulus construction. To determine these reduced dimensional aspects of auditory stimuli, participants engaged in pairwise similarity ratings of auditory stimuli.

### Method

**Participants.** Participants included 35 undergraduate students from Ohio University recruited through the Psychological Research Participant Pool who received course credit for participation. Only participants who report normal, un-assisted hearing and who are not professional musicians can participate in the study.

**Stimuli.** The five auditory dimensions examined in experiment 1 were: frequency/pitch (370 or 523 Hz; Krumshansl & Iverson, 1992), amplitude/loudness (64 dB or 70 dB; Melara & Mark, 1990), spectral centroid (no harmonic unit above fundamental frequency or three harmonic units above fundamental frequency; Caclin, et al., 2005), overall spectral structure (string instruments or wind instrument; Caclin, et al., 2005; Grey, 1977), and log-attack (15 ms or 200 ms; Caclin, et al., 2005). Table 1 presents each value a stimulus dimensions can take. According to the formula in Chapter 4, 5 dimension generate a total of 32 stimuli (Table 2 and 3).

Stimulus dimensions of frequency, amplitude, spectral structure, and log-attack time were generated using the ChuckK v.1.3 programming language's library of physical instrument models. Sonic Visualizer v.2.0 confirmed dimensional values for each stimulus in order to standardize the stimuli and decrease the integrated influence each dimension exerts over the others. Consistent with the findings of previous research concerning similarity judgments of harmonic units (Caclin et al., 2005), the spectral centroid of manipulated sounds is equivalent to the third harmonic unit. To manipulate the spectral centroid, audio derived from ChuckK and verified in Sonic Visualizer was loaded into Ableton Live v.8.2.6 where the spectral centroid of each sound was manipulated using SpectrumWorx's v.2.5.0 spectral centroid module. Output from Ableton Live and SpectrumWorx was re-examined in Sonic Visualizer to ensure that the original values remained consistent and not affected by the additional changes. After verification of dimensional consistency, each stimulus was output as a .wav file recorded in 16-bit and 44100Hz format.

**Procedure.** Researchers led participants into the research laboratory where participants sat at a table to review participation consent forms. After completing the forms, the researcher read the following experimental instructions:

In the following experiment you will be asked to rate the similarity between the two sounds. During the experiment, a pair of sounds will be played. After you have listened to both sounds, use the slider located below the icons to rate the similarity. A value of 1 means they are identical; a value of 11 means they are very different. You will have 20 seconds to provide a rating, so be prompt in your



response. After you have made a similarity judgment on the sounds you have heard, click the 'Enter' button on the computer keyboard to continue with the experiment.

After answering any participant questions, the researcher performed an informal hearing test. To identify that participants can adequately hear the frequency range of the experimental audio, participants listened to a series of eight sine wave harmonics beginning at 370 Hz and ascending to 2960 Hz. Each tone played at 64 dB and the timing between each sound randomized in order to dissuade automatic, temporal-based responding. Participants wore a pair of Koss SB/45 headphones to listen to the tones. If the participant heard all eight harmonic tones, they listened a second hearing test to determine the upper frequency bound of their hearing. The researcher played a series of seven sine wave tones play at 64 dB and beginning at 14 kHz and ending at 20 kHz for the participant. As in the previous experimental-tone hearing test, the program had randomized timing between each tone. With the upper hearing range identified, the researcher recorded the value and the participant was asked to self-report any music performance experience.

Following the hearing test, participants sat in front of a HP WX4600 workstation with a Dell 1798FP 15-inch flat panel LCD monitor (5-ms response time). At each experimental station, the participant listened to the audio through Koss SB/45 headphones. The experiment, programmed in PsychoPy version 1.78.01, began by presenting the experimental instructions. Participants can initiate the experimental trials with a response on the "Enter" key. During a single trial of the experiment, a target cue

appeared on the screen and two of the sounds played sequentially. After the sounds played, the target cue disappeared and a rating slider appeared on the screen. Participants used the slider to rate the similarity between the pair of sounds using a scale ranging from 1 (identical) to 11 (completely dissimilar). After confirming their similarity rating by pressing “Enter” on the keyboard, participants saw another target cue and two more sounds played in a sequence. Participants went through half the total number of possible stimulus pairings (512) and took a short 5-10 minute break halfway through the experiment to rest their ears. The volume of the auditory stimuli remained at a consistent, comfortable level throughout each trial. Overall, most participants completed the task in approximately an hour.

After completing the experiment, participants completed a form in which they rated the five auditory dimensions used in the experiment according to the amount of attention they allocated to each dimensions during the similarity judgment task. Each dimension included a definition in understandable terms such that the participant understood which dimension they are rating. For example, the spectral structure would be described as “Type of instrument playing,” thus differentiating between the two physical models. Below each description, participants assigned ratings to each described dimension using a scale from 0 to 10, where a rating of ‘0’ indicates no attention was given to that dimension and a rating of ‘10’ indicates attention resources were allocated predominately to that dimension. The researcher then debriefed participants on the experiment once they complete the rating form.

## Results

Subject data was examined to determine whether participants were responding and engaging in the task. Data demonstrating systemic, obvious indications that the participant was not engaging in the task appropriately were discarded from further analysis (e.g., giving ratings of ‘11’ to every stimulus pair, only rating using the extreme values). Furthermore, subjects that demonstrated a high level of musical experience such that it exceeded 10 years, regardless of instrument and time in which they stopped playing (if applicable), were also excluded due to their extensive training and expertise dealing with such stimuli (Tervaniemi, et al., 2005). Of the remaining participants, 10 had experience playing musical instruments with an average of 4.1 years of experience (median of 3.5,  $SD = 3.5$ ), and a range of 2 to 8 years’ experience.

The similarity ratings from the remaining subjects ( $N = 32$ ) were analyzed using non-metric multidimensional scaling (NMDS) and models were generated by iteratively increasing the dimensional space of the NMDS solution (Giguere, 2007). To assess the appropriate number of dimensions from which stimuli for the second experiment would be constructed, several metrics were used to examine each NMDS dimensional model; these were the stress benchmarks put forth by Kruskal & Wish (1978), visual examination of the stress scree plot (Borg & Groenen, 2005; Giguere, 2007; Groenen & Velden, 2005; Jaworska & Chupetlovska, 2009), examination of the Shepard diagrams (Bord & Groenen, 2005; Jaworska & Chupetlovska, 2009), and examining each solution according to its’ intuitiveness given the conducted metrics.

The stress of each dimensional solution of the NMDS was examined using the benchmarks of Kruskal and Wish (1978). These benchmarks indicate the fit of the NMDS model to the data given the dimensionality. The stress benchmarks, while lacking interpretability between datasets, allow for a comparison of the relative fit each dimensional NMDS model. As such, the stress values may fall within one of the following intervals in terms of fit:  $\text{stress} \geq 0.20$  (poor),  $0.10 \leq \text{stress} \leq 0.20$  (fair),  $0.05 \leq \text{stress} \leq 0.1$  (good),  $0.025 \leq \text{stress} \leq 0.05$  (excellent),  $\text{stress} < 0.025$  (perfect). Thus, an increasingly lower proportion of stress indicates that the model fits the data better than other, lower dimensional models (Gigeure, 2007; Kruskal & Wish, 1978).

As anticipated, increasing the dimensionality of the model results in increasingly lower stress metrics (Table 4) with the most pronounced decrease occurring between a one and two dimensional NMDS model. According to the benchmarks, the one and two dimensional solutions only provide “fair” and “good” fits to the data, respectively. Increasing the dimensionality to a three dimensional solution offers the first indication of an “excellent” fit to the data (stress = 0.034), while an increase to a four dimensional solution decreases stress such that it barely exceeds the next categorical benchmark cutoff (stress = 0.022). Thus, a three dimensional NMDS solution may provide a sufficient model from which to base further research. Beyond three dimensions, the decrease in model stress between the three and four dimensional NMDS is not as large as previous model. Any improvements in fit are likely just a function of higher dimensionality rather than meaningful. A four dimensional solution may be fitting the data better (perhaps even overfitting) at the expense of interpretability (Borg & Groenen, 2005; Kruskal,

1964); specifically, does the four dimensional solution accurately reflect participant's behavior or is this just a mathematically better fit with no behavioral underpinnings? Because determining the appropriate number of dimensions is integral to the second experiment, further tests were conducted to determine which dimensional model provided the closest meaningful interpretation of the behavioral data.

A scree plot of the stress values from each dimensional model indicates a similar, albeit less definitive, assessment (Figure 10) relative to the benchmark method. No clear "elbow" or leveling off indicating the appropriate number of dimensions exists in the scree plot; rather, there is a gradual decrease for stress between the three, four and five dimensional solutions. The visualized stress data lends an amount of ambiguity as to which dimensional solution is most appropriate, but based on the scree plot it would appear that the solution is potentially higher dimensional. Specifically, there does appear to be more leveling off in the scree plot after the three dimensional model than previous dimensional models.

To further examine the NMDS models, Shepard diagrams using linear and non-metric  $R^2$  fits of each NMDS dimensional solution were examined to determine the number of dimensions (Figures 11-15). Visually, the variance of points around the model fit decreases as the number of dimensions increases resulting in high  $R^2$  fits for even low-dimensional solutions. A scree plot of the  $R^2$  values for the linear (Figure 16) and non-metric (Figure 17) fits shows a similar, albeit exaggerated, pattern as compared to the stress scree plot. The "elbow" of these graphs distinctly occurs on the two dimensional NMDS solution. There is still some fit increase on the third dimension and little to no

increase for higher dimensional models. Figures 18 shows that when the three dimensional NMDS solution is graphed, three dimensions are clearly being used in order to assess similarity. Thus, it appears that the appropriate number of dimensions indicated by the Shepard diagrams is possibly two to three dimensions.

This result is consistent with the stress benchmarks of each NMDS dimensional solution. Recall that the stress benchmarks indicated that the two dimensional solution only provided a “good” fit to the data while the three dimensional solution provided an “excellent” fit. Contrast against the Shepard diagram data, it would seem that, while the two dimensional solution is accounting for a large amount of variance, the three dimensional solution is still contributing to the model and a three dimensional model make interpretable and intuitive sense.

Previous experiments have demonstrated that MDS solutions of auditory stimuli tend to demonstrate three dimensional solutions rather than lower dimensional solutions and the current one is no exception. In a previous MDS experiment, Gygi, Kidd, and Watson (2007) found that individuals assessed similarity based on three auditory dimensions when making similarity judgments of naturalistic sounds. Other research has found similar results; when given a set of stimuli and asked to judge similarity, individuals tended to use three dimensions to accomplish this task (Alrich, Hellier, & Edworthy, 2008; Bonebright, 2001; Howard & Silverman, 1976; McAdams, et al., 1995; Samson, Zatorre, & Ramsay, 1997). For example, Howard and Silverman (1976) performed multidimensional scaling on auditory stimuli and found that individuals were assessing stimulus similarity based on three dimensions (namely, fundamental frequency,

type of waveform, and presence of high frequency formants). Therefore, based on the obtained data and a review of the literature, it appears that the NMDS solution that most adequately satisfies the tests of fit and interpretability is a three dimensional model.

To determine which of the five dimensions were used by participants in the similarity assessment, participants rated each of the five dimensions on a scale of 0 to 10 to indicate to how often they used that dimension to make their similarity judgments. A value of '0' corresponded with a dimension that was never used to determine similarity; a value of '10' corresponded with a dimension that was using in every similarity comparison. Descriptive statistics from these ratings, presented in Table 5, show a qualitative pattern of dimensional number that corroborates the story told by NMDS. Namely that of the five dimension used in the audio stimuli, three dimensions were rated most highly and similarly- frequency, amplitude, and spectral envelope were rated most highly, respectively. To evaluate differences between the five auditory dimensions, a Kruskal-Wallis test was conducted on participant rating. Correcting for tied ranks, the test was significant  $X^2(4, 32) = 39.68, p < 0.001$ .

Follow-up tests using a Bonferroni adjustment were conducted to determine significant differences among the five auditory dimensions. The significance values are presented in Table 6. Of the comparisons, three were significant with one approaching significance. Pitch, rated highest by the participants, was significantly different from spectral centroid and log-attack time but not loudness or spectral envelope. Loudness was also significantly different from log-attack time. The difference between spectral envelope and log-attack time approached significance. It appears that the participant

ratings have nearly formed two groups: one group being pitch, loudness, and spectral envelope, and the other group being spectral centroid and log-attack time. There is some overlap between these groups in that loudness and spectral envelope are not strictly significant from spectral centroid and log-attack time. However, it may be that the integrality of dimensions is influencing participant rating such that spectral centroid and log-attack time are both influencing perceived loudness and timbre perception (Caclin, et al., 2005; Melara & Marks, 1990; Nelson, 1993; Pitt, 1994). Despite this potential influence of integrality, it appears that participant ratings correspond to the three dimensional NMDS model and that these three dimensions correspond to pitch, loudness, and spectral envelope.



## Chapter 10: Experiment 2

The second experiment was an unsupervised learning task using auditory stimuli dimensionally defined according participants' attention ratings and using a number of dimensions confirmed by the MDS results from experiment 1. The purpose of the experiment is to examine how individuals spontaneously sort a group of sounds given the experimental verified constraints on the number and type of auditory dimensions.

### Method

**Participants.** Participants included 49 undergraduate students from Ohio University recruited through the Psychological Research Participant Pool who received course credit for participation commiserate with experiment duration.

**Stimuli.** The stimuli were generated in the manner described above, although they were created using only the three dimensions identified in the MDS: frequency, amplitude, and spectral envelope.

**Procedure.** The researcher led participants into the research laboratory where participants reviewed participation consent forms. The experimenter read the following instructions, similar to those used by Pothos and Chater (2002), to the participants:

In the following experiment, you will be asked to categorize a set of auditory stimuli. Square icons will be presented on the computer screen. Using the left mouse button to click on an icon will activate the audio clip of a particular sound from the audio group; likewise clicking a different icon plays a different sound from the group. When listening to the sounds, use the slider below each icon to assign the sound to a group. For example, if you believe two sounds are similar

enough to be grouped together, you would assign a value of 1 to each of those sounds. When assigning each sound to groups, try not to use more groups than is necessary. You may change your group responses for any of the individual sounds during the trial. Once you have assigned a group value to each of the sounds, hit the “Enter” key on the keyboard to move to the next set of auditory stimuli.

The researcher answered any participant questions and sat the participant in front of a HP WX4600 workstation with a Dell 1798FP 15-inch flat panel LCD monitor (5-ms response time). At each experimental station, participants listened to the audio through Koss SB/45 headphones. The researcher initiated the unsupervised learning program written in PsychPy version 1.78.01. The experiment began with a prompt containing on-screen instructions similar to those read by the researcher. The participant started the experiment by pressing the “Enter” key or, if no response is made, the experiment began automatically after 60-s. The experiment will consist of a number of trials such that the participant categorizes all category types for the given number of dimensions (Feldman, 2003). During a single trial of the experiment, participants saw a set of white square icons arranged in a manner to dissuade grouping/patter bias. The exact arrangement of the set of square icons depended on the category type/number of stimuli. Each square icon was associated with a specific auditory stimulus in the current category set (Figure 19). Participants moved the mouse pointer around the screen and, by clicking the left mouse button when the mouse pointer is within a square icon, they triggered audio clips associated with each square icon. When participants clicked a square icon, the audio clip

played uninterrupted for its full duration (2-s); additional mouse presses during this time did not re-trigger the audio clip. Only when the audio completed playing could the participant re-trigger the audio clip. Participants could play each audio clip as often as they wanted within each category set.

Below each icon was a slider where individuals input the category label they wished to assign that specific sound. For example, assigning a label of '1' to two different sounds indicated that the participant found those stimuli sufficiently similar, according to their dimensions, to categorize them together. Accordingly, participants grouped the stimuli in any way that they saw fit and in as many groups as they deemed necessary. Participants were able to re-assign category labels within each categorization set without penalty if they wanted to change groupings.

After participants assigned each auditory stimulus a label, the participant could hit enter and continue to the next category set. In total, participants categorized the following categories: 3(2), 3(3), 3(4), 3(5), and 3(6). Each participant examined all the types within each category a total of four times. Cumulatively, participants sorted 156 categories total.

## **Model Analysis**

After determining the manner in which individuals categorize auditory stimuli a critical analysis, a model comparison of the predominant mathematical descriptions of unsupervised learning was performed. While other minimization models exist – namely, Feldman’s minimization (2000) and mental models (Goodwin & Johnson-Laird, 2011) – their current derivations are specific to supervised learning tasks. For that reason, analysis will consist of a comparison between two models of simplicity created specifically for the analysis of unsupervised learning; the simplicity model (Pothos & Chater, 2002) and the Supervised and Unsupervised Stratified Adaptive Incremental Network (SUSTAIN; Love, Medin, & Gureckis, 2004) and the Generalized Invariance Structural Model (GIST; Vigo, 2013, 2014). The following is a brief description of how these models function.

**Simplicity model.** Pothos and Chater (2002) developed a model of unsupervised categorization behavior based on a simplicity principle that the authors referred to as the simplicity model (henceforth referred to as SM). To achieve simplicity, specifically in relation to unsupervised learning, an individual would try to minimize within-group differences while maximizing between-group differences. In this case, simplicity would refer to a participant categorizing stimuli in the least number of groupings or clusters, while maximizing an internal consistency between within-group stimuli (Fleiss & Zubin, 1969; Pothos & Chater, 2002, 2005; Zubin, 1938). Such a method of categorization is similar to recently proposed theories in the field of supervised categorization, such as minimization complexity (Feldman, 2000), algebraic complexity (Feldman, 2006), and

mental models (Goodwin & Johnson-Laird, 2011). While the proposed cognitive mechanisms differ slightly between these models and that of SM, the underlying proposal is similar; that is, when an individual learns a category they attempt to minimize required knowledge in order to learn the category. Stated differently, an individual is guiding their categorization behavior through using the simplest possible mechanism to group objects - whether that is color, shape, or some simple, easily implemented combination of object features. For example, suppose a participant views a group of related stimuli, such as the category “chairs.” For simplicity, these stimuli – chairs - vary on three dimensions; namely, number of legs, wood grain, and height from the floor. Additionally, all vary on type of wood grain and height from the floor, but share the dimension four legs. According to minimization theory, success in learning this category would be dependent upon learning only one dimension: number of legs. Since the other dimensions vary between chairs, learning each set of features per chair would decrease the learnability of the category. Therefore, the participant can easily learn the category according a single dimension, and this single dimension would represent a minimized expression (Feldman, 2003).

Minimization is central to the simplicity model (SM). As previously mentioned, a central goal unsupervised learning tasks is to increase within-group similarity, and decrease between-group similarity during categorization. As a measure of overall similarity and learnability, or compression, Pothos and Chater (2002) adopted the usage of code length from information theory. Increased code length reflects a decrease in compression and a decrease of within-category consistency or *categorical coherence*;

categorization that results in a short code length reflects an increase in categorical coherence and may be the preferential categorization of stimuli by the individual. However, the SM makes no assumptions on the representation of the concept, therefore it is not exclusive to theories of mental representation such as prototypes (Rosch, 1975; Rosch & Mervis, 1975), exemplars (Medin & Schaffer, 1978, Nosofsky, 1984, 1986) process accounts of minimization (Goodwin & Johnson-Laird, 2011) or invariance theory (Vigo, 2009, 2013).

Similar to other models of minimization and similarity (Tversky, 1977), SM makes several assumptions about the nature of data. First, the data must adhere to minimality, or that the distance of stimulus A from itself is always equal to zero. Second, stimuli within the data must demonstrate symmetry. As a specific example, the distance of stimulus A to stimulus B must be equivalent to the distance of stimulus B to stimulus A. In contrast to previous theories of similarity, SM theory does violate transitivity for reasons noted by the authors, but accordingly does not affect optimality (see Pothos & Chater, 2002 for details). Of particular interest, SM theory makes no assumptions on the number of groupings, or category numbers, required by the participants. This lack of group distribution, or non-parametricity, is a particular strength of the model. In unsupervised learning tasks, there is often no “correct” response for categorization of a given stimulus; rather, as described previously, the purpose of unsupervised categorization is to determine conditions under which an individual will categorize in a particular manner. Therefore, the lack of assumptions regarding category number is congruent with natural categorization; rarely does an individual have pre-existing

knowledge concerning general category formation (Ashby, Queller, & Berretty, 1999; Pothos & Chater, 2002; Pothos, Perlman, Edwards, Gureckis, Hines, & Chater, 2008).

To determine these conditions, the SM computation includes in two parts: (1) the code length of grouping similarity, and (2) the code length of group. In determining part (1), Pothos and Chater define a grouping such that all within-group similarity is greater than any between-group similarity. First, the number of distances between stimuli is determined by:

$$s = r(r - 1)/2$$

where  $r$  is the number of objects in the unsupervised learning task and  $s$  is the total number of comparisons. Using this value, the total number of similarity constraints (inequalities) in the unsupervised learning task assuming no groups is:

$$u = s(s - 1)/2$$

Because unsupervised learning is concerned with how individuals form categories, only a certain number of the total inequalities will exist between groups. Each group will have a number of within-group inequalities ( $n_{group} * (n_{group} - 1)/2$ ) and between-group inequalities ( $n_{group1} * n_{group2}/2$ ). Multiplying these values gives the number of distances (inequalities) according to the number of groups.

Because groups are being used, the codelength necessary to specify these groups must be determined and the influence of errors in group specification must be taken into account. The formula for to compute this is:

$$\log_2(u + 1) + \log_2(uC_e)$$

where  $\log_2(u + 1)$  is the codelength to specify groupings and  $\log_2({}_u C_e)$  is the codelength to correct for errors. First, to specify the number of groupings, the formal computation of  $\log_2(u + 1)$  is:

$$\log_2 \sum_{v=0}^n (-1)^v ((n - v)^r / (n - v)! v!)$$

where  $r$  is the total number of objects and  $n$  is the number of categories. This computational cost to specify the number of categories results in a decrease in the number of total inequalities given a number of categories,  $u - \log_2(u + 1)$ . If the individual makes a particular error in the unsupervised learning task,  $\log_2({}_u C_e)$  is computed. Errors are instances in which an individual will incorrectly judge distances between pairs of stimuli (Pothos & Chater, 2002). For example, if in reality the distance between pairs was  $(A,B) < (C,D)$ , but an individual group pairs such that  $(A,B) > (C, D)$ , this would reflect a participant's error given the categorization constraints. The number of  $e$  errors varies between 0 to  $u$  and the codelength necessary to correct for these errors is:

$${}_u C_e = (u/e!(u - e)!)$$

The final codelength to correct for errors,  $\log_2({}_u C_e)$ , and the cost to specify groups,  $\log_2(u + 1)$ , affect the compression of the final codelength. A greater compression of the codelength (i.e., the smaller the resulting bit size) indicates more category coherence. As described by Pothos and Chater (2002), "smaller codelengths should correspond to more obvious groupings" (p. 317).

As an example, let's use the 3[4]-1 category of objects. As previously described, this category is linearly separable according to a single dimension; e.g., four of the eight



objects will be triangular and the remaining four objects will be circular. The number of similarities between pairs of these eight objects corresponds to:

$$s = r(r - 1)/2$$

where  $r$  is the number of objects. From this the resulting number of similarities within 3[4]-1 is  $8(8-1)/2 = 28$  distances between pairs. Furthermore the total number of inequalities is given by  $s(s - 1)/2$  resulting in a value of  $28(28-1)/2 = 378$  inequalities.

Because 3[4]-1 is linearly separable, let's assume that when analyzed by a clustering algorithm two distinct groups are established. Within each group, there will be four objects resulting in  $4(4-1)/2 = 6$  distances (12 distances overall) and  $4*4 = 16$  distances across groups. This gives us a total of  $(12 * 16) = 192$  inequalities. Because we're specifying categories, the number of bits required to create these categories is:

$$\log_2 \sum_{v=0}^n (-1)^v ((n - v)^r / (n - v)! v!)$$

resulting in approximately 7 bits of code length. If the participant makes no grouping errors (because these groups are so very distinct), the final compression from the 3[4]-1 category is a compressed code length of 185 with a full code length of 192 bits. Because these code lengths are small, the implication is that an individual may easily distinguish and categorize the available objects in such a way as to maintain a high level of category coherence.

Therefore in summary, the SM works on the similarity principle wherein an individual attempts to create a high degree of category coherence by decreasing the dissimilarity within-group and increase dissimilarity between-group (Fleiss & Zubin, 1969; Pothos & Chater, 2002, 2005; Zubin, 1938). In doing so, SM can be used to

determine the category that adheres closest to the similarity principle; information particularly relevant to human classification performance.

**SUSTAIN.** SUSTAIN (Love & Medin, 1998; Love, Medin, & Gureckis, 2004) is a connectionist network implementation of a clustering algorithm that models both supervised and unsupervised learning. Stimuli are categorized on a trial-by-trial basis and, when categorizing stimuli, SUSTAIN operates on a simplicity principle similar to that of the simplicity model (Pothos & Chater, 2002); simple solutions are considered first before more complex solutions are used. For example, if a group of stimuli can be clustered according to a single dimension (e. g, all blue stimuli or all red stimuli), SUSTAIN will use this linear separation to generate a clustering solution. The model uses similarity between the stimuli to determine category membership, and establish the fewest number of clusters possible for categorization.

When a novel stimulus does not conform to the current clustering, only then does SUSTAIN create different clusters or sub-clusters to accommodate the stimulus. Such adaptive clustering, initially beginning with simple solutions and progressing to complex clustering, allows the model to avoid issues relating to standard back propagation techniques and the bias-variance dilemma (Love, et al., 2004).

SUSTAIN compares the similarity of currently presented stimulus against that of each cluster. Clusters then, in essence, compete for inclusion of the presented stimulus through activation, which varies according to attention weights that adjust to direct the model's focus towards the most salient feature, or set of features, that facilitate consistent categorization behavior. When there is a cluster that is clearly more active, SUSTAIN

assigns that stimulus to the cluster. In addition, when proper clustering is less clear-cut based on activation, SUSTAIN creates a new cluster should a particular threshold value be exceeded. This parameter, referred to by Love and colleagues (2004) as the “cluster recruitment mechanism”, is a free parameter and allowed to vary between values of 0 and 1. Higher values indicate a higher threshold; therefore, such high values will decrease the likelihood that SUSTAIN creates a new cluster for a given stimulus.

Applied to data, SUSTAIN has been found to replicate human performance in unsupervised learning tasks, such as those conducted by Billman and Knutson (1996) and Medin and colleagues (1987). Through application of the SUSTAIN model to the experimentation data, it was found that unsupervised categorization behavior depended on the saliency of the stimulus features. Additionally, it was also found that SUSTAIN is somewhat unaffected by the initial parameter values; more specifically, the model is slightly insensitive to these values. Therefore, SUSTAIN used the parameter values reported by Love et al. (2004) in the current experiment. Such a tactic is not without precedent; Pothos and colleagues (2008) adopted a similar solution and subsequently only manipulated the threshold parameter. Thus, SUSTAIN will be used to predict categorization behavior of auditory stimuli according to these values in the current experiment.

**Generalized structural invariance model.** The generalized structural invariance model (GISTM; Vigo, 2013, 2014) is an extension of the categorical invariance model (CIM; Vigo, 2009) in which continuous dimensions are permissible for analysis. As previously discussed, the GISTM falls under the ideotype theory of concept-learning (see

Chapter 2). Currently, the model has primarily been applied to predict the degree of learning difficulty of categories in supervised learning tasks; although, in *Mathematical Principles of Conceptual Behavior* (Vigo, 2014) an extension of the model generates the probability of a correct classification for each item of a predefined category. GIST provides the theoretical framework for analyzing qualitative aspects of unsupervised categorization tasks (Vigo, personal communication, October 31, 2014).

First, it is necessary to understand how the theory works through a basic, informal example (for the cognitive underpinnings of the model, refer to Chapter 2 and to Vigo, 2013, 2014). To use Vigo's (2013) example, let's begin with a set of three objects each defined according to three dimensions: shape ( $x$ ), size ( $y$ ), and color ( $z$ ). For the purpose of this example, value assignment to the dimensions is binary, although recall that this can vary continuously between 0 and 1. If the presented categorical stimulus is defined by the Boolean rule  $xyz + x'yz + x'y'z'$ , then it can easily be encoded as the following set  $\{111, 011, 000\}$ , where 111 would refer to a small, black, triangle. Perturbing this category according to each dimension in isolation allows for comparison to the original set. If any objects in the original set remain in the perturbed set, Vigo refers to these objects as categorical invariants. Using the current example, perturbing the shape dimension would result in the new set  $\{011, 111, 100\}$ . Of these, 011 and 111 are present in both sets, thus two of the three objects are invariant to a change to the shape dimension and gives a partial invariance of  $2/3$ . Continuing this process with each

dimension results in a logical manifold of  $(2/3, 0/3, 0/3)$ <sup>5</sup>. According to Vigo's theory, the overall degree of variance is the square root of the sum of the square of the partial invariances:

$$\Phi = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{0}{3}\right)^2 + \left(\frac{0}{3}\right)^2}$$

and can be generalized by using the Minkowski distance (Vigo, 2011, 2013, 2014, p. 100) such that the distance metric becomes:

$$\Phi = \left[ \left[ \left(\frac{2}{3}\right)^s + \left(\frac{0}{3}\right)^s + \left(\frac{0}{3}\right)^s \right] \right]^{1/s}$$

The resulting value determines the degree of perceived difficulty of a categorical stimulus (Vigo, 2013, 2014) using an exponential form of:

$$\psi(X) = pe^{-k\Phi^2(X)}$$

where  $p$  is the number of dimensions (in this case '3') and  $k$  is a discrimination index.

Thus, this last computation of the GISTM is a measure of the degree of perceived concept learning difficulty based on the degree of perceived categorical invariance.

GISTM allows for three different measures of qualitative examination: the structural manifolds, the category difficulty, and the category invariance. The structural manifolds are the mathematical description of the ideotypes – which themselves are memory traces of concepts within psychological space – and they have utility in providing a qualitative representation of the underlying patterns within a category

---

<sup>5</sup> It is important to note that the produced manifold in this example is a logical manifold because the dimensions used are binary. When the dimensions occur along a continuum – values may be between 0 and 1 – it is a structural manifold. To remain consistent with the GISTM, I will refer to as structural manifolds beyond the example.

composed of continuous dimensions. Current research examining the structural manifolds with respect to the category it's representing shows the individuals categorize in a manner consistent with a parsimony principle (Vigo, 2013, 2014).

According to the parsimony principle, individuals have a tendency to place “disproportional emphasis... on structural kernels (SKs) (of ideotypes) with values of 0 and 1.” The reason for this, according to Vigo (2014) is that “the extreme SK values of 0 and 1 of the structural manifold representing the ideotype exert a much greater relative influence on the perceived learnability of a concept than SK values between 0 and 1.” As an example, the resulting structural manifolds of participant groupings would more likely take extreme patterns such as (0/4, 0/4, 0/4) rather than patterns which only show some invariances to perturbations such as (2/4, 2/4, 2/4). The notion is that individuals tend towards the most intuitive groupings possible given a set of objects and dimensions. A structural manifold of (0/4, 4/4, 4/4) is easily understood according to parsimony; every object within that category would remain invariant to two of the three dimensional perturbations. This would be akin to grouping all identical objects together. The structural manifold of (0/4, 0/4, 0/4) represents a set of objects that have no invariances with respect to dimensional perturbations. These two extremes act as anchors from which participants can construct their categories. In the current unsupervised learning experiment, I anticipate that the structural manifolds resulting from the participants' stimulus groupings would adhere to the parsimony principle.

Alternatively, the participants' groupings may demonstrate qualities of the structural equilibrium principle. When a group of stimuli are in structural equilibrium, all

dimensions are diagnostic necessary for an individual to correctly categorize the stimuli. Structural manifolds such as (0/4, 0/4, 0/4) exhibit this quality; no invariances are present in the dimensional perturbations and as a result all dimensions must be attended to by the perceiver in order to form the correct category concept. Vigo (2014) describes structural equilibrium in terms of the following relationship:

[A]s the degree of structural equilibrium of a categorical stimulus X increases, the easier it is to determine which dimensions should participate in rule formation.

(pg. 93)

In other words, an individual presented with a category associated with a structural manifold of (0/4, 4/4, 4/4) would be able to easily determine that the first dimension of this structural manifold demonstrates diagnosticity. That is, if the individual were to partition the whole category into multiple groups, it is likely that they would use that first dimension (that exhibits structural equilibrium) to form their categorization rule. If that first dimension were associated with size, the participant would create groups based on size rather than the other two dimensions.

The second GISTM measure is the category's difficulty. Within the context of supervised learning experiments, a category's difficulty relates to invariances present within the structural manifold, or:

$$e^{-\Phi}$$

Taking the negative exponential of phi (refer to the formula above for the calculation of phi) gives a category's difficulty. For an unsupervised learning task, the GISTM theory can examine the distance between the ideotype (the memory trace of the

structural manifold) and the zero ideotype within psychological space thereby giving the qualitative category difficulty. As put by Vigo (2013):

The shorter this distance is, the less homogenous the categorical stimulus is perceived to be and, consequently, the more difficult it is judged to be from the standpoint of concept formation. (pg. 95)

A manifold, such as the one given above (0/4, 4/4, 4/4), demonstrates some invariances to dimensional perturbations and is therefore more likely to be perceived as being more homogenous than categories lacking these invariances. As a result, an individual may be able to successfully form a concept of the perceived structure. Compare this to a structural manifold of (0/4, 0/4, 0/4): this manifold completely lacks invariances to dimensional perturbations. Because this ideotype matches, or is equivalent to, the zero ideotype in psychological space, an individual would perceive it as being less homogenous or coherent. They would therefore believe the category as being more difficult to form a concept from.

A final measure of the GISTM is categorical invariance. As previously stated, individuals prefer extreme structural kernel values – namely, 0 and 1 - The formula, given above, squares the value of phi and includes a discrimination index parameter  $k$ . By squaring the phi, relative contribution of these extremes is accentuated while the contribution of intermediate structural kernels that fall within the 0 to 1 interval are diminished (Vigo, 2013, 2014). Such a move models the tendency of individuals to gravitate to these extremes given with respect to the category difficulty according to the perceiver. Qualitatively, it becomes apparent using the measure of categorical invariance



where particular structural manifolds are in relation to each other. Thus, the measure of categorical invariance uses category difficulty and the structural manifold to provide a full, qualitative picture of a category's underlying structure and how this underlying structure relates to a participant's categorization behavior.

## **Results**

After excluding participants who had incomplete data, didn't follow task procedure, or had self-reported expertise with musical instruments (Tervaniemi, et al., 2005), we analyzed the data of 39 number of participants. Participants reported, on average, about 1.5 years of experience playing musical instruments (median = 0, SD = 2.4) with a range of 0 to 9 years of experience.

In the current experimental procedure, the amount of data is extensive and analyzing individual categorization behavior presents many logistical and practical issues (Love et al., 2004; Pothos & Chater, 2002; Pothos, et al., 2008) due to the number of unique classifications possible in the data. Therefore, the predominant method of categorization utilized by the participants – referred to by Pothos and colleagues (2008) as category intuitiveness – will be analyzed to determine representativeness of the SUSTAIN and simplicity model simulations.

Pothos (2008) proposed two metrics of category intuitiveness; examining classification variability and frequency of classification method. Category variability related the diversity of categorization solutions within a given category or dataset. If classification variability was low within a particular data set, the implication was that there would be more agreement in participants' categorization and therefore the manner

in which to categorize the stimuli must be more intuitive (Pothos, et al., 2008). Using this metric in the current experiment presents a number of problems. Most notably each category contains a different number of stimuli and as a result, the diversity of categories with large sets of stimuli will often be larger simply due to a larger number of possible responses. However, in the scheme of categorizations made compared to possible categorizations, smaller categories will always show a higher level of variability than larger categories. For example, using two stimuli there are four possible configurations (two of which are functionally equivalent) for group assignment: group1-group1, group1-group2, group2-group1, and group2-group2. It's likely that participants will use all these categorization types because it's tractable for human learning. Conversely, it's impossible for participants to categorize 8 stimuli using every possible configuration, especially given the number of trials. Therefore, with respect to the total number of possible categorizations, diversity will be lower. In summary, this metric is not particularly appropriate for the current experiment given how the categories are constructed and the categorization possibilities that entails.

The second metric is straightforward; the classification that occurs most frequently in the participants' data for a given dataset corresponds to the most intuitive classification or "considered more obvious" (Pothos, et al., 2008). Although diversity will not be used, the correlation between these two metrics was  $r = -0.76$  (when diversity was unstandardized) and  $r = 0.76$  (when standardized according to maximum possible categorizations). For the unstandardized metric, as categorization strategy diversity increased, the frequency of the most used categorization strategy decreased. This makes

sense; as the number of possible categorization strategies increased, fewer participants selected the same strategy. With respect to the standardized metric, as the number of possible categorizations increased based on the number of stimuli, it's less likely participants will respond according to the whole spectrum of categorization strategies. Thus, with respect to the maximum limit, participants focused on a smaller set of categorization strategies rather than diversifying their responses. As mentioned, model analysis used the most frequent classification as an evaluation metric rather than diversity.

Participants tended to categorize the auditory stimuli using only two groups rather multiple groups even when exposed to categories with a higher number of stimuli. In fact, across every possible set of category stimuli, participants exclusively created two categories based on the stimuli rather than multiple categories. Table 7 presents the most frequent categorization strategy for each category and category type – presented in Boolean form - along with the frequency for that particular categorization strategy. No category existed for which participants consistently elected to use a multi-group strategy; multi-group strategies account for less than 30% of the possible solutions for each category.

Also noteworthy is the method in which participants created the categories; in other words, which dimension or dimensions of the auditory stimuli were used to partition the stimuli. Across each category type, it appears as though the criteria for developing a category boundary changed. More specifically, as the context (dimensional composition and number of stimuli) changed, participants constantly re-assessed their

conceptualization of the stimuli and this was reflected in spontaneous categorization of the audio (Table 7). When the category contained a small set of stimuli, such as the 3(2) category, participants employed a categorization strategy using the relevant dimension to separate groups of stimuli. In the case of 3(2), timbre was – across all types – the consistently different dimensions between stimuli, and its likely participants used this dimension to determine category membership. In other small categories such as 3(3) and 3(4), participants continued to use uni-dimensional strategies of categorization, whether the diagnostic dimension be spectral envelope, loudness, or frequency. These results are consistent with those found by other wherein individuals tend to use one dimension in order to make a decision for categorization (Goudbeek, et al., 2009; Gygi, et al., 2007).

In contrast, when participants engaged in the free-sorting task and multiple stimuli were available, the categorization strategy participants used was more complex than a simple uni-dimensional split. For example, category types for 3(5) and 3(6) show no clear use of a single diagnostic dimension to separate the stimuli. Rather than use only the spectral envelope, loudness, or frequency, participants used a combination of these dimensions to partition the stimuli. For example, in the 3(6) category, it appears that participants grouped stimuli according to a similar frequency but also used the given loudness and spectral envelope of the stimuli to make their categorization.

Participants could have easily used a single dimension for any category type to partition the set of stimuli but instead chose more complex categorization strategies. This result hints that individuals are employing more than a simple A/B comparison of a single

dimension and, while biased towards using only two categories, they are not necessarily minimizing differences between stimulus dimensions.

The integrality of the auditory stimuli may in part be responsible for the complex categorization strategies; participants may have been processing the stimuli as a whole and distributing their attention across all dimensions of the stimuli (Little, Nosofsky, & Denton, 2011; Little, et al., 2013). By doing so, new categorization strategies opened up, so to speak; rather than simply dividing stimuli into two categories based on a category-relevant dimension, participants would now be engaged in determining the relative similarity between possible within-category stimuli. In doing so, some categorization strategies more so resemble the ideal categorization for the 3(4)-2 category type where individuals must learn an “exclusive or” rule (e.g., high frequency and low amplitude OR low frequency and high amplitude). Thus, individuals are employing both similarity and discrimination processes when creating category clusters, as opposed to simply unidimensional discrimination. To determine the possible cognitive processes used by the participants in the free-sorting tasks, we used three current mathematical models to evaluate their performance on the task and their plausibility (with respect to the mechanisms and theories on which they are based) according to their performance.

**SUSTAIN.** The SUSTAIN model has been previously used to examine unsupervised learning tasks (Gureckis & Love, 2002; Love, Medin, & Gureckis, 2004) and, of particular interest, free-sorting tasks (Pothos, et al., 2008). In examining the SUSTAIN model, Pothos and colleagues (2008) used specific parameters and assumptions in order to approximate the free-sorting task within SUSTAIN, which

typically functions trial-by-trial. One assumption made by Pothos and colleagues was “that subjects consider each stimulus one at time but that the order of item consideration is idiosyncratic.” Thus, even though SUSTAIN examines stimuli trial-by-trial, it’s assumed the random presentation is equivalent to how participants examined the audio stimuli.

Similar to the participants in the experiment, SUSTAIN was given 4 trials of training per category type with a random ordering of the stimuli. SUSTAIN takes the vector of each Boolean stimulus representation as input. Because participants were told to categorize using any dimension or dimensions, SUSTAIN attention parameters for the three dimensions was set at an initial value of  $\lambda = 1.0$  and could be adjusted per trial by SUSTAIN during learning. This assumes SUSTAIN has equal initial attention to every dimension of the stimuli. Other parameters – attentional focus, clustering competition, decision consistency, and learning rate – used the initial unsupervised learning values recommended by Love (2004). Additionally, the threshold parameter, which sets the decision criteria to recruit new clusters, took on values sampled from a normal distribution of values per trial in order to simulate variation in participant stimulus categorization (Love, et al., 2004; Pothos, et al., 2008). To the extent that SUSTAIN was able to replicate the observed participant categorization for each category type, the probability of SUSTAIN developing that solution was multiplied by 156 to obtain the equivalent frequency of the most popular response metric (Pothos, et al., 2008).

Overall, SUSTAIN was unable to develop categories resembling the ones created by participants (Figure 20). Only when categories were simple such as 3(2) would

SUSTAIN's categories - and the associated probabilities of obtaining that category - resembled the participants' categories. On the 3(2) categories, the dissimilarity between the two objects increased by type such that SUSTAIN computed these category separations. Beyond the 3(2) category, the most frequent or relevant categories created by SUSTAIN in the 4 trials would often fail to mimic participant behavior. The only exception to this appears to be the 3(4)-2 category type which, given the pattern of SUSTAIN's clustering solution, seems to correspond to the participant data by pure chance. Table 8 presents the most frequent – and relevant – SUSTAIN clustering solution per category type. As discussed, whenever SUSTAIN would arrive at a two cluster solution, it was often incorrect with respect to the participants.

The clustering behavior obtained by SUSTAIN indicates a difference from previous results. Using stimuli that varied according to two dimensions, Pothos and colleagues (2008) found that SUSTAIN adequately captured the category structure obtained from participants. Each category had different clusters based on these two dimensions and categories contained different number of clusters, from two clusters to five clusters to an ambiguous clustering. Frequently, SUSTAIN clustering solutions would correspond to those of participants, particularly when the category contained a low number of anticipated clusters. SUSTAIN clustering solutions, however, often predicted the frequency of obtaining that clustering solution to occur less frequently relative to the data. It is perhaps the simplicity of the stimuli in previous experiments that contributed to SUSTAIN's unsupervised learning performance; in the current experiment, the stimuli consisted of three dimensions rather than just two. The increase in dimensional

complexity resulted in SUSTAIN's unsupervised learning underperformance; an increase in stimulus dimensionality also increase in the number of possible ways to cluster the data. Because the clustering solutions obtained by participants were all two cluster solutions and lack the complexity of multi-cluster solutions, these clustering patterns should have occurred more frequently (if only just occurring) within the SUSTAIN model should there be no difference between the current study and the study by Pothos and colleagues (2008). Specifically, because SUSTAIN performed fairly well with two cluster solutions in Pothos' study, it should follow that SUSTAIN should also perform well with the two cluster solutions in the current experiment. However, as mentioned, the increase in stimulus dimensionality obviously played an important role in the SUSTAIN model; even a simple dimensional increase can cause the model to perform poorly relative to previous performance.

**Simplicity model.** Unlike SUSTAIN, the simplicity model (Pothos and Chater, 2002; Pothos, et al., 2008) requires no parameter estimation or initial parameterization. The simplicity model, based off the Rosch and Mervis (1975) similarity proposal, examines total distances between the stimuli in the superset category and assesses the "savings" in clustering stimuli into smaller via a cost function. The model outputs savings in terms of absolute gain (number of bits saved through clustering) and relative gain (percentage of bits saved with respect to superset bit-size). According to the simplicity model, individuals are constantly attempting to minimize the code-length of the superset by partitioning it into smaller categories. Therefore, a smaller percentage value associated with the relative gain corresponds to the non-reducible amount of



remaining code-length (e.g., 40% means only 40% of the original 100% code-length remains in the given reduction via clustering). Smaller code-length corresponds to increased category intuitiveness and learnability (Pothos, et al., 2008; Pothos, et al., 2011).

There are two computational mechanisms for determining code-length and categorization in the simplicity model. Much like SUSTAIN, the simplicity model can attempt to predict the categorization in an unsupervised learning task by determining the best clustering - one that maximally minimizes the original superset code-length - via determining distances in a similarity matrix. In this manner, simplicity model initializes with a trivial number of clustering (each stimulus “assigned” to its own category) and then minimizes the number of the trivial clusters to decrease code-length. Using this mechanism, we can assess to what extent the simplicity model is capable of generatively creating clustering solutions that mirror those created by the participants in the current free-sorting task.

The other computational mechanism of the simplicity model allows the researcher to input a vector of numbers associated with the observed category labels and, rather than determine the best clustering, use these labels to compute any gains or loss in code-length according to the clustering. This computation allows for an analysis of the participants’ observed clustering code-length. Additionally, I can use the simplicity model’s best clustering algorithm to compare the “ideal” clustering compared to the observed data. These computational methods allow for model comparisons between the observed data in addition to SUSTAIN.

To address the first computational method of ideal clustering, the stimulus values for each category type were entered individually into a matrix and a trivial clustering solution was initiated prior to running the best clustering algorithm. The predicted per-category type clustering solutions determined by the simplicity model are in Table 9. Overall, the model was often successful in minimizing the number of initial clusters although the extent to which it accomplished this was often poor.

Examining the percentage change in the form of relative gains (Table 10) reveals that often the clustering solutions simplicity model *increased* the code-length of the categories when generatively constructing categories. The relative gain values computed by the simplicity model used the formula:

$$\frac{[\text{code-length to specify clusters}]}{[\text{code-length without clusters}]}*100$$

This ratio provides an indication of the bit number required to cluster categorical stimuli with respect to the bit number required to specify the category without clusters. Multiplying this value by 100 gives the relative percentage gains. Recall that the simplicity model attempts to minimize code-length, and therefore a smaller code-length to specify clusters relative to the code-length without clustering would specify that a particular clustering solution required less bits to construct than no clustering (reflected as a percentage lower than 100%). Pothos and colleagues (2008, 2011) have previously associated gains and losses in code-length to increases and decreases in category intuitiveness; a decrease in relative code-length indicates a more intuitive categorization because less learning resources are necessary to encode the stimuli. Conversely, a large

percentage of relative gains ( $> 100\%$ ) means that the code-length of the clustering solution decreased the learnability or intuitiveness of the category.

Such examples of increased relative gain are readily apparent in the current experiment. For example, the relative gains of category type (3)3-1 increased with clustering the three stimuli into different categories so that the final code-length for clustering the stimuli resulted in a less intuitive categorization structure. Increasing in gain occurred in four different category types: 3(3)-1, 3(3)-2, 3(3)-3, and 3(4)-4. For the 3(3)-1 types, it seems that the model is unable to minimize the clustering structure because the overall category has few stimuli. The generated clustering solution for 3(4)-4 makes sense within the context of the model. When creating a similarity matrix based on the categorical stimuli, the between-stimulus distance would be identical and as a result the simplicity model would guess that each of these stimuli belonged in independent clusters. Therefore, the simplicity model has difficulty predicting sub-100% code-lengths when the categories are small and the categorical stimuli are equidistant. The simplicity model minimized – to some extent – the code-length for all other category types other than these four types.

Although the simplicity model generated ideal clusters for minimizing code-length per category type, these clustering solutions frequently lacked correspondence with the observed data (Figure 21). The simplicity model accurately predicted clustering solutions when the number of stimuli per category was small (i.e., 3(2) category types) and when the possible categorization rule was distinct (i.e., 3(4)-2, exclusive or “XOR” rule). Other than these four category structures, the predicted clustering by the simplicity

model did not replicate the category clustering solutions performed by the participants and demonstrated a poor fit,  $r = 0.18$ . Using the computational mechanism of category generation appears to be inappropriate for the current experiment.

Previous studies by Pothos (2008, 2011) have shown that the simplicity model performance more than adequately on unsupervised learning tasks, but these studies differed from the current task in two key ways. First, the constructed categories contained more stimuli (16) than the categories in the current experiment, a fact that as previously mentioned may have caused the simplicity model to generate clusters with poor relative gains for some category types. Second, these experiment predefined the ideal clustering strategy prior to the experiment. Specifically, researchers created stimulus such that they would generate certain clusters at a categorical level (e.g., two clusters, two ambiguous cluster, and three clusters). The current experiment was a pure free-sorting task; there was no predefined ideal categorization strategy that influenced stimulus construction, so that it could be observed if individuals tend to use a particular categorization over another. Due in part to these differences in experimental design, the current results of simplicity model performance do not corroborate those from previous research.

Because the simplicity model also computes the absolute and relative gains for a pre-selected cluster solution, these were also examined (Table 10). These results are notable with respect to the best clustering solutions generated by the simplicity model. For the human data, the relative gains for each category type are almost always higher than the best clustering solutions (with the exception of 3(4)-2, where the clustering was

accurately predicted). According to the simplicity model, the most frequently used categorization solution made by individuals in the current experiment increased the code-length of the category, implying that individuals were creating more difficult categories than necessary.

The vast amount of literature on similarity, categorization, and concept learning does not support such an implication (Pothos, et al., 2008, 2011; Tversky & Gati, 1978, 1982), and this highlights a potential issue with the simplicity model. Namely, given that individuals are categorizing stimuli in a manner that might be counter-intuitive according to the simplicity model, the underlying cognitive mechanisms used by individuals must differ from those assumed by the simplicity model. Rather than generating these categories from perceived similarity according to a metric distance and minimizing the categorical expression so that it is tractable, individuals must be utilizing similarity and discrimination in a manner not employed by the simplicity model. Therefore, for the current unsupervised learning task, it appears that participants are using the cognitive mechanisms of similarity assessment, discrimination, and attention in a more complex manner than assumed by the simplicity model.

**Simplicity model vs. SUSTAIN.** The SUSTAIN and simplicity models have been compared previously on unsupervised learning tasks and found to both be approximately equivalent in terms of their generated categorical solutions and performance with respect to the human data (Pothos, et al., 2008; Pothos, et al., 2011). In their studies, Pothos and colleagues (2008, 2011) used two dimensional stimuli resembling insects that varied continuously according to body length (long or short) and

leg length (long or short). Researchers created a priori categories with the intention of specifying ideal categorizations that participants should create. For example, Pothos and colleagues created stimuli in a manner to create two distinct clusters as the ideal solution by assigning all stimuli of that group approximately the same value. In total, Pothos and colleagues created 9 different data sets contain different numbers of clusters and, when graphed, different proximities of the clusters in order to create ambiguity.

In comparing these two models, Pothos found that the simplicity model and SUSTAIN generated clustering solutions that successfully corresponded to some of the category structures human categorization fairly well (Pothos, et al., 2008; Pothos, et al., 2011). The simplicity model was able to capture categorization performance similar to that of the humans when the a priori categories had three groups or arranged in an ambiguous structure. The simplicity model also had a tendency to over-predict the response frequency for simple categorizations such as the two group solutions. Stimulus distances determine group inclusion in the simplicity model; for human participants engaging in similarity and discrimination of continuous dimension stimuli, the distances between stimuli are not as apparent. Due in part to this ambiguity, the responses by human participants are just not as frequent as estimated by the simplicity model (Pothos, et al., 2011). Also notable is the simplicity model's performance on categories where a large number of clusters are created; when the optimal category solution was five clusters, human participants easily discriminated categorical boundaries while the simplicity model grossly under-predicted the participant response.

In comparison, SUSTAIN generated an equivalent number of cluster solutions to the human data except when the categorization task required two clusters (Pothos, et al., 2011). Under these circumstances, SUSTAIN – much like the simplicity model – over-estimated the frequency of the most popular clustering solution when only two clusters were necessary to sort the stimuli. In this experiment, SUSTAIN did not drastically under-estimate the frequency of the most popular solution as the simplicity model did. However, the previous experiment by Pothos and colleagues (2008) found that while the simplicity model performed about the same as their more recent (2011) experiment, SUSTAIN would sometimes under-estimate the frequency of the solution and – at worst – fail to generate the category at all. Based on these extreme differences in results, it seems that SUSTAIN is highly variable in terms of performance; more often than not the model will generate the correct clustering solution, but if the model fails it does so noticeably (Pothos, et al., 2008).

In order to examine these models, category clustering occurred generatively based on the stimulus vectors of each category type. Figures 20 and 21 shows the empirical data and model predictions per category type. SUSTAIN provided a better fit ( $r = 0.68$ ) than the simplicity model ( $r = 0.18$ ). However, these values still indicate weak performance with respect to predicting human categorization in the current unsupervised learning task. This result may be due to previously mentioned reasons; the categories presented in the current experiment contain a smaller number of stimuli than those used in other experiments and the category clusters were not pre-defined (Pothos & Chater, 2002; Pothos, et al., 2008; Pothos, et al., 2011). Therefore the categories created by the

participants may have been intuitive to them but were not intuitive to the simplicity model and SUSTAIN.

Overall the simplicity model and SUSTAIN are both poor descriptors and predictors of the human data from the current unsupervised learning free-sorting task. Their performance, although similar between the models, frequently demonstrated situations in which they inaccurately attempted to replicate humans, and this speaks to the validity of their internal mechanisms and assumptions as representations of human conceptual and categorization behavior.

**Generalized structural invariance theory.** As previously discussed, the GISTM derives predictions of category learning from the inherent invariances present within a given category set. In the supervised literature, researchers can utilize GISTM to determine the perceived learning difficulty of a category and how invariant that category is to perturbations, and in such a capacity, it has demonstrated much success and potential future applications (Vigo, 2009, 2013, 2014). Additionally, current research has examined the structural manifolds with respect to whether they adhere to structural equilibrium or the parsimony principle (Chapter 4), specifically in choice behavior (Doan & Vigo, *under review*). I adopt such a procedure in the current analysis as one method of GISTM analysis. Therefore, in the current usage, GISTM will not provide an output like SUSTAIN or the simplicity model with respect to the frequency of most popular categorization strategy. Instead, GISTM will essentially tell us the quality of the categorizations with respect to the underlying principles of invariance theory.



An examination of the structural manifolds per category type reveals an interesting pattern (Table 11). It appears that – as proposed by Vigo (2014) – individuals do prefer constructing categories that reflect extreme structural kernels such that in the current experiment participants demonstrated tendency towards structural manifolds of (0, 0, 0) for both categories. In fact, the participant created categories clearly demonstrate both the principle of invariance-parsimony and the principle of structural equilibrium. These state that when categorizing stimuli, an individual will demonstrate a disproportionate emphasis on extreme structural kernels ('0' or '1') and by doing so participants are able to discriminate which dimensions are relevant or diagnostic to categorization (Vigo, 2014).

For categories such as 3(2), the presence of structural equilibrium is relatively trivial; if participants separated the two stimuli, both categories of a single stimulus would always have a structural manifold of (0, 0, 0). In fact, other than the 3(2)-1 category type, types 3(2)-2 and 3(2)-3 would have (0, 0, 0) structural manifolds even when grouping both stimuli together. These types are only indicative of a very simple categorization task that regardless of categorization strategy results in the same underlying structure. The except to this is the 3(2)-1 category type; rather than categorize these stimuli together such that the dimension of spectral envelope was invariant – and therefore not diagnostic under the invariance-parsimony principle – participants grouped each stimulus separately such that both single stimulus categories adhered to the

structural equilibrium principle<sup>6</sup>. When examining the more complex categories, participants' tendency towards structural equilibrium becomes apparent.

In the more complex categories where more than two audio stimuli are present, participant behavior still exhibits a preference towards extreme structural kernels and structural equilibrium. In these categories, most participants constructed categories such that at least one manifold from the two groups maintained structural equilibrium and, when possible, participants constructed categories wherein one structural kernel was completely invariant to all perturbations (e.g., 3(3)-2). Categories constructed containing single invariance kernels allowed individuals to disregard this dimensional redundancy and instead categorize using the other two dimensions. As a specific example, in the 3(3)-2 category participants created category groupings such that in one group the dimension of spectral envelope was invariant to dimensional perturbations. Participant may have perceived the underlying structural invariance of spectral envelope and detecting this redundancy allowed participants to redirect attention towards the other dimensions when categorizing. The other grouping participants created for the 3(3)-2 category shows a structural manifold of (0, 0, 0) and an adherence to structural equilibrium. When presented with these complex categories, the trend of participants to either create groups in which one structural kernel is invariant (an extreme SK) or create

---

<sup>6</sup> It should be noted that according to GISTM, the participants' behavior in separating 3(2)-1 into two categories such that the structural manifolds equal (0, 0, 0) is fully supported by both the mathematical foundations and process account of the GISTM. Both SUSTAIN and the simplicity model were unable to account for the participants' tendency to engage in this behavior.

groups which exhibit structural equilibrium is apparent. Such a result is consistent with the principles of parsimony and structural equilibrium put forth by Vigo (2014).

Some unique structural manifolds are apparent in Table 11) such as manifolds in which the structural kernels are not at either extreme (e.g., 0.67). For example, 3(5)-1 demonstrates a group one manifold of (0, 0.67, 0.67) and a group two manifold of (0, 0, 0). While the structural manifold of the first group is not optimal according to either parsimony or structural equilibrium, taken in context it appears that individuals are sacrificing structural equilibrium and dimensional diagnosticity in order to attain structural equilibrium for the second category. In a way, these categories containing structural kernels that fall between the extreme 0 and 1 values are acting as conceptual “junk draws”; categories that have little in terms of cohesion or invariance but allow other constructed categories to maintain some form of structure with respect to extreme SKs. Categories 3(6)-1 and 3(6)-3 also show this pattern; it appears that it is easier for participants to develop and maintain a memory trace in which all dimensions of a category are in structural equilibrium.

From qualitatively assessing the structural manifolds of the most frequently used categorization strategy across each category type, it is very apparent that participants categorizing by using the underlying structure between stimuli to inform category sorting. Specifically individuals are using the principles of invariance-parsimony and structural equilibrium almost exclusively during the free-sorting task when encoding the category structure as an ideotype. The degree of structural equilibrium informs individuals of the ease in which they can discriminate between stimuli and when a category is in perfect

structural equilibrium, all dimensions provide relevant or diagnostic information. In the current experiment, stimuli grouped together and resulting in a high degree of structural equilibrium indicates that participants perceived the grouped stimuli to be highly discriminable to participants because the participants are attending to all dimensions. This results marks a departure from previous results in which participants created categories according to uni-dimensional criteria (Goudbeek, et al., 2009; Gygi, et al., 2007). Instead, participants are attending to all three dimensions in order to easily discriminate between stimuli.

The second GISTM measure examined in the current experiment is the perceived category difficulty of the each grouping created by the participants. As mentioned, the category difficult can be assessed by examining the distance between each category type's difficulty – psychologically manifested as the ideotype memory trace - and the zero ideotype in psychological space. As the distance between the ideotype and the zero ideotype decreases, a category becomes difficult to learn and subsequently more difficult for an individual to develop the concept of the category. Table 12 contains the difficulty values for each category type. Consistently across each category type, participants tended to create categories such that at least one of the two categories represented an ideotype furthest away from the zero ideotype. For example, participants' groupings on the 3(3)-1 category type represent such a situation in which one structural manifold lacks invariance across every dimension whereas the other structural manifold demonstrates a single invariant SK. Within the context of the current measure of category difficult, it appears that individuals are creating one, less homogenous category (that in turn doesn't

adhere to the notion of structural equilibrium) in order to create one category that demonstrates a relatively homogenous quality (Vigo, 2014). Categories with a high degree of difficulty are also more discriminable. While stimuli may lack within-category similarity and dimensional redundancy, participants seem to prefer creating categories that are highly discriminable across all stimulus dimensions.

The third measure used in the current experiment from the GISTM model examines the categorical invariance of each free-sorted category. Squaring the value of phi accentuates the contribution of the extreme SKs while diminishing the contribution of intermediary SKs. For the current analysis, the discrimination index parameter  $k = 1$  so that there was no preference given to an expanded or diminished psychological space per category type. The resulting invariance values correspond to the results of the first and second measures. By accentuating the extreme values, the categorical invariance of each category structure shows a gravitation towards extremes rather than a tendency to remain firmly within the interval. Based on the measure of categorical invariance, although some structural manifolds appeared to be almost squarely within the SK interval, the invariance from the structural manifolds shows that participants are still preferring extremes even when a structural manifold in perfect equilibrium is not possible. Participants inherently biased towards creating categories that are highly discriminable or as discriminable as they can make them given constraints.

Based on these three measures, GISTM presents a different and useful theoretical alternative to that of SUSTAIN and the simplicity model. In the current implementation, rather than fruitlessly attempt to predict or mimic the participants' categorization

strategies, GISTM uses its flexible mathematical framework to determine the underlying relationships between the stimuli of each category (Vigo, 2013, 2014). In doing so, GISTM allows researchers to determine how participants should categorize based on these underlying patterns and relationships. It may be that the exact structure of the free-sorted categories is un-important but that the stimulus relationships within group are what establish functional equivalence between free-sort responses. As a concrete example, imagine there are four stimuli with varying dimensions; SUSTAIN and the simplicity model would attempt to exactly predict how an individual would group these objects and they are therefore trying to determine the exact structure of each category. However, analyzing this scenario through the lens of invariance and the GISTM, there may be several combinations of stimulus groupings that satisfy structural equilibrium and parsimony. Grouping stimulus 1 and 2 in our example may satisfy these principles, but grouping stimulus 1 and 3 may also do this. In this respect, these two categories are functionally equivalent from an underlying structural standpoint; specifically, both groupings have the same SKs and structural manifold (granted, a participants' exact response would depend on whether the manifold of the second grouping changed based on these changes). The only reason an individual may select one categorization over another would depend on their ability to perceive, detect, and attend to particular patterns. One individual may be more sensitive to the grouping of stimulus 1 and 2 because there's a particular salient dimension that the individual's attention is biased or directed towards. Thus, it appears that GISTM provides the most relevant account of human unsupervised

learning, specifically through the mechanisms of pattern detection, perceived difficulty, and the capacity of individuals to discriminate stimulus dimensions.

## Chapter 11: Discussion

### General Discussion

Previous experiments in the human auditory categorization literature demonstrate that individuals typically attend to three dimensions when performing similarity assessment (Alrich, Hellier, & Edworthy, 2008; Bonebright, 2001; Gygi, et al., 2009; Howard & Silverman, 1976; McAdams, et al., 1995; Samson, Zatorre, & Ramsay, 1997). Result such as these are in correspondence to studies on limits of auditory attention; individuals are limited to approximately 3 to 4 auditory attributes or dimensions when attending to an auditory stimulus (Chen & Cowan, 2005; Cowan, Chen, & Rouder, 2004; Sauls & Cowan, 2009; Tulving & Patkau, 1962). The current experiment also had a similar result; when individuals were engaged in the similarity assessment task, they generally used three dimensions to make establish their assessment of similarity between the presented pair of auditory stimuli. From this result, participants reported using frequency/pitch, amplitude/loudness, and the spectral envelope as dimensions for comparison. These results make sense with respect to the literature within the area; the reported number of dimensions corroborate previous studies (Caclin, et al., 2005; Grey, 1977; Howard & Silverman, 1976; Krumshansl & Iverson, 1992; McAdams, et al., 1995; Melara & Mark, 1990; Samson, Zatorre, & Ramsay, 1997). Specifically, at most participants could only attend to approximately three dimensions when determining similarity between the stimuli.

These three dimensions were used to construct stimuli and examine unsupervised learning of Boolean categories (Feldman, 2000). When constructing categories for each



set of stimuli, participants preferred to partition the stimuli into two categories rather than multiple categories. Furthermore, participants often used multiple stimulus dimensions to determine categorization between these two groups, a result that differs from previous findings that individuals categorize uni-dimensionally (Goudbeek, et al., 2009; Gygi, et al., 2007). The participant constructed categories were then examined according to three current mathematical models in order to determine validity of these models with respect to the current unsupervised learning task and - to the extent that these models do approximate the data - to examine possible cognitive processes that are occurring when participants engage in the task.

Of the three models tests – SUSTAIN, the simplicity model, and GISTM – only GISTM explained the human categorization results in a manner completely consistent with the theory and associated principles (Vigo, 2013, 2014). The results of the simplicity model and SUSTAIN demonstrate that code-length reduction/minimization and exemplar based accounts of clustering and categorization are not capable of accurately representing human performance in the current free-sorting task. These are as a result not plausible explanations for behavior. Although the simplicity model and SUSTAIN may demonstrate the capacity to account for human behavior in unsupervised learning tasks involving a large number of stimuli (Pothos, et al., 2008, 2011), the failure to account for the data in the current experiment supposes that individuals are not engaging the task and using their cognitive abilities in the manner presupposed by SUSTAIN and the simplicity model. Therefore, the answer for how individuals develop

concepts and categorize stimuli appears to lie in the notions of invariance, structural equilibrium, and parsimony (Vigo & Doan, 2015).

### **Limitations and Future Directions**

The current experiment had some limitations that could be addressed in future experiments to expand the generalizability and strength of the current results.

Concerning the first experiment in which we conducted MDS, the results from the scree plot and other measures hinted at a three dimensional solution but this was rather ambiguous at best. This of course was due to the participant's similarity responses per pairwise stimulus comparison. One clear direction might resolve this ambiguity.

Although the dimensional values for each stimulus came directly from previous research (Caclin, et al., 2005; Grey, 1977; Krumshansl & Iverson, 1992; Melara & Mark, 1990) it might be that some dimensional values we're particularly salient to participants with no-to-low musical experience compared to those with a year or more of experience. As mentioned, there are documented differences between musicians and non-musicians (Tervaniemi, et al., 2005) and it may be that non-musicians weren't sensitive enough to discriminate dimensional changes. To address this, future studies may want to increase the distance between auditory stimulus dimensions to possibly further influence discriminability.

Another potential reason for the ambiguity in the MDS may relate to potential individual differences in the MDS dimensional solutions. The composition of the participant sample included some individuals who possessed musical ability and, though their years of experience were below previously established thresholds (Tervaniemi, et

al., 2005), this may have influenced the number of dimensions they attended and stored in memory. Individuals with musical experience may have been attending to more dimensions than those lacking in musical training and, accordingly, the MDS reflected these differences in dimensional attention in the form of an ambiguous dimensional solution. For example, suppose the individuals with musical training used four dimensions whereas individuals with no musical training used two dimensions; because MDS averages similarity judgments across participants, this distinction may have been lost. Future studies should examine these individual differences as an additional metric for determining the correct MDS dimensional solution. Specifically, if the results is ambiguous yet the majority of individuals use a particular dimensional solution, the most popular dimensional solution would provide more evidence towards a correct dimensional solution.

Additionally, different methods of altering the spectral centroid could allow researchers to determine whether these influence or changes human similarity judgments. In the current MDS experiment, I altered the spectral centroid using a spectral band-pass filter with a 12 dB slope centered on a particular harmonic unit of the audio source. An alternative approach to this would be to use additive synthesis to have greater control over each partial within the tone. This would allow the attenuation or reduction of specific partial harmonics by removing individual sine waves in the additive processes, allowing for the accentuation of only odd or even harmonics in the resulting tone. If we wanted to continue examining physical models, recent modifications of the Karplus-

Strong algorithm could potentially provide an increased level of control over the spectral components of the audio (Karjalainen, et al., 1998; Sullivan, 1990).

In the unsupervised learning task, future studies might address several limitations. First, time allowing, I could collect more data from participants across several days. In doing so, participants might eventually settle into a specific categorization strategy per category type. Over time, individuals may demonstrate less categorization response variability and show a stronger preference for a particular strategy in the form of frequency of the most popular strategy. It would then be necessary to further test the generalizability of the most frequent clustering solution across sample. As a first step, observing whether individuals focus on a particular strategy could provide more insight into the cognitive processes used during the unsupervised learning experiment.

Third, I did not test all possible instances of each category type. By re-assigning the dimensional values of the stimuli, new relationships and patterns may form. For example, under the current experiment the Boolean stimuli of 011 would represent a low frequency, a loud amplitude, and a woodwind timbre. Dimensional reassignment could create a different stimulus using the same Boolean code; the stimulus 011 in a different instance of a category type could refer to a sound with a low amplitude, a high frequency, and a woodwind timbre. Re-coding the stimuli and testing individuals on these categories would allow future researchers to establish whether the perception and attention to auditory dimensions is driving categorization or if the underlying structure – which is preserved – influences categorization. To provide a specific example: imagine we are studying a category containing the objects 000 + 001 + 011 and have assigned the

dimensions sequentially, thus: frequency, amplitude, and timbre. Under this configuration, perhaps participants develop a categorization based on timbre such that the groups are 000 | 001 + 011. After dimensional reassignment, would individuals still direct attention to the timbre dimensions or would they be able to re-direct attention to the new dimensions occupying the Boolean value that timbre currently does. That is, re-assigning dimensions allows researchers to determine the extent to which individuals are learning patterns or engaging in stimulus (dimensional) directed attention (Feldman, 2000; Vigo, 2009).

In future experiments, I can broaden the research goals to examine different aspects of unsupervised learning while maintaining a high level of experimental control. One possible direction is to examine multi-valued dimensional stimuli rather than simply examine binary values. Stimulus dimensions rarely take on discrete binary values and it's more common in daily life to experience stimuli that vary according to a continuous dimension. The introduction of multi-dimensional stimuli need not be associated with an overwhelming number of values; we can gradually introduce multi-valued dimension such that they can take on four values rather than two similar to previous experiments by Vigo (2013, 2014). By extending the number of values a dimension can take on, we are also getting closer to experiments examining natural sounds (such as Gygi, et al., 2007) but with rigorous pre-defined and generated stimuli. This can allow us to remove any noise or influence from unintended stimulus dimensions that are often an integral part of field recorded stimuli (e.g., spatialization).

Future studies can also further examine GISTM (Vigo, 2013, 2014) with respect to different unsupervised learning tasks. In current experiment, the GISTM qualitatively predicted how individuals would group stimuli according to the principles of parsimony, structural equilibrium, and invariance. There is much potential in using the GISTM as it has been found to provide excellent fits of the human supervised learning performance and choice behavior (Vigo, 2009, 2013, 2014; Vigo & Doan, 2015). However the model is relatively untested for unsupervised learning, and a good first step might be a replication of experiment performed by Pothos and colleagues (Pothos, et al., 2008, 2011) in order to continue establishing the validity of the GISTM as an accurate account of human concept learning and categorization for unsupervised learning.

Concerning the models of unsupervised learning, a unique direction would be to modify the code associated with them to include machine listening components to more closely mimic the human experience. Specifically, rather than using a vector of stimulus dimensions, the audio can be played for the program at which time auditory feature extractors and detectors (Hinton et al., 2012; Liu, Wang, & Chen, 1998) can attempt to determine the stimulus dimensions. With the dimensions then estimated by the feature extractors, the models then can compute associated similarity values and distances and other metrics. Applying these machine listening mechanisms for unsupervised learning could allow for the extended models to be applied to different domains including autonomous sensing agents (Magee, et al., 2004) and recommendation algorithms for auditory related programs and applications (Bu, et al., 2010; Huang & Jenor, 2004).

Finally, we can examine whether changes in categorization occurred according to different groups. Specifically, to date no researchers have examined difference in categorization and unsupervised learning between musicians and non-musicians. Only research examining differences between these two groups according to perception and discrimination of single and multiple dimensional auditory stimuli has been examined (Tervaniemi, et al., 2005). Examining differences between these groups could provide insight into the differences between experts and novices and help to uncover any cognitive processes that differ between these groups. Lastly, we can further investigate whether age and age-related perception of these stimuli influences categorization strategies of individuals from different age groups.

Table 1

*Auditory Values*

---

<u>Dimension</u>	<u>0</u>	<u>1</u>
Frequency (Hz)	370	523
Amplitude (dB)	64	70
Spectral Structure	String	Clarinet
Spectral Centroid (Harmonic)	-	3
Log-Attack Time (ms)	15	200

---



Table 2

*Auditory Stimulus Boolean Values*

	<u>Freq</u>	<u>Amp</u>	<u>SpecEnv</u>	<u>SpecCent</u>	<u>Log-Attack</u>
1	0	0	0	0	0
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	1	0
6	0	0	0	0	1
7	1	1	0	0	0
8	1	0	1	0	0
9	1	0	0	1	0
10	1	0	0	0	1
11	0	1	1	0	0
12	0	1	0	1	0
13	0	1	0	0	1
14	0	0	1	1	0
15	0	0	1	0	1
16	0	0	0	1	1
17	1	1	1	0	0
18	1	1	0	1	0
19	1	1	0	0	1
20	1	0	1	1	0
21	1	0	1	0	1
22	1	0	0	1	1
23	0	1	1	1	0
24	0	1	1	0	1
25	0	1	0	1	1
26	0	0	1	1	1
27	1	1	1	1	0
28	1	1	1	0	1
29	1	1	0	1	1
30	1	0	1	1	1
31	0	1	1	1	1
32	1	1	1	1	1

Table 3

*Auditory Stimuli According to Boolean Values*

	<u>Freq (Hz)</u>	<u>Amp (dB)</u>	<u>SpecEnv</u>	<u>SpecCent</u>	<u>Log-Attack</u>
1	370	64	Strings	-	15
2	523	64	Strings	-	15
3	370	70	Strings	-	15
4	370	64	Clarinet	-	15
5	370	64	Strings	3	15
6	370	64	Strings	-	200
7	523	70	Strings	-	15
8	523	64	Clarinet	-	15
9	523	64	Strings	3	15
10	523	64	Strings	-	200
11	370	70	Clarinet	-	15
12	370	70	Strings	3	15
13	370	70	Strings	-	200
14	370	64	Clarinet	3	15
15	370	64	Clarinet	-	200
16	370	64	Strings	3	200
17	523	70	Clarinet	-	15
18	523	70	Strings	3	15
19	523	70	Strings	-	200
20	523	64	Clarinet	3	15
21	523	64	Clarinet	-	200
22	523	64	Strings	3	200
23	370	70	Clarinet	3	15
24	370	70	Clarinet	-	200
25	370	70	Strings	3	200
26	370	64	Clarinet	3	200
27	523	70	Clarinet	3	15
28	523	70	Clarinet	-	200
29	523	70	Strings	3	200
30	523	64	Clarinet	3	200
31	370	70	Clarinet	3	200
32	523	70	Clarinet	3	200

Table 4

*Stress per NMDS Model*

	<u>k = 1</u>	<u>k = 2</u>	<u>k = 3</u>	<u>k = 4</u>	<u>k = 5</u>
Stress (proportion)	0.101	0.051	0.034	0.022	0.016

Table 5

*Participant ratings for each dimension*

	<u>Mean</u>	<u>Median</u>	<u>SD</u>
Pitch	8.5	8.5	1.39
Loudness	6.9	8	2.75
Spectral Envelope	6.4	7	3.58
Spectral Centroid	4.9	5	3.14
Log-Attack Time	4.0	4.5	2.73

Table 6

*Pairwise tests of audio dimensions.*

	<u>Pitch</u>	<u>Loudness</u>	<u>SpecEnv</u>	<u>SpecCent</u>	<u>Log-Attack</u>
Pitch					
Loudness	0.2882				
SpecEnv	0.3731	1			
SpecCent	3.36E-05*	0.0990	0.4794		
Log-Attack	3.69E-08*	0.0019*	0.0614	1	

\* $p < 0.05$

Table 7

*Participant Groupings per Category Type*

<u>Category</u>	<u>Type</u>	<u>Frequency</u>	<u>Group 1</u>	<u>Group 2</u>
3(2)	1	130	000	001
	2	130	000	011
	3	154	000	111
3(3)	1	63	000 + 010	001
	2	61	000 + 001	110
	3	69	000 + 011	101
3(4)	1	49	000 + 010 + 100	001
	2	102	000 + 001	110 + 101
	3	100	000 + 001	010 + 101
	4	50	000 + 001 + 100	010
	5	99	000 + 001	111 + 010
	6	99	000 + 110	101 + 011
3(5)	1	45	111 + 110 + 100	101 + 011
	2	48	111 + 010 + 100	101 + 011
	3	47	111 + 101	110 + 001 + 100
3(6)	1	34	111 + 011 + 100 + 010	110 + 101
	2	36	111 + 001 + 100 + 010	110 + 101
	3	34	010 + 011 + 100 + 001	110 + 101

Table 8

*SUSTAIN Categorization Predictions.*

<u>Category</u>	<u>Type</u>	<u>Frequency</u>	<u>Group 1</u>	<u>Group 2</u>
3(2)	1	0	000 + 001	
	2	78	000	011
	3	156	000	111
3(3)	1	0	000 + 010 + 001	
	2	0	001	000 + 110
	3	0	000 + 101	011
3(4)	1	0	000 + 010	011 + 001
	2	78	110 + 111	001 + 000
	3	0	000 + 010	001 + 101
	4	0	000 + 010 + 001 + 100	
	5	0	000 + 010 + 100	111
	6	0	011 + 110	101 + 000
3(5)	1	0	111 + 100 + 101 + 110	011
	2	0	111 + 011 + 010	101 + 100
	3	0	111 + 101 + 001 + 100	110
3(6)	1	0	010 + 011 + 110 + 111	100 + 101
	2	0	010 + 001 + 100 + 110 + 101	
	3	0	110 + 101 + 011	010 + 100 + 001

Table 9

*Simplicity Model Predictions.*

<u>Category</u>	<u>Type</u>	<u>Group 1</u>	<u>Group 2</u>	<u>Group 3</u>	<u>Group 4</u>
3(2)	1	000	001		
	2	000	011		
	3	000	111		
3(3)	1	000	010	001	
	2	000	001	110	
	3	000	101	011	
3(4)	1	000 + 010	001 + 011		
	2	000 + 001	110 + 111		
	3	000 + 001	010 + 101		
	4	000	001	100	010
	5	000 + 010 + 001	111		
	6	000 + 101 + 011	110		
3(5)	1	111 + 101 + 110 + 100	011		
	2	111 + 101 + 100	010 + 011		
	3	111 + 101 + 110 + 100	001		
3(6)	1	010 + 011	100 + 110 + 101 + 111		
	2	010 + 100 + 110	001 + 101 + 111		
	3	010 + 011	100 + 110	101 + 001	

Table 10

*Simplicity Model Computation.*

<u>Category</u>	<u>Type</u>	<u>Abs.Gain</u> <u>(Human)</u>	<u>Rel.Gain</u> <u>(Human)</u>	<u>Abs.Gain</u> <u>(SM)</u>	<u>Rel.Gain</u> <u>(SM)</u>
3(2)	1	1	Inf	1	Inf
	2	1	Inf	1	Inf
	3	1	Inf	1	Inf
3(3)	1	6	191.8	5	152.8
	2	6	191.8	5	152.8
	3	6	191.8	5	152.8
3(4)	1	19	128.7	15	99.8
	2	15	99.8	15	99.8
	3	21	140.7	15	99.8
	4	17	115.3	17	113.3
	5	18	119.8	14	64.2
	6	15	99.8	14	64.2
3(5)	1	49	108.7	40	88.8
	2	55	122.5	36	81.0
	3	52	116.0	40	88.8
3(6)	1	115	109.6	81	77.0
	2	125	119.5	83	78.7
	3	113	107.9	83	79.3

Table 11

*Structural Manifolds for Each Type as Computed by GISTM.*







<u>Category</u>	<u>Type</u>	<u>Group 1 Manifold</u>	<u>Group 2 Manifold</u>
3(2)	1	(0, 0, 0)	(0, 0, 0)
	2	(0, 0, 0)	(0, 0, 0)
	3	(0, 0, 0)	(0, 0, 0)
3(3)	1	(0, 1, 0)	(0, 0, 0)
	2	(0, 0, 1)	(0, 0, 0)
	3	(0, 0, 0)	(0, 0, 0)
3(4)	1	(0, 2/3, 2/3)	(0, 0, 0)
	2	(0, 0, 1)	(0, 1, 1)
	3	(0, 0, 1)	(0, 0, 0)
	4	(2/3, 0, 2/3)	(0, 0, 0)
	5	(0, 0, 1)	(0, 0, 0)
	6	(0, 0, 0)	(0, 0, 0)
3(5)	1	(0, 2/3, 2/3)	(0, 0, 0)
	2	(0, 0, 0)	(0, 0, 0)
	3	(0, 1, 0)	(0, 2/3, 0)
3(6)	1	(1/2, 0, 1/2)	(0, 0, 0)
	2	(0, 0, 0)	(0, 0, 0)
	3	(0, 1/2, 1/2)	(0, 0, 0)



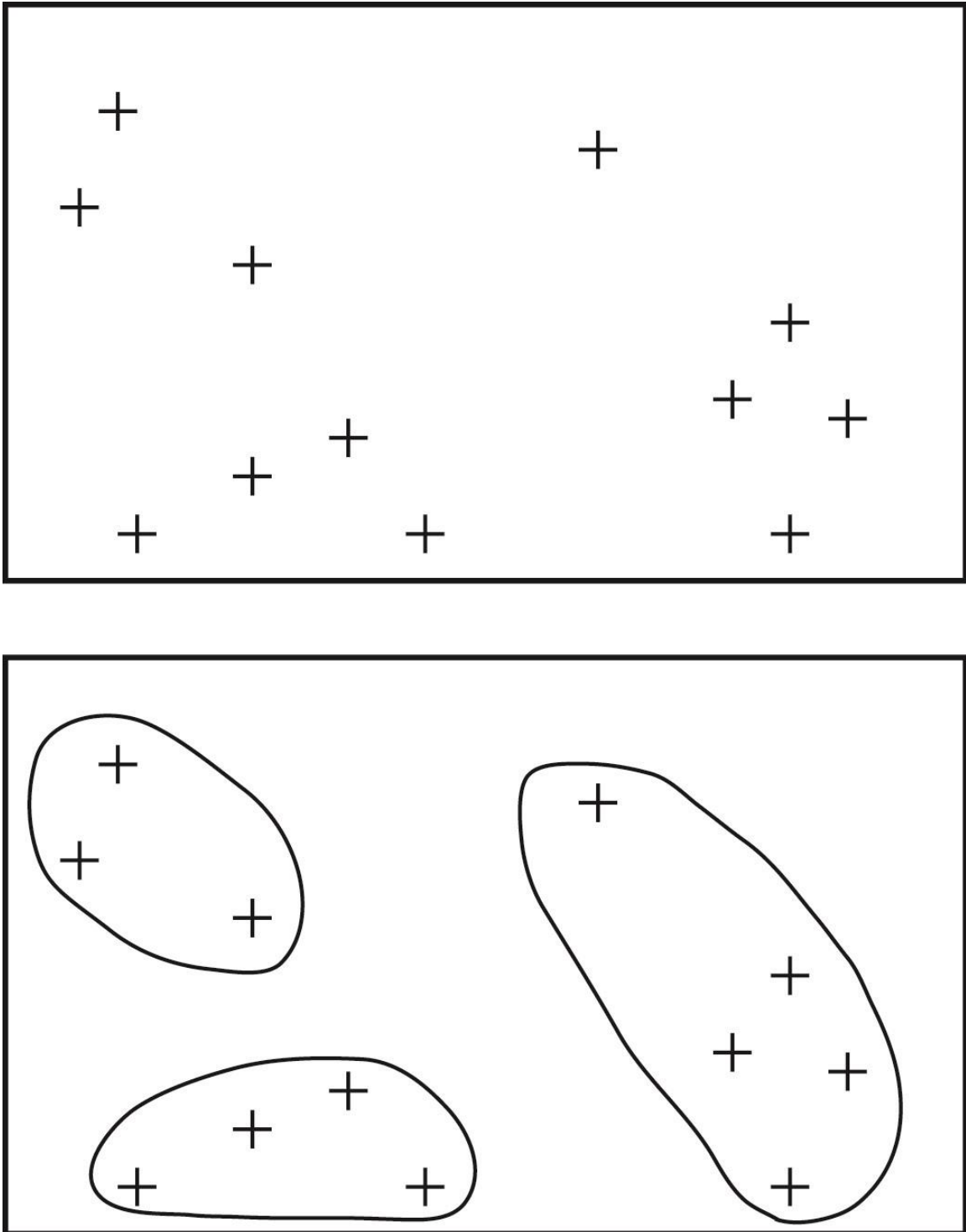
Table 12

*Difficulty and Invariance Values per Category..*

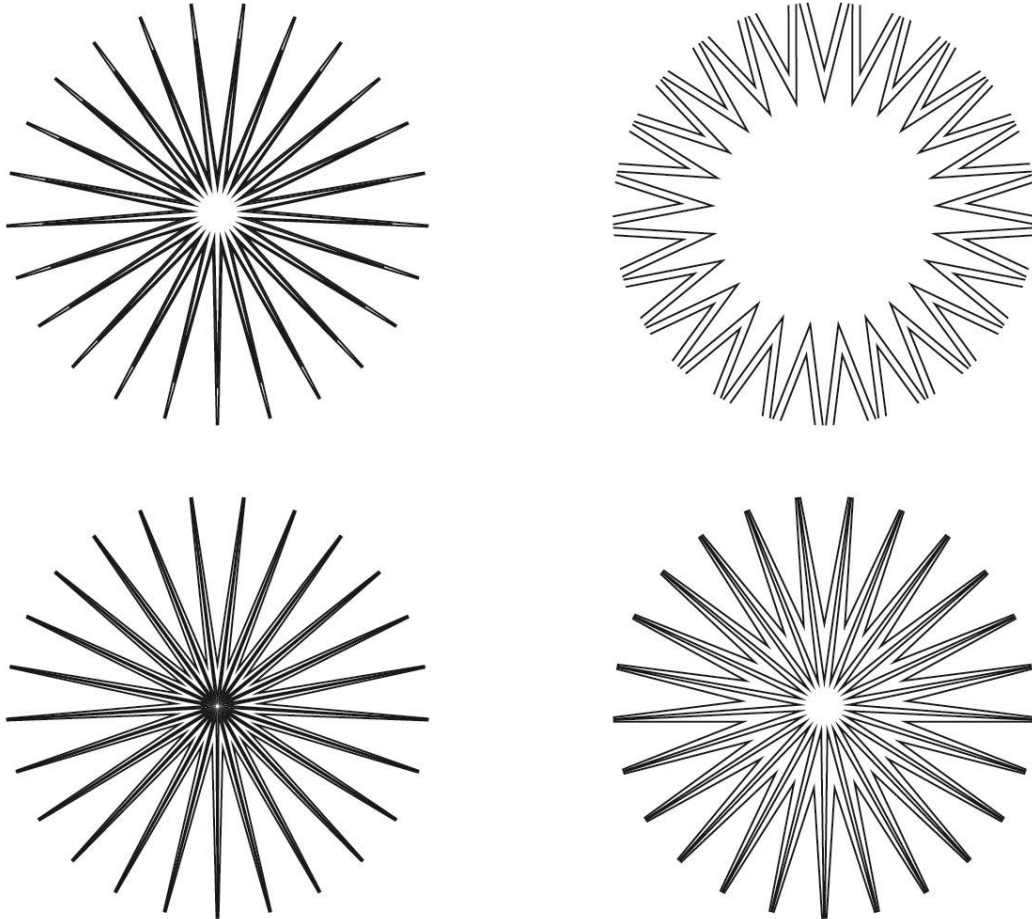
<u>Category</u>	<u>Type</u>	<u>Group 1</u>		<u>Group 2</u>	
		<u>Difficulty</u>	<u>Invariance</u>	<u>Difficulty</u>	<u>Invariance</u>
3[2]	1	1	1	1	1
	2	1	1	1	1
	3	1	1	1	1
3[3]	1	0.73	0.27	1	1
	2	0.73	0.27	1	1
	3	2	2	1	1
3[4]	1	1.17	0.45	1	1
	2	0.73	0.27	0.74	0.27
	3	0.73	0.27	1	1
	4	1.17	0.45	2	2
	5	0.73	0.27	2	2
	6	2	2	2	2
3[5]	1	1.17	0.45	2	2
	2	3	3	2	2
	3	0.73	0.27	1.54	0.79
3[6]	1	1.97	0.97	2	2
	2	4	4	2	2
	3	1.97	0.97	2	2

<b>3[4]</b>	<b>Example Structure</b>
<b>Type I</b> $3[4]-1$	 $x'y'z' + x'y'z + x'yz' + x'yz$
<b>Type II</b> $3[4]-2$	 $x'y'z' + x'y'z + xyz' + xyz$
<b>Type III</b> $3[4]-3$	 $x'y'z' + x'y'z + x'yz' + xy'z$
<b>Type IV</b> $3[4]-4$	 $x'y'z' + x'y'z + x'yz' + xy'z'$
<b>Type V</b> $3[4]-5$	 $x'y'z' + x'y'z + x'yz' + xyz$
<b>Type VI</b> $3[4]-6$	 $x'y'z' + x'yz + xy'z + xyz'$

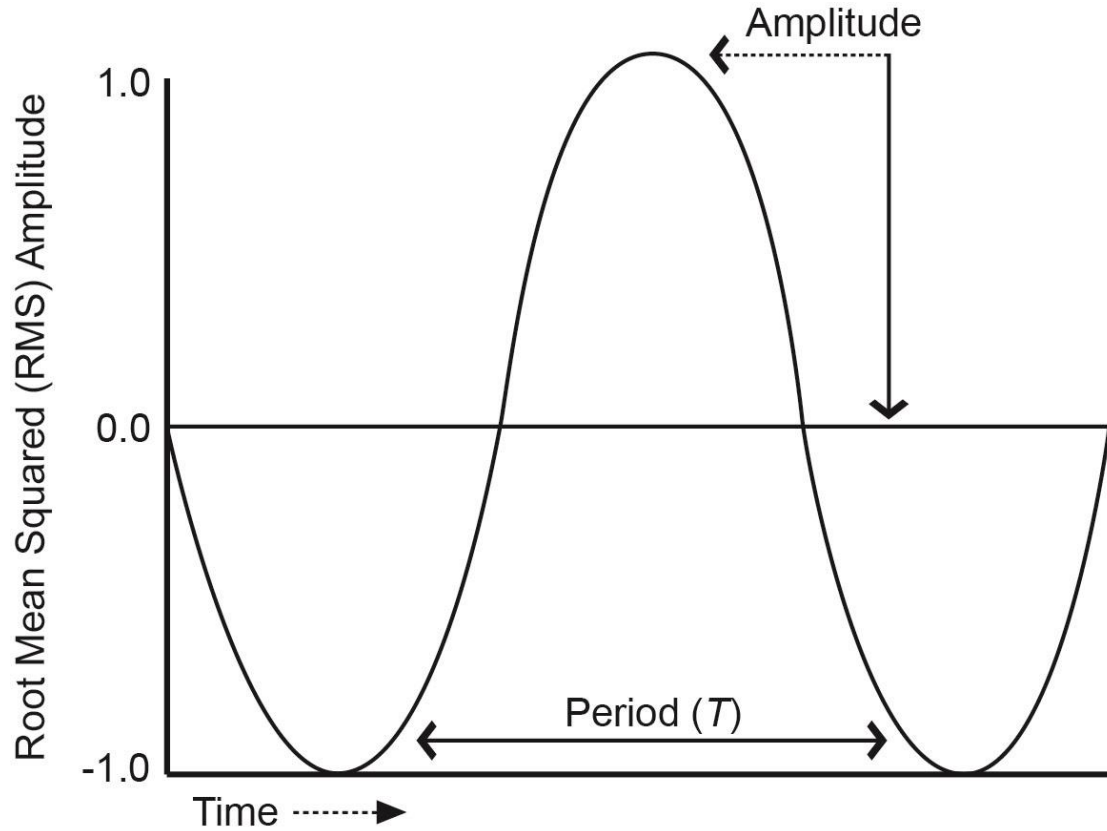
*Figure 1.* The 3[4] SHJ family. Each set of objects contains 4 positive objects defined according to three binary dimensions. In this example, those dimensions are shape, size, and color.



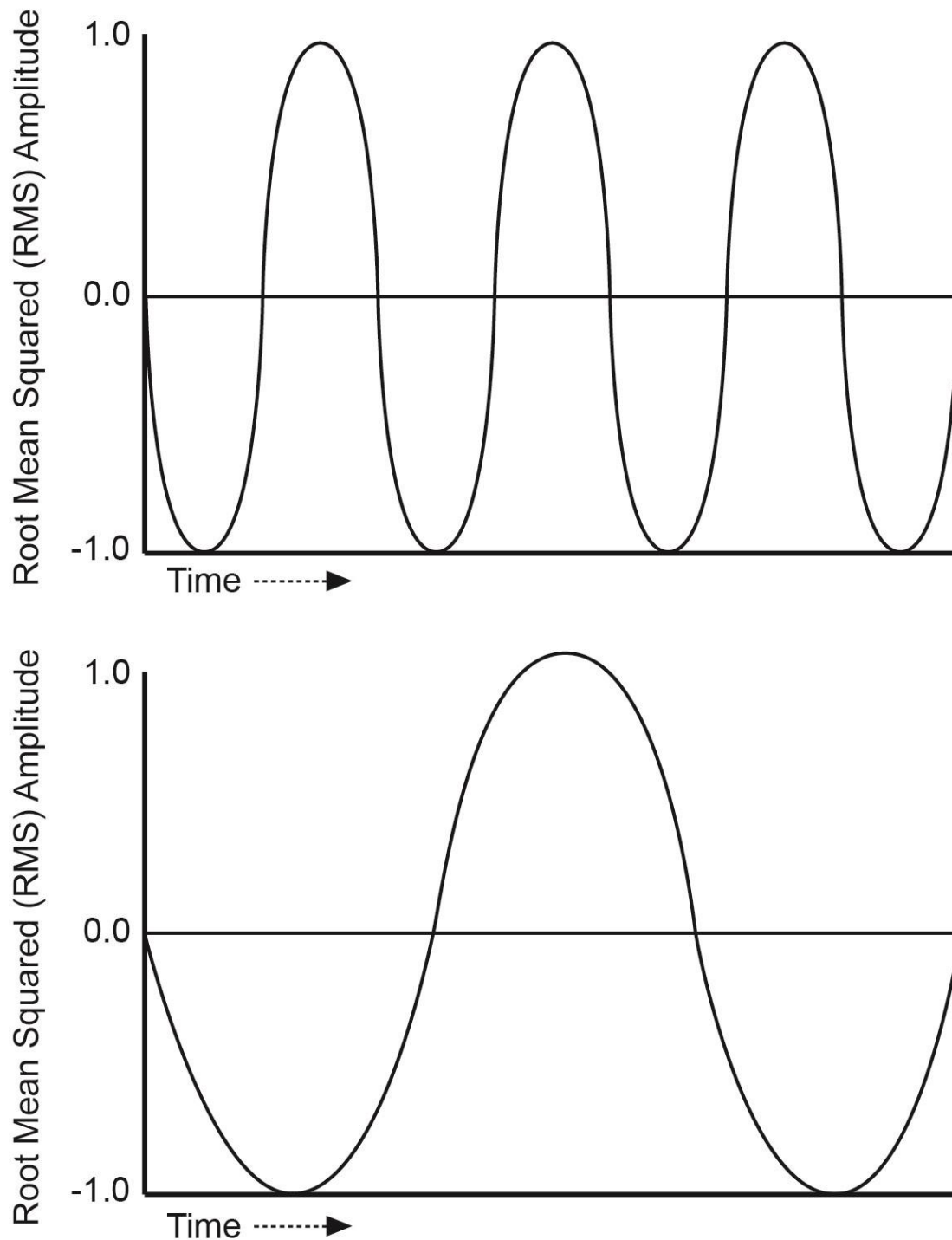
*Figure 2.* An example of how stimuli may be clustered in an unsupervised learning task. Each cross varies according to two dimensions; X-axis location and Y-axis location. The lower panel presents a possible grouping of the stimuli.



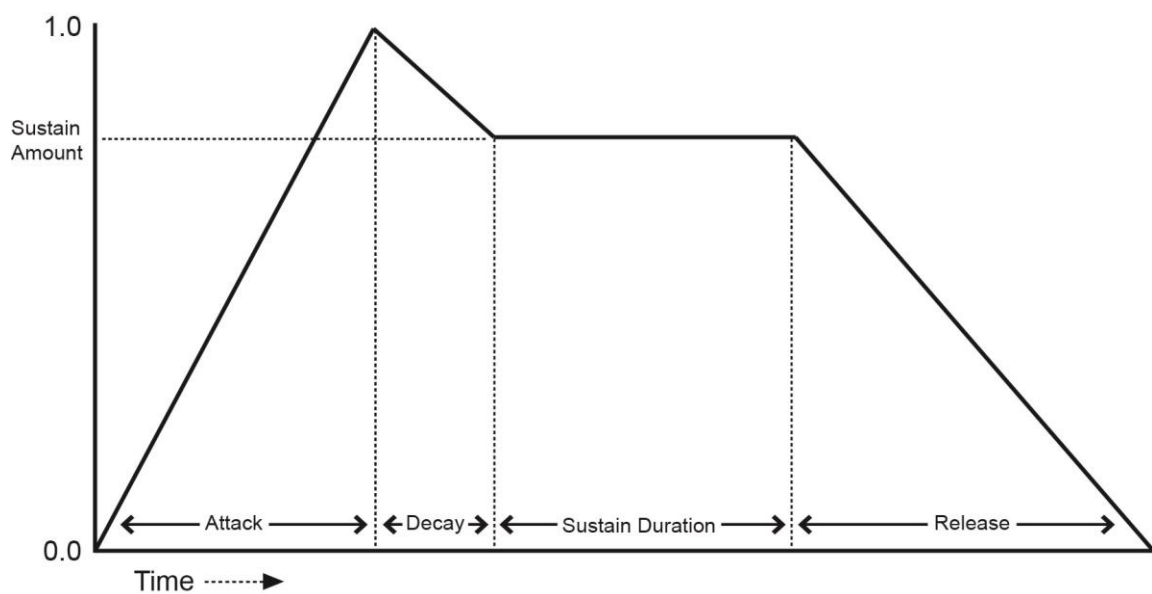
*Figure 3.* These stimuli are similar to those used by Pothos and Chater (2002). The X and Y dimensions of Figure 2 are translated in to different radius of the stars.



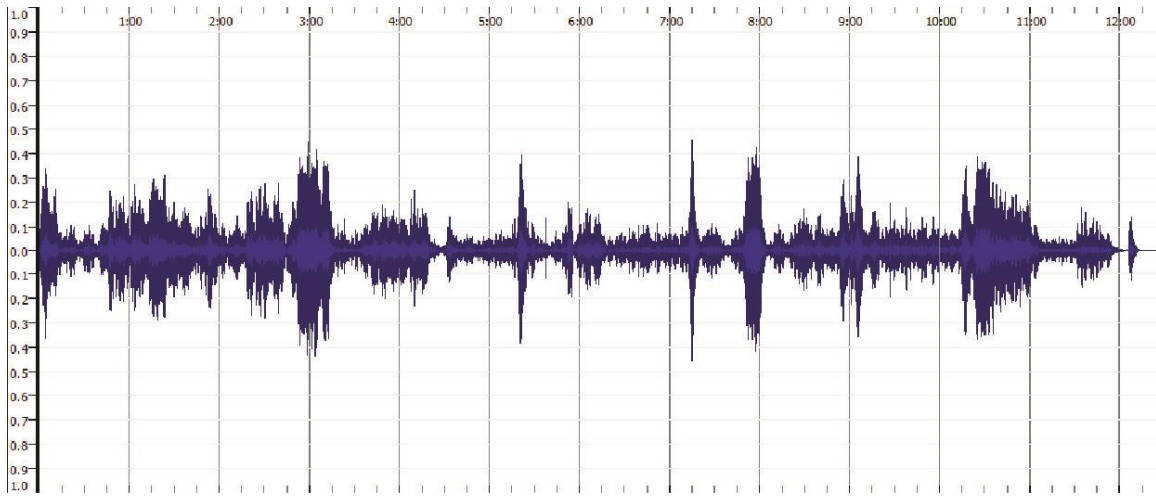
*Figure 4.* This presents a regular, repeating wave. The period of a wave is the amount of time necessary for the wave to return to the same state with respect to the zero crossing. The amount of deviation the wave makes from the zero crossing reflects the amplitude for the given period.



*Figure 5.* The top panel represents a wave where the periodicity is short. If the frequency of the waveform is within the range of human hearing, such a sound may be perceived as having a higher pitch. The panel below represents a waveform with a longer period. Relative to the waveform in the top panel, the pitch would be considered lower

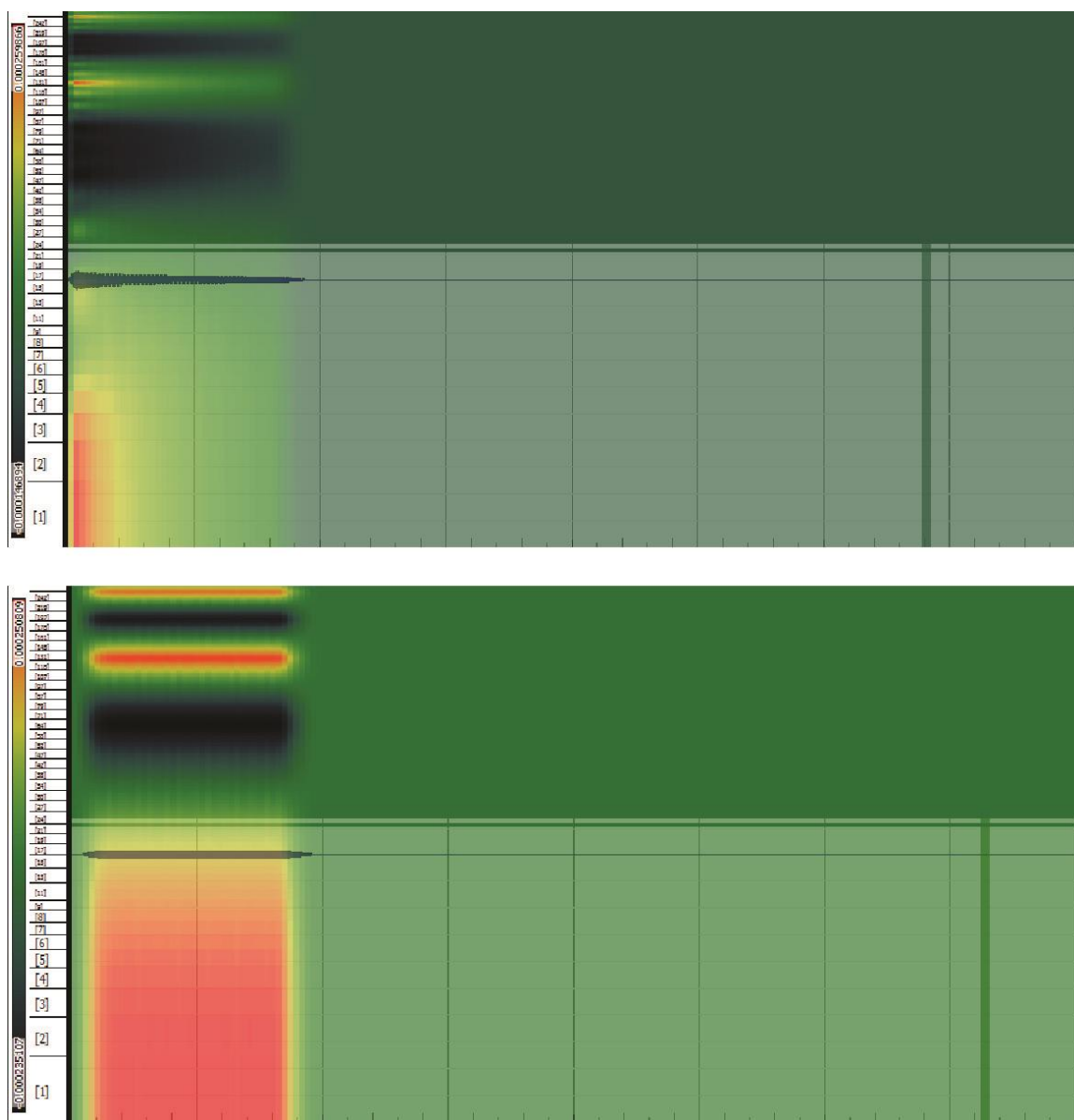


*Figure 6.* The components of an ADSR (attack, decay, sustain, and release). The amplitude of the audio changes through time.

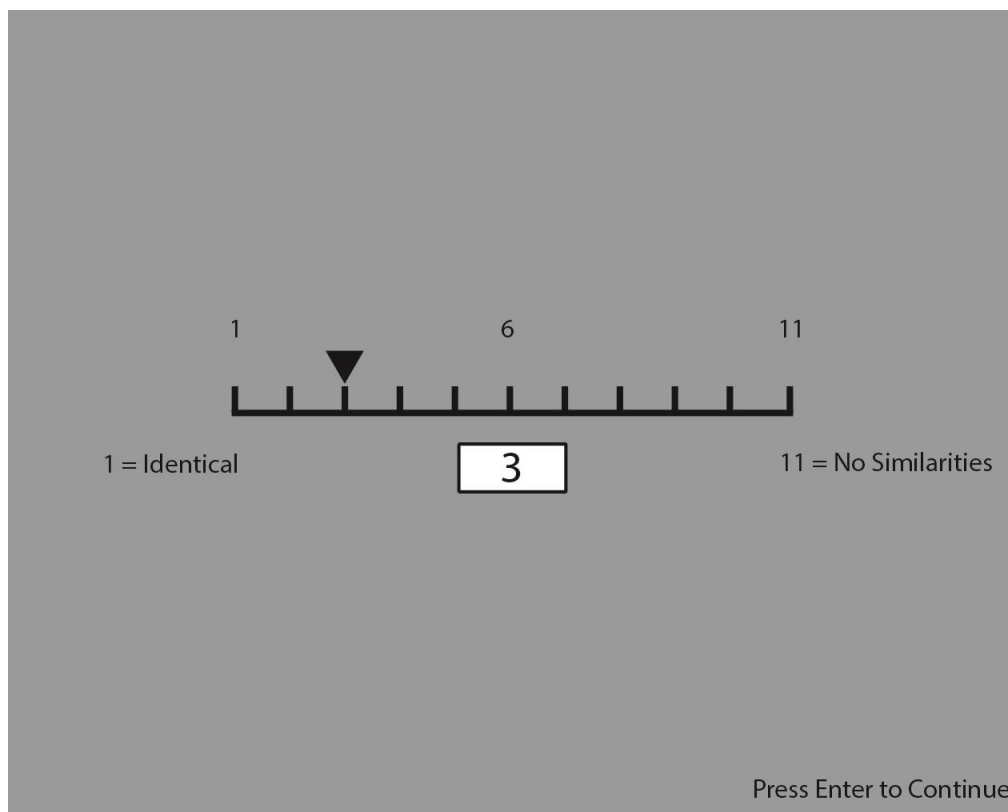


*Figure 7.* A waveform in the time-domain as taken from Sonic Visualizer. Intuitively, the X axis represents the progression of time while the Y axis represents the amplitude of the waveform. As can be discerned, the wave form contains several instances of amplitude dynamics – changes in the relative amplitude across time.

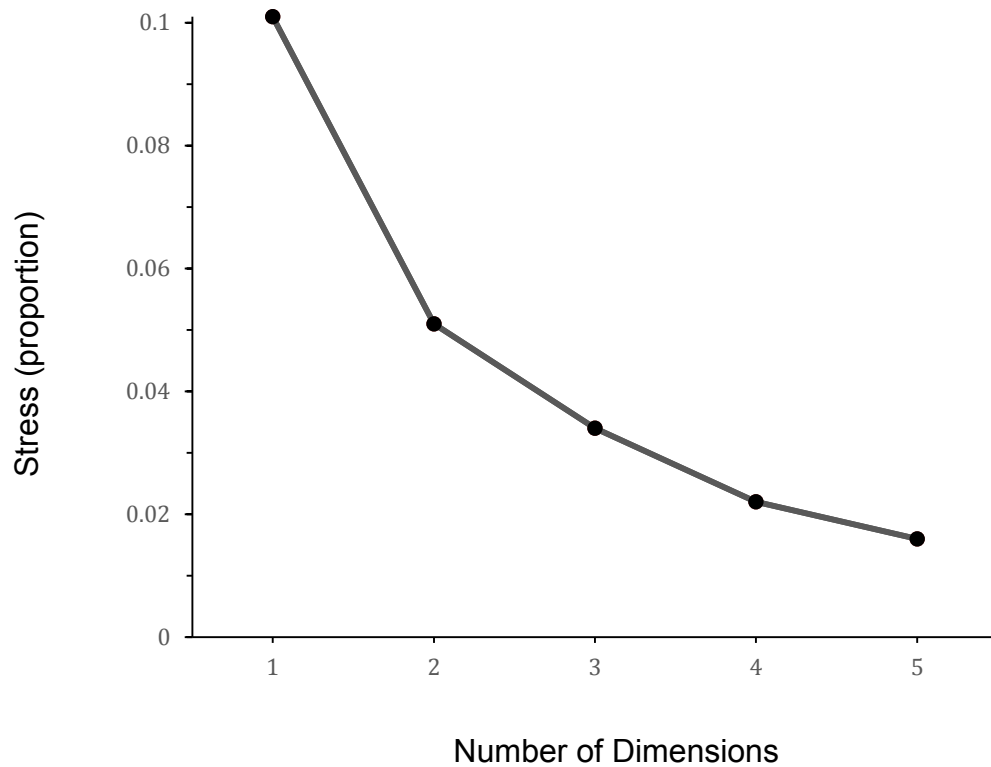




*Figure 8.* Waveforms in the frequency-domain. Instead of showing the waveform through time, the waveform is examined only for an instance of time (the window) for the frequency content. These particular waveforms demonstrate differences in their harmonics (identified by their color, duration, and frequency bin location on the X axis).



*Figure 9.* The computer screen during the MDS of experiment 1. Participants listen to the pair of sounds play. Immediately the slider is presented. Participants can then drag the slider to input their answer. A value of '1' is no difference/identical; A value of '11' is completely different/no similarities.



*Figure 10.* Scree plot of the stress measures per NMDS model.

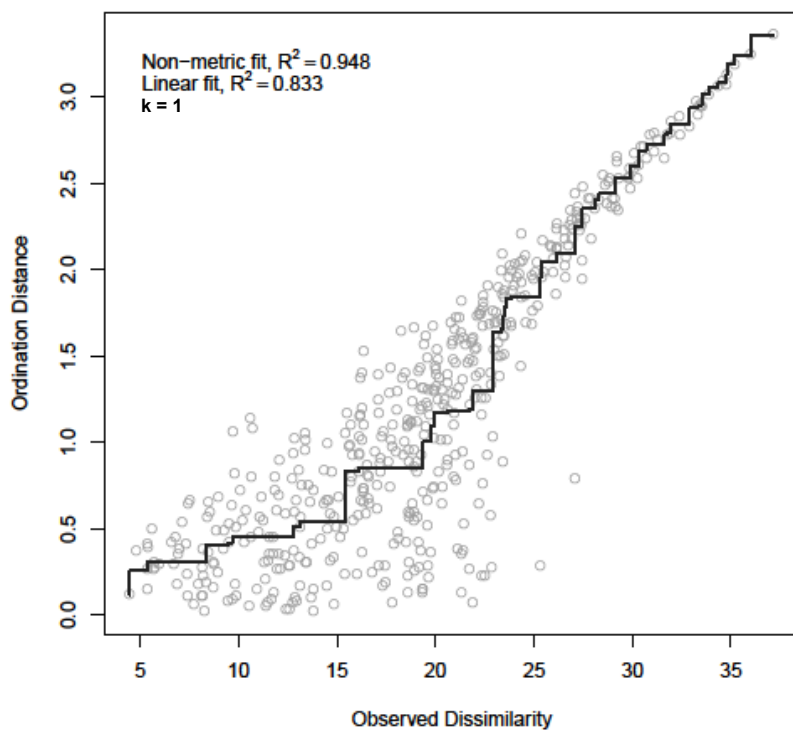


Figure 11. Shepard plot of one-dimensional NMDS.

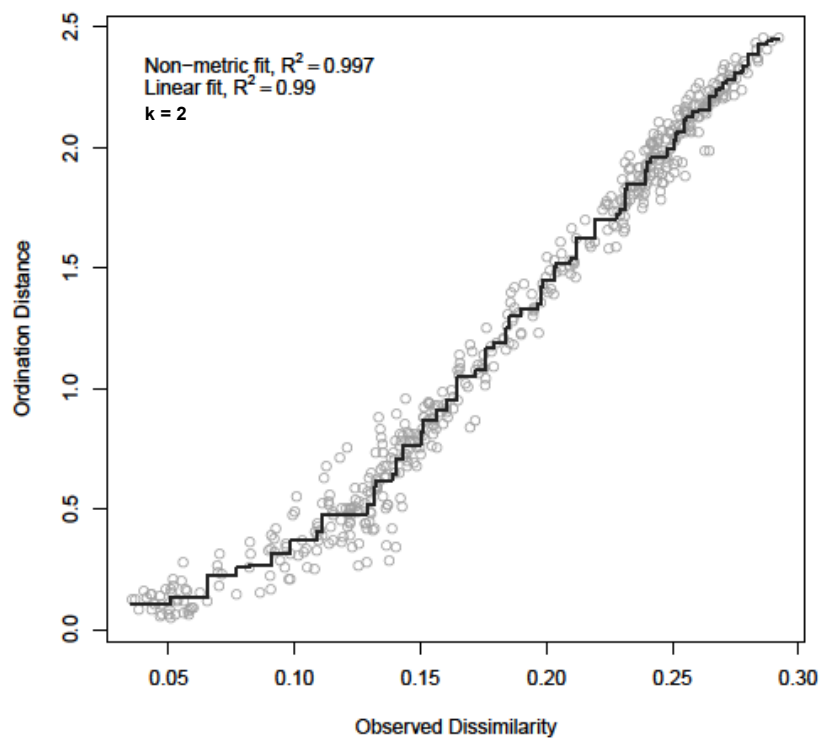


Figure 12. Shepard plot of two-dimensional NMDS.

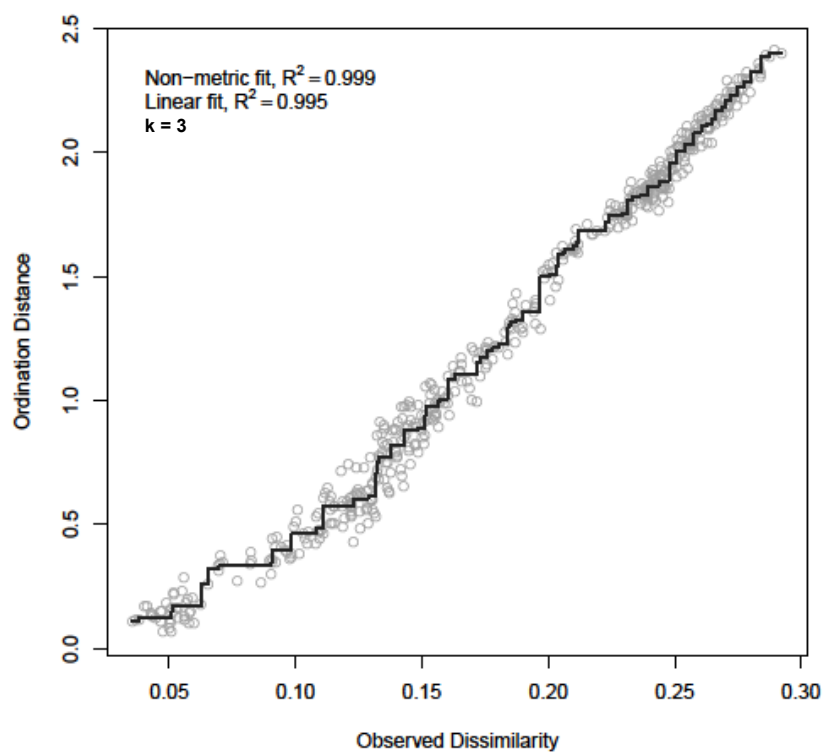


Figure 13. Shepard plot of three-dimensional NMDS.

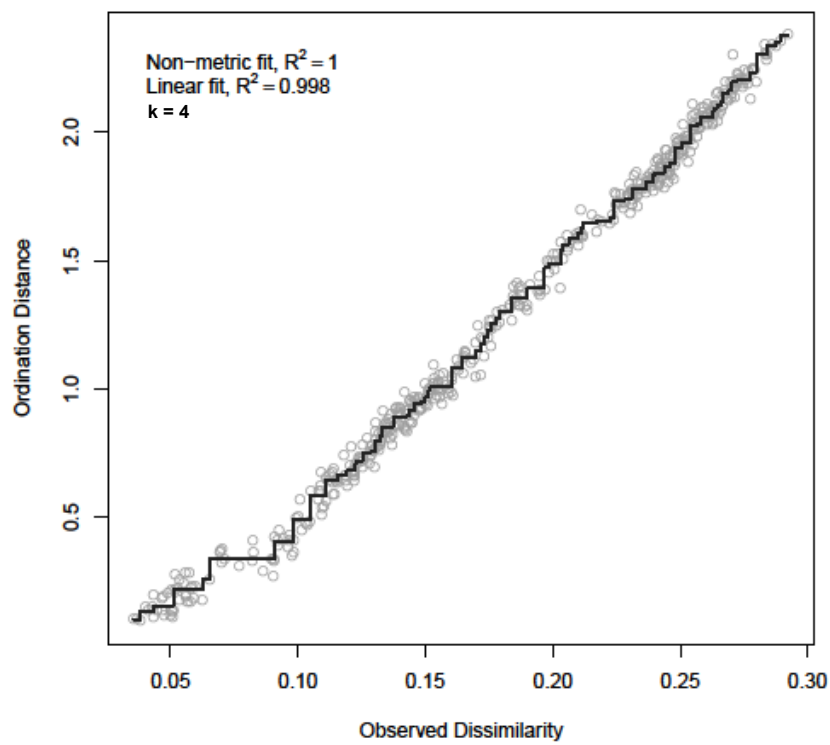


Figure 14. Shepard plot of four-dimensional NMDS.

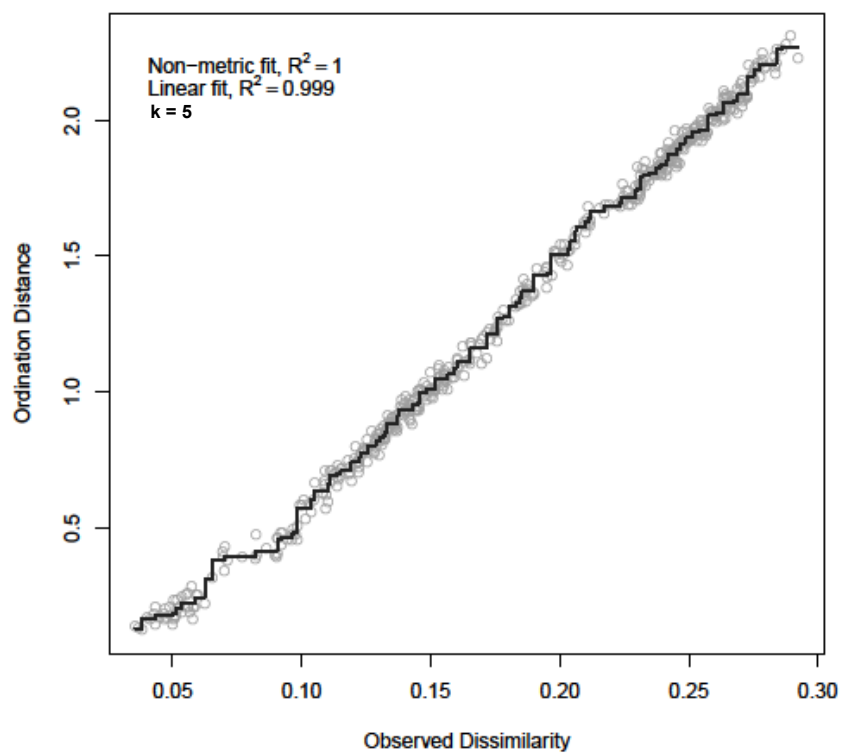
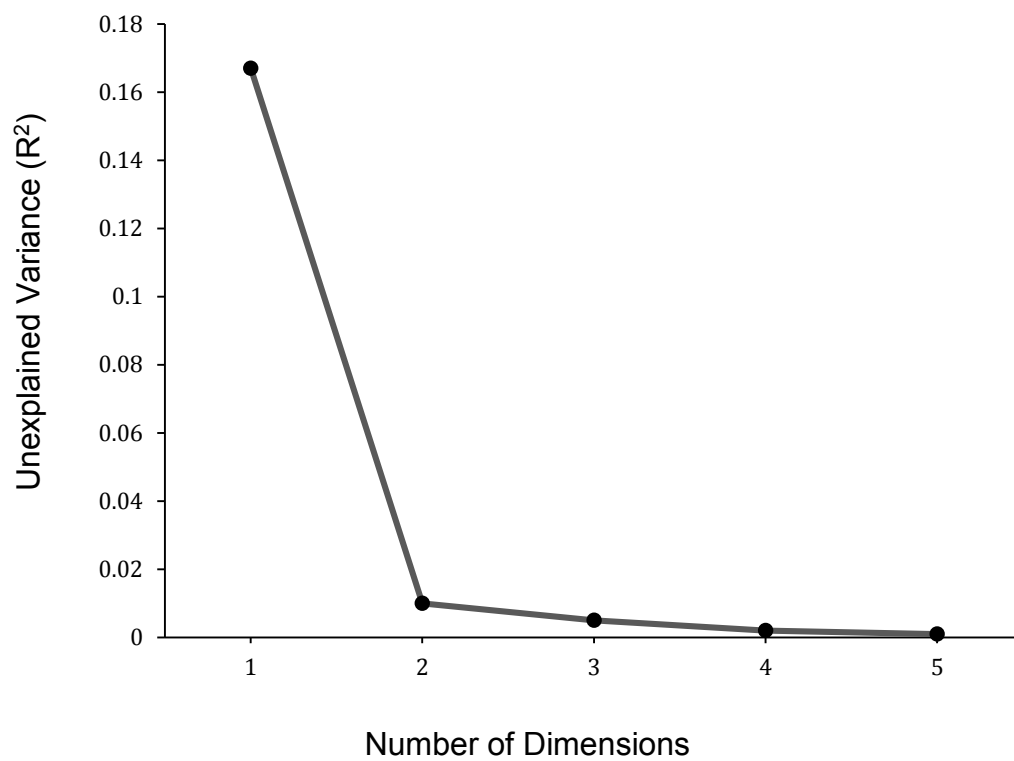
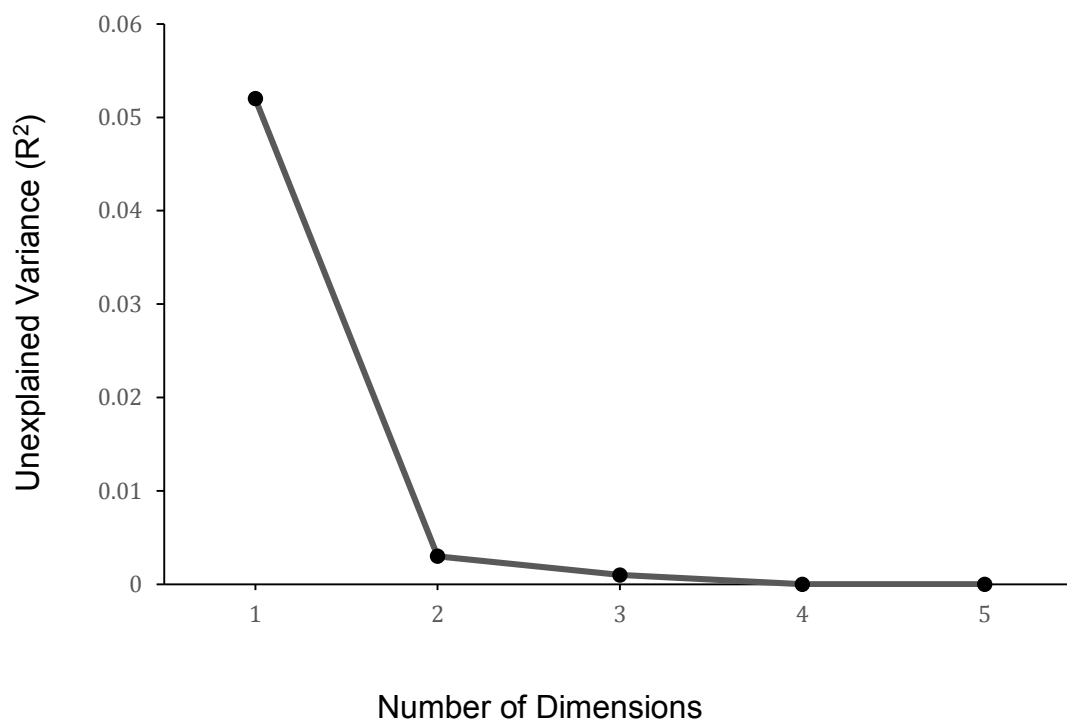


Figure 15. Shepard plot of five-dimensional NMDS.





*Figure 16.* Scree plot of unexplained variance for metric Shepard plot  $R^2$



*Figure 17.* Scree plot of unexplained variance for non-metric Shepard plot  $R^2$

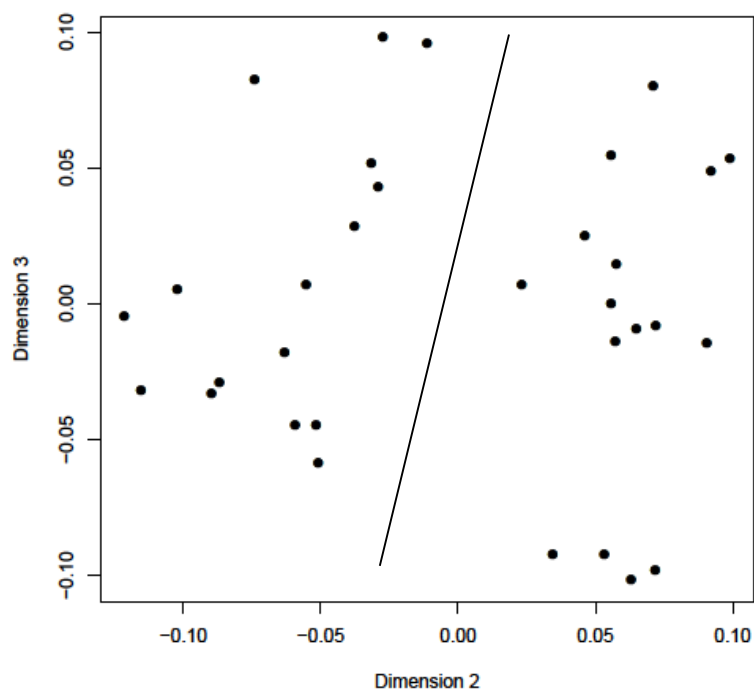
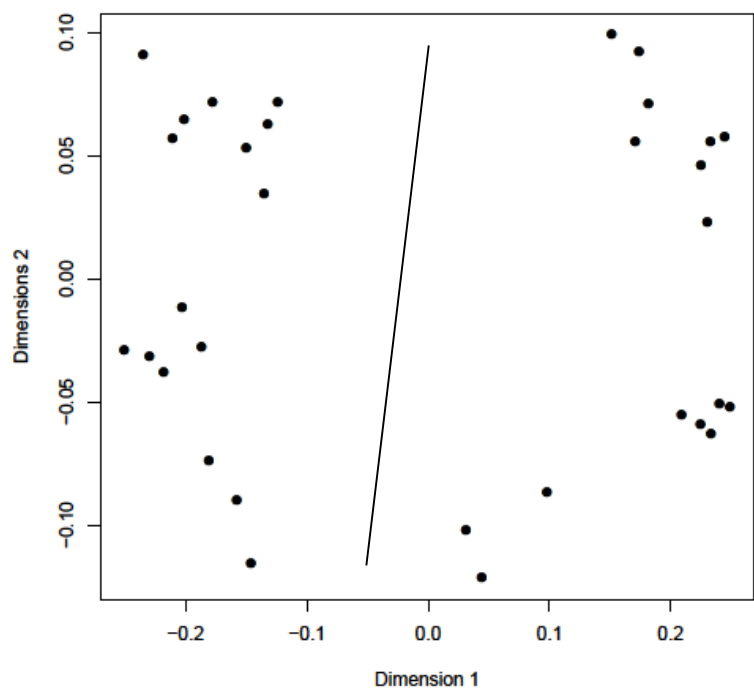
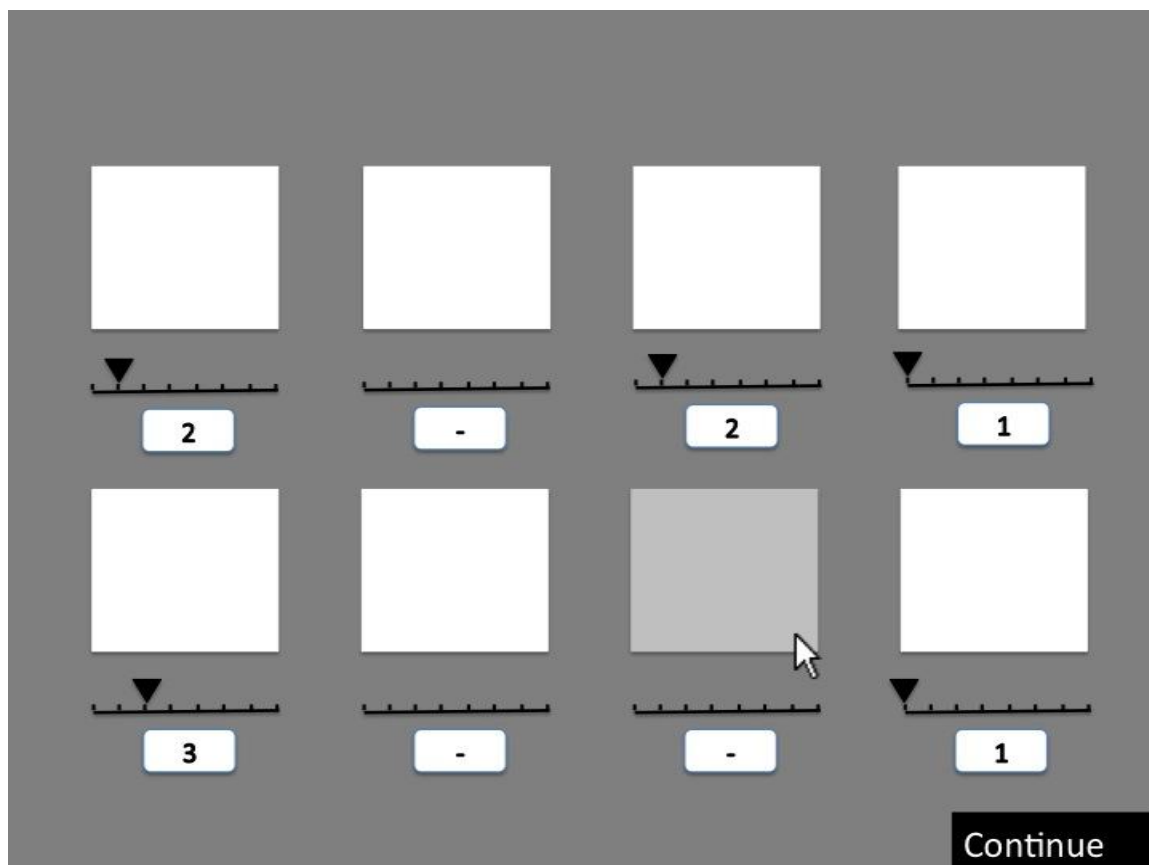


Figure 18. Graph of the 3 dimensional NMDS solution.



*Figure 19.* The computer screen during the unsupervised learning task. Each square icon is associated with a unique sound. Participants listen to the sounds by clicking the associated icon. They can then input the group they wish to associate that particular sound, and can do so in any order.

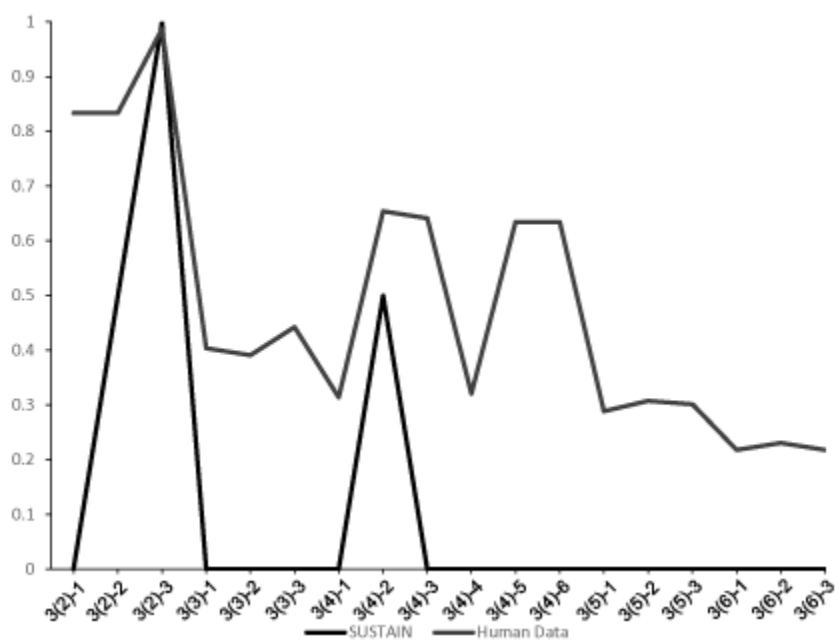


Figure 20. SUSTAIN predictions against the human data.

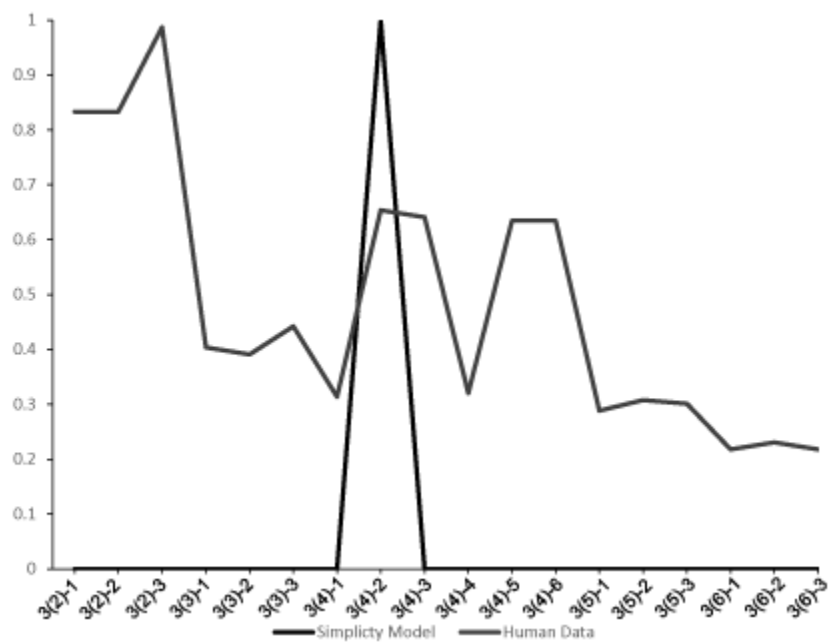


Figure 21. Simplicity model predictions against the human data.

## REFERENCES

- American Standards Association. (1960). *American Standard Acoustic Terminology*. New York: American Standards Association.
- Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*(6), 1178-1199.
- Balzano, G. J. (1986). What are musical pitch and timbre?. *Music Perception*, 297-314.
- Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*(2), 127.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(8), 1798-1828.
- Berger, K.W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, *36*, 1888-1891.
- Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 458.
- von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acta Acustica united with Acustica*, *30*(3), 159-172.
- Blumensath, T., & Davies, M. (2004, May). Unsupervised learning of sparse and shift-invariant decompositions of polyphonic music. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on* (Vol. 5, pp. V-497). IEEE.
- Bonebright, T. L. (2001). Perceptual structure of everyday sounds: A multidimensional scaling approach.
- Bourne, L. E. (1963). Factors affecting strategies used in problems of concept-formation. *The American journal of psychology*, 229-238.
- Bourne Jr, L. E., & Guy, D. E. (1968). Learning conceptual rules: II. The role of positive and negative instances. *Journal of Experimental Psychology*, *77*(3p1), 488.
- Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L., & He, X. (2010, October). Music recommendation by unified hypergraph: combining social media information and

- music content. In *Proceedings of the international conference on Multimedia* (pp. 391-400). ACM.
- Bulgarella, R. G., & Archer, E. J. (1962). Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology*, 63(3), 254.
- Bundy, R. S., Colombo, J., & Singer, J. (1982). Pitch perception in young infants. *Developmental Psychology*, 18(1), 10.
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, 118(1), 471-482.
- Cai, R., Lu, L., & Hanjalic, A. (2005, November). Unsupervised content discovery in composite audio. In *Proceedings of the 13th annual ACM international conference on Multimedia* (pp. 628-637). ACM.
- Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience letters*, 405(3), 164-167.
- Chen, Z., & Cowen, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1325-49.
- Cowan N., Chen Z., & Rouders J.N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller. *Psychological Science*, 1956, 634-640.
- Cheng, P.W. & Pachella, R.G. (1984). A psychophysical approach to dimensional separability. *Cognitive Psychology*, 16, 279-304.
- Clarkson, M. G., & Clifton, R. K. (1985). Infant pitch perception: Evidence for responding to pitch categories and the missing fundamental. *The Journal of the Acoustical Society of America*, 77(4), 1521-1528.
- Clarkson, B., & Pentland, A. (1999, March). Unsupervised clustering of ambulatory audio and video. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on* (Vol. 6, pp. 3037-3040). IEEE.
- Clarkson, M. G., Clifton, R. K., & Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. *Perception & psychophysics*, 43(1), 15-20.



- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2), 240-247.
- Colreavy, E., & Lewandowsky, S. (2008). Strategy development and learning differences in supervised and unsupervised categorization. *Memory & cognition*, 36(4), 762-775.
- Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics*, 53(4), 403-421.
- Compton, B. J., & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception & Psychophysics*, 61(7), 1320-1335.
- Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *The Journal of the Acoustical Society of America*, 95(5), 2720-2727.
- de Bruijn, A. (1978). Timbre classification of complex tones. *Acustica*, 40, 108-114.
- Doan, C.A., & Vigo, R. (under review). Constructing and deconstructing concepts: On the nature of category modifications and one-dimensional sorts.
- Edwards, D. J., Perlman, A., & Reed, P. (2012). Unsupervised categorization in a sample of children with autism spectrum disorders. *Research in developmental disabilities*, 33(4), 1264-1269.
- Farnell, A. (2010). *Designing sound* (pp. 310-312). Cambridge: MIT Press.
- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, 407(6804), 630-633.
- Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, 47(1), 75-89.
- Feldman, J. (2006). An algebra of human concept learning. *Journal of mathematical psychology*, 50(4), 339-368.
- Fleiss, J. L., & Zubin, J. (1969). On the methods and theory of clustering. *Multivariate Behavioral Research*, 4(2), 235-250.

- Freyman, R. L., Nerbonne, G. P., & Cote, H. A. (1991). Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition. *Journal of Speech, Language, and Hearing Research*, 34(2), 415-426.
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current opinion in neurobiology*, 17(4), 437-455.
- Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, 104(1), 505-510.
- Gao, S., Lee, C. H., & Zhu, Y. W. (2004). An unsupervised learning approach to musical event detection. In *Multimedia and Expo, 2004, IEEE International Conference*, 2, 1307-1310.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3), 225-241.
- Garner, W. R. (1974). *The processing of information and structure*. Psychology Press.
- Ghahramani, Z. (2004). Unsupervised learning. In *Advanced Lectures on Machine Learning* (pp. 72-112). Springer Berlin Heidelberg.
- Giguere, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. *Tutorials in Quantitative Methods for Psychology*, 2, 26-37.
- Glasberg, B. R., & Moore, B. C. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5), 331-342.
- Goodwin, G. P., & Johnson-Laird, P. N. (2011). Mental models of Boolean concepts. *Cognitive psychology*, 63(1), 34-59.
- Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, 23(8), 995-1019.
- Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16), 10911-10916.

- Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics*, *18*(1), 26-28.
- Goudbeek, M., Swingley, D., & Kluender, K. R. (2007). The limits of multidimensional category learning. In *INTERSPEECH* (pp. 2325-2328).
- Goudbeek, M., Swingley, D., & Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1913.
- Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *The Journal of the Acoustical Society of America*, *77*(2), 671-677.
- Grant, K. W., Braida, L. D., & Renn, R. J. (1991). Single band amplitude envelope cues as an aid to speechreading. *The Quarterly Journal of Experimental Psychology*, *43*(3), 621-645.
- Grant, K. W., Braida, L. D., & Renn, R. J. (1994). Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech. *The Journal of the Acoustical Society of America*, *95*(2), 1065-1073.
- Grau, J. W., & Nelson, D. K. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*(4), 347
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270-1277.
- Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, *63*(5), 1493-1500.
- Guastavino, C., & Katz, B. F. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of the Acoustical Society of America*, *116*(2), 1105-1115.
- Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & psychophysics*, *69*(6), 839-855.
- Hartmann, W. M. (1978). The effect of amplitude envelope on the pitch of sine wave tones. *The Journal of the Acoustical Society of America*, *63*(4), 1105-1113.

- Harshman, R.A, Green, P.E, Wind, Y., & Lundy, M.E. (1982). A model for analysis of asymmetric data in marketing research. *Marketing Science, 1*, 205-242.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors.
- Hoover, D. M., & Cullari, S. (1992). Perception of loudness and musical preference: Comparison of musicians and nonmusicians. *Perceptual and motor skills, 74*(3c), 1149-1150.
- Houtsma, A. J. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*(2), 104-115.
- Howard, D. M., & Angus, J. (2009). *Acoustics and psychoacoustics*. Taylor & Francis.
- Huang, Y. C., & Jenor, S. K. (2004, June). An audio recommendation system based on audio signature description scheme in mpeg-7 audio. In *Multimedia and Expo, 2004 IEEE International Conference, 2*, 639-642.
- Hull, C. L. (1920). Quantitative aspects of evolution of concepts: An experimental study. *Psychological monographs, 28*(1).
- Jain, A.K., & Pinson, C. (1976). The effect of order of presentation of similarity judgments on multidimensional scaling results: An empirical examination. *Journal of Marketing Research, 4*, 435-439.
- Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 92.
- Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and it's utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology, 5*, 1-10.
- Karjalainen, M., Välimäki, V., & Tolonen, T. (1998). Plucked-string models: From the Karplus-Strong algorithm to digital waveguides and beyond. *Computer Music Journal, 17*-32.
- Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole view.
- Khalifa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J. L., & Collet, L. (2004). Increased perception of loudness in autism. *Hearing research, 198*(1), 87-92.

- Kohler, K. J. (1987). Categorical pitch perception. In *Proc. 11th ICPHS* (pp. 331-333).
- Kollmeier, B., Brand, T., and Meyer, B. (2008). Perception of speech and sound. In *Springer handbook of speech processing* (pp. 61-82). Springer Berlin Heidelberg
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques.
- Krimphoff, J. (1993). Analyse acoustique et perception du timbre. *unpublished DEA thesis, Université du Maine, Le Mans, France.*
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 739.
- Krumhansl, C. L. (1989). Why is musical timbre so hard to understand. *Structure and perception of electroacoustic sound and music*, 9, 43-53.
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological review*, 99(1), 22.
- Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in cognitive sciences*, 9(1), 26-33.
- Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 1.
- Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 801.
- Liu, Z., Wang, Y., & Chen, T. (1998). Audio feature extraction and analysis for scene segmentation and classification. *Journal of VLSI signal processing systems for signal, image and video technology*, 20(1-2), 61-79.
- Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *The Journal of the Acoustical Society of America*, 70(2), 387-389.
- Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, 9(4), 829-835.

- Love, B. C., & Medin, D. L. (1998, July). SUSTAIN: A model of human category learning. In *AAAI/IAAI* (pp. 671-676).
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological review*, *111*(2), 309.
- Luce, R. D. (1963). Detection and recognition.
- Luce, R. D. (1977). Thurstone's discriminial processes fifty years later. *Psychometrika*, *42*(4), 461-489.
- Luo, X., & Fu, Q. J. (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. *The Journal of the Acoustical Society of America*, *116*(6), 3659-3667.
- Magee, D., Needham, C. J., Santos, P., Cohn, A. G., & Hogg, D. C. (2004). Autonomous learning for a cognitive agent using continuous models and inductive logic programming from audio-visual input. In *Proceedings of the AAAI workshop on Anchoring Symbols to Sensor Data*, 17-24.
- McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philosophical transactions: Biological sciences*, 383-389.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological research*, *58*(3), 177-192.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive psychology*, *19*(2), 242-279.
- Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*(2), 169-178.
- Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *The Journal of the Acoustical Society of America*, *58*(3), 711-720.
- Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, *80*(2), 479-483.

- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility head movement improves auditory speech perception. *Psychological science*, *15*(2), 133-137.
- Murphy, G. L. (2002). *The big book of concepts*. MIT press.
- Neuhoff, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: a central, analytic process. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1050.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, memory, and cognition*, *10*(1), 104.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*(1), 39.
- Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & cognition*, *22*(3), 352-369.
- Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, *3*(2), 222-226.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological review*, *101*(1), 53.
- Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1194.
- Ohl, F. W., Scheich, H., & Freeman, W. J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature*, *412*(6848), 733-736.
- Park, A. S., & Glass, J. R. (2008). Unsupervised pattern discovery in speech. *Audio, Speech, and Language Processing, IEEE Transactions*, *16*(1), 186-197.
- Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research*, *50*(3), 647-666.

- Pachella, R. G., Somers, P., & Hardzinski, M. (1980). A psychophysical approach to dimensional integrality. In D.J. Getty & J.H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 107-126). Hillsdale, NJ: Erlbaum.
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of experimental psychology: human perception and performance*, 20(5), 976.
- Plack, C. J., & Oxenham, A. J. (2005). The psychophysics of pitch. In *Pitch* (pp. 7-55). Springer New York.
- Pothos, E. M., & Chater, N. (2001). Categorization by simplicity: A minimum description length approach to unsupervised clustering.
- Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, 26(3), 303-343.
- Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *The Quarterly Journal of Experimental Psychology*, 58(4), 733-752.
- Pothos, E. M., Perlman, A., Edwards, D. J., Gureckis, T. M., Hines, P. M., & Chater, N. (2008). Modeling category intuitiveness. In *Proceedings of the 30th annual conference of the cognitive science society*. Mahwah, NJ: LEA.
- Pothos E.M., Perlman, A., Bailey, T.M., Kurtz, K., Edwards, D.J., Hines, P., and McDonnell, J.V. (2011). Measuring category intuitiveness in unconstrained categorization tasks. *Cognition*, 121, 83-100.
- Prior, M., & Troup, G. A. (1988). Processing of timbre and rhythm in musicians and non-musicians. *Cortex*, 24(3), 451-456.
- Puckette, M. (2007). The theory and technique of electronic music.
- Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science*, 12(5), 410-430.
- Rehder, B., & Hoffman, A. B. (2005). Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 811.
- Roads, C. (1996). *The computer music tutorial*. MIT press.
- Roads, C. (2004). *Microsound*. MIT press.



- Robinson, D. W., & Sutton, G. J. (1979). Age effect in hearing—a comparative analysis of published threshold data. *Audiology*, *18*(4), 320-334.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, *104*(3), 192.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573-605
- Rumelhart, D. E., & Ortony, A. (1976). *The representation of knowledge in memory* (pp. 99-135). Center for Human Information Processing, Department of Psychology, University of California, San Diego.
- Samson, S., Zatorre, R. J., & Ramsay, J. O. (1997). Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *51*(4), 307.
- Ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America*, *91*(5), 2872-2880.
- Tulving E., & Patkau J.E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, *16*, 83–95.
- Saults, J.S., & Cowan, N. (2009). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, *136*, 663-684.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325-345.
- Shepard, R.N. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*, 242-256.
- Shepard, R.N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*, 509-523.
- Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis.

- Shepard, R. N., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, 65(1), 94.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317-1323.
- Shinn, P., & Blumstein, S. E. (1984). On the role of the amplitude envelope for the perception of [b] and [w]. *The Journal of the Acoustical Society of America*, 75(4), 1243-1252.
- Smalley, D. (1994). Defining timbre - Refining timbre. *Contemporary Music Review*, 10, 35-48.
- Smoke, K. L. (1933). Negative instances in concept learning. *Journal of Experimental Psychology*, 16(4), 583.
- Speer, J. R., & Meeks, P. U. (1985). School children's perception of pitch in music. *Psychomusicology: A Journal of Research in Music Cognition*, 5(1-2), 49.
- Stevens, S. S., & Volkman, J. O. H. N. (1940). The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, 329-353.
- Stevens, S. S. (1960). The psychophysics of sensory function. *American Scientist*, 226-253.
- Stevens, S. S. (1970). Neural events and the psychophysical law. *Science*.
- Sullivan, C. R. (1990). Extending the Karplus-Strong algorithm to synthesize electric guitar timbres with distortion and feedback. *Computer Music Journal*, 26-37.
- Svirsky, M. A. (2000). Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues. *The Journal of the Acoustical Society of America*, 107(3), 1521-1529.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental brain research*, 161(1), 1-10.
- Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of vision*, 10(12), 14.
- Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, 7(3), 289-296.

- Trehub, S. E., Endman, M. W., & Thorpe, L. A. (1990). Infants' perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, *49*(2), 300-313.
- Tucker, L.R., & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, *28*, 333-367.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.
- Tversky, A., & Gati, I. (1978). Studies of similarity. *Cognition and categorization*, *1*(1978), 79-98.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological review*, *89*(2), 123.
- Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273-13278.
- Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Wouters, J., & Ghesquière, P. (2011). Impairments in speech and nonspeech sound categorization in children with dyslexia are driven by temporal processing difficulties. *Research in developmental disabilities*, *32*(2), 593-603.
- Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology*, *50*(5), 501-510.
- Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology*, *53*(4), 203-221.
- Van Segbroeck, M., & Van Hamme, H. (2009). Applying non-negative matrix factorization on time-frequency reassignment spectra for missing data mask estimation.
- Vigo, R. (2011). Towards a law of invariance in human concept learning. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 2580-2585). Cognitive Science Society.
- Vigo, R. (2013). The gist of concepts. *Cognition*, *129*(1), 138-162.
- Vigo, R., Barcus, M., Zhang, Y., & Doan, C. (2013). On the learnability of auditory concepts. *The Journal of the Acoustical Society of America*, *134*, 4064.

- Vigo, R., & Basawaraj. (2013). Will the most informative object stand? Determining the impact of structural context on informativeness judgements. *Journal of Cognitive Psychology*, 25(3), 248-266.
- Vigo, R., & Doan, C. A. (2015). Constructing and deconstructing concepts: On the nature of category modification and one-dimensional sorts (*under review*).
- Vigo, R., & Doan, C. A. (2015). The structure of choice. *Cognitive Systems Research*.
- Vigo, R., Zeigler, D. E., & Halsey, P. A. (2013). Gaze and informativeness during category learning: Evidence for an inverse relation. *Visual Cognition*, 21(4), 446-476.
- Vigo, R. (2014). *Mathematical principles of human conceptual behavior*. New York, NY: Psychology Press.
- Vurma, A., Raju, M., & Kuuda, A. (2010). Does timbre affect pitch? Estimations by musicians and non-musicians. *Psychology of Music*.
- Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, 24(4), 1052-1057.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, 18(2), 158-194.
- Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, 18(2), 221-281.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4), 420-422.
- Wülfing, J., & Riedmiller, M. (2012, October). Unsupervised Learning of Local Features for Music Classification. In *ISMIR* (pp. 139-144).
- Yates, J. T., Johnson, R. M., & Starz, W. J. (1972). Loudness perception of the blind. *International Journal of Audiology*, 11(5-6), 368-376.
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4), 1908-1919.

Zubin, J. (1938). A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology*, 33(4), 508.



**OHIO**  
UNIVERSITY

Thesis and Dissertation Services