Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A stimulation-orthogonal neural biomarker of therapeutic response in deep brain stimulation

Tanmoy Sil

sil_t@ukw.de

Department of Neurology, University Hospital Wuerzburg, Wuerzburg, Germany https://orcid.org/0009-0002-5041-7579

Pavel Navratil

Department of Neurology, University Hospital Wuerzburg, Wuerzburg, Germany

Ghadir Abbas

University Hospital Wuerzburg

Florian Lange

University Hospital Wuerzburg

Jens Volkmann

University Hospital Wuerzburg, Department of Neurology, Wuerzburg, Germany https://orcid.org/0000-0002-9570-593X

Muthuraman Muthuraman

University of Augsburg

Martin Reich

University Hospital Wuerzburg

Robert Peach

Department of Neurology, University Hospital Wuerzburg, Wuerzburg, Germany https://orcid.org/0000-0002-8738-5825

Article

Keywords:

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10206188/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. M.R. reports grant support and honoraria for speaking from Medtronic and Boston Scientific. J.V. reports grants and personal fees from Medtronic, grants and personal fees from Boston Scientific, and personal fees from Abbott. F.L. reports honoraria for speaking from Boston Scientific. These interests are outside the submitted work. The remaining authors (T.S., P.N., G.A., M.M. and R.P.) declare no competing interests.

A stimulation-orthogonal neural biomarker of therapeutic response in deep brain stimulation

Tanmoy Sil[a], Pavel Navratil[a], Ghadir Abbas[a], Florian Lange[a], Jens Volkmann[a], Muthuraman Muthuraman[a,b], Martin Reich[a]*, Robert Peach[a]*

*Senior authors

[a] Department of Neurology, University Clinic Würzburg, Germany

[b] Chair of Informatics for Medical Technology, University of Augsburg, Germany

Abstract

Closed-loop neuromodulation requires a feedback signal that reports therapeutic state, yet such signals are typically read from the same circuits that stimulation is perturbing. A neural feature that changes during stimulation may therefore simply reflect engagement of the circuit rather than clinical improvement. This entanglement problem is common to any stimulation paradigm in which the recording and stimulation share a substrate. Here, we show that the neural signature of stimulation and the signature of therapeutic response can be separated within a single learned representation, without supervision of clinical outcome. We use generative causal representation learning to learn a supervised latent dimension to detect stimulation presence rather than therapeutic state in a low-dose condition, where stimulation engages the circuit, but clinical benefits are minimal. This quarantines the stimulation signature, allowing an orthogonal subspace to carry therapeutic structure. We test the approach in subthalamic deep brain stimulation for Parkinson's disease, where the confound persists even after the exogenous stimulation artifact is removed. Applied unchanged to recordings at clinically-relevant stimulation amplitudes across four frequencies in nine patients, the stimulation axis classified stimulation state under leave-one-subject-out validation (AUC = 0.919 and 0.942 at half and full amplitude) but did not predict motor outcome, whereas the auxiliary subspace tracked graded motor improvement ($\rho = 0.50$), survived mixed-effects control for frequency, amplitude and patient identity, and identified low-frequency stimulation as a large anti-therapeutic displacement not captured by beta or gamma power. The frozen representation generalized to a pre-specified held-out patient (rigidity $\rho = 0.731$) and an independent prospective video-kinematic cohort ($n = 3$). Beyond DBS, the approach offers a template for discovering biomarkers of therapeutic effect in any closed-loop setting where stimulation contaminates the recording used for feedback, from other neurological and psychiatric DBS targets to spinal cord and non-invasive brain stimulation.

Introduction

Closed-loop neuromodulation promises to replace fixed stimulation settings with therapy that adapts to a physiological readout, but in most such systems that readout is recorded from the very circuit the stimulation perturbs[1-3]. Features that change when stimulation is on may therefore reflect genuine therapeutic benefit, or merely the presence of stimulation, and a feedback signal that cannot tell these apart can be optimized in the wrong direction. Subthalamic deep brain stimulation (DBS) for Parkinson's disease is a clear instance of this problem and the setting we study here: it is

among the most effective circuit therapies for the disease[4,5], yet its parameters are still set by empirical clinical titration, and the sensing-enabled devices now used to automate that process record the guiding signal from the stimulated tissue itself. For physiology-guided programming and adaptive DBS, a useful biomarker must therefore do more than detect that stimulation has changed neural activity; it must distinguish therapeutic from non-therapeutic change.

Beta-band activity in the subthalamic nucleus illustrates both the promise and the limitations of current DBS biomarkers. Pathological beta oscillations are linked to akinesia and rigidity[6,7], are attenuated by dopaminergic medication and high-frequency stimulation[8,9], and have become the leading feedback signal for adaptive DBS[1,10,11], now embodied in the first FDA-approved adaptive DBS system, which titrates stimulation to beta power in real time[11]. Yet beta also highlights why stimulation-responsive physiology cannot be equated with therapeutic physiology. Dopaminergic medication and DBS can produce comparable motor benefit despite divergent effects on STN beta activity[12,13], and lower-frequency stimulation can improve selected motor features without simply suppressing broadband beta power[14-16]. Beta suppression is therefore a powerful marker of pathological circuit state and stimulation engagement, but not a universal coordinate of graded clinical benefit[17]. The broader challenge is to separate the neural dimensions that identify the presence of stimulation from those that measure the therapeutic effects of stimulation.

However, such challenges are not limited to beta power. DBS alters spectral power[8], burst dynamics[18], connectivity and cross-frequency coupling[19,20], while also imprinting a neural signature of the stimulation state itself[21]. A decoder trained on such data can therefore succeed simply because it learns whether stimulation is ON rather than whether it is beneficial. This distinction becomes critical across stimulation parameters: because the features are dominated by the shared signature of stimulation presence, similarity in local field potential space need not imply similarity in clinical effect. The relevant object is therefore not a stimulation-responsive feature, but a representation in which the dimensions that capture the presence of stimulation are separated from the dimensions that explain therapeutic variation during stimulation.

Generative Causal Explanations provides a natural architecture for disentangling this problem[22]. GCE embeds high-dimensional neural features in a latent space partitioned into a classifier-aligned component and auxiliary components that preserve structure orthogonal to the classifier target. We invert the usual application of this architecture: rather than training the classifier-aligned dimension to identify therapeutic state directly as in previous works[22], we train it to detect stimulation presence, so that the dimension absorbs the stimulation signature and the orthogonal auxiliary space is left free to carry neural structure that tracks motor benefit across stimulation settings, without any supervision of outcome labels. To weaken the coupling between stimulation detection and clinical response, the model is trained on half-amplitude recordings, where stimulation-related neural changes remain detectable but therapeutic effects are smaller and less consistent. The trained encoder is then applied to full-amplitude recordings, where clinical effects are larger.

We test this prediction using STN local field potentials recorded across four stimulation frequencies and two amplitudes in patients with Parkinson's disease. The model is trained on a single contrast, the half-amplitude best-frequency ON versus OFF recordings from the nine-patient development cohort. Generalization is then evaluated, without any retraining, on conditions the model never saw: the remaining stimulation frequencies, the full-amplitude recordings, and held-out subjects under leave-one-subject-out cross-validation. Finally, the frozen model is applied to a single prospectively held-out patient (P11), excluded from training, hyperparameter selection and analysis development, as an out-of-sample validation. As a further, independent test, the frozen model is applied to a separate cohort acquired in a different study under a different stimulation protocol and assessed with video kinematics, providing out-of-sample evidence across both stimulation protocol and outcome modality. We hypothesize that the stimulation-aligned dimension generalizes as an ON/OFF marker across these unseen conditions and subjects, whilst therapeutic information resides in the orthogonal auxiliary space. We then ask whether this auxiliary geometry predicts protocol-defined motor improvement, distinguishes low-frequency stimulation from more effective settings, and transfers without retraining both to a prospectively held-out patient and a separate kinematic cohort.

Results

Study cohort and stimulation protocol

We studied ten participants with advanced Parkinson's disease who had received chronic bilateral STN-DBS via Medtronic Percept PC devices for at least one year. Nine participants formed the model-development cohort and one (P11) was reserved for prospective validation; demographic and clinical characteristics, including pre-operative levodopa-challenge MDS-UPDRS III scores, are summarized in Table 1. In each participant, STN local field potentials were recorded OFF medication across an OFF-stimulation baseline and four stimulation frequencies (125, 110, 85, and 55 Hz), each delivered at the participant's clinically programmed amplitude and at half that amplitude. Motor state was rated at every condition using a protocol-defined subset of MDS-UPDRS Part III items and normalized so that positive values denote improvement relative to OFF (see Methods). At the cohort level, motor improvement under stimulation was larger and more consistent at full amplitude than at half amplitude, where it was near zero and variable (mean ± SD normalized total improvement: +0.182 ± 0.236 at full amplitude, −0.001 ± 0.219 at half amplitude, n = 36 patient-frequency observations per amplitude). This amplitude-dependent separation of clinical effect motivated the training design described below, in which the stimulation classifier is learned from half-amplitude recordings and the resulting representation is evaluated at full amplitude.

A stimulation-supervised latent partition separates intervention detection from residual neural structure

We first asked whether the disentanglement of stimulation presence and therapeutic entrainment could be built directly into the model architecture. Our aim was not to train a decoder of clinical benefit, but instead to isolate the neural signature of stimulation itself (Fig. 1a). To achieve this, we trained a GCE model on a binary stimulation-detection task using half-amplitude recordings,

contrasting OFF stimulation with each participant's best-frequency ON condition (280 segments). In this setting, the classifier-aligned latent dimension was forced to capture information about stimulation presence, whereas the remaining auxiliary dimensions preserved neural structure orthogonal to that task (Fig. 1b-d).

The resulting model embedded the 70-dimensional STN-LFP feature space into a six-dimensional latent representation comprising one stimulation-aligned dimension and five auxiliary dimensions (see Table S5 for model architecture details). Finding a single classifier-aligned dimension was not a result of constraining the GCE model, instead it was a result of the GCE optimization and therefore sufficient to capture the stimulation-detection signal while preserving stable reconstruction of the neural feature space. Given this, we treated the classifier-aligned dimension as a stimulation-detection axis rather than as a therapeutic biomarker. Conversely, the auxiliary dimensions were not trained on clinical outcome, stimulation frequency, or motor improvement, and therefore provided the candidate space in which therapeutic structure could be tested.

After training, the encoder was frozen and applied without retraining to recordings from all stimulation frequencies and amplitudes (1008 segments). This design created a direct test of the architectural dissociation. If the classifier-aligned dimension has isolated the intervention signature, it should generalize as an ON/OFF marker but remain unrelated to motor benefit. If therapeutic information was present independently of stimulation detection, it should appear in the orthogonal auxiliary space (Fig. 1e, f).

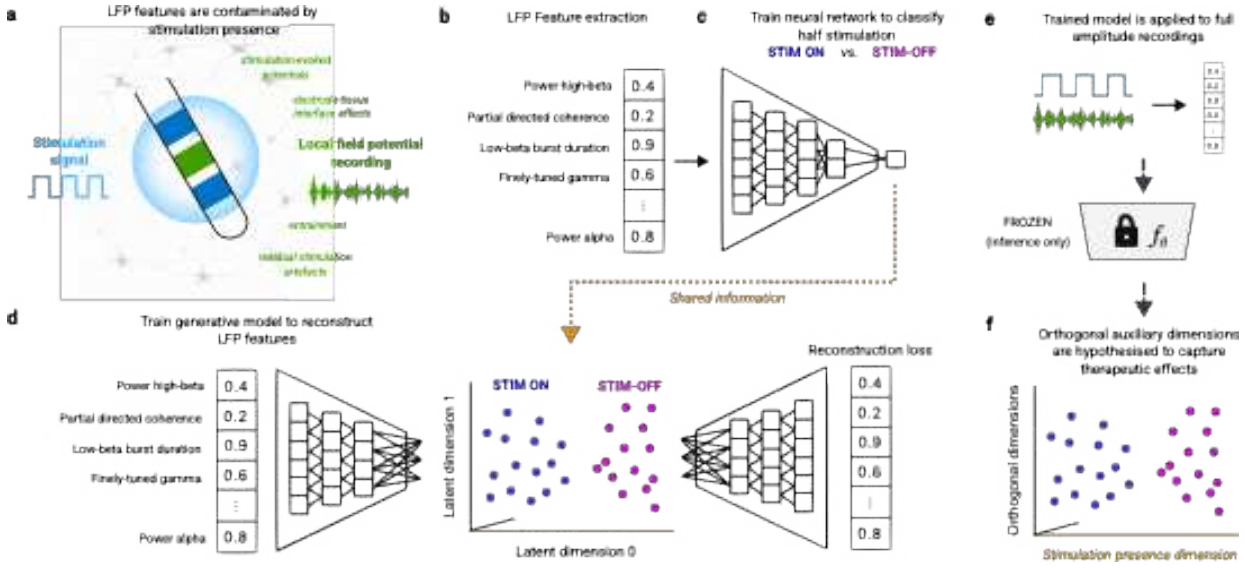


Figure 1. Architectural separation of stimulation detection and therapeutic encoding. ***a****, Active DBS creates an intrinsic biomarker confound in that neural features recorded from the stimulated structure may reflect stimulation presence, circuit engagement, or therapeutic benefit.* ***b****, Standard features are extracted from the STN-LFP and* ***c****, used to train a classifier between recordings of stimulation on and off (at half stimulation amplitude where we don't expect significant therapeutic effects).* ***d****, we train a generative causal explainer to map the 70-dimensional STN-LFP feature vectors into a 6-dimensional latent space which*

forces one dimension to align with the classifier, thereby encoding stimulation presence. ***e****, after training, the frozen encoder is applied to full-amplitude recordings across stimulation frequencies.* ***f****, the central prediction is that the stimulation-aligned dimension should classify stimulation state but not predict motor improvement, whereas therapeutic structure should emerge in the orthogonal auxiliary space.*

The stimulation-aligned dimension detects stimulation but not therapeutic response

An in-sample classifier trained on all 70 LFP features separated each patient's best-frequency ON condition from OFF with high accuracy (train AUC = 0.987, test AUC = 0.951; Fig. 2a), confirming that stimulation state is recoverable from the feature space before asking whether that signal generalizes across patients. Having trained the classifier-aligned dimension on stimulation presence, we next tested whether this dimension captured a patient-general stimulation signature rather than patient-specific variance. We performed leave-one-subject-out cross-validation on the half-amplitude binary dataset, using Dim 0 values from eight patients to train a logistic classifier and evaluating stimulation-state discrimination in the held-out patient. Dim 0 separated OFF stimulation from each patient's best-frequency ON condition with high accuracy across subjects, yielding a mean AUC of 0.919 (95% CI, 0.831–0.981; Fig. 2b). When the same fixed encoder was applied to full-amplitude recordings, Dim 0 retained similar discriminative performance (AUC = 0.942; 95% CI, 0.893–0.985; Fig. 2b), indicating that the stimulation signature learned at half amplitude generalized across both patients and amplitudes. Applying the same leave-one-subject-out procedure to the auxiliary subspace yielded near-chance discrimination at both half amplitude and full amplitude (AUC = 0.502 and 0.480, respectively; Fig. 2c). Thus, the GCE partition concentrated stimulation-detection information in Dim 0 while leaving the auxiliary dimensions largely uninformative about ON/OFF state.

We then asked whether successful stimulation detection was sufficient to explain clinical response. It was not. At full amplitude, OFF-referenced Dim 0 signed distance showed no association with protocol-defined total motor improvement (Spearman $\rho = -0.032$, $p = 0.854$). Dim 0 was similarly unrelated to rigidity ($\rho = 0.048$, $p = 0.783$), bradykinesia ($\rho = 0.034$, $p = 0.843$), and tremor ($\rho = 0.187$, $p = 0.274$) (Fig. 2d-g). The stimulation-aligned dimension therefore behaved as intended: it generalized as a marker of stimulation presence but remained clinically uninformative. This dissociation established Dim 0 as a negative-control axis for subsequent tests of therapeutic encoding in the orthogonal auxiliary space. Feature attribution was consistent with this interpretation. Dim 0 was driven primarily by interhemispheric partial directed coherence and high-beta spectral power (Fig. 2h), features that track stimulation presence rather than parkinsonian motor state.

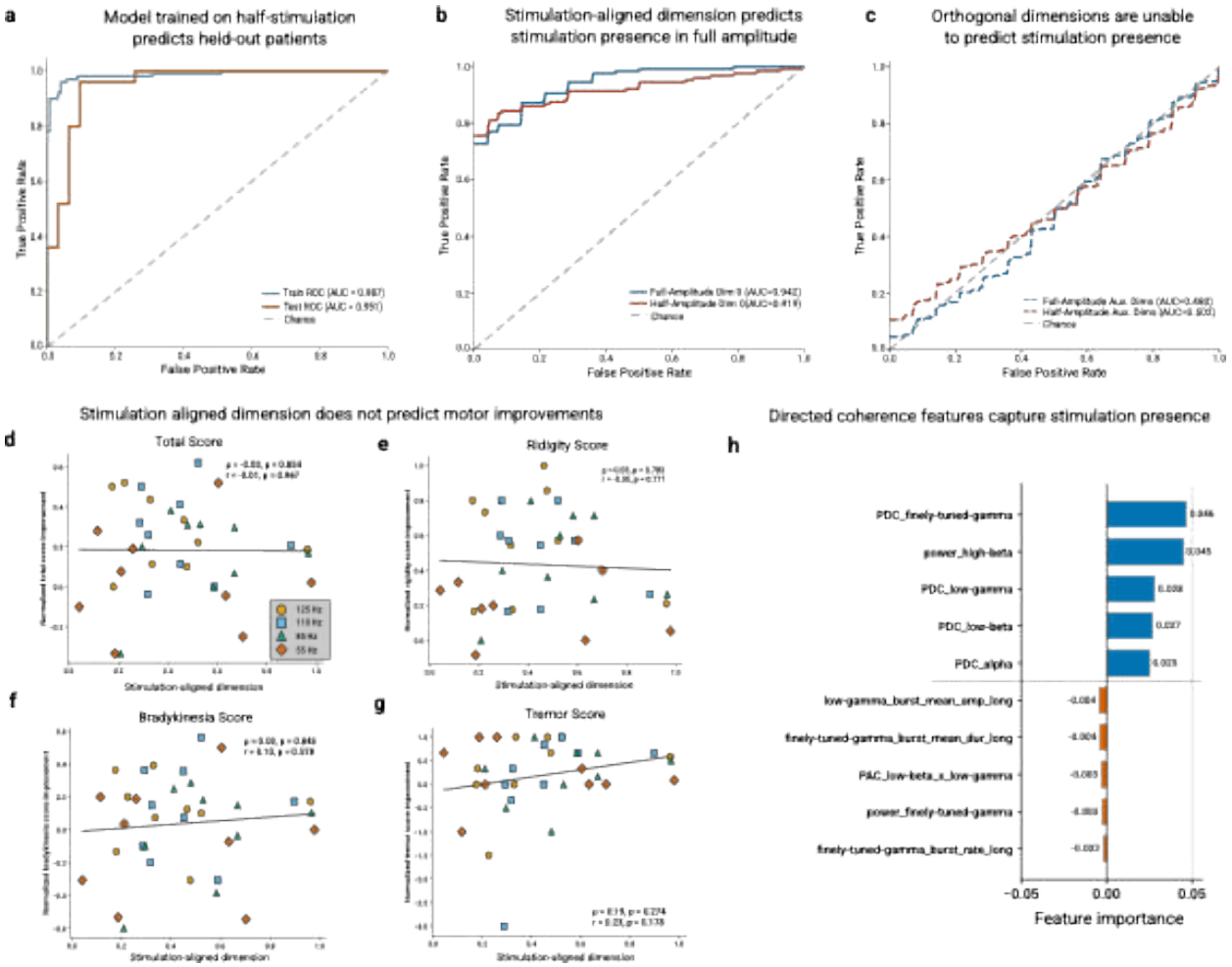


Figure 2. The stimulation-aligned dimension detects stimulation presence but carries no therapeutic information. a. *Classifier performance on the half-amplitude binary task (OFF vs each patient's clinically selected best-frequency ON; 280 segments from 9 patients) for an 80/20 segment-level train/test split, reflecting in-sample capacity (train AUC = 0.987, test AUC = 0.951).* ***b****. Leave-one-subject-out cross-validation of the stimulation-aligned dimension: the learned axis generalizes across held-out patients at both half amplitude (AUC = 0.919) and full amplitude (AUC = 0.942), confirming a patient-general signature of stimulation state.* ***c****. The same leave-one-subject-out procedure applied to the auxiliary orthogonal dimensions performs at chance (half-amplitude AUC = 0.502; full-amplitude AUC = 0.480), establishing that stimulation-state information is confined to Dim 0.* ***d–g****. OFF-referenced signed distance along the stimulation-aligned dimension shows no association with normalized motor improvement for total score (****d****; ρ = −0.03, p = 0.854), rigidity (****e****; ρ = 0.05, p = 0.783), bradykinesia (****f****; ρ = 0.03, p = 0.843), or tremor (****g****; ρ = 0.19, p = 0.274); points colored by stimulation frequency (n = 36 patient–condition pairs, 9 patients × 4 frequencies).* ***h****. Feature importance for the stimulation-aligned dimension, ranked by magnitude (top five positive, top five negative). The axis is driven primarily by interhemispheric partial directed coherence (PDC) and high-beta spectral power, consistent with a representation of stimulation presence rather than therapeutic state.*

Therapeutic response emerges in the stimulation-orthogonal auxiliary space

Having isolated stimulation detection in Dim 0, we next asked whether motor benefit was represented in the orthogonal auxiliary space. For each full-amplitude stimulation condition, we computed the OFF-referenced signed distance of each latent dimension and related this distance to protocol-defined motor improvement across 36 patient-frequency observations. Because the orientation of individual latent axes is arbitrary, the analysis tested whether relative movement along each dimension covaried with clinical response rather than assigning intrinsic meaning to positive or negative axis direction.

Therapeutic information was concentrated in the first two auxiliary dimensions. Individually, Dim 1 and Dim 2 were each associated with total motor improvement (Fig. 3a), whereas Dims 3–5 showed no consistent relationship with clinical outcome (Fig. 3d; Table S1). A combined signed distance across Dims 1 and 2 provided the strongest summary metric, correlating with total motor improvement at full amplitude (Spearman $\rho = 0.497$, $p = 0.002$; Pearson $r = 0.536$, $p = 0.0008$), and was consistent across Kendall and distance correlation (Table S2). Adding the remaining auxiliary dimensions produced little improvement, indicating that most clinically relevant variance was captured by the two-dimensional auxiliary subspace rather than distributed broadly across the latent representation. Of note, however, this pooled correlation aggregates within- and between-frequency variation; we decompose these contributions in the next section using a linear mixed effects model.

The auxiliary signal was not equally related to all motor domains (Fig. 3b). The strongest symptom-level association was with rigidity, where Dim 2 showed the clearest individual relationship ($\rho = 0.479$, $p = 0.003$) and the combined Dim 1+2 metric also tracked improvement; bradykinesia showed weaker and less consistent associations, and tremor showed no robust relationship with the auxiliary subspace. This symptom hierarchy is consistent with a therapeutic signal dominated by STN-linked akinetic-rigid physiology rather than by tremor-related circuits that may be less completely captured by STN local field potentials. The two therapeutic dimensions also drew on a different feature set from the stimulation-aligned axis, loading more heavily on low-beta burst dynamics, phase–amplitude coupling, and alpha-band power (Fig. 3c), features more closely tied to parkinsonian motor pathophysiology.

This auxiliary geometry was not specific to the half-amplitude recordings that were used for training. Although the model was trained only on half-amplitude recordings, the latent positions and OFF-to-ON distance vectors it produced were broadly preserved at full amplitude (Supplementary Figure S1), and the auxiliary metric still related to motor improvement when the frozen encoder was applied there, where clinical effects were larger. This supports using the half-amplitude-trained representation in the full-amplitude clinical domain. Furthermore, the latent geometry was reproducible across ten independent random weight initializations of the GCE (Supplementary Fig. S2 and Supplementary Methods).

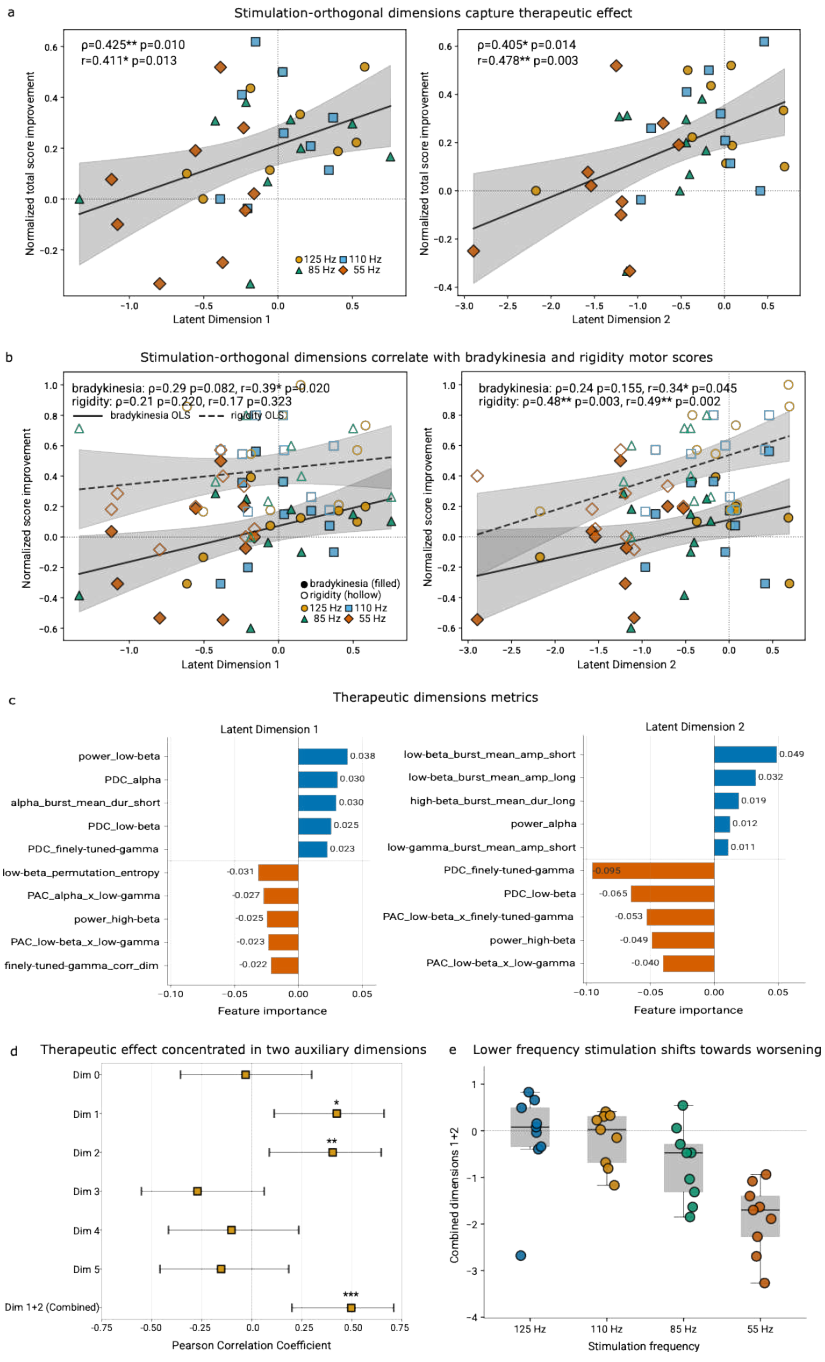


Figure 3. Stimulation-orthogonal dimensions encode a graded, frequency-dependent therapeutic signal. ***a,*** *OFF-referenced signed distance along auxiliary Latent Dimension 1 (left) and Dimension 2 (right) correlates with normalized total motor improvement at full amplitude (Dim 1: ρ = 0.425, p = 0.010; Dim 2: ρ = 0.405, p = 0.014; n = 36); points colored by stimulation frequency, solid line shows the linear fit.* ***b,*** *the same relationship for motor subscores, shown for Latent Dimension 1 (left) and Dimension 2 (right), with bradykinesia as filled markers and rigidity as hollow markers, with the strongest and most consistent association for rigidity along Dimension 2 (ρ = 0.479, p = 0.003).* ***c,*** *Feature importance for the two therapeutic dimensions (top five positive, top five negative). Dimensions 1 and 2 draw more heavily on low-beta burst dynamics, phase–amplitude coupling, and alpha-band features than the stimulation-aligned axis, linking them to parkinsonian motor pathophysiology.* ***d,*** *Pearson correlation between OFF-referenced signed distance and normalized total improvement for each latent dimension and for the summed Dim 1+2 metric, with 95% confidence intervals. Therapeutic information is concentrated in Dimensions 1 and 2 and is strongest in their combination, while the stimulation-aligned Dim 0 and Dims 3–5 are non-significant (**p < 0.01, ***p < 0.001).* ***e,*** *Combined signed distance in the Dim 1+2 subspace ($\Delta z_1 + \Delta z_2$) at full amplitude across stimulation frequencies (n = 9 patients per frequency); box shows median and interquartile range. Distance shifts systematically toward the anti-therapeutic direction as frequency decreases, with all nine patients showing negative distance at 55 Hz. This frequency ordering emerged without any frequency or outcome information in the training objective.*

We next asked whether the auxiliary therapeutic geometry differed systematically across stimulation frequencies. At full amplitude, all nine patients showed a strongly negative Dim 1+2 auxiliary metric at 55 Hz, with the higher frequencies graded toward zero (Fig. 3e; Table S3). The 55 Hz condition was also clinically least effective, producing the lowest mean motor improvement and clinical worsening in four of nine patients. This was not a sign of weak engagement: the Euclidean displacement of the auxiliary state was largest at 55 Hz (mean 1.49; Table S3), so low-frequency stimulation produced the most robust auxiliary perturbation while moving every patient in the same anti-therapeutic direction. Signed distance therefore captured what magnitude alone would miss: the least effective frequency drove the strongest and most uniform displacement toward the least therapeutic region of the auxiliary space.

Linear mixed-effects modeling localizes frequency-independent encoding to bradykinesia

As each patient contributed multiple stimulation conditions, we applied a more conservative linear mixed-effects analysis modeling patient as a random intercept with amplitude and stimulation frequency as fixed effects, pooling half- and full-amplitude observations for power (n = 72; 9 patients × 4 frequencies × 2 amplitudes). This tested whether latent dimensions predicted motor outcome beyond frequency and amplitude. Because the models used within-patient centered latent values rather than OFF-referenced signed distances, the sign of β is not directly comparable to the correlations reported above.

For the total motor score, Dim 0 remained null (β = +0.889, p = 0.344; negative control). Dim 2 was the strongest auxiliary predictor (β = −0.586, p = 0.037), and in the joint model with Dim 1 it both retained significance (β = −0.648, p = 0.019) and survived predictor-permutation testing (p_perm = 0.044), whereas Dim 1 reached only nominal parametric significance and did not survive permutation (β = −0.573, p = 0.044, p_perm = 0.081). Permutation results were stable across three seeds (Δ ≤ 0.008; N = 5,000) and Dims 3–5 were non-significant (all p ≥ 0.166). The subscore models localized this signal (Table S4). For bradykinesia, both dimensions survived permutation in the joint model (Dim 1 β = −0.802, p_perm = 0.037; Dim 2 β = −0.752, p_perm = 0.046). For rigidity, no auxiliary dimension predicted residual improvement once frequency was modeled (all p ≥ 0.263). Thus, rigidity's strong pooled correlation with auxiliary displacement is carried largely by covariation across frequencies, which the categorical frequency term absorbs, whereas the within-condition therapeutic signal that survives frequency control resides in bradykinesia. Dim 2 is therefore the primary therapeutic axis for overall motor state, with Dim 1 emerging as a co-contributor specifically for bradykinesia.

As benchmarks, we applied the same mixed-effects framework to two reference feature sets: unsupervised principal components of the original feature space and canonical spectral band-power measures. For PCA, only PC6 (3.7% variance explained, p = 0.029) and PC10 (2.7% variance explained, p = 0.050) reached nominal significance across the first 10 components. These associations arose from post hoc screening across multiple components and did not provide an architectural separation between stimulation detection and therapeutic variance. Similarly, canonical low-beta, high-beta and gamma power showed no graded association with normalized

total improvement after accounting for frequency and amplitude (all $p \geq 0.39$; non-significant under permutation; Table S6A), and full-amplitude OFF-referenced band-power changes were unrelated to total or symptom-specific improvement (Table S6B). Band power was also essentially flat across stimulation frequencies and did not reproduce the monotonic shift toward the anti-therapeutic 55 Hz region recovered by the auxiliary subspace (Table S6C). Thus, under the same statistical framework, neither unsupervised PCA nor canonical spectral power recovered the therapeutic structure identified by GCE. Feature-attribution analyses supported this dissociation: the therapeutic dimensions were driven by beta burst dynamics, phase–amplitude coupling and alpha-band features rather than by canonical beta or gamma power, indicating that GCE captured therapeutic structure not recovered by conventional spectral-power summaries.

The latent geometry transfers prospectively to an unseen subject

We next tested whether the learned geometry transferred to a patient who was not included in model training, hyperparameter selection, or analysis development. The pre-trained model was applied to P11, who underwent the same stimulation-frequency and amplitude protocol as the training cohort. Both therapeutically relevant dimensions showed generalization. The value along Dim 0 from the OFF baseline was positive across all eight ON conditions (Fig. 4a), indicating preservation of the stimulation-detection axis in an unseen patient. Across four frequencies and two amplitudes, Dim 1+2 signed distance correlated with rigidity improvement (Spearman $\rho = 0.731$, $p = 0.040$; Fig. 4b). The low-frequency pattern was concordant with the cohort result, with 55 Hz producing the largest negative auxiliary distance and the poorest rigidity outcome at both amplitudes. Total motor improvement showed the expected direction but was not significant, bradykinesia was not associated with the auxiliary metric, and additional tremor- and Dim 5-related effects were not present in the model-development cohort.

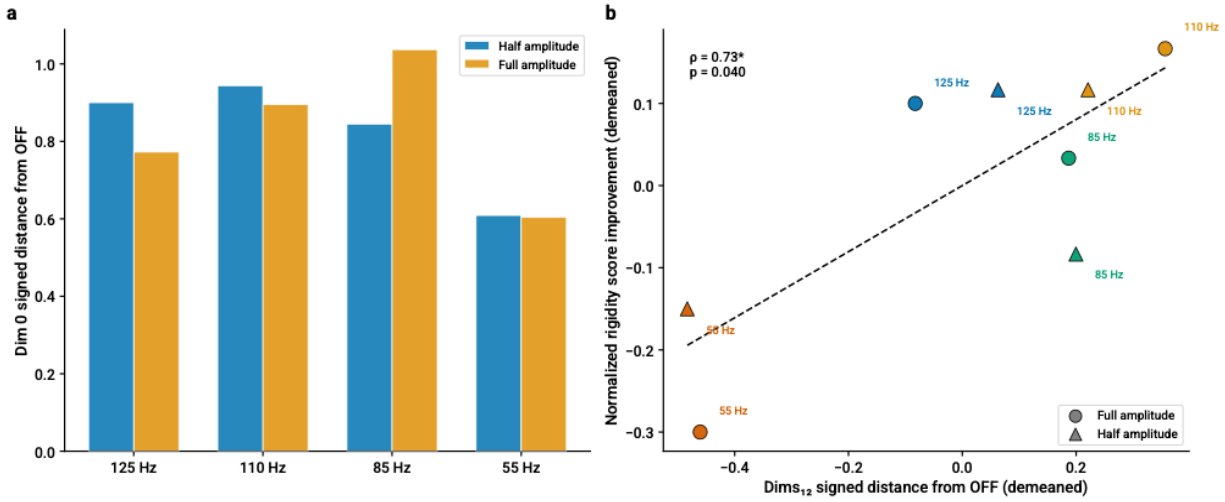


Figure 4. Prospective validation in a held-out patient (P11). *P11 was not included in model training, hyperparameter selection, or analysis development.* ***a,*** *Dim 0 signed distance from OFF across all four stimulation frequencies at both amplitudes. All eight values are positive (range 0.60–1.04), confirming transfer of the stimulation-detection axis to an unseen patient without retraining.* ***b,*** *Combined-amplitude correlation between dims$_{12}$ signed distance and normalized rigidity improvement (n=8; 4 frequencies × 2 amplitudes). The pre-specified primary validation criterion, a significant combined-amplitude Spearman correlation between dims$_{12}$ and rigidity at p<0.05 was met (ρ=0.731, p=0.040). Points are colored by frequency; circles indicate full amplitude; triangles indicate half amplitude. The 55 Hz condition produced the largest negative distance and the worst rigidity outcome at both amplitudes, replicating the pattern observed in the nine-patient training cohort. Axes are demeaned within patient.*

The therapeutic axis replicates in an independent kinematic cohort

Finally, we tested our model prospectively on an independent cohort (differing from the training dataset in both acquisition protocol and outcome modality). Concretely, three patients were recruited and evaluated with the frozen encoder. Their motor states were assessed not by clinical rating but by markerless video kinematics of finger tapping and hand movement, and subsequently scored by projection onto principal components published for a separate 154-patient cohort[23]. Because these components index continuously graded, bradykinesia-sensitive movement features, this cohort probes whether the auxiliary axis tracks therapeutic state through a different and higher-resolution readout than the ordinal scores used in the development cohort and in P11. We found that stimulation detection generalized: Dim 0 displacement was positive in all 14 ON recordings (per-patient AUC 0.54–0.82, below the development-cohort value as expected given that this protocol used only two stimulation frequencies and unilateral recordings). The combined auxiliary metric Dim 1 + 2 correlated with two of the kinematic components in the direction expected for therapeutic benefit (Fig. 5), finger-tapping performance consistency (FT-PC2: ρ = −0.763, p = 0.002, n = 14) and hand-movement speed (HM-PC1: ρ = −0.653, p = 0.011, n = 14); a third component, finger-tapping movement speed, trended in the same direction (FT-PC1: ρ = +0.481, p = 0.081), and no examined component moved against the direction expected for improvement. Both significant associations

remained significant on a balanced subset using a uniform number of recordings per patient (FT-PC2: ρ = −0.748, p = 0.005; HM-PC1: ρ = −0.629, p = 0.028; n = 12), indicating they did not depend on the unequal number of recordings per patient.

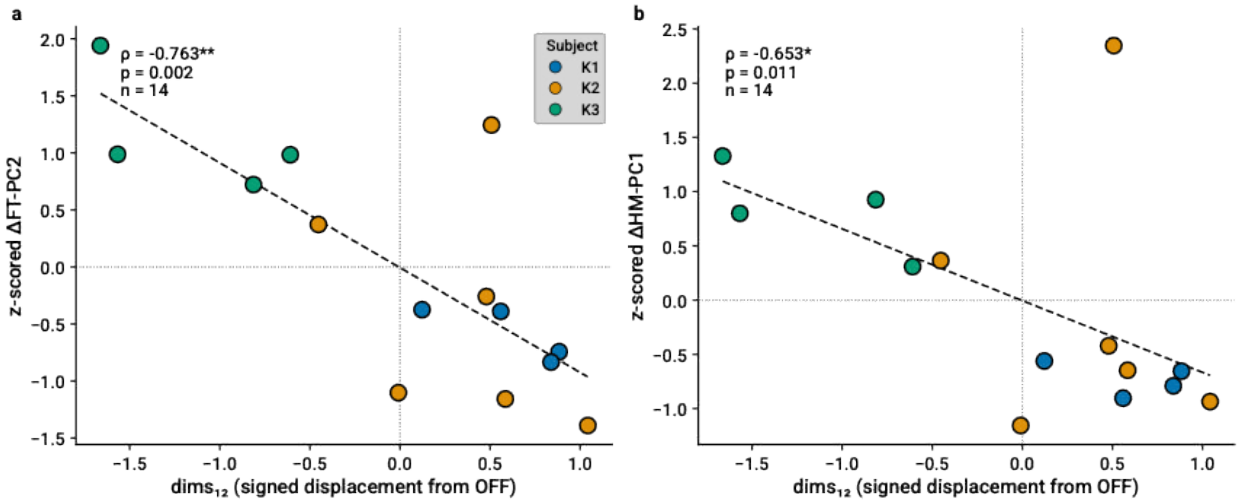


Figure 5. $dims_{12}$ correlates with kinematic motor improvement in an independent cohort from a separate study at the same center. *Each point represents one stimulation condition; points are colored by patient (K1–K3). The dashed line is an ordinary least-squares fit. Kinematic PC change scores (Δ relative to per-patient stimulation-OFF baseline) were z-scored within the cohort; PC loadings are frozen from Lange et al.[23] and were applied without refitting. (a) $dims_{12}$ vs finger-tapping performance consistency (FT-PC2); ρ = −0.763, p = 0.002, n = 14. (b) $dims_{12}$ vs hand-movement speed (HM-PC1); ρ = −0.653, p = 0.011, n = 14. In both panels the regression runs in the direction defined as therapeutic improvement in the levodopa reference cohort (decreasing FT-PC2 and decreasing HM-PC1). Stimulation detection transferred across cohorts: Dim 0 displacement was positive in all 14 ON recordings, with per-patient AUC 0.54–0.82. Both associations held in a balanced subset using a uniform four recordings per patient (FT-PC2: ρ = −0.748, p = 0.005; HM-PC1: ρ = −0.629, p = 0.028; n = 12). Correlations for all four examined kinematic components are reported in Supplementary Table S8.*

Discussion

The signal used to guide sensing-enabled deep brain stimulation is recorded from the same circuit that stimulation perturbs, so the presence of stimulation and its therapeutic effect are embedded in the same neural features[1,24]. We addressed this confound using Generative Causal Explanations, a deep learning framework which allows us to partition neural signals into (1) a subspace that determines a classification target and (2) an orthogonal subspace constrained to carry the remaining structure[22]. In contrast to standard approaches that use therapeutic labels, we trained a classifier to detect stimulation presence rather than therapeutic state, thereby repurposing the partition to quarantine the stimulation signature and expose therapeutic structure in the remaining dimensions. The stimulation-aligned axis classified stimulation state across patients and amplitudes (leave-one-

subject-out AUC = 0.94) yet was unrelated to motor outcome, whereas the auxiliary subspace tracked graded motor improvement ($\rho = 0.50$) and, without any supervision of outcome or frequency, recovered the anti-therapeutic signature of low-frequency stimulation.

This reframing departs from the two common strategies the field has used to confront the same confound. The first treats the stimulation contribution as an artifact to be subtracted. Methods such as period-based artifact reconstruction and removal[25] clean the recorded trace, but operate in signal space, so the covariation between stimulation presence and neural change survives into any decoder trained on the cleaned data. The second reads therapeutic information directly from the stimulation-evoked response, as in evoked resonant neural activity[26], a stimulation-locked pallido-subthalamic oscillation that reflects engagement of the indirect pathway, localizes to the dorsal subthalamic nucleus and predicts clinical response[27,28]. Evoked resonant activity is a powerful biomarker, but it binds the readout to the evoked response; our auxiliary subspace is its complement, defined by precisely what remains once the stimulation-locked component has been quarantined.

The same architecture was first applied clinically in the only prior use of generative causal models for neuromodulation, where the classifier-aligned dimension was treated as the biomarker itself, tracking recovery during subcallosal cingulate stimulation for depression[29]. That approach reads the therapeutic signal from the supervised dimension during brief stimulation-off windows, which is possible because the depressive state evolves over weeks and persists when stimulation is briefly interrupted. Our setting denies us that step: the therapeutic effect of DBS in Parkinson's is the on-stimulation state itself, so the biomarker must be read while stimulation is on. To our knowledge, this is the first use of this partition to quarantine the stimulation confound rather than to define the biomarker directly. Training the stimulation classifier on half-amplitude recordings, where stimulation reliably engages the circuit, but clinical benefit is small and inconsistent[1,8,24], biased supervision toward detectability, so that covariation between the auxiliary dimensions and motor outcome at full amplitude is less readily attributed to the classification target.

The strongest evidence that the auxiliary subspace encodes therapy rather than residual stimulation detectability came from the unseen stimulation frequencies. At 55 Hz (the least clinically effective setting) we observed the smallest mean improvement and frank worsening in four of nine patients. All nine patients moved in the same direction along the auxiliary metric, toward the anti-therapeutic region of the latent space. Because the training objective contained no information about frequency or outcome, this unanimity cannot have been inherited from the known frequency–outcome gradient; moreover, linear mixed-effects modeling confirmed that the auxiliary axis explained motor variance beyond frequency, amplitude and patient identity (Table S4). Low-frequency stimulation produced a larger displacement in the auxiliary subspace than the more effective higher frequencies yet the poorest outcomes, so circuit engagement and therapeutic benefit came apart in precisely the way that motivates our approach[15,30,31]. Set against the heterogeneous, anatomy-dependent geometry of effective settings, the convergence at 55 Hz suggests that the maladaptive response to low-frequency stimulation is mechanistically more stereotyped than therapeutic response itself.

The therapeutic axis was physiologically coherent. Its encoding followed a symptom hierarchy, strongest for rigidity, weaker and readout-dependent for bradykinesia, and absent for tremor, that mirrors the pathophysiology of the subthalamic nucleus, in which beta-band synchrony is tightly coupled to akinetic–rigid signs whereas tremor depends on cerebello-thalamic circuits poorly represented in subthalamic local field potentials[6,32,33]. The partition was equally coherent at the feature level. The stimulation-aligned axis was driven by interhemispheric directed coherence and high-beta power, whereas the therapeutic dimensions loaded on low-beta burst dynamics, phase–amplitude coupling and alpha-band activity, features[9,18,19] all of which are tied to parkinsonian motor state. The dissociation we built into the latent geometry thus reflects a dissociation in the underlying physiology, not a statistical orthogonality alone.

Both therapeutically relevant components of the representation transferred to a patient excluded from training, model selection and development. Applied without retraining, the stimulation-aligned axis remained positive across all this patient's stimulation conditions, and the auxiliary metric correlated with rigidity improvement ($\rho = 0.731$, $p = 0.040$). The 55 Hz condition again produced the largest anti-therapeutic displacement and the poorest rigidity outcome. Other domains were more variable, so a single held-out patient can motivate but not establish generalizability.

A second cohort, acquired prospectively under a different protocol and read out through video kinematics rather than clinical ratings, converged on the same auxiliary axis from a different outcome modality. Here the metric tracked the two components indexing bradykinetic improvement, finger-tapping performance consistency and hand-movement speed, both shifting in the direction defined as improvement in the levodopa reference cohort. Video kinematics of finger tapping and hand movement quantify bradykinesia and carry no rigidity readout, so this cohort tests bradykinetic tracking by construction rather than adjudicating between symptoms; that the axis tracks bradykinesia here is therefore expected. The symptom that shows the strongest association therefore shifts with measurement resolution while the underlying axis does not: the pre-specified Spearman endpoint aggregates across frequencies and is dominated by rigidity, the most frequency-sensitive sign, whereas the mixed-effects model removes the frequency main effect and isolates the within-condition residual, which bradykinesia (a more continuously graded sign) retains and the higher-resolution kinematics resolve. The three-patient cohort is too small to fit such a model reliably, so its associations are reported descriptively and corroborated on a balanced subset rather than treated as a population-level test.

These results sit within a broader shift in adaptive DBS away from beta-band power and toward multimodal, patient-specific biomarkers[2,10,34,35]. Machine-learning decoders have expanded what can be recovered from recordings during stimulation, including medication and motor state[36], controllable cortical markers[37], movement-triggered control[38] and video-derived behavioral markers[39,40]. Each asks which signal best predicts state; however, none address how to keep the presence of stimulation from contaminating that signal in the first place, and beta-guided closed-loop systems inherit the same entanglement[1,11,33]. The separation achieved here is architectural

rather than incidental, i.e., under the same mixed-effects framework, neither an unsupervised decomposition of these features nor canonical beta and gamma power reproduced it. The nearest precedent separates an external state, sleep, from the motor signal using a second detector[41]; our problem differs in kind, because the confound is the intervention's own signature, and we separate it within a single representation rather than by adding a parallel classifier.

The confound is not specific to subthalamic stimulation in Parkinson's disease but arises wherever recording and stimulation share a substrate. In spinal cord stimulation, the evoked compound action potential is itself a stimulation-evoked signal recorded from the stimulated tissue and is already used for closed-loop control[42,43]. In non-invasive brain stimulation, the electroencephalographic readout is contaminated by the stimulus, and the field's response has been to gate stimulation on brain state rather than to separate the two[44,45]. Psychiatric DBS faces the same dissociation between the modulated circuit and the clinical target[29]. In each setting the partition we describe, i.e., supervising on intervention presence to free an orthogonal space for intervention effect, offers a route to biomarkers that would otherwise be confounded.

Several limitations qualify these findings. First, all recordings were acquired at a single center. Out-of-sample evidence came from one prospectively held-out patient and from a second, independent cohort of three patients recorded under a different stimulation protocol and assessed with a different outcome modality. Although these analyses test transfer across stimulation protocol and measurement type, both validation sets remain small and single center. Larger multi-center cohorts are therefore needed to establish generalizability and to confirm the symptom-specific encoding hierarchy. Second, recordings were sampled at 250 Hz, reflecting the fixed sensing ceiling of the Medtronic Percept PC system used here rather than a limitation of the modeling framework. This sampling rate excludes higher-frequency activity, including the range in which evoked resonant activity is typically observed, and such activity may carry additional therapeutic information[10,11,27]. Higher-rate sensing could allow the same framework to separate stimulation presence from therapeutic structure in these higher-frequency regimes. Third, all recordings were acquired within a single session, leaving the longitudinal stability of the representation untested. Finally, feature attribution for the auxiliary dimensions was qualitative; the specific feature rankings should therefore be interpreted as indicative rather than definitive.

Adaptive DBS today feeds back on beta-band power, adjusting stimulation amplitude to suppress it. Because beta suppression indexes stimulation engagement, an amplitude controller cannot distinguish engagement from benefit, and would be insensitive to, or even reinforce, an engaging but anti-therapeutic setting such as 55 Hz. By construction, the auxiliary therapeutic axis is decoupled from stimulation presence, making it a candidate feedback signal, i.e., a continuous readout of whether a configuration is moving a patient toward or away from a beneficial state. In doing so it points beyond amplitude-only adaptation toward closed-loop optimization of the full stimulation parameter space. Realizing this will require prospective, multi-center evaluation built around a patient-general encoder with brief per-patient calibration. More broadly, by separating the signature

of intervention presence from the signature of intervention effect, the approach offers a template for closed-loop biomarker discovery wherever the recording site and the site of intervention coincide.

Methods

Subjects and clinical assessment

For this proof-of-concept study, ten (10) subjects with advanced PD and Medtronic Percept PC IPG[46,47] with bilateral Sensight™ electrodes were investigated. For model development, we report results from nine (9) recruited subjects, and one (1) subject reserved for prospective validation. Mean age at the time of surgery was 64.1±2.9, mean MDS-UPDRS III score from Levodopa challenge test was 36.9±4.38 MedOFF, and 14.33±3.43 MedON[48] (Table 1). Data presented as mean ± SEM (SEM – Standard Error of Mean).

Table 1. Demographics and clinical scores for subjects under investigation.

Subject	***Age at manuscript preparation***	***Age at diagnosis***	***Age at surgery***	***L_Dopa_test Pre DBS (MDS-UPDRSIII)***	
				MED OFF	**MED ON**
P1	61	35	44	Externally implanted[1]	
P2	70	43	68	46	20
P3	68	53	64	14	3
P4	73	65	69	52	37
P5	78	66	75	31	9
P6	57	50	53	27	6
P7	67	52	65	35	10
P8	73	52	69	35	15
P10	74	63	71	57	20
P11	66	50	63	35	9

The study was approved by the Medical Ethics Committee of the Julius-Maximilians-University Würzburg, Germany (reference no. 20240305 01). The dataset was collected during routine clinical programming. All patients provided written informed consent, and the protocol involved was in accordance with the Declaration of Helsinki[49].

Antiparkinsonian medications were ceased at least 12 hours before and stimulation was switched off at least 30 minutes before beginning the experimental protocol to allow washout. All experiments were conducted in the presence of a neurologist (PN). To limit patient burden during the off-medication assessment, subjects were rated only on some items of the MDS-UPDRS Part III namely

[1]no L-dopa challenge available because P1 was implanted externally.

items 3.3 (Rigidity), 3.4 (Finger Tapping), 3.5 (Hand Movements), 3.6 (Pronation-Supination Movements of Hands), 3.7 (Toe Tapping), 3.8 (Leg Agility), 3.15 (Postural Tremor of the Hands), 3.16 (Kinetic Tremor of the Hands), 3.17 (Rest Tremor Amplitude), 3.18 (Constancy of Rest Tremor). In this manuscript, the total motor score refers to the sum of the above rated items rather than the full MDS-UPDRS Part III. Rigidity, bradykinesia, and tremor sub-scores were defined as the sums of their respective item groups. Normalized clinical scores were derived as described below under Statistical analysis.

DBS programming and LFP acquisition

All subjects had bilateral electrodes implanted in the subthalamic nucleus (STN) and had been receiving chronic DBS therapy for at least 1 year before the study. A monopolar review was performed by a neurologist to determine the clinically effective stimulation contacts, current amplitude, and best stimulation frequency.

The study protocol explored four stimulation frequencies (125 Hz, 110 Hz, 85 Hz, 55 Hz) at two amplitudes: each participant's clinically programmed amplitude and half that amplitude. These were chosen as the IPG only supports a LFP sampling rate of 250 Hz for sensing and we wanted to avoid aliasing which would have plagued the signals above the Nyquist frequency. Pulse width was fixed at the clinical setting for all participants (60 μs). Concurrent stimulation and sensing were performed using sandwich configurations, in which the sensing contact lay between two stimulating contacts. As Percept permits sensing only in such a configuration, the chronic therapeutic contacts were retained whenever they allowed a sandwich sensing arrangement; otherwise, stimulation was reassigned to an adjacent contact pair that enabled concurrent sensing. For each stimulation condition, the selected clinical items were rated after the wash-in period. Resting-state LFPs were additionally acquired for 120 s in OFF medication/OFF stimulation using the same sensing configuration and indefinite streaming mode. A 600-s wash-in period was allowed between the OFF baseline and the first ON condition, and a minimum interval of 30–60 s was maintained between successive stimulation settings. LFPs were recorded at 250 Hz.

Signal processing and feature extraction

Signal processing and artifact suppression were performed using custom MATLAB scripts (v24.2, MathWorks) and FieldTrip toolbox[50]. All recordings were high-pass filtered at 1 Hz using a sixth-order zero-phase Butterworth filter to remove low-frequency drift. Stimulation-on recordings were then processed with PARRM using the appropriate stimulation frequency to suppress stimulation artifacts[25].

LFP recordings were segmented into 10-s windows before feature extraction. Features were computed through two parallel pipelines. In the first pipeline, normalized spectral power was estimated in canonical frequency bands after parameterization of the power spectrum with FOOOF to separate periodic from aperiodic 1/f components[51]. Band-specific power was normalized by

broadband power in the 2–95 Hz range. The same pipeline also quantified interhemispheric directed connectivity using partial directed coherence (PDC)[52,53].

In the second pipeline, raw LFPs were filtered with a zero-phase FIR filter using a Blackman window to derive temporal burst features, nonlinear dynamical features, and within-hemisphere phase-amplitude coupling. Features were extracted in the alpha (9–12 Hz), low-beta (13–20 Hz), high-beta (21–30 Hz), low-gamma (31–50 Hz), and finely tuned gamma (55–90 Hz) bands. Cross-frequency coupling was computed for the alpha–low-gamma, alpha–finely tuned gamma, low-beta–low-gamma, low-beta–finely tuned gamma, and high-beta–finely tuned gamma pairs.

Phase-amplitude coupling was quantified using the Tort algorithm as implemented in PACtools[54,55]. PDC was computed with the asympPDC package in MATLAB[52,53]. To capture nonlinear signal structure, four additional measures were extracted: the largest Lyapunov exponent, estimated with the Rosenstein algorithm via the nolds library[56,57]; correlation dimension, estimated with Theiler's implementation of the Grassberger–Procaccia algorithm[58,59]; sample entropy, computed using a Richman–Moorman implementation in MATLAB[60,61]; and permutation entropy, computed using the Bandt–Pompe algorithm[62]. In total, 70 features were extracted for each 10-s segment.

Disentangling stimulation presence from therapeutic effect

The central analytic objective was to separate neural variance associated with stimulation presence from neural variance associated with therapeutic response. To do so, we used an amplitude-separated design within the Generative Causal Explanations[22] (GCE) framework. The model was trained on a half-amplitude binary task contrasting OFF recordings with the half-amplitude recording acquired at each participant's clinically selected best-frequency ON condition. Half-amplitude stimulation was chosen as a domain in which stimulation-related neural signatures were expected to remain detectable while clinical benefit was attenuated and less consistent than at full amplitude.

Under this design, the stimulation-aligned dimension was optimized to encode stimulation-state information, whereas the auxiliary latent dimensions were constrained to represent variance orthogonal to the stimulation classifier. After training, the encoder was frozen and applied without retraining to all stimulation recordings across the tested frequencies, amplitudes and subjects. Associations between auxiliary latent geometry and clinical outcome in the full-amplitude dataset were interpreted as evidence that therapeutically relevant information was represented in variance not explained by the stimulation-detection target.

Model architecture and training

Before model fitting, features were z-score normalized within participant to reduce between-patient variation in baseline LFP characteristics, electrode impedance, and lead placement. For participant p, normalized features were computed as,

$$x_{norm,p} = \frac{x_p - \mu_p}{\sigma_p}$$

where μ_p and σ_p denote the mean and standard deviation estimated from all segments contributed by participant p within the dataset being analyzed. This approach preserved within-participant neural dynamics while reducing dependence on absolute feature magnitudes across participants. For held-out and independent validation data, the same participant-specific normalization procedure was applied before passing features through the frozen encoder; no outcome information was used in this step.

The GCE model consisted of an encoder that mapped the 70-dimensional feature vector into a latent space of $K + L$ dimensions and a decoder that reconstructed the original feature vector from the latent representation, i.e., a variational autoencoder architecture[63]. Network architecture details are provided in Supplementary Table S5.

Model parameters were optimized to minimize,

$$\mathcal{L} = \lambda[MSE(x, \hat{x}) + D_{KL}\big(q(z \mid x) \parallel p(z)\big)] - I(z_{\mathrm{dim0}}; Y)$$

where $\mathrm{MSE}(x, \hat{x})$ is the reconstruction loss, D_{KL} is the Kullback–Leibler divergence between the approximate posterior and the prior, $I(z_{\mathrm{dim0}}; Y)$ is the mutual information between the stimulation-aligned dimension and the binary stimulation label Y, and λ controls the trade-off between reconstruction fidelity and stimulation discriminability. In the half-amplitude training dataset, Y contrasted OFF recordings with the participant-specific best-frequency ON condition.

Hyperparameter selection

Hyperparameters were selected in two stages following the procedure of O'Shaughnessy et al.[22] First, with $K = 0$ and $\lambda = 0$, we increased the number of auxiliary dimensions L until reconstruction loss plateaued. This initial search identified $L^* = 6$ as the minimum latent dimensionality required for stable reconstruction. We then incrementally increased the number of causal dimensions K while decrementing L, adjusting λ to preserve reconstruction quality while maximizing the mutual information between the stimulation discriminative component and the binary classifier output. This stage suggested an approximate solution of $K \approx 1$, $L \approx 5$, and $\lambda \approx 0.07$. A fine-grained grid search was then performed in a local neighborhood around this solution over $K \in \{1,2\}$, $L \in \{3,4,5,6\}$, and $\lambda \in \{0.04, 0.045, \ldots, 0.10\}$. The final configuration was selected as the model that maximized $I(z_{\mathrm{dim0}}; Y)$ while preserving reconstruction quality within the plateau range. This yielded $K = 1$, $L = 5$, and $\lambda = 0.068$, with $I(z_{\mathrm{dim0}}; Y) = 0.481$ nats.

Latent space biomarker computation

The trained encoder was applied to OFF recordings and to all stimulation recordings at both half and full amplitude across 125, 110, 85 and 55 Hz, yielding latent representations for each 10-s segment.

For each participant p, condition c, and latent dimension i, we computed the mean latent position $\bar{z}_{p,c,i}$ by averaging across all segments within that condition.

Signed distance from the participant's OFF baseline was defined as,

$$\Delta z_{p,c,i} = \bar{z}_{p,c,i} - \bar{z}_{p,OFF,i}$$

Positive values indicate movement in the positive direction of latent dimension i, whereas negative values indicate movement in the opposite direction.

The primary combined auxiliary metric was the summed signed distance across auxiliary dimensions 1 and 2,

$$dims_{12}^{\text{signed}}(p,c) = \Delta z_{p,c,1} + \Delta z_{p,c,2}$$

For completeness, we also computed the summed signed distance across all auxiliary dimensions,

$$dims_{1-5}^{signed}(p,c) = \sum_{i=1}^{5} \Delta\, z_{p,c,i}$$

Distance along Dim 0 was computed identically and used as a negative control metric for stimulation detection.

Leave-one-subject-out cross-validation of the stimulation-aligned dimension

To test whether Dim 0 generalized across participants as a marker of stimulation state, we performed leave-one-subject-out cross-validation on the half-amplitude binary dataset comprising OFF recordings and participant-specific best-frequency ON recordings (280 samples from the 9-participant model-development cohort). In each fold, a logistic regression classifier was trained on Dim 0 values from 8 participants and evaluated on the held-out participant. Model discrimination was summarized as the area under the receiver operating characteristic curve (AUC). The mean AUC across folds was reported together with a 95% confidence interval obtained from 2,000 bootstrap resamples of the fold-wise AUC values.

In this implementation, the encoder was kept fixed and LOSO therefore evaluated the generalizability of the learned Dim 0 representation rather than full end-to-end retraining of the GCE model.

Cross-amplitude validation

To assess whether latent geometry learned from half-amplitude training data was preserved at clinical amplitude, we applied the trained encoder to full-amplitude recordings. For each participant and stimulation condition, mean latent positions were computed across all 10-s segments.

Cross-amplitude similarity was quantified in three ways. First, for each latent dimension, Pearson correlation was computed between half- and full-amplitude mean latent values across all

participant-condition pairs ($n = 36$; 9 participants × 4 conditions). Second, cosine similarity was computed between the OFF-to-ON distance vectors at half and full amplitude for each participant and each ON condition ($n = 36$; 9 participants × 4 ON conditions). Third, for each latent dimension, Pearson correlation was computed between half- and full-amplitude signed distance values across participant-condition pairs.

Statistical analysis

The primary inferential analysis used linear mixed-effects (LME) models to test whether latent dimensions predicted clinical improvement independently of stimulation frequency. Because the full-amplitude dataset alone contained only 36 participant-condition observations, LME analyses were performed on the combined half- and full-amplitude dataset ($n = 72$; 9 participants × 4 frequencies × 2 amplitudes), with amplitude included as a fixed covariate.

For each latent dimension, within-participant centered latent values were computed by subtracting each participant's mean across all stimulation conditions. The same model structure was fitted separately for normalized total motor improvement, normalized bradykinesia improvement and normalized rigidity improvement:

$$norm_{outcome} \sim dim_{i,wc} + amp + C(frequency) + (\,1 \mid patient\,)$$

where $\mathrm{norm}_{outcome}$ denotes one of the three normalized clinical outcomes, $\mathrm{dim}_{i,\mathrm{wc}}$ is the within-participant centered value of latent dimension i, amp denotes stimulation amplitude (half vs. full), and $C(frequency)$ models stimulation frequency as a categorical fixed effect. A random intercept for participant accounted for baseline between-participant differences. Models were fit by restricted maximum likelihood using the *lbfgs optimizer* in statsmodels[64]. Dim 0 was tested in the same framework as a negative control axis. When individual-dimension analyses suggested shared variance, joint models including multiple auxiliary dimensions were also fit.

Given the small number of participant clusters (n = 9), model-based p-values were supplemented with a two-sided predictor-permutation test using 5,000 permutations. In each permutation, the within-participant centered predictor was shuffled globally while preserving the patient-level outcome and covariate structure. Permutation p-values were defined as the proportion of permuted absolute regression coefficients greater than or equal to the observed absolute coefficient.

As descriptive and sensitivity analyses, pooled correlations were computed on the 36 full-amplitude participant-frequency observations (9 participants × 4 frequencies). Normalized improvement was defined as,

$$normalized\ \ improvement = \frac{score_{OFF} - score_{ON}}{score_{OFF}}$$

so that positive values indicated clinical improvement relative to OFF. Spearman rank correlation was used as the primary pooled correlation metric, with Pearson correlation, Kendall's τ, and distance correlation reported as secondary measures.

To benchmark GCE against an unsupervised dimensionality-reduction baseline, principal component analysis (PCA) was applied to the 70-dimensional participant-normalized feature space using all 1,008 ON-condition samples (all frequencies, both amplitudes, 9 participants). The first 10 principal components were extracted and subjected to the same LME framework described above (within-patient centered PC scores as predictors, with amplitude and frequency covariates and random patient intercepts).

As a further benchmark using the spectral features most familiar in this literature, the canonical band-power features were carried through the identical set of analyses applied to the latent dimensions. Low-beta (13–20 Hz), high-beta (21–30 Hz), low-gamma (31–50 Hz), and finely tuned gamma (55–90 Hz) power, normalized by broadband 2–95 Hz power as described above, were each entered into the same linear mixed-effects model (within-participant centered band power as the predictor, with amplitude and stimulation frequency as fixed effects and a random patient intercept) and supplemented by the same predictor-permutation test (5,000 permutations). In parallel, OFF-referenced signed change in each band was correlated with normalized clinical improvement at full amplitude using Spearman correlation, and mean signed change was summarized per stimulation frequency, mirroring the descriptive analyses applied to the auxiliary metric (Table S6).

Feature importance

To characterize which input features drive each latent dimension, we computed feature attributions using GradientShap (Captum library) with OFF-condition samples as the baseline distribution[65]. Attributions targeted the encoder posterior mean, μ, for each latent dimension and were computed across all 1,008 ON-condition segments before averaging. Convergence deltas were monitored to assess approximation quality. As convergence deltas for auxiliary dimensions were approximately 5–6× mean attribution magnitude, feature rankings are reported qualitatively rather than as precise indices.

Prospectively held-out validation

To provide initial out-of-sample evaluation, the trained model was applied to one held-out participant (P11), who was not included in model fitting, hyperparameter selection, or development. P11 underwent the same acquisition protocol as the model-development cohort, and features were extracted and normalized using the same processing pipeline and participant-specific z-score procedure.

Before inspection of the held-out outcomes, one primary success criterion was specified: the combined-amplitude Spearman correlation between $dims_{12}^{signed}$ and rigidity improvement had to reach $p < 0.05$. This endpoint was chosen to match the main auxiliary metric and the symptom

domain showing the strongest encoding in the model-development cohort. All other analyses in the held-out participant were treated as exploratory and were not corrected for multiplicity.

Independent kinematic validation cohort

To test whether the auxiliary axis transferred to an independent dataset acquired under a different protocol (approved by the Medical Ethics Committee of the Julius-Maximilians-University Würzburg, Germany, ref. no. 2024-100_1-dvh) and assessed with a different outcome modality (Table S7), we applied the frozen encoder to three additional patients from a separate study. This protocol used two stimulation frequencies and unilateral stimulation and recording. Recordings were acquired after overnight withdrawal of antiparkinsonian medication for at least 12 h, with the same stimulation washout procedure as in the development cohort. Local field potentials were processed using the identical artifact-suppression, feature-extraction and within-participant normalization pipeline used for the development cohort, and were passed through the frozen encoder without retraining, yielding the same six-dimensional latent representation. For each stimulation condition, OFF-referenced signed displacement along the auxiliary dimensions, $dims_{12}$, was computed as in the main analysis, and transfer of the stimulation-detection axis was assessed from Dim 0. Because stimulation and recording were unilateral, each recording was paired with the kinematic assessment of the contralateral hand.

Motor state was quantified not by clinical rating but by markerless video kinematics of finger tapping (MDS-UPDRS III item 3.4) and hand movement (item 3.5), extracted with VisionMD[40]. For each task, the kinematic feature vector of every recording was projected onto fixed principal-component loadings published by Lange et al. for an independent 154-patient cohort[23], without refitting. Analyses used the two components carrying a defined direction of levodopa-related improvement in that work, Movement Speed (PC1) and Performance Consistency (PC2), for each task. The Movement Timing and Scale component (PC3) was not analyzed: for finger tapping the feature pipeline does not export the rate and rate-decay measures that define it, and for hand movement it showed no significant levodopa effect and therefore no defined improvement direction. For each component a change score (ΔPC) was computed relative to that patient's and hand's stimulation-OFF baseline, mirroring the OFF-referencing of the latent displacements. Because the absolute units of the projected components are arbitrary, ΔPC scores were z-scored within the cohort before correlation. The direction corresponding to clinical improvement was taken from the published levodopa mixed-model coefficients: increasing Movement Speed and decreasing Performance Consistency for finger tapping and decreasing Movement Speed and increasing Performance Consistency for hand movement.

Auxiliary displacement ($dims_{12}$) was related to each component change score by Spearman correlation across the pooled set of recordings (14 recordings from three patients). Because the cohort comprised only three patients, the pooled coefficient is reported as a within-cohort

association rather than a population-level test, and a balanced subset with a uniform number of recordings per patient (n = 12) was analyzed as a sensitivity check; correlations for all four examined kinematic components are provided in Supplementary Table S8. Applying levodopa-derived kinematic axes to stimulation-induced change assumes that the two therapies shift movement along comparable dimensions, an assumption supported by the reported concordance between levodopa and subthalamic stimulation motor effects[66].

Acknowledgements

T.S. was supported by a PhD fellowship of the Deutscher Akademischer Austauschdienst (DAAD). F.L. was supported by the Interdisciplinary Center for Clinical Research (IZKF Würzburg; Project-ID Z-3BC/19). T.S., P.N., G.A., F.L., J.V., M.M., M.R., R.P. were supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; Project-ID 424778381, TRR 295).

Author contributions

T.S. and R.P. conceived the study. T.S., P.N., G.A., and F.L. collected the dataset. F.L. performed the kinematic-cohort principal-component analysis. M.R. and J.V. provided clinical oversight. T.S. implemented the software, performed all analyses and prepared the figures. G.A. contributed to data analysis. R.P. and M.M. supervised the work. M.R., M.M. and J.V. acquired funding. T.S. wrote the original draft. All authors reviewed and edited the manuscript and approved the final version.

Competing interests

M.R. reports grant support and honoraria for speaking from Medtronic and Boston Scientific. J.V. reports grants and personal fees from Medtronic, grants and personal fees from Boston Scientific, and personal fees from Abbott. F.L. reports honoraria for speaking from Boston Scientific. These interests are outside the submitted work. The remaining authors (T.S., P.N., G.A., M.M. and R.P.) declare no competing interests.

Data availability

Electrophysiology recordings and clinical scores that support the findings of this study are not publicly available owing to data-protection restrictions and the terms of patient consent. De-identified derived data (latent representations and condition-level summary metrics) sufficient to reproduce the reported analyses are available from the corresponding author on reasonable request, subject to a data-sharing agreement and approval by the Medical Ethics Committee of the Julius-Maximilians-University Würzburg. Data from the independent kinematic cohort are governed by the data-sharing terms of that separate study and are available from F.L. on reasonable request, subject to the same restrictions. Principal-component loadings used to score the kinematic data are published in Lange et al.[23]

Code availability

Code for the generative causal modeling framework, latent-space analyses and statistical procedures is available at https://github.com/tanmoysil/GCE. The implementation builds on the generative causal explanations framework of O'Shaughnessy et al.[22]

References

1 Neumann, W. J., Gilron, R., Little, S. & Tinkhauser, G. Adaptive Deep Brain Stimulation: From Experimental Evidence Toward Practical Implementation. *Mov Disord* **38**, 937-948 (2023). https://doi.org:10.1002/mds.29415

2 Shah, A. *et al.* Combining Multimodal Biomarkers to Guide Deep Brain Stimulation Programming in Parkinson Disease. *Neuromodulation* **26**, 320-332 (2023). https://doi.org:10.1016/j.neurom.2022.01.017

3 Wang, S. *et al.* Closed-Loop Adaptive Deep Brain Stimulation in Parkinson's Disease: Procedures to Achieve It and Future Perspectives. *J Parkinsons Dis* **13**, 453-471 (2023). https://doi.org:10.3233/JPD-225053

4 Schuepbach, W. M. *et al.* Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med* **368**, 610-622 (2013). https://doi.org:10.1056/NEJMoa1205158

5 Deuschl, G. *et al.* A randomized trial of deep-brain stimulation for Parkinson's disease. *N Engl J Med* **355**, 896-908 (2006). https://doi.org:10.1056/NEJMoa060281

6 Kuhn, A. A. *et al.* Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. *Exp Neurol* **215**, 380-387 (2009). https://doi.org:10.1016/j.expneurol.2008.11.008

7 Eusebio, A., Cagnan, H. & Brown, P. Does suppression of oscillatory synchronisation mediate some of the therapeutic effects of DBS in patients with Parkinson's disease? *Front Integr Neurosci* **6**, 47 (2012). https://doi.org:10.3389/fnint.2012.00047

8 Feldmann, L. K. *et al.* Toward therapeutic electrophysiology: beta-band suppression as a biomarker in chronic local field potential recordings. *NPJ Parkinsons Dis* **8**, 44 (2022). https://doi.org:10.1038/s41531-022-00301-2

9 Tinkhauser, G. *et al.* Beta burst dynamics in Parkinson's disease OFF and ON dopaminergic medication. *Brain* **140**, 2968-2981 (2017). https://doi.org:10.1093/brain/awx252

10 Oehrn, C. R. *et al.* Chronic adaptive deep brain stimulation versus conventional stimulation in Parkinson's disease: a blinded randomized feasibility trial. *Nat Med* **30**, 3345-3356 (2024). https://doi.org:10.1038/s41591-024-03196-z

11 Bronte-Stewart, H. M. *et al.* Long-Term Personalized Adaptive Deep Brain Stimulation in Parkinson Disease: A Nonrandomized Clinical Trial. *JAMA Neurol* **82**, 1171-1180 (2025). https://doi.org:10.1001/jamaneurol.2025.2781

12 Hill, M. E. *et al.* Paradoxical Modulation of STN beta-Band Activity with Medication Compared to Deep Brain Stimulation. *Mov Disord* **39**, 192-197 (2024). https://doi.org:10.1002/mds.29634

13 Mathiopoulou, V. *et al.* Modulation of subthalamic beta oscillations by movement, dopamine, and deep brain stimulation in Parkinson’s disease. *npj Parkinson's Disease* **10**, 77 (2024). https://doi.org:10.1038/s41531-024-00693-3

14 Xie, T. *et al.* Low-frequency stimulation of STN-DBS reduces aspiration and freezing of gait in patients with PD. *Neurology* **84**, 415-420 (2015). https://doi.org:10.1212/WNL.0000000000001184

15 Blumenfeld, Z. *et al.* Sixty-hertz stimulation improves bradykinesia and amplifies subthalamic low-frequency oscillations. *Mov Disord* **32**, 80-88 (2017). https://doi.org:10.1002/mds.26837

16 Ramdhani, R. A. *et al.* Differential spatiotemporal gait effects with frequency and dopaminergic modulation in STN-DBS. *Frontiers in Aging Neuroscience* **Volume 15 - 2023** (2023). https://doi.org:10.3389/fnagi.2023.1206533
17 Colombo, A. *et al.* Finely Tuned γ Tracks Medication Cycles in Parkinson's Disease: An Ambulatory Brain-Sense Study. *Movement Disorders* **40**, 881-895 (2025). https://doi.org:10.1002/mds.30160
18 Tinkhauser, G. *et al.* The modulatory effect of adaptive deep brain stimulation on beta bursts in Parkinson's disease. *Brain* **140**, 1053-1067 (2017). https://doi.org:10.1093/brain/awx010
19 Meidahl, A. C. *et al.* Synchronised spiking activity underlies phase amplitude coupling in the subthalamic nucleus of Parkinson's disease patients. *Neurobiol Dis* **127**, 101-113 (2019). https://doi.org:10.1016/j.nbd.2019.02.005
20 Muthuraman, M. *et al.* Cross-frequency coupling between gamma oscillations and deep brain stimulation frequency in Parkinson's disease. *Brain* **143**, 3393-3407 (2020). https://doi.org:10.1093/brain/awaa297
21 Olaru, M. *et al.* Deep brain stimulation-entrained gamma oscillations in chronic home recordings in Parkinson's disease. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* **18**, 132-141 (2025). https://doi.org:10.1016/j.brs.2025.01.011
22 O'Shaughnessy, M., Canal, G., Connor, M., Davenport, M. & Rozell, C. in *Proceedings of the 34th International Conference on Neural Information Processing Systems* Article 458 (Curran Associates Inc., Vancouver, BC, Canada, 2020).
23 Lange, F. *et al.* Computer vision uncovers three fundamental dimensions of levodopa-responsive motor improvement in Parkinson's disease. *npj Parkinson's Disease* **11**, 140 (2025). https://doi.org:10.1038/s41531-025-00999-w
24 Bouthour, W. *et al.* Biomarkers for closed-loop deep brain stimulation in Parkinson disease and beyond. *Nat Rev Neurol* **15**, 343-352 (2019). https://doi.org:10.1038/s41582-019-0166-4
25 Dastin-van Rijn, E. M. *et al.* Uncovering biomarkers during therapeutic neuromodulation with PARRM: Period-based Artifact Reconstruction and Removal Method. *Cell Rep Methods* **1**, 100010 (2021). https://doi.org:10.1016/j.crmeth.2021.100010
26 Sinclair, N. C. *et al.* Deep brain stimulation for Parkinson's disease modulates high-frequency evoked and spontaneous neural activity. *Neurobiol Dis* **130**, 104522 (2019). https://doi.org:10.1016/j.nbd.2019.104522
27 Steiner, L. A. *et al.* Neural signatures of indirect pathway activity during subthalamic stimulation in Parkinson's disease. *Nature Communications* **15**, 3130 (2024). https://doi.org:10.1038/s41467-024-47552-6
28 Van Bogaert, T. *et al.* Evoked Resonant Neural Activity Reveals an Electrophysiologic Sweet Spot for Directional Subthalamic Nucleus Deep Brain Stimulation in Parkinson's Disease. *Neuromodulation: Technology at the Neural Interface* (2025). https://doi.org:10.1016/j.neurom.2025.09.303
29 Alagapan, S. *et al.* Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130-138 (2023). https://doi.org:10.1038/s41586-023-06541-3
30 Johnson, L. A. *et al.* Closed-Loop Deep Brain Stimulation Effects on Parkinsonian Motor Symptoms in a Non-Human Primate - Is Beta Enough? *Brain Stimul* **9**, 892-896 (2016). https://doi.org:10.1016/j.brs.2016.06.051

31 Su, D. *et al*. Frequency-dependent effects of subthalamic deep brain stimulation on motor symptoms in Parkinson's disease: a meta-analysis of controlled trials. *Scientific Reports* **8**, 14456 (2018). https://doi.org:10.1038/s41598-018-32161-3

32 Helmich, R. C., Hallett, M., Deuschl, G., Toni, I. & Bloem, B. R. Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? *Brain* **135**, 3206-3226 (2012). https://doi.org:10.1093/brain/aws023

33 Lauro, P. M., Lee, S., Akbar, U. & Asaad, W. F. Subthalamic-Cortical Network Reorganization during Parkinson's Tremor. *J Neurosci* **41**, 9844-9858 (2021). https://doi.org:10.1523/JNEUROSCI.0854-21.2021

34 Gilron, R. *et al*. Long-term wireless streaming of neural recordings for circuit discovery and adaptive stimulation in individuals with Parkinson's disease. *Nat Biotechnol* **39**, 1078-1085 (2021). https://doi.org:10.1038/s41587-021-00897-5

35 di Biase, L. *et al*. Adaptive, personalized closed-loop therapy for Parkinson's disease: biochemical, neurophysiological, and wearable sensing systems. *Expert Rev Neurother* **21**, 1371-1388 (2021). https://doi.org:10.1080/14737175.2021.2000392

36 Sand, D. *et al*. Machine learning-based personalized subthalamic biomarkers predict ON-OFF levodopa states in Parkinson patients. *J Neural Eng* **18**, 046058 (2021). https://doi.org:10.1088/1741-2552/abfc1d

37 Castano-Candamil, S. *et al*. Identifying controllable cortical neural markers with machine learning for adaptive deep brain stimulation in Parkinson's disease. *Neuroimage Clin* **28**, 102376 (2020). https://doi.org:10.1016/j.nicl.2020.102376

38 Merk, T. *et al*. Electrocorticography is superior to subthalamic local field potentials for movement decoding in Parkinson's disease. *Elife* **11**, e75126 (2022). https://doi.org:10.7554/eLife.75126

39 Peach, R. *et al*. Head movement dynamics in dystonia: a multi-centre retrospective study using visual perceptive deep learning. *NPJ Digit Med* **7**, 160 (2024). https://doi.org:10.1038/s41746-024-01140-6

40 Acevedo, G. *et al*. VisionMD: an open-source tool for video-based analysis of motor function in movement disorders. *NPJ Parkinsons Dis* **11**, 27 (2025). https://doi.org:10.1038/s41531-025-00876-6

41 Gilron, R. *et al*. Sleep-Aware Adaptive Deep Brain Stimulation Control: Chronic Use at Home With Dual Independent Linear Discriminate Detectors. *Frontiers in Neuroscience* **Volume 15 - 2021** (2021). https://doi.org:10.3389/fnins.2021.732499

42 Will, A. *et al*. Improvements in Therapy Experience With Evoked Compound Action Potential Controlled, Closed-Loop Spinal Cord Stimulation-Primary Outcome of the ECHO-MAC Randomized Clinical Trial. *The Journal of Pain* **25** (2024). https://doi.org:10.1016/j.jpain.2024.104646

43 Vallejo, R., Chakravarthy, K., Will, A., Trutnau, K. & Dinsmoor, D. A New Direction for Closed-Loop Spinal Cord Stimulation: Combining Contemporary Therapy Paradigms with Evoked Compound Action Potential Sensing. *J Pain Res* **14**, 3909-3918 (2021). https://doi.org:10.2147/jpr.S344568

44 Zrenner, C. & Ziemann, U. Closed-Loop Brain Stimulation. *Biological Psychiatry* **95**, 545-552 (2024). https://doi.org:10.1016/j.biopsych.2023.09.014

45 Wischnewski, M. *et al*. Real-time TMS-EEG for brain state-controlled research and precision treatment: a narrative review and guide. *Journal of Neural Engineering* **21**, 061001 (2024). https://doi.org:10.1088/1741-2552/ad8a8e

46 Jimenez-Shahed, J. Device profile of the percept PC deep brain stimulation system for the treatment of Parkinson's disease and related disorders. *Expert Rev Med Devices* **18**, 319-332 (2021). https://doi.org:10.1080/17434440.2021.1909471

47 Thenaisie, Y. *et al*. Towards adaptive deep brain stimulation: clinical and technical notes on a novel commercial device for chronic brain sensing. *J Neural Eng* **18**, 042002 (2021). https://doi.org:10.1088/1741-2552/ac1d5b

48 Saranza, G. & Lang, A. E. Levodopa challenge test: indications, protocol, and guide. *J Neurol* **268**, 3135-3143 (2021). https://doi.org:10.1007/s00415-020-09810-7

49 World Medical, A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191-2194 (2013). https://doi.org:10.1001/jama.2013.281053

50 Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* **2011**, 156869 (2011). https://doi.org:10.1155/2011/156869

51 Donoghue, T. *et al*. Parameterizing neural power spectra into periodic and aperiodic components. *Nat Neurosci* **23**, 1655-1665 (2020). https://doi.org:10.1038/s41593-020-00744-x

52 Baccala, L. A. & Sameshima, K. Partial Directed Coherence and the Vector Autoregressive Modelling Myth and a Caveat. *Front Netw Physiol* **2**, 845327 (2022). https://doi.org:10.3389/fnetp.2022.845327

53 Sameshima, K. & Baccala, L. A. asympPDC Package. *MATLAB Central File Exchange* (2022). <https://www.mathworks.com/matlabcentral/fileexchange/116290-asymppdc-package?s_tid=srchtitle>.

54 Dupre la Tour, T. *et al*. Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS Comput Biol* **13**, e1005893 (2017). https://doi.org:10.1371/journal.pcbi.1005893

55 Tort, A. B., Komorowski, R., Eichenbaum, H. & Kopell, N. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *J Neurophysiol* **104**, 1195-1210 (2010). https://doi.org:10.1152/jn.00106.2010

56 Schölzel, C. Nonlinear measures for dynamical systems. (2019). https://doi.org:10.5281/zenodo.3814723

57 Rosenstein, M. T., Collins, J. J. & De Luca, C. J. A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets. *Physica D-Nonlinear Phenomena* **65**, 117-134 (1993). https://doi.org:10.1016/0167-2789(93)90009-P

58 Theiler, J. Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Phys Rev A Gen Phys* **36**, 4456-4462 (1987). https://doi.org:10.1103/physreva.36.4456

59 Grassberger, P. & Procaccia, I. Measuring the Strangeness of Strange Attractors. *Physica D* **9**, 189-208 (1983). https://doi.org:10.1016/0167-2789(83)90298-1

60 Richman, J. S. & Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* **278**, H2039-2049 (2000). https://doi.org:10.1152/ajpheart.2000.278.6.H2039

61 Lee, K. Sample Entropy. (2012). <https://www.mathworks.com/matlabcentral/fileexchange/35784-sample-entropy>.

62 Bandt, C. & Pompe, B. Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* **88**, 174102 (2002). https://doi.org:10.1103/PhysRevLett.88.174102

63 Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* (2013). https://doi.org:10.48550/arXiv.1312.6114

64 Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, 92-96 (2010). https://doi.org:10.25080/Majora-92bf1922-011

65 Kokhlikyan, N. *et al.* Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896* (2020). https://doi.org:10.48550/arXiv.2009.07896

66 Timmermann, L. *et al.* Differential effects of levodopa and subthalamic nucleus deep brain stimulation on bradykinesia in Parkinson's disease. *Movement Disorders* **23**, 218-227 (2008). https://doi.org:10.1002/mds.21808

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- natbmesuppldocs.docx