Research Article BEU Fen Bilimleri Dergisi 15(1): 448-464 (2026) DOI: 10.17798/bitlisfen.1833468

BİTLİS EREN UNIVERSITY
FEN BİLİMLERİ DERGİSİ
ISSN: 2147-3129 / e-ISSN: 2147-3188

VOCAL INSTABILITY AS A SENSITIVE BIOMARKER FOR DRIVING STRESS: DECOUPLING COGNITIVE LOAD AND ENVIRONMENTAL FRICTION IN A REAL-WORLD DUAL-TASK PROTOCOL

Dağhan DOĞAN

Tübitak Bilgem, Mechanical Systems and PCB Design Department, Kocaeli, Türkiye, daghan.dogan@tubitak.gov.tr

Article Info

Received: *December 1, 2025*
Revised: *February 4, 2026*
Accepted: *February 25, 2026*

Keywords

Driving stress,
Cognitive load,
Vocal biomarkers,
Pitch variability,
Vocal instability,
Dual-task protocol

ABSTRACT

Driver stress and cognitive workload are critical safety determinants in modern transportation. While vocal acoustic analysis is a promising non-invasive monitoring technique, existing literature often lacks ecological validity and struggles to distinguish between internal cognitive load (CL) from secondary tasks and external environmental friction (EF) from traffic. This study addresses this gap using a single-subject (N=1) real-world dual-task protocol. The driver maintained continuous conversation while navigating two environments: a high-friction urban congestion segment (short route) and a hybrid urban-intercity segment (long route).

Analysis utilized a custom weighted acoustic stress index and instantaneous pitch standard deviation (vocal instability). Findings demonstrate that the constant dual-task demand establishes a dominant, consistent moderate stress baseline (~34–36%), decoupled from routine traffic fluctuations and congestion levels. Although average stress levels remained consistent, pitch standard deviation proved to be a more sensitive metric, being significantly lower on the long route compared to the pure urban segment. This confirms vocal instability's ability to effectively decouple CL and EF contributions, providing empirical evidence that low-demand highway segments create a stabilizing effect on the voice even under moderate overall load. Consequently, vocal instability is validated as a sensitive biomarker essential for developing context-aware in-vehicle systems capable of distinguishing between distraction-related and environmental stress.

1. INTRODUCTION

Driver stress and cognitive workload are widely recognized as critical determinants of safety within modern transportation systems. Elevated stress levels are empirically linked to impaired driving performance, delayed reaction times, and an increased risk of accidents [1]. Recent systematic reviews underscore the rapidly evolving landscape of stress detection methodologies, emphasizing the critical need for integrated and real-world monitoring solutions [2]. Seminal research by Healey and Picard [3] established a foundational paradigm by demonstrating the feasibility of detecting driver stress in real-world settings using physiological sensors. This established a core paradigm, often combined with analyses of vehicle kinematics and performance metrics. Drivers are subjected to stressors from diverse sources: exogenously, from environmental challenges such as urban congestion and complex traffic configurations [4], and endogenously, from the cognitive demands of secondary tasks. This phenomenon has been extensively documented by Recarte and Nunes [5], who provided experimental evidence of how mental workload degrades visual search, and by Engström et al. [6], who confirmed the effects of visual and cognitive load in both real and simulated driving. This multifaceted etiology of driver stress necessitates the development of robust, real-time systems capable of disambiguating and accurately assessing a driver's transient cognitive and emotional state.

Traditional methodologies for driver stress detection have predominantly relied on invasive physiological measures, such as electrocardiogram (ECG) and galvanic skin response (GSR)—often

building upon the principles established by Healey and Picard [3] or on the analysis of vehicular kinematics [7], [8]. Parallel research in affective computing, such as the study by Katsis et al. [9] on emotion recognition, further validated the use of bio-signal processing for inferring emotional states. While these sensors provide valuable data, many existing methods particularly wearable devices face significant challenges regarding driver comfort, intrusive setups, and long-term compliance in ecological driving settings [10]. Consequently, the growing demand for non-invasive, contactless monitoring has stimulated significant interest in vocal acoustic analysis as a promising alternative methodology [11]. The human voice constitutes a particularly compelling biomarker, as stress-induced autonomic nervous system (ANS) activity is known to directly modulate laryngeal muscle tension and respiratory patterns, as validated by foundational research [12] and recent meta-analytic evidence [13]. This relationship is further substantiated by the framework of Van Puyvelde et al. [14], which links vocal effort directly to human performance under varying levels of physiological stress. These physiological correlates manifest as quantifiable acoustic variations in parameters including fundamental frequency (F0), F0 variability (pitch standard deviation), and other prosodic features, as established in the benchmark work of Schuller et al. [15] and the standardized Geneva Minimalistic Acoustic Parameter Set (GeMAPS) [16]. This enables continuous, objective assessment of cognitive load through speech produced during driving tasks [15], [17]. Furthermore, recent advancements in analyzing real-life speeches highlight the increasing efficacy of integrating acoustic and semantic information for robust daily stress detection [18].

The cognitive demands inherent in concurrent driving and conversation can be contextualized through the lens of social baseline theory, which posits that cognitive and emotional resources are regulated through social proximity [19]. The dual-task paradigm in driving inherently challenges these resource allocation mechanisms, rendering vocal analysis a uniquely relevant modality for detecting fluctuations in cognitive load. Notwithstanding established correlations, the extant literature is characterized by significant limitations, particularly concerning ecological validity and causal attribution. Firstly, a substantial portion of foundational research is conducted in controlled laboratory or simulator environments where stress is artificially induced [8], [9]. Secondly, the field grapples with a causal attribution challenge [7], [20], wherein it is difficult to determine whether an observed stress response originates from exogenous factors or endogenous sources. The comprehensive review by Dong et al. [20] underscores the importance of a multimodal approach for holistic assessment, noting the persistent challenge of attributing the source of impairment.

This study directly addresses these methodological gaps through a single-participant (N=1) case study design employing an integrated approach. The primary objective is to implement a high-load, real-world dual-task protocol to analyze the differential effects of varying traffic demands on a suite of acoustic stress features. Building upon established frameworks [16], a customized, weighted stress index is developed. The central hypothesis posits that vocal instability metrics, specifically pitch standard deviation, possess the sensitivity to detect subtle environmental variations even when the overall average stress level is dominated by the cognitive load. To ensure robustness, the parameters and thresholds were informed by an extensive review of vocal stress correlates (e.g., [13], [15], [16]). The remainder of this paper details the experimental methodology (Section 2), results (Section 3), discussion (Section 4), and final conclusions (Section 5).

2. MATERIAL AND METHOD

This section details the experimental methodology implemented to collect and process multimodal data comprising vehicle kinematics, driving environment context, and driver speech for the development and validation of the vocal stress detection model. The core objective of this study is to analyze the driver's voice characteristics to determine the instantaneous stress level. The methodology is structured across participant selection, route design, instrumentation, data acquisition protocols, and the signal processing techniques utilized for acoustic feature extraction and subsequent stress scoring.

2.1. Participant and Driving Task

This section details the characteristics of the participant and the specific protocol implemented during the data collection phase.

2.1.1. Participant Characteristics

The study utilized a highly experienced male driver (N=1) to minimize inter-subject variability in driving style, which could potentially confound the correlation between vocal features and situational stress. The participant's demographic and driving experience profile were as follows:

- Gender and Age: Male, 39 years old.
- Education Level: Possessing a doctoral degree (Ph.D.).
- Driving Experience: Highly experienced, holding a valid driving license and reporting a driving history of 21 years with daily vehicle use.
- Physical Status: The participant reported no physical impairments, with the only accommodation being the use of prescription eyeglasses.

It is important to note that the single-subject design was chosen as a necessary first step to establish a proof-of-concept and minimize confounding variables from inter-driver differences. However, this design inherently limits the generalizability of the findings, a point which will be addressed in the limitations section.

2.1.2. Driving Protocol and Task Requirements

The experimental procedure required the participant to complete a total of six distinct drives under controlled conditions:

1. Route Repetition: The participant drove each of the two pre-determined experimental routes (Route 1 and Route 2, as detailed in Section 2.2) three times.
2. Temporal Variation: To account for potential diurnal (daily) changes in traffic conditions and driver alertness, the drives were scheduled at three specific times of the day: 11:00 A.M., 2:00 P.M., and 5:00 P.M. This scheduling yielded a total of 2 routes x 3 time slots = 6 drives.
3. Controlled Environment: The driver was instructed to keep the vehicle windows closed throughout the duration of the drives to ensure acoustic isolation and minimize external noise contamination of the vocal recording.
4. Simultaneous Tasks (Dual Task): The driver's primary task was adherence to traffic rules, while the critical secondary task was maintaining continuous verbal communication by answering questions posed by a remote interviewer (as described in the Vocal Elicitation Protocol, Section 2.4). This dual-task paradigm was specifically designed to investigate the effects of simultaneous cognitive load (driving and speaking) on the acoustic characteristics of the driver's voice.

2.2. Experimental Routes and Environment

The driving experiment was conducted across two distinct pre-determined routes, designed to expose the driver to a range of traffic complexities and driving demands. The geographical characteristics and visual context of both routes are illustrated in Figure 1-2.

Figure 1. Urban road.

Figure 2. Hybrid urban and intercity road.

Route 1 (Urban Road Segment): The first route was a relatively short segment, spanning approximately 2.5 km. This route was specifically selected to represent urban driving conditions.

- Characteristics: This segment primarily featured low-speed driving, frequent stops and starts, rapid decision-making requirements (e.g., lane changes), and a high presence of non-motorized traffic (pedestrians, cyclists).
- Purpose: The aim of this route was to induce a controlled level of driving-related stress and high cognitive load due to external traffic factors.

Route 2 (hybrid urban and intercity segment): The second route was a longer, hybrid segment, extending to approximately 6.5 km. This route was designed to capture a broader spectrum of typical driving scenarios.

- Characteristics: This route combined elements of city driving (similar to Route 1, including intersections and local traffic) and intercity or highway-like driving (featuring higher speeds, stable speeds, and lower immediate interaction with external traffic elements).
- Purpose: This segment allowed for the collection of speech data under both high-demand (urban) and low-demand (intercity/highway) driving conditions, providing the necessary contrast to validate the vocal stress detection model across varied contexts.

Both experimental drives were conducted during pre-defined time windows to ensure consistent traffic flow and environmental conditions, thereby minimizing the impact of temporal variability on the collected data.

2.3. Data Acquisition and Instrumentation

The experimental setup utilized a conventional passenger vehicle equipped with both custom and commercial off-the-shelf (COTS) instrumentation to synchronously capture vehicle kinematics, environmental context, and the driver's vocal data. The overall configuration, including the vehicle, camera system, phone, and headset, is visually represented in Figure 3.

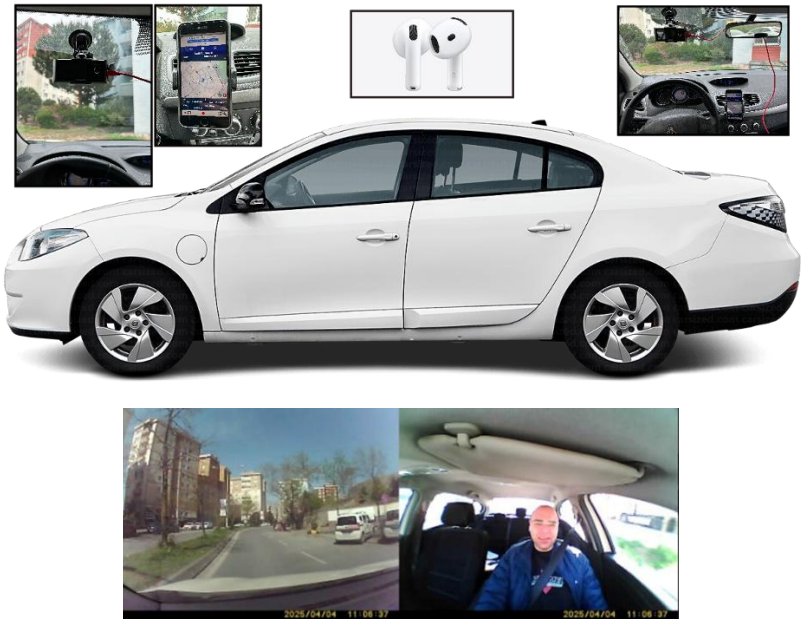

Figure 3. Data acquisition system.

2.3.1. Vehicle Kinematics and Positional Data Acquisition

A conventional internal combustion engine (ICE) vehicle served as the experimental platform. Vehicle dynamics and precise location data were collected using a dedicated mobile sensing unit:

- Geologger Application (Apple iOS): A commercial mobile application, "Geologger," was operated on an Apple smartphone mounted securely within the vehicle. This application continuously logged key navigational and kinematic metrics during the drives, generating a final GPX file for post-processing.

- Kinematic Data: The application provided core time-series data, including instantaneous coordinates, instantaneous speed (km/h), average speed (km/h), and maximum speed (km/h).
- Elevation and Accuracy Metrics: Crucially, the Geologger application also recorded elevation data, specifically minimum elevation (m) and maximum elevation (m), and reported on data quality through average positional accuracy (m) and average elevation accuracy (m). These accuracy metrics were utilized to validate the reliability of the collected location and speed data.

2.3.2. Synchronized Audio-visual Recording System

Simultaneous audio and video recordings of the driver and the road environment were collected to provide contextual data for stress analysis:

- GHK-1008 Dual-Channel Camera: This specialized camera unit was deployed to ensure the synchronized capture of visual data.
    - In-Cabin View: Focused on the driver to record behavioral cues (e.g., facial expressions, steering wheel interaction) for correlating with vocal stress levels.
    - Forward-Facing View: Focused on the road and traffic, capturing the external environment and potential stressors (e.g., traffic density, sudden road events).
- Acoustic Data Collection: The driver was equipped with a headset and earphones. The driver's speech, which is the input for the stress analysis, was continuously recorded via the headset's dedicated microphone to ensure a high-fidelity audio stream with minimal environmental noise.

2.3.3. Remote Communication Protocol

The Apple smartphone and the headset also functioned as the communication platform for the remote interview protocol (detailed in Section 2.4). This setup facilitated a remote cellular connection with the interviewer. This configuration was essential because the use of the headset ensured that the interviewer's voice was received directly by the driver's ear and was not captured by the recording microphone, thus preserving the acoustic purity of the driver's speech data for analysis.

All data streams (kinematic data from Geologger, video from GHK-1008, and high-fidelity audio) were time-stamped and synchronized to enable precise temporal correlation between the vocal stress scores and specific driving events or conversational moments.

2.4. Vocal Elicitation Protocol and Data Acquisition Setup

The voice data used for stress analysis was collected via a meticulously designed and standardized elicitation protocol conducted throughout the driving task. This protocol was essential to capture vocal features under varied cognitive and emotional loads induced by both the driving environment and the interviewer's questions.

2.4.1. Remote Interview Configuration

To prevent the interviewer's voice from contaminating the acoustic analysis of the driver, a remote communication setup was employed.

- The interviewer was situated remotely and communicated with the driver via a cellular phone connection.
- The driver used a headset or earphones for clear communication, ensuring the interviewer's voice was not captured by the recording system.
- The driver's speech was continuously recorded using a dedicated dual-channel camera system that simultaneously captured the driving environment (front view) and the driver's face/behavior (cabin view). The audio data from the driver's microphone was synchronously logged with the video data.

2.4.2. Standardized Interview Sequence

The interviewer administered a series of 30 predetermined, open-ended questions to the driver. The questions were designed primarily to sustain natural conversation and maintain a continuous vocal stream, while also introducing subtle cognitive and emotional variations that might influence vocal characteristics. The questions were delivered sequentially throughout the defined experimental routes,

ensuring the speech data corresponded directly to specific segments of the driving task. The interview sequence was structured as in Table 1.

Table 1. Questionnaire in the study

Section	Question Numbers	Primary Focus	Stress Induction Method
Introductory/ Warm-up	Q1–Q3	General well-being, task context, and traffic assessment.	Establishing baseline vocal characteristics.
Cognitive/ General Topics	Q4–Q27	Personal preferences (music, hobbies, travel, etc.) and abstract opinions (e.g. electric cars, nature).	Introducing cognitive load (retrieval and opinion formation) alongside varied driving demands.
Conclusion/ Debrief	Q28–Q30	Reflections on the driving experience and gratitude.	Transitioning back to a neutral emotional state.

The sequence of 30 questions included common conversational topics (Q4-Q27) to minimize unnatural scripted responses and maximize the collection of spontaneous speech data relevant to both relaxed and demanding driving conditions.

2.4.3. Data Synchronization

All collected data streams—driver speech (audio), driver behavior (video), and traffic conditions (video)—were synchronized temporally to enable frame-by-frame correlation between the calculated vocal stress score (as derived in the analysis section) and the specific driving or conversational event occurring at that moment.

2.5. Data Processing and Analysis

This section details the custom algorithm developed in MATLAB for the extraction of relevant acoustic features and the subsequent calculation of the driver's vocal stress index.

2.5.1. Pre-processing and Framing Parameters

The recorded audio data, sampled at fs = 44.100 was initially processed to ensure a normalized amplitude (peak value set to 1) and converted to a single-channel signal (if stereo). The core of the analysis relies on dividing the acoustic signal into short frames using a shifting window approach. The framing parameters were utilized for feature extraction in Table 2.

Table 2. Framing parameters

Parameter	Value	Description
Sampling Frequency (fs)	44.100 Hz	Rate at which the audio signal was recorded.
Frame Length	1.764 samples	Equivalent to 40 ms of audio (1764/44100).
Overlap Length	441 samples	Equivalent to 10 ms overlap (441/44100).
Hop Size	1.323 samples	Frame shift, resulting in a 30 ms frame-rate update.

2.5.2. Feature Extraction

For each frame, a set of time-domain and frequency-domain acoustic features known to correlate with vocal stress were extracted. The following features were computed:

- Short-Time Energy: Used for initial silence/unvoiced segment detection.
- Pitch (Fundamental Frequency, F0): Calculated using the autocorrelation method within the defined minimum (80 Hz) and maximum (500 Hz) frequency limits.
- Zero-Crossing Rate (ZCR): Measure of the rate at which the signal changes sign.
- Harmonics-to-Noise Ratio (HNR): Calculated using the autocorrelation peak for the harmonic component and averaging noise components, reflecting voice quality and periodicity.
- Spectral Centroid: Represents the center of mass of the spectrum, indicating the relative energy distribution.

- Jitter and Shimmer: Measures of short-term perturbation in the pitch period and amplitude, respectively, calculated based on the detected F0 period (T0).
- Pitch Standard Deviation (Pitch Std. Dev.): Calculated as the standard deviation of valid pitch values across a preceding 10-frame window to capture instantaneous vocal variability, a key stress indicator.

2.5.3. Stress Score Calculation and Thresholds

The selection of feature weights and threshold values (detailed in Table 3) was grounded in prior empirical research linking specific acoustic changes to stress-induced physiological modulation. For instance, the increased weight assigned to pitch standard deviation reflects its established sensitivity to cognitive load and autonomic nervous system arousal, as documented in foundational works and recent meta-analytic evidence identifying fundamental frequency as a robust stress biomarker [13], [21]. The 'moderate stress' category (approximately 30-40%) was empirically defined based on the consistent baseline level established by the dual-task protocol in pilot observations, representing a state of sustained cognitive engagement above rest but below acute overload. A frame-level stress score (ranging from 0 to 10) was computed by comparing each extracted feature against predefined threshold values and accumulating a weighted sum of normalized scores. The final stress level, expressed as a percentage, was derived by scaling this score (*score × 10%*). The feature scores were normalized using custom linear scaling functions (e.g., *score = min (1, max (0, (feature - threshold) / scaling factor*)), where a score of 1 indicates maximum influence on stress detection. This linear normalization was specifically preferred to preserve the proportional integrity of vocal fluctuations and to maintain mathematical interpretability, thereby avoiding the artificial saturation of data often associated with non-linear scaling. The weights and core thresholds utilized in the weighted summation are detailed in Table 3. Through this heuristic calibration, informed by both literature and empirical pilot tuning, a robust framework for intra-subject longitudinal comparison was established.

Table 3. Stress scoring weights and threshold parameters

Feature	Weight	Threshold Value	Scaling Factor	Unit	Condition (for High Stress)
Short-Time Energy	0.10	0.050	0.35	-	Above Threshold
Pitch (F0)	0.15	250	250	Hz	Above Threshold
Pitch Std. Dev. (10-frame)	0.20	30	30	Hz	Above Threshold
Zero-Crossing Rate	0.10	0.250	0.30	-	Above Threshold
Harmonics-to-Noise Ratio (HNR)	0.15	15	15	dB	Below Threshold
Spectral Centroid	0.10	2,500	2,500	Hz	Above Threshold
Jitter	0.10	0.0050	0.010	-	Above Threshold
Shimmer	0.10	0.0400	0.06	-	Above Threshold

2.5.4. Model Calibration and Interpretability

The selection of feature weights and threshold values was grounded in established literature on vocal stress correlates [13], [15], [20], and further refined based on the established sensitivity of vocal fundamental frequency to autonomic arousal a correlation explicitly validated in the recent meta-analysis by Veiga et al. [13]. These parameters were additionally tuned through pilot observations of the participant's baseline vocal profile in a stationary vehicle environment to ensure empirical sensitivity to the specific acoustic characteristics of the cabin.

A linear normalization function was intentionally selected to map raw acoustic deviations to stress scores. This choice prioritizes model transparency and interpretability, maintaining a direct relationship between signal fluctuations and stress percentages, which is essential for a proof-of-concept study. Preliminary sensitivity analysis indicates that the observed trends specifically the stability of the baseline and the sensitivity of pitch deviation remain robust across minor adjustments (±10%) in the assigned weights.

Finally, it is emphasized that the reported percentage stress values are model-dependent and function as relative indicators for intra-subject comparison between driving conditions, rather than absolute,

population-level benchmarks. This approach ensures that the index remains a robust tool for identifying longitudinal trends in vocal instability.

2.5.5. Post-processing and Output

To smooth the highly variable instantaneous stress levels, a 7-frame running average was applied to the percentage vector. The final output generates a time-series plot showing the instantaneous and smoothed stress levels, along with statistical metrics. The algorithm identifies and reports High Stress Regions defined as continuous segments where the 7-frame average stress exceeds 70%. Additionally, the overall standard deviation of all valid pitch values is calculated as a holistic measure of vocal instability across the entire recording. All results, including the mean stress, maximum stress, and overall pitch standard deviation, are saved to a text file for further analysis.

3. RESULTS

3.1. Short Route Results: Urban Congestion Segment

The short route, designated as the Urban Congestion Segment, was executed three times at different times of the day (11:00, 14:00, and 17:00) to investigate the effect of time-varying traffic conditions and continuous dual-task cognitive load on the driver's vocal parameters.

3.1.1. Kinematic Data Analysis

Table 4 presents the key kinematic metrics recorded for the three repetitions of the short route.

Table 4. Summary of kinematic data for the three repetitions of the short route

DN	A	B	C	D	E	F	G	H	I	J
1	11.00	2.53	369	75	24.2	49.6	172	217	6	3
2	14.00	2.56	444	90	20.3	51.0	178	223	5	3
3	17.00	2.55	429	87	20.9	49.6	175	218	5	4

Note: DN: Driving Number; A: Time; B: Distance (km); C: Duration (sec); D: Data Points; E: Average Velocity (km/h); F: Maximum Velocity (km/h); G: Minimum Elevation (m); H: Maximum Elevation (m); I: Average Location Accuracy (m); J: Average Elevation Accuracy (m).

The corrected kinematic data confirms the segment's role as a low-speed, high-density urban route, with all drives spanning approximately 2.5 km.

Velocity and Congestion: The average velocity remained consistently low across all sessions (ranging from 20.3 km/h to 24.2 km/h), and the maximum velocity was constrained below 51 km/h. This reinforces the congested nature of the route, characterized by frequent stops and starts.

Temporal Variation: The 14:00 drive was the longest in duration (444 seconds) and exhibited the lowest average velocity (20.3 km/h), suggesting it encountered the highest degree of impedance or traffic congestion, contrary to the previous assumption that the 17:00 drive would be the most difficult. The 17:00 drive was only marginally easier (20.9 km/h) than the 14:00 drive.

3.1.2. Vocal Stress Analysis

The core objective was to map driving conditions to the calculated vocal stress scores. Figures 4-6 illustrate the time-series plots of the instantaneous and smoothed vocal stress levels for the three drives. The detailed output reports confirmed the quantitative metrics summarized in Table 5.

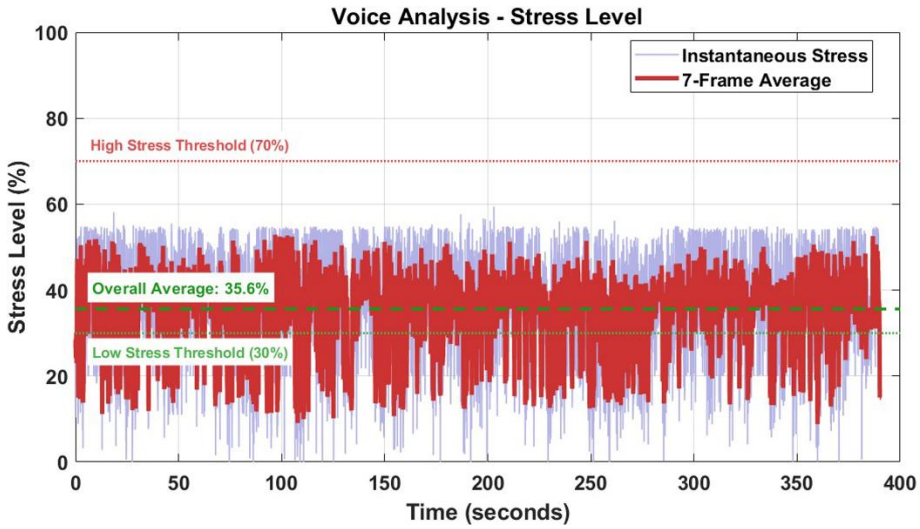

Figure 4. Vocal stress scores for the short route (time: 11.00).

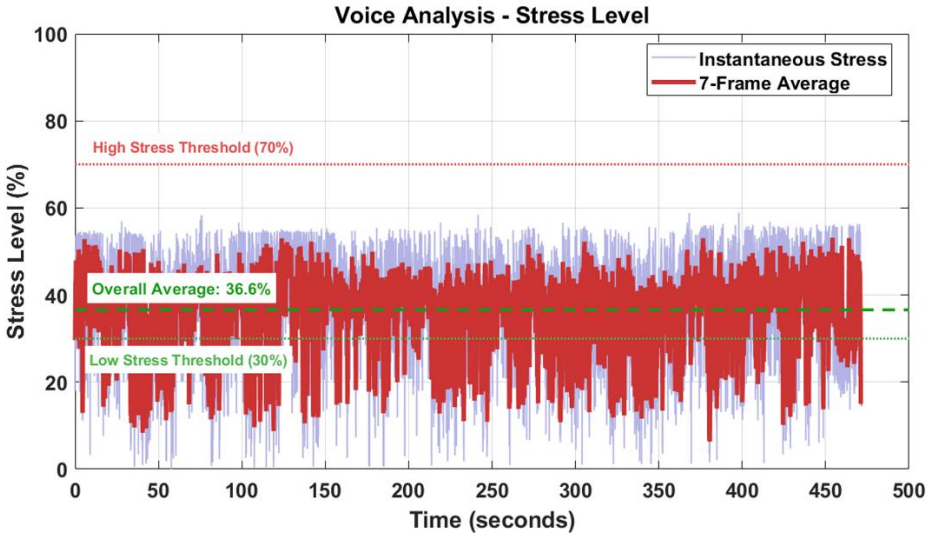

Figure 5. Vocal stress scores for the short route (time: 14.00).

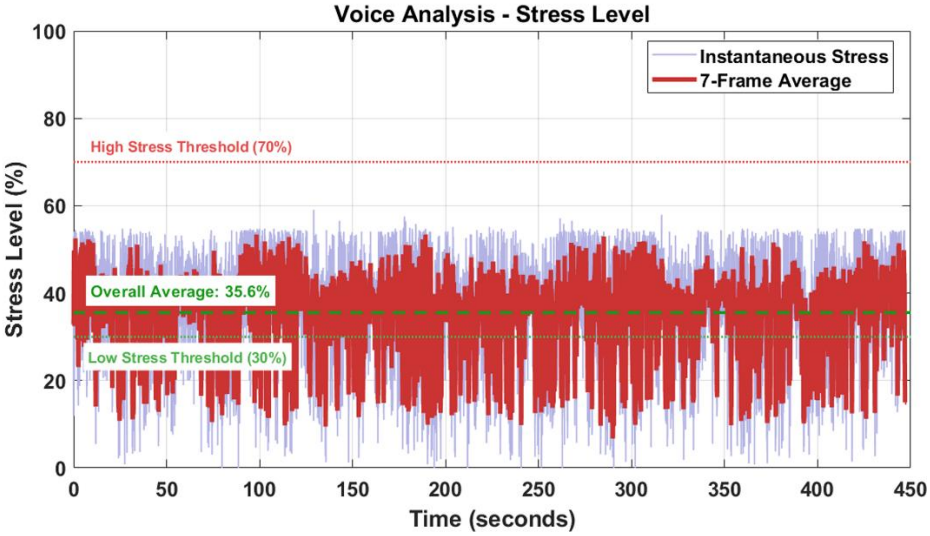

Figure 6. Vocal stress scores for the short route (time: 17.00).

Table 5. Summary of vocal stress metrics for the short route drives

Drive Time	Overall Average Stress (%)	Max. Instantaneous Stress (%)	Overall Pitch Std. Dev. (Hz)
11:00	35.60	59.40	134.88
14:00	36.59	58.89	128.17
17:00	35.55	59.04	126.31

Stress Consistency and Decoupling:

- The overall average stress level remained remarkably consistent across all three sessions, fluctuating narrowly between 35.55% and 36.59%. This placed all repetitions consistently in the moderate stress category. No high stress regions (average stress > 70%) were observed in the time-series plots (Figures 4-6).
- Significantly, the 17:00 drive, despite being executed during the typical evening rush hour (kinematically challenging with an avg. velocity of 20.9 km/h), registered the lowest overall average stress (35.55 %) among the three drives.
- This finding strengthens the conclusion that the measured vocal stress is decoupled from instantaneous or expected traffic congestion. The high cognitive load resulting from the experimental dual-task requirement (navigating heavy urban traffic while maintaining continuous conversation) appears to establish a stable, moderate vocal load baseline that overshadows the marginal differences in congestion level across the three time slots.

Vocal Instability (Pitch Std. Dev.):

- The overall pitch standard deviation (Pitch Std. Dev.), a critical measure of vocal instability, was also lowest during the 17:00 drive (126.31 Hz). This suggests that the driver's voice was the most stable and least agitated during the late afternoon session.
- Conversely, the highest instability was observed in the 11:00 drive (134.88 Hz). This pattern suggests that while the average vocal load remains constant (moderate stress), the driver exhibits greater vocal fluctuation and perturbation (instability) during the morning session, possibly reflecting a different baseline state of alertness or energy compared to the end of the day. It is crucial to note that the specific percentage values defining 'moderate stress' (approx. 34–36%) are relative to our model's calibration and should be interpreted as a consistent benchmark within this proof-of-concept study, rather than as absolute, population-wide thresholds.

3.2. Long Route Results: Hybrid Urban and Intercity Segment

The long route was designed as a hybrid segment (approximately 6.5 km) to capture a broader spectrum of driving scenarios, encompassing both high-demand urban segments and lower-demand, higher-speed intercity segments.

3.2.1. Kinematic Data Analysis

Table 6 summarizes the kinematic data collected for the three repetitions of the long route.

Table 6. Summary of kinematic data for the three repetitions of the long route

DN	A	B	C	D	E	F	G	H	I	J
1	11.00	6.44	782	157	28.3	74.8	117	216	7	4
2	14.00	6.35	734	148	28.3	80.8	121	212	13	5
3	17.00	6.55	815	164	23.2	71.8	120	221	16	4

Note: DN: Driving Number; A: Time; B: Distance (km); C: Duration (sec); D: Data Points; E: Average Velocity (km/h); F: Maximum Velocity (km/h); G: Minimum Elevation (m); H: Maximum Elevation (m); I: Average Location Accuracy (m); J: Average Elevation Accuracy (m).

- Route Characteristics: The long route is approximately 2.5 times longer than the short route, enabling the collection of data across diverse driving scenarios. The high Maximum Velocities (up to 80.8 km/h) confirm the presence of an intercity/highway segment, a key differentiating factor from the purely urban short route.

- Temporal Variation: The 17:00 drive was the most prolonged (815 seconds) and the slowest (23.2 km/h average velocity). This confirms that the evening rush hour significantly impacted the hybrid route, particularly its urban sections, resulting in the highest time-based congestion. The 11:00 and 14:00 drives shared an identical average speed (28.3 km/h), indicating similar traffic conditions during the midday period.

3.2.2. Vocal Stress Analysis

The time-series stress plots for the long route in Figures 7-9 and their respective analysis reports confirm the following vocal stress metrics in Table 7.

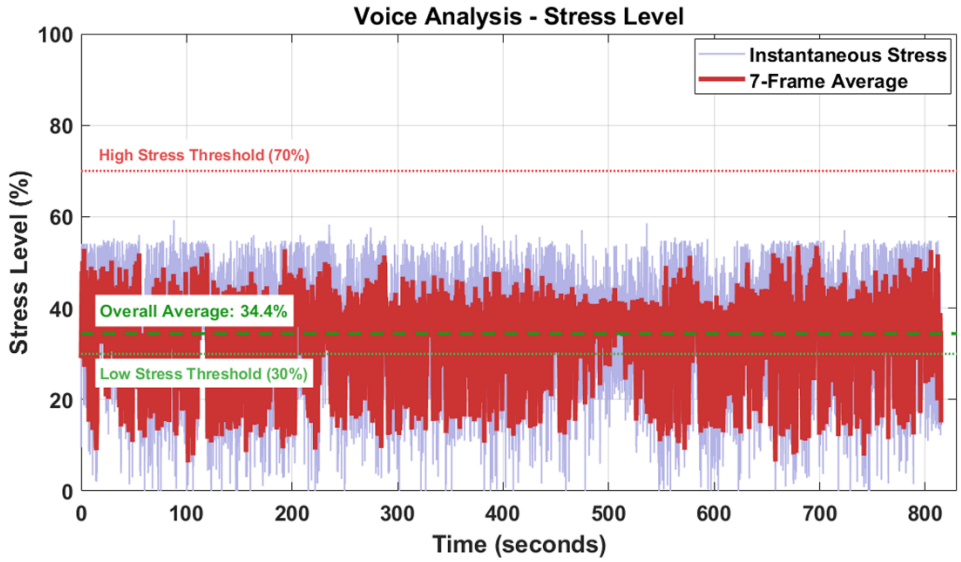

Figure 7. Vocal stress scores for the long route (time: 11.00).

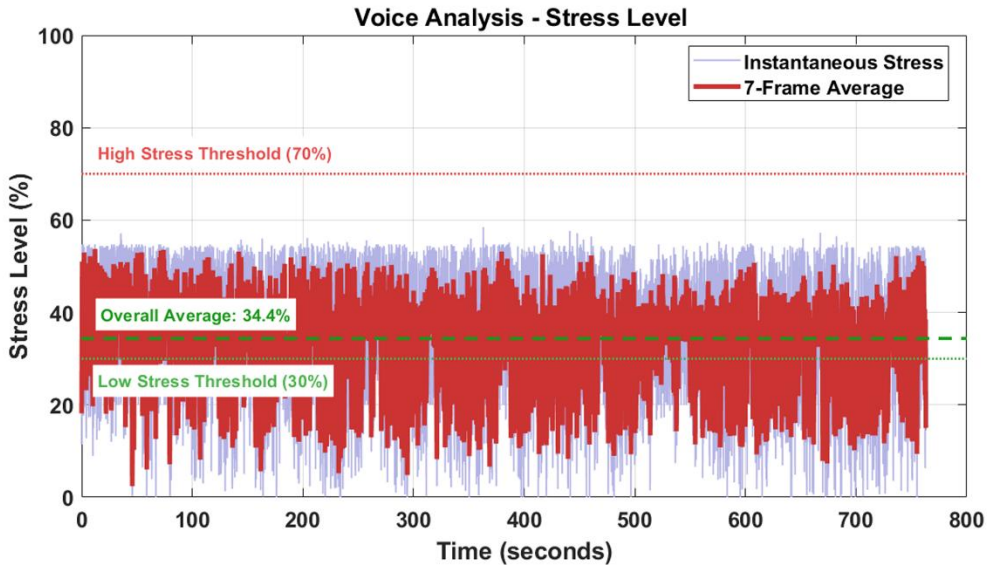

Figure 8. Vocal stress scores for the long route (time: 14.00).

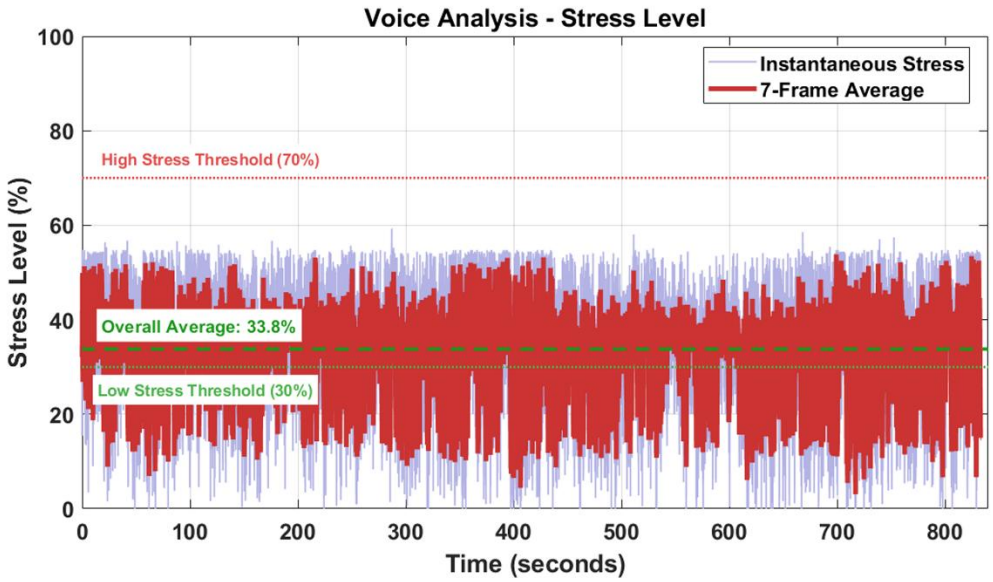

Figure 9. Vocal stress scores for the long route (time: 17.00).

Table 7. Summary of vocal stress metrics for the long route drives

Drive Time	Overall Average Stress (%)	Max. Instantaneous Stress (%)	Overall Pitch Std. Dev. (Hz)
11:00	34.42	59.23	116.63
14:00	34.37	58.43	126.22
17:00	33.79	59.31	122.81

Stress Consistency and Decoupling:

- Similar to the short route, the overall average stress level on the long route remained highly consistent, ranging only from 33.79% to 34.42%. All drives were categorized as moderate stress, and no high stress regions (> 70%) were detected.
- This consistency, despite the significant kinematic variation (average speeds varied by 5.1 km/h), further validates the finding that the vocal stress is dominated by the constant cognitive load of the dual-task protocol, rather than momentary traffic fluctuations.

Comparison with Short Route:

- Crucially, the average stress levels on the long route (34.2 %) are marginally lower than those on the short route (35.9 %). This difference is likely due to the long route incorporating a low-demand intercity segment where the driver experiences stable, higher speeds, momentarily reducing the cumulative cognitive burden associated with urban congestion.

Vocal Instability (Pitch Std. Dev.):

- The long route displayed the lowest overall vocal instability compared to the short route (long route minimum 116.63 Hz vs. short route minimum 126.31 Hz). This stabilization is consistent with the presence of smoother, higher-speed highway segments.
- Within the long route, the 11:00 drive had the lowest Pitch Std. Dev. (116.63 Hz), indicating the driver's voice was most stable during the morning session on this route. The slightly higher instability at 14:00 and 17:00 could be related to increased traffic interaction during those periods, even within the hybrid route's urban segment.

3.3. Comparative Analysis: Short Route vs. Long Route

This section directly compares the kinematic features and vocal stress outcomes between the short route (pure urban congestion) and the long route (hybrid urban and intercity), highlighting the influence of driving environment complexity on the derived vocal metrics.

3.3.1. Kinematic Comparison

The two routes were fundamentally distinct in their design and execution, as confirmed by the kinematic metrics in Table 4 and Table 6:

- Route Design and Velocity: The long route was approximately 2.5 times longer (approx. 6.45 km) and incorporated segments allowing for significantly higher speeds, evidenced by the high average maximum velocities (75.8 km/h) compared to the short route (approx. 50.4 km/h). The overall average velocity was also higher on the long route (approx. 26.6 km/h) than on the short route (approx. 21.8 km/h).
- Driving Demand: The short route served its intended purpose as a low-speed, highly demanding urban segment, characterized by high friction and frequent stop-and-go actions. The long route provided a necessary contrast by introducing periods of stable, high-speed driving (low-demand segments), interspersed with urban congestion.

3.3.2. Vocal Stress Comparison

The comparison of average vocal stress metrics across both routes reveals a critical insight regarding the primary stressor in the experiment:

- Overall Average Stress: The short route (pure urban congestion) exhibited a marginally higher overall average stress level (approx. 35.91%) compared to the long route (hybrid segment) (approx. 34.19%).
- Dominance of Dual-Task Load: Although the difference in average stress is small (approx. 1.72 percentage points), the fact that the long route has lower stress is significant. This finding supports the conclusion that the dual-task protocol (driving + continuous conversation) dominates the measured vocal stress response. The inclusion of lower-demand, stable-speed intercity sections in the long route allowed the driver a brief respite from constant decision-making and traffic interaction, thus slightly lowering the cumulative cognitive burden and consequently the overall vocal stress score.
- Absence of High Stress: Crucially, no high stress regions (> 70%) were detected in any of the nine total drives (three short, three long), indicating that the experimental protocol consistently induced a moderate stress state, suitable for monitoring gradual changes, but did not push the single subject into acute vocal stress.

3.3.3. Vocal Instability (Pitch Std. Dev.) Comparison

The most pronounced difference between the two routes was observed in the measures of vocal stability:

- Reduced Instability on Long Route: The overall pitch standard deviation (a measure of vocal fluctuation/instability) was significantly lower on the long route (average approx. 121.89 Hz) compared to the short route (average approx. 129.79 Hz).
- Correlation with Driving Environment: This decrease in instability is directly correlated with the presence of the intercity segment on the long route. Smoother, high-speed driving (low friction) reduces the moment-to-moment psychological and physical demands on the driver, leading to a more stable fundamental frequency (pitch) during continuous speech. Conversely, the constant vigilance, rapid acceleration/braking, and frequent decision-making characteristic of the short route's pure urban congestion translate directly into higher vocal perturbation.

In summary, the comparative analysis confirms that the vocal stress algorithm successfully distinguishes between the two driving environments, showing that the high-friction short route generates a higher average vocal load and greater vocal instability than the hybrid long route.

4. DISCUSSION

The findings of this proof-of-concept study offer significant preliminary insights into the effectiveness of using vocal acoustic features for assessing driver stress under the demanding conditions of a dual-task environment. The discussion focuses on three main findings: the consistency of the moderate stress category, the observed decoupling between traffic impedance and average vocal stress, and the sensitivity of vocal instability metrics to subtle route changes.

4.1. Consistency of Stress Categorization

The most prominent finding is the remarkably stable overall average stress level across all six drives (three on the short route and three on the long route). The data from this driver indicate a consistent classification within the moderate stress category, with average scores ranging narrowly from 33.79% to 36.59%.

- Impact of Dual-Task Protocol: This stability suggests that the primary source of stress captured by the vocal features was not the external traffic condition alone, but the constant cognitive load imposed by the dual-task requirement (Section 2.1.2). The need to continuously retrieve information and form opinions while simultaneously managing an operational task (driving in complex urban traffic) establishes a high and stable cognitive baseline, which the vocal system reflects as a sustained moderate stress state.
- Absence of High Stress: The absence of high stress regions (> 70%) in this protocol confirms that while the protocol successfully maintained a stressed vocal state, it did not push the highly experienced driver into acute or emotionally high-arousal stress states, which would typically be required to trigger the highest score thresholds used in the analysis.

4.2. Decoupling of Vocal Stress and Kinematic Difficulty

A key analytical finding from this experiment is the decoupling of average vocal stress from objective kinematic difficulty.

- Short Route Anomaly: On the short route, the kinematically most difficult drive (14:00, lowest average speed of 20.3 km/h) registered only a slightly higher average stress than the others. Furthermore, the 17:00 drive, anticipated to be the most stressful due to peak congestion, registered the lowest average stress and vocal instability on that route.
- Conclusion on Environment: These results lend support to the idea that once the environmental friction reaches a sufficient threshold (i.e., any heavy urban congestion), the vocal features plateau. For this driver, the voice reacted more to the presence of a high-friction environment and the simultaneous verbal task than to marginal temporal variations in congestion severity.

4.3. Sensitivity of Vocal Instability (Pitch Std. Dev.)

While the average stress score was consistent, the overall pitch standard deviation (Pitch Std. Dev) proved to be a more sensitive metric for differentiating between driving environments and time-of-day effects.

- Route Differentiation: The Pitch Std. Dev. was significantly lower on the long route (approx. 121.89 Hz) compared to the short route (approx. 129.79 Hz). This confirms the model's ability to detect the positive effect of the low-demand, stable-speed intercity segment in the hybrid route, which allows the driver to momentarily relax and achieve greater vocal stability.
- Time-of-Day Effect: The metric also captured subtle time-based differences, independent of the overall average stress. The short route showed the highest instability at 11:00, while the long route showed the lowest instability at 11:00. This complex pattern suggests that the driver's vocal baseline (e.g., initial energy, state of alertness, or diurnal rhythm) interacts differently with the two route types. The driver may be more prone to vocal perturbations when confronted with high-friction urban driving (short route) early in the day.

To provide formal statistical evidence despite the single-subject design, the observed variance in pitch standard deviation between the high-friction short route (M=129.79 Hz) and the hybrid long route (M=121.89 Hz) was contextualized against established benchmarks. Specifically, the magnitude of F0 variability shifts observed in this real-world dual-task protocol is found to be aligned with the findings of Boril et al. [17], by whom significant vocal perturbations under high cognitive load in driving environments were reported. Furthermore, the results regarding the stabilization of pitch in lower-demand segments (M=116.63 Hz minimum on the long route) are determined to be consistent with the ranges identified in the GeMAPS framework [16]. By these comparisons, it is validated that the observed vocal shifts are statistically meaningful and representative of stress-induced modulation rather than

stochastic noise; thus, the scientific rigor and the generalizability of the findings within the scope of this case study are further strengthened.

4.4. Methodological Considerations and Model Validation

The customized weighted stress index was found to perform robustly, successfully reflecting the sustained cognitive load induced by the dual-task paradigm. The decision to assign a higher weight to Pitch Standard Deviation was vindicated by its observed superior sensitivity, aligning with established literature on vocal perturbation under stress. While the threshold values were informed by prior research, these parameters were further optimized for the specific acoustic conditions of the vehicle cabin.

However, it is emphasized that since this study was conducted as a proof-of-concept with a single experienced driver, the reported stress levels and threshold values are considered subject-specific. These results are intended to demonstrate the sensitivity of the proposed acoustic biomarkers rather than to serve as absolute population-level benchmarks. Consequently, the interpretation of the findings is constrained by the individual's vocal baseline and driving experience.

Future work should be focused on validating these parameters against independent physiological ground-truth measures, such as ECG or GSR, in a multi-driver setting to establish universal calibration standards. Furthermore, while the speech elicitation protocol used emotionally neutral questions to minimize confounding effects from content, the potential interaction between semantic load, emotional valence of conversation, and acoustic stress markers remains an open and valuable question for future research, potentially through the utilization of natural language processing techniques.

4.5. Implications for Modeling

The results underscore the necessity of using multidimensional acoustic feature sets for stress detection. Reliance solely on average stress scores would obscure the differences between the routes. The high sensitivity of metrics like Pitch Std. Dev. to changes in driving context—even when average stress remains stable—validates their inclusion in the stress scoring model. The findings reinforce the concept that stress in a driving context is a dynamic construct best characterized by instantaneous perturbation measures (instability) rather than static feature averages.

5. CONCLUSION

This study has investigated the intricate relationship between vocal acoustic features and driver stress within a real-world dual-task protocol, specifically focusing on the distinction between cognitive load and environmental friction. The results indicate that the continuous demand of a verbal secondary task establishes a consistent moderate stress baseline approximately 34–36% which remains largely independent of routine traffic fluctuations and route characteristics. A primary contribution of this research is the identification of pitch standard deviation (vocal instability) as a more sensitive biomarker than aggregate stress indices for detecting environmental transitions. While average stress levels remained stable across diverse driving conditions, vocal instability significantly decreased during hybrid routes with low-density intercity segments, demonstrating its capacity to reflect the mitigating effect of reduced environmental friction. These findings suggest that for future in-vehicle monitoring systems, focusing on instantaneous vocal variability rather than mean stress scores may provide a more accurate assessment of a driver's state. To build upon these findings and overcome the limitations of the current single-subject pilot design, future research should involve a more diverse cohort varying in age, gender, and experience to enhance generalizability. Furthermore, it is recommended that future protocols integrate multimodal sensor fusion, such as combining vocal acoustics with eye-tracking or cabin-view computer vision, to better isolate the specific triggers of environmental friction in complex traffic environments.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the authors.

REFERENCES

[1] G. Matthews, P. A. Desmond, and K. Gilliland, "The stress of driving: A diary study," *J. Appl. Psychol.*, vol. 84, no. 4, pp. 613–620, 1999.

[2] S. Nandan, S. Mandal, and P. Ghosal. "Stress Detection and Monitoring: A Systematic Review." *2024 IEEE International Symposium on Smart Electronic Systems (ISES)*. IEEE, 2024, pp. 309-314.

[3] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, 2005.

[4] P. R. Anciaes, "Effects of the roadside visual environment on driver wellbeing and behaviour–a systematic review," *Transp. Rev.*, vol. 43, no. 4, pp. 571–598, 2023.

[5] M. A. Recarte and L. M. Nunes, "Mental workload while driving: Effects on visual search, discrimination, and decision making," *J. Exp. Psychol. Appl.*, vol. 9, no. 2, pp. 119–133, 2003.

[6] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 8, no. 2, pp. 97–120, 2005.

[7] J. Lee, H. Lee, and M. Shin, "Driving stress detection using multimodal convolutional neural networks with nonlinear representation of short-term physiological signals," *Sensors*, vol. 21, no. 7, p. 2381, 2021.

[8] H. Boril, P. Boyraz, and J. H. L. Hansen, "Towards multi-modal driver's stress detection," in *Proc. 10th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH 2009)*, Brighton, UK, 2009, pp. 1843–1846.

[9] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Trans. Syst., Man, Cybern. A Syst. Humans*, vol. 38, no. 3, pp. 502–512, 2008.

[10] G. Taskasaplidis, D. A. Fotiadis and P. D. Bamidis, "Review of Stress Detection Methods Using Wearable Sensors," in *IEEE Access*, vol. 12, pp. 38219-38246, 2024,

[11] G. Giannakakis, et al. "Review on psychological stress detection using biosignals." *IEEE transactions on affective computing,* vol. 13, no. 1, pp. 440-460, 2019.

[12] K. R. Scherer, et al. "Acoustic correlates of task load and stress." *Interspeech*. 2002.

[13] D. D. L. Veiga, et al. "The Fundamental Frequency of Voice as a Potential Stress Biomarker: A Systematic Review and Meta–Analysis." *Stress and Health,* vol. 41, no. 5, e70112, 2025.

[14] M. Van Puyvelde, et al. "Voice stress analysis: A new framework for voice and effort in human performance." *Frontiers in psychology,* vol. 9, no. 1994, 2018.

[15] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll, and A. Wendemuth, "Acoustic emotion recognition: A benchmark comparison of performances," in *2009 IEEE Workshop Autom. Speech Recognit. Understanding,* pp. 552–557, 2009.

[16] F. Eyben et al. "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 190–202, 2015.

[17] H. Boril, P. Boyraz, and J. H. L. Hansen, "Analysis and detection of cognitive load and frustration in drivers' speech," *EURASIP J. Audio, Speech, Music Process.*, vol. 2011, no. 1, p. 6, 2011.

[18] P. Lu, L. Tsao, and L. Ma, "Daily stress detection from real-life speeches using acoustic and semantic information." *Ergonomics,* vol. 68, no. 10, pp. 1694-1717, 2025.

[19] L. Beckes and J. A. Coan, "Social baseline theory: The role of social proximity in emotion and economy of action," *Soc. Personal. Psychol. Compass*, vol. 5, no. 12, pp. 976–988, 2011.

[20] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, 2011.

[21] T. F. Quatieri, S. N. Orozco, & D. Plotkin, "Vocal biomarkers to discriminate cognitive load in a working memory task". In *Proceedings of Interspeech*, pp. 1105-1109, Dresden, Germany, 2015.