Vrije Universiteit Brussel

FACULTY OF SCIENCE AND BIO-ENGINEERING SCIENCES
DEPARTMENT OF COMPUTER SCIENCE

# Learning Hierarchical Spectral Representations of Human Speech in the Information Dynamics of Thinking

Graduation thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Applied Sciences and Engineering: Computer Science

Steven Homer

Promoter: Prof. Dr. Dr. Geraint Wiggins
Advisor: Prof. Dr. Dr. Geraint Wiggins

Academic year 2018-2019

# Abstract

# Declaration of Originality

I hereby declare that this thesis was entirely my own work and that any additional sources of information have been duly cited. I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this thesis has not been submitted for a higher degree to any other University or Institution.

# Acknowledgements

# Contents

# List of Figures

# List of Tables

# 1
# Introduction

## 1.1 Context

## 1.2 Problems & Justifications

## 1.3 Objectives & Hypotheses

## 1.4 Research Method

## 1.5 Structure

# 2
# Background

## 2.1 Conceptual Spaces

In knowledge representation, the argument over the representation of cognition often falls into two camps: connectionism and symbolicism. As the lowest level of represention, connectionism is best exemplified by artificial neural networks, where cognition emerges from myrid connections between neurons. At the highest level, symbolicism views the mind as a Turing machine (Turing, 2009), where cognition is equivalent to compution by symbol manipulation. Conceptual spaces theory (Gärdenfors, 2004) argues for a middle way, through the use of so-called conceptual spaces. Conceptual spaces theory views cognition as the process of concept formation by means of similarity, so that the continuous representations of connectionism can be bridged to the discrete representations of symbolicism, hopefully gaining the best of both worlds.

Conceptual spaces theory places objects into so-called conceptual spaces – semantically rich spaces with geometric properties – which allow for intuitive geometric reasoning about related objects. For instance, in the conceptual space of color with dimensions of hue, saturation, and brightness, one can formally make a geometric claim that "orange lies between yellow and red." This simple claim cannot be made by connectionism or symbolicism without imposing ad-hoc external semantics on the neural connections or symbols respectively, highlighting the explanatory power of conceptual spaces for knowledge representation.

### 2.1.1 Quality Dimensions

Quality dimensions are the basic building blocks of a conceptual space. They can be thought of as the axes that give meaning to the elements in the space. In three-dimensional Cartesian space, when referring to a point, we specify it by its placement on each of the $x$, $y$, and $z$ -axes. By analogy, each of the $xyz$-axes would be a quality dimension in the 3D Cartesian space. However, quality dimensions are more than just orthogonal unit vectors, they can also have their own specific geometry that serves to constrain the dimension. For instance, a quality dimension may have the geometry of a circle, resulting in different behavior than the real number line. Quality dimensions also allow us to speak meaningfully about similarity between objects in a space since, by definition, they possess distance and betweenness relations. Finally, it is important to remember that the 'quality' of the quality dimension is what gives it inherent semantic content beyond just being a descriptive dimension.

### 2.1.2 Domains and Conceptual Spaces

Integral quality dimensions require one another to exist. For instance, the three qualities of sound: pitch, timbre, and loudness, are all integral to one another. It is impossible to identify a sound without specifying all three of these dimensions. On the other hand, most quality dimensions are separable, meaning that they are independent of one another. Though separable quality dimensions are independent, they may still be highly correlated, which may give the illusion that they are integral.

A domain is a set of integral dimensions that are separable from all other dimensions. In a sense, it is the minimum description needed for a given space. For example, the three qualities of sound form a domain. Often, different domains will be correlated with one another, and combining them will yield a richer description of a given object. This combination of multiple domains is what is referred to as a conceptual space, so that the specification of an object is nothing else but its location in a conceptual space.

### 2.1.3 Similarity, Distance, Betweenness

Humans have an innate sense of similarity without being able to fully describe why two things are similar (cite?). In simple cases, this similarity can be made explicit, for example that a rectangle is more similar to a square than a circle. However, this intuition for similarity extends to even very abstract realms. For instance, most people would naturally agree that country music is closer to rock-n-roll than it is to classical Indian ragas, but would be hard-pressed to give an exact definition or method of why this is so.

Conceptual spaces allow this innate sense of similarity to be codified in the intuition of geometry, betweenness, and distance. Given a betweenness relation for a conceptual space, one can say that a given object is between two other objects, which allows us to say that one is closer to another than the third. In our case, we will study conceptual spaces equipped with a distance measure. This allows us to examine similarity between objects, such that more similar onjects will be closer to one another than more different objects.

### 2.1.4 Convexity of Properties and Concepts

Given that Conceptual Spaces Theory posits that cognition is equivalent to the formation of concepts (Gärdenfors, 2004), one should define a concept. Defining a property or context as "an invariance across a range of contexts, [reifiable] so that it can be combined with other appropriate invariances"

(Kirsh, 1991), it is immediately clear that these correspond to regions of a conceptual space. Since all of the objects in a given region are similar to each other, by grouping them together as a region, we can see that the region corresponds to a property or concept. A property would be a region in a domain – for instance the red property corresponds to a region of the color space – and a concept would be a region in a conceptual space.

Gardenfors (Gärdenfors, 2004) posits that regions corresponding to properties and concepts are convex in nature. Though this does not fall directly from the theory itself, it is reasonable to think that the region of concept is not intruded upon by other concepts. For example, in the color space, the property of red is convex, since we don't see another color like blue interloping into the red region, which would appear as a small area of blue surrounded by a region of red.

### 2.1.5 Higher-order Conceptual Spaces

Higher-order conceptual spaces can be created from combinations and transformations of one or more lower-order conceptual spaces (Gärdenfors, 2004). The key notion of the higher-order nature of these spaces is that they are more abstract than the spaces they are generated from. This is to distinguish higher-order conceptual spaces from combinations that serve to more tightly constrain a space, similar to an intersection set operation. For instance, overlaying the full color space over the space of human phenotypes results in a restricted color space of human skin tones, not a more abstract space.

Since by nature, quality dimensions and domains often describe low-level quantities like color and sound, it is necessary to combine them into more abstract spaces that yield higher-level spaces with more explanatory power. Intuitively, the higher level a given conceptual space is, the richer its semantics, so to arrive at a space with sufficient descriptive capabilities for a given cognitive representation, it may be necessary to recursively abstract conceptual spaces into higher-order spaces to arrive at something nontrivial with interesting semantics.

## 2.2 Hilbert Spaces

> "Hilbert spaces are the means by which the ordinary experience of Euclidian concepts can be extended meaningfully into the idealized constructions of more complex math." (Bernkopf, 2008)

Originally, conceptual spaces (Gärdenfors, 2004) were formalized by using Euclidian or Manhattan distances to measure similarity in conceptual spaces modeled in constrained Cartesian space. Though this is useful in building an intuition as to how conceptual spaces operate in practice, it can be limiting in the description of relations between objects and the space itself. In order to allow for more flexibility in this regard, we need a more general notion of a space than a finite-dimensional Euclidian space. The generalization employed here is that of Hilbert spaces, most famously applied as the foundation of quantum mechanics (von Neumann, 1955), but with numerous applications in a variety of fields (Kennedy & Sadeghi, 2013).

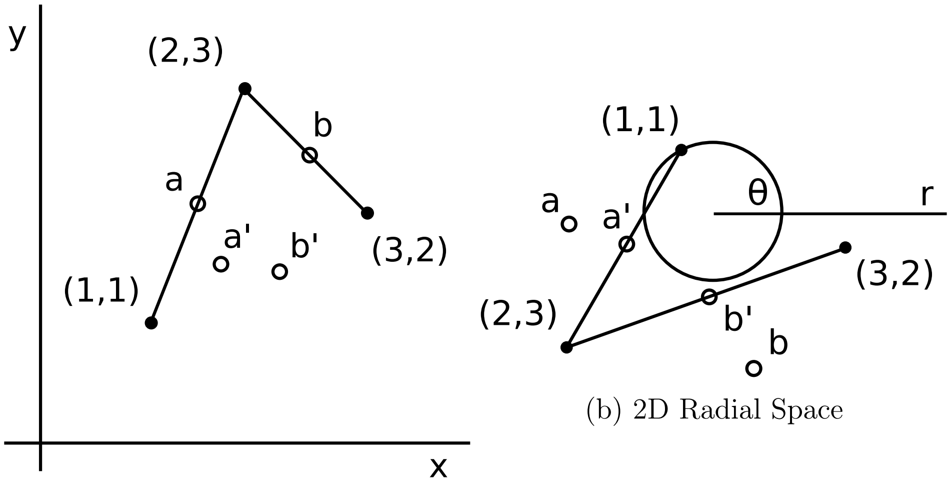
### 2.2.1 Complete Inner Product Space

A Hilbert space is defined as a complete inner product space. That is, a Hilbert space is a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$, but is also complete: the space is big enough to include the norm of converging sequences. In the case of an infinite-dimensional Hilbert space, this completeness criterion cannot be taken for granted, but in the finite-dimensional case, the space is always complete. The inner product of a Hilbert space induces a norm $\|f\| = \langle f, f \rangle^{1/2}$, which allows us to talk about distances between vectors, something that we require in a generalized formalism of conceptual spaces.

What makes Hilbert spaces powerful is the ability to represent a function as a point in the space. With the aid of the inner product, one can produce an (infinite) orthonormal sequence $\{\varphi_n\}_{n=0}^{\infty}$ for the Hilbert space. By decomposing any function $f$ into its Fourier series (equation 2.1) on that sequence, we can recover a corresponding coefficient $\langle f, \varphi_n \rangle$ for each element of the orthonormal sequence.

$$f = \sum_{n=0}^{\infty} \langle f, \varphi_n \rangle \varphi_n \tag{2.1}$$

By arranging each of these coefficients into a vector with dimensions $\varphi_n$, the function can be represented as a point in the Hilbert space. If that space is finite, we can equivalently represent a point $x$ in the Hilbert space as an array of complex numbers, $x \in \mathbb{C}^n$ where $n$ is the number of dimensions.




(a) 2D Cartesian Space

(b) 2D Radial Space

Figure 2.1: Example of effect of spatial geometry on location of coordinates

### 2.2.2 Application to Conceptual Spaces

Now, by defining a conceptual space formally in terms of a Hilbert space, we see that the geometry of the conceptual space can be fully defined by the inner product of the Hilbert space. This has two consequences. First, since for a given function, the Fourier series decomposition generates a vector representation, any number of different vectors can be produced from the same function for each inner product. This means that a given object can be represented in any number of conceptual spaces defined by their inner product.

Conversely, since it can be proven that all Hilbert spaces of the same number of dimensions are isomorphic (Kennedy & Sadeghi, 2013), given a vector, we can then choose an inner product to determine its representation. In this way, each inner product imposes a geometry that allows for different perspectives of the same "raw" data. By analogy, in figure 2.1 we see that coordinates represented on $(x, y)$ in a 2D Cartesian space have a different meaning and location than if they are represented on $(r, \theta)$ in a 2D radial space. This allows for complete flexibility in representation, as an object can be placed in a conceptual space simply by interpreting its vector representation according to the inner product of that space. (Wiggins, 2018)

## 2.3 Information Theory

Shannon's information theory (Shannon, 1948), though originally aimed at providing a theoretical foundation to signal processing and communication, has become foundational in all aspects regarding information. According to information theory, the amount of information in a signal can be thought of as the amount of surprise at seeing a given quantity in that signal. Put more simply, suppose you come in late to work one morning, and upon arriving into the office, your boss tells you about the weather. No information was gleaned from the conversation, since you already knew the weather from being outside. If instead, upon entering the office late, your boss fires you, you would be very surprised, learning something unexpected, and gaining a lot of information.

### 2.3.1 Entropy

Though a concept originally borrowed from statistical physics that represents the amount of disorder in a system (Boltzmann, 1970), entropy can be applied to any probability distribution. If applied to a discrete signal of symbols, it can be used to represent the amount of information of the signal (equation 2.2), by taking it over the probabilities of each symbol $s$ of the alphabet $A$.

$$H(s) = -\sum_{s \in A} p(s) \log_2 p(s) \tag{2.2}$$

### 2.3.2 Information Content

Though entropy is useful in describing the quantity of information of a whole signal, to find the information content $h(s)$ of a single symbol $s$ in an alphabet, one must use equation 2.3 below. Since some symbols are less likely to occur than others, observing them gives more information than more likely ones, which is reflected here.

$$h(s) = -\log_2 p(s) \tag{2.3}$$

Whereas information content looks at a symbol in isolation, conditional entropy looks at symbols in pairs. Conditional entropy is equivalent to the information content of a symbol, given that another symbol is known. If we are referring to a stream of symbols, the conditional entropy would be the information content of a symbol given that we just saw the previous symbol.

$$H(s|t) = -\log_2 p(s|t) \tag{2.4}$$

## 2.4 Information Dynamics of Thinking

Cognitive architecture blending information theory and conceptual spaces.

### 2.4.1 Boundary Entropy Segmentation

In natural language processing, n-gram models are often employed to track the frequency of chains of symbols in a given signal. For instance, if examining textual data, a unigram model would count the frequency of individual words, whereas a bigram model would track the frequency of pairs of words. It is clear that one can then utilize the information content measure on a unigram model and the conditional entropy measure on the bigram model in order to quantify the amount of information in that model. (Sproat, Gale, Shih, & Chang, 1996)

### 2.4.2 Sequential Memory

The sequential memory in IDyOT is represented as a hierarchy of chains of symbols. As the agent perceives a continuous stream of perception from the environment, it first discretizes that stream into moments, and links those discretized percepts as symbols in an ever-growing chain. By examining time-varying information-theoretic properties of this chain, it can be chunked into a series of segments composed of a sequence of symbols. This segment can then be abstracted by a representative symbol in the superior layer and can be said to subsume the inferior segment. This process is done recursively until no more abstraction can be performed.

### 2.4.3 Spectral Representations of Meaning

### 2.4.4 Semantic Memory

If the sequential memory can be though of as how concepts relate to eachother over time, semantic memory can be thought of as how those concepts relate to eachother outside of time. In IDyOT this is modeled using conceptual spaces. At any given layer of abstraction in the sequential memory, there is a "parallel" semantic space for the symbols of that layer. This allow us to think of a segment of symbols in sequential memory as a trajectory of concepts in semantic memory. By looking at a spectral representation of this trajectory, the time-varying properties of the segment are removed, and this spectral representation can be thought of as an abstraction of the segment, and therefore as the abstracted symbol in the superior layer of the sequential memory.

# 3
# Theory & Implementation

## 3.1 Abstraction

When a segment is produced in the segmentation process, it is abstracted to a single symbol in the superior abstraction layer. Since each dimension is composed of both an alphabet of symbols and a conceptual space in which those symbols live, a segment of symbols can be though of as a trajectory through a conceptual space. Viewing this trajectory as a sequence of points in a high-dimensional space, one can abstract out the time variant properties of the signal by viewing it in terms of its spectral properties. When viewed spectrally, this same trajectory can be thought of as a single point, creating a mapping from the segment in the inferior dimension to the a point in the superior abstraction layer.

### 3.1.1 Fourier Transform

In the Information Dynamics of Thinking, the Fourier Transform is used to take the time-domain trajectory of the inferior layer to the frequency-domain point of the superior layer. Though there are other spectral representations of a signal available besides the Fourier transform, there is evidence, especially in the auditory domain, that the brain operates on frequency transformations of time-varying signals (organ of corti), and so it is employed here.

Since the segment is a trajectory of discrete points in a Hilbert space, we use the discrete fourier transform to take the trajectory from time-domain to frequency-domain. It is important to note that the (discrete) Fourier transform is a linear operator on the Hilbert space, meaning essentially that the *shape* of the signal does not change in the tranformation, only its domain. Therefore, use the independence of the frequency bins resultant from the DFT to represent those bins as dimensions in the superior abstraction layer.

### 3.1.2 Tensor Rank Promotion

To clarify this point, consider figure (below). When the time-varying sound signal is transduced at the base perceptual level, it can be represented as a row vector of real values, where each element in the row represents the signal amplitude dimension at that time step. This sounds signal is of finite length, and can be thought of as a moving window of attention. Since it is a linear operator, performing the DFT on this signal results in a row vector of the same length, only now with complex elements in the frequency domain. Since each element of this row vector represents an orthonormal frequency bin, we can alternatively think of this row vector as a column vector, where each row in the column represents a different dimension. This transposition of row vector to column vector is what takes the frequency-domain trajectory to a single point in its dual space.

Now, when we have a trajectory of in this frequency space, instead of operating on a row of real values as in the sound signal, we are operating on a row of complex column vectors, and the trajectory now looks like a matrix. Taking the DFT of this trajectory and transposing it to be represented as a point, the abstracted point is now a matrix. Taking a trajectory of these matrices results in a point represented by a cube (3d rectangle?). Taking a trajectory of these cubes results in a point represented by a hypercube, and so on.

Therefore, when moving from one abstraction layer to its superior, the complex tensor representing the contents of the symbol increases in rank, which is what is meant by tensor rank promotion.

(Figure of tensor rank promotion)

### 3.1.3 Component-wise Fourier Transform

When performing the Fourier transform on a tensor of any rank, it should be noted that each component in the tensor is independent from every other component. Though this is not necessarily true in general, due to the particular hierarchical construction of these spaces, each component is decoupled from the rest. Starting at the bottom, the time-domain sound signal is transformed into a frequency-domain signal of coefficients in independent frequency bins. It is this independence that allowed us to represent it as a point with those frequency bins as dimensions. Performing the DFT on the trajectory of column vectors results in independent frequency bins filled with column vectors with independent entries, meaning all components are independent one another.

Recursively performing the DFT on tensors with independent components results in higher rank tesnors with independent components. This component-independence of the tensor means that the DFT should be taken component-wise, since all other cross-component terms would involve orthgonal components, resulting in zero terms.

## 3.2 Segmentation

Segmentation is the process of dividing an input stream into discrete chunks according to a difference function.

### 3.2.1 Difference Function

The difference function operates on the stream of symbols entering a given dimension and decides where segments begin and end. Though there are a variety of ways in which to determine where to cut in this chain of symbols, here we will look at their information-theoretic properties to determine a meaningful cut. Primarily, we will look at moving entropy of the signal. If the entropy rises, then a cut should be a made, marking the end of the current segment and the beginning of a new one.

Since entropy represents the amount of uncertainty at what comes next, it makes sense that a jump in entropy would mark the beginning of a new segment. For instance, at the beginning of a sentence, entropy is high because the listener has little idea what the speaker will say next. As the sentence proceeds, the speaker will be better able to predict what comes next, meaning entropy is decreasing until the end of the sentences. Once the sentence is finished, the listener again is less sure what will come next, and so entropy rises. Therefore, at this rise in entropy, we would make a cut, resulting in a segment naturally representing the sentence just spoken.

This segmentation process results in two problems stemming from the two types of sparsity inherent in the segmentation and subsequent abstraction.

### 3.2.2 Symbol Sparsity

The first kind of sparsity arises from the length of the segment produced in the segmentation process. Since the abstraction of the segment is its spectral transform, and we use the Discrete Fourier Transform (see section ?) to find the representative symbol in the superior abstraction layer, we run into precision problems due to the uncertainty principle of signal sparsity. When performing a DFT, the signal precision is limited by the number of non-zero coefficients in either the time or frequency domain. Therefore, if the number of symbols in a given segment is small, its spectral representation will be imprecise. Therefore, it is necessary to perform interpolation (see section ?) to fill out the signal, so that high precision is maintained in the spectral transformation.

### 3.2.3 Content Sparsity

The second kind of sparsity is due to the nature of the symbol contents. Since in the case of audio perception, the content of a symbol is a high-dimensional tensor of complex coefficients, not only are the possible values of each dimension uncountably infinite, there are a high number of uncountably infinite dimensions for each symbol. Therefore, it is incredibly unlikely that any two symbols have exactly the same value for every dimension in the conceptual space in which they live.

This poses a problem for determining the entropy and information content of symbols, which rely on the probability of a symbol. If every symbol is unique, in the limit, the probability of seeing one in the signal is 0, meaning the entropy and information content each symbol in the signal would be 0. Really, the the probability of a unique event is meaningless, and therefore, even talking about its entropy is also meaningless.

Though the space in which the symbols live is so sparse, we would still like to say that if two symbols are close enough together, for all intents and purposes (intensive porpoises), they are the same symbol. This is accomplished through categorization (see section ?), where a label is attached to each symbol according to its category. If two symbols have the same label, then they are equal, even though they may have different contents, that is, different values in their complex tensors.

### 3.2.4 Symbol Generation

When a segment is cut from the stream of symbols in a given dimension, a new symbol is generated that represents that segment in the superior abstraction layer. At least in the case of all dimensions supervenient on audio perception, this means that the abstracted symbol is the spectral transformation of the segment.

## 3.3 Categorization

### 3.3.1 Information Content Reduction Criterion

Since the goal of an IDyOT is to be as information-efficient as possible in its representation of concepts, the primary way to do this is categorize two different symbols together if they lead to an overall reduction in information content of the space.

However, if this reduction by information content measure was the only method used to determine categories, there would be nothing to stop all symbols from being categorized together. If all the symbols are the same the information content is maximally reduced, but the result is a meaningless stream of monotony. Obviously, this is unrealistic and undesirable.

### 3.3.2 Categorical Convexity Criterion

To push back against the reduction by information content is the categorical convexity criterion. In section (?) we saw that categories in conceptual spaces are convex regions of the space, which translates to an infininite-dimensional hyperellipsoid in the corresponding Hilbert space. What the convexity criterion guarantees is that for any two symbols in a given category, there is no symbol from a different category between those two symbols.

### 3.3.3 Betweenness Relation

Though this criterion is simple in formulation, the definition of what "between" actually means can vary greatly depending on the space in question. Even when the space is unidimensional, the definition of between is somewhat arbitrary. For instance, take the space that is wrapped aorund a circle. Any point is between any other two points, depending on which direction around the circle you move. Things get even less clear when moving into higher dimensions, where oftentimes, only a partial ordering is possible.

Therefore, instead of looking at betweenness at all, we instead incrementally build up categories by way of an inclusion radius around each point. If another point falls within the inclusion radius, those poiints are categorized together. In this way, we can ensure that ther eis never an interloper in a category, since if it was intruding on the region of the category, it would already be a member.

### 3.3.4 Other Categorization Schemes

Chinese Restaurant Process is a nonparametric bayesian method that allows for unbounded clustering of new points in a space. Since for any given space, the number of "natural" categories is unknown, or perhaps unbounded, it makes sense to use a process that has the ability to add more categories as more data is seen in a given space, in the same way that humans are able to see more nuance the more trained they are in a given subject.

Locality-Sensitive Hashing allows for only a portion of the space to be searched for candidate categories. Since points are only categorized together when they are close to eachother, there is no need to examine distant points as candidates. Locality-sensitive hashing allows us to quickly determine what symbols are similar to a target symbol, and check if they should be categorized together. This is the same as how humans are able to immediately sense similarity, though may require closer inspection to see if two subjects are actually of the same kind.

## 3.4 Interpolation

When talking about trajectories through conceptual spaces, we refer to interpolation as the use of virtual concepts that factor into the abstraction to the superior space. Since we formalized conceptual spaces as Hilbert spaces, we can analogously talk about drawing a curve or curve through points in the space. Hence, the process of interpolation is simply sampling from a regression.

### 3.4.1 Symbol Sparsity

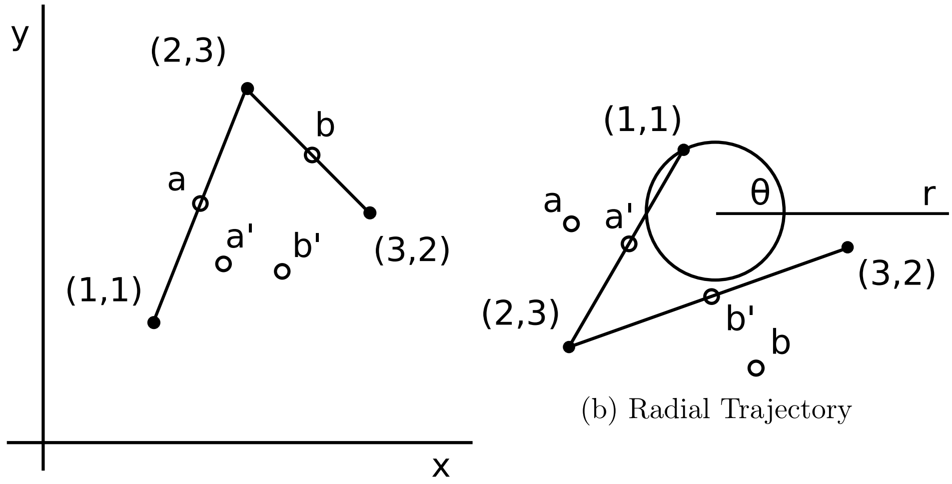
As discussed in section (?) on segmentation, interpolation is necessary to deal with symbol sparsity in a given segment. To reiterate, this problem arises due to the use of the Discrete Fourier Transform in the abstraction process. Since the precision of a signal, either in the time or frequency domains, is determined by the number of non-zero coefficients describing that signal, if the segmentation process results in a relatively short segment, the number of coefficients will be small, and the resulting transformation imprecise.

The way to deal with this imprecision is fill in coefficients in an informed way such that the transform of the signal accurately represents the original signal while maintaining higher precision. This is valid because we can think of the sparse signal as a sequence of samples from a function, and by intelligently regressing on those samples, we can obtain the representative function. By interpolating, we are simply drawing more samples from that regressed function to deal with the precision limits of the DFT.

### 3.4.2 Spatial Geometry

Though interpolation is necessary primarily to deal with symbol sparsity of a segment, it is also the main vehicle in which the geometry of the conceptual space induces the path of the trajectory of symbols. Since in the Information Dynamics of Thinking, the interpretation of any space is determined by its inner product, we need to interpolate through that space in order to "see" the geometry.

For instance, suppose we have three 2D symbols of which we want to draw a trajectory through. Consider two different spaces for conceiving those points: Cartesian (x, y) and Radial (r, $\theta$). Suppose we take the DFT of three points, i.e. as a trajectory without interpolation, in each of the two spaces. Since the values of each point are exactly the same for both spaces, when examining the transform, there would be no way to tell which space it came from. If instead we draw lines through the points and interpolate




(a) Cartesian Trajectory

(b) Radial Trajectory

Figure 3.1: Example of effect of spatial geometry on interpolation

before taking the transform, we see clearly that the interpolated points differ between the two spaces, and so the transform of each trajectory will also be different. In this way, the information about the geometry of the space can have an effect on the abstraction of the trajectory in that space.

### 3.4.3 Types

Since our conceptual spaces are represented as Hilbert spaces, interpolating a trajectory through a sequence of concepts is equivalent to regressing through the points of the space and sampling from that regression. As such, we can employ any number of regression techniques from the literature to determine an appropriate trajectory.

#### Connect the Dots

Though not really regression per say, the simplest method for interpolating between points would be to just draw a straight line between them and sample from that line. This method has the benefit of being very simple, and may prove to be near-equivalent to other methods if the number of interpolated points is low enough.

The problem with this method is that it is not continuously differentiable and therefore quite "sharp". This would manifest in the DFT by resulting in high amplitudes at higher frequencies. It is difficult to say a priori if this is desirable, but it seems in general that we would prefer smoother functions through these points to avoid this effects.

### Gaussian Regression (Kriging)

One method for creating a smooth curve through these points is Gaussian Regression. By thinking of the sequence of points as a multivariate Gaussian distribution, we can think of a distribution of trajectories or functions, i.e. a process, through these points. By finding the mean function of this process we can find the maximum a posteriori trajectory of the process, resulting in the highest likelihood function through a sequence of points. This method automatically avoids problems of overfitting inherent in higher-order transformation linear regression methods, though has the problem being computationally expensive.

By utilizing the square exponential kernel for Gaussian regression, the resulting curve will not only by the MAP curve, but also infinitely differentiable, meaning that it is hella smooth. This smoothness results in a reduction of large high frequency amplitudes that could potentially happen in simpler methods like Connect-the-dots.

# 4
# Empirical Analysis

## 4.1 Methodology

### 4.1.1 Parameterization

### 4.1.2 Data

### 4.1.3 Experiments

## 4.2 Results

### 4.2.1 Sequential Memory Category Flow

### 4.2.2 Semantic Memory Spatial Plot

### 4.2.3 Categorical Consistency Matrix

# 5
# Evaluation & Discussion

# 6
# Conclusion

## 6.1 Contributions

## 6.2 Limitations

## 6.3 Future Work

### 6.3.1 Inner Products

### 6.3.2 Spectral Projectors

# A
# Appendix

## References


Bernkopf, M. (2008). *Schmidt, erhardt.* Retrieved 2019-07-12, from `https://www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/schmidt-erhard`

Boltzmann, L. (1970). Weitere studien über das wärmegleichgewicht unter gasmolekülen. In *Kinetische theorie ii* (pp. 115–225). Springer.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought.* MIT press.

Kennedy, R. A., & Sadeghi, P. (2013). *Hilbert space methods in signal processing.* Cambridge University Press.

Kirsh, D. (1991). Today the earwig, tomorrow man? *Artificial intelligence*, *47*(1-3), 161–184.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

Sproat, R., Gale, W., Shih, C., & Chang, N. (1996). A stochastic finite-state word-segmentation algorithm for chinese. *Computational linguistics*, *22*(3), 377–404.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the turing test* (pp. 23–65). Springer.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton university press.

Wiggins, G. A. (2018). Creativity, information, and consciousness: the information dynamics of thinking. *Physics of life reviews*.