TRIDENT: A Redundant Architecture for Caribbean-Accented Emergency Speech Triage

Galbraith, E., Sutherland, C., and Morgan, D.

SMG Labs Research Group

December 12, 2025

Abstract

Emergency speech recognition systems exhibit systematic performance degradation on non-standard English varieties, creating a critical gap in services for Caribbean populations. We present TRIDENT (**T**ranscription and **R**outing **I**ntelligence for **D**ispatcher-**E**mpowered **N**ational **T**riage), a three-layer dispatcher-support architecture designed to structure emergency call inputs for human application of established triage protocols (the ESI for routine operations and START for mass casualty events), even when automatic speech recognition fails.

The system combines Caribbean-accent-tuned ASR, local entity extraction via large language models, and bio-acoustic distress detection to provide dispatchers with three complementary signals: transcription confidence, structured clinical entities, and vocal stress indicators. **Our key insight is that low ASR confidence, rather than representing system failure, serves as a valuable queue prioritization signal—particularly when combined with elevated vocal distress markers indicating a caller in crisis whose speech may have shifted toward basilectal registers.** A complementary insight drives the entity extraction layer: trained responders and composed bystanders may report life-threatening emergencies without elevated vocal stress, requiring semantic analysis to capture clinical indicators that paralinguistic features miss.

We describe the architectural design, theoretical grounding in psycholinguistic research on stress-induced code-switching, and deployment considerations for offline operation during disaster scenarios. This work establishes a framework for accent-resilient emergency AI that ensures Caribbean voices receive equitable access to established national triage protocols. Empirical validation on Caribbean emergency calls remains future work.

Keywords: speech recognition, Caribbean English, emergency dispatch, vocal stress, triage

1 Introduction

When a caller dials emergency services during a crisis, modern automatic speech recognition (ASR) systems exhibit well-documented performance disparities across demographic groups [10]. For Caribbean English speakers—a population of over 40 million—these disparities compound with a linguistic phenomenon: under acute stress, speakers tend to shift toward basilectal (more creole-heavy) speech registers, precisely the varieties on which ASR systems perform worst.

Caribbean health ministries have adopted internationally-validated triage protocols: the Emergency Severity Index (ESI) for routine operations and START (Simple Triage and Rapid Treatment) for mass casualty events. These protocols assume dispatchers can accurately capture caller information—an assumption that fails systematically when ASR systems cannot reliably transcribe Caribbean speech.

1.1 TRIDENT: Dispatcher-Empowered Architecture

This paper presents TRIDENT (Transcription and Routing Intelligence for Dispatcher-Empowered National Triage), designed to ensure Caribbean-accented callers receive equitable access to established triage protocols. Rather than attempting to eliminate ASR errors—an unrealistic goal—we build a **dispatcher-support system** that remains functional when transcription fails.

Our central contribution is a three-layer framework providing dispatchers with structured inputs for protocol application:

1. **Transcription confidence:** Flags unreliable transcripts so dispatchers know to listen directly to audio

2. **Structured entity extraction:** Extracts clinical indicators (location, mechanism, breathing status, vulnerable populations) even from degraded transcriptions

3. **Bio-acoustic distress detection:** Provides physiological stress markers independent of transcript content

1.2 Key Insights

Two complementary insights motivate this design:

1. **Content beyond voice:** Trained responders and composed bystanders may report life-threatening emergencies without elevated vocal stress. Semantic extraction captures information that paralinguistic features miss—ensuring "children trapped in burning building," spoken calmly, provides dispatchers with structured data for appropriate triage classification.

2. **Uncertainty as prioritization signal:** Low ASR confidence, rather than representing failure, serves as a queue prioritization indicator—particularly when combined with elevated vocal distress marking a caller in crisis whose speech may have shifted toward basilectal registers. This reframes accent-induced transcription errors from bugs into features correlating with genuine distress.

TRIDENT addresses critical gaps in existing emergency AI—cloud dependency with accent-agnostic ASR, text-only analysis ignoring paralinguistic signals, dialect blindness to stress-induced register shifting, and infrastructure fragility during disasters—while **respecting the clinical authority of established protocols**. The system structures inputs and prioritizes queues, but triage decisions remain with trained professionals applying Ministry of Health-mandated frameworks.

2 Related Work

TRIDENT's dispatcher-support architecture draws on research across five domains: ASR for Caribbean varieties, AI in emergency dispatch, vocal stress detection, dialect reversion under cognitive load, and edge computing for disaster resilience.

2.1 The Accent Gap in Automatic Speech Recognition

Modern ASR systems exhibit systematic performance degradation on non-standard English varieties. Koenecke et al. [10] evaluated five commercial ASR systems, finding word error rates averaged 0.35 for Black speakers compared to 0.19 for White speakers, with performance gaps traced to acoustic model limitations rather than language models.

Caribbean English remains especially underserved. Madden et al. [11] developed the first substantial Jamaican Patois corpus (42.58 hours) and derived scaling laws for Whisper performance. Pre-trained Whisper Large achieved 89% WER on Patois, while fine-tuned Whisper Medium reduced this to 30% WER. Critically, their scaling law (WER = $158.06 \times \mathrm{M}^{-0.255} \times \mathrm{D}^{-0.269}$) demonstrates that dataset increases yield greater gains than model scaling for underrepresented varieties, informing our choice of Whisper Medium with Caribbean-specific fine-tuning.

2.2 AI-Assisted Emergency Dispatch and Clinical Protocols

Emergency services worldwide are exploring AI to improve call handling, but these systems must support established clinical triage protocols rather than replace human judgment.

Clinical Triage Protocols. The Emergency Severity Index (ESI) is a five-level acuity scale (Level 1: immediate lifesaving intervention to Level 5: no resources needed) widely used in the United States and internationally [6]. Jamaica's Ministry of Health implemented ESI across all 19 public hospital emergency departments in 2016 [7]. For mass casualty events such as hurricanes, the START (Simple Triage and Rapid Treatment) protocol provides rapid four-category sorting: BLACK (deceased/expectant), RED (immediate), YELLOW (delayed), and GREEN (walking wounded). The ESI handbook explicitly notes that ESI should not be used during mass casualty incidents [6].

Current AI Systems. Existing emergency AI systems (e.g., ECA [1], Corti [2]) achieve promising classification accuracy but rely on cloud-dependent, accent-agnostic ASR and process only transcribed text, ignoring paralinguistic signals. A scoping review of 106 AI studies in prehospital care identified underutilization of multimodal inputs and absence of infrastructure-independent systems as key gaps [4].

Gaps for Caribbean Deployment. Three limitations motivate TRIDENT's design: (1) no accent adaptation for Caribbean varieties or stress-induced register shifting, (2) no integration of vocal stress detection with text classification, and (3) cloud dependency that fails during disasters when emergency services are most needed. TRIDENT addresses these gaps while maintaining the principle that AI should empower dispatchers to apply ESI/START protocols more effectively, not replace clinical judgment.

2.3 Vocal Stress Detection

The bio-acoustic layer builds on research establishing acoustic correlates of psychological stress. A systematic review of 38 studies found fundamental frequency (F0) as the most consistent stress marker, with 15 of 19 studies reporting significant mean F0 increases under stress [15].

Research on emergency communications provides direct validation. Van Puyvelde et al. [18] analyzed real-life emergency recordings including cockpit voice recorders and 911 calls, documenting F0 increases from 123.9 Hz to 200.1 Hz during life-threatening emergencies—a 62% increase. However, Deschamps-Berger et al. [5] found that while benchmark IEMOCAP data yielded 63% emotion recognition accuracy, real emergency calls achieved only 45.6%—a substantial domain shift. This finding reinforces our design decision to use bio-acoustic analysis as a triage signal routing high-distress calls to human dispatchers, rather than attempting fully automated classification.

2.4 Dialect Reversion Under Cognitive Load

Psycholinguistic research establishes that for Caribbean speakers navigating the creole continuum—from basilect (most creole features) through mesolect to acrolect (Standard English)—maintaining acrolectal speech requires sustained executive function. The inhibitory control model establishes that non-target languages remain continuously active and must be suppressed through cognitive

effort [9]. Under high cognitive load, this inhibition fails, causing speakers to revert toward their dominant variety.

Patrick's [12] sociolinguistic analysis of the Jamaican Creole continuum establishes that stress levels influence speakers' positioning on this spectrum, with most speakers being mesolectal under normal conditions but capable of shifting toward either pole. The implications for emergency services are significant: a professional who speaks Standard English at work may revert toward basilectal Patois when their house is flooding. Standard ASR systems will exhibit precisely the performance degradation documented in the accent gap literature at the moment when accurate recognition is most critical.

2.5 Edge Computing for Disaster Resilience

Infrastructure failure during disasters makes the case for offline-capable emergency AI. Hurricane Maria's impact on Puerto Rico saw 95% of cell towers fail, with the entire island losing power [14]. Communication infrastructure failure contributed to a disputed death toll ultimately estimated at approximately 3,000, with recovery requiring over 200 days for full power restoration.

Recent model compression advances make edge deployment feasible. Pre-positioned edge computing resources at hospitals, shelters, and emergency coordination centers, loaded with Caribbean-tuned models, could maintain triage capability even during complete grid and network failure.

2.6 Summary: Positioning Our Contribution

TRIDENT addresses four critical gaps in existing emergency dispatch AI for Caribbean deployment:

- **Caribbean-adapted ASR:** Fine-tuned Whisper models (informed by Madden et al.'s scaling laws) provide transcription accuracy for Caribbean speech varieties, enabling viable downstream entity extraction.

- **Multimodal distress detection:** Parallel bio-acoustic analysis provides a signal pathway that functions even when ASR fails, transforming low transcription confidence from a limitation into a queue prioritization feature.

- **Stress-aware design:** Accounts for stress-induced register shifting along the creole continuum—routing calls with elevated vocal distress and low ASR confidence to immediate human attention.

- **Offline operation:** Complete system deployment on edge hardware (Raspberry Pi 5) enables function during infrastructure failures when emergency services are most critical.

The result is the first dispatcher-support system designed specifically for Caribbean emergency services—not to make triage decisions, but to ensure Caribbean-accented callers receive equitable access to the ESI and START protocols that their health ministries have adopted. TRIDENT empowers dispatchers with better information and intelligent queue prioritization; clinical judgment remains with trained human professionals.

3 Theoretical Foundations

Fine-tuning Whisper on Caribbean speech improves transcription but cannot eliminate the accent gap. Madden et al. [11] achieved 30% WER on Jamaican Patois—dramatic improvement from 89% baseline, but still far above the <5% WER typical for standard English. Moreover,

fine-tuning on broadcast speech cannot capture emergency acoustics: elevated noise, emotional qualities, and stress-induced basilectal reversion. ASR alone will fail when needed most.

Conversely, bio-acoustic distress detection cannot provide semantic information needed for dispatch. A caller may exhibit extreme vocal stress while saying "my house is on fire" or "I lost my keys"—identical distress signals but dramatically different responses. Furthermore, Deschamps-Berger et al. [5] found laboratory emotion recognition accuracy (63%) drops substantially in real emergency calls (45.6%). Bio-acoustic features provide gradient information about caller state but cannot substitute for semantic content.

3.1 The Integration Thesis

Our architecture integrates these complementary information sources based on the following thesis: **In emergency contexts, the correlation between ASR failure and genuine distress creates an opportunity to use recognition uncertainty as a routing signal rather than an error to be minimized.**
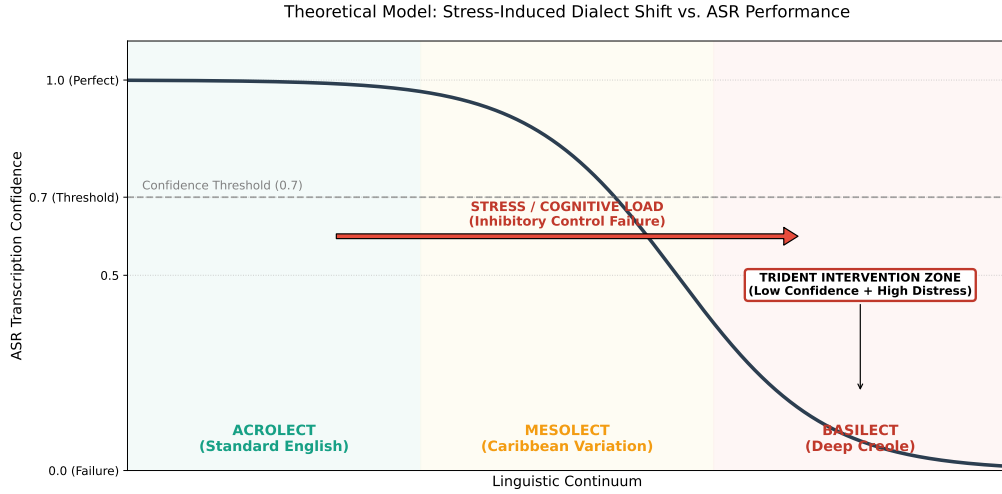

Figure 1: The TRIDENT Integration Thesis: Stress-Induced Dialect Shift vs. ASR Performance. The model illustrates the system's theoretical foundation. Under acute stress (red arrow), speakers experience inhibitory control failure, shifting along the continuum from acrolectal (standard) to basilectal (creole) registers. As speech becomes more basilectal, ASR confidence (blue line) degrades below the usable threshold of 0.7. The "Intervention Zone" highlights TRIDENT's novel contribution: identifying calls where low transcription confidence coincides with high bio-acoustic distress, thereby converting a technical failure into a high-priority (Q1) routing signal.

This thesis rests on the psycholinguistic literature establishing that:

1. Stress triggers cognitive load effects that impair executive function [8]

2. Impaired executive function leads to reduced inhibition of dominant language varieties [9]

3. For Caribbean speakers, dominant varieties include basilectal forms underrepresented in ASR training [12, 11]

4. Stress simultaneously elevates bio-acoustic markers (F0, intensity) that can be detected independently of speech content [18]

The logical conclusion: when ASR confidence drops and bio-acoustic distress rises, the system has detected a caller in genuine crisis whose speech has shifted beyond standard recognition capabilities. This combination should trigger immediate human review—not because the system has failed, but because it has successfully identified a caller who needs human attention most.

4 System Architecture

TRIDENT implements a three-layer dispatcher-support architecture where each component provides independent value while contributing to intelligent queue prioritization. The system does not make clinical triage decisions—those remain with trained dispatchers applying ESI or START protocols—but ensures dispatchers receive the highest-priority calls first along with structured information to support protocol application. Figure 2 illustrates the system flow.
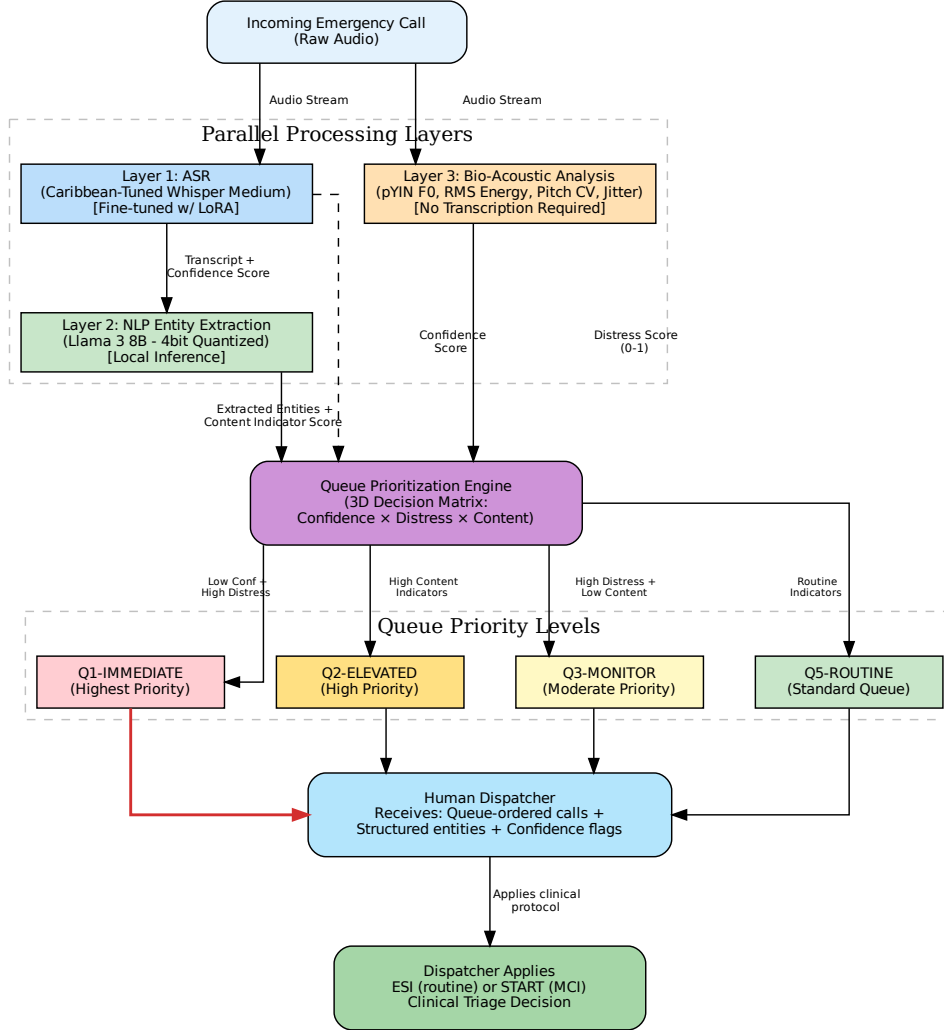

Figure 2: The TRIDENT architecture. The system processes raw audio through two parallel streams: (Left) A Caribbean-adapted ASR and NLP pipeline for entity extraction and content analysis, and (Right) a bio-acoustic analysis layer for detecting physiological distress markers. The **Queue Prioritization Engine** integrates three independent signals—transcription confidence, extracted clinical indicators, and vocal distress—to determine queue position for dispatcher attention. This ensures that (1) calls with low transcription confidence but high vocal distress receive immediate human review, and (2) semantically urgent calls from calm reporters are not delayed due to absent vocal stress markers. The dispatcher then applies established triage protocols (ESI for routine operations, START for mass casualty events) using both TRIDENT's extracted entities and direct audio review.

4.1 Design Philosophy: Enabling Protocol Application

TRIDENT's architecture reflects a core principle: **AI should empower dispatchers to apply established protocols more effectively, not replace clinical judgment**. Caribbean health ministries have adopted validated triage frameworks, ESI for emergency departments, START for mass casualty incidents, that represent decades of clinical refinement. TRIDENT's role is to solve the *input problem*: ensuring these protocols can be applied equitably to Caribbean-accented callers whose speech current ASR systems fail to transcribe accurately.

Each architectural layer addresses a specific input challenge:

- **Layer 1 (ASR):** Produces transcripts and confidence scores, enabling dispatchers to know when to trust text versus listen directly to audio.

- **Layer 2 (NLP):** Extracts structured clinical entities—location, mechanism of injury, breathing status, vulnerable populations—that map directly to ESI/START decision points.

- **Layer 3 (Bio-Acoustic):** Detects physiological distress markers that indicate caller crisis state, providing a signal not currently captured by standard protocols but valuable for queue prioritization.

The following subsections detail each layer's implementation.

4.2 Layer 1: Caribbean-Tuned ASR

The ASR layer employs OpenAI's Whisper Medium fine-tuned with Low-Rank Adaptation (LoRA) on Caribbean broadcast speech. We selected Whisper Medium over Large based on Madden et al.'s [11] scaling law, which demonstrates that domain-specific data yields greater gains than model size for Caribbean varieties. Whisper Medium is also more efficient for Raspberry Pi 5 edge deployment.

Fine-tuning Configuration:

- Base model: openai/whisper-medium

- Adaptation: LoRA (rank=16, alpha=32)

- Training data: BBC Caribbean broadcast corpus (∼28,000 clips)

- Trainable parameters: ∼0.5% of total model

Confidence Scoring: The system computes utterance-level confidence as the mean log-probability across all decoded tokens, normalized to 0-1:

$$\text{confidence} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log P(t_i|t_1\ldots t_{i-1}, \text{audio})\right) \tag{1}$$

We use utterance-level rather than token-level confidence because emergency triage requires holistic assessment of transcription reliability. The low confidence threshold is set at 0.7 based on initial calibration.

4.3 Layer 2: Local NLP Entity Extraction

When ASR produces usable transcription (confidence ≥ 0.7), the NLP layer extracts structured emergency information using Llama 3 8B running locally via Ollama. The extraction schema targets entity types that map directly to ESI and START triage protocol decision points.

4.3.1 Entity Extraction Schema

The schema targets four entity categories:

- **LOCATION:** Street addresses, landmarks, geographic references

- **MECHANISM/HAZARD:** Emergency type (fire, flood, medical, violence, traffic)

- **CLINICAL INDICATORS:** Breathing status, consciousness, bleeding, mobility

- **SCALE:** Number of people involved, vulnerable populations

4.3.2 Mapping to Triage Protocols

TRIDENT entities support ESI and START protocol application. For ESI, extracted entities inform the four decision points: Point A (lifesaving intervention) captures "not breathing," "choking," "unresponsive"; Point B (high-risk situation) captures mechanism of injury and altered status; Point C (resource needs) uses hazard type and complexity; Point D (vital signs) uses reported vitals and distress indicators [6].

For mass casualty events using START, entities support rapid sorting: GREEN captures "walking," "minor injuries"; YELLOW captures "injured but stable," "conscious"; RED captures "trapped," "not breathing," "heavy bleeding"; BLACK captures cessation indicators.

Protocol	Decision Point	Example Extraction Target
ESI Level 1	Immediate lifesaving intervention?	"not breathing," "choking," "heavy bleeding," "unresponsive"
START RED	Not walking, breathing issues	"trapped," "not breathing," "unresponsive," "heavy bleeding"

Table 1: Example entity extraction targets supporting ESI and START protocols. Full protocol mappings detailed in extended version.

4.3.3 Handling Garbled Input

The NLP layer handles low-quality transcriptions through confidence-aware prompting. When ASR confidence is below 0.7, the system instructs the LLM to mark uncertain extractions, avoid hallucination, prioritize location extraction, and note phonetically similar alternatives. When confidence is very low (<0.4), minimal structured output is produced and the call is flagged for immediate human review.

4.3.4 Content Indicator Scoring

The NLP layer computes a **Content Indicator Score** ($S_c \in [0, 100]$) quantifying urgency implied by semantic content, independent of how the caller sounds. This addresses a critical gap: a trained first responder may report a mass casualty event calmly, producing low bio-acoustic distress despite extremely urgent content. Without content analysis, such calls would be deprioritized.

Rather than keyword matching, we leverage the LLM's semantic understanding to classify transcript content. This approach handles Caribbean creole variants ("mi granmodda drop dung an she nah move" conveys the same urgency as "my grandmother collapsed and she's not moving"), negation, and indirect references.

The LLM outputs structured classifications:

```
{
  "hazard_category": "violent_crime" | "medical" | "fire" |
                     "flood" | "traffic" | "infrastructure" | "other",
  "life_threat_level": "imminent" | "potential" | "none",
  "vulnerable_population": true | false,
  "situation_status": "escalating" | "stable" | "resolved",
  "persons_affected": <integer>
}
```

A deterministic function maps classifications to the score:

$$S_c = \min\left(100,\ S_{\text{hazard}} + S_{\text{threat}} + S_{\text{vuln}} + S_{\text{scale}}\right) \tag{2}$$

Scoring components: Hazard category weights range from 30 (violent crime) to 5 (other). Life-threat level contributes +30 (imminent), +15 (potential), or +0 (none). Vulnerable population adds +15. Scale combines persons affected (+5 per person, capped at +20) and escalation status (+10 if escalating).

Example calculations:

Transcript	Classification	S_c
"Pothole on Nelson Street"	infrastructure, none, false, stable, 0	10
"House fire, spreading to neighbor's yard"	fire, potential, false, escalating, 0	50
"Pickney dem trap inna di fire"	fire, imminent, true, stable, 2+	80

Table 2: Content indicator scoring via LLM classification. Semantic understanding captures urgency from Caribbean creole variants. High scores elevate queue priority; clinical triage remains with dispatchers.

The Content Indicator Score feeds into queue prioritization (Section 4.6), ensuring semantically urgent calls reach dispatchers promptly even when vocal distress markers are absent. Weights are tunable parameters that should be calibrated with local emergency services to reflect institutional priorities and regional hazard profiles.

4.4 Layer 3: Bio-Acoustic Distress Detection

The bio-acoustic layer operates on raw audio, independent of ASR success, extracting features correlated with psychological distress. Based on the vocal stress literature [15, 18, 19], we focus on features that capture physiological arousal through vocal production changes.

4.4.1 Feature Extraction

Using librosa, we extract the following acoustic features:

1. **Fundamental Frequency (F0):** Mean pitch extracted via autocorrelation method

   - Typical baseline: 85–180 Hz (male), 165–255 Hz (female) [16]
   - Stress indicator: Elevation above speaker baseline

2. **F0 Coefficient of Variation (CV):** Pitch instability measure

   - Computed as $CV = \sigma_{F0}/\mu_{F0}$
   - Normalizes for baseline differences across speakers
   - Stress indicator: $CV > 0.3$ suggests vocal instability

3. **Energy (RMS amplitude):** Mean intensity across utterance

   - Normalized to 0–1 scale relative to recording gain
   - Stress indicator: Elevated intensity during distress vocalizations

4. **Jitter:** Cycle-to-cycle variation in F0 period

   - Relatively independent of prosodic patterns [18]
   - Pathology threshold: $>1.04\%$ [3]

4.4.2 Distress Score Calculation

The distress score combines multiple acoustic indicators into a composite metric. We weight features according to their documented reliability and sex-independence:

$$D = w_{\text{pitch}} \cdot P + w_{\text{var}} \cdot V + w_{\text{energy}} \cdot E + w_{\text{jitter}} \cdot J \tag{3}$$

where:
The pitch elevation component now uses sex-adaptive parameters:

$$P = \min\left(1.0, \max\left(0, \frac{\bar{F}_0 - B}{R}\right)\right) \quad \text{(pitch elevation)} \tag{4}$$

where (B, R) adapts based on estimated speaker sex:

$$(B, R) = \begin{cases} (120, 80) & \text{if } \bar{F}_0^{(\text{init})} < 165 \text{ Hz (estimated male)} \\ (200, 100) & \text{otherwise (estimated female)} \end{cases} \tag{5}$$

The baseline B and range R parameters adapt based on a heuristic sex estimation from the initial 3 seconds of speech. A male speaker at 170 Hz (stressed) now contributes $P = (170 - 120)/80 = 0.625$ rather than the previous formulation's 0.0, addressing the male pitch penalty.
The remaining components are:

$$V = \min\left(1.0, \frac{CV_{F0}}{0.5}\right) \qquad \text{(pitch instability)} \tag{6}$$

$$E = \min\left(1.0, \frac{\bar{E}}{0.1}\right) \qquad \text{(energy)} \tag{7}$$

$$J = \min\left(1.0, \frac{\text{jitter}}{0.02}\right) \qquad \text{(perturbation)} \tag{8}$$

The weights reflect relative reliability from the literature:

- $w_{\text{pitch}} = 0.30$ — F0 elevation is the most consistent stress marker but is sex-dependent

- $w_{\text{var}} = 0.35$ — F0 coefficient of variation is sex-normalized and robust

- $w_{\text{energy}} = 0.20$ — intensity elevation accompanies distress

- $w_{\text{jitter}} = 0.15$ — perturbation measures are prosody-independent

4.4.3 Threshold Classification

- **High Distress:** $D > 0.5$

- **Low Distress:** $D \leq 0.5$

These thresholds are calibrated against Van Puyvelde et al.'s [18] findings on vocal markers in emergency versus baseline speech.

Note on sex differences: The distress score prioritizes sex-normalized features (CV, jitter) over absolute F0 elevation to mitigate the substantial baseline differences between male (85–175 Hz) and female (165–270 Hz) speakers. See Section 6.1 for detailed discussion of remaining bias risks.

4.5 The Complementarity Principle

The theoretical foundation for our multi-layer design rests on what we term the **Complementarity Principle**: the three signal dimensions capture distinct failure modes and urgency indicators that compensate for each other's blind spots, ensuring dispatchers receive the most critical calls first regardless of which individual signal might fail.

Dimension 1: Transcription Confidence. The conditions that degrade ASR performance (high stress, code-switching to basilect, environmental noise) are precisely the conditions that often accompany genuine emergencies. Low confidence is not merely a technical limitation to be hidden—it correlates with caller distress and should elevate queue priority while flagging the call for direct audio review.

Dimension 2: Content Indicators. Semantic analysis of transcript content captures urgency that vocal characteristics may miss. Trained professionals, repeat callers, and composed bystanders often report critical emergencies without elevated vocal stress—their calm delivery masks the urgency that only content analysis reveals. When transcription confidence is high, extracted entities map directly to ESI/START decision points.

Dimension 3: Bio-Acoustic Distress. Vocal stress markers (elevated pitch, intensity, instability) provide a parallel assessment channel that operates on raw audio, independent of transcription success. A caller whose speech is entirely unintelligible to ASR will still produce detectable distress signals. This dimension captures information not currently used by ESI or START protocols, representing TRIDENT's novel contribution to dispatcher awareness.

This creates a robust prioritization space with complementary coverage:

Dimensional ordering. The three dimensions are evaluated in deliberate sequence: *Confidence*, *Content*, *Concern*. This ordering reflects operational logic: (1) *Can we understand the caller?*—ASR confidence determines whether transcription is reliable enough for downstream analysis; (2) *What is being reported?*—semantic content establishes the substance of the emergency; (3) *How distressed does the caller sound?*—bio-acoustic indicators validate and can elevate priority, but do not override content. This sequence ensures that a composed professional reporting a mass casualty event receives appropriate priority based on content, while a highly distressed caller reporting a minor issue is not over-prioritized based on vocal expression alone.

- **High Confidence + Low Content + Low Concern:** Routine call; dispatcher applies ESI using extracted entities at normal pace

- **High Confidence + High Content + Low Concern:** The composed reporter—urgent content from a calm caller requires elevated queue position; dispatcher reviews entities and applies ESI, likely assigning ESI-2 or ESI-3

- **High Confidence + Low Content + High Concern:** Anxious caller, possibly minor issue—dispatcher assesses whether distress reflects emergency or anxiety

- **High Confidence + High Content + High Concern:** All signals aligned; immediate queue position for rapid ESI/START application

- **Low Confidence + Low Content + Low Concern:** Likely technical issue; dispatcher reviews audio quality before processing

- **Low Confidence + High Content + Low Concern:** Garbled but fragments suggest urgency—elevated priority; dispatcher listens directly

- **Low Confidence + Low Content + High Concern:** Distressed caller with unintelligible speech—immediate priority; dispatcher listens and applies protocol based on direct assessment

- **Low Confidence + High Content + High Concern:** Maximum queue priority—all indicators suggest crisis; immediate dispatcher attention

Two cells represent our key insights. The **High Confidence + High Content + Low Concern** cell captures callers whose semantic content demands urgent attention despite calm delivery: the trained first responder, medical professional, or composed bystander whose measured voice belies the severity of their report. The **Low Confidence + Low Content + High Concern** cases capture the complementary pattern—callers in crisis whose speech has shifted toward basilectal registers, where ASR failure combined with vocal stress becomes valuable prioritization information rather than system failure.

Together, these insights ensure that neither semantic nor paralinguistic signals alone determine queue position—and that clinical triage decisions remain with trained dispatchers who can assess the full context of each call.

4.6 Queue Prioritization Engine

The Queue Prioritization Engine integrates three independent signals to determine the order in which calls receive dispatcher attention. **Critically, this system determines queue position, not clinical triage category.** Clinical triage—assigning ESI levels 1–5 or START colors (RED/YELLOW/GREEN/BLACK)—remains the responsibility of trained dispatchers applying Ministry of Health protocols.

The prioritization logic ensures that:

1. Callers most likely to need immediate intervention reach dispatchers first

2. Dispatchers receive structured information to support rapid protocol application

3. Calls with unreliable transcriptions are flagged for direct audio review

4.6.1 Three-Dimensional Prioritization Space

Each call is mapped to a point in prioritization space defined by:

- **Transcription Confidence** (C): High (≥ 0.7) or Low (< 0.7)

- **Content Indicators** (S_c): High (≥ 50) or Low (< 50)

- **Bio-Acoustic Distress** (D): High (> 0.5) or Low (≤ 0.5)

The $2 \times 2 \times 2$ combination yields eight queue priority cells, shown in Table 3.

Confidence	Content	Concern	Queue	Dispatcher Action
High	Low	Low	**Q5-ROUTINE**	Apply ESI using extracted entities
High	High	Low	**Q2-ELEVATED**	Priority review; calm reporter, urgent content*
High	Low	High	**Q3-MONITOR**	Review for anxiety vs. emergency
High	High	High	**Q1-IMMEDIATE**	Immediate attention; apply ESI/START
Low	Low	Low	**Q5-REVIEW**	Check audio quality; possible technical issue
Low	High	Low	**Q2-ELEVATED**	Listen to audio; fragments suggest urgency
Low	Low	High	**Q1-IMMEDIATE**	Priority audio review; possible dialect shift†
Low	High	High	**Q1-IMMEDIATE**	Highest priority; all indicators elevated

Table 3: Three-dimensional queue prioritization matrix. *Addresses trained responder/composed bystander scenario. †Preserves core insight: low ASR confidence + high vocal concern may indicate stress-induced basilectal shift requiring human ears.

4.6.2 Queue Priority Levels

Q1-IMMEDIATE: Top of queue. Dispatcher reviews within seconds. System flags call for potential crisis requiring direct audio assessment.

Q2-ELEVATED: High priority queue. Dispatcher attention within 1–2 minutes. Extracted entities displayed prominently to support rapid ESI/START application.

Q3-MONITOR: Moderate priority. May indicate anxious caller with non-urgent situation. Dispatcher assesses and de-escalates if appropriate.

Q5-ROUTINE: Standard queue. Extracted entities available; dispatcher applies ESI at normal pace.

Q5-REVIEW: Standard queue but flagged for audio quality check. May indicate technical issues rather than emergency content.

Note on Q4: The current matrix does not produce a Q4 outcome. Future refinement with real operational data may identify scenarios warranting an intermediate priority level. A theoretical case: High Confidence + Low Content + Moderate Concern (anxious caller, minor issue).

4.6.3 Relationship to Clinical Triage Protocols

Table 4 illustrates how TRIDENT's queue prioritization relates to—but does not replace—clinical triage protocols.

4.6.4 Dispatcher Interface

Figure 3 illustrates the dispatcher interface for a high-priority scenario. The interface presents:

- Queue priority level with visual urgency coding

- Transcription confidence (with recommendation to review audio if low)

TRIDENT Output	Dispatcher Action	Protocol Application
Q1-IMMEDIATE	Immediate audio review; assess caller state	Dispatcher determines ESI-1/2 or START-RED based on clinical assessment
Q2-ELEVATED	Review extracted entities; listen if uncertain	Dispatcher applies ESI using structured data; may be ESI-2 through ESI-4
Q3-MONITOR	Assess distress source; de-escalate if needed	Often ESI-4/5 after dispatcher determines no emergency
Q5-ROUTINE/REVIEW	Process normally using extracted metadata	Full ESI protocol application; typically ESI-3 through ESI-5

Table 4: TRIDENT queue priority does not determine clinical triage level. Dispatchers apply ESI or START protocols after reviewing TRIDENT's structured outputs and/or call audio.

- Extracted clinical entities mapped to ESI/START decision points

- Bio-acoustic distress indicators

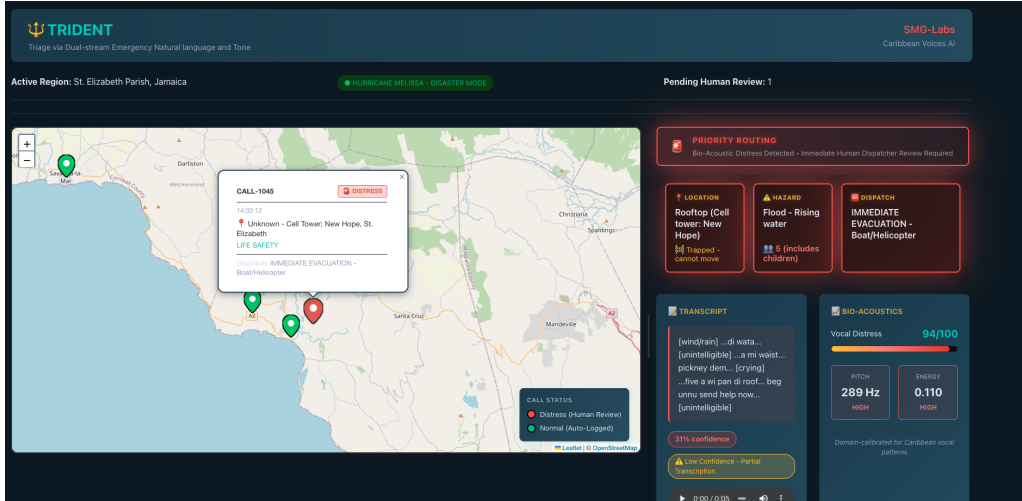- One-click access to call audio for direct assessment

Figure 3: Dispatcher interface for a high-priority scenario (Q1-IMMEDIATE). Elevated distress markers combined with low transcription confidence trigger immediate queue placement. The interface prominently recommends audio review and displays partial entity extraction with uncertainty markers. The dispatcher will listen directly and apply ESI or START protocol based on their clinical assessment.

5 Deployment Considerations

5.1 Operational Context: Supporting Protocol Application

TRIDENT integrates with existing emergency dispatch workflows to support standardized triage protocols—ESI for routine operations, START for mass casualty incidents. **Day-to-day (ESI context):** TRIDENT processes incoming calls to extract structured entities (location, mechanism, clinical indicators) and assigns queue priority. Dispatchers apply ESI to determine clinical acuity level (1–5) and appropriate response. **Mass casualty events (START context):**

During hurricanes or earthquakes, TRIDENT's queue prioritization manages call surges when volume exceeds dispatcher capacity, enabling rapid caller sorting even when transcription quality degrades. **Key principle:** TRIDENT determines which calls dispatchers see first and what structured information they receive; clinical triage decisions remain with trained professionals applying Ministry of Health protocols.

5.2 Primary Deployment: Surge Queue Prioritization

TRIDENT's greatest value emerges during **disaster surge conditions**—hurricanes, earthquakes, floods—when call volume exceeds dispatcher capacity and callers must wait in queue. TRIDENT's processing latency (45–60 seconds on edge hardware) precludes real-time transcription, but surge queues provide ideal operational context.

   Operational flow:

1. Caller dials emergency services; all dispatchers engaged

2. Caller enters queue and hears automated message requesting description

3. Caller provides initial statement (15–30 seconds)

4. TRIDENT processes audio while caller waits (45–60 seconds)

5. Queue reordered by priority (Q1-IMMEDIATE through Q5-ROUTINE)

6. Highest-priority call routes first when dispatcher becomes available

7. Dispatcher receives transcription, extracted entities, and distress indicators to support ESI/START application

   Why this context maximizes value: Calls are waiting regardless—TRIDENT uses wait time productively. Queue prioritization ensures most critical callers reach dispatchers first. Extracted entities enable faster protocol application. Low ASR confidence flags alert dispatchers to potential dialect shift or audio quality issues before engagement.

   This deployment model represents TRIDENT's primary design target. Caribbean emergency services face predictable annual surge events (hurricane season, June–November) where this capability would directly impact response effectiveness.

5.3 Early Exit for Critical Cases

To provide faster routing for clearly distressed callers, the system implements early exit when:

1. **High Distress + Low Confidence:** If $D > 0.8$ and $C < 0.4$, route immediately to Q1-IMMEDIATE. This captures callers exhibiting extreme vocal stress whose speech has likely shifted to basilectal registers.

2. **Extreme Distress:** If $D > 0.9$ regardless of confidence, route to Q1-IMMEDIATE.

   Under early exit, ASR and bio-acoustics complete in approximately 12 seconds (with bio-acoustic extraction parallel to transcription), reducing Time-to-Q1 from 55 seconds to 12 seconds for clearly distressed callers—a critical improvement for surge queue scenarios.

5.4 Offline Operation

All components operate without internet connectivity: Whisper model weights and Llama 3 stored locally, bio-acoustic analysis uses standard signal processing libraries, and queue logic implemented in local Python. This enables deployment at emergency coordination centers that may lose connectivity during disasters while maintaining local power (generator/battery backup). Offline capability ensures TRIDENT can support ESI/START protocol application precisely when infrastructure degradation makes accurate call processing most difficult.

5.5 Integration with Existing Dispatch Systems

TRIDENT operates as a **pre-processing layer** integrating with existing Computer-Aided Dispatch (CAD) systems. The system accepts audio streams, processes them through the three-layer architecture, and outputs structured data packages (queue priority, transcription with confidence, extracted entities, distress indicators) to CAD systems. Dispatchers receive calls in priority order and apply ESI or START protocols using TRIDENT's structured data and/or direct audio review. This requires no changes to clinical protocols—only familiarization with TRIDENT's output format.

5.6 Hardware Requirements

The complete system deploys on Raspberry Pi 5 (8GB RAM) or equivalent edge hardware:

Component	Model	Size	Inference Speed
ASR	Whisper Medium (INT4)	~400MB	~10s per 30s audio
NLP	Llama 3 8B (4-bit)	~4GB	2-5 tokens/sec
Bio-acoustic	librosa + numpy	<50MB	Real-time

Table 5: Hardware requirements for edge deployment

Total system footprint: ~4.5GB, well within Raspberry Pi 5 8GB capacity.

6 Limitations and Future Work

6.1 Current Limitations

Validation gap (most critical). This paper presents an architectural framework with theoretical grounding but limited empirical validation on real emergency calls. Performance claims are based on component evaluations and related literature rather than end-to-end system testing. The three-dimensional queue prioritization matrix has not been validated against expert dispatcher judgments.

Protocol integration. While TRIDENT is framed as supporting ESI and START protocols, the entity extraction schema and queue prioritization logic were developed independently of clinical stakeholder input. Full Ministry of Health integration requires validation that extracted entities map correctly to ESI decision points and that queue priorities align with operational workflows.

Training data constraints. Caribbean emergency speech corpora do not exist. ASR fine-tuning was performed on broadcast speech, which differs from emergency call acoustics in noise profiles, emotional content, and register distribution.

Sex differences in F0 baseline. Fundamental frequency is sexually dimorphic: male voices typically range 85–175 Hz while female voices range 165–270 Hz [16, 17]. We mitigate this by prioritizing sex-normalized features (F0 coefficient of variation, jitter) over absolute F0 elevation in distress score calculation. Research confirms that stress manifests with "striking parallels in men and women" [13]—both sexes show increased pitch mean and variation under acute stress. However, residual bias risks remain: relaxed female speakers near upper baseline may contribute to elevated distress scores, while stressed male speakers with naturally low F0 may not contribute sufficiently. A validation study with sex-stratified analysis on Caribbean emergency calls is essential to calibrate population-appropriate thresholds and confirm normalized measures maintain sensitivity across demographics.

Content indicator classification. The Content Indicator Score depends on LLM classification quality. Caribbean creole expressions not well-represented in training data may be

misclassified. Empirical evaluation of classification accuracy on Caribbean transcripts is needed, particularly for false negatives that could delay critical calls.

Single-speaker assumption. Multi-party calls are not handled. Speaker changes mid-call could confuse bio-acoustic analysis and entity extraction.

Threshold sensitivity. Multiple thresholds (ASR confidence 0.7, distress 0.5, content indicators 50) were selected based on literature but have not been rigorously optimized. Sensitivity analysis examining precision-recall tradeoffs is needed.

6.2 Future Work

Clinical stakeholder collaboration. Partnership with Caribbean emergency services to validate TRIDENT's utility in real dispatch workflows, including observation studies of current ESI/START challenges, dispatcher feedback on extracted entity usefulness, and iterative schema refinement based on clinical input.

Caribbean Emergency Speech Corpus. A dedicated corpus combining Caribbean-accented speech with emergency domain content and stress annotations is critical. We are exploring *VoicefallJA*, a gamified speech elicitation platform designed to collect stressed Caribbean speech through game-induced cognitive load rather than acted performance. The Progressive Web App targets 100–300 speakers via church network distribution, with Q2–Q3 2026 data collection. However, game-induced stress differs fundamentally from genuine emergency distress; this approach should be viewed as a stepping stone toward real-call annotation under appropriate ethical frameworks, not a replacement.

Empirical validation. End-to-end evaluation with emergency dispatch professionals assessing whether TRIDENT's queue prioritization aligns with expert judgment, including sex-stratified analysis of bio-acoustic accuracy and entity extraction accuracy on Caribbean creole transcripts.

Ablation studies. Quantifying the marginal contribution of each architectural component (bio-acoustic analysis, content indicators, Caribbean-tuned ASR).

Sex-adaptive distress detection. Implementing within-call F0 change detection rather than absolute thresholds, and ensemble approaches combining multiple normalization strategies.

7 Conclusion

TRIDENT presents a dispatcher-support architecture that ensures Caribbean-accented emergency callers receive equitable access to ESI and START triage protocols. By combining accent-adapted speech recognition, local NLP entity extraction, and bio-acoustic distress detection, the system empowers dispatchers to apply established protocols even when automated transcription fails.

The architecture operationalizes two complementary insights established in Section 1.2: that ASR uncertainty combined with vocal distress signals priority callers requiring human attention, and that calm delivery of urgent content must not delay dispatcher response. These insights drive the three-dimensional queue prioritization matrix that routes calls based on confidence, content, and concern signals.

Critically, TRIDENT respects the clinical authority of established protocols. The system determines which calls dispatchers see first and provides structured information to support rapid protocol application—but triage decisions remain with trained human professionals. This design philosophy reflects a broader principle for emergency AI: technology should empower human expertise, not attempt to replace it.

We hope this architectural framework contributes to more equitable emergency services—not just for Caribbean populations, but for the billions of speakers worldwide whose accents and

dialects remain underserved by current speech technology. When a caller dials for help, the system that answers should understand them. TRIDENT is a step toward that goal.

References

[1] Afraa Attiah and Manal Kalkatawi. AI-powered smart emergency services support for 9-1-1 call handlers using textual features and svm model for digital health optimization. *Frontiers in Big Data*, 8:1594062, 2025.

[2] Stig Nikolaj Blomberg et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*, 138:322–329, 2019.

[3] Paul Boersma and David Weenink. *Praat: doing phonetics by computer*, 2013. Version 5.3.51.

[4] Marcel Lucas Chee, Mark Leonard Chee, Haotian Huang, Katelyn Mazzochi, Kieran Taylor, Han Wang, Mengling Feng, Andrew Fu Wah Ho, Fahad Javaid Siddiqui, Marcus Eng Hock Ong, Nan Liu, et al. Artificial intelligence and machine learning in prehospital emergency care: A scoping review. *iScience*, 26(8):107407, 2023.

[5] Théo Deschamps-Berger, Lori Lamel, and Laurence Devillers. End-to-end speech emotion recognition: Challenges of real-life emergency call centers data recordings. In *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8, 2021.

[6] Emergency Nurses Association. *Emergency Severity Index (ESI): A Triage Tool for Emergency Departments, Version 5*. Emergency Nurses Association, Schaumburg, IL, 2020. Available at https://www.ena.org/practice-resources/resource-library/esi.

[7] Simone French, Georgiana Gordon-Strachan, Kevon Kerr, Jacquiline Bisasor-McKenzie, Lambert Innis, and Paula Tanabe. Assessment of interrater reliability of the emergency severity index after implementation in emergency departments in jamaica using a learning collaborative approach. *Journal of Emergency Nursing*, 46(6):875–882, 2020.

[8] Tamar H. Gollan and Victor S. Ferreira. Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3):640–665, 2009.

[9] David W. Green. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2):67–81, 1998.

[10] Allison Koenecke et al. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689, 2020.

[11] Jordan Madden, Matthew Stone, Dimitri Johnson, and Daniel Geddez. Towards robust speech recognition for Jamaican Patois music transcription. *arXiv preprint arXiv:2507.16834*, 2025.

[12] Peter L. Patrick. *Urban Jamaican Creole: Variation in the Mesolect*. John Benjamins Publishing, Amsterdam, 1999.

[13] Katarzyna Pisanski, Joanna Nowak, and Piotr Sorokowski. Multimodal stress detection: Testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. *Hormones and Behavior*, 106:52–61, 2018.

[14] Carlos Santos-Burgoa, John Sandberg, Erick Suárez, Ann Goldman-Hawes, Scott Zeger, Alejandra Garcia-Meza, Cynthia M. Pérez, Kenneth Rivera, Adriana Colón Ramos, Jose Figueroa, et al. Differential and persistent risk of excess mortality from hurricane maria in puerto rico: A time-series analysis. *The Lancet Planetary Health*, 2(11):e478–e488, 2018.

[15] Lilien Schewski, Mathew Magimai Doss, Guido Beldi, and Sandra Keller. Measuring negative emotions and stress through acoustic correlates in speech: A systematic review. *PLOS ONE*, 20(7):e0328833, 2025.

[16] Ingo R. Titze. Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85(4):1699–1707, 1989.

[17] Hartmut Traunmüller and Anders Eriksson. The frequency range of the voice fundamental in the speech of male and female adults. *Journal of the Acoustical Society of America*, 97(4):2634–2639, 1995.

[18] Martine Van Puyvelde, Xavier Neyt, Francis McGlone, and Nathalie Pattyn. Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in Psychology*, 9:1994, 2018.

[19] André Veiga et al. The fundamental frequency of voice as a potential stress biomarker: A systematic review and meta-analysis. *Stress and Health*, 2025.

A   Implementation Details

Repository: https://github.com/smg-labs/project-filter *(to be made public upon acceptance)*

Dependencies:

- Python 3.11+

- openai-whisper

- transformers, peft (LoRA fine-tuning)

- ollama (Llama 3 serving)

- librosa (audio feature extraction)

- jiwer (WER evaluation)

Hardware requirements:

- Training: NVIDIA GPU with 16GB+ VRAM recommended

- Inference: CPU-only operation supported; 8GB RAM minimum

B   Acknowledgments