SymbolSight: Minimizing Inter-Symbol Interference for Reading with Prosthetic Vision

Jasmine Lesner[1,*], Michael Beyeler[1,2]

Abstract—Retinal prostheses restore limited visual perception, but low spatial resolution and temporal persistence make reading difficult. In sequential letter presentation, the afterimage of one symbol can interfere with perception of the next, leading to systematic recognition errors. Rather than relying on future hardware improvements, we investigate whether optimizing the visual symbols themselves can mitigate this temporal interference. We present SYMBOLSIGHT, a computational framework that selects symbol-to-letter mappings to minimize confusion among frequently adjacent letters. Using simulated prosthetic vision (SPV) and a neural proxy observer, we estimate pairwise symbol confusability and optimize assignments using language-specific bigram statistics. Across simulations in Arabic, Bulgarian, and English, the resulting heterogeneous symbol sets reduced predicted confusion by a median factor of 22 relative to native alphabets. These results suggest that standard typography is poorly matched to serial, low-bandwidth prosthetic vision and demonstrate how computational modeling can efficiently narrow the design space of visual encodings to generate high-potential candidates for future psychophysical and clinical evaluation.

Index Terms—Prosthetic Reading, Symbol Set Selection, Retinal Prostheses, Bionic Vision, Visual Rehabilitation, Simulated Prosthetic Vision, Computational Sensory Substitution, Assistive Technology,

I. INTRODUCTION

Retinal prostheses provide a crucial restorative option for individuals with degenerative retinal diseases, but functional tasks like reading remain difficult because the effective spatiotemporal channel is severely constrained. To date, PRIMA is the only retinal implant with published evidence of form vision [1] in the sense of reliably perceiving structured spatial patterns. At the same time, earlier devices such as Alpha-IMS and Argus II showed that letter-level reading is not categorically impossible for all users [2], [3]. With training and careful stimulus presentation, some participants could identify isolated characters presented one symbol at a time.

These mixed outcomes point to two interacting bottlenecks that limit sequential reading in practice. First, spatial nonlinearities can break the assumption that multi-electrode stimulation produces a percept that is the linear sum of single-electrode percepts [4], [5]. This complicates any attempt to render conventional letterforms, since increasing spatial detail by recruiting more electrodes does not necessarily yield predictable, compositional structure at the perceptual level [6].

*Corresponding author: jlesner@ucsb.edu

[1]J.L. and M.B. are with the Department of Computer Science, University of California, Santa Barbara, CA 93106, USA.

[2]M.B. is also with the Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106, USA.
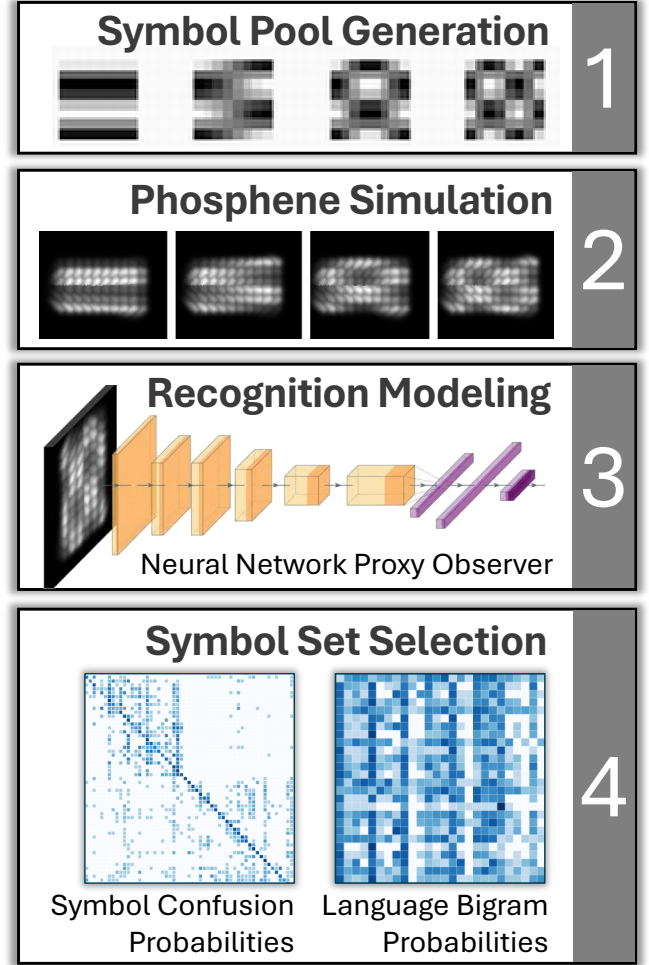
Fig. 1. SYMBOLSIGHT pipeline: candidate symbols undergo phosphene simulation and recognition modeling, then are assigned to letters based on confusion probabilities and bigram statistics.

Second, temporal nonlinearities introduce inter-symbol interference. Percepts can persist and fade over hundreds of milliseconds or longer [7], [8], so the afterimage of one symbol can distort the next. In this regime, the discriminability of a symbol depends not only on its own geometry but also on the preceding symbol.

Prior work on reading in bionic vision has largely focused on modeling and accommodating known distortions [9]–[11]. Complementary work in applied perception and display design has explored how symbols and typefaces can remain legible

under degraded viewing conditions [12], [13]. These efforts provide important foundations, but they typically assume a fixed alphabet and attempt to recover it through better hardware, better stimulation, or better decoding. In contrast, here we ask a different question: *given that the channel is low-bandwidth and exhibits spatial and temporal nonlinearities, what symbol set should we transmit in the first place?*

To test this, we present SYMBOLSIGHT (Fig. 1), a framework that optimizes the symbol-to-letter mapping for specific languages. The pipeline consists of: (1) generating a diverse pool of candidate glyphs; (2) simulating prosthetic spatial and temporal distortions; (3) estimating pairwise confusability using a neural network as a consistent proxy observer; and (4) assigning symbols to letters to minimize the expected confusion cost based on language-specific letter transition probabilities.

Our primary contributions are:

- We introduce a modular computational pipeline that combines simulated prosthetic vision (SPV), a neural proxy observer, and combinatorial optimization. This approach allows us to navigate the vast search space of potential symbol mappings, filtering for candidates that maximize distinguishability prior to costly human-subject testing.
- We formulate a language-aware cost function that specifically targets temporal persistence, a key bottleneck in sequential reading. By weighting symbol confusability against language-specific bigram frequencies, we minimize perceptual interference for the most frequent letter transitions rather than treating all errors equally.
- We demonstrate that optimal legibility in prosthetic vision may require abandoning typographic cohesion. In simulations across Arabic, Bulgarian, and English, we find that mixed pools containing diverse glyphs (e.g., Latin, Braille, discrete cosine transform (DCT) basis functions) reduce predicted confusion by a median factor of 22 compared to native alphabets. This finding challenges the assumption that standard letters are the appropriate baseline for prosthetic interfaces.

II. METHODS

We render high-contrast symbols and present them one at a time, consistent with standard SPV paradigms. Due to implant limitations, symbols are shown sequentially rather than in parallel. Depending on electrode placement and axonal activation, users may perceive small focal spots, elongated streaks, or irregular shapes rather than uniform patterns [14], [15]. Our goal is to model how these spatiotemporal distortions affect symbol discrimination and to optimize symbol assignments accordingly.

A. Simulated Prosthetic Vision

We use pulse2percept [16], [17] to simulate a 16×16 electrode grid, a resolution commonly used in prior SPV studies. While higher-density arrays may provide additional spatial detail, current spread and axonal activation often limit effective resolution [4].

Spatial Distortion. We simulate three levels of spatial distortion (shown top to bottom in the middle column of Fig. 3) following Han et al. [18]: *Low distortion*: no axonal stimulation ($\rho = 100\,\mu\mathrm{m}$, $\lambda = 0\,\mu\mathrm{m}$); *Medium distortion*: intermediate axonal stimulation ($\rho = 300\,\mu\mathrm{m}$, $\lambda = 1000\,\mu\mathrm{m}$); *High distortion*: strong axonal stimulation ($\rho = 500\,\mu\mathrm{m}$, $\lambda = 5000\,\mu\mathrm{m}$).

Temporal Distortion. In sequential prosthetic reading, the percept of a current symbol x_t is degraded by the visual persistence of the preceding symbol x_{t-1}. While biological persistence involves complex, nonlinear decay dynamics [19], [20], the dominant recognition bottleneck is inter-symbol interference, where the fading afterimage of the previous symbol obscures the next.

To capture this effect without full time-domain simulation, we approximate temporal interference as linear superposition. We use MixUp augmentation [21] to generate convex combinations of symbol pairs:

$$\hat{x} = \gamma x_i + (1 - \gamma)x_j, \quad \gamma \sim \mathrm{Beta}(2.0, 2.0), \qquad (1)$$

where x_i and x_j denote pixel intensity vectors for the current and preceding symbols. Unlike standard MixUp, we set $\alpha = \beta = 2.0$ to emphasize high-overlap cases ($\gamma \approx 0.5$), reflecting strong temporal ghosting. This forces the observer to discriminate symbols based on features that remain robust under perceptual overlap.

B. Simulated Observer

We quantify how overlapping percepts affect recognition using a convolutional neural network (CNN) pre-trained on ImageNet as a simulated observer. We chose MobileNetV3Large [23] for computational efficiency. Under each distortion condition, we fine-tune a separate classifier and estimate pairwise confusion probabilities across the full symbol pool.

While CNNs do not replicate human psychophysics in detail and may differ in decision strategies for specific stimuli, prior work has shown that task-optimized deep networks operating on simulated prosthetic vision exhibit strong correspondence with human performance across multiple functional visual tasks, including letter and symbol recognition under phosphene distortions [24]. We do not use the network to predict absolute human error rates. Instead, the resulting confusion matrix F is used to rank the relative separability of symbol pairs under overlap. This acts as a filter for structural ambiguity: if a high-capacity feature extractor cannot reliably distinguish two superimposed symbols under simulated distortion, the signal itself likely lacks sufficient information for robust discrimination by a human observer. We therefore treat the CNN-derived confusion matrix as a consistent metric of *structural distinguishability*.

C. Symbol Families

For testing, we construct a diverse pool of visual symbols drawn from established alphabets and simple synthetic patterns: *Latin Symbols*: The 26-letter alphabet used by English
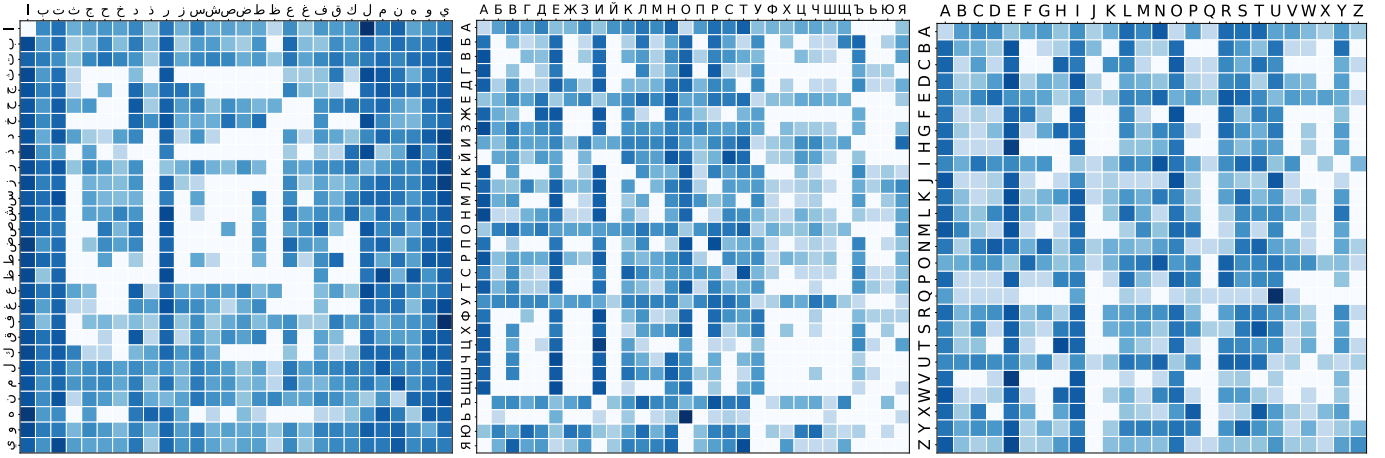
Fig. 2. **Letter transition probabilities across three languages.** Heatmaps show $P(L_{n+1} \mid L_n)$ for **Arabic** (left), **Bulgarian** (middle), and **English** (right). The vertical axis is the leading (current) letter; the horizontal axis is the following (next) letter. Darker cells indicate higher probability. Language-specific bigram probabilities were estimated from November 2023 Wikipedia database dumps [22] for Arabic, Bulgarian, and English. For each language, we processed the first 100,000 Wikipedia documents. English text was case-folded and filtered to the 26 Latin letters using regex word extraction. Bulgarian text was Unicode NFC-normalized and case-folded across its 30-letter Cyrillic alphabet, with the accented variant (Cyrillic I with grave accent) mapped to plain Cyrillic I. Arabic text underwent NFKD decomposition with diacritical marks stripped, *tatweel* removed, *alif maqsura* normalized to *yeh*, *ta marbuta* mapped to *ta*, and standalone *hamza* treated as a word boundary, yielding a 28-letter alphabet. Documents with fewer than 500 valid letters were excluded to ensure reliable statistics. The resulting co-occurrence matrices were normalized to estimate conditional bigram probabilities $P(L_{n+1} \mid L_n)$ used in the assignment optimization. For visualization, each row of the heatmap is normalized to sum to 1.

and many other languages. *Braille Symbols*: A tactile system of raised dots in 2×3 cells, optimized for fingertip discrimination rather than visual recognition. *Arabic Symbols*: The 28-letter Arabic script with cursive, connected letterforms. *DCT Symbols*: Basis functions from the two-dimensional discrete cosine transform [25], sinusoidal gratings that efficiently capture fundamental visual features. *Cyrillic Symbols*: The 30-letter Bulgarian alphabet, representative of the broader Cyrillic family. Adapted from Greek, it includes letters shared with Latin (e.g., A, O, T) alongside distinctly Slavic forms (e.g., Zhe, Sha, Yu, Ya).

D. Symbol Assignment Optimization

We combine two empirical inputs: (i) a symbol confusion matrix $F \in \mathbb{R}^{S \times S}$ measured under simulated prosthetic distortions, and (ii) a language-specific letter bigram matrix $C \in \mathbb{R}^{L \times L}$. We seek a mapping π from letters to symbols that minimizes expected confusion between successive letters:

$$\text{Cost} = \sum_{i=0}^{L-1} \sum_{\substack{j=0 \\ j \neq i}}^{L-1} C_{i,j} \, F_{\pi(i),\pi(j)} \tag{2}$$

This cost increases when frequently co-occurring letters are assigned to visually confusable symbols. Minimizing Eq. 2 reduces expected reading errors due to temporal interference. The formulation treats all letter confusions equally and is intended as a baseline; future extensions could incorporate semantic or task-dependent error weights.

We solve the assignment using the Hungarian algorithm [26] for the reported experiments. To isolate the effect of optimization from symbol diversity alone, we compare SYMBOLSIGHT

against two baselines: *Native* alphabets and *Randomized* assignments drawn from the full symbol pool while ignoring F and C. We evaluate 201 random seeds to estimate both mean performance and optimistic bounds. For each candidate alphabet, we renormalize predicted symbol distributions so probabilities sum to one over the retained symbol set before computing the cost in Eq. 2.

E. Evaluation Languages and Letter Statistics

We evaluate the framework on Arabic, Bulgarian, and English, which differ in script topology and alphabet size. For each language, we estimate letter bigram probabilities $P(L_{n+1} \mid L_n)$ from Wikipedia text and treat these statistics as fixed, language-specific inputs defining the matrix C. Fig. 2 visualizes the resulting bigram matrices used in all experiments.

III. RESULTS

A. Symbol Confusion Structure

Fig. 3 summarizes the pairwise symbol confusion probabilities estimated from our proxy observers under increasing distortion, revealing how symbol families interact under spatial and temporal interference.

First, at low distortion, the matrices exhibit a faint block structure along the diagonal, with each block corresponding to a symbol family: Latin (indices 0–25), Braille (indices 26–51), Arabic (indices 52–79), DCT (indices 80–115), and Cyrillic (indices 116–145). This block structure indicates that symbols tend to be confused with others from the same family, reflecting within-family visual similarity. As distortion increases, this block structure progressively dissolves: family boundaries become less distinct, and off-diagonal confusion
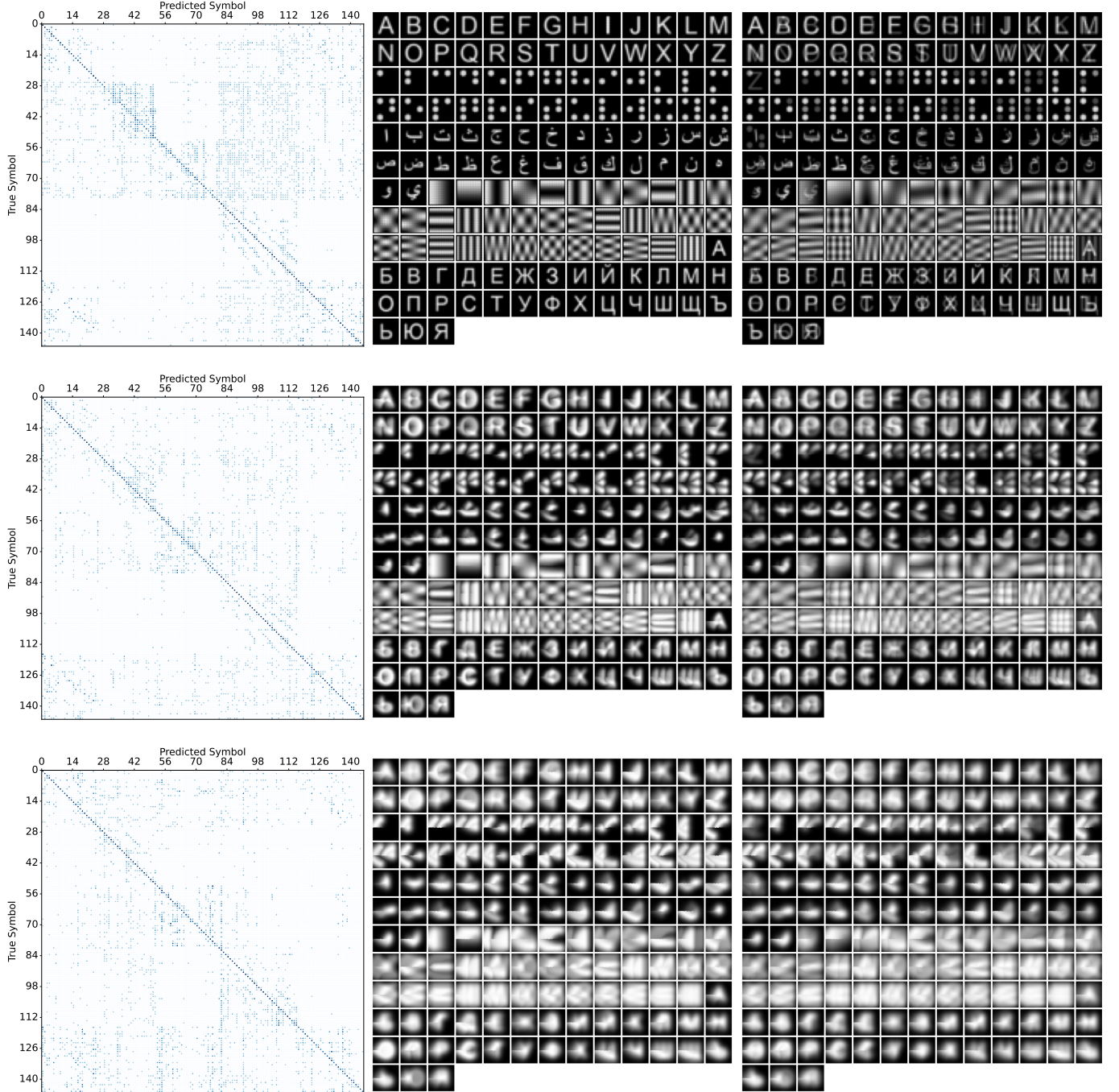
Fig. 3. **Top row**: low distortion; **middle row**: medium distortion; **bottom row**: high distortion. **Middle column**: 146 symbols with spatial distortion (Latin 0–25, Braille 26–51, Arabic 52–79, DCT 80–115, Cyrillic 116–145). **Right column**: example of temporal distortion showing perceptual residue when symbols are presented sequentially. **Left column**: symbol confusion probability heatmaps from fine-tuned neural networks evaluated on a held out validation dataset. Darker cells indicate higher confusion; white indicates near-zero. Our proxy observer networks were trained on a symbol pool of 146 glyphs comprising 26 Latin letters, 26 Braille characters, 28 Arabic letters, 36 DCT basis patterns, and 30 Cyrillic letters. Each symbol was rendered at 64×64 pixels and processed through four distortion conditions using the pulse2percept library [16]: undistorted, and three prosthetic vision simulations on a 16×16 electrode grid with phosphene sizes $\rho \in \{100, 300, 500\}$ μm and axon streak parameters $\lambda \in \{0, 1000, 5000\}$ μm. For each distortion level, a separate MobileNetV3Large [23] classifier was trained with ImageNet-pretrained weights frozen and a custom head consisting of global average pooling, 20% dropout, and a 146-way softmax output layer with L1-L2 regularization ($\ell_1 = 0.001$, $\ell_2 = 0.02$). Training used the Adam optimizer with learning rate of 10^{-4}, batch size of 64, and MixUp augmentation [21] with $\alpha = \beta = 2.0$ to generate 500 augmented training samples per symbol (73,000 images) and 100 validation samples per symbol (14,600 images). Early stopping monitored validation loss with patience of 20 epochs, and learning rate was reduced by half after 3 epochs without improvement. To match MobileNetV3Large's input requirements, images were resized to 224×224 RGB. The train and validation sets contained all 146 symbol classes with independently augmented samples, ensuring no leakage between splits.

Distortion	Native (baseline)	Randomized (mean ± std)	Randomized (best ablation)	SymbolSight Optimized
Arabic				
Low	715	607 ± 242	201	**54**
Medium	987	327 ± 194	124	**45**
High	1264	474 ± 403	156	**47**
Bulgarian				
Low	811	615 ± 251	231	**27**
Medium	422	308 ± 103	134	**34**
High	673	459 ± 232	158	**31**
English				
Low	700	646 ± 264	246	**32**
Medium	458	329 ± 176	133	**24**
High	563	475 ± 226	127	**38**

Distortion	Native (baseline)	Randomized (mean ± std)	Randomized (best ablation)	SymbolSight Optimized
Arabic				
Low	1.0	1.3 ± 0.5	3.6	**13.3**
Medium	1.0	3.5 ± 1.2	8.0	**21.7**
High	1.0	3.4 ± 1.4	8.1	**27.2**
Bulgarian				
Low	1.0	1.5 ± 0.6	3.5	**29.6**
Medium	1.0	1.5 ± 0.5	3.1	**12.5**
High	1.0	1.8 ± 0.8	4.3	**21.9**
English				
Low	1.0	1.3 ± 0.5	2.8	**21.6**
Medium	1.0	1.7 ± 0.6	3.4	**19.2**
High	1.0	1.5 ± 0.7	4.4	**14.7**
Median:	1.0	—	3.6	**21.6**

disperses more uniformly across the matrix rather than concentrating within family-specific blocks. At high distortion, the confusion pattern approaches a sparser, more scattered distribution, suggesting that prosthetic distortion obscures the fine-grained features that distinguish symbol families.

Second, cross-family confusion occurs predictably between Latin and Cyrillic symbols, reflected in off-diagonal activity linking these two blocks. This pattern persists across distortion levels, as shared letterforms such as A, E, K, M, O, and T remain visually similar even under prosthetic distortion.

Third, DCT symbols exhibit a distinctive diagonal banding pattern within their block: because DCT basis functions vary systematically in spatial frequency and orientation, neighboring indices often represent similar frequencies, leading to mutual confusion along diagonals. However, DCT symbols are rarely confused with letter-based symbols from other sets, as their periodic grating structure is visually distinct from the curved and angular strokes of alphabetic characters.

Fourth, Braille symbols show elevated off-diagonal confusion at low distortion due to temporal smearing: the linear superposition of two sparse dot patterns often inadvertently creates a geometry resembling a third valid Braille symbol. Unintuitively, this systematic confusion *decreases* at higher spatial distortion levels. As pulse2percept blurs individual dots into streaks, the precise combinatorial structure required for these specific misidentifications dissolves. This yields a critical insight: spatial prosthetic distortions do not uniformly degrade information but can occasionally mitigate specific temporal interference patterns.

Taken together, these observations show that symbol confusability in SPV depends strongly on symbol family and the distortion regime, and that the resulting collision patterns are predictable and can be exploited algorithmically rather than dismissed as noise.

B. Quantitative Performance Across Assignments

To assess how these symbol collision patterns translate into reading performance, we compare native alphabets, randomized heterogeneous assignments, and optimized mappings produced by SYMBOLSIGHT. Tables I and II show our findings.

First, *native baseline costs exhibit non-monotonic behavior with respect to distortion.* Notably, native Bulgarian and English symbols yield higher predicted confusion at low distortion than at medium levels (e.g., Bulgarian: 811 vs. 422). This mirrors the Braille phenomenon noted earlier, where prosthetic distortions may increase separability for certain symbol pairs. This counterintuitive finding suggests that legibility may not degrade linearly with signal fidelity, underscoring the necessity of computational modeling to identify optimal operating points that intuition alone might discard.

Second, *diversity alone improves legibility.* Despite the occasional resilience of native symbols, randomized assignments from the mixed pool consistently outperform the native baseline (e.g., Arabic Medium: 327 vs. 987). This suggests that standard alphabets, constrained by typographic cohesion, are generally suboptimal for prosthetic vision compared to a mixed pool containing distinct shapes like DCT gratings and Braille.

Third, *optimization provides the critical gain.* SYMBOLSIGHT outperforms even the *best* randomized seeds by factors of 3–8×. This confirms that minimizing Eq. 2, specifically avoiding collisions for high-frequency bigrams, yields gains far beyond what chance selection can provide.

Finally, *the improvement is robust.* Table II shows SYMBOLSIGHT reduces predicted confusion by a median factor of 21.6× compared to native text (range 12.5–29.6×). This improvement holds across all three languages and distortion levels, demonstrating the framework's effectiveness regardless of orthography or specific distortion severity.

C. Qualitative Structure of Optimized Alphabets

Figures 4, 5, and 6 show the symbols selected for each language at each distortion level. Several observations stand out. First, the optimized sets are heterogeneous: they draw from Latin, Cyrillic, Braille, and DCT sources, mixing alphabets and synthetic patterns within a single assignment. This mixing exploits cross-set distinctiveness visible in the confusion matrices (Fig. 3), where symbols from different families are rarely confused.
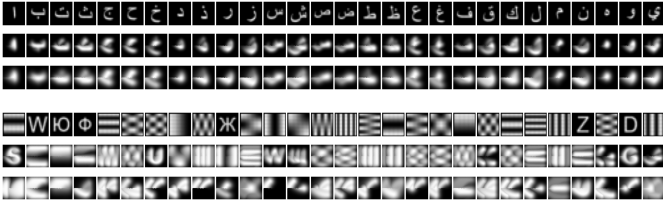
Fig. 4. **Arabic symbol comparison.** Top three rows: native Arabic at low, medium, and high distortion. Bottom three rows: optimized symbols at each distortion level in matching order. Note how the native characters blur into indistinguishable blobs at high distortion (Row 3), whereas the optimized glyphs maintain distinct structural footprints (Row 6).
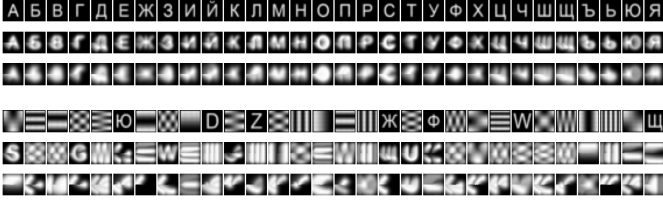

Fig. 5. **Bulgarian symbol comparison.** Top three rows: native Cyrillic at low, medium, and high distortion. Bottom three rows: optimized symbols at each distortion level in matching order.
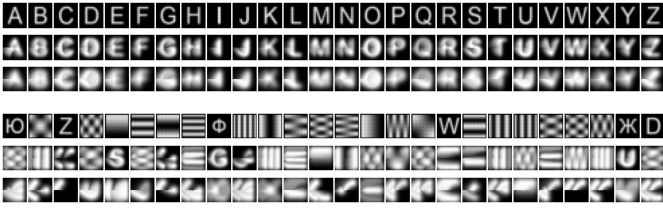

Fig. 6. **English symbol comparison.** Top three rows: native Latin at low, medium, and high distortion. Bottom three rows: optimized symbols at each distortion level in matching order.

Second, as distortion increases, the algorithm shifts toward symbols with coarse, high-contrast structure, preferring Braille and DCT symbols, as their low-frequency content survives blur better than fine strokes.

Third, comparing native symbols at high distortion (third row of each figure) to optimized symbols at high distortion (sixth row) makes the benefit apparent: many native letters become visually inseparable under the simulated distortion, whereas optimized symbols retain clearly different gross shapes.

Finally, the benefit at low and medium distortion is less visually obvious in static images because temporal distortion, the ghosting of sequential symbols, cannot be depicted in a single frame. The confusion cost reductions (Tables I and II) confirm that optimized sets also outperform baselines in these conditions.

IV. DISCUSSION

SYMBOLSIGHT is motivated by the observation that in serial, symbol-at-a-time prosthetic reading, the dominant bottleneck is not single-symbol recognition, but interference from the persistence of the *previous* symbol. By explicitly penalizing visual similarity between frequently adjacent letters (high-probability bigrams), the framework targets the error modes introduced by temporal nonlinearities rather than attempting to reconstruct fine spatial detail. The observed $22\times$ reduction in predicted confusion cost suggests that native alphabets are poorly matched to this viewing modality and that substantial gains are achievable through software-level optimization alone.

A. Optimizing the Symbol Codebook Matters

The central finding of this study is that symbol choice and assignment have a first-order impact on predicted reading performance under prosthetic vision constraints, independent of hardware changes. Across three languages and multiple distortion regimes, optimized symbol-to-letter mappings consistently outperformed native alphabets and randomized heterogeneous baselines (Table II). This indicates that standard typography, designed for high-acuity parallel reading, constitutes a suboptimal codebook for serial, low-bandwidth prosthetic vision, and that heterogeneous symbol sets can substantially reduce temporal interference.

B. Modeling Assumptions and Limitations

A primary challenge in prosthetic vision research is the combinatorial explosion of parameter spaces, which renders exhaustive human testing impractical. We position this framework not as a replacement for clinical trials, but as a computational sieve to identify high-potential symbol sets. We do not claim our neural network observer replicates human psychophysics or the idiosyncratic phosphene maps of individual patients. Rather, the network serves as a consistent upper-bound discriminator: if a deep convolutional model cannot reliably distinguish two superimposed symbols under simulated distortion, it is unlikely that a human observer operating under comparable spatiotemporal constraints will do so [24]. Our simulation also relies on approximations. Temporal persistence is modeled via MixUp-based linear superposition, which abstracts complex decay dynamics [19] into feature-space interference. In addition, the resulting confusion matrices depend on rendering choices such as font weight, size, and stroke geometry. While we standardized these parameters to isolate symbol geometry, alternative typographic instantiations would alter the confusion structure. Importantly, these factors do not limit the framework itself, as both temporal models and typographic parameters can be incorporated directly into the optimization loop.

C. Clinical Interpretation and Path Forward

While natural reading relies on suprasegmental cues such as word shape, these cues require parallel letter processing and are unavailable in the symbol-at-a-time modality imposed by current low-resolution prostheses [11]. Consequently, the heterogeneity of the optimized symbol sets does not sacrifice word-level information, because temporal sequencing has already eliminated it. From a clinical perspective, the relevant trade-off is therefore not visual cohesion, but reliability.

Although the optimized sets introduce a learned mapping that mixes Latin letters with Braille- and DCT-like symbols, prosthetic vision already demands substantial rehabilitation. In this context, learning an optimized symbol code represents a redirection of training toward a more reliable input signal rather than an added cognitive burden [27]. Looking forward, this work can be extended beyond letter-level optimization to word-level encodings and re-instantiated using device- or user-specific phosphene maps [28] as these become available for retinal and cortical implants.

V. CONCLUSION

We introduced SYMBOLSIGHT, which treats sequential prosthetic reading as communication over a noisy channel with memory and optimizes the symbol codebook rather than the hardware. Using SPV and a proxy observer to estimate pairwise interference costs, then weighting these costs by language-specific bigram statistics, SYMBOLSIGHT reduced the expected confusion cost by a median factor of $22\times$ across Arabic, Bulgarian, and English, consistently outperforming native alphabets and randomized heterogeneous baselines. In future work, we will validate these findings through human psychophysical experiments under sequential presentation and extend the framework to incorporate device- and user-specific percept maps.

REFERENCES

[1] F. G. Holz, Y. L. Mer, M. M. K. Muqit, L.-O. Hattenbach, A. Cusumano, S. Grisanti, L. Kodjikian, M. A. Pileri, F. Matonti, E. Souied, B. V. Stanzel, P. Szurman, M. Weber, K. U. Bartz-Schmidt, N. Eter, M. N. Delyfer, J. F. Girmens, K. A. v. Overdam, A. Wolf, R. Hornig, M. Corazzol, F. Brodie, L. O. d. Koo, D. Palanker, and J.-A. Sahel, "Subretinal Photovoltaic Implant to Restore Vision in Geographic Atrophy Due to AMD," *New England Journal of Medicine*, vol. 394, pp. 232–242, Oct. 2025.

[2] E. Zrenner, K. U. Bartz-Schmidt, H. Benav, D. Besch, A. Bruckmann, V. P. Gabel, F. Gekeler, U. Greppmaier, A. Harscher, S. Kibbel, J. Koch, A. Kusnyerik, T. Peters, K. Stingl, H. Sachs, A. Stett, P. Szurman, B. Wilhelm, and R. Wilke, "Subretinal electronic chips allow blind patients to read letters and combine them to words," *Proc Biol Sci*, vol. 278, pp. 1489–1497, May 2011.

[3] L. da Cruz, B. F. Coley, J. D. Dorn, F. Merlini, E. Filley, P. Christopher, F. K. Chen, V. Wuyyuru, J. Sahel, P. Stanga, M. Humayun, R. J. Greenberg, G. Dagnelie, and Argus II Study Group, "The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss," *British Journal of Ophthalmology*, vol. 97, pp. 632–636, May 2013.

[4] M. Beyeler, A. Rokem, G. M. Boynton, and I. Fine, "Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies," *J Neural Eng*, vol. 14, p. 051003, June 2017.

[5] Y. Hou, D. Nanduri, J. Granley, J. D. Weiland, and M. Beyeler, "Axonal stimulation affects the linear summation of single-point perception in three Argus II users," *Journal of Neural Engineering*, vol. 21, p. 026031, Apr. 2024.

[6] B. Christie, R. Sadeghi, A. Kartha, A. Caspi, F. V. Tenore, R. L. Klatzky, G. Dagnelie, and S. Billings, "Sequential epiretinal stimulation improves discrimination in simple shape discrimination tasks only," *Journal of Neural Engineering*, vol. 19, p. 036033, June 2022.

[7] A. Horsager, G. M. Boynton, R. J. Greenberg, and I. Fine, "Temporal interactions during paired-electrode stimulation in two retinal prosthesis subjects," *Invest Ophthalmol Vis Sci*, vol. 52, pp. 549–57, Jan. 2011.

[8] A. Pérez Fornos, J. Sommerhalder, L. da Cruz, J. A. Sahel, S. Mohand-Said, F. Hafezi, and M. Pelizzone, "Temporal Properties of Visual Perception on Electrical Stimulation of the Retina," *Investigative Ophthalmology & Visual Science*, vol. 53, no. 6, pp. 2720–2731, 2012.

[9] A. P. Fornos, J. Sommerhalder, and M. Pelizzone, "Reading with a simulated 60-channel implant," *Frontiers in Neuroscience*, vol. 5, p. 57, 2011.

[10] M. Vurro, A. M. Crowell, and J. S. Pezaris, "Simulation of thalamic prosthetic vision: reading accuracy, speed, and acuity in sighted humans," *Frontiers in Human Neuroscience*, vol. 8, 2014.

[11] N. Paraskevoudi and J. S. Pezaris, "Full gaze contingency provides better reading performance than head steering alone in a simulation of prosthetic vision," *Scientific Reports*, vol. 11, p. 11121, May 2021. Number: 1.

[12] C. D. Wickens, W. S. Helton, J. G. Hollands, and S. Banbury, *Engineering Psychology and Human Performance*. Routledge, 5th ed., 2021.

[13] P. M. Garvey, M. T. Pietrucha, and D. Meeker, "Development of a new guide sign alphabet," *Ergonomics in Design*, vol. 6, no. 3, pp. 7–11, 1998.

[14] M. Beyeler, D. Nanduri, J. D. Weiland, A. Rokem, G. M. Boynton, and I. Fine, "A model of ganglion axon pathways accounts for percepts elicited by retinal implants," *Scientific Reports*, vol. 9, no. 1, p. 9199, 2019.

[15] N. C. Sinclair, M. N. Shivdasani, T. Perera, L. N. Gillespie, H. J. McDermott, L. N. Ayton, and P. J. Blamey, "The Appearance of Phosphenes Elicited Using a Suprachoroidal Retinal Prosthesis," *Investigative Ophthalmology & Visual Science*, vol. 57, pp. 4948–4961, Sept. 2016.

[16] M. Beyeler, G. M. Boynton, I. Fine, and A. Rokem, "pulse2percept: A python-based simulation framework for bionic vision," in *Proceedings of the 16th Python in Science Conference (SciPy 2017)*, pp. 81–88, 2017.

[17] J. Granley and M. Beyeler, "A Computational Model of Phosphene Appearance for Epiretinal Prostheses," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4477–4481, Nov. 2021.

[18] N. Han, S. Srivastava, A. Xu, D. Klein, and M. Beyeler, "Deep learning–based scene simplification for bionic vision," in *Augmented Humans International Conference 2021 (AHs '21)*, pp. 45–54, ACM, 2021.

[19] Y. Hou, L. Pullela, J. Su, S. Aluru, S. Sista, X. Lu, and M. Beyeler, "Predicting the Temporal Dynamics of Prosthetic Vision," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4, July 2024.

[20] J. M. Kasowski, A. Varshney, R. Sadeghi, and M. Beyeler, "Simulated prosthetic vision confirms checkerboard as an effective raster pattern for epiretinal implants," *Journal of Neural Engineering*, vol. 22, no. 4, p. 046017, 2025.

[21] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations (ICLR)*, 2018.

[22] Wikimedia Foundation, "Wikimedia downloads: Wikipedia database dumps," 2023.

[23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314–1324, IEEE, October 2019.

[24] J. Skaza, S. Murlidaran, A. Varshney, Z. Wen, W. Wang, M. P. Eckstein, and M. Beyeler, "A Deep Learning Framework for Predicting Functional Visual Performance in Bionic Eye Users," June 2025. Pages: 2025.06.23.660990 Section: New Results.

[25] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 23, no. 1, pp. 90–93, 1974.

[26] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[27] K. E. K. Rassia and J. S. Pezaris, "Improvement in reading performance through training with simulated thalamic visual prostheses," *Scientific Reports*, vol. 8, p. 16310, Nov. 2018. Number: 1.

[28] J. Granley, T. Fauvel, M. Chalk, and M. Beyeler, "Human-in-the-loop optimization for deep stimulus encoding in visual prostheses," in *Advances in Neural Information Processing Systems* (A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds.), vol. 36, pp. 79376–79398, 2023.