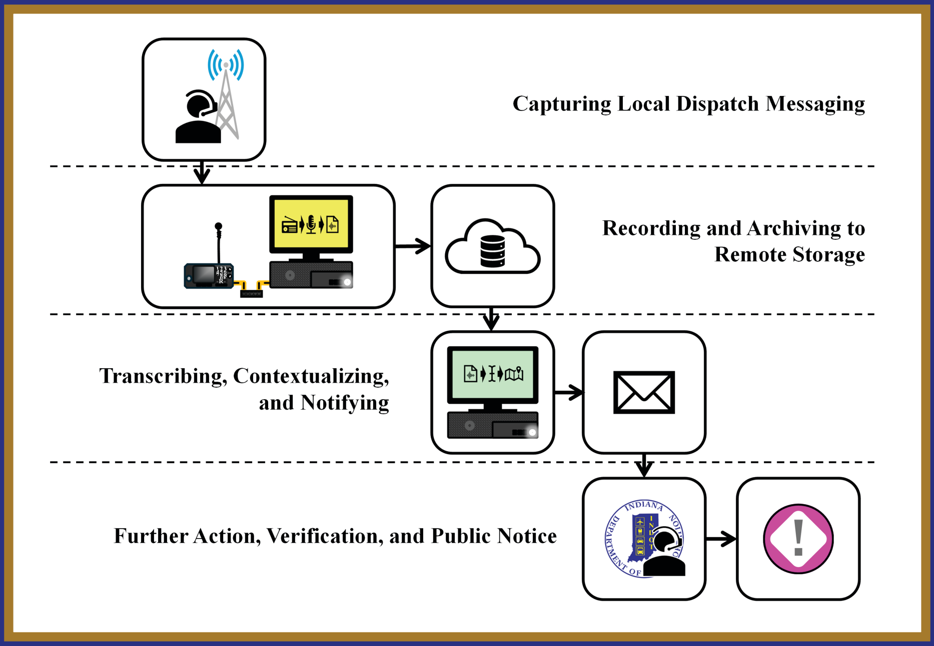
JOINT TRANSPORTATION
RESEARCH PROGRAM
INDIANA DEPARTMENT OF TRANSPORTATION
AND PURDUE UNIVERSITY

Utilizing Emergency Management Audio Messages for Traffic Incident Management Performance Measures



Christopher Gartner, Vihaan Vajpayee,
Jairaj C. Desai, and Darcy M. Bullock

SPR-4931 • Report Number: FHWA/IN/JTRP-2026/15 • DOI: 10.5703/1288284318623

RECOMMENDED CITATION

Gartner, C., Vajpayee, V., Desai, J. C., & Bullock, D. M. (2026). *Utilizing emergency management audio messages for Traffic Incident Management performance measures* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2026/15). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284318623

AUTHORS

Christopher Gartner
JTRP Transportation Research Engineer
Lyles School of Civil and Construction Engineering
Purdue University

Vihaan Vajpayee
Purdue University

Jairaj C. Desai, PhD
JTRP Transportation Research Engineer
Lyles School of Civil and Construction Engineering
Purdue University

Darcy M. Bullock, PhD, PE
Lyles Family Professor of Civil Engineering
JTRP Director
Lyles School of Civil and Construction Engineering
Purdue University
(765) 494-2226
darcy@purdue.edu
Corresponding Author

ACKNOWLEDGMENTS

ProScan (by Bob Anue) was utilized for programming, interfacing, recording, and indexing all compressed audio data. Crowdsourced information from RadioReference was used to select appropriate talk groups, frequencies, and channels for programming all radio equipment. The authors would like to acknowledge Enrique Daniel Saldivar-Carranza, Howell Li, Andrew Tully, Christopher Montagna, Landon Jacome-Mendez, and Daniel Chen for their integral efforts to this program that enabled this study.

JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education institutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. Learn more at engineering.purdue.edu/JTRP.

Published reports of the Joint Transportation Research Program are available at docs.lib.purdue.edu/jtrp/.

NOTICE

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2026/15		

4. Title and Subtitle
Utilizing emergency management audio messages for Traffic Incident Management performance measures

5. Report Date
March 6, 2026

6. Performing Organization Code

7. Author(s)
Christopher Gartner (https://orcid.org/0009-0008-8448-2356)
Vihaan Vajpayee (https://orcid.org/0009-0005-4212-8744)
Jairaj C. Desai, PhD (https://orcid.org/0000-0003-2885-203X)
Darcy M. Bullock, PhD, PE (https://orcid.org/0000-0002-7365-1918)

8. Performing Organization Report No.
FHWA/IN/JTRP-2026/15

9. Performing Organization Name and Address
Joint Transportation Research Program
Hall for Discovery and Learning Research (DLR), Suite 204
207 S. Martin Jischke Drive
West Lafayette, IN 47907

10. Work Unit No.

11. Contract or Grant No.
SPR-4931

12. Sponsoring Agency Name and Address
Indiana Department of Transportation (SPR)
State Office Building
100 North Senate Avenue
Indianapolis, IN 46204

13. Type of Report and Period Covered
Final Report

14. Sponsoring Agency Code

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract
Timely and accurate information regarding roadway incidents is critical for Traffic Management Centers (TMCs) to maintain safety and mobility. While Intelligent Transportation Systems (ITS) cameras and crowdsourced data provide significant visibility, rural interstate corridors often suffer from coverage gaps, or "blind spots," where incidents go undetected until significant queuing occurs. Public safety radio communications serve as the primary coordination method for first responders, creating a "verbal infrastructure" that contains vital information regarding incident location and severity. This study evaluates the feasibility of using Automatic Speech Recognition (ASR) to bridge these gaps by monitoring and analyzing radio traffic. The project developed and validated a prototype system capable of capturing, transcribing, and geolocating public safety radio transmissions for Traffic Incident Management (TIM) along a 71-mile segment of I-65 between Indianapolis and Chicago. The findings include a system latency assessment, transcription accuracy analysis, noise filtering, and a case study. Further recommendations are made based on the findings for improving results and larger implications.

17. Key Words
traffic management, traffic incident management, automatic speech recognition, public safety answering point

18. Distribution Statement
No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	35, including appendices	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

EXECUTIVE SUMMARY

Motivation

Timely, accurate information on roadway incidents is crucial for Traffic Management Centers (TMCs) to coordinate incident response and cleanup. While intelligent traffic system (ITS) cameras and crowdsourced connected vehicle data help identify problems after they occur, rural interstate corridors often have coverage gaps, or "blind spots," where incidents go unnoticed until queues form (Mathew et al., 2024; Sakhare et al., 2024). Public safety radio communications are the primary coordination method for first responders, forming a "verbal infrastructure" that is difficult to integrate with automated systems due to its unstructured nature. This study explores using Automatic Speech Recognition (ASR) to fill these gaps by monitoring and analyzing radio traffic (Alharbi et al., 2021; Hawkins, 2007; Ibrahim & Varol, 2020; Stein et al., 2012).

Study

The study developed and validated a prototype system that captures, transcribes, and geolocates public safety radio transmissions for Traffic Incident Management (TIM) along a 71-mi segment of Interstate 65 between Indianapolis and Chicago. Using digital trunking scanners and remote-control software, more than 100,000 transmissions were recorded from May 12 to June 30, 2025. This report describes:

- System Architecture: Integration of OpenAI's Whisper automated speech recognition (ASR) model to transcribe unstructured voice data into text.
- Keyword Analysis: Implementation of logic-based filtering using "Interstate," "Location," and "Incident" keywords to identify relevant events.
- Performance Evaluation: Assessment of transcription latency, Word Error Rate (WER), and model confidence to determine the system's suitability for real-time operations.
- Case Study Validation: Comparison of ASR-generated timelines against ITS camera imagery during a tractor-trailer fire event.

Results

The following results were observed from the analysis:

- Latency: The system demonstrated an average end-to-end processing time (from transmission end to notification delivery) of 59 s, with 90% of alerts delivered within 96 s.
- Accuracy: There is an inverse relationship between message length and error frequency; while short utterances (0–10 words) had high error rates, substantive transmissions (>20 words) achieved a stabilized Mean WER of 0.332.
- Filtering: Filtering transmissions by minimum word count and model confidence effectively segregated valuable dispatch narratives from short utterances such as 10-4 and radio noise.
- Operational Value: In a case study of a tractor-trailer fire, the system captured critical incident details, providing situational awareness well before significant queuing was visible on ITS cameras and well before the TMC became aware of the incident.

Recommendations

The study confirms that transforming unstructured voice data into structured, geocoded text provides a scalable solution to extend TMC visibility into rural corridors where sensor density is limited. Future research should focus on fine-tuning models specifically for public safety lexicons, including "ten-codes" and local vernacular. While ASR serves as a vital bridge for interoperability gaps, a Centralized Computer-Aided Dispatch (CAD) database that aggregates direct dispatcher notes represents a superior long-term solution for incident awareness across multiple agencies.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. PROJECT OVERVIEW

Timely and accurate information regarding roadway incidents is critical for Traffic Management Centers (TMCs) to maintain safety and mobility. While intelligent traffic system (ITS) cameras and crowdsourced data provide significant visibility, rural interstate corridors often suffer from coverage gaps. This study was initiated to evaluate the feasibility of using automated speech recognition (ASR) to bridge these gaps by monitoring public safety radio traffic. By capturing, transcribing, and analyzing dispatch communications along a 71-mi segment of Interstate 65 (I-65), this research demonstrates a novel, scalable approach to incident detection. This report is structured as follows:

- Section 1: Provides an overview of the project's scope, objectives, and dissemination of results.
- Section 2: Details the operational environment, including Public Safety Answering Points (PSAPs), the Project Hoosier SAFE-T radio network, and the hardware/software infrastructure used for signal capture.
- Section 3: Describes the system architecture, including the integration of ProScan software, ASR model selection (OpenAI Whisper vs. Google speech-to-text [STT]), and the logic-based keyword analysis used to filter noise and identify incidents.
- Section 4: Presents a case study of the system during a tractor-trailer fire on I-65, providing a comparative timeline of radio dispatch versus ITS camera observations.
- Section 5: Summarizes the system's operational viability, highlighting latency metrics and the effectiveness of model confidence as a quality control measure.
- Section 6: Outlines necessary steps for scalability, including model fine-tuning on public safety vernacular and the potential for integration with centralized Computer-Aided Dispatch (CAD) systems.

1.1 Introduction

Traffic Incident Management (TIM) relies heavily on the speed at which incidents are detected and verified. Traditionally, TMCs rely on a combination of inductive loop sensors, ITS cameras, and third-party data (e.g., Waze, Google Maps) to monitor network conditions. However, infrastructure constraints often leave rural interstate segments unmonitored, creating "blind spots" where incidents may go undetected until significant queueing occurs.

1.1.1 The Role of Public Safety Audio

Public safety radio communications serve as the primary coordination channel for first responders. Dispatchers and field units continuously exchange critical information regarding incident location, severity, and roadway status (e.g., lane closures). TMCs in urban areas typically monitor relevant radio channels/talk groups in the background. Although this works well in urban areas with Traffic Management Staff monitoring a limited geographic area and agencies, it does not scale statewide.

1.1.2 Advancements in ASR Technology

Recent advancements in deep learning and ASR have enabled high-fidelity transcription of noisy, technical audio streams. By leveraging models such as OpenAI's Whisper, unstructured voice data can be converted into text, allowing for the application of Natural Language Processing (NLP) techniques to extract actionable intelligence. This study explores the application of these technologies to create a low-cost, automated sensor network that leverages existing public safety broadcasts to detect roadway incidents faster than traditional methods.

1.2 Scope and Objectives

The primary objective of this study is to develop and validate a prototype system capable of capturing, transcribing, and geolocating public safety radio transmissions for TIM applications. The research focuses on a 71-mi segment of Interstate 65 (I-65) between Indianapolis and Chicago, covering White, Tippecanoe, Clinton, and Boone counties.

Key objectives include:

- System Architecture: Assemble and deploy a prototype audio logging system utilizing digital trunking scanners (Uniden SDS200) and remote-control software (ProScan).
- Data Management: Develop robust procedures for logging audio files and designing a relational database to manage metadata, transcriptions, and keyword extractions.
- Model Evaluation: Assess the performance of various ASR models (e.g., Whisper Turbo, Google STT) based on transcription latency and Word Error Rate (WER) to determine suitability for real-time applications.
- Incident Detection: Develop and refine a keyword filtering protocol (e.g., "I-65," "Mile Marker," "Crash") to automatically identify relevant traffic incidents while filtering out routine administrative traffic.
- Performance Metrics: Quantify system performance by measuring notification latency (the time elapsed between a radio transmission and the delivery of an automated alert) and validating accuracy against ground-truth events.

1.3 Dissemination of Research Results

The findings and methodologies developed during this project have been documented to facilitate technology transfer to public agencies and private sector stakeholders. A significant portion of the methodology and case study analysis presented in this report has been adapted from the following open-access journal article (Gartner et al., 2025):

> Gartner, C. M., Vajpayee, V., Desai, J., & Bullock, D. M. (2025). Automatic speech recognition of public safety radio communications for interstate incident detection and notification. *Smart Cities*, *8*(5), 157. https://doi.org/10.3390/smartcities8050157

Furthermore, the findings from this study were presented to stakeholders throughout this project at various Study Advisory Committee (SAC) meetings and engagement events. The following sections of this technical report summarize the background, methodology, and key findings of this research.

2. BACKGROUND INFORMATION

2.1 Public Safety Answering Points

PSAPs answer all 9-1-1 and E9-1-1 calls (IN911, n.d.). These centers are formally categorized based on their call routing function: a primary PSAP receives 9-1-1 calls directly from the 9-1-1 Control Office (such as a selective router), while a secondary PSAP receives 9-1-1 calls only after they have been transferred from a primary PSAP (Federal Communications Commission [FCC], n.d.a).

In this report, it is important to consider PSAP as the intermediary between motorists and first responders. Further, motorists are crucial sources of information about the condition of the roadway and are often the primary source in rural or underserved/under-monitored sections of the interstate (either due to limited resources or infrastructure). Therefore, it is desirable to capture and analyze communication between dispatch and first responders to obtain timely information on roadway conditions.

2.1.1 Indiana Public Safety Answering Points

PSAPs are the critical first point of contact for emergency services in Indiana, responsible for receiving and dispatching 9-1-1 calls. The state maintains 117 PSAPs operating across 91 counties, with Fountain and Warren counties consolidating their 911 operations into a single authority (IN911, n.d.). The organizational details of these centers, including location and contact information, are tracked in the Master PSAP Registry via an FCC-assigned identification number, ensuring regulatory compliance and facilitating system coordination across the state (FCC, n.d.). Figure 2.1 displays the location of PSAPs within Indiana.

2.2 Public Safety Radio

The public safety spectrum supports mission-critical communications for first responders such as police, firefighters, and EMS. It also fulfills the telecommunications needs of various levels of government (FCC, n.d.b). Throughout Indiana, there are several implementations of Public Safety Radio systems that facilitate encrypted and non-encrypted communications within operational groups, county-level communications, and multi-regional communications across the state (Integrated Public Safety Commission, n.d.b).

2.2.1 Trunked and Conventional Systems

The main difference between trunked and conventional radios is how they assign frequencies. Conventional systems give each talk group a fixed, dedicated frequency, which can lead to inefficiency and congestion. Trunked systems, like Project 25 (P25), assign frequencies dynamically from a shared pool, allowing many talk groups to efficiently share limited frequencies. This enhances channel utilization and interoperability.

2.2.2 Project Hoosier SAFE-T

Project Hoosier SAFE-T (Safety Acting For Everyone-Together) is a statewide initiative by the Integrated Public Safety Commission (IPSC), created by the Indiana General Assembly in 1999 (IPSC, n.d.a), to implement a unified, interoperable, digital 800 MHz trunked voice and mobile data communications network for public safety officials across Indiana's 92 counties (State of Indiana, 2024). Historically, disparate radio systems (low-band VHF, UHF, and 800 MHz) prevented agencies from communicating when an incident crossed jurisdictional lines (IPSC, n.d.a). The SAFE-T system was specifically designed to solve this critical interoperability gap by implementing a unified, digital 800 MHz trunked network that provides at least 95% statewide mobile coverage through its expansive 126-site infrastructure (IPSC, n.d.a, n.d.b; State of Indiana, 2024).

2.2.3 Operating Frequency Bands

Public safety communication utilizes various frequency bands, each possessing unique propagation properties. VHF (below 470 MHz) covers large rural areas but struggles with building penetration and is highly congested. UHF (470–512 MHz) offers better urban penetration at the expense of shorter range. These lower-band systems are now primarily retained for backup purposes due to their limited channel capacity and vulnerability to interference (FCC, n.d.b).

The 700 MHz and 800 MHz bands are standard for high-capacity, reliable communication, like in P25 systems such as Hoosier SAFE-T (FCC, n.d.b; State of Indiana, n.d.). They require denser towers but provide better urban coverage and are used in trunked systems to reduce congestion. In Indiana, a combination of the above spectrums is used throughout the state and must thus be considered to fully "cover" a given region.

2.3 Public Safety Talk Groups

Talk groups are the core units of communication in digital P25 trunked radio systems, like Indiana's Hoosier SAFE-T Network. They function as logical groups of users rather than having permanently assigned radio frequencies (State of Indiana, n.d.; RadioReference, n.d.; National Pubic Safety Telecommunications Council, n.d.). The system greatly improves efficiency by temporarily assigning a voice channel from a limited pool of frequencies to a specific talk group during transmission, making sure only the designated radios tune in. This allows hundreds of agencies to share channels without constant interference (Hawkins, 2007; IPSC, n.d.b). Additionally, talk groups are vital for interoperability by using predefined mutual aid channels (e.g., FIRE or EMS mutual aid), enabling local, county, and state agencies to establish seamless cross-jurisdictional communication rapidly during large-scale events.

2.3.1 Frequency Separation and Range

Monitoring a 70-mi section of interstate spanning multiple jurisdictions poses significant challenges due to frequency overlap and incomplete coverage. While the SAFE-T 800 MHz

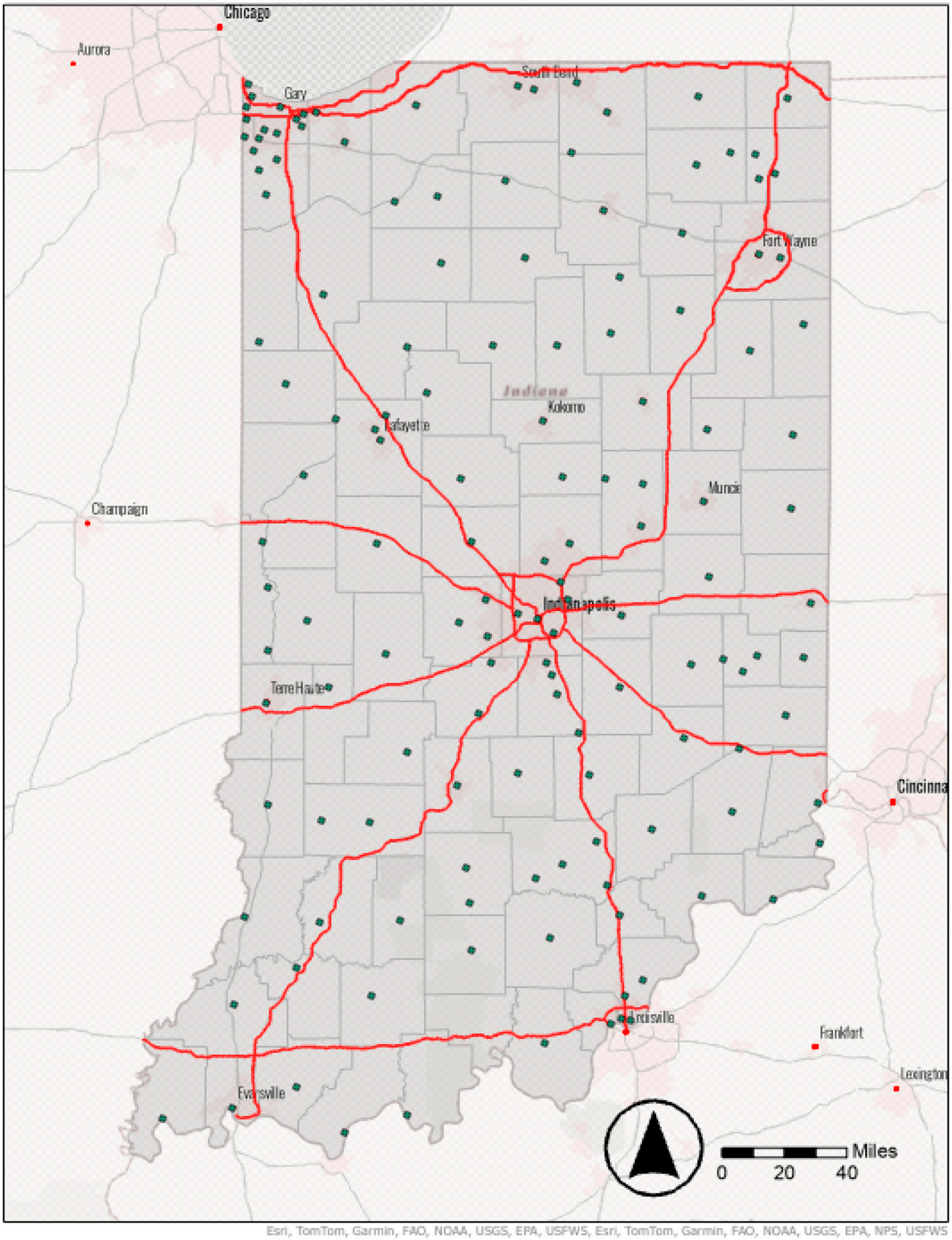


Figure 2.1 Map of PSAP Locations Across Indiana Counties. Data Retrieved October 2025 (FCC, n.d.a).

trunked network mitigates static frequency assignments via dynamic channel allocation (Hawkins, 2007; Integrated Public Safety Commission, n.d.b; Computer Assisted Pre-Coordination Resource and Database System, n.d.), this solution is not universal. Many adjacent jurisdictions still rely on older, conventional VHF systems. VHF's superior signal propagation causes co-channel interference where one county's traffic can bleed into and "appear as another's" communication, creating confusion about call origin. This fundamental physical phenomenon necessitates careful system planning to distinguish local, overlapping transmissions across the corridor (Bercovivi, 2006).

2.4 Digital Police Scanner Sites

The purpose of identifying digital police scanner sites is to strategically collect regional dispatch communications over the public, unencrypted spectrums. In the course of this research, several initial sites were identified and established, and future sites of interest were proposed.

2.4.1 Digital Police Scanner

The Digital Police Scanner is the physical hardware used to capture and record audio and metadata, as detailed in later sections. Two units were implemented: The Uniden BCD996P2 (Figure 2.2; Uniden, n.d.a) and the Bearcat SDS200 (Figure 2.3; Uniden, n.d.b). The former is an older yet capable model, but it faced certain challenges in programming and data interfacing. It was decided that the newer SDS200 would better suit the research program's needs, enabling better control and communication over Internet Protocol (IP).

2.4.2 Deployed and Proposed Sites

At the time of this report, three sites were established: one at the Indianapolis TMC, one at the Hampton Hall of Civil and Construction Engineering (HAMP) on Purdue University's campus, and one at the Gary TMC (near Chicago, Illinois). Figure 2.4 shows each location within Indiana. The SDS200s were allocated to the Purdue and Gary sites, and the Indianapolis site was established with the BCD996P2 unit.

Figure 2.2 The Uniden BCD996P2 Digital Trunking Scanner Used for Initial Prototype Sites and Legacy System Monitoring (Uniden, n.d.a).

Figure 2.3 The Uniden SDS200 Advanced Digital Scanner, Selected for Widespread Deployment Due to its Integrated LAN Port and IP-Based Remote Control Capabilities (Uniden, n.d.b).

As a result of the SAC meeting on September 29, 2025, it was determined that the following sites would be of interest: I-70 near Terre Haute, Indiana, perhaps Brazil, Indiana; I-70 near Richmond, Indiana, perhaps Cambridge City; and I-65 near Louisville, Kentucky. These sites were proposed due to their proximity to sites lacking continued coverage due to limited ITS infrastructure and high volume of incidents on adjacent interstates.

2.5 Digital Scanning Software

Digital Scanning Software controls scanner units. For this research, the software *ProScan* was used. ProScan is designed to interface with digital trunking systems (such as those monitoring the Hoosier SAFE-T Network) to decode and analyze P25 control channel data, as well as conventional systems. It helps achieve the monitoring goals along the interstate corridor by logging and identifying active Talk Groups and Radio IDs in real time across multiple counties. Its features for remote control and data logging enable continuous, unattended collection and analysis of communication patterns, crucial for assessing regional coverage and activity. ProScan also supports administrative tasks like talk group and frequency selection, programming, and storing metadata in an accessible database (Proscan, n.d.).

2.6 Speech-to-Text and Automatic Speech Recognition

STT, also known as ASR, is a technology that analyzes audio signals using mathematical models to derive likely spoken words. As Alharbi et al. (2021) note, speech remains "the most important way for humans to communicate with each other and acquire information." Consequently, this research utilizes ASR to transform unstructured spoken data into a textual format; while the resulting text remains unstructured, it provides a data medium that is significantly more convenient for analysis than raw audio.

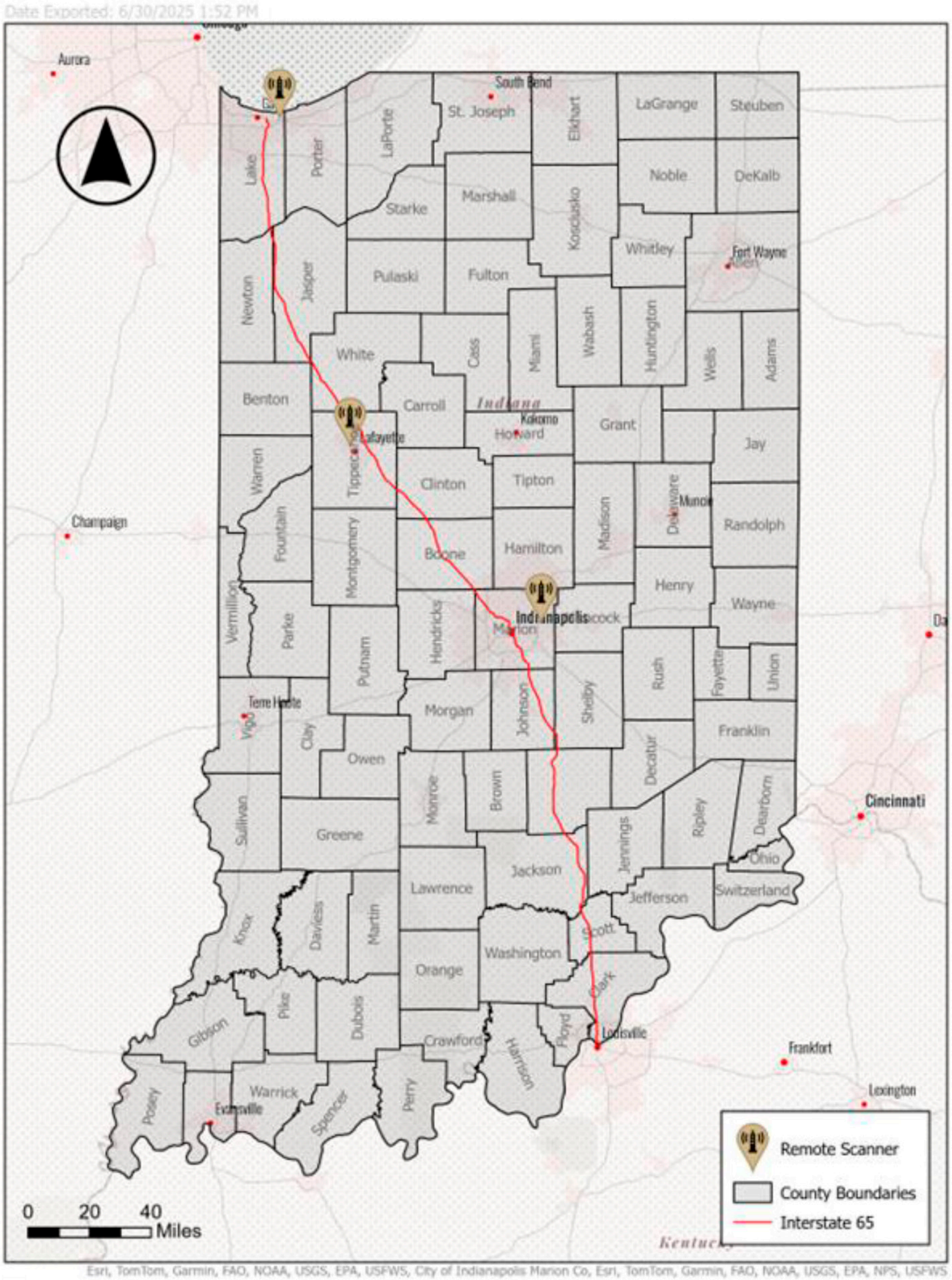


Figure 2.4 Map of Deployed Digital Police Scanner Sites Along the I-65 Corridor in Indiana. Icons Indicate the Physical Location of Listening Hardware at Traffic Management Centers and University Facilities.

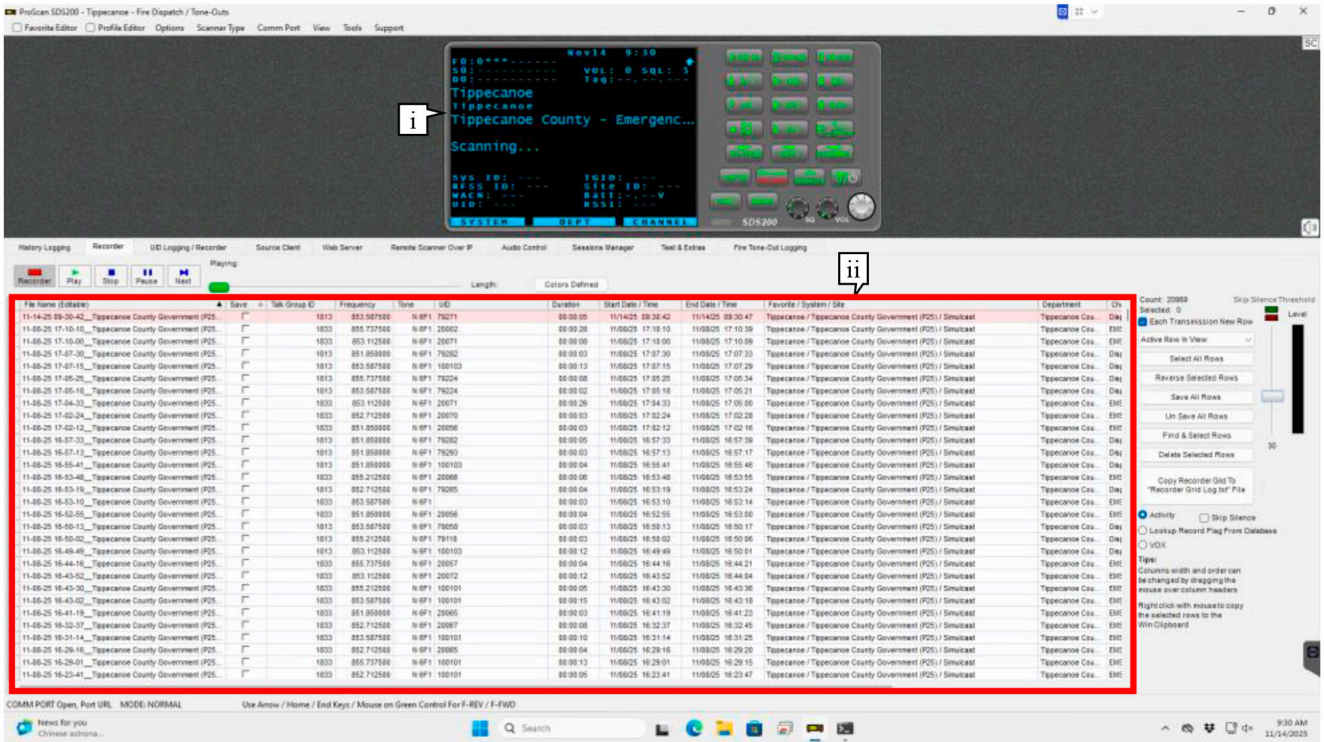


Figure 2.5 Interface of the ProScan Software Used for Remote Scanner Control and Data Logging. Callout i Shows the Virtual Scanner Display Emulation, and Callout ii Highlights the Real-Time Metadata Logging Grid.

2.6.1 Technology Overview

Foundational ASR systems rely on specific modeling techniques to process audio data. Alharbi et al. (2021) identify the Hidden Markov Model (HMM) as a traditional "statistical classification technique for speech recognition". However, the integration of Deep Neural Networks (DNN) has "contributed massively towards the development" of modern recognition systems (Alharbi et al., 2021). Contemporary Artificial Intelligence (AI) applications frequently employ hybrid architectures combining DNNs and HMMs to optimize acoustic and linguistic analysis (Alharbi et al., 2021).

ASR systems are generally categorized by their capabilities regarding utterances, vocabulary size, and speaker dependency. Alharbi et al. (2021) outline four primary speech input classifications:

- Isolated words: Systems that accept single utterances separated by short breaks.
- Connected words: Systems that allow single utterances to run together with marginal pauses.
- Spontaneous speech: Natural speech characterized by imperfections such as filled pauses, repetitions, and false starts.
- Continuous speech: A system where natural voice input is recognized by the machine.

For the purposes of this research, the selected model is designed for continuous speech. Furthermore, the complexity of an ASR system is often defined by its vocabulary capacity. These range from "small-size" vocabularies restricted to tens of words, up to "very large-size" vocabularies capable of processing millions of words (Alharbi et al., 2021).

2.6.2 Measuring Performance

Evaluating the accuracy and utility of ASR systems requires standardized metrics. The primary metric for assessing model performance is the WER. Complementing this is the concept of model confidence. Jiang (2005) describes confidence measures (CM) as a method to evaluate the reliability of recognition results. These measures generally compute a score between 0 and 1 to quantify the model's certainty about its transcription (Jiang, 2005). In this research, these CMs are used based on the native model's confidence output from the selected ASR model, Whisper. For validation, a dataset of 2,907 files (554 min) was manually transcribed and used to assess the WER of different models.

2.6.3 Applications in Literature

The utility of ASR spans a diverse range of sectors. Alharbi et al. (2021) identify key fields of application, including healthcare, education, telecommunications, military, forensics, and law enforcement. Within the domain of public safety and emergency response, Stein et al. (2012) highlight specific use cases for "large-scale emergencies," noting that radio operators are responsible for transcribing information submitted over public safety networks to pass to operations control. Their research suggests systems employing ASR, Speaker Identification (SID), and keyword highlighting can assist in "firefighter broadcast communication" and broader "intelligent resource management," including the detection of abnormal events in audio material (Stein et al., 2012).

3. METHODOLOGY

3.1 System Architecture Overview

Figure 3.1 is the system flow diagram. This figure shows the system/data flow starting in the top left corner at the PSAP, culminating in notifications being sent to the Indiana Department of Transportation (INDOT) TMC. After the PSAP receives their first 9-1-1 call, they typically broadcast dispatch communication over public frequencies. These transmissions can be captured by devices, such as those shown in Figure 2.2 and Figure 2.3, and recorded. Remote systems process those captured audio files and index the audio in a central database. During that process, the transcribed audio is analyzed for keywords. If specific keywords are detected, an automated notification is generated. The last two steps, reception by the TMC and response, are opportunities for further implementation.

3.2 Audio and Metadata Capture

The system's data ingestion relies on synchronizing the physical scanner with ProScan software. Unlike traditional audio logging that records continuous data with silence, this research uses a transmission-based capture protocol. This method stores data relationally, dividing the stream into unstructured audio and structured metadata.

3.2.1 Discrete Transmission Capture

The ProScan software connects to the Uniden SDS200 over a local area network (LAN), monitoring the control channel in real time. When an active transmission on a monitored Talk Group (TG) is detected, a recording is triggered, lasting 3–10 s and stopping when the carrier drops. Each file captures a single vocal exchange, reducing storage and processing needs for the ASR engine.

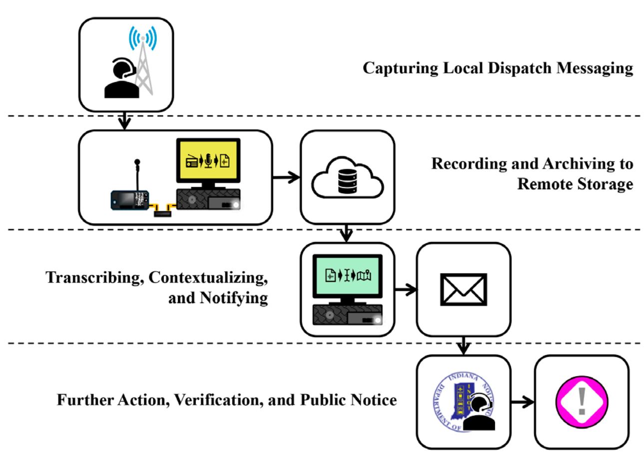


Figure 3.1 High-Level System Architecture Diagram Illustrating the Data Flow From Local Radio Capture (Top) to Cloud-Based Transcription, Keyword Analysis, and Automated Notification Generation (Bottom).

3.2.2 Relational Data Structure

The system separates physical audio files from descriptions while maintaining a strict link through dual logging for efficient querying and analysis.

- Audio: The voice data is saved as a compressed audio file.
- Metadata: The digital headers associated with that transmission are logged into a structured database.

The Audio Object connects to the Metadata Log via a file naming convention. Each audio file is named with a timestamp and unique ID, which serves as the Primary Key in the log, allowing automatic linking of audio and metadata. The unique ID is a derivation of the following attributes:

- Timestamp: The precise date and time (down to the second) the transmission began, serving as the primary temporal index for synchronization with traffic data.
- System: The high-level trunked radio network hosting the call (e.g., "Project Hoosier SAFE-T").
- Department: The specific organizational grouping or agency sub-category (e.g., "ISP District 14" or "Tippecanoe Co Fire").
- Channel: The alphanumeric display name, or "Alpha Tag," associated with the talk group (e.g., "ISP D14 Dispatch").
- Frequency: The specific voice frequency assigned dynamically from the trunked pool for the duration of the call (e.g., 854.1125 MHz).
- RSSI: The Relative Signal Strength Indicator, recorded to monitor reception quality and identify potential coverage gaps at the sensor location.
- Talk Group ID: The unique decimal identifier for the specific agency resource (e.g., "10024"). This serves as the primary key for filtering dispatch versus tactical operations.

3.3 Audio Transcription

This study investigated multiple ASR models; those explored in further depth were:

- OpenAI's Whisper (in multiple sizes): a locally based model.
- Google Speech-to-Text: an API-based, cloud model.

The primary considerations when selecting an appropriate model for analysis are as follows: Model size, Model speed, Word Error Rate. These primary considerations are detailed in the following sections. Additionally, for implementation, additional considerations should be made for cost (device, storage, API services, etc.); it should be noted that the researchers opted to use local models and thus the cost associated with devices capable of such.

3.3.1 Model Selection

The first consideration for model selection was the transcription latency relative to the audio duration (i.e., "how long does it take the model to transcribe an x-second audio file?").

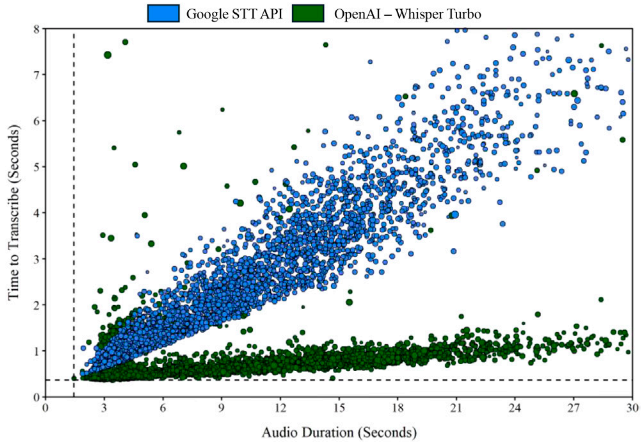


Figure 3.2 Scatter Plot Comparing Transcription Latency (y-axis) Against Audio File Duration (x-axis) for the Cloud-Based Google STT API (Blue) and the Local OpenAI Whisper Turbo Model (Green).

Figure 3.2 is the scatter plot of such a comparison with latency (time to transcribe) on the second axis, and audio duration on the first axis, both in seconds. Recall that the Google model is accessed via an API, while the Whisper model is hosted locally on the machine. Transcription latency comparisons were performed on identical subsets of audio files, 3,673 in total. The two best-performing models tested in each category were the standard Google API model and the Whisper Turbo model.

Of the two, the local model outperformed the API model, but results suggest that model selection based on speed does not significantly affect the performance of the proposed system.

Further, Figure 3.3 shows the distribution (density) of each model's transcription latency. Additionally, two different Whisper models of varying sizes are provided. This is done to analyze whether larger models outperform smaller ones in terms of transcription latency. As before, the results of this comparison were obtained from the same 3,673-file subset.

The results suggest that, across different-sized models (i.e., OpenAI Whisper models), larger model size generally leads to lower transcription latency. Although, Figure 3.3 shows there is difference in the latency between models, those differences have negligible impact on the overall performance needed by TMCs.

3.3.2 Assessment of Word Error Rate

The WER serves as the standard metric for evaluating the accuracy of ASR systems. By quantifying the divergence between a machine-generated hypothesis and a human ground truth, WER provides a normalized measure of performance across different acoustic environments and model architectures.

Mathematically, WER is computed using the Levenshtein distance (Yujian & Bo, 2007), which measures the minimum number of edits required to transform the hypothesized text (model output) into the reference text. The formula is defined as:

$$WER = \frac{D + I + S}{WC} \qquad \text{(Equation 3.1)}$$

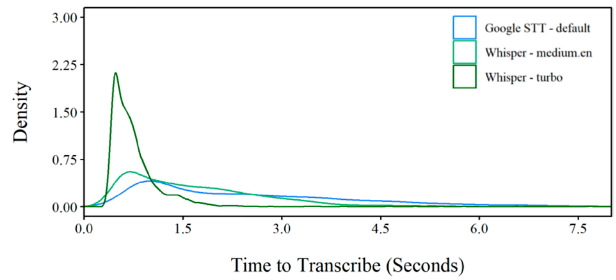


Figure 3.3 Probability Density Function of Transcription Latency for Three Different Model Architectures. The "Whisper-Turbo" Model Demonstrates the Highest Density of Low-Latency Processing.

Where:

- S represents Substitutions: The number of words in the hypothesis that were incorrectly transcribed as different words (e.g., "fire" transcribed as "tire").
- D represents Deletions: The number of words present in the reference but missing from the hypothesis.
- I represents Insertions: The number of words added in the hypothesis that were not present in the reference.
- WC represents the Total Number of Words in the reference (ground truth) transcript.

A lower WER indicates higher accuracy; 0.0 represents a perfect match. It is important to note that because the number of insertions (I) can theoretically exceed the number of words in the reference (N), the WER can exceed 1.0 (or 100%) in cases of severe model hallucination or noise misinterpretation (Wang et al., 2003).

Figure 3.4 illustrates the density distribution of WER across three ASR models: Google STT (Blue), Whisper Medium.en (Light Green), and Whisper Turbo (Dark Green). The aggregate data shown in (a) reveals a distinct bimodal distribution, characterized by two primary clusters: one area of high accuracy (Callout i, low WER) and one area of total failure (Callout ii, WER ≈ 1). To isolate the cause of this polarization, the dataset was filtered by word count.

Figure 3.4b and Figure 3.4c demonstrate that "short" transmissions (1–10 words) are the primary drivers of high error rates. In these instances, the models often lack sufficient semantic context, resulting in a WER of 1: either failing to transcribe the audio entirely or hallucinating (adding words not present). However, as transmission length increases, the opposite is observed. Figure 3.4d and Figure 3.4e show that for "long" transmissions (> 10 words), the density curve shifts dramatically toward the left. This indicates that as models are provided with more context, WER dramatically improves and stabilizes. Notably, all three models perform comparably across these thresholds, though Whisper "Medium.en" shows a slightly higher density of lower WERs at longer durations (e).

Table 3.1 quantifies the inverse relationship between transmission length and transcription error. Descriptive statistics show that including short transmissions (Word Count < 10) skews error rates upward due to bimodal distribution, with extreme Skewness values (e.g., Whisper Turbo at 18.740) indicating a distribution distorted by outliers and short transcriptions. Filtering for substantive communications improves performance: Whisper

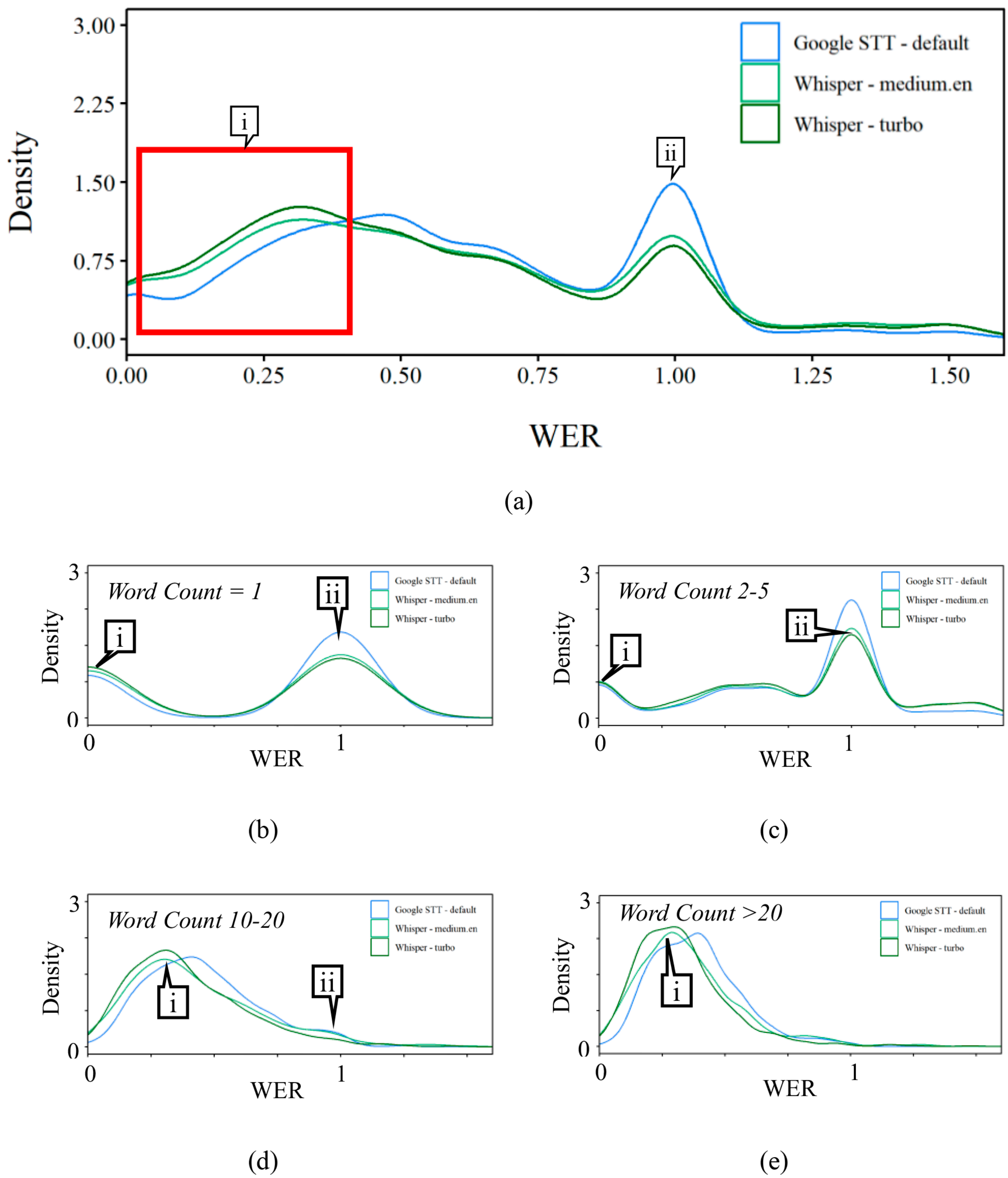


Figure 3.4 Word Error Rate (WER) Density Plots by Transmission Word Count. (a) Aggregate Data Showing a Bimodal Distribution; Callout i Indicates Accurate Transcriptions, While Callout ii Indicates Model Failure. Subplots b–e Illustrate the Stabilization of WER as Word Count Increases From Single Utterances (b) to > 20 Words (e).

TABLE 3.1
Descriptive Statistics of Word Error Rate (WER) Across Three ASR Models by Minimum Word Count Thresholds.

		Word Count		
		>0	>10	>20
Turbo	Mean	0.717	0.399	0.332
	Median	0.500	0.353	0.300
	Skewness	18.740		
Medium.en	Mean	0.825	0.451	0.372
	Median	0.571	0.375	0.320
	Skewness	11.250		
Google STT	Mean	0.638	0.458	0.387
	Median	0.563	0.429	0.375
	Skewness	3.070		

Turbo's Mean WER drops from 0.717 to 0.332 for transmissions of 20 or more words. While Google STT was slightly better with unfiltered data (Mean WER of 0.638), Whisper Turbo achieved the lowest error rates (> 20 words), indicating it provides better fidelity for transcribing longer messages.

Figure 3.5 shows an inverse relationship between message length and transcription accuracy (WER). In the 0–10 words group, most data points have a WER of 1.0, indicating short utterances often cause model failure. As word count increases, data shift toward lower error values, with 11–20 words showing a peak WER of 0.35. In the 20+ words group, density peaks at a WER of about 0.30, confirming that longer messages improve accuracy by providing more semantic context, helping the model resolve phonemes and generate correct text.

3.3.3 Assessment of Model Confidence as a Quality Control Measure

To utilize model confidence as an automated Quality Assurance/Quality Control (QA/QC) mechanism, the relationship between message length (word count) and model

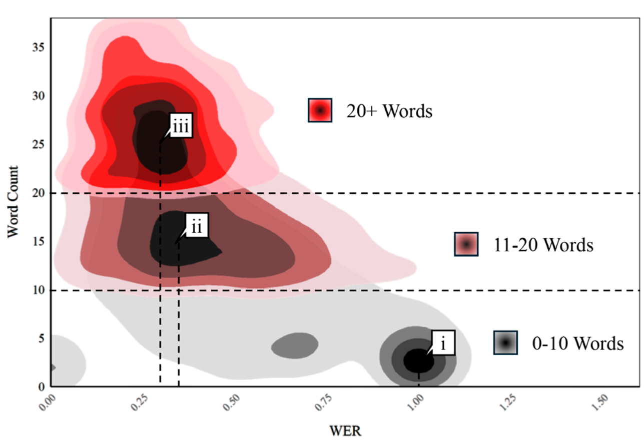


Figure 3.5 Bivariate Density Plot of Word Count Versus WER. Callout (i) Shows High Error Rates for Short Utterances (0–10 words). The Density Center Shifts to Lower Error Rates for Medium (ii) and Long (iii) Transmissions, Confirming the Inverse Relationship Between Message Length and Error.

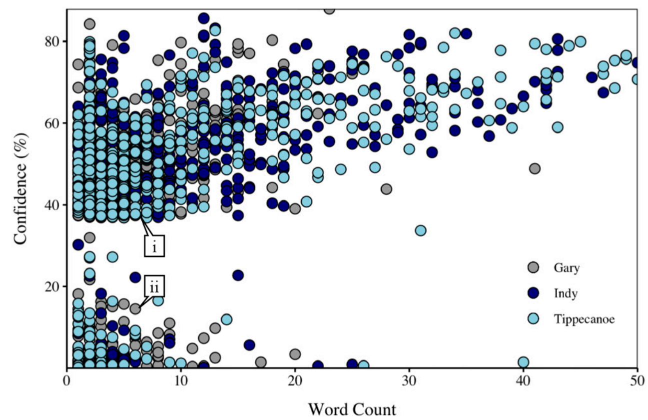


Figure 3.6 Scatter Plot of Model Confidence (%) Versus Word Count. Callout i Represents High-Confidence Short Commands, While Callout ii Represents Low-Confidence Noise. The Variance Decreases as Word Count Increases, Regardless of the Monitoring Site.

confidence was analyzed. Figure 3.6 illustrates the distribution of confidence scores against word count. The plot reveals a critical bifurcation in short-duration transmissions (0–10 words), highlighted by two distinct clusters:

- Cluster i (High Confidence): Short, clear procedural commands (e.g., "10-4," "Copy that").
- Cluster ii (Low Confidence): Static bursts, digital noise, or unintelligible fragments.

This bimodal distribution confirms that raw audio ingestion captures a significant volume of noise. However, as the message length increases, the variance decreases. Figure 3.6 reinforces this trend across all three sites, showing that long-form narratives are more likely to score higher in confidence.

The analysis indicates that ASR performance is not uniform across the state but is heavily influenced by local environments and their operational protocols. Table 3.2 and Figure 3.7 present a statistical comparison of confidence scores across the three study sites: Gary TMC, Indianapolis TMC, and Tippecanoe. The Gary TMC site exhibits the lowest performance metrics, with a Median confidence of 46.36% and a lower quartile (Q1) of 38.21%. In contrast, both Indianapolis TMC and Tippecanoe show higher reliability, with medians exceeding 51%. This discrepancy suggests that the "Gary" dataset contains a higher frequency of nonspeech noise, static, or short-burst transmissions that the model struggles to resolve. This finding emphasizes that CM is location-dependent; a filtering threshold appropriate for Tippecanoe may be too aggressive for Gary, potentially discarding valid but lower-quality audio.

TABLE 3.2
Statistical Summary of Model Confidence Scores Across the Gary, Indy, and Tippecanoe Monitoring Sites, Highlighting Location-Dependent Performance Variations.

	Minimum	Q1	Median	Q3	Maximum
Gary TMC	4.79	38.21	46.36	54.61	87.94
Indy TMC	4.58	43.82	52.33	62.05	85.63
Tippecanoe	17.96	40.97	51.03	62.65	82.64

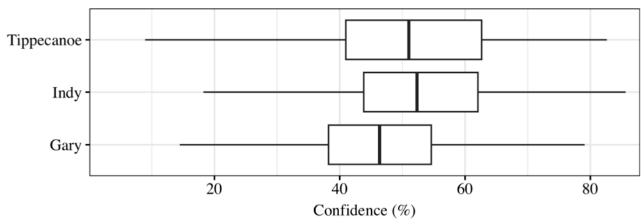


Figure 3.7 Box-and-Whisker Plots Comparing the Distribution of Confidence Scores by Site.

The use of model confidence as a noise filter is further visualized in the density evolution plots (Figure 3.8). In Figure 3.8a, the distribution is skewed by low-confidence noise (<40%). As the message length increases from >5 to >20 words (Figure 3.8b, e), the low-confidence curve flattens. This suggests that filtering by minimum word count and confidence effectively reduces noise, discarding radio noise while keeping important dispatches for incident detection.

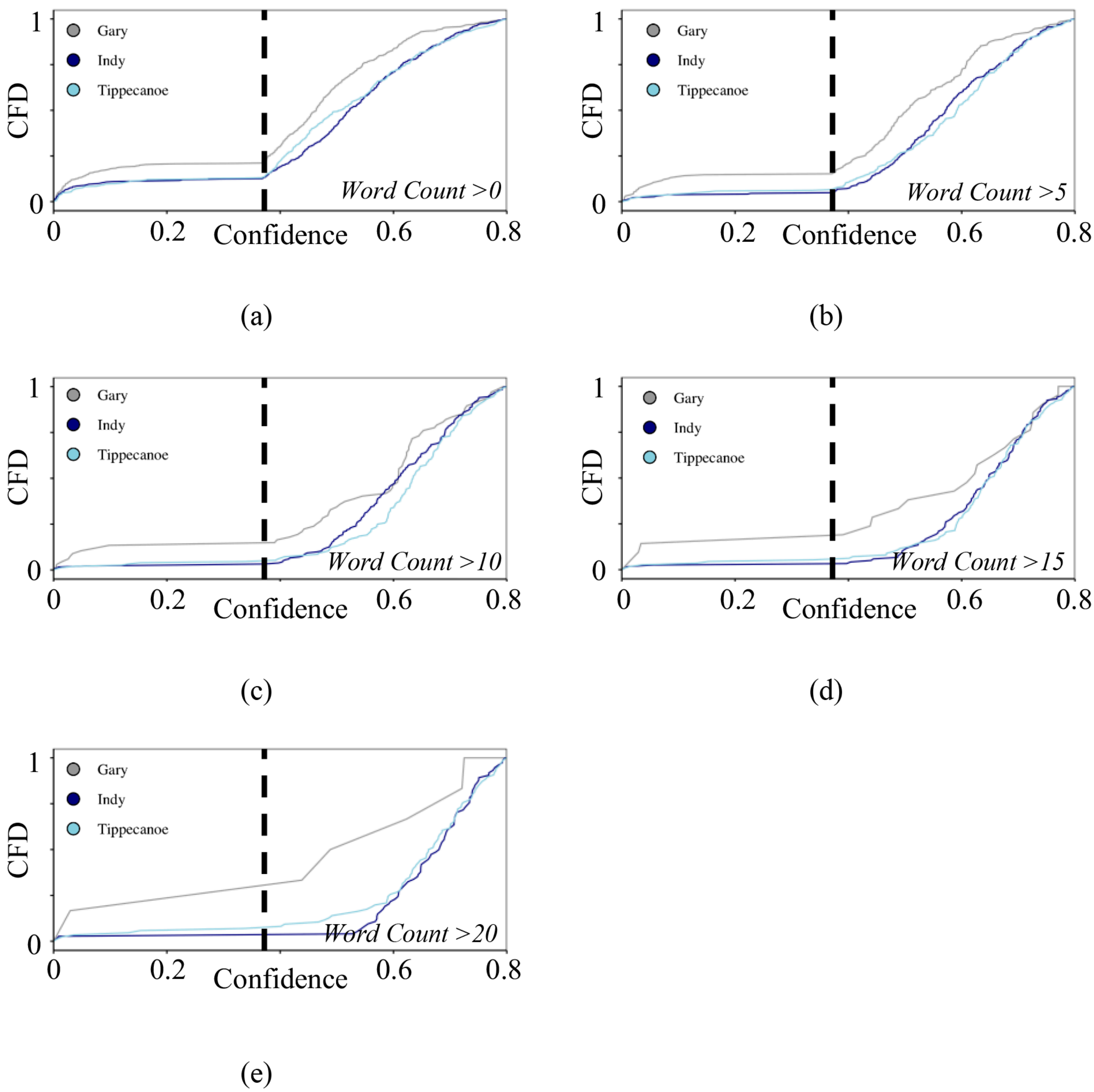


Figure 3.8 Cumulative Density Functions (CDF) of Model Confidence Scores at Varying Word Count Thresholds (a–e). The Flattening of the Curve in Plot (e) Indicates the Successful Filtration of Low-Confidence Noise for Longer Transmissions.

3.4 Keyword Analysis

From the transcribed audio (a string of text), Regular Expressions (RegEx) are used in combination with simple logic to extract instances of keywords belonging to one of three groups, as shown in Table 3.3 below.

- "Interstate" keywords refer to specific references to an interstate. For the study location, the only interstate of interest is I-65; hence, the examples are limited to I-65 (including the local phrase "the I"). "Interstate" keywords serve as the primary indicator of an incident occurring on the interstate.
- "Location" keywords refer to a particular location along the interstate. The desired "location" is the mile marker (MM; e.g., "168.5"); however, sufficient information is contained within specific exits/entrances to infer most locations along a given route. Additional context, including which lane or shoulder, is also considered valuable as it will expedite the locating time for a TMC operator.
- "Incident" keywords refer to additional context about the specific incident (e.g., "fire" or "crash"). These keywords help indicate the type and severity of the incident.

In addition to the common phrases presented in the table, coded communications are also included. Commonly referred to as "ten-codes," these phrases convey information such as units on scene, scenes being cleared, or other frequently used messages.

TABLE 3.3
Taxonomy of Keyword Extraction Groups Used for Logic-Based Filtering, Including Specific Examples for the I-65 Study Corridor.

Keyword Group	Examples	Description
Interstate	"interstate," "the i," "i-65," "65"	Common and local phrase extraction for I-65 running through the study area.
Location	"168.5 northbound," "155 mile marker," "exit," "left shoulder"	The locational information contained within the call. Information includes interchange features, linear reference markers, and directionalities.
Incident	"10-50/51," "PI (Personal Injury)," "crash," "collision," "wreck," "fire," "medical"	Common coded communications and contextual information about the incident contained within the call.

A simple logical structure is used to determine if the keywords present in any given audio message warrant a notification to relevant parties. Figure 3.9 depicts this structure.

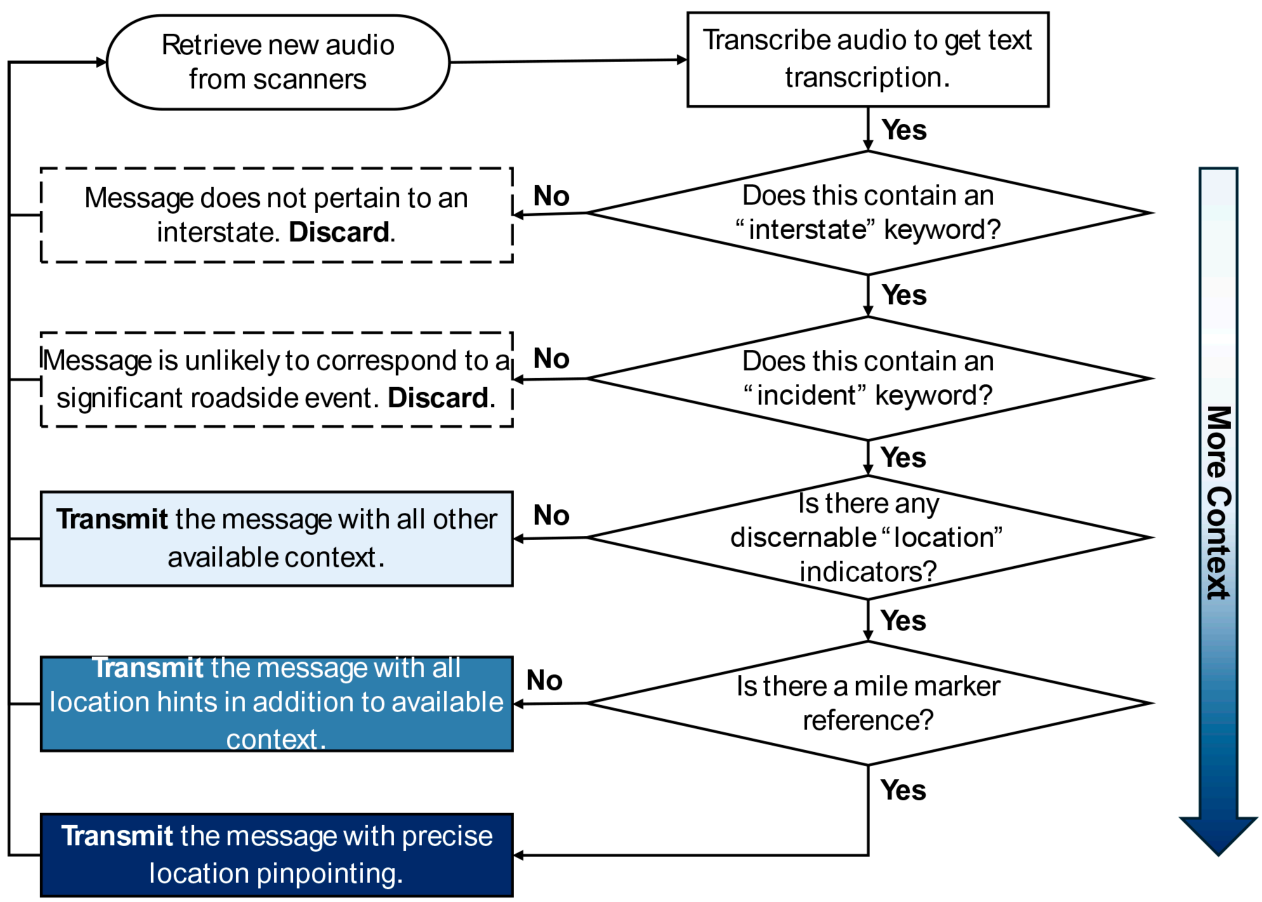
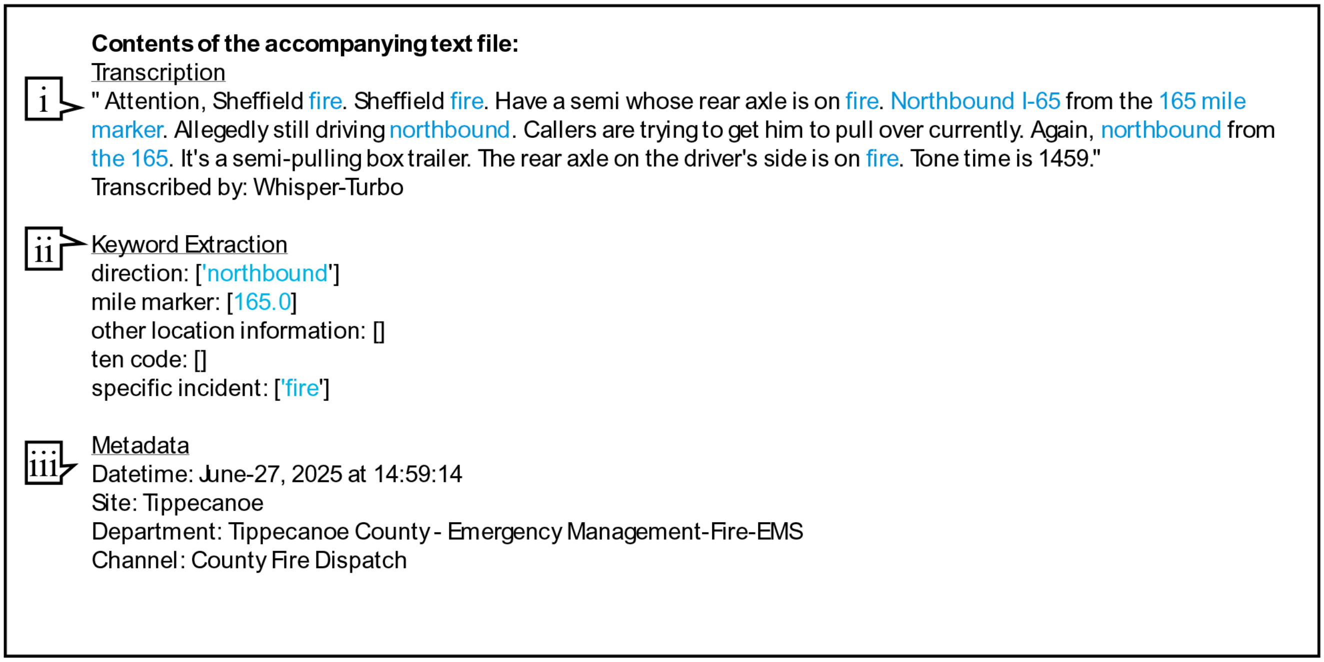


Figure 3.9 Logic Flow Diagram Illustrating the Hierarchical Keyword Checking Process Used to Determine if a Transcribed Message Contains Sufficient "Interstate" and "Incident" Context to Warrant an Automated Notification.



Contents of the accompanying text file:

Transcription

" Attention, Sheffield fire. Sheffield fire. Have a semi whose rear axle is on fire. Northbound I-65 from the 165 mile marker. Allegedly still driving northbound. Callers are trying to get him to pull over currently. Again, northbound from the 165. It's a semi-pulling box trailer. The rear axle on the driver's side is on fire. Tone time is 1459."

Transcribed by: Whisper-Turbo

Keyword Extraction

direction: ['northbound']

mile marker: [165.0]

other location information: []

ten code: []

specific incident: ['fire']

Metadata

Datetime: June-27, 2025 at 14:59:14

Site: Tippecanoe

Department: Tippecanoe County - Emergency Management-Fire-EMS

Channel: County Fire Dispatch

Figure 3.10 Example of an Automated Email Notification Generated by the System. Callout i Shows the Raw ASR Transcription; Callout ii Displays the Extracted Structured Keywords; Callout iii Provides the Technical Metadata (Timestamp and Talk Group). Resource Available at https://youtu.be/pJouXVLzutI.

Starting in the new audio section, transcriptions are checked for an “interstate” reference to determine whether it pertains to an interstate. Any additional context is then pulled from the “incident” keywords. Finally, “location” indicators are extracted, with preference given to mile marker utterances. For a message to be considered “important enough” to be sent, it must contain an “interstate” reference and an “incident” reference; “location” indicators are not strictly required as per this criterion.

3.5 Automated Notification

Figure 3.10 is the automated email notification corresponding to the incident in this case study. It was generated entirely autonomously using Microsoft Power Automate, Microsoft SharePoint, Python, and the recording devices and software described in the Capturing Audio Section and corresponds to the incident described in Figure 3.10.

Figure 3.10 is an example of an ideal transcription. Callout iii is the transcription generated from the ASR engine. Blue words indicate identified keyword indicators. In the observed audio clips, repetition of information is common in dispatch transmissions. Thus, Figure 3.9 is used to identify the “best” indicators and present them in the section denoted by Callout iv. Finally, the recording contains useful information within its metadata. Callout v is data captured during the recording process, including the date and time of capture, as well as the site, department, and transmission channel. Information from Callouts iv and v allows for the localization of transmissions to a specific geographic area, and in ideal cases, a particular location along an interstate.

The median latency between the time an audio file was recorded (Callout i) and the time the email was received (Callout ii) was approximately 52 s. This suggests that the proposed methodology relays important information quickly enough to prevent it from becoming outdated.

3.5.1 Notification Latency Analysis

To assess the operational viability of the system for real-time TIM, the end-to-end processing speed was evaluated. Notification latency is the elapsed time from the start of radio transmission to the timestamp of the alert delivery (email).

Figure 3.11 presents the temporal distribution of n = 270 validated notification events captured between June 20 and October 18. The system demonstrates consistent temporal stability throughout the deployment period, maintaining an average latency of approximately 59 s (00:00:59). To quantify the system’s reliability, a One-sided Upper Tolerance Limit (UTL) was calculated. The analysis indicates with 99% confidence that 90% of all notifications will be processed and delivered within 96 s (00:01:36).

The probabilistic performance is further detailed in the cumulative distribution function (CDF) shown in Figure 3.12. The steep ascent of the curve highlights the efficiency of the transcription and keyword-matching pipeline, with the median latency at approximately 59 s. Furthermore, the distribution shows that 90% of all detected incidents triggered a notification within 90 seconds. The outliers exceeding 120 s represent a negligible fraction of the total operational dataset and are likely due to network latency in the automated triggering mechanism.

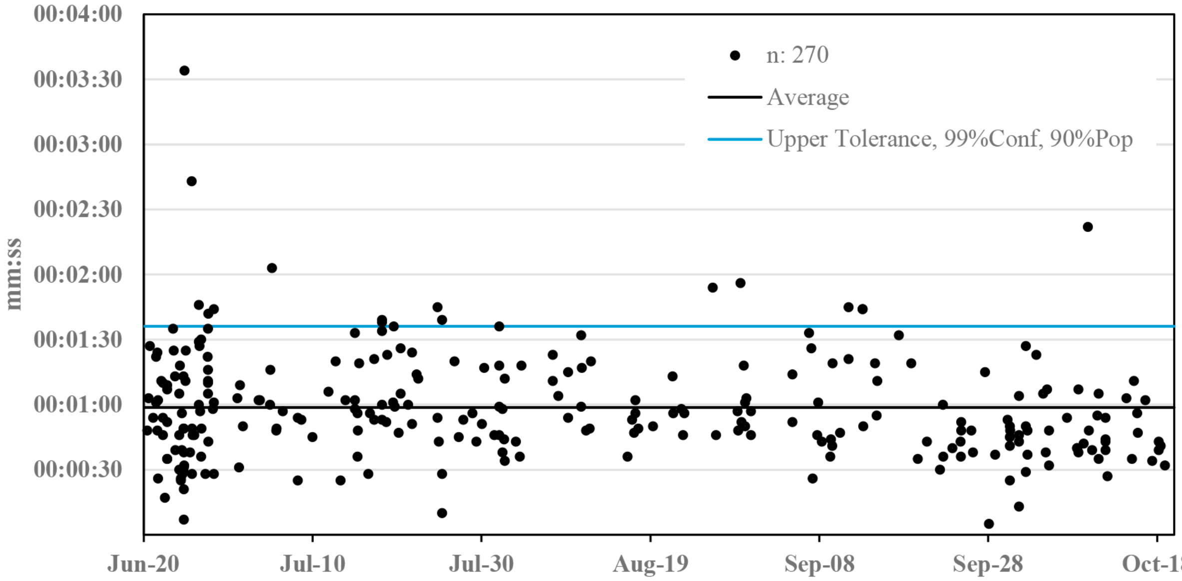


Figure 3.11 Longitudinal Scatter Plot of System Latency (Time From Transmission End to Notification Delivery). The Blue Line Indicates the Upper Tolerance Limit (96 seconds) at a 99% Confidence Level.

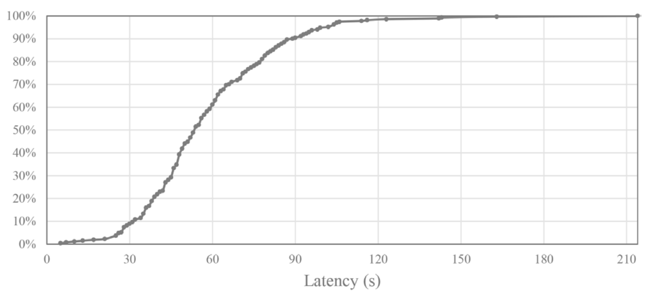


Figure 3.12 Cumulative Distribution Function (CDF) of Notification Latency (n = 270). The Curve Indicates That 90% of All Notifications Are Processed and Delivered Within 90 s.

3.6 Relational Database Structure

To transform unstructured audio into actionable intelligence, the system uses a normalized relational database schema, either in the cloud (Google BigQuery) or, with some modifications, locally. This structure is designed to decouple metadata, transcription, and geospatial analysis into distinct, interrelated tables, ensuring scalability and efficient queries. A generalized overview is given by Figure 3.13. There are likely attributes that have been overlooked (e.g., transcription length), and thus, the format should be strongly considered before implementation. The researchers here leave the "properties" attribute to serve as an "overflow" JSON to capture these attributes.

3.6.1. Audio Files

This table acts as the main reference point for the entire dataset. It directly catalogs the raw MP3 files stored in the cloud storage bucket (public_safety_audio_data). Its primary role is to store metadata derived from scanner hardware and ProScan software. Key details include the spatiotemporal context of each transmission, such as system_site (e.g., "Project Hoosier SAFE-T"), department (e.g., "Tippecanoe Fire"), and talk_group. Additionally, it records the physical URI (or location) linking to the object storage. To enhance performance for time-series analyses, the table is partitioned daily based on the datetime field.

3.6.2. Model Transcriptions

This table captures the output from the ASR engine. It connects to the audio_files table via a Many-to-One relationship via file_id, enabling multiple transcriptions (e.g., from different models or human validation) to be linked to a single audio file. Its main purpose is to store the raw transcribed text (transcribed_text) along with performance metrics. Important attributes include the transcriber name (e.g., "Whisper-Large"), processing latency, model confidence, WER, and an is_asr boolean that indicates whether the text is machine-generated or human-verified.

3.6.3. Keyword Analysis

This table serves as the system's semantic filter, linked to audio_transcripts via a one-to-one relationship via transcript_id. It holds structured data derived from the raw text. Its primary role is to store boolean indicators and JSON arrays for specific keyword detections as defined by the system's regex library. The data is categorized into three logical groups: 'interstate' for highway name detections (e.g., "I-65"), 'location' for spatial references like mile markers and travel directions, and 'incident' for identifying event types such as "collision," "fire," or "10-50."

3.6.4. Linear Reference Match

This table links semantic analysis with physical geography by connecting the extracted keyword data (audio_kwi) to a

audio_files	
id	int
uri	string
filename	string
datetime	timestamp
scanner	string
duration	float
system_stie	string
department	string
channel	string
talk_group	string
tone	string
frequency	float
properties	json

audio_transcripts	
id	int
file_id	int
transcribed_text	string
datetime_transscribed_	timestamp
transcriber	string
latency	float
is_asr	bool
properties	JSON

audio_kwi	
transcript_id	int
datetime_transcribed	timestamp
interstate	bool
general_interstate_reference	json
specific_interstate_reference	json
location	bool
direction	json
mile_marker	json
other_location	json
incident	bool
ten_code	json
specific_incident	json

smart_poly (lrs)	
id	inte
route	string
mm	float
state	string
active_date	timestamp
deactive_date	timestamp
geog	geography

audio_location	
id	int
poly_id	int
transcript_id	int
route	string
mm	float
state	string
datetime	timestamp

Figure 3.13 The Relational Database Schema Illustrates the Linkages Between Raw Audio Files, Transcriptions, Keyword Analysis, and Geospatial Reference Tables.

linear-referencing system (LRS) through a Many-to-One relationship. Its main function is to translate text-based location descriptions into accurate linear references. Key features include normalizing valid location hits into a standard LRS format (route [directional], mm, state). By joining with smart_polys.id, this table enables visualization of the textual audio data on a map and facilitates spatial integration with other traffic datasets.

4. CASE STUDY

This section presents a case study and further information on system integration. The contents of this section are heavily derived from those published in the following open-access journal (Gartner et al., 2025):

> Gartner, C. M., Vajpayee, V., Desai, J., & Bullock, D. M. (2025). Automatic speech recognition of public safety radio communications for interstate incident detection and notification. *Smart Cities*, *8*(5), 157. https://doi.org/10.3390/smartcities8050157

4.1 Study Location

The study location surrounding the Purdue University site is given in Figure 4.1. For this site, four counties were monitored,

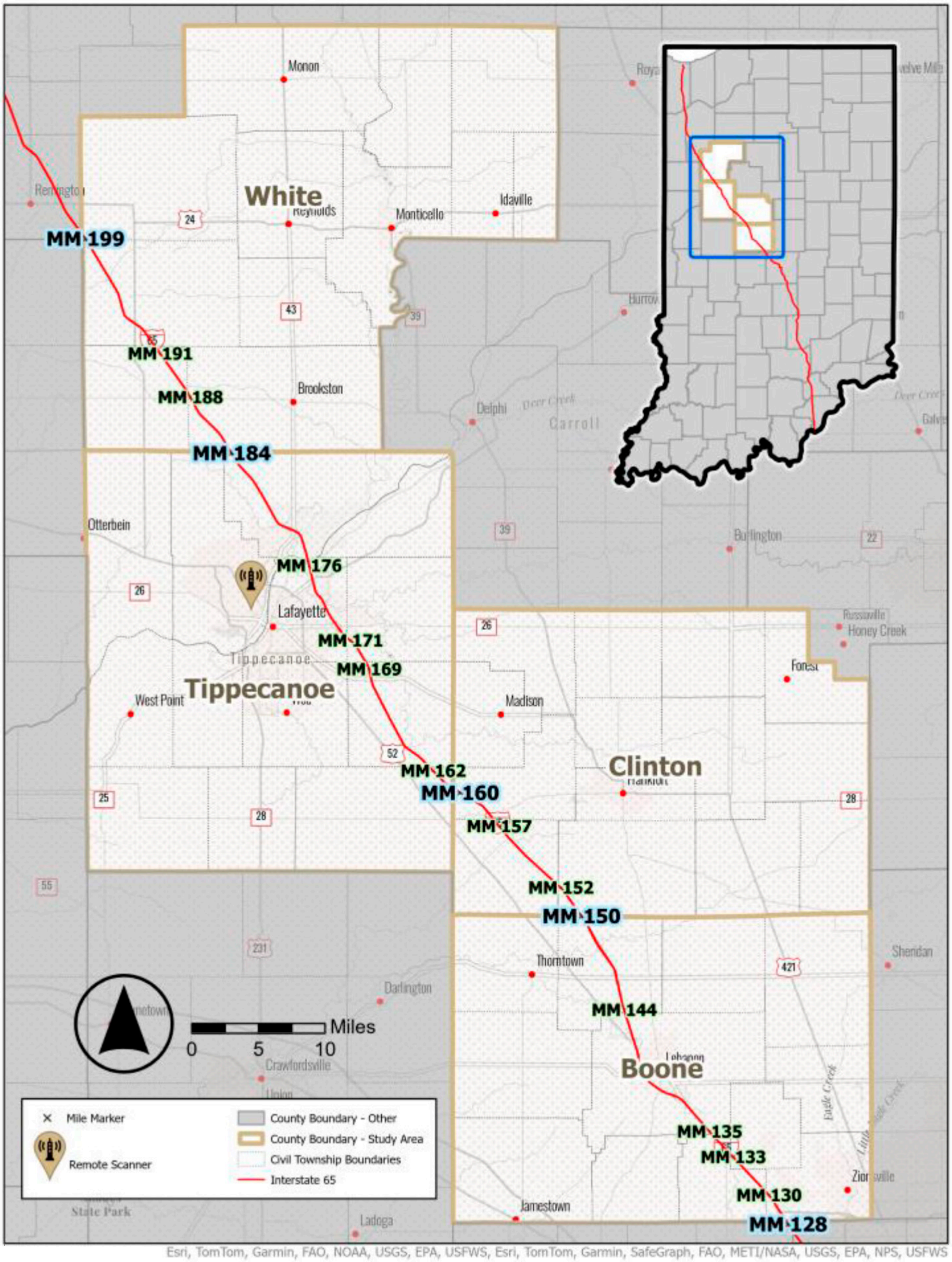


Figure 4.1 Detailed Map of the Purdue University Study Site Coverage Area. The Red Line Indicates the I-65 Corridor; Labels Denote Mile Markers (MM). Colored Boundaries Indicate County and Township Jurisdictions Relevant to Frequency Selection.

TABLE 4.1
Inventory of Monitored Talk Groups and Conventional Frequencies for the Purdue Study Site, Mapped to Their Approximate Linear Reference Coverage on I-65.

Sys.	Department	Channel	Type		MM	Conv	Trun
SAFE-T	State Police Region 1—Lowell	ISP D14 Lafayette Disp.	P	🚨	150–199		✓
		ISP D14 Lafayette Multigroup	P	🚨			✓
		ISP D14 Lafayette Ops. [1–3]	P	🚨			✓
SAFE-T.	White Co.	Co. Fire Disp.	F	🔥	184–199	✓	✓
		Co. (FD) Ops. [1–7]	F	🔥			✓
		Sheriff Disp.	P	🚨			✓
		Co. EMS Ops. 2	E	🚑			✓
Tippecanoe Co.	Tippecanoe Co. Emergency Ambulance Service	EMS Disp. and Alt.	E	🚑	160–184	✓	✓
	Tippecanoe Co. Emergency Mngmt.	Disp.	G	🏛			✓
	Tippecanoe Co. S.O.	Disp.	P	🚨			✓
	Tippecanoe Co. Emergency Mngmt./Fire/EMS	Fire Disp./Tone-Outs	F	🔥		✓	
	Tippecanoe Co. Fire	Disp.	F	🔥			✓
		Fire Mutual Response Area	F	🔥			
		Battleground FD	F	🔥	176–184		✓
	Lafayette FD	Sta. 6 Ops.	F	🔥	174–176		✓
		Sta. 5 Ops.	F	🔥	172–174		✓
		Sta. 9 Ops.	F	🔥	170–172		✓
		Sta. [2,7] Ops.	F	🔥	*		✓
	Tippecanoe Co. Fire	Sheffield FD	F	🔥	160–170		✓
		Clark Hill FD	F	🔥	*		✓
SAFE-T	Clinton Co.	Co. Fire Disp.	F	🔥	150–160	✓	✓
		Co. Law Enforcement Disp. [1,2]	P	🚨		✓	✓
SAFE-T	Boone Co.	Co. EMA/EMS Ops.	E	🚑	128–150	✓	✓
		Co. Fire Disp.	F	🔥		✓	✓
		Co. Fire Ops. [1–5]	F	🔥			✓
		Sheriff Disp. [1,2]	P	🚨			✓

Note: (F) 🔥: Fire; (P) 🚨: Law Enforcement; (E) 🚑: EMS; (G) 🏛: Government/Other; * Mutual Aide Only

each intersected by I-65, a primary north–south corridor between Indianapolis, Indiana, and Chicago, Illinois. The site successfully monitored White, Tippecanoe, Clinton, and Boone counties, providing a total of 71 linear miles of coverage along I-65. Figure 4.1 shows the MM or LRS and the relative location at county borders (blue) and townships (green). These boundaries are used to derive the spectrum of frequencies of interest, in which the assigned talk groups have some jurisdiction over the route of interest.

4.2 Study Talk Groups

Channels and talk groups were selected that were most likely to capture a 9-1-1 dispatch to an interstate incident. Selected channels are presented below in Table 4.1 with the County/Radio system to which they belong; the department, channel, and type; the approximate MM; and whether the system is conventional (conv.) or trunked (trun.). With the addition of 9-1-1 dispatch channels, a select group of operations (ops) channels was included, as was the state police for the region. Since this area of the state still utilizes a combination of legacy conventional fixed frequencies and modern digital trunked systems, it was essential to monitor both to provide comprehensive coverage. Once the appropriate systems, departments, and channels were identified, the Uniden Sentinel Software could be used to program the digital scanner.

4.3 Tractor-Trailer Fire Case

A case study is presented to demonstrate the vital information that 9-1-1 dispatches contain. Figure 4.2 depicts a tractor-trailer fire occurring on June 27, 2025, on northbound I-65 near the MM 165. The fire-damaged vehicle, first responders, and their

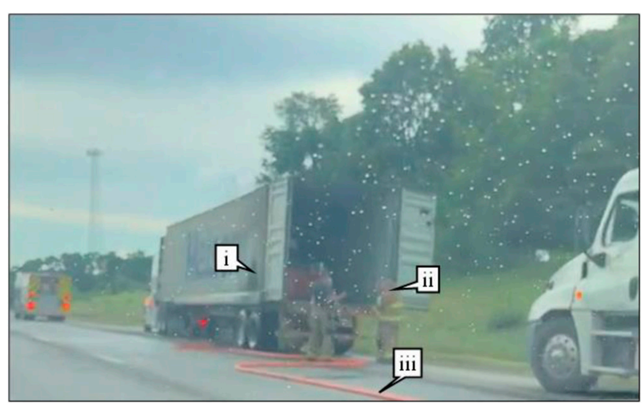


Figure 4.2 Ground-Level View of the June 27, 2025, Tractor-Trailer Fire on I-65. Callout i Identifies the Damaged Trailer; Callout ii Indicates Fire Personnel; Callout iii Shows Fire Hoses Obstructing the Roadway. (Photo: Kameron Pelfree).

TABLE 4.2
Chronological Log of Transcribed Radio Communications for the June 27 Incident. The “Action” Column Summarizes the Operational Context Derived From the Audio Available at https://youtu.be/pJouXVLzutI.

Time	Channel	Transcription	Action	WER
14:59:14	Co. Dispatch	Attention Sheffield fire. Sheffield fire. Have a semi whose rear axle is on fire. Northbound I-65 from the 165 mile marker. Allegedly still driving northbound. Callers are trying to get him to pull over currently. Again northbound from the 165. It’s a semi-pulling [a] box trailer. Rear axle on the driver’s side is on fire. Tone time is 1459.	Dispatch broadcast [Automated Notification Trigger]	0.02
15:06:12	Co. Dispatch	204 [in] route [with] 2. 204.	Units in route	0.5
15:08:08	Sheffield FD	204 on Sheffield.		0.0
15:08:20	Sheffield FD	To the [interstate] until we find out where we’re going.		0.0
15:09:40	Co. Dispatch	I’m on the phone with the semi-driver now. He’s pulled over. Looks like near the [Wyandot] overpass. [Advising] tandems are still on fire.	Location Confirmed	0.09
15:10:05	Co. Dispatch	204 is clear [Wyandot] overpass. 206 go ahead and make access to the 65 now.		0.13
15:11:14	Co. Sheriff’s Office	23. Go ahead. Do you know where the semi-drivers [at]? Traffic’s already backed up to the 162. He’s right at the [Wyandot] overpass. Sorry.	Driver Located	0.13
15:13:20	Co. Dispatch	Container fire		0.50
15:13:40	Co. Dispatch	Tippe. Fire 220. 204 is arriving on scene working fire. 220 will have IC on 65. Northbound they’re going to be shut down by ISP. [].	Units on Scene	0.08
15:14:53	Sheffield FD	220 40 on Sheffield. We have three on station. What do you want?		0.00
15:16:11	Co. Dispatch	206 where do you want me to stage? Stay high [and] forward. Clear.		0.08
15:16:29	Sheffield FD	2040 on Sheffield. You want to throw a one or a tanker?		0.00
15:16:43	Co. Dispatch	We got a good knockdown on it.	Fire out	0.13
15:22:08	Co. Dispatch	In the northbound lanes we’ve got it completely shut down for now.	NB closure	0.00
15:22:26	Sheffield FD	Be southbound and northbound.	SB closure	0.00
15:23:42	Co. Dispatch	One when you aarrive justgo ahead and get turned around and face back north in front of the... Come on.		0.00
15:24:05	Co. Dispatch	2-0-1.		0.00
15:28:08	Co. SO	We can open up the left lane. Sure.	Reopening [NB] lane	0.00
15:28:39	Co. Dispatch	Go ahead. Are they able to open up both lanes or just northbound or southbound? You’ve got a definite block in northbound. He can let the high-speed side go. We’ll keep the low speeds shut down still.		0.00
15:28:58	Co. Dispatch	Clear. Further southbound slowing is just slow for some reason. Been like that for almost an hour now.	Confirmation SB open	0.00
15:29:16	Co. SO	Got it. We’ll be clear. Clear.		0.00
15:45:03	Co. Dispatch	Tippe. fire 220. 220. Command [survey] on I-65, all [seems to] be cleared in the scene and service. I’ll move the pressure. Thank you.	Units’ clear scene	0.09

equipment are denoted by Callouts i, ii, and iii, respectively. This incident resulted in a full closure of both directions of travel and an estimated 10-mi queue in the northbound travel lanes. Despite creating a 10-mi queue, this incident was never reported in the state Motorist Information System, most likely due to its rural location.

Table 4.2 displays the 9-1-1 dispatch and operations communications relevant to the incident shown in Figure 4.2. The caption’s hyperlink links to the audio recording of the dispatch. The initial dispatch was recorded at 14:59:14 local time and transcribed with a WER of 2%. It includes details on “who” (Sheffield Fire [Tippecanoe County]), “what” (a semi-fire), “when” (14:59), and “where” (I-65 northbound from the 165 MM). This transmission provides all essential information for a TMC operator to stay informed and monitor travel impacts, prompting motorist alerts via the MIS. While the initial message offers comprehensive details, subsequent operation conversations add more specific information, such as location confirmation at 15:09 for a particular exit, lane closures at 15:13, lane openings at 15:28, and scene clearance at 15:45.

Spatiotemporal traffic speed plots (heatmaps) present linearly referenced traffic speed trajectories, color-coded by speed (Sakhare et al., 2024). Heatmaps are important tools for

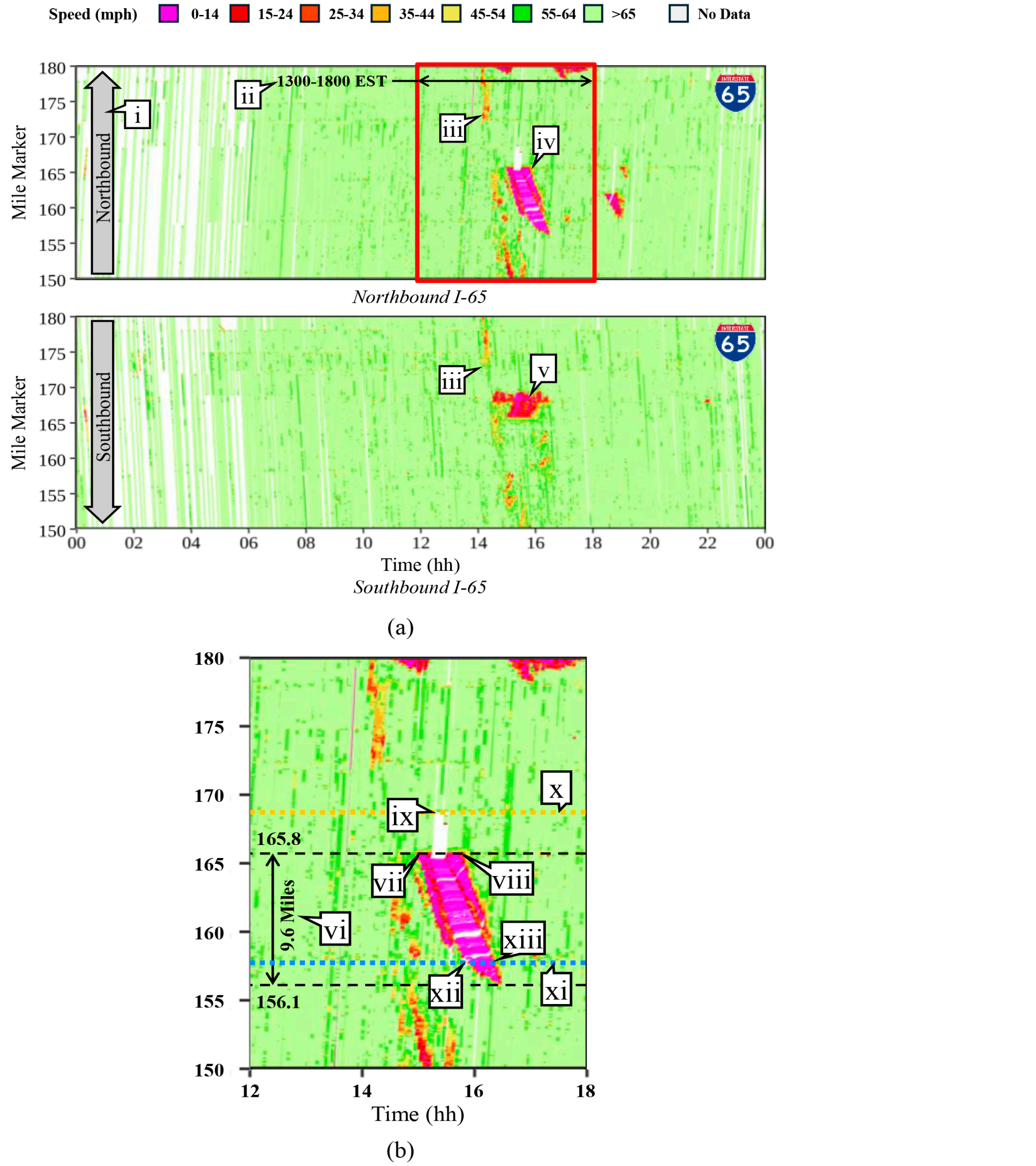
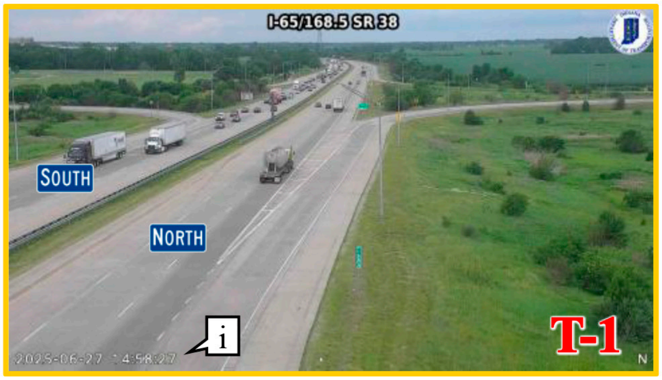


Figure 4.3 Spatiotemporal Traffic Speed Heatmaps for I-65. (a) Overview of Northbound and Southbound Lanes; the Red Box (iv) Highlights the Incident Congestion. (b) Detailed View of the Incident Timeline.

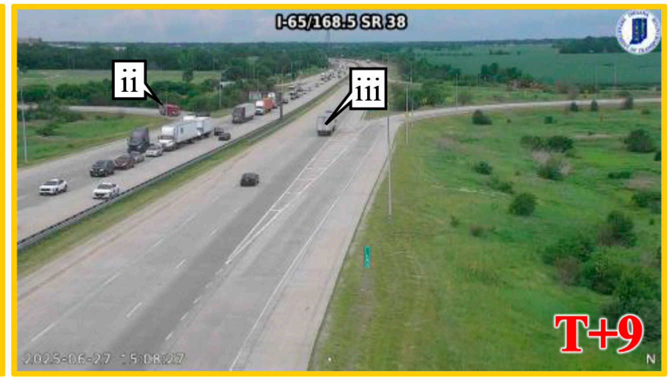
both real-time monitoring of freeway conditions and detailed "after-action" analysis of roadway incidents. Figure 4.3 is the heatmap corresponding to the incident in Figure 4.2.

Figure 4.3a presents the heatmaps for both northbound and southbound directions. Callout i indicates the direction of travel; Callout ii indicates the period of interest corresponding to the case study. Prior to the incident, a storm affected travel behavior in the same segment. This is observed as 'patches' on the heatmap and is noted by Callout iii for both directions of travel. Callout iv indicates reduced speeds and queuing due to the incident blocking traffic in the northbound lanes. Callout v indicates reduced speeds and queuing due to the incident response in the southbound lanes. Figure 4.3b isolates the primary incident in the northbound lanes. From the heatmap, the following observations are made about the incident, and verified through the events of Table 4.2:

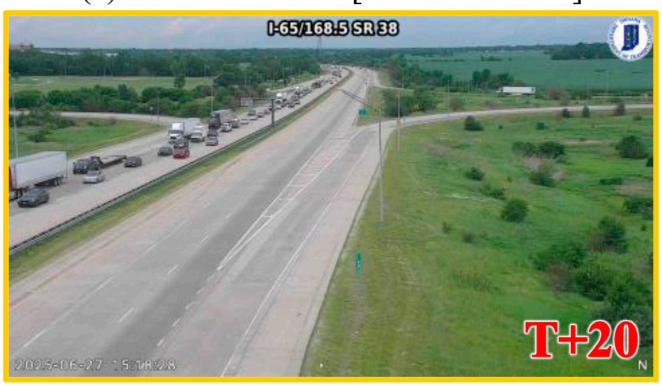
- The incident occurred at MM 165.8, northbound (Callout vii).
- The longest queue reached MM 156.1, an estimated 9.6-mi queue (Callout vi).
- The incident began at 14:59 (Callout vii).
- There is a short-term northbound closure at approximately 15:22 near MM 165.8 (Callout ix). The transcribed text for this can be seen in Table 4.2 at 15:22:08, and the lack of vehicles passing through this area is shown as the white area on the northbound heatmap at this approximate time.
- The incident cleared at 15:45 (Callout viii).



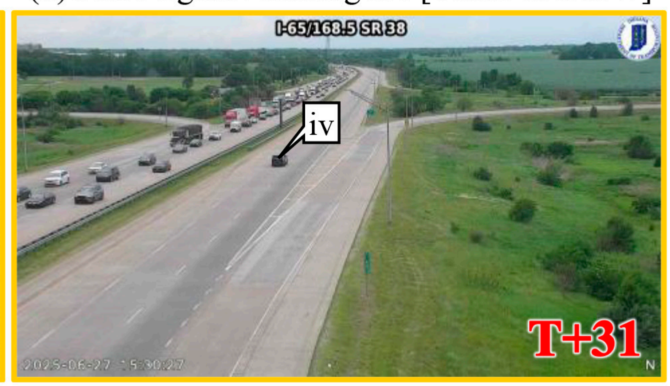
(a) Incident Occurs [Time: 14:58:27]



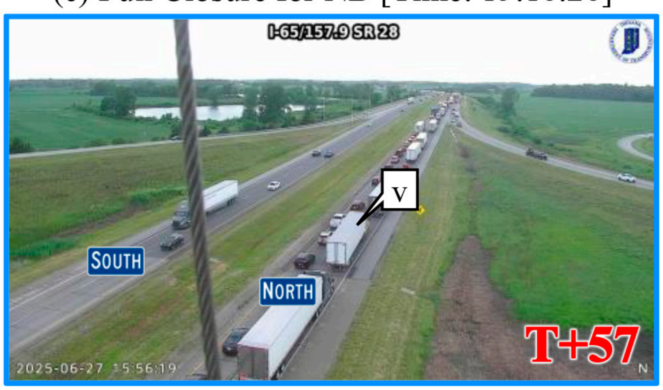
(b) Fire Engine Entering SB [Time: 15:08:27]



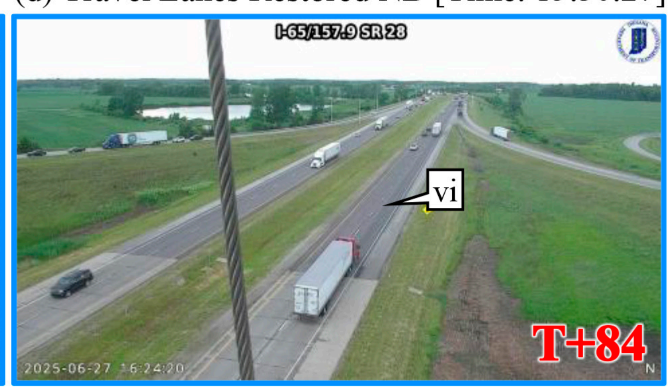
(c) Full Closure for NB [Time: 15:18:28]



(d) Travel Lanes Restored NB [Time: 15:30:27]



(e) Queue Reaches Rear Camera [Time: 15:56:19]



(f) Queue Clears Rear Camera [Time: 16:24:20]

Figure 4.4 Chronological Sequence of ITS Camera Images Capturing the Incident Timeline. (a–d) View From the Downstream Camera Showing Lane Closures and Reopening. (e–f) View From the Upstream Camera Showing the Back-of-Queue Formation and Dissipation.

- There were two ITS cameras near the incident (Callout x and Callout xi).
- The queue reached the upstream camera (Callout xi) and cleared the same, as shown by Callout xii and xiii, respectively.
- ITS images from those cameras are shown in Figure 4.4 and are described in more detail below.

Figure 4.4 shows the images from two ITS cameras, downstream (Figure 4.3, Callout x) and upstream (Figure 4.3, Callout xi) of the incident. Figure 4.4 approximates the information a TMC operator would have to investigate the occurrence of an incident without 9-1-1 dispatch audio or phone calls. For this case study, the incident occurred outside the viewable area for both cameras. Notable events from the subfigures of Figure 4.4 are as follows:

a. Incident occurrence, see the capture time (Callout i).
b. A fire engine entered the southbound lanes (Callout ii, being the same as Table 4.2, row "15:08:20"). Vehicles are traveling in the northbound lanes (Callout iii) at 9 min after the incident starts.

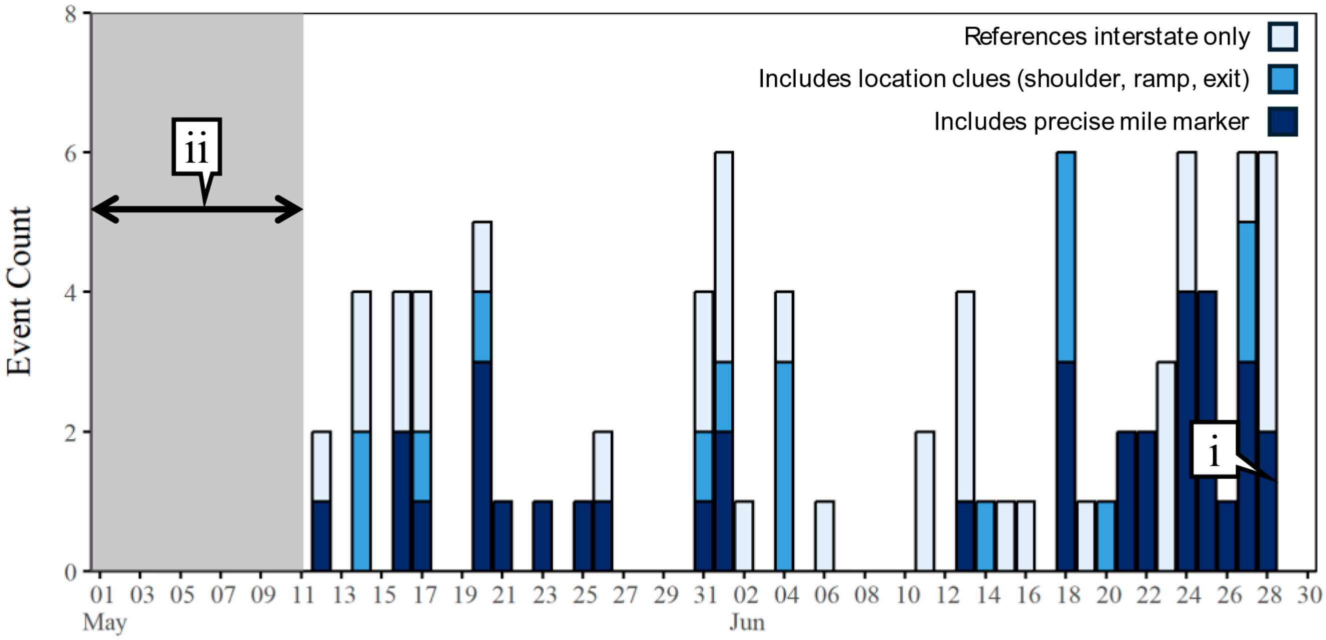


Figure 4.5 Daily Count of Detected Incidents During the May–June Study Period, Categorized by Keyword Specificity. Callout i Marks the June 27 Tractor-Trailer Fire Case Study; Callout ii Indicates the Pre-Deployment Period.

c. Full closure of the northbound lanes 20 min after the incident starts.
d. Northbound lanes reopened 31 min after the incident started (Callout iv).
e. Upstream camera captures the queue exceeding its location (allout v) at 57 min after the incident occurs.
f. The queue clears the upstream camera (Callout vi) at 84 min after the incident starts.

While the ITS camera images of Figure 4.4 provide some information on the queuing. The full lane closures of the northbound lanes and southbound lanes were not visible. Recorded 9-1-1 dispatch and operations communications (Table 4.2) provide a detailed chronology (even when not complete), unmatched by the heatmap (Figure 4.3) or ITS images (Figure 4.4).

The results of the study suggest that the ASR of public safety dispatch centers, particularly in remote areas, can extend TMC visibility in rural sections of interstates. In fact, ASR is a natural extension of what TMC operators already do when monitoring dispatch channels in metropolitan areas.

4.4 Map of Automatically Identified Incidents

The following section presents the study's results. First, all incidents detected in the study area (Figure 4.1) are presented, followed by a detailed case study (as introduced in Figure 4.2).

This study was conducted over a period of two months, from May 12 to June 30, 2025. In total, more than 100,000 audio transmissions were captured across the four counties in Figure 4.1. Figure 4.5 summarizes the incidents identified having an "interstate" keyword with (or without) a "location" keyword. Callout i refers to the incident detailed in the case study that follows. Callout ii denotes the period prior to the start of data collection. From the more than 100,000 transmissions, 87 transmissions were captured and indicated with some variation in "interstate" and other keyword groups from Table 3.3.

Of the 87 identified transmissions, 37 were assigned a geospatial position using the mile marker. Another 40 were manually assigned a geospatial position (whereas the majority were not automatically placed because they were adjacent to the interstate), and 10 were unable to be located. These 76 identified transmissions are plotted in Figure 4.6 below, where the incident referenced in Figure 4.5, Callout i, is given by Callout i. This result suggests that the quality of transcriptions and the RegEx extraction provide a feasible solution for monitoring 9-1-1 dispatch channels and automatically extracting relevant information accurately.

5. CONCLUSION

This study successfully demonstrated the feasibility and operational value of integrating ASR with public safety radio monitoring to enhance TIM. By deploying a system capable of capturing, transcribing, and analyzing discrete radio transmissions, this research addresses a critical information gap for TMCs, particularly in rural corridors where camera coverage and sensor density are limited.

The technical evaluation of ASR models indicated that while raw radio audio is inherently noisy, applying logic-based filtering significantly improves transcription utility. The analysis of WER identified a distinct inverse relationship between message length and error frequency. While short utterances (0–10 words) frequently resulted in high error rates (WER ≈ 1.0) due to a lack of semantic context, substantive transmissions (> 20 words) achieved a stabilized Mean WER of 0.332 using the Whisper

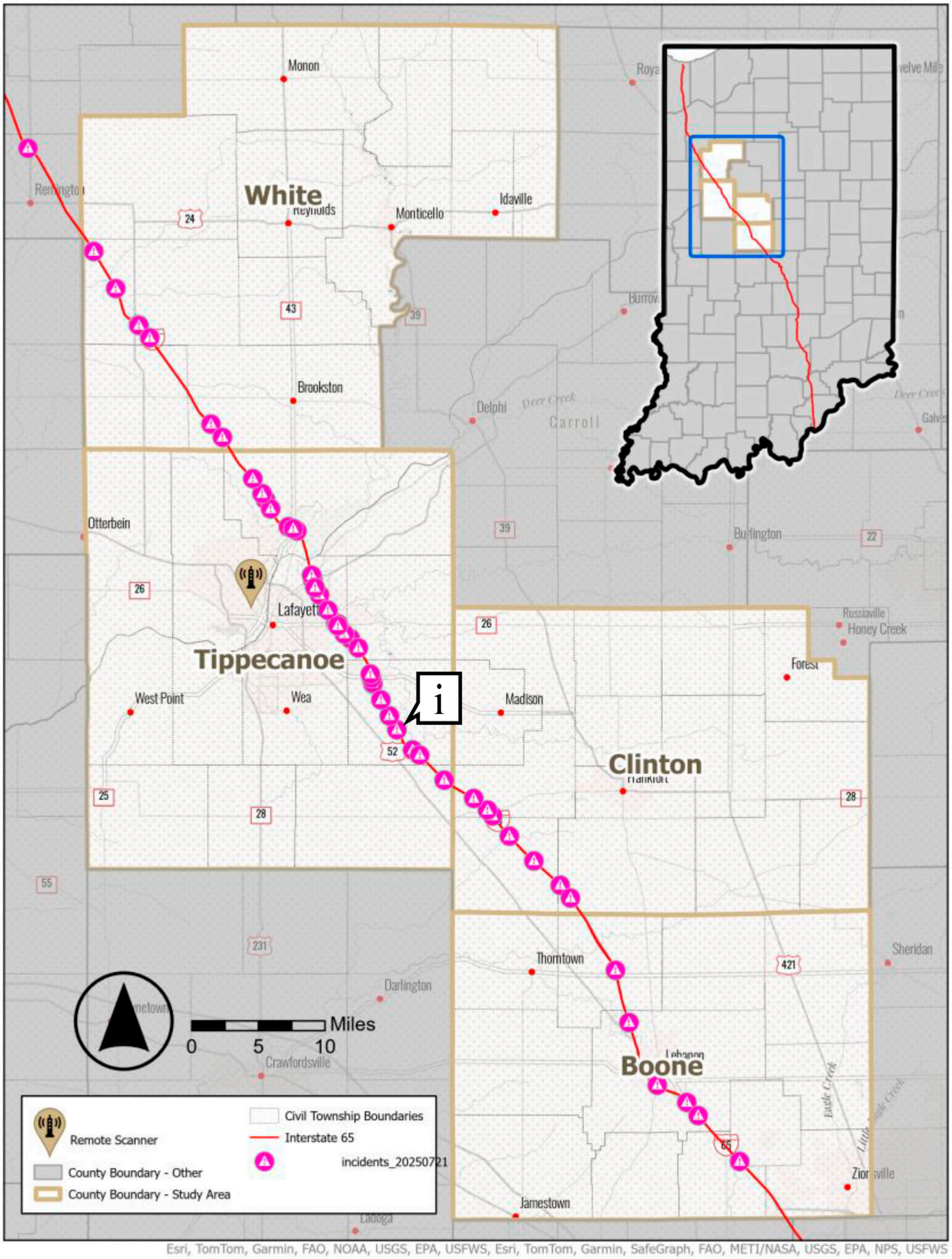


Figure 4.6 Geospatial Distribution of Automatically Identified Incidents Along the I-65 Corridor. Callout i Identifies the Location of the June 27 Case Study Incident.

Turbo model. Furthermore, this research validated the use of Model Confidence (MC) as an automated Quality Assurance mechanism. The data showed that filtering transmissions by a minimum word count and confidence threshold effectively segregates valuable dispatch narratives from static and digital noise, necessitating location-specific calibration to account for varying radio system environments.

From an operational perspective, the system demonstrated the ability to deliver actionable intelligence with minimal latency. The end-to-end processing time (from the start of a radio transmission to the delivery of an automated notification) averaged 59 s, with 90% of alerts delivered within 96 s. The case study of the I-65 tractor-trailer fire highlighted the system's ability to capture critical incident details (location, vehicle type, and lane status), providing situational awareness well before significant queuing was visible on ITS cameras or reported via traditional data streams.

In conclusion, transforming unstructured voice data into structured, geocoded text provides a scalable solution to extend TMC visibility. By leveraging the existing "verbal infrastructure" of first responders, transportation agencies can detect incidents faster, verify locations more accurately, and improve the overall responsiveness of the traffic management network.

6. FUTURE RESEARCH

While this study establishes the feasibility of monitoring a 71-mi section of interstate via public safety radio, significant opportunities exist to mature this technology into a robust, statewide traffic management tool. Future research efforts should prioritize rigorous validation, model optimization, and investigation of direct data-integration alternatives.

Further research is needed before large-scale deployment to establish standardized metrics for system performance. While this study showed feasibility, it lacks a detailed statistical comparison of detection methods. Future work should focus on assessing "recall" to ensure critical broadcasts are not missed; comparing system-inferred locations with verified crash sites to measure spatial accuracy; and creating a "ground truth" dataset to evaluate transcription accuracy amid background noise for key components of the audio.

To improve the utility of the transcribed text, future iterations of the system should move beyond off-the-shelf ASR models. Research should focus on fine-tuning models specifically for public safety lexicons, including "ten-codes," "signal codes," and local vernacular used in dispatch communications. Furthermore, integrating Large Language Models (LLMs) offers significant potential for "knowledge management." LLMs could be employed to assess the context of a transmission, distinguishing between a routine traffic stop and a major crash more effectively than simple keyword matching, thereby reducing false positives.

Finally, it must be acknowledged that ASR serves as a proxy for data that already exists in a structured format. The most robust long-term solution for incident detection is not transcription of voice, but the direct ingestion of dispatcher notes. A Centralized CAD database that aggregates text inputs from dispatchers across multiple jurisdictions in real time represents a superior approach to incident detection.

Direct CAD integration eliminates the latency and error potential inherent in speech-to-text conversion (e.g., misinterpretation of homophones or interference from background noise). While ASR provides a vital bridge to address interoperability gaps and monitor encrypted or fragmented radio systems, future research should develop administrative and technical frameworks to federate local CAD data into a single, state-level stream for TMC consumption.

REFERENCES

Alharbi, S., Alrazgan, M., Alrashed, A., Alnomasi, T., Almojel, R., Alharbi, R., Alharbi, S., Alturki, S., Alshehri, F., Almojil, M. (2021). Automatic speech recognition: systematic literature review. *IEEE Access, 9*, 131858–131876. https://doi.org/10.1109/ACCESS.2021.3112535

Bercovivi, M. W. (2006). *FCC narrowbanding mandate: A public safety guide for compliance.* International Association of Fire Chiefs and the International Municipal Signal Association. https://transition.fcc.gov/pshs/docs/clearinghouse/guidelines/Narrowbanding_Booklet.pdf

Computer Assisted Pre-Coordination Resource and Database System. (n.d.). *Region 14 – Indiana.* CAPRAD. https://www.caprad.org/800Mhz/Sites800MHz.aspx

Federal Communications Commission. (n.d.a). *911 master PSAP registry.* https://www.fcc.gov/general/9-1-1-master-psap-registry

Federal Communications Commission. (n.d.b). *Public safety licensing.* https://www.fcc.gov/public-safety-licensing

Gartner, C. M., Vajpayee, V., Desai, J., & Bullock, D. M. (2025). Automatic speech recognition of public safety radio communications for interstate incident detection and notification. *Smart Cities, 8*(5), 157. https://doi.org/10.3390/smartcities8050157

Hawkins, D. (2007, May). Project 25: The quest for interoperable radios. *COPS Interoperable Communications Technology Program Issue Brief, 6*. https://portal.cops.usdoj.gov/resourcecenter/content.ashx/cops-w0426-pub.pdf

Ibrahim, H., & Varol, A. (2020). *A study on automatic speech recognition systems* [Paper presentation]. 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon. https://doi.org/10.1109/ISDFS49300.2020.9116286

IN911. (n.d.). *Public safety answering points.* https://www.in911.net/psap-general-information.html

Integrated Public Safety Commission. (n.d.a). *25 years of SAFE-T.* https://www.in.gov/ipsc/25-years-of-safe-t/

Integrated Public Safety Commission. (n.d.b). *Systems and technology.* https://www.in.gov/ipsc/system-information/

Jiang, H. (2005). Confidence measures for speech recognition: A survey. *Speech Communication, 45*(4), 455–470. https://doi.org/10.1016/j.specom.2004.12.004

Mathew, J. K., Malackowski, H., Koshan, Y., Gartner, C., Desai, J., Li, H., Cox, E., Habib, A., & Bullock, D. M. (2024). *Development of latitude/longitude (and route/milepost) model for positioning traffic management cameras* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2024/03). Purdue University. https://doi.org/10.5703/1288284317720

National Public Safety Telecommunications Council. (n.d.). *Public safety Project 25 standards*. https://www.npstc.org/project25.jsp

ProScan. (n.d.). ProScan (Version 42.10). https://www.proscan.org/

RadioReference. (n.d.). *Indiana Project Hoosier SAFE-T.* https://www.radioreference.com/db/sid/8084

Sakhare, R. S., Desai, J., Mathew, J. K., McGregor, J., Kachler, M., & Bullock, D. M. (2024). *Measuring and visualizing freeway traffic conditions: Using connected vehicle data* (Joint Transportation Research Program Publication No. TPF-5(514)). Purdue University. https://doi.org/10.5703/1288284317751

State of Indiana. (2024). *State of Indiana tactical interoperable communications plan: State base plan with mutual aid and interoperability regions* (Version 2.0). https://www.in.gov/ipsc/files/IN-TICP-V2.0-MAY-2024.pdf

State of Indiana. (n.d.). *What is Project Hoosier SAFE-T?* https://faqs.in.gov/hc/en-us/articles/115005056187-What-is-Project-Hoosier-SAFE-T

Stein, D., Schwenninger, J., &Usabaev, B. (2012). *Automatic chat transcription on a firefighter TETRA broadcast channel* [Speech]. Speech Communication; 10. ITG Symposium, Braunschweig, Germany. https://ieeexplore.ieee.org/document/6309605

Uniden. (n.d.). *BCD996P2: Powerful digital base/mobile scanner.* https://uniden.com/products/bcd996p2

Uniden. (n.d.). *SDS200: High-performance digital base/mobile scanner.* https://uniden.com/products/sds200

Wang, Y.-Y., Acero, A., & Chelba, C. (2003). *Is word error rate a good indicator for spoken language understanding accuracy* [Paper presentation]. IEEE Workshop on Automatic Speech Recognition and Understanding, St. Thomas, US Virgin Islands.

Yujian, L., & Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(6), 1091–1095. https://doi.org/10.1109/TPAMI.2007.1078

APPENDICES

Appendix A. List of Acronyms

Appendix A. List of Acronyms

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
CAD	Computer-Aided Dispatch
CDF	Cumulative Distribution Function
DNN	Deep Neural Networks
DOT	Department of Transportation
EMS	Emergency Medical Services
FCC	Federal Communications Commission
FHWA	Federal Highway Administration
HMM	Hidden Markov Model
INDOT	Indiana Department of Transportation
IP	Internet Protocol
IPSC	Integrated Public Safety Commission
ISP	Indiana State Police
ITS	Intelligent Transportation Systems
LAN	Local Area Network
LLM	Large Language Model
LRS	Linear Referencing System
MIS	Motorist Information System
MM	Mile Marker
NLP	Natural Language Processing
P25	Project 25
PSAP	Public Safety Answering Point
QA/QC	Quality Assurance/Quality Control
RegEx	Regular Expressions
RSSI	Relative Signal Strength Indicator
SAFE-T	Safety Acting for Everyone-Together
SID	Speaker Identification
STT	Speech-to-Text
TG	Talk Group
TIM	Traffic Incident Management
TMC	Traffic Management Center
TSMO	Transportation System Management Operations
UHF	Ultra High Frequency
URI	Uniform Resource Identifier
UTL	Upper Tolerance Limit
VHF	Very High Frequency
WER	Word Error Rate

About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at docs.lib.purdue.edu/jtrp/.

Further information about JTRP and its current research program is available at engineering.purdue.edu/JTRP.

About This Report

An open access version of this publication is available online. See the URL in the citation below.

Gartner, C., Vajpayee, V., Desai, J. C., & Bullock, D. M. (2026). *Utilizing emergency management audio messages for Traffic Incident Management performance measures* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2026/15). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284318623