# A Compressive Sensing Approach to Inferring Cognitive Representations with Reverse Correlation

Benjamin W. Roop[1], Benjamin Parrell[2], Adam C. Lammert[1,3]*

**Affiliations:**

[1]Program of Neuroscience, Worcester Polytechnic Institute; Worcester, Massachusetts, USA

[2]Department of Communication Sciences & Disorders, University of Wisconsin-Madison; Madison, Wisconsin, USA

[3]Department of Biomedical Engineering, Worcester, Polytechnic Institute; Worcester, Massachusetts, USA

*Corresponding author. Email: alammert@wpi.edu

**Abstract:**

Uncovering cognitive representations is an elusive goal that is increasingly pursued using the reverse correlation method. Employing reverse correlation often entails collecting thousands of stimulus-response pairs from human subjects, a burdensome task that limits the feasibility of many such studies. This methodological barrier can potentially be overcome using recent advances in signal processing designed to improve sampling efficiency, specifically compressive sensing. Here, compressive sensing is shown to be directly compatible with reverse correlation, and a trio of simulations are performed to demonstrate that compressive sensing can improve the accuracy of reconstructed representations while dramatically reducing the required number of samples. This work concludes by outlining the potential of compressive sensing to improve representation reconstruction throughout the field of neuroscience and beyond.

**One-Sentence Summary:**

Perceptual representations are uncovered more efficiently than ever before, removing a major barrier to cognitive studies.

**Main Text:**

Human perceptual experience is mediated by internal representations of the world around us. Representations bridge the gap between the raw information available from the senses and the abstract and often categorical nature of our perceptual experience by acting as reference patterns for everything from low-level spatial and temporal features (e.g., edges and shapes in the retina image (2)) to high-level characteristics of cognitive categories (e.g., faces (3)) or social constructs (e.g., the trustworthiness of a face (4)). Full understanding of perceptual experience therefore depends on the ability to characterize these internal representations, as they encapsulate important abstractions that are necessary to make sense of complex sensory inputs (Figure 1). While substantial progress has been made toward fully describing lower-level representations (e.g., visual and auditory receptive fields), similar characterization of higher-level representations has proven to be more elusive. This is largely because, despite their central importance to perceptual experience and behavior (5), such representations are notoriously difficult to measure (1, 3, 6-8).

One of the more promising methods to overcome this difficulty attempts to uncover higher-level cognitive representations using a standard method for characterizing lower-level neural representations: reverse correlation (10). Reverse correlation reveals the latent representation in these low-level contexts by eliciting neural responses to richly varying stimuli (e.g., white noise) and regressing the observed responses against the stimuli over many trials. The same method can be utilized in the higher-level context by substituting voluntary responses for neural ones, in much the same way that one can report whether they "see" faces and other familiar shapes in clouds and other vague visual stimuli. Vague stimuli can thus be used to force the internal representations to exert their influence directly on elicited responses. In a pioneering study, Gosselin & Schyns sought to find representations of letters by asking participants that identify a letter "S" in visual noise stimuli (3); critically, the stimuli were purely white noise (and thus did not contain any actual letters), yet reverse correlation based on subject responses revealed a definitive "S" image recovered from each participant. Reverse correlation has subsequently been demonstrated to avoid the shortcomings of other techniques that bias expectations (11). Other groups have used reverse correlation to characterize the top-down processes of perception to abstract psychological categories, including the trustworthiness in voices (12), self-image (13), and "male" vs "female" faces (5) and shown that reverse correlation can quantitatively compare human decision-making to that of computers (14, 15).

Despite this initial success in certain contexts, reverse correlation has severe limitations that have prevented its wide-spread use for uncovering higher-level internal representations. Most critically, standard applications of reverse correlation require a burdensome number of stimulus-response samples for accurate reconstruction of cognitive representations with an adequate signal-to-noise ratio (8). For example, Gosselin & Schyns' 2003 experiment required 20,000 trials from each participant; another required 11,400 (6). In cognitive and psychological experiments, the large number of trials combined with the long latency for measurable behavioral responses from subjects results in data collection protocols that can last weeks 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 <5 (1, 3, 7)). 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 in (13). 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 in uncovering cognitive representations, it is necessary to reduce the number of trials required without introducing such biases.

An alternative approach to 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 (e.g., (3)). 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 (16, 17).

One constraint that seems especially 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 (18), with strong empirical and theoretical support (18-21). 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 (17). Mineault (17) 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.

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* (16, 17). To date, however, compressive sensing has not been applied to the reverse correlation paradigm. Here, we outline the underlying mathematical connections between reverse correlation and compressive sensing and provide a demonstration of the potential for dramatic efficiency improvements using the combined method. We will also show that, unlike previous approaches to incorporating sparsity constraints, it is possible to obtain reconstructions from compressive sensing analytically, using a closed-form solution. One such method by Zhang (22) is presented in detail and applied here in a novel way.

Compressive sensing begins with the assumption that signals of interest, x, which can include latent cognitive representations (Figure 2), are sparse (or, in the language of compressive sensing, *compressible*). This means specifically that they can be represented by a small number of functions from an appropriately-selected basis set ($s = \Psi^T x$, for basis $\Psi$ and

weights s). If one assumes that responses, y, stem from a process of comparing stimuli to the latent representation ($y=\Phi x$, for stimuli $\Phi$), it is possible to estimate the latent representation using only a small number of measurements by acquiring the basis function representation directly (i.e., $y=\Phi\Psi s$) via sparse optimization approaches to find s. 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 (e.g., the discrete cosine transform, wavelet transform).

Critically for the applicability of this method for reverse correlation, it has been demonstrated both theoretically (23, 24) and empirically (17) 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 (22, 25, 26). 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 the reverse correlation paradigm.

To directly assess the applicability of compressive sensing to recovering cognitive representations, we compare the results of simulations of Gosselin & Schyns' original study on representations of "S" (Figure 3 top) using both conventional reverse correlation (3) and compressive sensing approaches. We used an image of the letter "S" as the target cognitive representation for reconstruction (Figure 3 top A). Specifically, the reshaped, vectorized "S" image was used as the representation vector, x (Figure 2), the basis for generating simulated subject responses in light of random stimuli contained in the stimulus matrix, $\Phi$. The elements of $\Phi$ were generated by randomly sampling from a uniform distribution over the interval 0 and 1 and rounding those values to the nearest integer. A total of 10,000 subject responses were simulated using this procedure, representing the subject responses in (3). Additionally, a similar simulation was run using auditory spectrograms as stimuli. In both simulations, as in most cognitive experiments, we assume binary responses.

The simulated subject responses were used as the basis for reconstructing the latent representation, x, using both conventional regression-based estimation and compressive sensing-based estimation. Reconstructions were generated using both the full set of stimulus-response pairs (10,000 or 20,000) and on a limited subset of a randomly-selected 10% of pairs (1,000 or 2,000). Estimation accuracy for a given reconstructed representation was assessed using 2-dimensional correlation coefficient ($r^2$) between the target representation and the reconstructed estimate. Results (Figure 3) indicate that the image quality obtained using 10% of the samples via compressive sensing is effectively equivalent to that obtained using the full cohort of samples via conventional reconstruction, implying a 90% improvement in sampling efficiency. When allowed to operate on the full complement of samples, compressive sensing additionally shows improved accuracy over conventional reconstruction of approximately 15%.

The simulations reconstructing "S" and spectrograms showcase compressive sensing's utility for analyzing novel data, yet compressive sensing is not limited to use solely in new studies. Rather, compressive sensing can also be retrospectively applied to existant data to improve results. To demonstrate this, further simulations were conducted to replicate the work of Smith, Gosselin, & Schyns, in which they used reverse correlation to reconstruct internal

templates of faces (1). In that study, five subjects were each shown 10,500 white noise images and asked to identify those stimuli with a face in them; their results are recapitulated in Figure 4A. Compressive sensing simulations were implemented with the same number of samples as in the original work. To estimate the original stimulus-response data, a candidate measurement matrix, $\widehat{\Phi}_i$, was randomly generated and used as the basis for estimating subject responses, $\hat{y}$, via the assumed data generating process $\hat{y} = sign(\widehat{\Phi}_i x)$, where $x$ was the reconstructed face representation presented by (1). The estimated subject responses were then used to reconstruct an estimated face representation, which was subsequently subjected to smoothing using a Gaussian-shaped image filter as described by (1) to produce $\hat{x}_i$. This procedure was repeated 1000 times, with the final measurement matrix and simulated subject responses being selected from the iteration which maximizes the image correlation between $x$ and the estimate from that iteration, $\hat{x}_i$. As may be expected in a sparse image, the compressive sensing reconstructions have a smoother, more face-like appearance (Fig 4B). Furthermore, Smith, Gosselin, & Schyns identified several facial structures in their reconstructions (e.g., a hairstyle on S3 and a nose, mouth, and chin outline for S1, S2, and S3), and these are more clearly observable in the reconstructions using compressive sensing. The reconstructions presented here are notably symmetric and lack high frequency variation. These properties are not enforced by the compressive sensing method. Rather, faces themselves are largely symmetric, and these reconstruction properties are driven by the data.

While validation on real data is necessary, these preliminary simulation results suggest that compressive sensing can reduce the number of trials required for accurate estimation of cognitive representations using reverse correlation by up to 90% without any loss of accuracy. Alternatively, more accurate reconstructions could be generated even while still reducing the number of trials, though by a smaller amount. Such a drastic increase in time- and cost-effectiveness would substantially increase reverse correlation's potential for wide-spread use. 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 (e.g., (13)) could be conducted with assumptions only about sparsity; this will lead to clearer, less biased representations. 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., (10, 28, 29)), calcium imaging (30), and auditory system function (31). That the reverse correlation responses can be continuous (e.g., neural firing rates (29)), ordinal (e.g., similarity scores (32)), or binary (e.g., yes/no (22, 25, 26)) gives compressive sensing widespread utility in areas of science even where reverse correlation is not currently in common use.

**References and Notes**


1. M. L. Smith, F. Gosselin, P. G. Schyns, Measuring internal representations from behavioral and brain data. *Current Biology* **22**, 191-196 (2012).
2. A. T. Hendrickson, R. L. Goldstone, in *Proceedings of the thirty-first annual conference of the cognitive science society*. (Cognitive Science Society Amsterdam, Netherlands, 2009), pp. 1084-1089.
3. F. Gosselin, P. G. Schyns, Superstitious perceptions reveal properties of internal representations. *Psychological science* **14**, 505-509 (2003).
4. R. Dotsch, A. Todorov, Reverse correlating social face perception. *Social Psychological and Personality Science* **3**, 562-571 (2012).
5. L. Brinkman, A. Todorov, R. Dotsch, Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology* **28**, 333-361 (2017).
6. E. Barth, B. L. Beard, A. J. Ahumada Jr, in *Human Vision and Electronic Imaging IV*. (International Society for Optics and Photonics, 1999), vol. 3644, pp. 88-96.
7. B. C. Hansen, B. Thompson, R. F. Hess, D. Ellemberg, Extracting the internal representation of faces from human brain activity: An analogue to reverse correlation. *NeuroImage* **51**, 373-390 (2010).
8. L. Varnet, K. Knoblauch, F. Meunier, M. Hoen, Using auditory classification images for the identification of fine acoustic cues used in speech perception. *Frontiers in Human Neuroscience* **7**, (2013).
9. R. L. Gregory, Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **352**, 1121-1127 (1997).
10. E. De Boer, P. Kuyper, Triggered correlation. *IEEE Transactions on Biomedical Engineering*, 169-179 (1968).
11. A. Chauvin, K. J. Worsley, P. G. Schyns, M. Arguin, F. Gosselin, Accurate statistical tests for smooth classification images. *Journal of vision* **5**, 1-1 (2005).
12. E. Ponsot, J. J. Burred, P. Belin, J.-J. Aucouturier, Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences* **115**, 3972-3977 (2018).
13. K. Moon, S. Kim, J. Kim, H. Kim, Y.-g. Ko, The Mirror of Mind: Visualizing Mental Representations of Self Through Reverse Correlation. *Frontiers in Psychology* **11**, 1149 (2020).
14. F. Jäkel, B. Schölkopf, F. A. Wichmann, Does cognitive science need kernels? *Trends in cognitive sciences* **13**, 381-388 (2009).
15. G. Okazawa, L. Sha, B. A. Purcell, R. Kiani, Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nature communications* **9**, 1-16 (2018).
16. R. F. Murray, Classification images: A review. *Journal of vision* **11**, 2-2 (2011).
17. P. J. Mineault, S. Barthelmé, C. C. Pack, Improved classification images with sparse priors in a smooth basis. *Journal of Vision* **9**, (2009).
18. B. A. Olshausen, D. J. Field, Sparse coding of sensory inputs. *Current opinion in neurobiology* **14**, 481-487 (2004).
19. D. H. Hubel, T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology* **195**, 215-243 (1968).
20. B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607-609 (1996).

21. A. Srivastava, A. B. Lee, E. P. Simoncelli, S.-C. Zhu, On advances in statistical modeling of natural images. *Journal of mathematical imaging and vision* **18**, 17-33 (2003).
22. L. Zhang, J. Yi, R. Jin, in *International Conference on Machine Learning*. (PMLR, 2014), pp. 820-828.
23. J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, R. G. Baraniuk, Beyond Nyquist: Efficient sampling of sparse bandlimited signals. *IEEE transactions on information theory* **56**, 520-544 (2009).
24. S. Yotsukura, T. Omori, K. Nagata, M. Okada, Sparse Estimation of Spike-Triggered Average. *IPSJ Online Transactions* **7**, 52-58 (2014).
25. P. T. Boufounos, R. G. Baraniuk, in *2008 42nd Annual Conference on Information Sciences and Systems*. (IEEE, 2008), pp. 16-21.
26. L. Jacques, J. N. Laska, P. T. Boufounos, R. G. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory* **59**, 2082-2102 (2013).
27. N. Mesgarani, S. V. David, J. B. Fritz, S. A. Shamma, Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of neurophysiology* **102**, 3329-3339 (2009).
28. L. Stark, The Pupillary Control System: Its Nonlinear Adaptive and Stochastic Engineering Design Characteristics. *IFAC Proceedings Volumes* **2**, 519-534 (1968).
29. D. Ringach, R. Shapley, Reverse correlation in neurophysiology. *Cognitive Science* **28**, 147-166 (2004).
30. P. Ramdya, B. Reiter, F. Engert, Reverse correlation of rapid calcium signals in the zebrafish optic tectum in vivo. *Journal of neuroscience methods* **157**, 230-237 (2006).
31. A. Ahumada Jr, J. Lovell, Stimulus features in signal detection. *The Journal of the Acoustical Society of America* **49**, 1751-1756 (1971).
32. A. Zymnis, S. Boyd, E. Candes, Compressed sensing with quantized measurements. *IEEE Signal Processing Letters* **17**, 149-152 (2009).

**Acknowledgments:**

We thank Karen Coghlan and Lori Ostapowicz-Critz of the Worcester Polytechnic Institute George C. Gordon Library for their assistance.

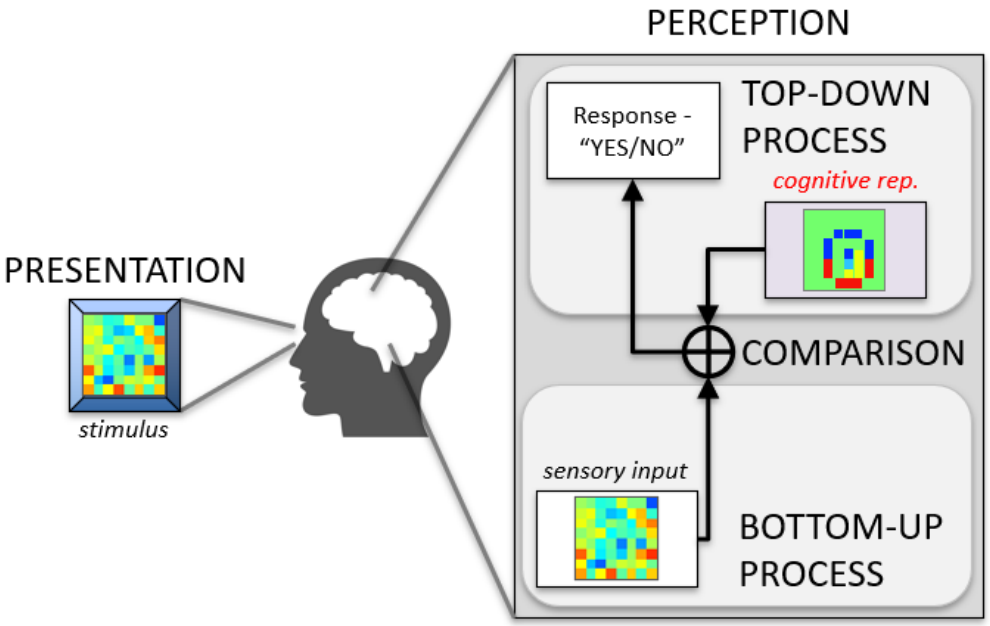


**Fig. 1. A schematic overview of perception.** At its most fundamental level, perception is maintained by a complex cognitive system in the perceiver, involving the combined efforts of bottom-up and top-down processes that bridge the gap between sensory input and cognitive representations. Bottom-up processes extract relevant features from the raw sensory signal, while top-down processes relate incoming features to perceptual categories by comparing against prior expectations in the form of internal, cognitive representations. Such top-down theories are widely posited across multiple domains of human perception (e.g., (9)).

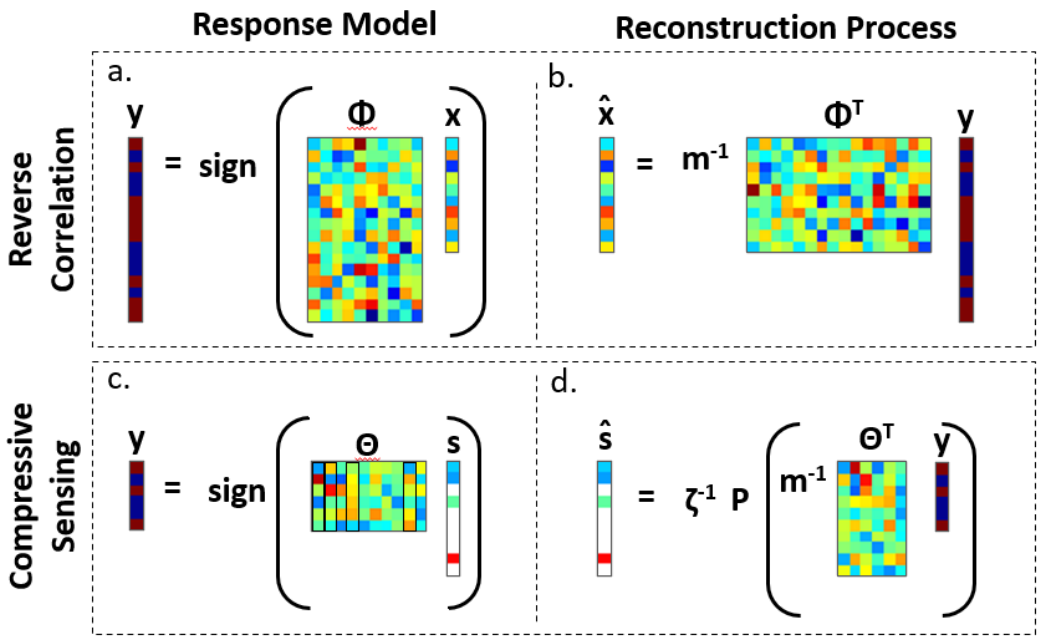


**Fig. 2. Process of reconstruction in reverse correlation and compressive sensing.** (**A)** In reverse correlation, the vector of subject responses is modeled as resulting from the multiplication of a latent representation vector (x) and a stimulus matrix (Φ), which can be thought of as a similarity 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 (Θ). The compressive sensing matrix is formed by multiplying a matrix of basis functions by the stimulus matrix (Θ=ΦΨ), 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 $\zeta=\|P(m^{-1}\Theta^{T}y)\|$,where m is the dimensionality of the stimulus vector and P is a soft thresholding function. The full representation estimate ($\hat{x}$) can be subsequently calculated using the estimated sparse representation and the known basis functions ($\hat{x} = \Psi\hat{s}$). 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.

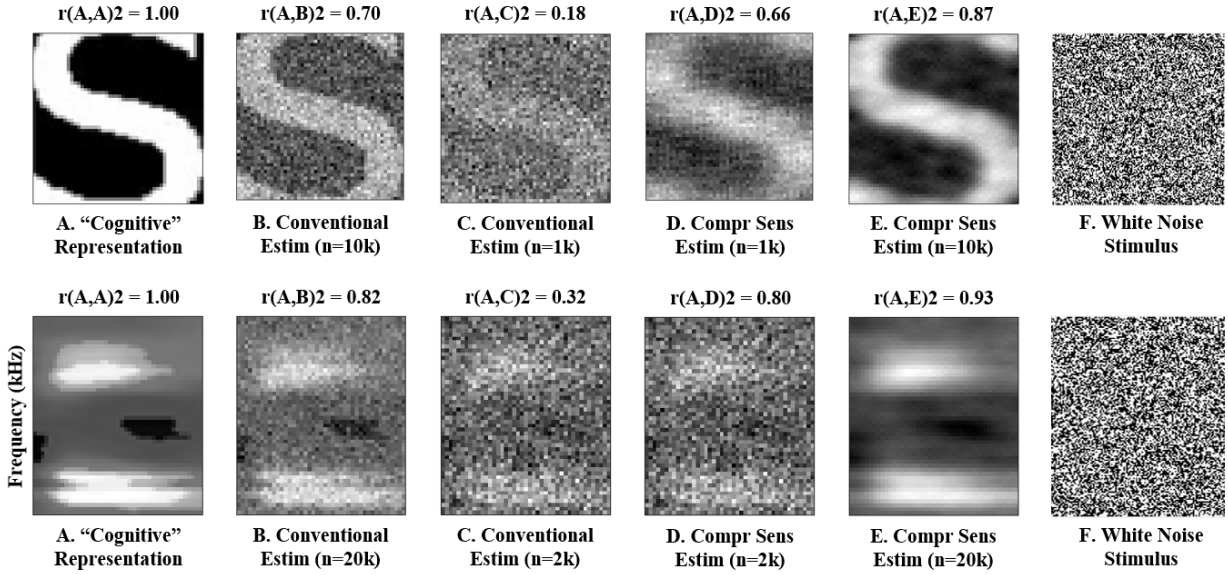


**Fig. 3. Comparison of conventional regression-based estimation and compressive sensing estimation.** The template image **(A)** is estimated in **(B-E)**. An example white noise stimulus for each domain is shown in **(F)**. The method of estimation and number of samples used (n) is indicated below the image. The correlation coefficient ($r^2$) between the template and the estimate is shown above the image, indicting estimation quality. Top row shows a simulation of the "S" from Gosselin & Schyns (3), and bottom row depicts a simulated time-frequency auditory spectrogram (an average phoneme spectrogram presented by Mesgarani et al. (27) is used as the template for reconstruction). In both simulations, compressive sensing provides equivalent accuracy as the conventional approach with only 10% as many trials.

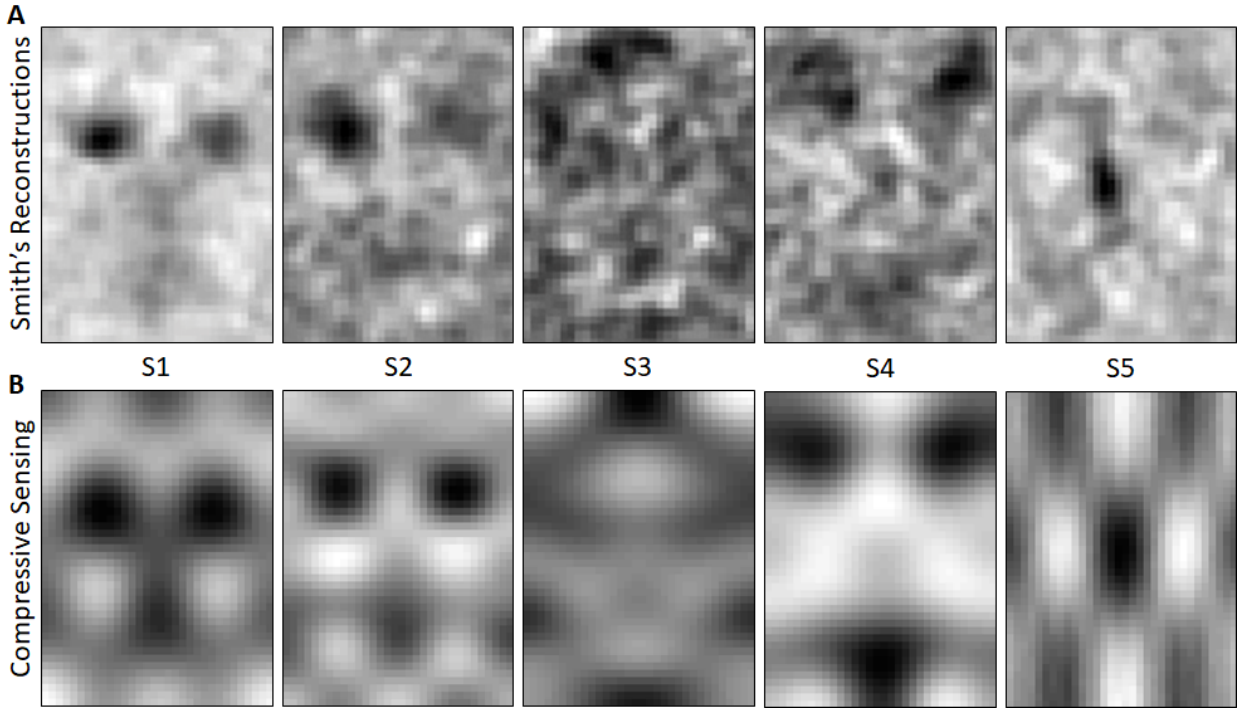


**Fig. 4. Comparison of conventional regression-based estimation and compressive sensing estimation of faces. (A)** shows the reconstructions from (Smith, Gosselin, & Schyns 2012 (1)), and **(B)** depicts the compressive sensing reconstruction. Structural features identified in (A), such as a nose, mouth, and chin outline on S1, S2, and S3, are more noticeable in (B).