Sampling Enhanced Reverse Correlation for the Reconstruction of Latent Representations

by

Benjamin William Roop

A Thesis
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Master of Science
in
Neuroscience

May 2022

APPROVED:

Professor Adam C. Lammert, Advisor

Professor Songbai Ji, Committee Member

Professor Ali Yousefi, Committee Member, Thesis Reader
ABSTRACT

Internal representations are widely characterized using reverse correlation, a technique capable of producing unconstrained estimates of the representation, all on the basis of subjects’ responses to random stimuli. However, employing reverse correlation often entails collecting thousands of stimulus-response pairs, which severely limits the breadth of studies that are feasible using the method. Current techniques to improve efficiency bias the outcome and ultimately limit the truth to the reconstruction. Here, three techniques are used to increase the efficiency of reverse correlation. Stimulus whitening, a statistical procedure that decorrelates stimuli, provides greater than 85% improvement in efficiency for a given estimation accuracy and a two- to five-fold increase in accuracy for a given sample size. Compressive sensing, an advanced signal processing technique designed to improve sampling efficiency based on sparsity, improves the accuracy of reconstructed cognitive representations and dramatically reduces the required number of stimulus-response pairs in both simulations and on human subject response data. Autoencoders, a type of artificial neural network, increase the reconstruction quality to a level that necessitates only collecting 10% of the previously needed samples. Improving the efficiency of reverse correlation in such ways may enable a broader scope of investigations of perceptual mechanisms and could improve representation reconstruction throughout the field of neuroscience and beyond.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Adam Lammert, for guiding me through this process and my entire academic journey at WPI. Thank you to my thesis committee, Professor Ali Yousefi and Professor Songbai Ji, for providing advice, expertise, and invaluable feedback.

Thank you to Professor Ben Parrell and Alexis Compton for their contribution to this work, and thanks to Apiwat Ditthapron for insight into deep neural networks at this project’s onset. Thank you to Professor Jagan Srinivasan and Dean Jean King for organizing the WPI Neuroscience Program; it has been everything I wanted from graduate school and more. Finally, thank you to my family and friends for providing me with motivation while still keeping me grounded to the important parts of life; I love you from Australia to Mars, and back again.
# TABLE OF CONTENTS

A. Abstract........................................................................................................... ii
B. Acknowledgements.......................................................................................... iii
C. Table of Contents............................................................................................... iv

1. Introduction........................................................................................................ 1
2. Methods............................................................................................................. 9
   2.1. Reverse Correlation....................................................................................... 9
       2.1.1. Modelling Subject Responses............................................................... 9
       2.1.2. Reverse Correlation Reconstruction................................................. 10
   2.2. Compressive Sensing.................................................................................. 12
       2.2.1. Background.......................................................................................... 12
       2.2.2. Compressive Sensing Reconstruction............................................... 12
       2.2.3. Simulating Gosselin & Schyns (2003).............................................. 14
       2.2.4. Reverse Correlation and Compressive Sensing at Varied $n$ and with noise...... 14
       2.2.5. Evaluation of Compressive Sensing on Human Response Data............. 15
       2.2.6. Cross-Validation of Reconstructions from Human Response Data........... 16
   2.3. Whitening Stimuli...................................................................................... 17
       2.3.1. Justification......................................................................................... 17
       2.3.2. Whitening Procedure......................................................................... 18
       2.3.3. Evaluation of Reverse Correlation with Whitened Stimuli...................... 19
   2.4. Compressive Sensing Combined with Whitening......................................... 19
   2.5. Artificial Neural Network Enhancement of Reverse Correlation Reconstructions... 20
       2.5.1. Justification........................................................................................ 20
       2.5.2. Data Used.......................................................................................... 20
       2.5.3. Making Reconstructions of MNIST Digits........................................... 20
       2.5.4. ANN Construction.............................................................................. 21
       2.5.5. ANN Hyperparameter Tuning............................................................ 22
           2.5.5.1. MLP Autoencoder........................................................................ 22
           2.5.5.2. Stacked MLP Autoencoder......................................................... 23
           2.5.5.3. Convolutional Autoencoder......................................................... 24
       2.5.6. Evaluation of ANNs Trained on Reverse Correlation Reconstructions........ 27
       2.5.7. Evaluation of ANNs Trained on Compressive Sensing Reconstructions...... 27
3. Results............................................................................................................. 29
   3.1. Compressive Sensing Reconstructions of the Letter “S”............................... 29
   3.2. Simulating Noisy Subject Responses.......................................................... 29
   3.3. Compressive Sensing on Human Data......................................................... 30
   3.4. Cross Validation of Reconstructions............................................................. 32
   3.5. Whitened Data.......................................................................................... 32
   3.6. Whitening with Compressive Sensing........................................................ 33
   3.7. Evaluation of ANNs Trained on Reverse Correlation Reconstructions........... 33
       3.7.1. MLP Autoencoder............................................................................. 33
       3.7.2. Stacked MLP Autoencoder............................................................... 35
       3.7.3. Convolutional Autoencoder.............................................................. 36
3.8. Evaluation of ANNs Trained on Compressive Sensing Reconstructions

3.8.1. MLP Autoencoder

3.8.2. Stacked MLP Autoencoder

3.8.3. Convolutional Autoencoder

4. Discussion

4.1. Compressive Sensing

4.1.1. Compressive Sensing Reconstructions of the Letter “S”

4.1.2. Robustness of Compressive Sensing to Noise

4.1.3. Compressive Sensing on Human Data

4.1.4. The Potential Impact of Compressive Sensing on Reverse Correlation

4.2. Whitening

4.2.1. Whitening Stimuli in Reverse Correlation

4.2.2. Whitening with Compressive Sensing

4.3. Artificial Neural Networks

4.3.1. ANNs for the Improvement of Reverse Correlation Reconstructions

4.3.2. Hyperparameter Tuning

4.3.3. The Potential for ANNs in Reverse Correlation Research

4.3.4. Filtering Reconstructions

4.4. Conclusion

5. References

6. Appendix
1. INTRODUCTION

One of the fundamental goals in neuroscience research is understanding how neural systems will respond when presented with arbitrary stimuli. Reverse correlation is a technique used for this purpose that has held an important place in neuroscientific studies for over half a century, dating at least as far back as De Boer and Kuyper in 1968 [3]. Reverse correlation’s origins in studying the brain began with transductive neuroscience, in which the neuron is modeled as a transducer – i.e., its output changes purely as a function of the input stimulus, much like a linear transducer [4]. Based on this model of the neuron, reverse correlation proceeds by inducing a measurable response to a controlled stimulus, and then using the relationship between the stimulus and response to infer a mathematical model of the transductive process itself. Use of this technique in neuroscience was first proposed for studying the auditory system [3], where the electrical response of cochlear cells to auditory stimulation were used to calculate the signal transformation within the inner ear [5]. Subsequent reverse correlation studies mapped response properties of the optic nerve across many species [6].

A critical aspect of reverse correlation is that it allows for unconstrained and unbiased characterization of latent representations directly from stimulus-response data by eliciting responses to richly varying stimuli, such as white noise [4, 7, 8]. For decades prior to reverse correlation, scientists studying the visual domain had made subjective descriptions of how visual neurons responded to stimuli, or they published images of spike trains without much mathematical analysis [9-12]. The combination of reverse correlation with neural recording led to a revolution of characterizing visual neurons [13, 14]. The response properties mapped with this approach became known as a spike triggered average (STA) [15, 16]. Computing the STA involves recording the activity of a neuron or group of neurons in response to a stimulus or series
of stimuli [2]. In pioneering work for understanding mammalian vision, cortical neurons recorded responding to grating images or to visual noise helped to map response properties in the visual cortex, be those properties firing patterns, orientation selectivity, or edge detection [17-21]. Although reverse correlation continues to be used to analyze neural spike trains [22], recent work in neurophysiology has also used reverse correlation for such tasks as finding the pattern of electrical activity in water that triggers a catfish’s electroreceptors [23], mapping cortical responsiveness based on signals collected with electrodes in the brains of epilepsy patients [24], or determining the navigational systems in fly larvae [25].

Reverse correlation has also been applied to domains beyond neurobiology. In a technique analogous to STA called pulse-triggered averaging, pulses in gene expression were analyzed over time to determine the relationship between two proteins [26]. Additionally, reverse correlation was used to infer how stimuli are represented within artificial neural networks.

Figure 1 Example spike-triggered average calculation. In this example, the STA is reconstructed from the average of segments that trigger spikes (the red boxes) minus the full stimulus. t indicates time. Taken from [2].
(ANNs) and other machine learning classifiers, in an approach analogous to receptive field analysis sometimes known as activation maximization [27-29].

Using reverse correlation for neurophysiological investigations, such as inferring the latent spatial and temporal receptive fields with STA, comprises the bottom-up portions of perception. However, reverse correlation has also emerged as a useful tool in cognitive neuroscience and neuropsychology for studying top-down mechanisms of perception [7, 30, 31]. In the cognitive domain, a person’s primed response to a stimulus (e.g., “do you see a face in this image?”) can be used to reconstruct their internal representation of the object for which they were primed (e.g., a face) [32].

As with neurophysiology [5], reverse correlation began to be used in the cognitive domain in studies of the auditory system [31]. Subjects were played either pure auditory Gaussian noise or a tone overlaid with the noise and were asked if they heard a tone. For each trial, a subject would reply yes or no to whether they heard a tone. In trials where only noise was played, subjects compared their own latent representation of the tone to the noise, and, if the noise matched closely enough, they would give a “yes” response. This was repeated over thousands of trials, and the sum of the noise signals that elicited a “no” response was then subtracted from the summed “yes” signals to generate what was to represent the subject’s internal representation of the tone. Other work has sought to reconstruct representations of phonemes [33] and determine the trustworthiness in voices [34].

Reverse correlation research extends to visual cognition as well (Figure 2). Gosselin & Schyns were able to discover a subject’s latent representation of the letter “S” through reverse correlation [35]. In this pioneering study, the subjects were shown a white noise image and were asked to report whether or not they saw an “S” in the image. As with representations of tones in
auditory studies, each subject has a latent representation image of what an “S” looks like. Although subjects were never presented with an S-containing image for a stimulus, only pure white noise, the resulting image, often referred to as a classification image, strongly resembled an “S”. Reverse correlation can be used to generate classification images ranging from letters to faces [1, 36]. Further studies have looked at more abstract representations, such a behavior [36] or one’s own self-image [37].

Figure 2 Cognitive reverse correlation experimental method. This example illustrates the experimental design of a reverse correlation task in which the subject is asked to identify whether they see an “S” in a visual noise stimulus. After many trials, the stimuli in which an “S” was not detected can be subtracted from the stimuli in which an “S” was detected to generate the classification image of the subject’s internal representation of an “S”. Adapted from [1].

A major problem with reverse correlation studies is the large number of trials required to generate an adequate reconstruction of the latent representation, be that representation a receptive field or an internal image of an “S”. For example, Gosselin & Schyns’ 2003 experiment required 20,000 trials from each participant [35]; another study classifying patterns required 11,400 trials [38]. In cognitive and psychological experiments, where each trial can take up to several seconds to complete, a large number of trials may require infeasibly lengthy data collection protocols that can last weeks or months for each individual subject. This inefficiency limits the feasibility of applying reverse correlation to only those experimental protocols where
subject participation and motivation can be maintained over long timelines. Even in such cases, the large amount of data required from each participant means the number of participants in a given study is typically very low (usually less than five \([1, 35, 39]\)). This severely limits any insight into the generalizability of existing findings to a broader population.

Existing methods attempt to overcome this fundamental limitation of reverse correlation by biasing the input stimuli to decrease the number of required trials. This can be accomplished, for example, by generating stimuli through adding noise to known examples of some real-world category (e.g., an image of a face) rather than using pure noise, as done by \([37]\). This increases the probability that subjects will report the stimulus as belonging to the category of interest but also biases the results toward representations consistent with the initial example. In order to realize the full impact of reverse correlation for uncovering cognitive representations, it is necessary to reduce the number of trials required without introducing such biases.

The first contribution of the present work is to demonstrate a method to condition the stimuli without biasing them towards the target. Normally, stimuli that are correlated with each other may reduce the number of effective trials by presenting overlapping or redundant information. Random stimuli will tend to have a small but non-zero degree of correlation, especially under the conditions in which reverse correlation is typically applied – i.e., many stimuli that are low- to moderate-dimensional in size. Using a standard statistical procedure called whitening, stimulus images are generated that are not correlated with one another, meaning that there is less redundancy in the information gained from each stimulus.

An alternative approach for improving the efficiency of reverse correlation is to impose constraints on the reconstruction process. Adding constraints can be a powerful way to improve reconstruction quality but can also severely limit the richness of the reconstruction in ways
prespecified by the constraints themselves. For example, current approaches constrain reconstructions by smoothing with a 2D Gaussian kernel [1] or, more commonly, through low-pass filtering [35]. While this can eliminate unwanted high-frequency noise from the reconstruction, it also presupposes that such high-frequency information does not form part of the underlying representation. For this reason, it is critical to assess the assumptions inherent in any added constraint and to be careful in deciding whether to utilize that constraint [40, 41].

One constraint that seems appropriate for reconstructing perceptual representations is that of sparsity. Sparsity is the notion that representations are composed of a finite and relatively small set of essential features. Importantly, sparsity is widely considered to be a fundamental principle of organization in perceptual systems at all levels [42], with strong empirical and theoretical support [42-45]. Thus, sparsity is a fairly innocuous assumption that is highly likely to preserve the essential aspects and important variation in cognitive representations. This assumption can also be exploited to improve estimation of internal representations [40]. Mineault and colleagues incorporated sparsity constraints in a binary logistic regression model to reconstruct visual-domain representations more efficiently, demonstrating an approximately 80% reduction in the number of trials to produce an equivalent quality reconstruction [40].

Recent advances in signal processing have resulted in a proliferation of efficient sampling methods based on the idea of sparsity, collectively known as compressive sensing [40, 41]. To date, however, compressive sensing has not been applied to the reverse correlation paradigm. The second contribution of the present work is to explicate the underlying mathematical connections between reverse correlation and compressive sensing and to provide a demonstration of the potential for dramatic efficiency improvements using the combined method. It will also be shown that, unlike previous approaches to incorporating sparsity constraints, it is possible to
obtain reconstructions from compressive sensing analytically, using a closed-form solution that is highly efficient and guaranteed optimal. One such method in [46] was adapted as part of this work and is presented here and in [47] in detail.

Critically for the applicability of compressive sensing for reverse correlation, it has been demonstrated both theoretically [48, 49] and empirically [40] that using random (e.g., white noise) stimuli to elicit responses is a highly effective way to ensure accurate reconstruction of latent representations within the compressive sensing framework. Moreover, a significant portion of the substantial literature on compressive sensing has focused on the problem of inferring representations from binary responses (e.g., yes-no), a variation of classical compressive sensing called 1-bit compressive sensing [46, 50, 51]. In other words, the somewhat unusual model of sampling assumed by compressive sensing maps directly onto the reverse correlation paradigm. This opens the door to potentially providing all the efficiency benefits of compressive sensing to reverse correlation.

The third contribution of the present work is to explore how the reconstructed images can be enhanced after reconstruction. In reverse correlation, the goal of performing more trials is to reduce the signal-to-noise ratio between the reconstruction and the internal representation [52]. The more trials that are performed (and, thus, the more stimuli that are presented), the lower the noise will be, and the reconstruction will resultantly have a higher correlation with the internal representation. As mentioned above, some currently used approaches filter the reconstructions to reduce the noise [1, 35, 53], but this filters out potentially relevant information. Outside of the reverse correlation paradigm, other approaches have sought to denoise images through the use of shrinkage fields [54] or total variation regularization [55], to name a few. However, the most successful modern image denoising approach is with artificial neural networks (ANNs) [56, 57].
Artificial neural networks (ANNs) are currently the leading approach for denoising images [56]. One popular variety of deep ANN for image denoising is an autoencoder [58]. Autoencoders are an ANN composed of three parts: an encoder, a coding, and a decoder [59]. The encoder converts a high dimensional input to a lower-dimensional representation, known as the coding. The decoder then generates a new output by converting the coding back into a high dimensional output. Autoencoders are generally symmetrical so that the input to the encoder is the same size as the output from the decoder, and multiple layers can comprise both the encoder and decoder in what is called a stacked autoencoder. Here, autoencoders are used to denoise noisy reconstructions of representations that were generated with few trials, with the goal of creating a noise-free output that resembles a reconstruction generated from many trials. The goal of the autoencoders is to generate a denoised version of the reconstruction. The network output can then be compared to reconstructions generated from many trials to evaluate the ability of autoencoders to transform low-trial reconstructions to high-trial reconstructions.

In this work, a set of computational simulations are used to model reverse correlation in subject responses to whitened or unwhitened visual stimulation. Internal representations are reconstructed based on these simulated responses using either the traditional reverse correlation approach or compressive sensing, and compressive sensing improves both the efficiency of the sampling process and the quality of the reconstructions. Next, compressive sensing is applied to human subject response data from a prior reverse correlation study ([1]), again demonstrating the ability of compressive sensing to enhance reverse correlation experiments. Furthermore, combining whitened stimuli with compressive sensing is shown to produce better reconstructions than using either technique alone. Finally, ANNs are constructed that are capable of further enhancing the simulated reconstruction quality.
2. METHODS

2.1 Reverse Correlation

2.1.1 Modeling Subject Responses

Reverse correlation studies generally model subject behavior with a linear observer model [41]. In this model, responses (an n-by-1 vector, \(y\)) are a result of linear comparisons between a latent representation (\(x\)) and a collection of random stimuli (\(\Phi\)):

\[
y = \text{sign}(\Phi x + \epsilon),
\]

where the sign function causes each value in \(y\) to be either a 1 for a yes response or a -1 for no. The latent representation \(x\) is reshaped into a m-by-1 vector, and \(\Phi\) is a n-by-m "measurement matrix" of stimuli that are presented. The values in \(\Phi\) are drawn randomly from a normal distribution with mean zero and variance one. The error term \((\epsilon)\) represents the irreducible error from noise in the subject response process [41]. The dimension \(n\) corresponds to the number of trials that will be performed, and the dimension \(m\) represents the size of the latent representation and of each stimulus.

Here, this model was modified slightly to simulate subject responses to random stimuli. Because it is irreducible, the error term is ignored. The decision process is further modified as a similarity calculation such that it is

\[
y_n = \begin{cases} -1, & \Phi_n x \leq 0 \\ 1, & \Phi_n x > 0' \end{cases}
\]

where \(n\) is the row of the measurement matrix. In simulated reverse correlation studies such as this, the threshold for the decision process in Equation 2 can be set so that the simulated subject
is biased towards choosing one response more than the other [41]. Here however, a threshold of zero is used.

2.1.2 Reverse Correlation Reconstruction

The subject response model in Equation 1 is inverted in conventional reconstruction of the latent representation (e.g., [35]), in a method closely related to regression of the responses against the stimuli, and similar to that used in the spike-triggered averaging (STA) approach to uncovering neural tuning (e.g., receptive fields):

$$\hat{x} = (\Phi^T \Phi)^{-1} \Phi^T y.$$  [3]

This equation is commonly simplified, using the knowledge that the stimuli in typical reverse correlation experiments are uncorrelated, making it unnecessary to calculate the full Sum of Squares and Cross Products matrix, and instead simply normalize by the number of responses gathered:

$$\hat{x} = (1/I) \Phi y,$$  [4]

where $I$ is the identity matrix. This process essentially amounts to an averaging procedure, whereby the mean of stimuli eliciting a no (-1) response is subtracted from the mean of stimuli eliciting a yes (1) response. The reconstruction process is presented graphically in Figure 3a and 3b.

The reconstruction accuracy of $\hat{x}$ was assessed as the two-dimensional correlation coefficient between the latent representation ($x$) and the estimated latent representation ($\hat{x}$), such that:
\[ r(x, \hat{x}) = \frac{\sum m(x_m - \bar{x})(\hat{x}_m - \bar{\hat{x}})}{\sqrt{\sum m(x_m - \bar{x})^2(\sum m(\hat{x}_m - \bar{\hat{x}})^2)}} \]

where \( \bar{x} \) is the mean of \( x \), and \( \bar{\hat{x}} \) is the mean of \( \hat{x} \). Correlations were reported as \( r^2 \) values.

**Figure 3** Response model and template reconstruction with reverse correlation and compressive sensing. (A) In reverse correlation, the vector of subject responses is modeled as resulting from the multiplication of a latent representation vector \( (x) \) and a stimulus matrix \( (\Phi) \), which can be thought of as a similarity calculation between the latent representation and a vector representation of each presented stimulus. (B) An estimate of the latent representation \( (\hat{x}) \) is then reconstructed by regressing responses against the stimuli. (C) In compressive sensing, the vector of subject responses is modeled as resulting from the multiplication of a sparse latent representation vector \( (s) \) and a compressive sensing matrix \( (\Theta) \). The compressive sensing matrix is formed by multiplying a matrix of basis functions by the stimulus matrix \( (\Theta=\Phi\Psi) \), which amounts to a similarity calculation between the stimuli and the known basis functions. (D) An estimate of the sparse latent representation \( (\hat{s}) \) is then reconstructed by regressing the responses against the compressive sensing matrix, soft-thresholding the resulting regression coefficients, and then normalizing by \( \zeta = ||P(m^{-1}\Theta^Ty)|| \), where \( m \) is the dimensionality of the stimulus vector, and \( P \) is a soft thresholding function. The full representation estimate \( (\hat{x} = \Psi\hat{s}) \). Note that the response vector \( y \) in compressive sensing is generally assumed to contain many fewer entries than in reverse correlation, and without sacrificing reconstruction accuracy.
2.2 Compressive Sensing

2.2.1 Background

Compressive sensing begins with the assumption that signals of interest, which can include latent cognitive representations, are sparse or “compressible”. This means specifically that they can be represented by a small number of functions from an appropriately-selected basis set. If one assumes that responses stem from a process of comparing stimuli to the latent representation, it is possible to estimate the latent representation using only a small number of measurements by acquiring the basis function representation directly via sparse optimization approaches to find the sparse representation. In practice, sparse representations can be found even when the chosen basis domain is quite general and incorporates no specific prior knowledge of the signal characteristics, such as the discrete cosine transform or wavelet transform.

2.2.2 Compressive Sensing Reconstruction

Compressive sensing uses a closely related model of subject behavior as is used above but assumes that the latent representation is sparse in some basis domain; that is, the latent representation is compressible. The objective of compressive sensing here is to estimate the sparse representation directly, which will be expected to be more efficient than reconstructing the full representation itself. Given this assumption, the model of subject behavior becomes:

\[ y = \text{sign}(\Phi \Psi s), \]

where \( s \) is the sparse representation, and \( \Psi \) is a matrix of orthonormal basis functions, specifically the 2D discrete cosine basis. It is helpful to calculate the similarity between the stimuli and the basis functions via matrix multiplication, producing a matrix of such similarity (i.e. inner product) values, \( \Theta \), which then leads to:
\[ y = \text{sign}(\Theta x). \]

The matrix \( \Theta \), sometimes known as the "compressive sensing matrix", is \( m \)-by-\( n \) and can be used as the basis for estimating \( s \). The process of estimation has been an active area of study since the beginning of compressive sensing research. Many compressive sensing algorithms have been proposed, depending on assumptions about the nature of the responses, the error criteria to be optimized, and the method of optimizing a given criterion. In the present work, the method of [46] is used, which is intended to operate on binary, yes/no responses and features a simple calculation and closed-form solution for finding an estimate of \( s \):

\[
\hat{s} = P\left( \frac{1}{m} \Theta^T y \right),
\]

where \( P(\cdot) \) is a soft thresholding operation that zeros all but the top 64 elements of the vector, and \( m \) refers to the height of \( \Theta \). The value of 64, referred to as \( \gamma \) in [46], was determined through cross-validation; however, \( \gamma \) could be any integer less than \( m \). Because \( \Theta \) is \( m \)-by-\( n \) and \( y \) is \( n \)-by-1, \( \hat{s} \) will be \( m \)-by-1.

Once \( s \) has been estimated (as \( \hat{s} \)), it can be used in conjunction with \( \Psi \) to reconstruct the latent representation \( x \) using the known basis functions:

\[
\hat{x} = \Psi \hat{s}.
\]

Classical compressive sensing assumes measuring of continuous response values. However, compressive sensing of binary responses (termed "binary compressive sensing" or "one-bit compressive sensing") has also been studied extensively, leading to the reconstruction process
incorporated in the framework used here [46] and in several other techniques [51]. The compressive sensing reconstruction process is presented graphically in Figure 3c and 3d.

2.2.3 Simulating Gosselin & Schyns (2003)

To evaluate the accuracy of the reconstructions generated with conventional reverse correlation and with compressive sensing, simulations were run in which there was a known template latent representation (x), which was used to generate simulated subject responses using the method outlined in section 2.1.1. These simulated responses were then used to estimate the latent representation (x̂) using either reverse correlation or compressive sensing. Because the entire simulation process is known, x̂ can be directly compared to x.

The template latent representation, x, was generated by recreating the “superstitious” perception of “S” for subject NL from Gosselin & Schyns’ original work [35]. The “S” was recreated by horizontally scaling a lowercase “S” in the Verdana font, as described in the original manuscript. This “S” was converted to grayscale and resized to 50x50 pixels using MATLAB (R2021a, The MathWorks, Inc.). Random stimuli (Φ) were generated by drawing 1,000 or 10,000 random values from a normal distribution. Subject responses to the stimuli were simulated as described in section 2.1.1. Reconstructions (x̂) were generated from the responses using conventional reverse correlation (section 2.1.2) and reverse correlation with compressive sensing (section 2.2.2). Correlation was calculated between the reconstruction and x for both techniques using Equation 5.

2.2.4 Reverse Correlation and Compressive Sensing at Varied n and with noise

Human subjects have internal noise that makes their responses vary from trial to trial partly independently of the stimulus [30], and subjects show uncertainty about the spatial
location of the target in the stimulus [60]. To model this in the simulated subject decision process, a percentage of the responses were flipped.

Simulations were conducted using the procedure described in section 2.2.3 but with 1250, 2500, 5000, 10000, and 20000 samples. For each number of samples, eight simulations obtained representation estimates using conventional reverse correlation, and another eight used compressive sensing. For all simulations at a given number of samples, the mean estimation accuracy and 95% confidence intervals were calculated separately for reverse correlation and compressive sensing estimates.

To simulate noisy subject responses, a random subset of specified size $e$ (a percentage of the total number of responses) within the response vector $y$ were switched prior to generating $\hat{x}$; that is, $e$ percent of the yes (1) responses became no (-1) and vice versa. Sixteen additional simulations were conducted at each number of samples with this added noise (eight simulations at $e = 10$ and eight at $e = 20$, each with 1250, 2500, 5000, 10000, and 20000 samples) for both reverse correlation and compressive sensing.

2.2.5 Evaluation of Compressive Sensing on Human Response Data

Human data from Smith, Gosselin, & Schyns [1] was obtained for use in this thesis and related work [47]. In the original paper, five subjects were presented with white noise visual stimuli and asked to identify if they perceived a face in the image; none of the images contained a face. Reverse correlation was used to reconstruct each subject’s internal representation of a face based on the 10,000-15,000 trials per subject. Here, conventional reverse correlation and reverse correlation with compressive sensing were performed on the stimulus-response pairs for each subject. For reverse correlation, reconstructions of the representations were generated by multiplying the matrix of stimuli by the vector of responses, as in Equation 4. For the
compressive sensing analysis, $\hat{s}$ was calculated and multiplied by $\Psi$ to reconstruct the representation, as in Equation 9.

2.2.6 Cross-Validation of Reconstructions from Human Response Data

Unlike in simulations, the quality of the generated reconstruction could not be directly compared to the template because the template was unknown. To assess the quality of the simulation estimation procedure, human data from each of the five subjects (S1-S5) from [1] was used for a five-fold cross-validation. For a subject, 4/5 of the stimuli and corresponding responses were used to generate a reconstruction using Equation 4. That reconstruction then served as a template for the simulated response generating procedure (Equation 2) using the remaining 1/5 of the stimuli. That is, the remaining 1/5 of the stimuli were regressed (compared) against the template estimated from 4/5 of the data to generate estimated subject responses. This procedure was repeated for each fifth of the stimulus-response pairs for a subject. The resulting estimated subject responses for all cross-validations were compared to the actual human responses, and balanced accuracy was calculated. The cross-validation was conducted independently on data for each of the five human subjects.

A nearly identical cross-validation was performed on the data to assess the compressive sensing simulation model. The matrix of orthonormal basis functions, $\Psi$, was generated, as outlined in section 2.2.2. For each subject, $\Psi$ was regressed against 4/5 of the stimuli to produce $\Phi$. Using Equation 8, $\Phi$ and 4/5 of the subject responses were used to produce $s$, the sparse weights. The weights were then regressed against $\Psi$ (Equation 9) to generate a compressive sensing estimate of the subject’s internal representation of a face, $\hat{x}$. That reconstruction then served as a template for the simulated response generating procedure; again, the remaining 1/5 of the stimuli were regressed against the estimated template to generate estimated subject
responses, as in Equation 2. This procedure was repeated for each fifth of the stimulus-responses for a subject. The resulting estimated subject responses for all cross-validations were compared to the actual human subject responses, and balanced accuracy was calculated. The cross-validation was conducted independently on data for each of the five human subjects.

A comparison of reconstruction quality from compressive sensing to that obtained from conventional reverse correlation was conducted by performing a hypothesis test on the balanced accuracies produced by cross-validation, and which were described above. Specifically, a paired t-test was performed across subjects under the null hypothesis that the mean difference in balanced accuracies obtained from compressive sensing versus those obtained from reverse correlation was zero. A level of $\alpha = 0.05$ was used as the significance threshold.

2.3 Whitening Stimuli

2.3.1 Justification

In standard reverse correlation, columns of the measurement matrix ($\Phi$) are uncorrelated, allowing for Equation 3 (known as the Normal equation) to be simplified to Equation 4. Despite the columns of $\Phi$ being uncorrelated, the elements within the rows of $\Phi$ may be correlated. That is, a particular point within a stimulus may be correlated with the same point in another stimulus. Having correlated stimuli leads to redundant information by having overlapping points across multiple stimuli. This shortcoming was evaluated by reducing this overlap using stimulus whitening.

Whitening is a well-known statistical procedure that eliminates the covariance in data. Here, the data to be whitened is the measurement matrix, $\Phi$. Whitening eliminates the covariance by transforming the known covariance matrix into an identity matrix. It is called whitening because $\Phi$ is transformed into a white noise matrix. As such, all dimensions of $\Phi$ become
statistically independent. Whitening is commonly used with principal components analysis, but this and related work have been the first to pair whitening with reverse correlation [61]. The intuition behind this is prediction that the correlated stimuli would overlap, in which case multiple stimuli would contain redundant information.

2.3.2 Whitening Procedure

Prior to running the simulations, a whitening matrix \( W \) was defined for the unwhitened \( \Phi \) as:

\[
W = \left( \frac{\Phi \Phi^T}{m-1} \right)^{-\frac{1}{2}}
\]

[10]

where \( C \) is the centering matrix, defined by:

\[
C = I - \frac{1}{m} \mathbf{1}.
\]

[11]

In Equation 11, \( I \) is the identity matrix, and \( \mathbf{1} \) is an all-ones matrix. Using these values, a new, whitened measurement matrix \( \Phi_w \) can be calculated as:

\[
\Phi_w = C \Phi^T W.
\]

[12]

This whitened matrix has the desired property that its rows are now uncorrelated, meaning that the images are uncorrelated.

However, performing matrix inversions, such as is done for whitening, can be numerically difficult when \( n \) is large because of the numerical precision, even when done by a
computer. In a way conceptually similar to ridge regression [62, 63], a slight bias (λ) can be added to the matrix that will be inverted, such that Equation 10 becomes:

\[ W = \left( \frac{\Phi C \Phi^T}{m-1} + \lambda I \right)^{-\frac{1}{2}}. \]

Here, a λ value of 0.001 is used for all simulations. This value was determined through cross-validation. The whitening matrix computed using Equation 13 was used in Equation 12 to produce the whitened measurement matrix, Φ_{\text{w}}, for analysis.

### 2.3.3 Evaluation of Reverse Correlation with Whitened Stimuli

Simulations of Gosselin & Schyns letter “S” experiment were carried out as described in section 2.2.3 using both whitened and unwhitened data. To assess the effect of whitening across varied numbers of trials, n was varied from 1000-5000, in increments of 1000. Correlation was calculated between the reconstructions and the template.

### 2.4 Compressive Sensing Combined with Whitening

To explore if whitened stimuli and compressive sensing could be combined synergistically, further simulations reconstructing the letter “S” were conducted. Correlations of reconstructions were correlated with the template for each four trial groups (standard reverse correlation, reverse correlation with whitened stimuli, compressive sensing, and compressive sensing with whitened stimuli) using a range of low-n, moderate-n, and high-n values, where n is the number of stimuli presented. Low-n values were varied from 100-1000 in increments of 100, and each n was repeated ten times. Moderate-n values were varied from 1000-5000 by 1000, and each n was repeated ten times. High-n values were varied from 1000-21000 by 4000, and each n was repeated four times. Average correlation was calculated at each n for each of the four groups.
2.5 Artificial Neural Network Enhancement of Reverse Correlation Reconstructions

2.5.1 Justification

When comparing the reconstructions generated with reverse correlation and compressive sensing to the template images, the reconstructions subjectively appear to be noisier versions of the templates. Image denoising techniques were therefore used on low-n reverse correlation reconstructions to assess if denoising improves the correlation of the low-n reconstructions with the template. ANNs, particularly autoencoders, have been shown to have the ability to effectively denoise images [64]. Furthermore, the decoding portion of autoencoders are a type of generative network that can be used for image-generating tasks like this. Autoencoders therefore seemed like an ideal candidate to generate denoised reconstructions. Rather than reconstruct just the letter “S” again, digits from the popular MNIST dataset were used as templates for this experiment to increase the number of and variability between templates.

2.5.2 Data Used

The MNIST dataset is a database of handwritten numerical digits collected from U.S. census data [65]. The digits range from 0-9 and are approximately evenly distributed across the ten digits (Figure 4). The dataset has 70,000 images, each with dimensions 28x28. The MNIST dataset is divided into a 60,000-digit training set and a 10,000-digit test set. The dataset was downloaded from its online repository (http://yann.lecun.com/exdb/mnist/).

2.5.3 Making Reconstructions of MNIST Digits

Each of the 70,000 MNIST images were used as a template for reverse correlation simulations using the simulation procedure outlined in Figure 3. Images were converted from size 28x28 to vectors of size 784x1. For each image, the number of trials was randomly chosen between 100-1100, such that one image may have \( n=234 \) while another has \( n=1087 \). For all
images in the training set, reconstructions were generated using both compressive sensing and conventional reverse correlation. Reconstructions were made using both the random \( n \)-number of trials and with 50,000 trials; thus, each image had a low-\( n \) and high-\( n \) reconstruction. For each of the 10,000 images in the test set, a low- and high-\( n \) image were generated using reverse correlation, reverse correlation with whitened stimuli, compressive sensing, and compressive sensing with whitened stimuli. Thus, there were eight reconstructions made for each image in the test set.

2.5.4 ANN Construction

Artificial neural networks were trained and built in Python version 3.8.8 using Keras. Hyperparameter tuning was conducted on both a CPU and GPU. For the CPU, an Intel Core i7 processor with 16GB of RAM was used, and GPU computation was run using Google Colab.
Final tuned networks used in analyses were trained on the GPU, and weights were stored locally for use in the analyses.

2.5.5 ANN Hyperparameter Tuning

To determine the optimum ANN configuration for generating higher quality low-n MNIST image reconstructions, network hyperparameter tuning was undertaken on three network architectures. The training set of reverse correlation reconstructions was split into a 50,000-digit training set and a 10,000-digit validation set, stratified on the image labels. Networks were trained with the 50,000-image training set and evaluated on the validation set. Loss and validation loss were assessed as the mean squared error (MSE) between the low-n network input and its corresponding high-n desired output. Networks were trained for 70 epochs with the Adam optimizer and the default learning rate of 0.001, the batch size was 32, and early stopping was used with a minimum delta of 0.01 and a patience of 5. The weights that led to the lowest validation loss before stopping had occurred were used as the final weights for that network configuration, and the network with those weights was used to again generate output reconstructions from the validation set inputs. Network outputs were correlated with the high-n reconstructions, and the average correlation of the validation set was used to determine the optimum hyperparameters.

2.5.5.1 MLP Autoencoder

The first ANN created was a multilayered perceptron (MLP) autoencoder with a 784-unit input layer, a fully-connected encoding layer, a coding, a fully connected decoding layer, and a 784-unit output layer. For symmetry, the encoding and decoding layers were designated to have the same number of units. The encoding and decoding layers were assessed with 20 to 1120 units (inclusive), in increments of 20. The coding was evaluated with 10 to 150 (inclusive) units, in
increments of 10. In all, 868 configurations were evaluated. All activation functions were Leaky ReLU.

Hyperparameter tuning of this autoencoder showed a correlation plateau as a factor of both coding size and the size of the encoding and decoding layers (Appendix C). To maximize network simplicity while still getting the best model performance on the validation set, the simplest architecture on the plateau was selected as the best. This resulted in the ANN having a 784-unit input layer, a 440-unit fully-connected encoding layer, a 120-unit coding, a 440-unit fully connected decoding layer, and a 784-unit output layer (Figure 5). This network had an average correlation of 0.797. The maximum correlation of any network was 0.801, and the minimum was 0.617, indicating that this simpler network was not far off from the best.

2.5.5.2 Stacked MLP Autoencoder

The second ANN created was a stacked MLP autoencoder with a 784-unit input layer, a fully-connected encoding comprised of two sequential layers (the outer and inner encoding layers), a coding, a fully connected decoding comprised of two sequential layers (the inner and outer decoding layers), and a 784-unit output layer. For symmetry, the outer encoding and outer decoding layers were designated to have the same number of units, and the inner encoding and inner decoding layers had the same number of units. The outer encoding and outer decoding layers were assessed with 60 to 580 units (inclusive), in increments of 20. The inner encoding and inner decoding layers were assessed with 40 to 520 units, in increments of 20; to preserve the autoencoder structure, the inner layers were prevented from having more units than the outer layers. The coding was evaluated with 20 to 120 (inclusive) units, in increments of 10. All activation functions were Leaky ReLU.
Tuning this network achieved a maximum correlation with the validation set of 0.798 (Appendix D). The architecture of this network was a 784-unit input layer, an encoding of sequential 480- and 460-unit fully connected layers, a 120-unit coding, a decoding of sequential 460- and 480-unit fully connected layers, and a 784-unit output layer (Figure 6).

2.5.5.3 Convolutional Autoencoder
The third ANN created was a convolutional autoencoder with input shape 28x28x1; because of this input shape, the reconstructions were resized back to 28x28 (from 784x1) for input into the network. Convolutional autoencoders are well suited for image processing tasks such as this because the convolutional layers are able to learn important features about the image structure [66]. The encoder had two 2D convolutional layers with a 2D max pooling layer after each. The number of filters in the convolutional layers was varied from 1 to 105, in increments of 4; the square kernel sizes varied from 1 to 11 in increments of 1 and from 15 to 31 in increments of 4; the padding for both was evaluated as same and valid. The max pooling layers were evaluated with pool size varied from 1 to 11 in increments of 1; padding was evaluated as same and valid and could be different from the convolutional layers’ padding. Following the encoder, the coding was evaluated from 10 to 150 (inclusive) units, in increments of 10. After the coding, the encoder had two 2D transposed convolutional layers. The transposed convolutional layers were evaluated with the number of filters, kernel size, and padding varied the same as the convolutional layers in the encoding. The output layer was a 2D convolutional layer with a filter size of 1, a kernel size of 1x1, and same padding. All activation functions were Leaky ReLU, the stride was 1 for the convolutional and pooling layers, and the stride was 2 for the transpose convolutional layers. To accommodate the input and output shape of the network, training and validation images were reconverted back to size 28x28x1.

The first hyperparameters of the convolutional autoencoder analyzed were the paddings of the convolutional and max pooling layers. Same/same and same/valid padding in the convolutional/pooling layers, respectively, performed nearly identically on the validation set
(Appendix E). Therefore, same padding was chosen as the best option. A pool size of 2 for the max pooling layers was the only size found to be compatible with the data, so 2 was the pool size used. Subsequent investigation of the optimum coding size revealed that a larger coding was not better; the optimum coding size was 60. Therefore, fewer coding sizes were evaluated at higher kernel sizes. Further analysis revealed that the correlation with a coding size of 60 was the

Figure 6 Architecture of stacked MLP autoencoder. The network had a 784-unit input layer, an encoding of sequential 480- and 460-unit fully connected layers, a 120-unit coding, a decoding of sequential 460- and 480-unit fully connected layers, and a 784-unit output layer.
highest performing size as kernel size and number of filters increased, so 60 was chosen as the best coding size. Of the networks with those parameters, the best had 101 filters in the convolutional and transposed convolutional layers, and a kernel size of 11x11 (Figure 7). The mean correlation of the validation set output from this network was 0.853.

2.5.6 Evaluation of ANNs Trained on Reverse Correlation Reconstructions

After the three ANNs had been tuned, each was trained on the 50,000-image reverse correlation reconstruction training set, with validation loss again monitored on the validation set. Low-n images from each of the four test sets (reconstructions made with reverse correlation, reverse correlation with whitened stimuli, compressive sensing, and compressive sensing with whitened stimuli) were input into the network. For a given test set, the correlations were found between the network output and both the high-n reverse correlation reconstructions and the MNIST digit test set.

2.5.7 Evaluation of ANNs Trained on Compressive Sensing Reconstructions

To determine the robustness of the network architectures, the three ANNs tuned on generating high-n reverse correlation reconstructions were each trained on the 50,000-image compressive sensing reconstruction training set, with validation loss monitored on the validation set. Low-n images from each of the four test sets were input into the network. For a given test set, the correlations were found between the network output and both the high-n compressive sensing reconstructions and the MNIST digit test set.
Figure 7 Architecture of convolutional autoencoder. The encoder had two pairs of alternating convolutional and max pooling layers with same padding and a stride size of 1. These led to a coding layer of size 60. The decoder consisted of two convolutional transpose layers with a stride of 2 and same padding. All convolutional and convolutional transpose layers had 101 filters and a kernel size of 11.
3. RESULTS

The structure of this results section parallels that of the methods. First, reconstructions of the letter “S” and faces using compressive sensing are compared to reconstructions generated with conventional reverse correlation. Next, the effects of whitening stimuli prior to performing reverse correlation are conveyed. Finally, the performance of each autoencoder on generating higher quality reconstructions from low-n reconstructions is outlined. The section number of the corresponding methods section is parenthetically included in these subsection headings.

3.1 Compressive Sensing Reconstructions of the Letter “S” (2.2.3)

A simulation was used to reconstruct a subject’s internal representation of a letter “S”, as was done in [35]. The reverse correlation reconstruction with 1,000 trials had an \( r^2 \) correlation with the template image of 0.18 (Figure 8). Increasing the number of trials to 10,000 increased the correlation to 0.70. Reconstructions made with compressive sensing had correlations of 0.66 and 0.87 with 1,000 and 10,000 trials, respectively.

![Reconstructions using compressive sensing and reverse correlation.](image)

**Figure 8** Reconstructions using compressive sensing and reverse correlation. Reconstructions of the letter “S” template were generated using compressive sensing and conventional reverse correlation with 1,000 and 10,000 stimuli. Correlation was calculated between each reconstruction and the template.

3.2 Simulating Noisy Subject Responses (2.2.4)

To simulate noisy subject responses while reconstructing the letter “S”, \( e \) of the responses in \( y \) were switched, with \( e \) being either 0, 10, or 20 percent. Average correlation over eight repeat trials for reverse correlation reconstructions with \( e \) equal to 0 (10, 20) were 0.106 (0.073, 0.046),
0.189 (0.132, 0.086), 0.321 (0.237, 0.147), 0.488 (0.374, 0.251), and 0.660 (0.546, 0.409) for \( n \) equal to 1,250, 2,500, 5,000, 10,000, and 20,000 trials, respectively. Average correlation over eight repeat trials for compressive sensing reconstructions with \( e \) equal to 0 (10, 20) were 0.581 (0.467, 0.344), 0.701 (0.650, 0.542), 0.765 (0.730, 0.657), 0.809 (0.787, 0.750), and 0.841 (0.822, 0.800) for \( n \) equal to 1,250, 2,500, 5,000, 10,000, and 20,000 trials, respectively (Figure 9).

![Figure 9 Simulation of noisy subject responses. Reconstructions were generated of the letter “S” using reverse correlation (RC) and compressive sensing (CS) at values of \( n \) varied from 1,250 to 20,000. To simulate noisily responding subjects, a proportion \( e \) of the responses were flipped, where \( e=0.1 \) indicates that 10% of the responses were flipped. Error bars indicate the 2.5 and 97.5 percent bootstrapped confidence intervals over the 8 repeat trials at each value of \( n \). Solid lines indicate plots for CS, and dashed lines indicate plots for RC.](image)

3.3 Compressive Sensing on Human Data (2.2.5)

Human responses to visual noise were used to reconstruct internal representations of faces with data from [1] using Equation 4. Compared to reverse correlation reconstructions with
the same number of trials, compressive sensing reconstructions were less noisy and bore higher resemblance to faces (Figure 10).

*Figure 10 Reconstructions of human internal representations of faces. For each of the five subjects (S1-S5) from the data collected in [1], reconstructions were generated using reverse correlation (RC) or compressive sensing (CS) on 10% of the trials and on the full dataset.*
3.4 Cross Validation of Reconstructions (2.2.6)

To assess the quality of the simulation estimation procedure, human data from each of the five subjects in [1] was used for a five-fold cross-validation with reverse correlation and compressive sensing reconstructions. Reconstructions with compressive sensing had a higher accuracy than their corresponding reverse correlation reconstructions (Figure 11). Most substantially, using compressive sensing increased the accuracy for subject S2 from 56 to 61%.

![Figure 11 Balanced accuracies of cross-validation for five human subjects.](image)

Figure 11 Balanced accuracies of cross-validation for five human subjects. Human subject response data was used to generate a reconstruction with reverse correlation or compressive sensing, and simulated responses to stimuli were generated using the reconstruction as a template. A five-fold cross-validation was done to compare the subject responses to the simulated responses for each subject. For all subjects, compressive sensing provides a more accurate estimate of responses than conventional reverse correlation.

3.5 Whitened Data (2.3.3)

Simulations of Gosselin & Schyns letter “S” experiment were carried out using both whitened and unwhitened data. The correlation to the template of the reconstruction with 1,000 whitened stimuli was 0.245, whereas the reconstruction with 1,000 random stimuli had a
correlation of 0.081 (Figure 12). The correlations were 0.772 and 0.320 for the reconstructions using 5,000 whitened and unwhitened stimuli, respectively.

![Figure 12 Comparison of reconstructions with whitened and unwhitened stimuli. Reconstructions of the template were generated using reverse correlation of responses to random stimuli and whitened stimuli. Reconstructions were generated in response to 1,000 and 5,000 stimuli. Correlation is shown between the reconstruction and the template.](image)

### 3.6 Whitening with Compressive Sensing (2.4)

Reconstructions were generated with standard reverse correlation, reverse correlation with whitened stimuli, compressive sensing, and compressive sensing with whitened stimuli using a varied number of stimuli (Figure 13). These reconstructions were correlated with the letter “S” template. Average $r^2$ for reverse correlation reconstructions was 0.285, 0.572, 0.681, 0.742, 0.782, and 0.820 at 1,000, 5,000, 9,000, 13,000, 17,000, and 21,000 samples, respectively. Average $r^2$ for whitened reverse correlation reconstructions was 0.504, 0.885, 0.930, 0.948, 0.960, and 0.967 at 1,000, 5,000, 9,000, 13,000, 17,000, and 21,000 samples, respectively. Average $r^2$ for compressive sensing reconstructions was 0.677, 0.883, 0.905, 0.916, 0.926, and 0.934 at 1,000, 5,000, 9,000, 13,000, 17,000, and 21,000 samples, respectively. Average $r^2$ for whitened compressive sensing reconstructions was 0.856, 0.943, 0.953, 0.955, 0.957, and 0.957 at 1,000, 5,000, 9,000, 13,000, 17,000, and 21,000 samples, respectively. Comparison of the techniques at lower n values can be found in Appendix B.

### 3.7 Evaluation of ANNs Trained on Reverse Correlation Reconstructions (2.5.6)

#### 3.7.1 MLP Autoencoder (2.5.5.1)
The MLP autoencoder was trained to generate high-n reverse correlation reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.795 (range: 0.601-0.847) and 0.818 (range: 0.618-0.871) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.813 (range: 0.643-0.855) and 0.836 (range: 0.661-0.879) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.657 (range: 0.458-0.730) and 0.677 (range: 0.473-0.751) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images had an average correlation...
correlation of 0.759 (range: 0.463-0.833) and 0.781 (range: 0.476-0.857) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively (Figure 14).

![Figure 14 Performance of MLP autoencoder trained on reverse correlation reconstructions. Reconstructions generated with reverse correlation (RC), reverse correlation with whitened stimuli (RC+W), compressive sensing (CS), or compressive sensing with whitened stimuli (CS+W) with 100-1100 trials were input into the network and correlated with either high-n (50,000 trial) reverse correlation reconstructions and MNIST digits.]

3.7.2 Stacked MLP Autoencoder (2.5.5.2)

The stacked MLP autoencoder was trained to generate high-n reverse correlation reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.796 (range: 0.609-0.846) and 0.819 (range: 0.628-0.871) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.812 (range: 0.647-0.852) and 0.835 (range: 0.665-0.876) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.663 (range: 0.435-0.733) and 0.683 (range: 0.448-0.755) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images had an average correlation of 0.757 (range: 0.452-0.830) and 0.779 (range: 0.463-0.854) with
high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively (Figure 15).

![Figure 15 Performance of stacked MLP autoencoder trained on reverse correlation reconstructions. Reconstructions generated with reverse correlation (RC), reverse correlation with whitened stimuli (RC+W), compressive sensing (CS), or compressive sensing with whitened stimuli (CS+W) with 100-1100 trials were input into the network and correlated with either high-n (50,000 trial) reverse correlation reconstructions and MNIST digits.]

3.7.3 Convolutional Autoencoder (2.5.5.3)

The convolutional autoencoder was trained to generate high-n reverse correlation reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.836 (range: 0.624-0.890) and 0.860 (range: 0.643-0.916) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.845 (range: 0.680-0.887) and 0.869 (range: 0.700-0.912) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.639 (range: 0.410-0.709) and 0.658 (range: 0.424-0.729) with high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images had an average correlation of 0.759 (range: 0.412-0.842) and 0.781 (range: 0.423-0.865) with
high-n reverse correlation reconstructions and the corresponding MNIST digits, respectively (Figure 16).

![Correlation of Network Output with High-n RC and Correlation of Network Output with MNIST](image)

*Figure 16 Performance of convolutional autoencoder trained on reverse correlation reconstructions. Reconstructions generated with reverse correlation (RC), reverse correlation with whitened stimuli (RC+W), compressive sensing (CS), or compressive sensing with whitened stimuli (CS+W) with 100-1100 trials were input into the network and correlated with either high-n (50,000 trial) reverse correlation reconstructions and MNIST digits.*

3.8 Evaluation of ANNs Trained on Compressive Sensing Reconstructions (2.5.7)

3.8.1 MLP Autoencoder (2.5.5.1)

The MLP autoencoder was trained to generate high-n compressive sensing reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.807 (range: 0.593-0.859) and 0.743 (range: 0.474-0.813) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.828 (range: 0.620-0.873) and 0.768 (range: 0.503-0.836) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.832 (range: 0.574-0.894) and 0.721 (range: 0.470-0.797) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images
had an average correlation of 0.853 (range: 0.620-0.897) and 0.763 (range: 0.524-0.830) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively (Figure 17).

3.8.2 Stacked MLP Autoencoder (2.5.5.2)

The stacked MLP autoencoder was trained to generate high-n compressive sensing reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.815 (range: 0.599-0.865) and 0.741 (range: 0.485-0.809) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.832 (range: 0.622-0.876) and 0.763 (range: 0.508-0.832) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.831 (range: 0.599-0.890) and 0.719 (range: 0.494-0.791) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images
had an average correlation of 0.853 (range: 0.622-0.898) and 0.758 (range: 0.528-0.822) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively (Figure 18).

![Figure 18 Performance of stacked MLP autoencoder trained on compressive sensing reconstructions. Reconstructions generated with reverse correlation (RC), reverse correlation with whitened stimuli (RC+W), compressive sensing (CS), or compressive sensing with whitened stimuli (CS+W) with 100-1100 trials were input into the network and correlated with either high-n (50,000 trial) compressive sensing reconstructions and MNIST digits.](image)

### 3.8.3 Convolutional Autoencoder (2.5.5.3)

The convolutional autoencoder was trained to generate high-n compressive sensing reconstructions from low-n inputs. Network outputs from low-n reverse correlation images had an average correlation of 0.825 (range: 0.561-0.874) and 0.771 (range: 0.430-0.838) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened reverse correlation images had an average correlation of 0.849 (range: 0.589-0.893) and 0.797 (range: 0.458-0.861) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n compressive sensing images had an average correlation of 0.857 (range: 0.595-0.916) and 0.758 (range: 0.494-0.830) with high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively. Network outputs from low-n whitened compressive sensing images had an average correlation of 0.873 (range: 0.606-0.915) and 0.796 (range: 0.514-0.858) with
high-n compressive sensing reconstructions and the corresponding MNIST digits, respectively (Figure 19).

**Figure 19** Performance of convolutional autoencoder trained on compressive sensing reconstructions. Reconstructions generated with reverse correlation (RC), reverse correlation with whitened stimuli (RC+W), compressive sensing (CS), or compressive sensing with whitened stimuli (CS+W) with 100-1100 trials were input into the network and correlated with either high-n (50,000 trial) compressive sensing reconstructions and MNIST digits.

4. DISCUSSION

Reverse correlation is a technique used widely across neuroscience, in areas ranging from electrophysiology to cognitive psychology. Conducting a reverse correlation study requires collecting a high number of samples, which severely limits the utility of the technique. In cognitive psychological experiments, wherein the goal is to determine a subject’s internal representation of an object, the procedures require collecting thousands of subject behavioral responses to noise stimuli. Collecting sufficient samples to create an adequate reconstruction of a cognitive representation can take weeks or months of data collection, and this burden often reduces the number of subjects included in a study, which limits the generalizability of the results.

Current approaches to address this inefficiency often involve overlaying noise over the target object so that the subject is more likely to perceive the object in the stimulus. Doing so
reduces the necessary number of trials but also biases the resulting reconstruction towards resembling the underlaid object instead of the subject’s own representation of the object. Other approaches smooth the reconstructions to eliminate unwanted noise, but doing so assumes that the noise is not important for the reconstruction. The present work proposes several alternative approaches to improving the efficiency of reverse correlation.

4.1 Compressive Sensing

4.1.1 Compressive Sensing Reconstructions of the Letter “S”

Compressive sensing is an advanced signal processing technique that here is coupled with the reconstruction process of reverse correlation. Compressive sensing was able to produce a reconstruction with a correlation of 0.66 in only 1,000 trials, whereas the reconstruction from conventional reverse correlation had a similar correlation of 0.70 in 10,000 trials (Figure 8). This suggests that compressive sensing can reduce the number of trials required for accurate estimation of cognitive representations using reverse correlation by up to 90% with negligible loss of accuracy. Because estimating representations with traditional reverse correlation typically requires the collection of several thousands of stimulus response pairs, very few participants are examined in any given study, which limits the possible analyses and inferences about potential universal aspects of human cognitive representation, as well as examination of individual differences. The 90% reduction in required trials demonstrated here opens the door to conducting more studies with more participants. Cognitive studies that would presently take weeks to complete could instead be performed in one sitting, and those studies that would otherwise impose limiting constraints on either the stimuli or the reconstructions could be conducted with assumptions only about sparsity; this will lead to clearer, less biased representations. Given the central importance of cognitive representations in mediating human perceptual experience and
behavior, this presents a substantial increase for reverse correlation’s potential for wide-spread use in the psychological and cognitive sciences.

4.1.2 Robustness of Compressive Sensing to Noise

Human subjects have internal noise that makes their responses vary from trial to trial partly independently of the stimulus [30], and subjects show uncertainty about the spatial location of the target in the stimulus [60]. Noise was modelled in the simulated subject decision process by flipping a certain percentage of the responses, and reconstructions were made based on those responses with both compressive sensing and reverse correlation. At all numbers of samples, compressive sensing drastically outperformed reverse correlation (Figure 9). Even with noise in 20% of the subject responses, compressive sensing had a higher average $r^2$ than noiseless reverse correlation. This suggests that generating reconstructions of subjects’ latent representations with compressive sensing is robust enough to perform well even with subjects responding imperfectly. To demonstrate this further, compressive sensing was evaluated on actual human subject data.

4.1.3 Compressive Sensing on Human Data

Stimulus-response pairs from [1] were used to reconstruct internal representations of faces using reverse correlation (as was done in the original paper) and using compressive sensing. Because each of the five subjects’ templates are not known, a correlation cannot be calculated to assess the quality of the reconstructions. Subjectively however, the faces reconstructed using the full measurement matrix more clearly show face-like features. For example, Smith identified features such as a nose, mouth, and chin on subjects S1, S2, and S3 (Figure 10, [1]), and these features are more noticeable on the reconstructions generated with compressive sensing. Furthermore, when the reconstructions from human data were used as
templates for a simulated subject response process, the simulated responses using the compressive sensing reconstructions as a template were closer to the actual responses than were the responses using the reverse correlation reconstructions as a template (Figure 11). This suggests that compressive sensing produced a better reconstruction of the subjects’ representations of a face than did reverse correlation.

4.1.4 The Potential Impact of Compressive Sensing on Reverse Correlation

The success of compressive sensing in generating reconstructions of faces also demonstrates the ability of the technique to be applied retroactively to existent datasets. Any researcher who has performed a reverse correlation experiment can apply compressive sensing to their data to improve the reconstruction quality. What’s more, the applicability of compressive sensing to fields beyond cognitive science continues to grow. One-bit compressive sensing, the form of compressive sensing used here, is an active area of research (e.g., [67]). In experimental paradigms where continuous responses are collected from subjects, classical compressive reconstruction techniques (e.g., [68]) may turn out to be appropriate, and perhaps superior, to use. In the context of perceptual experiments, eliciting yes/no responses from subjects is likely to remain the standard method because it is most straightforward for subjects and experimenters to implement. Other studies have made use of rating scales (e.g., Likert scales), and this approach may be treated in the present framework by further quantizing ratings into binary responses. Further work should evaluate the performance of compressive sensing in these other experimental paradigms.

The simulations reconstructing “S” and MNIST digits showcase compressive sensing’s utility for analyzing novel data, and reanalysis of human response data from [1] shows that compressive sensing can also be retrospectively applied to existent data to improve results. Real
human subject responses are the gold standard data when proposing a technique intended to improve the processing of human subject responses, and while this work demonstrates compressive sensing’s utility on real data, the simulations offer the benefit of being an entirely observable process, including the access to underlying representations that the responses are based off of. This means that the quality of the reconstruction could be assessed directly by comparing the representation to the reconstruction. The subject response model used in simulation here is consistent with that assumed in other applications of reverse correlation [30], and this work takes the additional step of modeling subject noise in the responses to further represent a human subject (Figure 9). The performance of compressive sensing on these robust, accepted simulations and on the human data suggests that it holds promise in other reverse correlation studies in cognitive science.

Outside of cognitive science, compressive sensing holds promise for drastically improving the efficiency and accuracy of analysis in any domain where reverse correlation is used, including a broad range of electrophysiology paradigms (e.g., [3, 4, 69]), auditory system function [31], calcium imaging [70], and gene expression [26]. That the reverse correlation responses can be continuous (e.g., neural firing rates [4]), ordinal (e.g., similarity scores [71]), or binary or yes/no (e.g., [46, 50, 51]) gives compressive sensing widespread utility in areas of science even where reverse correlation is not currently in common use.

4.2 Whitening

4.2.1 Whitening Stimuli in Reverse Correlation

Whitening the stimuli before presenting them to the simulated subjects greatly improved the subsequent reconstruction process. The results indicate that the number of trials required to produce reconstruction quality with an adequate signal-to-noise ratio is greatly improved by
whitening stimuli. As shown in the mean correlation values in Figures 12 and 13, using whitened stimuli produces a substantially higher reconstruction quality for all of the reconstructed images than using the conventional random stimuli. This improvement in reconstruction quality is apparent for all numbers of stimuli evaluated and is further reinforced by the example reconstructions shown in Figures 12 and 20.

Using 5000 whitened and unwhitened stimuli, the correlations of the reconstructions with the template were 0.77 and 0.32, respectively. Additionally, the correlation of reconstructions from whitened data at 5,000 trials (0.78) was greater than that from unwhitened stimuli after 21,000 trials (0.67). That a simple procedure like whitening could more than double the correlation in the same number of trials is a striking finding. The sole limitation of whitening is that the procedure must be applied before the subject responses are collected because the measurement matrix must be whitened before presentation. For this reason, whitening could not be evaluated on the human subject data from [1]. Especially given whitening’s simplicity, future work can and should evaluate stimulus whitening in human or other in vivo studies.

4.2.2 Whitening with Compressive Sensing

The improvements seen with whitening the stimuli and generating reconstructions with compressive sensing were substantial on their own, but the benefits were shown to be additive (Figure 13). At 1,000 samples, whitening and compressive sensing each increased the correlation from 0.08 to 0.25 and 0.46, respectively. When whitening and compressive sensing were combined, the correlation increased further to 0.73. Compressive sensing with whitening passed the 0.9 $r^2$ mark between 5,000 and 9,000 trials. Between 5,000 and 9,000 trials, reconstructions from reverse correlation had a correlation between 0.33 and 0.46, meaning that combining whitening with compressive sensing more than doubled the reconstruction quality in that many
trials for the stimulus size used. Reconstructions with an $r^2$ correlation over 0.9 have an $r$ correlation of at least 0.95, making them nearly indistinguishable from the templates, as exemplified in Figure 20.

Interestingly, reconstructions made with whitening both with and without compressive sensing had average $r^2$ correlations that plateaued around 0.9. Furthermore, compressive sensing appeared to be approaching 0.9 but had not yet reached that correlation in 21,000 trials. Future investigations should be done to determine if compressive sensing would eventually reach that level of performance and if reverse correlation, which was only at 0.67 but still rising at 21,000 trials, would ever reach 0.9.

![Figure 20 Reconstructions at high-n. Reconstructions of the letter “S” using 21,000 stimuli were generated using reverse correlation (RC), whitening, compressive sensing (CS), and compressive sensing with whitened stimuli.](image)

4.3 Artificial Neural Networks

4.3.1 ANNs for the Improvement of Reverse Correlation Reconstructions

Subjectively, the low-n reconstructions appeared to be noisy versions of the high-n reconstructions. That was the intuition behind using a denoising neural network like an autoencoder. Autoencoders have previously been used to denoise images ranging from medical imaging to faces [64, 72]. Compared to the low-n reconstructions, the network outputs were clearer, substantially less noisy images for reconstructions created with each of the four combinations of techniques. All six of the ANNs increased the correlation of the reconstructions,
in some cases by as much as 0.7. Such an increase in correlation would require a tenfold increase in the required number of trials in a conventional reverse correlation experiment.

For all networks trained to generate high-n reverse correlation reconstructions, average correlations were higher between the network outputs and the MNIST digits than the network outputs and the high-n reconstructions. The opposite was true for the networks trained to generate high-n compressive sensing reconstructions; correlation was higher between the network outputs and the high-n reconstructions than the network outputs and the MNIST digits. This may be because the outputs from the reverse correlation network have clear edges on a black background like the MNIST digits, whereas the compressive sensing network outputs less well-defined digits, like the high-n compressive sensing reconstructions (Figure 21).

On networks trained to generate high-n reverse correlation reconstructions, the correlation of the network output was much higher for inputs made with reverse correlation than with compressive sensing. Additionally, the correlation of the network outputs was similar between inputs made with reverse correlation and with compressive sensing for the network trained to generate high-n compressive sensing reconstructions. To put that concisely, the network trained on compressive sensing was better at improving reverse correlation than the network trained on reverse correlation was at improving compressive sensing. This may be explained by the low-n data. Prior to being input into the network, the low-n compressive sensing reconstructions had higher correlations with the high-n reverse correlation reconstructions, the high-n compressive sensing reconstructions, and the MNIST digits than did the low-n reverse correlation reconstructions (Appendix F). The networks were trained to minimize the difference between the low- and high-n reconstructions they were trained on. Because this difference was large when training on reverse correlation reconstructions, the
network learned to make more substantial changes to the input, which likely diminished the generation of compressive sensing reconstructions. Furthermore, the low-n compressive sensing reconstructions are also shown, as well as the MNIST digit that served as the template for reconstruction.
reconstructions are often wavier than the low-n reverse correlation reconstructions (because they are made up of basis functions), and the networks trained on reverse correlation were not trained to remove this waviness; this would diminish the quality of compressive sensing reconstructions from the reverse correlation-trained networks.

The goal of the networks was to create a high-n reconstruction, not the MNIST digit, so the correlation with high-n reconstruction is the more important metric for how well the network performed. That the networks trained to generate high-n reverse correlation reconstructions produced outputs with higher average correlations with the MNIST digits than the network outputs was surprising. However, this may ultimately be a boon for future work using this method because it demonstrates that the ANNs are able to generate representations of the template, not just a denoised low-n reconstruction.

At $n$ values between around 600 and 1,100, the correlation of high-n compressive sensing reconstructions with reconstructions output from the autoencoders trained on compressive sensing images deceased compared to at $n$ values between 400 and 600. That is, using more trials to generate the reconstruction decreased the post-network reconstruction quality. This may be because of how the high-n compressive sensing images, which the network is trained to produce, visually compare to high-n reverse correlation (with and without whitening) reconstructions. As illustrated in Figures 20 and 21, letters in high-n reverse correlation reconstructions have clear edges, whereas the letters in compressive sensing reconstructions have blurrier boundaries.

Beyond 600 trials, as the mismatches between pre-network reverse correlation and compressive sensing reconstructions become more apparent, the networks blur the already clear reverse correlation reconstructions to bias them towards resembling the compressive sensing reconstructions; doing so diminishes the reconstruction quality.
4.3.2 Hyperparameter Tuning

Tuning the autoencoder hyperparameters was an iterative process, but optimizing the hyperparameters proved to be worthwhile. Initially, the activation function of the output layer was sigmoid. Although the outputs from these networks were visibly denoised and had a higher correlation with the high-n reconstruction than the inputs did, the reconstructions had a puffy appearance without fine lines or edges. Changing the output activation from sigmoid to ReLU increased the average correlation from 0.6 to 0.7 and made the reconstructions cleaner and clearer. A comparison of the network outputs with sigmoid and ReLU activation functions can be found in Appendix G.

Another change that took place during hyperparameter tuning was to change the activation function in the other layers from ReLU to Leaky ReLU. This was done because the autoencoders’ coding layers were showing many units completely equal to zero. This is a phenomenon known as “dying ReLU” or “dead neurons”, where the activation function becomes saturated and cannot update the gradients during back propagation [73, 74]. As a result, the dead neurons will only output zero and will thus not learn any valuable information. Leaky ReLU solves this problem because it is a non-saturating activation function, and, thus, neurons with this activation function cannot “die”. After changing to leaky ReLU, the coding layers stopped containing dead neurons.

4.3.3 The Potential for ANNs in Reverse Correlation Research

This is not the first work to merge reverse correlation with ANNs. Reverse correlation has been used to evaluate how ANNs make decisions [29], to compare how objects are represented in humans and in neural networks [75], and to map the “receptive fields” in ANNs [28]. Other work has used neural networks to generate reconstructions of cognitive
representations based solely on stimulus response pairs then compared those reconstructions to reconstructions generated from reverse correlation [76]; [76] did not use reconstructions as input into the network. The present work is the first to use ANNs to enhance the quality of reconstructions generated from reverse correlation, and it holds promise in reducing the number of trials needed to generate an accurate reconstruction.

A limitation of the work using autoencoders to improve reconstruction quality is that the networks required supervised training. In simulation studies, generating sufficient training data is easily done, but collecting such data on human subjects is not feasible. For this work to be useable in experimental settings, networks need to be created that are capable of meeting at least one of two criteria: networks trained on simulation data need to be robust enough to accurately reconstruct representations from in vivo data, or the denoising networks must be unsupervised, the latter of which has been done in [77]. Future work should be undertaken to demonstrate that ANNs can meet at least one of these criteria in generating reverse correlation reconstructions. Nonetheless, this work lays the groundwork for this by establishing that ANNs are capable of both reducing the number of trials required to generate a reconstruction and increasing the reconstruction quality in reverse correlation studies.

4.3.4 Filtering Reconstructions

Besides ANNs, another common image denoising technique is filtering. Filtering is a process by which an input signal is transformed into a modified version of that signal. In image denoising, the input signal is the image, and the output is a denoised version of that image. Filtering has been used before in reverse correlation studies. Gosselin & Schyns presented their original letter “S” reconstructions after passing the images through a low-pass filter [35]. As noted above, using such filters presupposes that the underlying representation does not contain
the high-frequency filtered information. Yet if autoencoders are serving to denoise the reconstructions, it is important to compare the networks to filtering. As a supplementary analysis to determine if the autoencoders outperformed simple filters, low-n reconstructions were run through mean blur and Gaussian blur filters. In both cases, filtering increased the correlation with the high-n reconstructions, but the increase was not as substantial as was seen in any of the autoencoders (Appendix H). Had the two techniques shown similar performance, it could be proposed that the autoencoders are simply acting as filters. This would help to unveil the black box of the neural networks. However, filtering was not as effective as inputting the reconstructions into the network. This suggests that the ANNs are learning a more complex computation. Nonetheless, in situations where enough training data cannot be collected, filtering could serve as a helpful option.

4.4 Conclusion

In this work, a set of computational simulations are used to model a technique called reverse correlation in subject responses to whitened or unwhitened visual stimulation. Internal representations are reconstructed based on these simulated responses with reverse correlation or with compressive sensing. Compressive sensing was shown to improve both the efficiency of the sampling process and the quality of the reconstructions in both simulated data and, additionally, in human subject responses. Whitening the stimuli increased the resulting reconstruction quality as well and saw further benefits when combined with compressive sensing. Finally, ANNs were constructed that increased the quality of simulated reconstructions.
5. REFERENCES


Shen, J. *One-Bit Compressed Sensing via One-Shot Hard Thresholding*. in *Conference on Uncertainty in Artificial Intelligence*. 2020. PMLR.


6. **APPENDIX**

*Appendix A* Variables used throughout the thesis.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>The vector of subject responses</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Measurement matrix aka matrix of stimuli</td>
</tr>
<tr>
<td>$x$</td>
<td>The latent representation</td>
</tr>
<tr>
<td>$\hat{x}$</td>
<td>The estimated latent representation</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Irreducible error from noise</td>
</tr>
<tr>
<td>$n$</td>
<td>The number of trials performed</td>
</tr>
<tr>
<td>$m$</td>
<td>The size of the latent representation</td>
</tr>
<tr>
<td>$I$</td>
<td>The identity matrix</td>
</tr>
<tr>
<td>$r^2$</td>
<td>Squared correlation</td>
</tr>
<tr>
<td>$s$</td>
<td>The sparse latent representation</td>
</tr>
<tr>
<td>$\hat{s}$</td>
<td>The estimated sparse latent representation</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>A matrix of orthonormal basis functions</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>Similarity between stimuli and basis functions (i.e., $\Phi \Psi$)</td>
</tr>
<tr>
<td>$l$</td>
<td>The height of the compressive sensing matrix</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>The number of elements not thresholded to zero in compressive sensing</td>
</tr>
<tr>
<td>$e$</td>
<td>The noise added to simulated subject responses</td>
</tr>
<tr>
<td>$W$</td>
<td>A whitening matrix</td>
</tr>
<tr>
<td>$C$</td>
<td>A centering matrix</td>
</tr>
<tr>
<td>$I$</td>
<td>An all-ones matrix</td>
</tr>
<tr>
<td>$\Phi_w$</td>
<td>The whitened measurement matrix</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The ridge regression penalty term</td>
</tr>
</tbody>
</table>
Appendix B  Quality of reconstructions generated from reverse correlation, reverse correlation with whitened stimuli, compressive sensing, and compressive sensing with whitened stimuli over a varied number of trials.
Appendix C Hyperparameter tuning of MLP autoencoder.
Appendix D Hyperparameter tuning of stacked MLP autoencoder.
Appendix E Hyperparameter tuning of convolutional autoencoder.

Padding
Convolutional / Pooling

valid/valid

valid/same

same/valid

same/same

Model Complexity

Mean Correlation

Mean Correlation

Figure A

Figure B
Appendix F Correlation of network inputs with high-n reverse correlation reconstructions, high-n compressive sensing reconstructions, and MNIST digits before being input into the networks.
Appendix G Comparison of network outputs when activation function is Sigmoid and ReLU.

Low-n Reconstruction

High-n Reconstruction

Network Output (Sigmoid)

Low-n Reconstruction

High-n Reconstruction

Network Output (ReLU)
Appendix H Correlations as a result of filtering the low-n reconstructions.