Deep Learning–Based Control of Electrically Evoked Activity in Human Visual Cortex

Pehuén Moure[1,*], Jacob Granley[2,*], Fabrizio Grani[3,*],
Leili Soo[3], Antonio Lozano[3], Rocío López-Peco[3], Adrián Villamarín-Ortiz[3],
Cristina Soto-Sánchez[3], Shih-Chii Liu[1,†], Michael Beyeler[2,4,†], Eduardo Fernández[3,5,6,7,†]

[1]Institute of Neuroinformatics, ETH Zurich and University of Zurich, Switzerland.
[2]Department of Computer Science, University of California, Santa Barbara, CA, USA.
[3]Bioengineering Institute and Cátedra Bidons Egara, University Miguel Hernández, Elche, Spain.
[4]Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA, USA.
[5]CIBER Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER BBN), Madrid, Spain.
[6]John Moran Eye Center and Biomedical Engineering, University of Utah, Salt Lake City, UT, USA.
[7]Radboud University, Donders Centre for Neuroscience, Nijmegen, Netherlands.

[*]These authors contributed equally as co-first authors.
[†]These authors jointly supervised this work.
Lead contact: Eduardo Fernández (e.fernandez@umh.es).

SUMMARY

Visual cortical prostheses offer a promising path to sight restoration, but current systems elicit crude, variable percepts and rely on manual electrode-by-electrode calibration that does not scale. These limitations reflect a deeper challenge: electrical microstimulation evokes nonlinear, state-dependent population responses in human visual cortex, complicating the link between stimulation and perception. Here we present a deep learning framework that leverages a bidirectional cortical implant to causally shape stimulation-evoked population activity in human visual cortex. The framework, trained on trial-resolved neural recordings, supports two complementary control strategies, a learned inverse network for real-time stimulation synthesis and a gradient-based optimizer for precise targeting. Both outperform conventional methods, achieve targets at lower stimulation currents, and elicit more consistent perception. Achievable responses lie on the intrinsic low-dimensional manifold of cortical activity, and recorded population activity predicts reported percepts substantially better than stimulation parameters alone. Together, these results provide a population-level foundation for linking microstimulation, cortical activity, and perception in the human visual system.

KEYWORDS

cortical visual prosthesis, microstimulation, deep learning, neural shaping, closed-loop neuroprosthetics

INTRODUCTION

Implanted neuroprostheses offer a promising route to restoring lost sensory function by electrically stimulating neurons in brain areas responsible for perception. Devices targeting vision[? ? ?], audition[?], and somatosensation[? ?] are in various stages of development, yet all face a fundamental challenge: interfacing with highly non-linear networks of biological neurons whose role in perception is not fully understood. Due to the limited spatial resolution of electrical stimulation, prostheses evoke neural response patterns that are unnatural to the brain, often resulting in percepts that are artificial, distorted, or variable[? ? ? ?].

One prominent example of neuroprostheses is the visual prosthesis, which electrically stimulates neurons in the retina[? ? ? ?], optic nerve[?], lateral geniculate nucleus[?], or early visual cortex[? ? ? ?] to restore visual perception. Cortical visual prostheses, which bypass the eye and optic nerve entirely, can both stimulate and record neurons in the primary visual cortex using microelectrode arrays such as the Utah Electrode Array (UEA)[? ?]. This bidirectional interface allows the delivery of patterned stimulation and the measurement of evoked neural responses (Figure **??**A), enabling a neuroscientific testbed for linking patterned microstimulation to trial-resolved population responses in human visual cortex and for quantifying the influence of cortical resting state.

Despite their potential, cortical implants to date have rarely elicited percepts beyond simple shapes or flashes of light[? ? ? ?]. This limitation is not only practical but mechanistic: microstimulation perturbs the ongoing activity within the cortical network, therefore, percepts depend on both the evoked population response and cortical state. Consequently, phosphenes (i.e., percepts induced by electrical stimulation) are often distorted, difficult to interpret, and inconsistent. Their size, brightness, shape, and color vary across time, electrodes, and individuals[? ?]. Percepts from both simultaneous and temporally interleaved multi-electrode stimulation typically do not correspond to the linear sum of individual phosphenes[? ?], reflective of not only overlapping electrical current interactions (*i.e.*, for simultaneous), but also population-level response dynamics shaped by local connectivity and nonlinear integration[? ? ?] (present in simultaneous, interleaved, or sequential stimulation). Selecting stimulation patterns that produce meaningful and stable percepts is, therefore, nontrivial.

The conventional '1-to-1 mapping' strategy assumes a direct visuotopic correspondence between visual space and the electrode grid, mapping camera pixel intensities directly to stimulation amplitudes. This implicitly assumes a stable, approximately linear mapping from stimulation to a perceptual population code, ignoring nonlinear dynamics and state dependence. While alternative stimulation strategies

have been proposed[?,?,?,?,?,?,?,?,?,?,?,?,?,?,?], none have been validated in vivo in humans with perceptually relevant pulse trains, and most do not address the substantial day-to-day variability in cortical states that can limit reliability.

Recent evidence points to a tight coupling between neural responses and perception. Together, these findings motivate treating evoked population activity, rather than stimulation parameters, as the key intermediate variable linking microstimulation to perception. In macaques, intracortical electrode recordings in V4 predicted the current threshold for perception[?], and cortical microstimulation could bias visual perception[?]. In humans, recorded neural activity from a UEA correlated with perceived brightness, detection thresholds, and percept duration[?,?,?]. Additionally, pre-stimulus activity has been shown to influence both neural responses and perceptual thresholds[?,?,?,?]. These findings suggest that closed-loop approaches, which monitor trial-resolved population activity and adapt stimulation accordingly, could improve the consistency and quality of stimulation-evoked perception[?].

Here, we introduce a deep learning-based closed-loop framework for neural activity shaping[?] in a human cortical visual prosthesis. Using multi-session recordings from a participant implanted with a UEA in early visual cortex (Figure **??**A), we trained a forward neural network (NN) to predict single-trial evoked responses from multielectrode stimulation parameters, incorporating pre-stimulus baseline activity to account for day-to-day drift. This model served as an in silico testbed for two complementary control strategies: (1) a gradient-based optimizer that iteratively refines stimulation parameters to minimize response error, and (2) a learned inverse model that generates stimulation patterns from target neural responses in real time (Figure **??**B). In vivo experiments demonstrate that both approaches improve control of population responses compared to 1-to-1 mapping and linear baselines, elicit more consistent percepts, and require lower currents. These findings establish the feasibility of model-driven shaping of neural population responses in a human cortical implant, providing a foundation for closed-loop studies of stimulation, cortical dynamics, and perception.

RESULTS

A State-Dependent Model of Stimulation-Evoked Population Responses in Human Visual Cortex

We first sought to characterize and predict how patterned microstimulation drives trial-resolved population responses in human visual cortex, and how this mapping depends on ongoing cortical state. To this end, we collected a large dataset of stimulation-evoked activity from a 96-channel Utah Electrode Array (UEA) implanted in the early visual cortex of a blind participant (see Supplementary Figure **??**).

Stimulation patterns spanned both random ($N = 5818$, Supplementary Figure **??**A) and structured ($N = 484$, Supplementary Figure **??**B) multi-electrode configurations and were delivered over the course of four months across 26 separate days. On a subset of trials ($N = 5515$), the participant reported phosphene presence (detection task, $N = 4811$) or provided percept descriptions (shape, color, brightness, size; $N = 704$). Neural responses to stimulation were recorded using the same UEA, from which we extracted the envelope multi-unit activity (MUA_e). This signal reflects the pooled spiking activity of multiple neurons near each electrode tip and provides a robust estimate of local population activity[?]. To reduce variability introduced by slow fluctuations in ongoing neural activity and highlight stimulation-evoked changes, we computed change in envelope multi-unit spiking activity ($\Delta\text{MUA}_\text{e}$), defined as the difference between the mean MUA_e in a post-stimulation window ($100\,\text{ms}$ to $200\,\text{ms}$ after stimulus offset) and a per-trial pre-stimulation window ($-110\,\text{ms}$ to $-10\,\text{ms}$). These intervals were selected to avoid contamination by stimulation artifacts while capturing short-latency evoked responses (see Supplementary Figure **??** for effects of different time windows). Because responses to identical stimulation patterns varied substantially across days, we also incorporated average pre-stimulus baseline activity from each day as an operational measure of cortical state to improve across-session generalization (see Supplementary Figure **??**A for the distribution of baseline rates). For more details, see STAR Methods.

The same stimulation generally evoked similar responses but with day-to-day variations (Figure **??**A). For repeated stimulation patterns, the neural response's variance across days was significantly larger

than the variance within a day (Figure **??**B, $p<0.0001$, Wilcoxon signed-rank test; see STAR Methods for details on statistical tests). We found that ΔMUA_e exhibited lower across-day variability for repeated stimuli than post-stimulation MUA_e (Figure **??**C, $p<0.001$), suggesting reduced sensitivity to slow, day-to-day changes in cortical state (see Supplementary Figure **??**A-C). Principal Component Analysis revealed that only 48 components were required to explain 95% of the variability in ΔMUA_e, whereas the stimulation space required all dimensions (Figure **??**D). This suggests that stimulation-evoked population activity occupies a low-dimensional manifold, whereas the stimulation parameterization is high-dimensional.

We trained a deep neural network (DNN) (1.2M parameters, 'Forward NN', see STAR Methods) to predict single-trial ΔMUA_e from stimulation patterns, with average pre-stimulus baseline activity as an additional input to capture day-to-day state. The model was trained on data from 19 days and evaluated on held-out data from 5 non-consecutive days (5123 train, 1179 test), minimizing mean squared error (MSE) by gradient descent.

We compared the model against baselines that test common assumptions about stimulation-evoked responses (focal activation, linear superposition, and dependence on baseline state; see STAR Methods): (1) a conventional receptive field model assuming focal activation and thus a one-to-one mapping between stimulation amplitude and neural activity ('1-to-1'); (2) linear regression with access to average pre-stimulation activity ('Linear'); (3) a linear-nonlinear model that applies a saturating pointwise nonlinearity to a linear projection of the input ('Linear Non-linear'); (4) a dictionary-based model using an unweighted average of recorded single-electrode responses for each electrode in the multi-electrode stimulation pattern ('Dictionary'); (5) an optimally weighted combination of single-electrode responses ('Optimal Dictionary'); (6) the forward NN without access to pre-stimulation activity ('NN Without Pre-Activity'); and (7) a random baseline using responses to random stimuli from the training set ('Random'), as an estimated lower performance bound.

Performance was evaluated using MSE, R^2 score averaged across electrodes, and an adapted Wasserstein Earth Mover's Distance (EMD)[?] , which accounts for the spatial layout of the electrodes (STAR Methods). Only channels which i) elicited spikes on at least one day and ii) had a variance in neural response across all days greater than a threshold were included in the analysis, resulting in 54/96 valid channels (STAR Methods and Supplementary Figure **??**).

The forward NN model's predictions aligned well with recorded responses from the test set (Figure **??**E). Although the MSE was low on most channels, R^2 was higher on the channels near the bottom or upper right of the array, *i.e.* the channels with more spiking activity (Figure **??**F). Our forward NN model predicted responses with significantly higher R^2 and lower MSE and EMD than all baselines (Figure **??**G, $p < 0.0001$). This trend was also consistent regardless of the time window chosen for neural activity prediction (Supplementary Figure **??**). The neural network outperformed linear baselines, linear-nonlinear baselines, and even the idealized optimal dictionary, providing strong evidence that neural responses from hybrid simultaneous/interleaved multi-electrode stimulation are not linearly related to either stimulation parameters or a superposition of single-electrode responses.

Deep Learning Enables Causal Shaping of Electrically Evoked Population Activity

Having established a state-dependent forward model, we next asked whether targeted stimulation patterns can be synthesized to causally shape stimulation-evoked population responses in awake humans. Prior attempts to evoke specified population responses have had mixed success and have largely been evaluated in simulation[? ?] , ex vivo preparations[?] , or non-human primates[?] , leaving open whether comparable shaping is achievable in awake humans.

To achieve this goal, we evaluated two complementary approaches that treat the forward model as an in silico proxy for the stimulation-to-population-response mapping. Inspired by the success of gradient-based optimization in controlling populations of neurons via visual stimuli[? ?] , the first approach was a gradient-based method that directly searched the stimulation parameter space (*i.e.* the current to stimulate with on each electrode; 'Gradient'). An iterative gradient descent algorithm optimized a stimulation pattern so that, when the stimulus was put through the forward model, the predicted neural response minimized the MSE with the target response.

Gradient descent is computationally expensive and may not be fast enough for real-time deployment.

Therefore, the second NN approach consisted of training an inverse neural network ('Inverse NN') to map directly from target neural responses to stimuli in a single pass. This inverse network was trained alongside the forward model (with frozen weights), similar to an autoencoder[? ? ?] . The network learned to output stimuli that, when input through the forward model, yielded neural responses that minimized MSE with the desired target (STAR Methods). Both the Gradient and Inverse NN methods additionally utilized the average (per-session) pre-stimulus baseline activity as an additional input, to allow adjustment for day-to-day variations.

To evaluate these shaping approaches in vivo, optimized stimuli were generated and delivered using the prosthesis over two evaluation days, with the participant reporting whether they perceived a phosphene. Below, we compare the performance of these approaches for both natural and synthetic target responses. Natural targets were trial-resolved responses from held-out data for which a stimulus had previously been observed to evoke the response on another day (Supplementary Figure **??**). Synthetic targets were structured activity patterns designed to probe responses outside the distribution of recorded activity (Supplementary Figure **??**).

The gradient and inverse network approaches were benchmarked against several baselines: (1) the conventional one-to-one mapping ('1-to-1'), (2) a linear regression model ('Linear'), and (3) an inverse dictionary approach ('Dictionary'), where stimulation was computed as a weighted average of the previously recorded stimuli, weighted by target electrode activity (similar to a spike triggered average). For natural targets, the direct replay of stimuli that previously evoked the target responses ('Replay Stimulus') was included as a reference quantifying day-to-day consistency of responses evoked by the same stimulus. A random approach was also included as a lower performance bound, where performance was averaged across responses from random stimuli in the training dataset. The neural activity shaping methods were evaluated using MSE, R^2 score averaged across electrodes, and adapted Wasserstein EMD. All metrics were computed on single trials (no trial averaging).

Reproducing Natural Target Responses

For natural target responses, the goal was to reproduce known evoked neural responses on a new day, without knowing the original stimulation pattern. Both the inverse NN and gradient optimization significantly outperformed all baselines in reproducing the target neural responses (Figure **??**A,C, $p< 0.05$, more examples in Supplementary Figure **??**). Gradient optimization yielded the best overall performance, achieving errors comparable to replaying the original stimuli. Although gradient optimization achieved the best overall results, it required $\sim 10-20\,\mathrm{s}$ per example on a desktop GPU, whereas the inverse NN inferred stimuli in $\sim 50\,\mu\mathrm{s}$, a difference that could prove critical in real-time applications. Importantly, the perception reported by the participant for DNN-optimized stimulation more closely matched the participant's report of the original stimuli than baselines for the detection task (see STAR Methods), shown in Figure **??**C.

Generalizing to Synthetic Target Responses

A key question is whether stimulation can shape responses beyond those previously observed, or whether shaping is constrained to the endogenous response repertoire. To test generalization, we used synthetic targets: structured patterns that may lie outside the distribution of recorded evoked activity. These targets were manually selected to test a variety of discrete shapes, commonly used response patterns (*e.g.*, focal activation), and spanned both inhibitory and excitatory responses (Supplementary Figure **??**).

Compared to natural targets, synthetic targets had worse MSE but, interestingly, improved R^2 and EMD for all methods (Figure **??**). However, the relative improvement compared to random stimulation was smaller for synthetic targets than natural, suggesting that neural activity shaping was not as effective as for natural targets. Both the inverse NN and gradient approaches significantly outperformed the baseline methods in shaping neural responses ($p < .05$, Figure **??**A,B, see Supplementary Figure **??** for more examples). MSE was higher for synthetic targets than for natural responses, suggesting constraints imposed by the stimulation paradigm, the training data, or cortical dynamics. Notably, gradient inversion again provided the closest match to the targets, while the inverse NN again showed slightly inferior performance.

Optimized stimulation patterns produced with both gradient-based optimization and the inverse NN achieved target neural responses while requiring lower overall stimulation currents (Figure **??**C), representing a substantial improvement in stimulation efficiency and safety. To test whether this improvement was simply due to the use of lower amplitudes, we examined the relationship between stimulation current and model performance. Only a weak correlation was observed ($r = 0.25$, Figure **??**D), indicating that the superior performance of the DNN-based approaches cannot be explained by current reduction alone.

Performance was instead governed by a geometric constraint: the intrinsic low-dimensional structure of stimulation-evoked population activity. Similar to observations in visually evoked activity[?] and other brain regions[?] , electrically evoked responses were confined to a low-dimensional neural manifold, quantified using latent factor analysis[? ?] (STAR Methods). After accounting for noise, 95% of the manifold variance in evoked responses could be explained with just 10 latent factors, compared with 86 factors required to explain 95% of the variance in stimulation parameters. Natural target responses clustered near the manifold, while synthetic targets deviated further (Figure **??**E, $p < 0.001$, t-test). The farther a target response lay from the neural manifold, the harder it was to control, with a strong correlation between manifold distance and model error ($r = 0.85$, Figure **??**F). These results suggest that precise control of evoked neural activity is fundamentally bounded by intrinsic response geometry, not simply by stimulation parameters.

Together, these results demonstrate targeted shaping of stimulation-evoked population responses in an awake human cortical implant.

Evaluating Simulated Neural Responses to Optimized Stimuli

A critical requirement for stimulus optimization is the ability to accurately predict neural responses before delivering stimulation in vivo. If the forward model provides a reliable approximation of trial-resolved responses, it can serve as an in silico surrogate for comparing candidate stimulation strategies and targets. At the same time, any mismatch between simulated and recorded responses must be quantified to define the model's limits.

To evaluate predictive accuracy, we compared simulated responses (predicted by the forward model for optimized stimuli) with the corresponding in vivo responses recorded from the participant (Figure **??**A,B).

Across all inversion approaches, simulations systematically overestimated shaping performance relative to in vivo outcomes (lower MSE and EMD, higher R^2). Importantly, this bias was consistent across approaches, and the relative ordering of stimulation synthesis methods was preserved.

Despite this bias in absolute performance, simulated and in vivo MSE were highly correlated across targets ($r = 0.86$ for natural and $r = 0.95$ for synthetic targets; Figure **??**C). Thus, while simulations underestimate absolute error, they reliably predict which targets and strategies are more achievable in vivo, supporting the forward model as a useful surrogate for stimulation synthesis.

Consistent with our in vivo findings, target difficulty in simulation was strongly linked to manifold distance: simulated MSE increased with the simulated target's distance from the evoked-activity manifold (Figure **??**D).

Together, these results reinforce that intrinsic response geometry constrains the achievable space of stimulation-evoked population activity in both in-silico and in-vivo settings.

Decoding Perception from Neural Responses

A central question is whether trial-resolved neural population responses provide a meaningful intermediate variable linking stimulation to perception. In principle, one could attempt to predict percepts directly from stimulation parameters, bypassing neural recordings. We therefore tested whether recorded neural activity predicts perceptual reports better than stimulation parameters alone.

Deep neural network (DNN) decoder models were trained to predict the participant's behavioral reports of phosphene detection (seen or not seen, $N = 4416$), brightness (subjective rating 1-5, $N = 427$), and color (4 color categories, $N = 427$). Decoders were trained using different input features: stimulation parameters alone, neural responses alone ($\Delta\text{MUA}_\text{e}$ or $\Delta\text{MUA}_\text{e}$ plus average pre-stimulus baseline

MUA_e), or combinations of these inputs. Models were trained via gradient descent to minimize prediction loss, with performance evaluated using leave-one-day-out cross-validation to assess generalization across recording sessions (see STAR Methods for details).

Neural activity significantly outperformed stimulation parameters alone in predicting detection, color, and brightness ratings (Figure **??**A-C, $p<0.05$). Including neural activity (pre-stimulus MUA_e and ΔMUA_e) alongside stimulation parameters improved detection accuracy from 77.0% to 87.8% (Figure **??**D-F, $p < 0.05$), corresponding to a 47% reduction in classification error. The same trends in accuracies of different predictors were observed across varying time windows, with maximum predictive accuracy occurring near the chosen 100ms - 200ms interval (Supplementary Figure **??**). Similar improvements were observed for color (49.4% to 76.7%, 54% error reduction) and brightness (44.8% to 60.9%, 29% error reduction) prediction.

Together, these results show that trial-resolved population activity carries substantially more information about perceptual reports than stimulation parameters alone, supporting neural responses as the appropriate intermediate variable linking microstimulation to perception.

DISCUSSION

This study demonstrates that stimulation-evoked population activity in human visual cortex can be predictably modeled and deliberately shaped despite strong nonlinearity and state dependence, using data-driven models trained on intracortical recordings. Using a forward model trained on months of recordings, we accurately predicted neural responses across substantial day-to-day variability, enabling in-silico exploration of stimulation strategies. Our deep learning-based inverse models shaped neural activity toward target patterns while requiring lower stimulation currents and eliciting more consistent perception than conventional approaches. Together, these findings identify neural population responses, rather than stimulation parameters alone, as the critical intermediate variable linking microstimulation to perception. Although demonstrated here in the context of cortical visual prostheses, this framework provides a generalizable foundation for adaptive, closed-loop control across bidirectional neural interfaces. Importantly, this study does not seek to demonstrate improved visual perception directly, but instead addresses a more fundamental prerequisite that has not previously been established in humans: that stimulation-evoked cortical population responses can be predictably controlled across substantial neural variability. Any future improvement in prosthetic perception is likely to depend on this controllability. In this sense, the present work establishes the missing intermediate step between microstimulation and perceptual optimization.

Necessity of Nonlinear Neural Activity Models When predicting neural responses, our deep NN forward model significantly outperformed multiple linear approaches (Figure **??**G), demonstrating that stimulation-evoked responses are strongly nonlinear and not well captured by the linear receptive-field assumptions often used in cortical prostheses. These results highlight the need for nonlinear, scalable models to capture the relationship between stimulation and population activity, particularly as implants increase in channel count and density.

This nonlinearity was mirrored in the inverse problem, where deep learning-based approaches substantially outperformed conventional stimulus optimization methods (Figures **??** and **??**). Nonlinear response structure therefore constrains not only prediction but also the synthesis of stimulation patterns required to achieve desired population responses.

Both inverse approaches achieved superior neural control compared to 1-to-1 mappings and linear models, with gradient-based optimization yielding the best performance. However, its computational cost ($\sim$20 seconds per stimulus) currently limits real-time feasibility, whereas the inverse network performs inference in $\sim$50 microseconds. Hybrid strategies combining rapid inference with periodic optimization may offer a practical compromise between speed and accuracy.

Activity shaping showed smaller improvements for synthetic than for natural targets. This discrepancy may reflect either modeling limitations outside the training distribution or more fundamental constraints of cortical microstimulation, such as current spread, circuit organization, or activation of passing axons. The fact that the forward model generalized well to out-of-distribution stimuli (Figures **??** and **??**), and that

the inverse models relied on this forward model, suggests that part of the error reflects intrinsic limits of achievable cortical responses rather than model failure.

We found that stimulation-evoked responses lie on a low-dimensional neural manifold, and that distance from this manifold strongly predicts elicitation error. While low-dimensional structure has been described for visually evoked activity ([?], but see also[?]), it is notable that the same constraint governs electrically evoked responses. This constraint likely reflects intrinsic circuit organization and may limit what stimulation optimization alone can achieve for visual prostheses.

Importantly, the neural activity window analyzed here captures responses well after stimulation has ceased, reflecting network-mediated cortical dynamics rather than direct electrical artifacts. Although this introduces additional complexity, these distributed dynamics are the substrate of perception. The fact that responses within this regime remain predictable and shapeable demonstrates that the relevant neural substrate for perception is accessible to model-based control.

Together, these findings indicate that the space of achievable stimulation-evoked responses is constrained by intrinsic cortical response geometry, imposing limits that cannot be overcome by stimulation optimization alone. Future work will need to determine whether these constraints can be expanded through adaptive training paradigms or broader stimulation strategies.

Generalization and the Forward Model as a Predictive Tool A central result of this study is the strong correspondence between simulated and in vivo neural responses (Figure **??**), establishing the forward model as a reliable predictive tool for evaluating stimulation strategies. Even for synthetic targets, the model produced consistent and generalizable approximations of neural activity. This predictive accuracy enables systematic comparison of candidate stimulation patterns without requiring exhaustive in vivo testing, providing a practical framework for refining stimulation protocols and adapting prostheses to individual users.

A persistent challenge in neurostimulation is that identical stimulation patterns can evoke different neural responses across days, contributing to variability in perception reported in prior studies[?]. Our approach mitigates this issue by modeling the change in neural activity before and after stimulation (ΔMUA_e), which is substantially more stable across days (Figure **??**A-C). In addition, conditioning stimulation on the average daily pre-stimulation baseline allows the model to adapt to slow changes in latent cortical state. Conditioning on per-trial pre-stimulus activity did not further improve performance, indicating that compensating for daily state drift is sufficient for shaping neural responses.

Perceptual Relevance and Implications for Visual Prostheses Stimuli synthesized by our inverse models (i.e., stimuli generated to match target neural population responses) evoked percepts that more closely resembled those associated with the natural target responses (Figure **??**C), while requiring lower stimulation amplitudes (Figure **??**C). In addition, perceptual descriptions were decoded substantially more accurately when neural responses were included than when using stimulation parameters alone (Figure **??**). Together, these results indicate that neural population activity, rather than stimulation parameters per se, is a more proximal predictor of perceptual outcomes.

A remaining challenge is determining which neural population response corresponds to a desired percept. One possibility is to leverage image-computable neural networks that predict primate and human V1 activity[? ?]. Alternatively, the perception decoders developed here could be incorporated directly into the optimization loop, enabling stimulus synthesis that is jointly informed by neural activity and behavioral reports. This approach is consistent with prior demonstrations of model-informed microstimulation for modulating perception in sighted primates[?] and suggests a pathway toward perception-driven stimulation strategies that integrate neural and behavioral feedback.

Limitations and Future Directions Before considering broader implications, it is important to acknowledge several limitations of the present study. First, the study was conducted in a single participant, and generalization across individuals remains to be established.

Second, the present study used a Utah Electrode Array because it provided a rare bidirectional human cortical interface capable of both stimulation and recording. However, rigid penetrating microelectrode ar-

rays face important chronic stability challenges, including tissue encapsulation, changes in stimulation thresholds, loss of recording yield, mechanical degradation, and variability across individuals and arrays[? ? ? ?]. Thus, the present work should not be interpreted as demonstrating that the specific array technology used here is sufficient for chronic clinical vision restoration. Rather, it provides a proof of principle for model-based control of stimulation-evoked cortical population activity. The framework itself is interface-agnostic: it requires patterned stimulation and concurrent neural recording, and could in principle be applied to future bidirectional multi-electrode interfaces with improved long-term stability, broader cortical coverage, or higher channel density.

Third, although our methods improved the predictability and shaping of stimulation-evoked population responses, they did not produce qualitatively richer percepts or measurable gains in visual acuity. This limitation reflects the design and constraints of the present study rather than a general impossibility of acuity testing with prosthetic vision. The experiments were designed to test neural controllability and perceptual report consistency; and not to estimate spatial resolution. Acuity-like measurements would require a dedicated psychophysical protocol with calibrated percept locations and repeated forced-choice discrimination of spatial patterns, such as gratings or separated shapes. In the present participant, such measurements were further complicated by bilateral enucleation after traumatic injury, substantial oculomotor impairment, and percepts that were confined to a small region of visual space and often appeared in close spatial proximity. Under these conditions, the behavioral measures reported here are best interpreted as probing perceptual detection, brightness, color, and consistency rather than clinical visual acuity. Phosphene location can also shift with eye position even in participants without remaining vision[? ? ?], so a stable oculocentric reference frame is typically required to construct calibrated spatial discriminanda. In the present participant, severe oculomotor impairment made such calibration impractical within the existing protocol.

These constraints make the present work a conservative test of the approach. The limited cortical coverage and resolution of a single Utah array likely restrict the range of achievable percepts, suggesting that the benefits of neural activity shaping demonstrated here would be expected to be substantially greater in future systems with denser arrays or broader V1 coverage. The current results therefore establish feasibility under minimal conditions rather than representing the upper bound of what this framework could enable.

Fourth, while the model captured nonlinear neural responses across days, it does not account for long-term plasticity or adaptive changes in cortical processing. Finally, although the inverse models outperformed traditional methods, further improvements are needed to make gradient-based optimization fully practical for real-time use.

These limitations define the boundary conditions of the current clinical interface and our presented experimental design, but do not diminish the broader implications of being able to model and control cortical population activity. They also clarify the next steps required to move from neural activity shaping toward clinically meaningful perceptual improvement.

Future Directions for Systems Neuroscience Beyond prosthetic applications, this framework enables a new experimental paradigm for probing causal population dynamics in the human cortex. The ability to synthesize targeted neural activity patterns and measure their downstream effects opens opportunities for studying functional connectivity, hierarchical processing, and the geometry of cortical representations using combined stimulation and recording across areas. This suggests that model-based neural control may become a powerful tool not only for neuroprosthetics but also for answering fundamental neuroscience questions about how cortical circuits transform and propagate information.

Resource Availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Eduardo Fernández (e.fernandez@umh.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All neural recording and stimulation data reported in this paper have been deposited at GitHub (`https://github.com/jgranley/deep_neural_shaping`) and are publicly available as of the date of publication. The repository URL is also listed in the key resources table.
- All code is also available at GitHub (`https://github.com/jgranley/deep_neural_shaping`).
- Any additional information required is available from the lead contact upon request.

Acknowledgments

This work was financially supported by the Ministerio de Ciencia, Innovación y Universidades (Grants DTS19/00175 and PDC2022-133952-100); the Generalitat Valenciana, Directorate General of Science and Research (PROMETEO Grant CIPROM/2023/25); the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 899287 (NeuraViPeR) and Marie Skłodowska-Curie Grant Agreement No. 861423 (enTRAIN Vision); the Innovative Neurotechnology for Society (INTENSE) Dutch Neurotechnology Consortium; and the Swiss National Science Foundation (project CA-DNNEdge 208227). This work was also supported by the National Library of Medicine of the National Institutes of Health under Award Number DP2-LM014268 to M.B. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

Conceptualization, P.M., J.G., F.G., M.B., S.-C.L., and E.F.; data collection and curation, E.F., F.G., L.S., C.S.-S., P.M., J.G., A.V.-O., and R.L.-P.; methodology, J.G., P.M., F.G., and A.V.-O.; data analysis, F.G., P.M., and J.G.; writing – original draft, F.G., J.G., and P.M.; writing – review & editing, all authors; supervision, S.-C.L., M.B., and E.F.

Declaration of Interests

The authors declare no competing interests.

Main figure titles and legends

Figure 1. Schematic representation of the experimental setup and modeling framework. **A**) A stimulation pattern across electrodes is selected and sent to a UEA, which delivers the stimulus to the primary visual cortex of a human participant. The electrical activity under each electrode is recorded, amplified, and multi-unit spiking activity is extracted. **B**) Visualization of three approaches for neural response targeting: the conventional 1-to-1 mapping, the proposed inverse neural network method, and the proposed gradient-based optimization. Generated stimuli can then be played *in vivo* or input to the forward model to obtain real and simulated neural responses.

Figure 2. Modeling the neural response to stimulation. **A)** Example neural responses ($\Delta\text{MUA}_\text{e}$) over 3 days to the same stimulation pattern and the corresponding average prestimulus activity (MUA_e) for each day. **B)** Comparison of the variance of $\Delta\text{MUA}_\text{e}$ neural responses to repeated stimulations across days and within days, averaged across all channels. Red triangles represent the mean across all repeated

stimulation patterns. **C)** Comparison of the variance across days of neural responses to repeated stimulations using (*left*) the difference in activity with respect to the activity before stimulation (ΔMUA$_e$) and (*right*) the activity after stimulation (post MUA$_e$), averaged across all channels. **D)** PCA analysis showing explained variance as a function of number of components, for stimulation patterns and neural responses, across the collected dataset. **E)** Examples of forward model predictions and in vivo neural responses on held out samples. **F)** Forward model MSE and R^2 for each channel (selected spiking channels are outlined in black). **G)** Performance of the forward model in predicting the neural response to stimulation, as measured by MSE, R^2 and EMD, and compared to baseline approaches. Bars with asterisks denote significant p values: *, **, *** are p $<$ 0.05, 0.01, and 0.001 respectively. Mean and standard error of the mean (SEM) shown.

Figure 3. Evaluation of neural activity shaping methods. **A)** Optimized stimuli (*top*) and in vivo neural responses (*bottom*) for each shaping method for an example natural target response. **B)** Comparison of errors for each activity shaping method using MSE, R^2, and EMD. Bars with asterisks denote significant p values: *, **, *** are p $<$ 0.05, 0.01, and 0.001 respectively. Error bars represent SEM. **C)** Confusion matrices showing whether reported perceptions matched the previously observed perceptions corresponding to the same target neural activity.

Figure 4. Evaluation of neural activity shaping methods for synthetic targets. **A)** Optimized stimuli (*top*) and in vivo neural responses (*bottom*) for each activity shaping method for an example synthetic target response. **B)** Comparison of errors for each method using MSE, R^2, and Wasserstein EMD. Bars with asterisks denote significant p values: *, **, *** are p $<$ 0.05, 0.01, and 0.001 respectively. Error bars represent SEM. **C)** Average stimulation amplitude for each method across natural and synthetic targets. **D)** Amplitude was only slightly correlated with neural response reconstruction error. **E)** Kernel density estimate plot showing the distance to the low dimensional (D=10) natural neural response manifold across target neural responses. The manifold was quantified using latent factor analysis. Synthetic targets lied further from this manifold than natural targets. **F)** Target responses which were further from the latent manifold had larger reconstruction error (shown for 'Gradient' responses).

Figure 5. Generalization of neural activity shaping strategies from simulated models to in-vivo measurements. **A)** Simulated neural responses (*bottom*) and *in vivo* responses (*top*) for an example target. Simulated responses were obtained by running the synthesized stimulus through the forward neural network model. **B)** Comparison of errors for simulated and in vivo responses across methods. **C)** Simulated vs in vivo neural response reconstruction error (MSE) for each shaping method, for natural (*left*) and synthetic (*right*) targets. **D)** Target responses with a larger distance to the latent neural manifold also had larger simulated errors.

Figure 6. Perceptual relevance of evoked activity patterns. **A-C)** Accuracy of decoder models for phosphene detection ($N = 4416$), brightness ($N = 427$), and color ($N = 427$) perception, decoding using different input features. Leave-one-day-out cross validation was used, and performance aggregated across predictions from different folds (mean $\pm$ SEM). Asterisks denote significant differences when compared to stimulus only: *, **, *** are p $<$ 0.05, 0.01, and 0.001 respectively. The dashed line is the stimulus-only condition, repeated for clarity. **D-F)** F1 scores for the same perceptual attributes and model inputs as A-C.

STAR METHODS

Experimental Model and Study Participant Details

Human Study Participant The volunteer was a 27-year-old blind male with good physical and mental health. He had lost his vision due to a traumatic head injury, after which bilateral enucleation was performed one year prior to his participation. Due to the sensitive nature of this single-subject study ($N = 1$) and to ensure the preservation of the participant's anonymity, specific details regarding ancestry, race, ethnicity, and socioeconomic status have been omitted. We acknowledge that data restricted to a single male participant limits the generalizability of these findings across sexes, genders, and diverse demographic backgrounds. Following bilateral enucleation, the participant wore cosmetic ocular prosthe-

ses and exhibited substantial oculomotor impairment with limited and unstable voluntary eye movements. These constraints precluded reliable construction of an oculocentric reference frame for spatial percept mapping. Behavioral measures in this study therefore focused on phosphene detection, brightness, color, and consistency rather than acuity-style spatial discrimination.

Ethical Oversight The experiments were conducted as part of a clinical research trial for the development of a cortical visual prosthesis for the blind based on intracortical microelectrodes. The study adhered to the principles outlined in the Declaration of Helsinki, and all relevant ethical guidelines associated with clinical trial regulation were followed, including EU No. 536/2014 (repealing Directive 2001/20/EC) and EU Commission Directives 2005/28/EC and 2003/94/EC. The clinical trial was approved by the Clinical Research Committee of the General University Hospital of Elche. The participant provided written informed consent prior to their participation. The confidentiality and privacy of the participant were maintained throughout the study, and all procedures were designed to minimize any potential risks to their well-being.

Method Details

Surgical Procedure and UEA Implantation At the start of the trial, one Utah Electrode Array (UEA) (Blackrock Neurotech, NeuroPort Electrode) was implanted into the early visual cortex as an interface for the cortical visual prosthesis[??]. The implant was placed in the right occipital cortex, likely in or near V1 based on cortical anatomy (see Supplementary Figure **??**). However, fMRI verification of its precise location was not possible due to the presence of metal fragments in the skull.

The UEA consists of 100 $1.5\,\mathrm{mm}$-long electrode shanks (of which 96 are effectively connected to the pedestal connector) that can electrically stimulate the visual cortex to induce visual perception and record the neural activity. A robot-assisted procedure was followed in the UEA implantation[?] in addition to the standard surgical procedure[?]. A platinum wire implanted under the dura was used as reference electrode. After 6 months, the subject underwent another surgery where the UEA was explanted.

Neural Recording and Stimulation The Ripple Neuromed Summit Explorer processor was used for both neural recording and electrical stimulation. It was connected to the UEA connector through three Micro2+Stim front-ends (32 channels each). Neural signals were sampled at $30\,\mathrm{kHz}$ from 96 channels applying a $0.1\,\mathrm{Hz}$–$7.5\,\mathrm{kHz}$ analog filter. The resolution of the analog-to-digital conversion was set to $0.25\,\mu\mathrm{V}$ (16 bit). The stimulation step-size was set to $1\,\mu\mathrm{A}$. The trigger of electrical stimulation was recorded as a digital event in the stimulation data stream. We used the fast settle option of the processor for all recording channels. With this option, the recording of each channel was blanked for the duration of each stimulation pulse on any electrode, and an additional $0.5\,\mathrm{ms}$ after. Electrical stimulation was controlled through a customized Python interface which allows easy configuration of the stimulation parameters and timing for each channel.

Stimulation Protocols A set of stimulation parameters was used that has proven effective in inducing visual perception with intracortical electrodes, based on previous literature[??]: a train of 50 pulses lasting 167 ms (300 Hz), where each pulse is rectangular, cathodic first, and charge-balanced (cathodic and anodic duration of $170\mu s$, interphase duration $60\mu s$).

For safety reasons, pulses from individual electrodes of multielectrode stimulation patterns were temporally interleaved to minimize the instantaneous current delivered to the neural tissue (*i.e.*, stimulating during the interpulse gap of other stimulating electrodes, Supplementary Figure **??**C). For a stimulation train at 300 Hz and a single-pulse duration of 400 μs there are 2.93 ms between pulses. We distributed up to 6 pulses from different electrodes in this interpulse period. In case more than 6 electrodes were included in the stimulation pattern, hybrid simultaneous/interleaved stimulation would distribute the first 6 electrodes' pulses during the inter-pulse interval, then overlay the next electrodes' pulses to be simultaneous with the first few electrodes' pulses (*i.e.*, the seventh electrode's pulses would simultaneously coincide with first electrode's pulses, and so on, see Supplementary Figure **??**C). Under this method, exactly 2 electrodes delivered pulses at the same time. The order of interleaved stimuli was the ascending

order of stimulating electrode numbers. Across all data, 3697 out of 7398 (49.9%) of stimuli were hybrid simultaneous/interleaved (number of electrodes > 6). For optimized stimuli, 741 out of 789 (93.9%) and 731 out of 856 (85.4%) of optimized stimuli were hybrid simultaneous/interleaved for natural and synthetic target responses, respectively.

During the 6 months of implantation (after which the subject was explanted), the subject performed daily sessions where electrical stimulation of the visual cortex induced visual perceptions (phosphenes). Experimental sessions usually lasted 4 hours per day, 5 days per week. The data collected for this work span across 26 daily sessions, with the last 2 sessions dedicated to the optimized stimulation patterns. In general, the visual perceptions he described were dots or half-circle shapes, and varied in size, color and brightness, around 2 degrees visual angle. Phosphene locations moved slightly as the stimulating electrode was varied, although this change was often reported to be difficult to detect. Single electrode thresholds were found on 27/96 electrodes. Thresholds ranged from 5 to $90\,\mu\mathrm{A}$ (mean $24\,\mu\mathrm{A}$, standard deviation $18\,\mu\mathrm{A}$).

Three types of stimulation patterns were used:

1. **Random ($N = 5818$):** A random group of electrodes (between 1 and 10 total) was selected. For each selected electrode, a random stimulation current was chosen between $5\,\mu\mathrm{A}$ to $50\,\mu\mathrm{A}$. Examples of random stimulation patterns are shown in Supplementary Figure **??**A.
2. **Structured ($N = 484$):** Occasionally, the patterns were not random but followed structured shapes such as rectangles, triangles, circles or lines. For these structured patterns the current from all the electrodes was set to $15\,\mu\mathrm{A}$ or $45\,\mu\mathrm{A}$. Examples of structured stimulation patterns are shown in Supplementary Figure **??**B.
3. **Single Electrode ($N = 2389$):** Additionally, we collected data to determine the current threshold and neural response for each individual electrode, which was then used as a single-electrode dataset to build the forward dictionary models. To find the minimum current needed to induce perception from each electrode, a binary search procedure was used**?** . This led to the collection of a dataset of neural responses to each single electrode stimulation for different currents. Note that this was not included in the training data for the forward and inverse neural network models.

Behavioral Tasks In the experiments, the participant was sitting comfortably in a chair, and a sequence of stimulation patterns was delivered to the UEA. After each stimulation, there was a minimum delay of $1\,\mathrm{s}$ before the next stimulus. Three tasks were performed, differing in the answers requested from the participant:

1. **No answers needed:** This allowed the collection of the neural response to a large number of stimulation patterns in a short amount of time. $N = 787$.
2. **Detection task:** The participant reported via keyboard whether the stimulation pattern induced a visual perception or not. $N = 4811$ (25 % of which led to perception).
3. **Description of the shape, brightness, color, and size of perception:** The participant reported the description by voice, and the experimenter collected the answer in an Excel table. Shape and size were free-form responses which did not vary much throughout data collection, brightness was rated between 0 and 5, and color was a free-form response, which was later grouped into 4 color categories (green, yellow, white, and ambiguous). This protocol took the longest amount of time, so fewer stimulation parameters were tested. $N = 704$.

When using optimized stimulation patterns, the detection task was used, except the patterns of stimulation were no longer random but determined by the output of the various stimulation synthesis methods.

Neural Activity Extraction We calculated the neural activity for each stimulation using $\mathrm{MUA_e}$, an averaged representation of the aggregate spiking activity from several neurons near the electrode tip, as defined in**?** . The raw signal is first filtered between 0.5 and $9\,\mathrm{kHz}$, full-wave rectified, low-pass filtered at $200\,\mathrm{Hz}$, and down-sampled at $1\,\mathrm{kHz}$. Such multi-unit activity approaches are computationally efficient to estimate online and have been show to provide an accurate representation of neural activity**?** as well as robust performance on decoding tasks**?** .

As a measure of the neural response induced by each stimulation pattern, we used the average $\mathrm{MUA_e}$

from $100\,\mathrm{ms}$ to $200\,\mathrm{ms}$ after the end of stimulation, normalized by subtracting the average $\mathrm{MUA_e}$ in a baseline window from $-110\,\mathrm{ms}$ to $-10\,\mathrm{ms}$ before stimulation ($\Delta\mathrm{MUA_e}$). This time window was chosen to be the closest to stimulation offset without any stimulation artifacts, even on the stimulating electrode.

Forward Model We developed a deep neural network ('Forward NN') to predict the $\Delta\mathrm{MUA_e}$ elicited by each stimulation pattern. The model takes as input the stimulation pattern $\mathbf{s}$ (currents on each of 96 electrodes) and the average pre-stimulation activity for each electrode on a given day $\bar{\mathbf{y}}_{pre}$ (computed from $-110\,\mathrm{ms}$ to $-10\,\mathrm{ms}$), as a proxy for ongoing cortical state. The network predicts $\hat{\mathbf{y}} = F_{\theta_1}(\mathbf{s}, \bar{\mathbf{y}}_{pre})$ and was trained to minimize mean squared error:

$$\hat{\mathbf{y}} = F_{\theta_1}(\mathbf{s}, \bar{\mathbf{y}}_{pre}) \tag{1}$$

$$\mathcal{L}_{forward} = \frac{1}{n}||\mathbf{y} - \hat{\mathbf{y}}||_2^2. \tag{2}$$

The architecture consisted of ten fully connected layers with 20% dropout, batch normalization, and residual connections. The stimulation pattern was first passed through four fully connected layers (residual connections every two layers), concatenated with $\bar{\mathbf{y}}_{pre}$, and then passed through six additional fully connected layers (residual connections every three layers). ReLU activations were used throughout, with a final 96-unit output layer. In total, the model had approximately 1.2 million trainable parameters.

This model (and all other forward and inverse networks) was trained using Tensorflow version 2.12[?] using the Adam optimizer (learning rate 0.0001, batch size 256), with a scheduler that reduced the learning rate by a factor of 0.5 every 200 epochs and early stopping based on validation performance (911 epochs).

The train and test data for the forward model consisted of random and structured stimulation patterns from the first 24 days of data collection. Importantly, forward-model performance was evaluated on 5 nonconsecutive, fully held-out recording days: all trials from a given test day were excluded from training, so the model never saw stimulation–response pairs from the test day. The only day-specific quantity provided at test time was $\bar{\mathbf{y}}_{pre}$, computed from pre-stimulation windows on that day, in addition to the day specific stimulation parameters $\mathbf{s}$.

The Forward NN was compared against several baselines:

- **Forward without pre-activity:** To test the effect of ablating the average pre-stimulation activity, we trained an otherwise identical neural network to the forward model, but with only the stimulus as input.
- **1-to-1**: Each stimulating electrode was assumed to focally activate only the nearby neurons. Thus, the predicted response on each channel was directly proportional to the stimulating current on that electrode. The maximum observed neural activity was assumed to be proportional to the maximally delivered current ($50\,\mu\mathrm{A}$).
- **Dictionary:** Loosely inspired by[?] , this method predicts multi-electrode stimulation responses via a lookup 'dictionary' of responses to single-electrode stimuli. Given an input multi-electrode stimulation $\mathbf{s} \in \mathbb{R}^{N_{elecs}}$ and selected channel $i \in [0, N_{elecs}]$, then $\mathcal{D}\{\mathbf{s}, i\}$ is a dictionary function which outputs the average response in the training dataset across single-electrode stimuli on electrode i with the nearest amplitude to s_i (*e.g.* if s_2 was $40\,\mu\mathrm{A}$, and the training set had 5 responses to single-electrode stimuli on electrode 2 at both 45 and $25\,\mu\mathrm{A}$, then $\mathcal{D}\{\mathbf{s}, 2\}$ would return the average of the five $45\,\mu\mathrm{A}$ responses).

  We evaluated two dictionary variants. First, the neural responses for each individual single-electrode stimulus were averaged to get the predicted multi-electrode response ('Dictionary'):

$$\hat{\mathbf{y}}_{avg} = \frac{1}{|\mathbf{s} > 0|} \sum_{s_i \in \mathbf{s} > 0} \mathcal{D}\{\mathbf{s}, i\}; \tag{3}$$

  and second, the ideal linear combination of the single electrode responses was computed ('Optimal Dictionary'):

$$\hat{\mathbf{y}}_{opt} = \sum_{s_i \in \mathbf{s} > 0} w_i \cdot \mathcal{D}\{\mathbf{s}, i\}, \tag{4}$$

where w_i are learned weights optimized for each given stimulus and its true neural response $\mathbf{y}$. This idealized dictionary would be infeasible to deploy in a real device, but gives a baseline of the maximum performance assuming linear combination of single-electrode responses.

- **Linear:** A linear regression model was trained on the same dataset as the deep neural network. This model was given both the stimulus vector and the pre-activity to enable a fair comparison to the neural network forward model. The model can be formulated as:

$$\hat{\mathbf{y}}_{linear} = \mathbf{W}\mathbf{s} + \mathbf{b}, \tag{5}$$

where $\mathbf{s}$ is the stimulus concatenated with the average pre-stimulation activity for that day, $\mathbf{W}$ is a weight matrix and $\mathbf{b}$ is a bias term.

- **Linear-Nonlinear:** Following approaches previously used to predict electrically evoked retinal responses[? ?] , we tested whether a single static saturating nonlinearity could account for the evoked responses. As with the Linear model, we trained on the same dataset as the deep neural network. A linear projection of the input was passed through a saturating nonlinearity ($\tanh$) and rescaled on each channel:

$$\hat{\mathbf{y}}_{LN} = \mathbf{a} \odot \tanh(\mathbf{W}\mathbf{s}) + \mathbf{b}, \tag{6}$$

where $\mathbf{s}$ is the stimulus, $\mathbf{W}$ is a weight matrix, $\mathbf{a}$ and $\mathbf{b}$ are per-channel scale and bias terms, and $\odot$ denotes element-wise multiplication. Because our recordings reflect change in multi-unit activity rather than the single-unit firing probability of previous methods[?] , we used a tanh nonlinearity; other variants, including an explicit double-sided sigmoid or tanh, did not improve performance. Further, including both the neural activity and the prestimulus neural activity as predictors did not improve performance.

- **Random:** This approach randomly paired stimuli with neural activities from the training set, and averaged across 500 random pairings, serving as a lower performance bound.

Inverse Models We developed several inversion methods using the forward model to target particular neural activity patterns. We compare our inversion methods with several baseline and state-of-the-art inversion methods:

- **Gradient:** We developed a gradient-based optimization method, which freezes the training parameters of the forward model and uses back propagation to maximize the input stimulation pattern for a desired target neural response. This method directly optimizes a stimulation pattern $\mathbf{s}$ to minimize the difference between the forward model's predicted response $F(\mathbf{s})$ and the target response $\mathbf{y}$, with L1 regularization on the stimulus to minimize stimulation amplitudes and the number of active electrodes:

$$\mathcal{L}_{\text{grad}}(\mathbf{s}) = ||F(\mathbf{s}, \bar{\mathbf{y}}_{pre}) - \mathbf{y}||_2^2 + \lambda \cdot ||\mathbf{s}||_1 \tag{7}$$

$$\mathbf{s}_{grad} = \underset{\mathbf{s}}{\operatorname{argmin}}\ \mathcal{L}_{grad}(\mathbf{s}) \tag{8}$$

Where again $\bar{\mathbf{y}}_{pre}$ is the average pre-stimulation activity for the given day. We employed the Adam optimizer with learning rate of 0.001 and ran optimization for 10000 iterations or until convergence.

- **Inverse NN:** We also built an inverse neural network model to directly map from stimuli to target neural responses. In this method, the weights of the forward model were frozen, and a secondary neural network was trained together with the frozen forward model in an autoencoder fashion, to approximate the forward model's inverse function F^{-1}, with weights θ_2. The architecture was the same as our forward model with three fully-connected layers of 96 units each, residual connections every other layer, and ReLU activations throughout the network, with dropout layers (rate=0.3) applied after each hidden layer to prevent overfitting. The inverse network also took as input the average pre-stimulation activity $\bar{\mathbf{y}}_{pre}$ to account for daily variations in neural activity. The networks output was

scaled by a factor α to ensure generated stimulation patterns fell within the valid range ($0 - 50\,\mu\text{A}$). The weights of the network θ_2 were trained to minimize:

$$\mathbf{s}_{inv} = F_{\theta_2}^{-1}(\mathbf{y}, \bar{\mathbf{y}}_{pre})$$
$$\mathcal{L}_{\text{inv}} = ||F(\alpha \cdot \mathbf{s}_{inv}, \bar{\mathbf{y}}_{pre}) - \mathbf{y}||_2^2 + \lambda \cdot \|\mathbf{s}_{inv}||_1$$

where F represents the frozen forward model, F^{-1} is our inverse network, and $\mathbf{y}$ is the target neural response pattern. With this loss, the inverse network is trained to generate stimulation patterns that, when passed through the forward model, reproduce the target response.

- **1 to 1 mapping:** The first baseline method was a conventional approach that assumes a direct relationship between stimulation and neural response. This is based on many previous works in neurostimulation as well as traditional clinical stimulation strategies with visual prostheses[?] , which explicitly or implicitly assume that activating an electrode produces focal excitation. In this method, stimulating amplitude was directly proportional to the desired activity on the underlying channel. Thus, there was a 1-1 mapping between target activations on each channel and stimulating currents on the same electrode. A target activity of 0 would have a current of 0, and a target activity of 4 (the maximum among our targets) would have a stimulating amplitude of $50\mu A$. We do note that this method will simply choose an amplitude of 0 when the target response on the corresponding electrode is negative. In practice, however, it is unclear how to elicit inhibition reliably, so we argue this methodological limitation is true to what the 1-1 method is meant to be: a simple baseline reflective of real-world practice.

  Dictionary: We evaluated an inverse dictionary approach, which leverages the dictionary ($\mathcal{D}$) of recorded neural responses ($\tilde{\mathbf{y}}$) and stimulation pattern ($\tilde{\mathbf{s}}$) pairs to optimize stimulation patterns for a new target $\mathbf{y}$. For a given target activation pattern, the dictionary stimulus was defined as the weighted average of every stimulus in the dictionary, weighted by the mean recorded response on channels active in the target response, similar to a traditional spike-triggered average.

$$\tilde{w} = \frac{1}{|y > 0]} \sum_{y_i \in \mathbf{y} > 0} \tilde{y}_i \quad \forall \tilde{\mathbf{y}}, \tilde{\mathbf{s}} \in \mathcal{D} \tag{9}$$

$$\mathbf{s}_{dictionary} = \frac{1}{\sum_{\tilde{\mathbf{y}}, \tilde{\mathbf{s}}, \tilde{w} \in \mathcal{D}} \tilde{w}} \sum_{\tilde{\mathbf{y}}, \tilde{\mathbf{s}}, \tilde{w} \in \mathcal{D}} \tilde{w} \cdot \tilde{\mathbf{s}} \tag{10}$$

- **Linear:** We also evaluated a linear regression model trained to map from neural responses back to stimulation patterns:

$$\mathbf{y} = \mathbf{W}\mathbf{s} + \mathbf{b}$$
$$\mathbf{s}_{linear} = \mathbf{W}^{\dagger}(\mathbf{y} - \mathbf{b}) \tag{11}$$

  where W and b are learned parameters optimized to minimize MSE on the training data and $W^{\dagger}$ is the pseudo inverse.

- **Replay Stimulus:** We evaluated the activity generated by replaying the stimulation that evoked the target activity for Natural targets. This approach serves as an upper performance bound (the best achievable performance is limited by the variability of the response to the same stimulus).

- **Random:** The performance of a random stimulation synthesis strategy for each target response was calculated using the average distance (measured using each metric) from each target to each response to random stimuli in the forward model training dataset.

For safety constraints, for all methods we only select the ten electrodes of $\mathbf{s}$ with the largest current amplitude. Stimuli for the various inversion methods were delivered for 15 target responses (5 natural, 10 synthetic) on day 25, and for 25 target responses (15 natural, 10 synthetic) on day 26 (out of 26 total days). Trial order was randomized with respect to both the target response and the inversion method, and each trial was repeated 4-5 times.

Latent Factor Analysis To quantify the manifold underlying neural activity we used latent factor analysis[?] , which has been widely used previously (*e.g.*[?]). Under this model, neural responses $\mathbf{y} \in \mathcal{R}^{N_{elecs}}$ are assumed to be distributed according to a number of latent factors $\mathbf{z} \in \mathcal{R}^{D}; \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$:

$$\mathbf{y}|\mathbf{z} \sim \mathcal{N}(\mathbf{\Lambda z} + \mu, \mathbf{\Psi}) \tag{12}$$

where $\mu \in \mathcal{R}^{N_{elecs}}$ is the mean response for each channel, $\mathbf{\Psi} \in \mathcal{R}^{N_{elecs} \times N_{elecs}}$ is a diagonal covariance matrix, and $\mathbf{\Lambda} \in \mathcal{R}^{N_{elecs} \times D}$ contains loadings mapping latent factors back to the neural activity.

$\mathbf{\Lambda}$ and $\mathbf{\Psi}$ were estimated using expectation-maximization with the python package scikit-learn[?] , fit to the same dataset of neural responses used to fit the forward model. We chose $D = 10$, which captured 95% of the manifold variance in the observed responses (*i.e.*, after accounting for noise). Similar to[?] , the neural manifold was defined as the column space of $\mathbf{\Lambda}$. The reported manifold distance was calculated as the Euclidean distance from each neural response to its orthogonal projection onto the columns of $\mathbf{\Lambda}$ (Equation **??**). Note that even for the observed neural responses themselves, there is still a nonzero distance to the underlying manifold due to noise and variance not captured in the 10 components (Figure **??**E, green dotted line). Natural targets had similar manifold distances as the neural manifold responses, and synthetic targets had significantly larger manifold distances.

$$d_m(\mathbf{y}) = ||\mathbf{y} - \mathbf{\Lambda}(\mathbf{\Lambda}\mathbf{\Lambda}^T)^{-1}\mathbf{\Lambda}^T\mathbf{y}||_2 \tag{13}$$

Perception Decoding Model To validate the relevance of neural responses to perceptual measurements, we trained a decoder to predict verbal perceptual responses from recorded activity (descriptions of phosphene detection, color, and brightness rating). See Behavioral Tasks for more details on the task description.

The perception model used five fully connected layers of 256 units each with residual connections to every other layer; batch norm layers, dropout layers, and ReLU activations after each hidden layer. Training was performed using the Adam optimizer and cross-entropy loss. For each model type, dropout, batch size, change in envelope multi-unit spiking activity ($\Delta\text{MUA}_\text{e}$) window, dropout rates, and learning rate were tuned independently for each model configuration, and number of parameters was balanced across model types, to ensure fair comparison. Models were tested using leave-one-day-out cross validation to evaluate generalization across sessions.

The model receives as input neural data, either pre- and post-stimulation $\Delta\text{MUA}_\text{e}$, or the stimulation amplitudes per electrode. The model outputs were categorical variables corresponding to the perceptual attributes: binary classification for phosphene detection (seen/not seen), multi-class classification for color perception, and ordinal classification for brightness ratings. The model was trained using PyTorch version 2.9.

Quantification and Statistical Analysis

Evaluation Metrics We evaluated our models using three complementary metrics. All models were trained using mean squared error (MSE) and additionally evaluated using R^2 and Earth Mover's Distance (EMD). Note that all metrics were computed on single-trial neural responses without averaging.

For a batch of n samples, the MSE between neural responses $\mathbf{y}_i$ and $\hat{\mathbf{y}}_\mathbf{i} \in \mathbb{R}^{N_{elecs}}$ was defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} |\mathbf{y}_i - \hat{\mathbf{y}}_i|_2^2 \tag{14}$$

For forward models predicting neural responses, $\mathbf{y_i}$ was an observed neural response, and $\hat{\mathbf{y}}_\mathbf{i}$ was a predicted response. For inverse models synthesizing new stimuli, $\mathbf{y_i}$ was the target neural response, and $\hat{\mathbf{y}}_\mathbf{i}$ was an observed response to the stimulus synthesized for that target response.

The coefficient of determination (R^2) measures explained variance for each channel. R^2 was calculated for each channel individually and averaged across the array:

$$R^2 = \frac{1}{N_{elecs}}\sum_{j=1}^{N_{elecs}} 1 - \frac{\sum_i (y_{ij} - \hat{y}_{ij})^2}{\sum_i (y_{ij} - \bar{y_j})^2} \tag{15}$$

Here, y_{ij} and $\hat{y}_{ij}$ represent the true and predicted neural response of the jth channel for the ith sample, and $\bar{y_j}$ is the mean of all responses on the jth channel.

We also utilized Earth Mover's Distance (EMD) to evaluate the proximity of two responses. EMD calculates the cost of "transporting" one sample into another, *i.e.*, treating both neural responses as "piles of dirt", and calculating the cost of moving the dirt from one response to the next. This metric accounts for the spatial characteristics of the electrode array–it considers responses that are "close" (e.g., off by one electrode) to have a small error–but is computationally expensive for use during training. EMD between two neural responses $\mathbf{y}, \hat{\mathbf{y}} \in \mathbb{R}^{N_{elecs}}$ is defined as:

$$\text{EMD}(\mathbf{y}, \hat{\mathbf{y}}) = \min_{\mathbf{\Gamma} \in G(\mathbf{y}, \hat{\mathbf{y}})} \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_{ij} d_{ij}, \tag{16}$$

where

$$G(\mathbf{y}, \hat{\mathbf{y}}) = \left\{ \mathbf{\Gamma} \in \mathbb{R}^{N_{elecs} \times N_{elecs}} \;\middle|\; \sum_{j=1}^{N_{elecs}} \Gamma_{ij} = y_i \text{ for all } i \text{ , } \sum_{i=1}^{N_{elecs}} \Gamma_{ij} = \hat{y}_j \text{ for all } j \right\}.$$

Here, d_{ij} is the physical distance between electrodes i and j, and $G(\mathbf{y}, \hat{\mathbf{y}})$ is the set of valid transport plans between two given recordings[? ?] . This metric captures both the spatial relationship between electrodes and the relative magnitudes of neural responses.

Channel Inclusion Criteria For the evaluation of forward and inverse models, we used only the channels in which any spiking activity was observed (Supplementary Figure **??**A shows the electrodes with spiking activity resulting from spike sorting of 9 recordings lasting 5 minutes each in different days) and were 'responsive' to stimulation, meaning a variance in the neural response larger than 0.2 uV (Supplementary Figure **??**B shows the variance in the neural response for each electrode). The total number of channels meeting these criteria is 54 out of 96 (Supplementary Figure **??**C). These inclusion criteria ensure the models were evaluated only on channels that recorded spiking activity in response to stimulation. Note that all electrodes were still used for stimulation, target activities, and predicting perception from neural responses, the selected channels were only used to ensure forward model and inverse model performance was evaluated only on channels with relevant spiking activity.

Statistical Testing Four different statistical tests were used to assess the significance of our results.

1. Wilcoxon signed-rank tests[?] were used in Results to show that repeated stimulations had less variance within the same day than across days (Figure **??**B) and that $\Delta\text{MUA}_\text{e}$ had less variance across repeated stimulations than post-stimulation MUA_e (Figure **??**C).
2. A one-way ANOVA, with post-hoc independent Student's t tests for individual differences, was used to test for differences between the forward neural network model and baselines (Results, Figure **??**F). When controlling for multiple comparisons, we found that individual pairwise tests were not independent, prohibiting family-wise error rate based methods (e.g., Bonferroni), and thus instead corrected for multiple comparisons by controlling the false discovery rate[?] .
3. A one-way repeated measures ANOVA was used to test for significant differences between stimulation synthesis methods (Results, Figure **??**B, Figure **??**B), with post-hoc paired Student's t tests, correcting for multiple comparisons by controlling the false discovery rate[?] .
4. A one-way ANOVA was used to test for differences across predictor models (Figure **??**). For comparing each perception model's performance with the stimulus-only baseline, paired Student's t-tests were used; p-values were corrected for multiple comparisons by controlling the false discovery rate[?] .

Additional resources

The clinical trial associated with this study is registered at ClinicalTrials.gov under identifier NCT02983370.

References

. Lewis, P.M., Ackland, H.M., Lowery, A.J., and Rosenfeld, J.V. (2015). Restoration of vision in blind individuals using bionic devices: A review with a focus on cortical visual prostheses. Brain Research *1595*, 51–73. URL: https://www.sciencedirect.com/science/article/pii/S0006899314015674. doi: 10.1016/j.brainres.2014.11.020.

. Liu, X., Chen, P., Ding, X., Liu, A., Li, P., Sun, C., and Guan, H. (2022). A narrative review of cortical visual prosthesis systems: the latest progress and significance of nanotechnology for the future. Annals of Translational Medicine *10*, 716. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9279795/. doi: 10.21037/atm-22-2858.

. Pio-Lopez, L., Poulkouras, R., and Depannemaecker, D. (2021). Visual cortical prosthesis: an electrical perspective. Journal of Medical Engineering & Technology *45*, 394–407. URL: https://doi.org/10.1080/03091902.2021.1907468. doi: 10.1080/03091902.2021.1907468. Publisher: Taylor & Francis.

. Zeng, F.G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (2008). Cochlear Implants: System Design, Integration, and Evaluation. IEEE Reviews in Biomedical Engineering *1*, 115–142. URL: https://ieeexplore.ieee.org/abstract/document/4664429. doi: 10.1109/RBME.2008.2008250.

. Flesher, S.N., Collinger, J.L., Foldes, S.T., Weiss, J.M., Downey, J.E., Tyler-Kabara, E.C., Bensmaia, S.J., Schwartz, A.B., Boninger, M.L., and Gaunt, R.A. (2016). Intracortical microstimulation of human somatosensory cortex. Science Translational Medicine *8*. URL: https://www.science.org/doi/10.1126/scitranslmed.aaf8083. doi: 10.1126/scitranslmed.aaf8083.

. Bjånes, D.A., and Moritz, C.T. (2019). A Robust Encoding Scheme for Delivering Artificial Sensory Information via Direct Brain Stimulation. IEEE Transactions on Neural Systems and Rehabilitation Engineering *27*, 1994–2004. URL: https://ieeexplore.ieee.org/abstract/document/8809212. doi: 10.1109/TNSRE.2019.2936739.

. Erickson-Davis, C., and Korzybska, H. (2021). What do blind people "see" with retinal prostheses? Observations and qualitative reports of epiretinal implant users. PLOS ONE *16*, e0229189. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229189. doi: 10.1371/journal.pone.0229189. Publisher: Public Library of Science.

. Murray, C.D. (2008). Embodiment and Prosthetics. In P. Gallagher, D. Desmond, and M. MacLachlan, eds. Psychoprosthetics pp. 119–129.. London: Springer. ISBN 978-1-84628-980-4 pp. 119–129. URL: https://doi.org/10.1007/978-1-84628-980-4_9.

. Barry, M.P., Armenta Salas, M., Patel, U., Wuyyuru, V., Niketeghad, S., Bosking, W.H., Yoshor, D., Dorn, J.D., and Pouratian, N. (2020). Video-mode percepts are smaller than sums of single-electrode phosphenes with the Orion® visual cortical prosthesis. Investigative Ophthalmology & Visual Science *61*, 927.

. Dagnelie, G., Jiang, A., Sadeghi, R., Barry, M.P., Stipp, K., Towle, V.L., and Troyk, P.R. (2023). Constructing a phosphene map for the inaugural recipient of the intracortical visual prosthesis (ICVP). Investigative Ophthalmology & Visual Science *64*, 5520.

. Luo, Y.H.L., and da Cruz, L. (2016). The Argus® II Retinal Prosthesis System. Progress in Retinal and Eye Research *50*, 89–107. URL: https://www.sciencedirect.com/science/article/pii/S1350946215000701. doi: 10.1016/j.preteyeres.2015.09.003.

. Ayton, L.N., Blamey, P.J., Guymer, R.H., Luu, C.D., Nayagam, D.A.X., Sinclair, N.C., Shivdasani, M.N., Yeoh, J., McCombe, M.F., Briggs, R.J., Opie, N.L., Villalobos, J., Dimitrov, P.N., Varsamidis, M., Petoe, M.A., McCarthy, C.D., Walker, J.G., Barnes, N., Burkitt, A.N., Williams, C.E., Shepherd, R.K., Allen, P.J., and for the Bionic Vision Australia Research Consortium (2014). First-in-Human Trial of a

Novel Suprachoroidal Retinal Prosthesis. PLoS ONE *9*, e115239. URL: https://dx.plos.org/10.1371/journal.pone.0115239. doi: 10.1371/journal.pone.0115239.

. Stingl, K., Bartz-Schmidt, K.U., Besch, D., Chee, C.K., Cottriall, C.L., Gekeler, F., Groppe, M., Jackson, T.L., MacLaren, R.E., Koitschev, A., Kusnyerik, A., Neffendorf, J., Nemeth, J., Naeem, M.A.N., Peters, T., Ramsden, J.D., Sachs, H., Simpson, A., Singh, M.S., Wilhelm, B., Wong, D., and Zrenner, E. (2015). Subretinal Visual Implant Alpha IMS – Clinical trial interim report. Vision Research *111*, 149–160. URL: https://www.sciencedirect.com/science/article/pii/S0042698915000784. doi: 10.1016/j.visres.2015.03.001.

. Palanker, D., Le Mer, Y., Mohand-Said, S., and Sahel, J.A. (2022). Simultaneous perception of prosthetic and natural vision in AMD patients. Nature Communications *13*, 513. URL: https://www.nature.com/articles/s41467-022-28125-x. doi: 10.1038/s41467-022-28125-x. Publisher: Nature Publishing Group.

. Veraart, C., Wanet-Defalque, M.C., Gérard, B., Vanlierde, A., and Delbeke, J. (2003). Pattern Recognition with the Optic Nerve Visual Prosthesis. Artificial Organs *27*, 996–1004. URL: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1525-1594.2003.07305.x. doi: 10.1046/j.1525-1594.2003.07305.x.

. Panetsos, F., Sanchez-Jimenez, A., Diaz-de Cerio, E.R., Diaz-Guemes, I., and Sanchez, F.M. (2011). Consistent Phosphenes Generated by Electrical Microstimulation of the Visual Thalamus. An Experimental Approach for Thalamic Visual Neuroprostheses. Frontiers in Neuroscience *5*. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2011.00084/full. doi: 10.3389/fnins.2011.00084. Publisher: Frontiers.

. Fernández, E., Alfaro, A., Soto-Sánchez, C., Gonzalez-Lopez, P., Lozano, A.M., Peña, S., Grima, M.D., Rodil, A., Gómez, B., Chen, X., Roelfsema, P.R., Rolston, J.D., Davis, T.S., and Normann, R.A. (2021). Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex. Journal of Clinical Investigation *131*, e151331. URL: https://www.jci.org/articles/view/151331. doi: 10.1172/JCI151331.

. Barry, M.P., Sadeghi, R., Towle, V.L., Stipp, K., Puhov, H., Diaz, W., Grant, P., Collison, F.T., Lane, F.J., Szlyk, J.P., Dagnelie, G., and Troyk, P.R. (2023). Preliminary visual function for the first human with the Intracortical Visual Prosthesis (ICVP). Investigative Ophthalmology & Visual Science *64*, 2842.

. Musk, E., and Neuralink (2019). An Integrated Brain-Machine Interface Platform With Thousands of Channels. Journal of Medical Internet Research *21*, e16194. URL: https://www.jmir.org/2019/10/e16194. doi: 10.2196/16194. Publisher: JMIR Publications Inc., Toronto, Canada.

. Beauchamp, M.S., Oswalt, D., Sun, P., Foster, B.L., Magnotti, J.F., Niketeghad, S., Pouratian, N., Bosking, W.H., and Yoshor, D. (2020). Dynamic Stimulation of Visual Cortex Produces Form Vision in Sighted and Blind Humans. Cell *181*, 774–783.e5. URL: https://www.sciencedirect.com/science/article/pii/S0092867420304967. doi: 10.1016/j.cell.2020.04.033.

. Maynard, E.M., Nordhausen, C.T., and Normann, R.A. (1997). The Utah Intracortical Electrode Array: A recording structure for potential brain-computer interfaces. Electroencephalography and Clinical Neurophysiology *102*, 228–239. URL: https://www.sciencedirect.com/science/article/pii/S0013469496951760. doi: 10.1016/S0013-4694(96)95176-0.

. Youssef, D., Wittig, J.H., Jackson, S., Inati, S.K., and Zaghloul, K.A. (2023). Neuronal Spiking Responses to Direct Electrical Microstimulation in the Human Cortex. Journal of Neuroscience *43*, 4448–4460. URL: https://www.jneurosci.org/content/43/24/4448. doi: 10.1523/JNEUROSCI.1666-22.2023. Publisher: Society for Neuroscience Section: Research Articles.

. Kumaravelu, K., and Grill, W.M. (2024). Neural mechanisms of the temporal response of cortical neurons to intracortical microstimulation. Brain Stimulation *17*, 365–381. URL: https://www.sciencedirect.com/science/article/pii/S1935861X24000482. doi: 10.1016/j.brs.2024.03.012.

. Histed, M.H., Bonin, V., and Reid, R.C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. Neuron *63*, 508–522. doi: 10.1016/j.neuron.2009.07.016.

. Ghaffari, D.H., Chang, Y.C., Mirzakhalili, E., and Weiland, J.D. (2021). Closed-loop Optimization of Retinal Ganglion Cell Responses to Epiretinal Stimulation: A Computational Study. In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER). pp. 597–600. URL: https://ieeexplore.ieee.org/document/9441437/?arnumber=9441437. doi: 10.1109/NER49283.2021.9441437 iSSN: 1948-3554.

. Shah, N.P., Madugula, S., Grosberg, L., Mena, G., Tandon, P., Hottowy, P., Sher, A., Litke, A., Mitra, S., and Chichilnisky, E. (2019). Optimization of Electrical Stimulation for a High-Fidelity Artificial Retina. In 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER). pp. 714–718. URL: https://ieeexplore.ieee.org/abstract/document/8716987. doi: 10.1109/NER.2019.8716987 iSSN: 1948-3554.

. Shah, N.P., Phillips, A., Madugula, S., Lotlikar, A., Gogliettino, A.R., Hays, M.R., Grosberg, L., Brown, J., Dusi, A., Tandon, P., Hottowy, P., Dabrowski, W., Sher, A., Litke, A.M., Mitra, S., and Chichilnisky, E. (2024). Precise control of neural activity using dynamically optimized electrical stimulation. eLife *13*, e83424. URL: https://doi.org/10.7554/eLife.83424. doi: 10.7554/eLife.83424. Publisher: eLife Sciences Publications, Ltd.

. Micera, S., Romeni, S., De Luca, D., Pierantoni, L., Toni, L., Marino, G., and Moccia, S. (2023). A computational model to design wide field-of-view optic nerve neuroprostheses. . URL: https://www.researchsquare.com/article/rs-3218482/v1. doi: 10.21203/rs.3.rs-3218482/v1.

. Vasireddy, P.K., Gogliettino, A.R., Brown, J.B., Vilkhu, R.S., Madugula, S.S., Phillips, A.J., Mitral, S., Hottowy, P., Sher, A., Litke, A., Shah, N.P., and Chichilnisky, E. (2023). Efficient Modeling and Calibration of Multi-Electrode Stimuli for Epiretinal Implants. In 2023 11th International IEEE/EMBS Conference on Neural Engineering (NER). pp. 1–4. URL: https://ieeexplore.ieee.org/document/10123907. doi: 10.1109/NER52421.2023.10123907 iSSN: 1948-3554.

. Granley, J., Relic, L., and Beyeler, M. (2022). Hybrid Neural Autoencoders for Stimulus Encoding in Visual and Other Sensory Neuroprostheses. Advances in Neural Information Processing Systems *35*, 22671–22685. URL: https://proceedings.neurips.cc/paper_files/paper/2022/hash/8e9a6582caa59fda0302349702965171-Abstract-Conference.html.

. Granley, J., Fauvel, T., Chalk, M., and Beyeler, M. (2023). Human-in-the-Loop Optimization for Deep Stimulus Encoding in Visual Prostheses. Advances in Neural Information Processing Systems *36*, 79376–79398. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/fb06bc3abcece7b8725a8b83b8fa3632-Abstract-Conference.html.

. Spencer, M.J., Kameneva, T., Grayden, D.B., Meffin, H., and Burkitt, A.N. (2019). Global activity shaping strategies for a retinal implant. Journal of Neural Engineering *16*, 026008. URL: https://dx.doi.org/10.1088/1741-2552/aaf071. doi: 10.1088/1741-2552/aaf071. Publisher: IOP Publishing.

. Castro, D., Grayden, D.B., Meffin, H., and Spencer, M. (2023). Neural Activity Shaping in Visual Prostheses with Deep Learning. bioRxiv. URL: https://www.biorxiv.org/content/10.1101/2023.12.20.572123v1. doi: 10.1101/2023.12.20.572123.

. de Ruyter van Steveninck, J., Güçlü, U., van Wezel, R., and van Gerven, M. (2022). End-to-end optimization of prosthetic vision. Journal of Vision *22*, 20. URL: https://doi.org/10.1167/jov.22.2.20. doi: 10.1167/jov.22.2.20.

. Relic, L., Zhang, B., Tuan, Y.L., and Beyeler, M. (2022). Deep Learning–Based Perceptual Stimulus Encoder for Bionic Vision. In Proceedings of the Augmented Humans International Conference 2022. AHs '22. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9632-5 pp. 323–325. URL: https://dl.acm.org/doi/10.1145/3519391.3524034. doi: 10.1145/3519391.3524034.

. Van Der Grinten, M., De Ruyter Van Steveninck, J., Lozano, A., Pijnacker, L., Rueckauer, B., Roelfsema, P., Van Gerven, M., Van Wezel, R., Güçlü, U., and Güçlütürk, Y. (2024). Towards biologically plausible phosphene simulation for the differentiable optimization of visual cortical prostheses. eLife *13*, e85812. URL: https://elifesciences.org/articles/85812. doi: 10.7554/eLife.85812.

. Küçükoğlu, B., Rueckauer, B., Steveninck, J.d.R.v., Grinten, M.v.d., Güçlütürk, Y., Roelfsema, P.R., Güçlü, U., and Gerven, M.v. (2025). End-to-end Learning of Safe Stimulation Parameters for Cortical Neuroprosthetic Vision. bioRxiv. URL: https://www.biorxiv.org/content/10.1101/2025.01.23.634543v1. doi: 10.1101/2025.01.23.634543 pages: 2025.01.23.634543 Section: New Results.

. Tafazoli, S., MacDowell, C.J., Che, Z., Letai, K.C., Steinhardt, C.R., and Buschman, T.J. (2020). Learning to control the brain through adaptive closed-loop patterned stimulation. Journal of Neural Engineering *17*, 056007. URL: https://dx.doi.org/10.1088/1741-2552/abb860. doi: 10.1088/1741-2552/abb860. Publisher: IOP Publishing.

. Romeni, S., Zoccolan, D., and Micera, S. (2021). A machine learning framework to optimize optic nerve electrical stimulation for vision restoration. Patterns (New York, N.Y.) *2*, 100286. doi: 10.1016/j.patter.2021.100286.

. Chen, X., Wang, F., Fernandez, E., and Roelfsema, P.R. (2020). Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. Science *370*, 1191–1196. URL: https://www.science.org/doi/10.1126/science.abd7435. doi: 10.1126/science.abd7435.

. Papale, P., Luca, D.D., and Roelfsema, P.R. (2024). Deep generative networks reveal the tuning of neurons in IT and predict their influence on visual perception. bioRxiv. URL: https://www.biorxiv.org/content/10.1101/2024.10.09.617382v1. doi: 10.1101/2024.10.09.617382 pages: 2024.10.09.617382 Section: New Results.

. Grani, F. Improving cortical visual prostheses using closed-loop stimulation approaches. Ph.D. thesis (2024). URL: http://dspace.umh.es/handle/11000/35610.

. Grani, F., Val-Calvo, M., Peco, R.L., Doblado, A.R., Waclawczyk, D., Ruiz, R.M., Soo, L., Sanchez, C.S., Ortiz, M.V., and Jover, E.F. (2023). Advantages of bidirectional cortical visual prostheses. In IBRO Neuroscience Reports vol. 15. pp. S948–S949. URL: https://linkinghub.elsevier.com/retrieve/pii/S2667242123020651. doi: 10.1016/j.ibneur.2023.08.2003.

. Grani, F., Soto-Sánchez, C., Doblado, A.R., Peco, R.L., Gonzalez-Lopez, P., and Fernandez, E. (2025). Neural correlates of phosphene perception in blind individuals: A step toward a bidirectional cortical visual prosthesis. Science Advances *11*, eadv8846. URL: https://www.science.org/doi/abs/10.1126/sciadv.adv8846. doi: 10.1126/sciadv.adv8846. arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.adv8846.

. Allison-Walker, T., Hagan, M.A., Meikle, S.J., Price, N.S.C., and Wong, Y.T. (2025). Local field potential phase modulates the evoked response to electrical stimulation in visual cortex. Journal of Neural Engineering *22*, 016009. URL: https://dx.doi.org/10.1088/1741-2552/ada828. doi: 10.1088/1741-2552/ada828. Publisher: IOP Publishing.

. Allison-Walker, T.J., Ann Hagan, M., Chiang Price, N.S., and Tat Wong, Y. (2020). Local field potential phase modulates neural responses to intracortical electrical stimulation. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3521–3524. URL: https://ieeexplore.ieee.org/abstract/document/9176186. doi: 10.1109/EMBC44109.2020.9176186 iSSN: 2694-0604.

. Dugué, L., Marque, P., and VanRullen, R. (2011). The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception. The Journal of Neuroscience *31*, 11889–11893. URL: https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1161-11.2011. doi: 10.1523/JNEUROSCI.1161-11.2011.

. Grani, F., Soto-Sánchez, C., Fimia, A., and Fernández, E. (2022). Toward a personalized closed-loop stimulation of the visual cortex: Advances and challenges. Frontiers in Cellular Neuroscience *16*, 1034270. URL: https://www.frontiersin.org/articles/10.3389/fncel.2022.1034270/full. doi: 10.3389/fncel.2022.1034270.

. Leo, J., Ge, E., and Li, S. (2023). Wasserstein Distance in Deep Learning. Rochester, NY: Social Science Research Network. URL: https://papers.ssrn.com/abstract=4368733. doi: 10.2139/ssrn.4368733.

. Minai, Y., Soldado-Magraner, J., Smith, M.A., and Yu, B.M. (2024). MiSO: Optimizing brain stimulation to create neural activity states. Advances in Neural Information Processing Systems *37*, 24126–24149. URL: https://proceedings.neurips.cc/paper_files/paper/2024/hash/2af641762dc02035c31a9314b2d090b6-Abstract-Conference.html.

. Bashivan, P., Kar, K., and DiCarlo, J.J. (2019). Neural population control via deep image synthesis. Science *364*, eaav9436. URL: https://www.science.org/doi/10.1126/science.aav9436. doi: 10.1126/science.aav9436.

. Walker, E.Y., Sinz, F.H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P.G., Ecker, A.S., Reimer, J., Pitkow, X., and Tolias, A.S. (2019). Inception loops discover what excites neurons most using deep predictive models. Nature Neuroscience *22*, 2060–2065. URL: https://www.nature.com/articles/s41593-019-0517-x. doi: 10.1038/s41593-019-0517-x. Publisher: Nature Publishing Group.

. Morales-Gregorio, A., Kurth, A.C., Ito, J., Kleinjohann, A., Barthélemy, F.V., Brochier, T., Grün, S., and Albada, S.J.v. (2024). Neural manifolds in V1 change with top-down signals from V4 targeting the foveal region. Cell Reports *43*. URL: https://www.cell.com/cell-reports/abstract/S2211-1247(24)00699-5. doi: 10.1016/j.celrep.2024.114371. Publisher: Elsevier.

. Langdon, C., Genkin, M., and Engel, T.A. (2023). A unifying perspective on neural manifolds and circuits for cognition. Nature Reviews Neuroscience *24*, 363–377. URL: https://www.nature.com/articles/s41583-023-00693-x. doi: 10.1038/s41583-023-00693-x. Publisher: Nature Publishing Group.

. Bartholomew, D.J., Knott, M., and Moustaki, I. (2011). Latent Variable Models and Factor Analysis: A Unified Approach. John Wiley & Sons. ISBN 978-1-119-97370-6. Google-Books-ID: 5YxvE1S56DwC.

. Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. Nature *512*, 423–426. URL: https://www.nature.com/articles/nature13665. doi: 10.1038/nature13665. Publisher: Nature Publishing Group.

. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional geometry of population responses in visual cortex. Nature *571*, 361–365. URL: https://www.nature.com/articles/s41586-019-1346-5. doi: 10.1038/s41586-019-1346-5. Publisher: Nature Publishing Group.

. Rousche, P.J., and Normann, R.A. (1998). Chronic recording capability of the Utah Intracortical Electrode Array in cat sensory cortex. Journal of Neuroscience Methods *82*, 1–15. doi: 10.1016/s0165-0270(98)00031-4.

. Patel, P.R., Welle, E.J., Letner, J.G., Shen, H., Bullard, A.J., Caldwell, C.M., Vega-Medina, A., Richie, J.M., Thayer, H.E., Patil, P.G., Cai, D., and Chestek, C.A. (2023). Utah array characterization and histological analysis of a multi-year implant in non-human primate motor and sensory cortices. Journal of Neural Engineering *20*. doi: 10.1088/1741-2552/acab86.

. Bullard, A.J., Hutchison, B.C., Lee, J., Chestek, C.A., and Patil, P.G. (2020). Estimating Risk for Future Intracranial, Fully Implanted, Modular Neuroprosthetic Systems: A Systematic Review of Hardware Complications in Clinical Deep Brain Stimulation and Experimental Human Intracortical Arrays. Neuromodulation: Technology at the Neural Interface *23*, 411–426. URL: https://linkinghub.elsevier.com/retrieve/pii/S1094715921021516. doi: 10.1111/ner.13069.

. Hughes, C.L., Flesher, S.N., Weiss, J.M., Downey, J.E., Boninger, M., Collinger, J.L., and Gaunt, R.A. (2021). Neural stimulation and recording performance in human sensorimotor cortex over 1500 days. Journal of Neural Engineering *18*. doi: 10.1088/1741-2552/ac18ad.

. Caspi, A., Barry, M.P., Patel, U.K., Salas, M.A., Dorn, J.D., Roy, A., Niketeghad, S., Greenberg, R.J., and Pouratian, N. (2021). Eye movements and the perceived location of phosphenes generated by intracranial primary visual cortex stimulation in the blind. Brain Stimulation *14*, 851–860.

. Silva, A.E., Tsang, K., Hasan, S.J., and Thompson, B. (2021). Precise oculocentric mapping of transcranial magnetic stimulation-evoked phosphenes. NeuroReport *32*, 913–917.

. Waclawczyk, D., Soo, L., Morollon Ruiz, R., Caspi, A., and Fernandez, E. (2025). Integrating eye-tracking with cortical visual prostheses in patients without eyes: A case study. IEEE Transactions on Neural Systems and Rehabilitation Engineering *33*, 3998–4007.

. Normann, R.A., and Fernandez, E. (2016). Clinical applications of penetrating neural interfaces and Utah Electrode Array technologies. Journal of Neural Engineering *13*, 061003. URL: https://iopscience.iop.org/article/10.1088/1741-2560/13/6/061003. doi: 10.1088/1741-2560/13/6/061003.

. Rocca, A., Lehner, C., Wafula-Wekesa, E., Luna, E., Fernández-Cornejo, V., Abarca-Olivas, J., Soto-Sánchez, C., Fernández-Jover, E., and González-López, P. (2024). Robot-assisted implantation of a microelectrode array in the occipital lobe as a visual prosthesis: technical note. Journal of Neurosurgery *140*, 1169–1176. URL: https://thejns.org/view/journals/j-neurosurg/140/4/article-p1169.xml. doi: 10.3171/2023.8.JNS23772.

. Trautmann, E.M., Stavisky, S.D., Lahiri, S., Ames, K.C., Kaufman, M.T., O'Shea, D.J., Vyas, S., Sun, X., Ryu, S.I., Ganguli, S., and Shenoy, K.V. (2019). Accurate Estimation of Neural Population Dynamics without Spike Sorting. Neuron *103*, 292–308.e4. URL: https://www.cell.com/neuron/abstract/S0896-6273(19)30428-3. doi: 10.1016/j.neuron.2019.05.003. Publisher: Elsevier.

. Ahmadi, N., Constandinou, T.G., and Bouganis, C.S. (2021). Robust and accurate decoding of hand kinematics from entire spiking activity using deep learning. Journal of Neural Engineering *18*, 026011. URL: https://dx.doi.org/10.1088/1741-2552/abde8a. doi: 10.1088/1741-2552/abde8a. Publisher: IOP Publishing.

. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv. URL: http://arxiv.org/abs/1603.04467. doi: 10.48550/arXiv.1603.04467 arXiv:1603.04467 [cs].

. Maturana, M.I., Apollo, N.V., Hadjinicolaou, A.E., Garrett, D.J., Cloherty, S.L., Kameneva, T., Grayden, D.B., Ibbotson, M.R., and Meffin, H. (2016). A Simple and Accurate Model to Predict Responses

to Multi-electrode Stimulation in the Retina. PLoS Computational Biology *12*, e1004849. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004849. doi: 10.1371/journal.pcbi.1004849.

. Pedregosa, F., Pedregosa, F., Varoquaux, G., Varoquaux, G., Org, N., Gramfort, A., Gramfort, A., Michel, V., Michel, V., Fr, L., Thirion, B., Thirion, B., Grisel, O., Grisel, O., Blondel, M., Prettenhofer, P., Prettenhofer, P., Weiss, R., Dubourg, V., Dubourg, V., Vanderplas, J., Passos, A., Tp, A., and Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

. Aoun, A., Shetler, O., Raghuraman, R., Rodriguez, G.A., and Hussaini, S.A. (2024). Beyond correlation: optimal transport metrics for characterizing representational stability and remapping in neurons encoding spatial memory. Frontiers in Cellular Neuroscience *17*. URL: https://www.frontiersin.org/journals/cellular-neuroscience/articles/10.3389/fncel.2023.1273283/full. doi: 10.3389/fncel.2023.1273283. Publisher: Frontiers.

. Rosner, B., Glynn, R.J., and Lee, M.L.T. (2006). The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics *62*, 185–192. doi: 10.1111/j.1541-0420.2005.00389.x.

. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics *29*, 1165–1188. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-4/The-control-of-the-false-discovery-rate-in-multiple-testing/10.1214/aos/1013699998.full. doi: 10.1214/aos/1013699998. Publisher: Institute of Mathematical Statistics.