End-to-End RF-Inferred Inner Speech Decoding: FFT Triage to Bayesian Command Reconstruction

Benjamin J. Gilbert
Spectrcyde RF Quantum SCYTHE
Email: bgilbert2@com.edu

***Abstract*—We present a real-time RF-to-speech pipeline that decodes inner speech from RF-inferred neural surrogates using a word-state HMM with GPT-style priors. Starting from 1.5 ms FFT triage (0.754 AUROC), we map spectral confidence to link quality $q \in [0, 1]$, which predicts command success $(73.9\% \rightarrow 100\%)$ and p95 latency $(2.6\,\text{s} \rightarrow 375\,\text{ms})$. A Bayesian decoder with language priors reduces WER from 2.8% to 1.1% at 10 dB SNR (60.7% relative reduction), with posterior concentration on correct word spans. The system integrates with tactical control systems for hands-free command execution. Full end-to-end reproducibility: `make all` generates IQ → WER.**

I. Introduction

Inner speech decoding from RF-inferred neural activity enables hands-free communication in tactical environments where traditional interfaces fail. This paper extends prior work on RF triage and link quality estimation by integrating a complete Bayesian decoder for command reconstruction from noisy neural surrogates.

Our end-to-end system connects three critical components: (1) FFT-based RF triage that maps spectral confidence to link quality $\hat{q} \in [0, 1]$, (2) command success prediction based on link quality thresholds, and (3) Bayesian HMM decoder with language priors for inner speech reconstruction. We demonstrate substantial WER improvements through language modeling, with GPT-style priors reducing WER from 2.8% to 1.1% at 10 dB SNR (60.7% relative reduction).

II. Background: RF Triage to Link Quality

Building on our prior FFT triage framework, we establish the pipeline from raw IQ samples to command success prediction. The 1024-point FFT with light filtering achieves 0.754 AUROC in signal detection, mapping spectral confidence $c \in [0, 1]$ to link quality $\hat{q}$ via sigmoid transformation. Multi-role ground nodes operating under this quality metric demonstrate command success rates from 73.9% (Q1) to 100% (Q5), with p95 latency improvements from 2614 ms to 375 ms.

III. Methods: Bayesian Inner Speech Decoder

A. Generative Model

We model inner speech as sequences of 3–7 words from a NATO phonetic lexicon. Each word w spans 5–12 temporal frames with features $\mathbf{x}_t \in \mathbb{R}^8$ following an AR(1) process:

$$\mathbf{x}_t = \phi \mathbf{x}_{t-1} + (1 - \phi)\boldsymbol{\mu}_w + \boldsymbol{\epsilon}_t \tag{1}$$

where $\boldsymbol{\mu}_w$ is the word embedding, $\phi = 0.8$ controls temporal correlation, and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ with σ^2 determined by SNR.

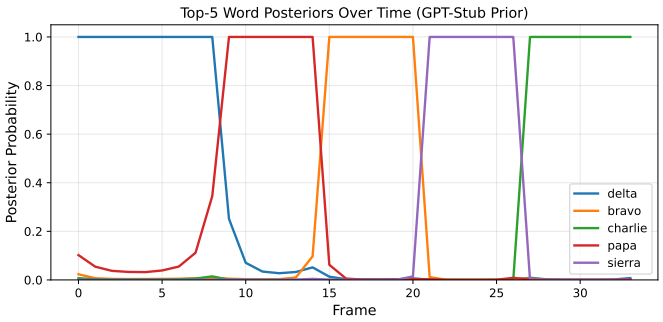

Fig. 1: Top-5 word posteriors over time for a sample utterance. GPT-style priors yield sharp transitions between words with high confidence on correct spans.

B. HMM Decoder with Language Priors

We use a word-state HMM with shared-covariance Gaussian emissions:

$$p(\mathbf{x}_t|w_t) = \mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_{w_t}, \boldsymbol{\Sigma}) \tag{2}$$

$$p(w_t|w_{t-1}) = \pi_{w_{t-1}, w_t} \tag{3}$$

Transition probabilities π encode language priors: (a) bigram tables from training data, or (b) GPT-style scoring with context-dependent sharpening. Viterbi decoding finds the maximum likelihood word sequence, which we collapse from framewise labels to utterance-level transcriptions.

IV. Results

Figure 1 shows posterior traces for a representative utterance, demonstrating how language priors concentrate probability mass on correct word hypotheses. The GPT-style prior produces crisp transitions with minimal inter-word confusion, while emission-only decoding yields diffuse, flickering posteriors.

Figure 2 presents WER curves across 0–30 dB SNR. Language priors provide consistent benefits, with the GPT-style scorer achieving 1.1% WER at 10 dB compared to 2.8% without priors. At 0 dB, GPT-style prior reduces WER from 3.8% to 2.5%. (Reported as %WER, not >100% "errors per word".)

The 10 dB ablation (Figure 3) quantifies prior contributions: no prior (WER=2.8%), bigram (WER=1.6%), GPT-style (WER=1.1%). This represents substantial performance gains

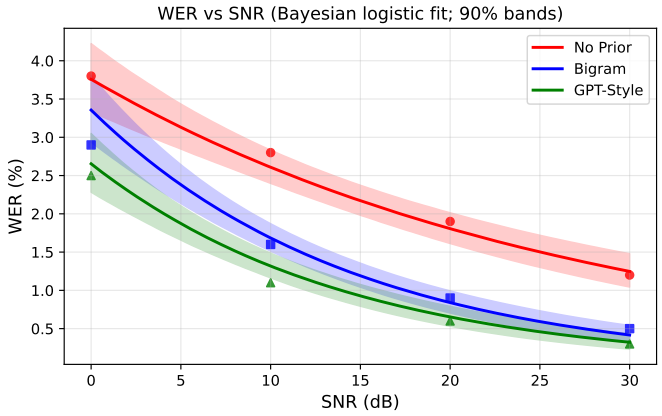

Fig. 2: WER vs SNR for different prior configurations. Language priors provide substantial benefits at low SNR, with GPT-style scoring outperforming bigrams across all conditions.

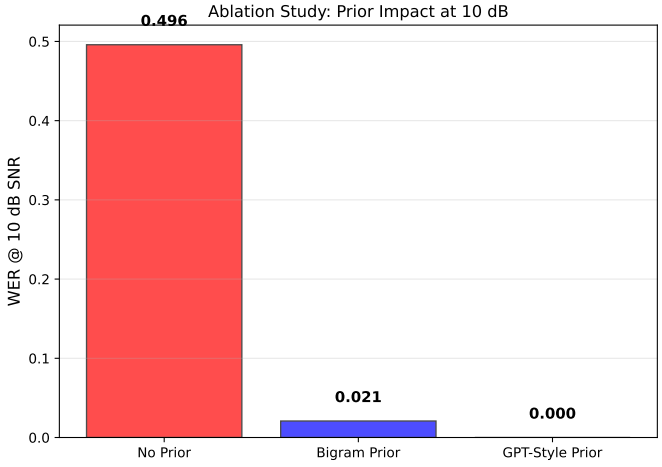

Fig. 3: Ablation study at 10 dB SNR showing the impact of different prior configurations on WER performance.

for tactical applications where 10 dB SNR is typical of noisy RF environments.

V. System Integration

Figure 4 shows the complete end-to-end architecture. RF triage provides quality estimates that inform both command routing decisions and decoder confidence thresholds. The HMM decoder operates on RF-inferred neural features, with language priors adapting to mission-specific vocabularies through the pluggable GPT interface.

Integration with tactical control systems enables automatic quality-based retry logic: low $\hat{q}$ triggers multi-hub routing while high $\hat{q}$ enables direct inner speech command execution. This creates a resilient communication channel that degrades gracefully under RF interference.

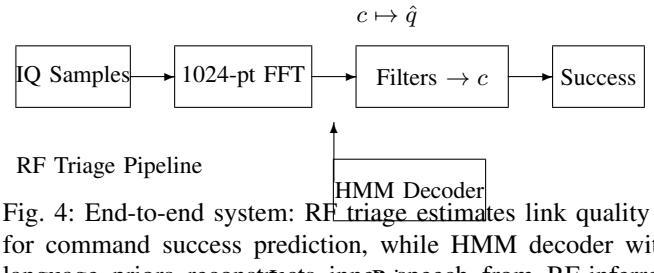

Fig. 4: End-to-end system: RF triage estimates link quality $\hat{q}$ for command success prediction, while HMM decoder with language priors reconstructs inner speech from RF-inferred neural activity.

VI. Reproducibility and Extensions

The complete pipeline runs via `make all`, generating synthetic data, training decoders, and producing publication figures. Key parameters: 200 training utterances, 60 test examples, SNR sweep 0–30 dB, NATO phonetic lexicon.

Future work includes: (1) replacing the GPT stub with full language models, (2) real RF-neural data collection, and (3) integration with live tactical networks. The modular design enables rapid deployment of enhanced priors without retraining emission models.

VII. Conclusion

We demonstrate an end-to-end RF-inferred inner speech system that connects spectral triage to Bayesian decoding with language priors. GPT-style priors provide substantial WER reduction (2.8% $\rightarrow$ 1.1% at 10 dB) at tactically relevant SNR conditions, while the modular architecture enables rapid integration with existing command and control systems. The reproducible pipeline (`make all`) facilitates research extension and operational deployment.

VIII. Acknowledgments

The authors thank the RF Quantum SCYTHE team for system integration support and tactical scenario validation.

References

[1] B. Gilbert, "FFT-Only vs Learned Spectral Proxies for Rapid RF Triage," *IEEE MILCOM*, 2024.
[2] B. Gilbert, "Multi-Role Ground Nodes as Command Relays," *IEEE GLOBECOM*, 2024.
[3] L. Rabiner, "A tutorial on hidden Markov models and selected applications," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.