

PROSODI: Paralinguistic Radar Observation for Semantic Operational State Detection and Intent Characterization — Adapting Multimodal Vocal Emotion LLMs for Cognitive RF Sensing

Sponsor Information:

Thomas E. Woodward, AFRL/RFE; thomas.woodward.9@us.af.mil

AFRL/RF Bldg. 620, Room 202,

2241 Avionics Circle

WPAFB, OH 45433-7333

Academic Area/Field and Education Level:

- Electrical and Computer Engineering / Digital Signal Processing, Artificial Intelligence, and Cognitive Radar (MS or PhD Level)

Objectives:

- Objective 1 (RF-to-Speech Alignment): Design and evaluate a physics-aware RF tokenization architecture (e.g., Vector Quantized Variational Autoencoder - VQ-VAE) that maps raw Radio Frequency (RF) in-phase and quadrature (IQ) data or Pulse Descriptor Word (PDW) sequences into a latent representation space aligned with pre-trained vocal speech large language models (LLMs) (such as SenseVoice, Whisper, or Wav2Vec2).
- Objective 2 (Semantic Intent and Sentiment Characterization): Investigate the zero-shot and parameter-efficient fine-tuning (PEFT/LoRA) capabilities of speech-oriented multimodal LLMs to decode the operational "tone" of a radar, specifically identifying transitions between search, acquisition, track, and track-while-scan states.
- Objective 3 (Physics-Bounded Interactive Dialogue Modeling): Formulate and model the dynamic, closed-loop interaction between a radar (probe) and a friendly platform (return) as an interactive "physical dialogue" or conversational exchange to predict future operational state transitions.

Description:

- Background and Motivation:
Modern radar systems operate in highly congested electromagnetic environments.

Traditional electronic support measures (ESM) identify RF emitters by comparing observed RF pulse parameters (e.g., Pulse Width, Pulse Repetition Interval, Carrier Frequency) against static, pre-defined libraries. However, modern cognitive radars dynamically adapt their waveforms in real-time, easily circumventing static rule-based ESM systems.

- This research proposes a paradigm shift: treating the electromagnetic emissions of adversary radars as a structured physical language and the EW engagement as a conversation. Just as human speakers convey emotional state, urgency, and underlying intent through paralinguistic acoustic cues—such as pitch, tempo, volume, and conversational turn-taking—cognitive radar platforms convey their tactical intent through analogous RF parameters.
- The Vocal-to-Radar Analogy:
To bridge the gap between human vocal communication and RF sensing, this research relies on key scientific analogies:

Human Vocal Feature	Analogous Radar Parameter	Tactical EW Interpretation (RF State / Intent)
Tone / Pitch Contour	Pulse Repetition Frequency (PRF) and Carrier Frequency Modulation	Search to Track Transition: Shifts from low, steady PRF (search) to high, rapid PRF/modulation (track lock).
Volume / Energy Contours	EIRP and Mainlobe Beam Dwell Time	Target Focus: Long dwell and high EIRP on a target (direct interrogation) vs. low sidelobes (clandestine search).
Speech Rate / Tempo	Beam Scan Rate and Pulse Width	Urgency of Engagement: Fast sector scan rate represents tactical urgency.
Turn-Taking	Interlocking Pulses and Radar Returns	Interactive RF Dialogue: Active feedback loops of probing and returning signals.
Arousal / Sentiment	RF Operational State Transitions	Platform Aggression: Tracking transition from benign search to precise track.

Technical Approach and Methodology:

The graduate student will approach this problem through a structured, three-phase plan:

1. Phase 1: RF Data Set Excursion: The student will be provided with sample RF datasets that contain a variety of radar and communications emitters/signals. Using this data as a template, the student will generate derivative data sets as needed to create a training set of multi-state radar-target RF interactions, labeled with operational ground-truth states.
2. Phase 2: RF-Acoustic Alignment (Tokenization): The student will develop an "RF Tokenizer" (such as a VQ-VAE or a contrastive encoder) to discretize continuous RF signals into "RF phonemes." This tokenizer will align RF features with the acoustic latent spaces of foundational speech models.
3. Phase 3: Fine-Tuning Vocal LLMs for Sensing: The student will utilize Parameter-Efficient Fine-Tuning (PEFT), such as LoRA/QLoRA, to fine-tune open-source multimodal audio-text models (e.g., *SenseVoice* or *Qwen-Audio*) on the aligned RF tokens to output real-time natural language intent reports and reaction-time estimations.

Key Challenges & Mitigations:

1. Temporal Latency: LLM inference is relatively slow compared to microsecond RF pulses. *Mitigation:* Frame the LLM as a high-level "cognitive intelligence aid" for tactical dashboards (operating at 100ms scale) to provide pilot/operator intent awareness rather than high-speed inline signal processing.
2. Sim-to-Real Domain Gap: *Mitigation:* Domain randomization during simulated data generation (introducing phase noise, multipath clutter, Doppler spreads) and few-shot calibration using Software Defined Radios (SDRs).
3. Adversarial Robustness: *Mitigation:* Adversarial contrastive training and physics-based spectrum constraints to filter out physically impossible RF inputs.

Research Questions (RQs):

Question ID	Graduate Research Question	Key Investigation Area
RQ 1	How can raw RF IQ data and PDW sequences be tokenized into discrete "phonemes" mapped to vocal LLM latent spaces?	RF Tokenization & Vocab Mapping
RQ 2	Can audio-text LLMs pre-trained on vocal emotion recognition generalize to RF signals with zero-shot or PEFT?	Cross-Modal Transfer Learning
RQ 3	How do we model the closed-loop RF interaction (probe and return) as a game-theoretic dialogue to predict state transitions?	Physics-Bounded Dialogue Modeling
RQ 4	What is the impact of multipath clutter, noise, and jamming on the LLM's intent classification robustness?	Signal Robustness & Security

Research Classification/Restrictions:

- Unclassified / Distribution A: Approved for public release. To maintain compliance, the academic research will exclusively use unclassified radar standards (e.g., marine radars, civil aviation transponders, open-source pulsed-Doppler models). The developed models will be designed to allow seamless, air-gapped transfer learning to classified military threat libraries in secure environments.

Eligible Research Institutions:

- Open to all member institutions of the Defense Associated Graduate Student Innovator (DAGSI) and accredited research universities in the State of Ohio (including the Air Force Institute of Technology (AFIT), Wright State University, University of Dayton, The Ohio State University, University of Cincinnati, Case Western Reserve University, Ohio University, and the University of Toledo).