UNIVERSITÀ CATTOLICA DEL SACRO CUORE
PhD In Psychology – XXXVIII Cycle
GSD 11/PSIC-01

Beyond the Black Box:
Vocal Biomarkers for Monitoring Stress in Real World Pilot–ATC Communication

Coordinator:
Professor Margherita Lanz
Supervisor:
Professor Federica Biassoni

PhD Candidate:
Martina Gnerre
ID number:
5214061

Submitted in fulfilment of the requirements for the degree
of Doctor of Philosophy
Academic Year 2024/202

All’eco dei miei avi

e alle voci da cui la mia discende

Acknowledgements

Writing a doctoral thesis is a solitary exercise. Yet over these years I have never felt alone.

My first thanks go to my supervisor, Professor Federica Biassoni. She believed in my ideas when they were still only intuitions and accompanied their development with patience, rigor, and intellectual generosity, granting me the most precious privilege: the freedom to try and to err.

I am grateful to Professors Moisés Betancort, a methodologist, and Jonathan Delgado Hernández, a speech analysis specialist, for their generous time and invaluable conversations. Our discussions on how to build a rigorous methodology for spontaneous speech analysis strengthened my conviction that voice research truly benefits from a multidisciplinary approach.

My path has also been enriched by dialogue with a vibrant community of scholars, met at conferences and in doctoral school lectures, and by the company of the authors who have lived on my desk, whose pages broadened my mind and sharpened my questions. To all of them, I owe my gratitude.

I also want to thank two colleagues who became friends along the way. To Ilaria, who taught me to reach beyond my limits. To Rossella, who taught me to welcome them with tenderness: thank you for your authenticity and steady support.

To my family and to Tommaso, I owe the belief in my stride.

Abstract

The human voice is a sensitive channel for inferring psychophysiological states and offers a noninvasive, ecologically valid biomarker of stress in safety-critical settings. Aviation provides a paradigmatic testbed, as pilot–air traffic controller officer (ATCO) exchanges frequently occur under time pressure, high workload, and unexpected emergencies. Despite accumulating evidence, findings on vocal stress markers remain fragmented by methodological heterogeneity, inconsistent results across contexts, and limited ecological validity. This dissertation advances understanding through three complementary studies.

Study 1 systematically reviewed 20 empirical investigations of vocal markers of workload, stress, fatigue, and sleepiness in aviation, identifying the most consistent acoustic correlates alongside recurrent limitations in design, measurement, and validation.

Study 2 analyzed a large archival corpus of authentic pilot–ATCO communications spanning emergency and routine phases. Twenty-seven speaker-normalized acoustic parameters were extracted with Parselmouth, and feature selection combined Welch ANOVAs with FDR control, collinearity screening, and the Boruta algorithm. Stress was then classified using supervised models (LDA, LASSO, Random Forest, and XGBoost), revealing both shared and role-specific vocal markers.

Study 3 introduced a geometric-morphometrics approach to speech, capturing stress-related spectrotemporal deformations beyond conventional acoustic features.

Taken together, these studies show that acoustic representations can index stress-induced vocal adaptations in ecologically valid aviation contexts and offer methodological guidance for future work.

The results refine psycholinguistic accounts of speech under stress and support the development of voice-based monitoring tools to enhance communication safety in aviation and other safety-critical domains.

Keywords**:** voice biomarkers; acoustic analysis; stress; human factors; pilot–ATC communication; safety-critical communication

TABLE OF CONTENTS

Preface

For most people, what happens inside an aircraft cockpit (and how language operates within it) remains hidden behind the closed cockpit door. As passengers, we hear only polite greetings and short announcements about weather or arrival times. The true linguistic and operational core of flight, the words through which coordination and decision-making occur, usually becomes visible only when something goes wrong. After an accident, investigators turn to the cockpit voice recorder, the so-called *black box*, to reconstruct not only what the pilots did but what they said. In aviation, the black box is an instrument of retrospection: it records voices and data but remains a mute archive until tragedy gives it meaning. Its purpose is forensic: to explain events after they occur.

This work, however, proposes a different approach, already declared in its title, 'Beyond the Black Box'. It is to listen to the voice not as posthumous evidence, but as a living signal of cognitive and emotional dynamics unfolding in real time. To move beyond the black box is to shift from reconstruction to anticipation, from analyzing failure to understanding the conditions that precede it. At the same time, the word "*beyond*" signals a broader aspiration: to transcend the reductionist view of voice as a mere technical signal and to restore it to its full complexity—embodied, relational, and situated—where data, emotion, and interaction converge. In this sense, *Beyond the Black Box* is not only a title but an epistemic stance.

This doctoral project grew from the conviction that voice deserves closer attention: not only as a channel of communication but as a psychophysiological marker of states that matter profoundly for aviation safety. The work that follows is organized into three complementary studies, each addressing a gap encountered along the way.

Looking back was the first step: a systematic review of the existing literature on vocal markers of stress, fatigue, workload, and sleepiness in aviation was performed. A central objective was to determine how stress can be differentiated from these adjacent constructs, clarifying both shared and distinct acoustic manifestations. This review revealed promising markers, but also fragmentation, methodological inconsistencies, and a reliance on laboratory simulations rather than real emergencies. Yet findings have remained fragmented, methodologies heterogeneous, and ecological validity limited. More often than not, investigations were confined to laboratory simulations, anecdotal reports, or narrowly focused acoustic analyses unable to account for the multidimensional nature of voice as a psychophysiological signal.

The second step was to look outward to the field itself, to authentic communications between pilots and controllers during aviation emergencies. By examining a corpus of 90 real recordings, I sought to look at the data as it unfolds in real time and environment, capturing voice not as a retrospective trace but as a living expression of action under pressure. Using supervised learning, these analyses identified robust, role-specific acoustic markers of stress in real-world conditions, revealing

not only what changes under strain but also how pilots and controllers regulate differently, shaped by their tasks, training, and the very nature of the risks they face.

Finally, the third step was to look forward. I experimented with geometric morphometrics, a methodology that does not originate in psychology or the sciences of voice but in fields such as biology, zoology, and palaeontology. In this framework, speech is modelled not as a set of isolated parameters but as a shape, a continuous surface spanning time, frequency, and amplitude. This line of inquiry arose from a paradox familiar to many scholars of phonation. Psychological states are often audible to human listeners, yet empirical models have struggled to capture them with reliability and consistency. This exploratory study indicates that acute stress produces measurable deformations in the geometry of the vocal signal that conventional features fail to represent, pointing to a new frontier in the study of speech as a marker of changes in cognitive and affective conditions. The work is a pilot and has clear limitations, yet it offers a proof of concept and outlines a program for future research.

Together, these three strands outline a path that reflects not only a scientific investigation but also a personal journey, an effort to know the human voice more deeply, to understand what it reveals when life hangs in the balance, and to imagine how such knowledge might help make aviation safer. I hope this dissertation will encourage others to carry this idea forward, expanding, adapting, and improving it across new contexts and realities.

I conclude with a brief reminder. Immersed in spectrograms and parameter estimation, one can forget that every acoustic trace is the expression of a living subject. Abstraction risks reducing voice solely to a hollow object of measurement rather than an act of presence. There is no voice without a hearer. For a time I lost sight of this. May I not forget it again.

CHAPTER 1

The rationale

The literature on voice and critical operational states in aviation is extensive yet fragmented. Three main issues hinder the translation of findings into practice: (1) conceptual heterogeneity, with overlapping constructs and non-uniform definitions; (2) low ecological validity, as many results derive from simulated tasks or controlled speech rather than real pilot–air traffic controller officer (ATCO) communications; and (3) metric reductionism and weak validation, characterized by a reliance on summary indices and limited estimation of event-level generalizability. There is a lack of cumulative synthesis, shared measurement standards, and sufficient attention to the temporal and structural dimensions of the vocal signal.

This combination of inconsistencies yields a crucial consequence: in the absence of robust and validated markers tested in real contexts, the potential of voice as an operational monitoring tool remains largely unrealized. The scientific and applied communities are thus left with promising yet fragmented findings, insufficiently integrated to be translated into reliable instruments for error prevention and safety assurance.

Against this backdrop, a central question remains unresolved: which vocal markers are robust, interpretable, and transferable under real operational conditions, and to what extent do they vary as a function of operational role (pilot vs. controller)? To address this gap, it is necessary to: (i) impose conceptual and methodological order; (ii) test markers on real speech with out-of-sample, event-level validation; and (iii) move beyond pointwise indices by adopting representations that capture the signal's temporal dynamics and shape structure. With this in mind, the present thesis develops across three studies, each building upon and justifying the next, to compose a coherent picture of the relationship between situational stress and vocal behavior. The thread uniting these three studies is a shared epistemic aim: to build a science of voice that integrates measurement, meaning, and context.

The systematic review clarifies the theoretical boundaries of the field; the empirical analysis tests its validity on ecological data; and morphometrics opens a novel avenue for representing vocal transformations. Together, these levels articulate a coherent proposal: to develop reliable tools for monitoring stress in real operational contexts and, more fundamentally, to redefine voice as an object of psychological and communicative inquiry. In this perspective, the meta-objective of the thesis is not merely to accumulate empirical evidence or refine analytical techniques, but to contribute to an epistemological reframing of how voice is studied: not as an epiphenomenon of language or a

physiological by-product of stress, but as an embodied and embedded form in which body, mind, and relationship are intertwined.

1.1 Scientific relevance

The present research aims to advance this field on three interconnected levels: theoretical, methodological, and practical.

At the theoretical level, it further situates voice as a *situated marker*—a signal that is at once biological and pragmatic—thus grounding a non-reductionist account of vocal indices that integrates physiology, interactional goals, and context. This entails a paradigm shift: from "universal markers" to adaptive, role- and context-dependent profiles, in which the same indices acquire different meanings as a function of pragmatic goals, workload, and environment.

At the methodological level, this study is grounded in real-world data and advances a measurement framework that prioritizes interpretability and replicability. It further introduces a shape-based approach (geometric morphometrics) to capture structural transformations of the acoustic signal that are not recoverable from scalar features alone.

On the practical level, it highlights the potential of early detection: the ability to recognize subtle vocal shifts linked to stress, in its various forms and origins, may enable the development of non-invasive monitoring systems for safety-critical environments. Despite decades of progress in human factors, accident analysis in aviation remains largely retrospective. This work pivots to anticipation: testing whether real-time changes in pilot–ATCO voice may serve as measurable stress signalers. If validated, voice analytics can be embedded within the Human Factors Analysis and Classification System (HFACS) informed architectures (Shappell & Wiegmann, 2001) to trigger proactive, role-sensitive interventions before errors or near-misses occur.

Ultimately, the challenge is epistemic rather than merely technical: to transform partial voice measures into a coherent, validated, and transferable framework—hence the scientific and applied relevance of this thesis.

1.2 The structure of the thesis

The thesis comprises six chapters. Two introductory chapters establish, respectively, the logic of measurement and the operational problem. Chapter 2 theorises what it means to measure voice and related critical aspects. Chapter 3 situates vocal measurement in aeronautical operations. Together, these chapters supply the context-specific requirements that guide the subsequent studies. Chapter 4 offers a systematic review of the literature, conducted according to PRISMA guidelines, identifying twenty relevant studies categorized by design, context, acoustic parameters, and psychological constructs investigated (stress, workload, fatigue, sleepiness). This review highlights two major

obstacles that have hindered cumulative progress so far: a *definitional problem*—since stress, workload, and fatigue are multidimensional and interdependent constructs—and a *metrological problem*, related to the high heterogeneity of tools, measures, and tasks. Building upon this foundation, the study presented in chapter 5 analyzes real-world communications between pilots and ATCOs during critical events. The use of ecological data serves two complementary aims: *ecological validity*, since only speech embedded in real operational activity reflects the full profile of situational pressures; and *role specificity*, as pilots and controllers operate within different contexts as far as role, function and risk level are concerned, and under distinct communicative constraints and responsibilities. The analysis integrates classical acoustic parameters informed by vocal physiology (e.g. F0, jitter, shimmer, HNR, spectral and rhythmic indices) with intra-speaker normalization and leave-one-event-out validation procedures, ensuring robustness and generalizability. Chapter 6 presents the third study, which introduces a conceptual and methodological shift. Voice is no longer represented as a sum of isolated parameters, but as a dynamic configuration analyzed through geometric morphometrics applied to fundamental frequency (F0) contours and spectrographic surfaces. This approach captures the overall shape of the signal and its transformations over time, revealing, even in small samples, interpretable differences between routine and emergency phases and distinct patterns across pilots and controllers. This is not a merely graphical exercise but a genuinely psychometric operation: reducing the complexity of the vocal signal into a space of forms where distances are statistically interpretable and physiologically and pragmatically meaningful.

1.3 Research questions

The present doctoral project investigates how vocal signals can serve as sensitive indicators of cognitive and emotional states in high-stakes operational environments. Each study addresses a distinct yet complementary research question.

Study 1 – A Systematic Review**.** This study aims to identify which acoustic parameters consistently reflect stress-related states in aviation. The guiding questions are:

1. Which vocal features reliably index workload, stress, fatigue, and sleepiness across studies?
2. How do methodological differences—such as acoustic parameters, software, and recording contexts—influence the identification of these markers?

Study 2 **–** Real-World Corpus Study. Building on empirical recordings from actual aviation events, this study explores the manifestation of stress in speech during real emergencies. The research questions are:

1. Which acoustic parameters significantly differ between routine and emergency speech in pilot and ATCO communications?
2. Are there role-specific stress patterns such that pilots and ATCOs exhibit distinct acoustic profiles due to their operational roles and perspectives during emergencies?
3. Can stress-related vocal changes be reliably detected and classified using acoustic features alone, and with what accuracy?

Study 3 – Methodological Case Study. This exploratory case study examines whether geometric morphometrics can capture subtle deformations in the acoustic "shape" of stressed speech that escape conventional analysis. The study asks:

1. Can geometric morphometrics detect stress-related deformations in the global spectro-temporal structure of speech?
2. Do geometric morphometrics-based descriptors provide greater sensitivity or interpretability than traditional acoustic features in distinguishing emergency from routine conditions?

Together, these studies aim to advance a unified framework for understanding how stress modulates vocal expression in operational contexts—integrating evidence synthesis, quantitative modelling, and shape-based analysis to identify reliable vocal biomarkers of situational stress.

CHAPTER 2

Toward a Measurement of Voice

Measuring voice means converting a living, complex phenomenon into selected numerical data; this conversion raises questions that are not merely technical but also theoretical and epistemological, as it entails decisions about what to measure and how to measure it. The act of measurement thus raises a central question: how can the human voice—simultaneously physiological, psychological, and relational —become an object of scientific inquiry without losing the complexity that defines it?

This chapter addresses that question looking at voice with a gradual movement from the sound to the symbolic, and from the symbolic back to the measurable. The first section, *From Sound to Sign*, examines the acoustic and semiotic foundations of vocal expression, considering how physiological processes give rise to perceptual and communicative indices. The second, *From Sign to Sense*, expands the discussion to the domain of meaning and sense, exploring how voice operates not only as a vehicle for linguistic content but also as a bearer of affective, relational, and pragmatic information. The third, *From Sense to Measurement (and from Measurement to Sense)*, confronts the epistemological and methodological challenges of quantifying vocal phenomena in relation to psychological constructs. Finally, *Challenges in Measurement* outlines the main conceptual, technical, and ecological issues that continue to shape the emerging science of vocal markers. Taken together, these sections delineate the theoretical framework underlying the empirical studies presented in the subsequent chapters. They aim to articulate the conditions under which voice can be rigorously analyzed as both a physical signal and a psychological phenomenon—a dual perspective that is indispensable for any science seeking to measure, interpret, and ultimately understand the human voice.

2.1 From sound to sign

To conduct an informed discussion of the phonetic and phonological phenomena that act as markers in speech, one must begin with the physical nature of voice. Vocal production results from the highly refined coordination of interconnected muscular systems: the breath providing energy and modulating airflow; the larynx initiating or inhibiting phonation; the pharynx and velopharyngeal port regulating the size and pathway of resonance; the tongue shaping the oral cavity into rapidly changing configurations; and the lips and mandible refining the overall articulatory profile (Magnani & Fussi, 2021). As Laver (1975) observed, none of these actions occurs in isolation; each motor gesture co-implies synergistic adjustments across the entire apparatus. From this dynamic, measurable acoustic variations emerge, which in turn correspond to auditory perceptual dimensions: pitch as the correlate of

F0, loudness as the correlate of intensity, and duration as the temporal correlate (Nguyen & Madill, 2013).

Other dimensions include timbral quality, frication, and noise, with the well-documented contributions of jitter and shimmer—that is, cycle-to-cycle irregularities in F0 and amplitude that lend the voice a "rough" or "hard" texture (Rabinov et al.,1995). A sound becomes a sign when it is not merely registered as an acoustic event, but is recognized as standing for something beyond itself, prompting the listener to attribute it to a source, a state, or an intention within a shared interpretive frame (Pesina & Solonchak, 2014).

What physiology organizes as sound, perception receives as sign: patterned acoustics become interpretable indices (Latinus & Belin, 2011). The acoustic outcomes of coordinated articulation function as indices that listeners routinely interpret; however, this step from sound to sign is only apparently short and simple (Pisanski & Bryant, 2019). Compared with verbal language, whose pairing of signifier and signified is stabilised by convention, as described by De Saussure (1985) and often illustrated by the Ogden and Richards triangle (1923), the nonverbal vocal domain is certainly less formalised and remains highly context-dependent.

From a semiotic perspective, voice is not merely a vehicle for verbal language but a system of indices that "stand for" something else, in accordance with Peirce's well-known triadic model of the sign (Peirce 1965)[1]. Explicitly referencing Peirce, Abercrombie (1967) proposed an equally explicit typology of identity indices in speech: Group-membership indices, which locate the speaker socially and geographically; Individualizing indices, which distinguish them as a person; Condition indices, which are sensitive to contingent fluctuations, for instance, in affective state. A useful, cross-cutting classification can be superimposed upon this tripartite scheme, one organized by social, physical, and psychological properties. Voice marks affiliations and roles, age and sex, health and temperament, personality traits and emotional states (Ciceri & Anolli, 2000). Some markers are relatively stable, tracing the speaker's biography; others are transient, reflecting momentary conditions. This conceptual map is further refined by the distinction between communicative and informative signals. A signal is communicative when a speaker deliberately produces it to make a listener aware of something; it is informative when, irrespective of the speaker's intention, it nonetheless makes knowledge available to the listener (Lyons, 1977). The words a speaker selects typically fall within the communicative domain; the manner in which they are spoken—timbre, intonational contour, a barely perceptible tremolo, accent—primarily fuels the informative domain. Elaborating on this, Lyons revisits and expands the concept of the "symptom": an index that not only informs but signals a state of the emitter, be it emotional (fear, anger), physiological (laryngitis), pharmacological, or toxicological. The symptom is,

[1] For a more detailed discussion of the ideas developed in this section, see Ciceri, M. R., & Anolli, L. M. (2000). *La voce delle emozioni. Verso una semiosi della comunicazione vocale non-verbale delle emozioni*; and Laver, J., & Trudgill, P. (Eds.). (1979). *Phonetic and Linguistic Markers in Speech*. Academic Press.

so to speak, a diagnosis in progress that exceeds the subject's conscious control. This distinction is conceptually fruitful but not entirely tenable. For instance, a sigh can be an involuntary symptom of fatigue (informative) or a communicative gesture to express resignation or disapproval. The crux of the matter is that there is no direct access to a speaker's intention[2]. We can rely only on observable evidence within the relevant contexts.

This foundation explains the rapid ascent of the concept of a vocal biomarker: the systematic analysis of voice as a measurable proxy for states of health, disease, and psychological condition, including transient states such as emotions (Scherer, 2013b). Depending on the objective, researchers employ sustained vocalizations, read speech, or spontaneous speech. In neurology, many efforts focus on Parkinson's disease, seeking acoustic signatures of hypokinetic dysarthria (e.g., reduced F0 variability, increased aperiodicity) for screening and disease monitoring (Gnerre et al., 2023; Skodda et al., 2012; Suppa et al., 2022). Parallel lines target other medical conditions such as schizophrenia (Abbas et al., 2022), autism (Rybner et al., 2022), depression (Abbas et al., 2021) and acute infections such as COVID-19 (Ismail et al., 2021). The field is thus moving beyond descriptive effects toward robust, generalizable markers suitable for early detection, stratification, and tracking.

2.2 From sign to sense

Since antiquity, and especially since Aristotle, voice has been conceived primarily as an instrument of language: the *phoné semantiké*, the "meaningful voice." In this perspective, voice is regarded as a medium through which a speaker conveys something to a listener, who in turn seeks to understand. Yet such a reduction overlooks what voice contributes beyond words. Voice is not only a vehicle for language; it also conveys bodily, emotional, relational, and aesthetic dimensions that cannot be reduced to the simple act of "saying something to someone" (Ciceri & Anolli, 2000). To systematically account for this complexity, theoretical models like Klaus Scherer's TEEP (Tripartite Emotion Expression and Perception) are invaluable (Scherer, 2013a). Building on Bühler's organon model (1934) and Brunswik's lens model (1952), the TEEP framework distinguishes three primary functions of paralinguistic voice:

1. Symptom (of internal states): Voice as an immediate, often involuntary, reflection of the speaker's physiology and emotion.

[2] From a sociolinguistic perspective, what a dominant community decides to treat as a "symptom" can convey value judgments and function as a mechanism of exclusion (Lippi-Green, 2012). A regional accent, a prosody deemed "too loud" for a woman, or an interactional rhythm judged as "vulgar" can be retrospectively pathologized. This serves as a methodological caveat: no one has direct access to a speaker's intentions; many signs are hybrid and slide from the informative to the communicative domain depending on context, the listener's attribution, and the strategic use speakers make of their vocal resources.

2. Appeal (to the listener): Voice as a tool to influence, persuade, or elicit specific inferences and reactions in the listener.

3. Symbol (of convention): Voice as a carrier of culturally shared and learned meanings (e.g., a specific tone denoting sarcasm).

These functions are not mutually exclusive but are interwoven in every vocal utterance, providing a robust grid for parsing voice's multifaceted nature. Crucially, this does not imply that meaning is literally carried inside the vocal signal as portable content. Rather, the acoustic form functions as a physical sign that triggers interpretation: what reaches the listener is the material trace, while sense is reconstructed within the listener's own cognitive domain through decoding, inference, and contextual integration (Pesina & Solonchak, 2014). In this view, voice can be treated as a sign, a physical acoustic trace, yet what matters is the sense that listeners construct from it in context (Ciceri & Anolli, 2000). This distinction can be read as broadly resonant with Frege's classic separation between '*Sinn and Bedeutung'* (1892), while adopting a different theoretical aim: rather than grounding meaning in a strictly logical account of reference, it foregrounds how sense is situationally reconstructed by listeners through inference and contextual integration. This sense is not limited to the lexical semantics of words: it includes what can be inferred from vocal cues as symptoms of inner states or as appeals directed to the listener, it includes meanings stabilized by convention, and it also includes a more resistant layer linked to embodiment and individuality. For this reason, voice can not be confined to a purely auxiliary role within the verbal code (Ciceri & Anolli, 2000). Rather, it is at once the embodied sign of individuality and the singular imprint of being-in-the-world, which both accompanies and exceeds the semantic content of speech. This surplus becomes evident in the irreducible singularity of voice. Like a fingerprint or a person's gait, it cannot be fully captured by the logic of linguistic intersubjectivity, which requires neutrality and standardization in order to function. As Federico Albano Leoni observes (2024), language and its structures (e.g., grammar), as a shared code, require that individual idiosyncrasy not intrude upon the exchange. Voice, however, even while sustaining language, preserves a surplus that resists absorption—the singular timbre of a body and the unrepeatable imprint of a person. The first conclusion is therefore clear: voice is not merely a support for linguistic content but also a bearer of sense in its own right. Conversely, language also exceeds voice, taking the forms of writing, sign systems, and digital inscriptions. From this perspective arises the theoretically plausible idea of constructing lexicons even for nonverbal systems: repertoires linking signals to meanings. Yet the task remains difficult. Segmenting signals and isolating their elementary units is inherently complex, and the map of meanings is never fixed once and for all. It must be reconstructed case by case, by examining what information a system actually transmits, through which modalities, and under which conditions. For this reason, any "lexicon of voice" can only be a provisional map: useful, but partial. Finally, voice

also operates on a pragmatic level: it does not only carry meanings but also performs actions. Through intonation, rhythm, pauses, or vocal gestures such as laughter and sighs, voice regulates turn-taking, signals stance, and shapes the force of speech acts. Voice does not simply signify; it actively performs actions in the world.

2.3 From sense to measurement (and from measurement to sense)

Once we acknowledge that voice does not merely convey meaning but is itself an embodied form of sense—capable of communicating intentions, states, and relationships along a gradient of intentionality ranging from involuntary symptom to deliberate expressive act—the question of measurement inevitably arises. In my view, this is not an easy issue to navigate: the epistemological foundations of measuring vocal aspects in relation to psychological constructs rest on hybrid ground, where the measurement of the signal (objective in principle) meets the psychological construct, which is always permeated by subjectivity. The epistemological challenge lies in developing instruments capable of honoring this dual belonging.

From voice, we can grasp traces, signs, and configurations of psychic processes, but never their entirety. Epistemologically, this serves as a reminder that there is an ontological distance between the psychological phenomenon and its vocal manifestation. When we measure voice, we are not directly measuring anger, fear, or sadness; we are measuring an acoustic expression that accompanies them. This distinction preserves conceptual rigor. Without it, we risk a reductionist confusion—as if voice were the psyche itself.

In reality, voice is an embodied index, a way in which the inner world reveals itself through the body and incarnates in the outside world. Measuring voice therefore involves three complementary levels that, rather than competing, inform one another: perceptual evaluation, which provides a descriptive and phenomenological measure; acoustic analysis, which translates that experience into explicit physical parameters; and AI-driven analysis, which allows hypotheses derived from the previous two levels to be tested and integrated on a larger scale. Individual self-reporting is, of course, another recognised approach; however, it will not be addressed in the present work, as its application pertains primarily to clinical settings (Carding et al., 2009).

Instrumental acoustic measures rest on the assumption that the functioning of the human vocal system is reflected in measurable variations of the acoustic signal: the physiological modifications are thus inscribed in the changeable organization of voice in the sound itself, and the challenge lies in deciphering it without impoverishing it (Kent & Kim, 2008). The perceptual level offers the experiential anchor: it does not measure what “is” in the signal, but what listeners subjectively perceive (Biassoni et al., 2022). This requires a precise definition of the listening unit and the task (global judgments, interval scales, continuous evaluations, paired comparisons), clear semantic anchors, and control for order, fatigue, and familiarization effects. The resulting data must be treated according to psychometric

criteria—estimating inter- and intra-rater reliability, internal consistency, sensitivity to change, and verifying convergent and discriminant validity. Acoustic analysis, by contrast, marks a change of domain with respect to listening: a passage from the phenomenology of experience to the physics of sound. By decomposing the physical signal into measurable quantities, it offers undeniable advantages: it can detect sub-perceptual variations (such as micro-fluctuations of fundamental frequency or energy within specific spectral bands) that escape human hearing. Perception, in fact, integrates dozens of vocal parameters unconsciously into a single global percept, expressed by the listener through a gestaltic judgment ("tense voice," "calm voice"). Acoustic analysis allows these dimensions to be disentangled, making it possible to ask whether a judgment of "tension" depends more on an increase in F0, a change in spectral tilt, or jitter. This decomposition is essential for testing theoretical models describing how different physiological mechanisms—muscular tension, respiratory control, vocal tract configuration—translate into acoustic modifications. Yet it is also here that the zone of risk opens up. An acoustic parameter is "objective" only within its own physical domain; its interpretation is far from objective. What does a 20 Hz increase in F0 mean? Does it stem from anger, joy, or mere vocal effort? Acoustics tells us what changes, but not why it changes. Without anchoring in perception and psychological theory, we risk constructing a catalog of ghosts—precise numbers devoid of meaning. AI-driven analysis makes it possible to explore vast amounts of data, identifying regularities and patterns that would otherwise escape human observation (Liu et al., 2024).

In principle, this ability to scale and generalize opens a new phase in vocal research: the opportunity to test psychological hypotheses on a large scale with quantitative rigor and replicability. However, AI-driven analysis does not observe a new domain of voice—it processes it through a different inferential regime. Unlike traditional acoustic analysis, which is hypothesis-driven, AI-driven analysis is data-driven: it relies on algorithms capable of detecting recurring patterns within large datasets, often beyond human perceptual or analytical capacities. This perspective does not replace theoretical analysis but subjects it to large-scale empirical testing, providing a means to verify the coherence of psychological and physiological hypotheses. Yet a change in technique does not imply a change in object. Algorithms learn correlations, not meanings. They operate on numerical representations of acoustic phenomena and, if deprived of an adequate theoretical and perceptual framework, risk producing statistically robust but conceptually hollow distinctions.

From this perspective, the three levels of analysis do not represent alternative approaches but rather different degrees of approximation to the same phenomenon. Perception provides the criterion of relevance, acoustics supplies the explanatory mechanisms, and automation tests the coherence of these relations on a broader empirical scale.

2.4 Challenges in measurement

Despite the abundance of studies and the methodological enthusiasm that characterize the current "golden age" of vocal markers, it remains difficult to identify parameters that are genuinely robust, reliable, and generalizable (Kalia et al., 2025). Research continues to grapple with a series of conceptual, methodological, and technical knots that, if left unresolved, risk condemning the field to oscillate between promising regularities and recurrent contradictions. These knots are not mere obstacles to be eliminated but structural tensions that reveal the complexity of studying voice as both signal and phenomenon. These reflections presented in this section represent a personal perspective developed in the context of studying vocal and psychological measurement. Although grounded in empirical literature, what follows should be read as a theoretical stance on the challenges inherent in quantifying complex psychological constructs (such as emotion, stress, or cognitive load) through vocal indicators.

Epistemological knot: voice has many souls

The first knot is epistemological. Literature reveals a plurality of perspectives: for some, voice is primarily body—a mirror of physiology (Casper & Leonard, 2006); for others, it is behavior—a communicative action sustained by intention (Engeström, 1995); for still others, it is expression—a trace that escapes volition, revealing inner states (Anikin & Lima, 2018). This plurality is not a defect to be corrected but a fact to be acknowledged and integrated. Voice can certainly be studied from a single vantage point, yet such a stance must be made explicit. In my view, the phenomenon itself exceeds any single framework and demands cross-disciplinary competence capable of bridging physiology, phonetics, pragmatics, and psychology. Without such integration, the risk is a form of blind specialization that develops increasingly precise instruments for increasingly narrow questions, at the cost of losing sight of the multifaceted nature of the vocal phenomenon.

Knot of universality: there are no invariant vocal markers

The second knot concerns the enduring belief in universal, invariant vocal markers of emotion or psychological states. A central issue lies in the assumption of fixed expressive prototypes for each emotion. Scherer and Ellgring (2007) demonstrated that even in facial expression research—where coding is more discrete than in voice and universal configurations have been hypothesized since Darwin (Darwin, 1872; Ekman, 1992)—evidence for such prototypes remains weak. As Scherer (2009) notes, emotion processes guided by appraisal produce highly variable expressive configurations across modalities, shaped by individual differences and situational factors, making the existence of stable prototypes unlikely. The literature reports some consistent trends (for instance, increased F0 under

stress) but each regularity is accompanied by a constellation of exceptions related to context, stimulus, instruction, language, or register. It is no coincidence that F0 appears to correlate with almost any psychological state; this classificatory paradox forces us to reconsider universalistic claims. Under such conditions, defending the notion of fixed, general markers becomes untenable. It is more fruitful to think in terms of adaptive profiles—patterns sensitive to physiology, context, and task—and to redefine a "marker" as a constellation of indices that gains meaning only within a specific vocal history, for a specific person, in a given moment.

Knot of temporality: voice as a dynamic process

The third knot concerns temporality. Voice is intrinsically non-stationary, constituting a dynamic process rather than a static object (Ramalingam, 1995). Its reduction to a set of discrete, punctual measurements fundamentally misrepresents this continuous flow (Magnani & Fussi, 2021). While this temporal continuity facilitates instrumental capture, it presents significant challenges for rigorous analysis, including problems of segmentation, serial dependence, parameter co-variation, and non-stationarity, wherein the properties of the signal are contingent upon both its past and future states. Furthermore, voice operates on a dual temporal scale, simultaneously reflecting transient states—such as fleeting emotions or momentary cognitive load—and enduring traits, such as those linked to age, physiology, or long-term sociocultural marking (Murray et al., 1996; van Mersbergen & Lanza, 2019). A failure to account for this duality leads to a critical, symmetrical error: the risk of misinterpreting an ephemeral fluctuation as a stable characteristic, or conversely, of attributing a stable, identity-marking feature to a transient state. A rigorous epistemological and methodological approach must therefore be sensitive to these multiple temporal layers and develop analytical techniques capable of disentangling the transient from the durable to construct a valid and nuanced interpretation of the vocal signal.

Knot of wholeness: voice as a gestaltic process

The meaning emerges not from isolated acoustic segments, but from temporally extended perceptual gestalts (Di Salle, 2003; Tenney & Polansky, 1980). Listeners perceive and interpret speech not as a sequence of discrete units, but as an integrated stream where segmental cues, prosodic contours, coarticulation patterns, lexical knowledge, and contextual constraints interact synergistically. Word recognition, for instance, is a global computation that relies on this dynamic synthesis rather than on local, atomistic analysis. This gestaltic nature demands methodological approaches that respect the architecture of the signal as a gestalt. As Magnani and Fussi (2021) aptly note, measuring a single parameter is like examining one tree to infer the health of the entire forest. Local measures can be legitimate and useful, but when detached from the whole, they risk becoming abstract or misleading. The task is not to abandon quantitative indices but to integrate them within models that preserve the

dynamic and perceptual integrity of the signal—reconciling index and gestalt, detail and whole, measurement and perception—so that the richness of voice is not lost in its reduction to data points.

Knot of ecological speech: the need to study the living voice

The fourth knot exposes the limits of experimental design and the persistent divide between phonetics and expressive voice. Much of current research, rooted in laboratory paradigms, ends with the ritual call to study natural contexts—yet few actually do so. The obstacle is not technical but epistemic: an attachment to the order and repeatability of the lab, even at the cost of overlooking the richness and unpredictability of living speech. On the other hand, it would be simplistic to oppose laboratory speech as "artificial" and real-life speech as "authentic." As Scherer's push–pull model reminds us, everyday emotions are never purely spontaneous but are regulated by cultural and situational constraints (Scherer, 1985, 1986, 1987, 1988). The key issue is not authenticity but embeddedness: spontaneous speech differs structurally from induced or read speech because it unfolds in interaction, marked by pragmatics and context. Understanding how voice truly functions in lived communication requires moving beyond the safety of the laboratory, even at the cost of analytic messiness (Tawari et al., 2010; Zhang et al., 2017) . While phonetically balanced material is easier to analyze with precision, perhaps a modest loss in accuracy could yield a gain in ecological validity. Equally problematic is the tendency to separate linguistic articulation from affective modulation, as if phonetics and expression belonged to distinct domains. This is a false dichotomy. Language and emotion are not parallel channels but interwoven dimensions of a single vocal act, in which structure and affect co-construct the acoustic event. The task, then, is not to abandon phonetics but to reconceive it within an integrated framework—refining traditional measures while complementing them with metrics that capture temporal form, distributional structure, and contextual dynamics. Only in such a framework we can approach voice as a unitary phenomenon, where body, language, and affect simultaneously converge in global expression.

Technical–infrastructural knot: every measure is dependent on its measurement pipeline

The fifth knot is technical and technological—and perhaps the most underestimated. What we call an "acoustic measure" is always the outcome of a methodological chain: microphone quality, recording environment, distance, noise and digitization all shape the signal. This produces a structural gap between controlled, professional recordings and data collected in ecological contexts such as social media, smartphones, or teleconferencing. Variability also stems from discrepancies among analytical tools: a parameter computed by Praat may not coincide with the same parameter estimated by MDVP or other systems, and even small changes in window size can alter results. Unless pipelines are transparent, documented, and replicable, cross-study comparisons become arbitrary. Technical

literature has long emphasized these issues, showing how variability arises from recording setups (Deliyski et al., 2005), software and algorithms (Amir et al., 2009; Bielamowicz et al., 1996; Hirose et al., 1992) or microphone choice (Parsa et al., 2001). When research moves from the lab to real-world speech, these sources of variability multiply—intersecting with differences in accent, language, and interactional style. Any claim to universality must therefore be tempered by an awareness that every acoustic measure is, by nature, situated.

Knot of listening: the forgotten dimension

The sixth knot concerns listening—a dimension too often neglected in the science of voice. As De Mauro once observed, there is a profound asymmetry between the lexicon of speaking, which is rich and articulated, and that of listening, which is sparse and elusive (De Mauro 1994). Speaking is an external, observable act: one can see the movements of the larynx, tongue, and lips. Listening, by contrast, is an interior act, invisible and difficult to thematize (Leoni, 2001). We cannot observe our auditory system—or that of others—while it functions. Yet in any interaction we are double listeners: we listen to the other and simultaneously to ourselves listening to the other. Ignoring this dimension undermines any attempt to link acoustic parameters with psychological states, for the bridge between the physics of the signal and the phenomenology of experience passes not only through curves and numbers but through the act of listening itself. It is in this subtle and often unnameable space that voice acquires meaning and its variations become shared experience.

CHAPTER 3

Measuring Stress in the Aviation Environment through Voice[3]

In cockpits and control towers, voice does not play the ornamental role of a mere expressive vehicle; it fulfills a genuinely operational function (Prinzo & Britton, 1993). It coordinates sequences of action, allocates multilocal attention, disambiguates priorities, and modulates access to cognitive resources along tightly timed trajectories. Within this space of technical and temporal constraints, a range of psychophysiological states such as stress, workload increase, fatigue, and sleepiness are not occasional incidents but structural components of professional practice (Terenzi et al., 2024; Vagner et al., 2018). These factors, can precipitate human error, generating consequences that span a spectrum from minor inefficiencies to catastrophic disasters (Stokes et al., 2017). Furthermore, research has established that chronic exposure to flight-related stress can lead to long-term health consequences, including post-traumatic stress disorder, anxiety, depression, and musculoskeletal issues such as back and neck pain (Masi et al., 2023).

Their consequences on vocal behavior are far from epiphenomenal. Alterations in intonation, energy, harmonicity, and speech rhythm reflect adaptive adjustments of the phonorespiratory system and, in many cases, signal (often with remarkable immediacy) an impending degradation of operational performance (Hagmüller et al., 2006). In particular, stress represents a significant concern in aviation, as it can critically compromise human performance (Masi et al., 2023). It is a multi-level phenomenon that spans physiology, cognition, and emotion, and its effects are inscribed in voice across physiological, psychological, and psychosocial dimensions (Hagmüller et al., 2006). It induces respiratory–laryngeal adjustments, reshapes attentional allocation and decision processes, and reorganizes interactional registers and roles. These changes surface as diffuse acoustic correlates which render voice a uniquely integrative indicator of situational stress at the interface of body, mind, and context.

The study of vocal communication in aviation thus lies at the crossroads of language sciences, cognitive psychology, and safety disciplines. Despite the growing attention devoted to human factors

[3] This chapter is drawn from: Gnerre, M., & Biassoni, F. (2024). Marcatori vocali per la detezione di stati di stress nel contesto dell'aviazione: Verso un nuovo framework di analisi. In F. Biassoni (Ed.), *Il fattore umano in aviazione: Sfide e frontiere in una prospettiva interdisciplinare* (pp. 71–91). EDUCatt. https://hdl.handle.net/10807/314099.

in aeronautics, voice remains both omnipresent and underestimated: omnipresent because it is the primary channel through which pilots and controllers coordinate action and manage critical situations; underestimated because it is too often reduced to a vehicle of verbal content, overlooking its nature as a sensitive indicator of psychophysiological states.

3.1 The air–ground communication

Air–ground communication is the backbone of aeronautical operations (Farris & Molesworth; Mahmoud et al., 2014; Mosier et al., 2013). It enables the transmission of commands, the continuous exchange of flight status, and the management of decisions across the entire operational profile. It is the channel through which pilots, ATCOs, command and maintenance centers, and other stakeholders braid intentions and situational representations into coordinated action. Early avionics relied on analog radio to make voice-at-a-distance practicable, allowing pilots and airports to "understand" one another in real time despite environmental friction (Gangl, 2006). As technology evolved, the introduction of data links and, later, satellite infrastructures added a more structured circulation of information alongside voice, improving the monitoring and management of the flight process (Mahmoud et al., 2014). Yet in time-critical windows, voice remains the principal interface: no other medium combines comparable temporal immediacy with such a dense capacity for coordination. This centrality becomes clear when the technical and human planes are considered together.

On the technical side, the communication architecture serves diverse needs—air–air, air–ground, space–earth, and satellite—and employs multiple modes, from voice calls to data exchange and networked services. Frequencies, bandwidth, and service profiles vary with the operational context, but radio voice continues to anchor tactical coordination. Aeronautical radiotelephony, however, operates far from ideal conditions: limited bandwidth, equipment imperfections, hiss and static, and, above all, elevated cockpit noise (Özmen et al., 2024, Jang et al., 2014). Speaker and listener face a persistent problem of auditory discrimination; extracting words, codes, and signals amid noise and interference is intrinsically challenging (Wu et al., 2019). Historical operational literature has pointed to enduring levers of improvement: targeted training in speaking and lexical choice, better headset shielding and fitting, higher fidelity reproduction, specific signal-processing transforms where appropriate to enhance discriminability, and the careful selection of signals and codings that maximize perceptual distinctiveness (Alketbi & Sipos, 2025; Geacăr, 2010).

On the human side, air–ground communication inhabits a structural tension. Voice is rapid and information rich, yet vulnerable to the very forces it must help govern. Noise masks; bandwidth constrains; transmissions can overlap or be blocked; and, crucially, the human factor imposes cognitive limits that tighten precisely when time does (Hagmüller et al., 2006). Under load and stress, working memory thins, attention fragments, and the perceptual field narrows; reliance on expectations and schemas grows, increasing the risk of hearing what one anticipates rather than what was actually said.

Standard radiotelephony phraseology exists to contain this risk (Drayton & Coxhead, 2023; Estival et al., 2023). The International Civil Aviation Organisation (ICAO) prescribes this standardized usage and associated procedures through its standards and recommended practices and guidance material, and requires operational English proficiency to reduce ambiguity, ensure mutual intelligibility among crews and controllers, and preserve safety margins in both routine and emergency communications (ICAO, 2004, 2010).

By reducing lexical entropy, stabilizing message formats, and making the lexicon uniform, standardized, and repeatable, it improves the reliability of the readback/hearback cycle and disciplines call-sign usage. Unlike plain language, whose meaning can shift with culture and context, these fixed expressions are built to deliver an exact operational meaning. In some circumstances, like emergencies, however, time pressure can still drive elisions, repairs, and noncanonical forms (Molesworth & Estival; 2015).

A stratified picture thus emerges. At the physical–technical level, signal and peripheral quality, frequency selection, link reliability, and message format design determine how much information survives the environment. At the psycho-linguistic level, message structure, phraseological competence, rhythm and prosody of delivery, and listening training directly affect the probability that what is said will be correctly understood and transformed into appropriate action. Added to this are the internal interpretive schemas—cognitive, affective, and cultural—that filter perception and decoding through expectations, communicative habits, and shared models of reality. At the organizational level, procedures and roles orchestrate who speaks, when, how, and to what end, distributing attention across concurrent tasks and narrowing the margin for error (Hagmüller et al., 2006). Within accident reports and the research literature, when radio communication is cited as a contributing factor, the most immediate and recurring explanation is often linguistic: a “language problem,” insufficient proficiency in aeronautical English, or a misunderstanding attributed to poor language skills on one or both sides (Alderson, 2009; Molesworth & Estival, 2015; Tajima, 2004; Wu et al., 2019). This interpretation, while never wholly wrong, is reductive. From the author’s perspective, many so-called ‘language-related’ events arise from misalignments between code-level form, prosodic realization, perceptual conditions, and interactional management under load. These are better framed as failures of coordination in which agents share an operational code but not the same models for interpreting it in context. Within this frame, standard radiotelephony phraseology contributes to safety by reducing degrees of freedom, making the sequence and format of information predictable, and enforcing robust readback/hearback loops. The ICAO Manual states that pilots and controllers should use standard radiotelephony phraseology whenever it is available, and resort to plain language only when the standard wording cannot convey the intended message, recognizing that in some situations fixed phrases are inadequate for successful communication (Alketbi & Sipos, 2025). This design increases tolerance to acoustic adversity, attentional limits, and cross-cultural heterogeneity. It does not remove human

vulnerability; it builds bounded robustness while preserving the flexibility required for off-nominal events where plain language is appropriate.

3.2 Stress in aviation

In 1903, the era of aviation began with an inception that—although marked by challenges and tragedies—has today achieved remarkably high safety standards, mainly thanks to advances in technology and aeronautical knowledge. Despite the considerable increase in the number of flights over the past decades, the rate of accidents caused by technical malfunctions has drastically decreased. Nowadays, safety risks are primarily associated with human factors, which include both individual characteristics of the actors involved and organizational dynamics. Indeed, most experts estimate that between 60% and 80% of aviation accidents are partially or entirely attributable to human error. The human element thus represents not only the most flexible, adaptable, and valuable component of the aviation environment, but also the most vulnerable one. A review of air accidents over the past fifty years shows that the majority have been caused by at least one of the following factors: lack of experience, insufficient training, voluntary risk-taking, distraction, reduced decision-making capacity, misjudgment of phenomena, communication errors, or stress (Lyssakov & Lyssakova, 2019; Shappell et al., 2006).

Among these, stress appears to be one of the factors with the most detrimental impact on safety in the aviation environment. The etymological origin of the word *stress* is related to the idea of "pressing" or "tightening" and derives from the Latin *strictus*, metaphorically suggesting a sense of tension, constraint, and even anguish. Stress can be defined as a subjective experience that arises when a person is subjected to environmental pressures that require adaptation or change. This description implies that stress encompasses both environmental demands and individual reactions. A broad definition that highlights the interaction between external stimuli and the subjective evaluation of environmental demands is provided by Cooper and Payne (1980), who describe stress as a "phenomenon that occurs when a person encounters events, or characteristics of events, perceived as significant for their well-being and as exceeding their coping resources" (p. 2). Recent studies suggest that work-related stress can have negative consequences on the health and well-being of aviation personnel, and consequently on performance and flight safety (Cahill et al., 2020). The aircraft environment itself—at high altitude, with noise, vibrations, ionizing radiation, and poor cabin air quality—can already induce a potentially stressful condition (Hagmüller et al., 2006). To this, we can add other specific stressors inherent to this context: some directly stem from workload, such as multitasking or unexpected events, while others are related to working conditions, such as circadian rhythm disruption and sleep deprivation. Moreover, the complexity of managing one's personal life

may introduce additional challenges, leading to relational or family tensions (Cahill et al., 2020). Even the frequent medical and psychological assessments required in the field can contribute to a state of chronic tension (Lempereur & Lauri, 2006). These stressors can have long-term consequences both at the individual level—potentially contributing to the development of psychotic conditions—and at the collective level, as evidenced by the high absenteeism rates in this sector. In light of this scenario, current efforts are converging toward the development of tools specifically designed to measure stress in the aviation industry, taking into account its contextual specificities and the various stressors that may occur within the aircraft or the control tower. The assessment of stress in aviation therefore aims to decode and measure the psychological stress experienced by pilots and other professionals, with the ultimate goal of enabling timely interventions to restore safe operating conditions.

The assessment of stress and workload typically relies on self-report measures, performance evaluations, and objective assessments (Hagmüller et al., 2006). Within this framework, vocal behavior analysis emerges as a promising tool—non-invasive, cost-effective, automatable, and capable of providing objective data that are difficult to manipulate. This potential was already well recognized in September 1995, during the NATO "Speech Under Stress" seminar. Since then, this methodology has been repeatedly applied to investigate the psychological state of both astronauts and pilots (Johannes et al., 2000; Kikuchi & Oagawa, 2018). For instance, as early as 1965, it was used to assess the mental condition of the first Russian astronaut during the first spacewalk. The National Transportation Safety Board (NTSB) also adopted the analysis of F0, retrieved from cockpit voice recorders, to examine the emotional state of a suicidal pilot at the time of a crash (National Transportation Safety Board, 2001).

However, two main issues currently emerge. First, vocal analysis is still predominantly used as a retrospective tool—that is, after an accident—rather than as a means to detect stress in real time and potentially prevent critical events. Second, the extensive literature on the topic, often lacking a strong psychological perspective, converges toward a conceptualization and phenomenology of stress that remain vague, fragmented, and sometimes overly polysemic. Within this blurred paradigm, the distinctive, multilayered, and dynamic nature of stress is often overlooked.

3.3 Different stressors

The twentieth-century endocrinologist Hans Selye, while attempting to isolate a new sex hormone in his laboratory, observed that laboratory mice exhibited a non-specific reaction to various harmful stimuli—such as extreme temperatures or chemical intoxication—showing, for example, enlargement of the adrenal glands (Selye, 1936; Selye, 1976). These findings laid the foundation for a deeper understanding of the nature and consequences of stress and opened the way for a vast and expanding field of research (Perdrizet, 1997; Rochette et al., 2023). Yet, contemporary literature provides ample evidence that each stressor elicits specific physiological and behavioral responses, and

that individual characteristics determine the unique pattern, duration, and intensity of the reaction (Goldstein & Kopin, 2007; Lu et al., 2021; Pacak et al.,1998).

Murray (1996) and his colleagues proposed a simple classification of stressors into four distinct categories.

Zero-order stressors exert a direct physical influence on the vocal production system. A classic example is *vibration*, which directly affects the vocal tract. Consider a helicopter pilot engaged in a search and rescue mission: during high-intensity operations—such as mountain rescues or flights in adverse weather—helicopters are exposed to strong vibrations and turbulence. These not only make flight physically demanding but also directly impact the voice production mechanism. When the pilot communicates via radio with air traffic control or the rescue team, the vibrations can cause involuntary fluctuations in voice, reducing its clarity and intelligibility.

First-order stressors induce physiological changes, such as chemical effects or fatigue. Imagine a pilot conducting a long-haul international flight. During such missions, the pilot is exposed to several first-order stressors, among which fatigue is one of the most significant. Prolonged duty hours can lead to both physiological and cognitive alterations—reduced concentration, slower reflexes, and impaired decision-making. Moreover, the pressurized cabin environment and low humidity may have chemical and physical effects on the body, such as dehydration and decreased blood oxygen levels. These combined factors increase the likelihood of errors or suboptimal decisions, particularly during critical phases such as take-off and landing. Managing these first-order stressors effectively is crucial to maintaining safety and performance on long-duration flights.

Second-order stressors lead to changes that arise from the perception and interpretation of the stressor, prompting conscious adjustments in speech production. These are primarily *perceptual* in nature. Consider, for instance, an ATCO who, disturbed by background noise during communication with a pilot, deliberately modifies his speech pattern—speaking more clearly, slowly, or with greater emphasis—to ensure his instructions are correctly understood. Such adaptations reflect the cognitive appraisal of the stressor and the need to adjust communicative behavior to maintain safety and efficiency.

Third-order stressors act at the highest levels of the speech production and cognitive-emotional regulation systems, and are therefore termed *psychological*. They may be external or internal in origin. Examples include workload, emotional distress, or anxiety. *Situational stress* clearly belongs to this third-order category in Murray's classification. These stressors involve the mental processing of events, the perception of control, and the individual's sense of competence and emotional state. Situational stress typically arises in unexpected circumstances that undermine control or social stability—such as emergencies, errors, or interpersonal conflict—and often generates feelings of helplessness and lack of support.

In essence, situational stress represents a third-order stressor, as it operates primarily at the psychological level. It influences both voice and communication through cognitive appraisal processes, emotional responses, and the individual's perceived control over the situation.

It is important to underscore that stressors rarely align with a single level (Finch & Stedmon, 1998; Keränen et al., 2004). They frequently span levels or unfold over time as cascades. For example, although situational stress is primarily third-order, its manifestations may include first order physiological responses and second order perceptual adjustments, depending on timing, context, and individual differences.

3.4 Different stages in the stress process

Selye (1951) described three stages, collectively referred to as the general adaptation syndrome (GAS), to illustrate the nonspecific changes that occur in an organism exposed to a stressor. These stages are:

Alarm Stage: This is the moment when stress is first perceived by the individual, triggering the activation of the physiological mechanisms previously described. In this phase, the body becomes alert and prepares to respond to the threat or stressful stimulus.

Resistance or Adaptation Stage: This represents the phase of active response to stress. During this period, the body attempts to restore homeostasis. The individual's reaction depends on the organism's adaptive capacity, the intensity and duration of the stressor, and any pre-existing stress conditions. This stage can last for varying periods: in some cases, homeostasis is successfully regained, while in others, if the stressor persists, the process evolves into the exhaustion stage.

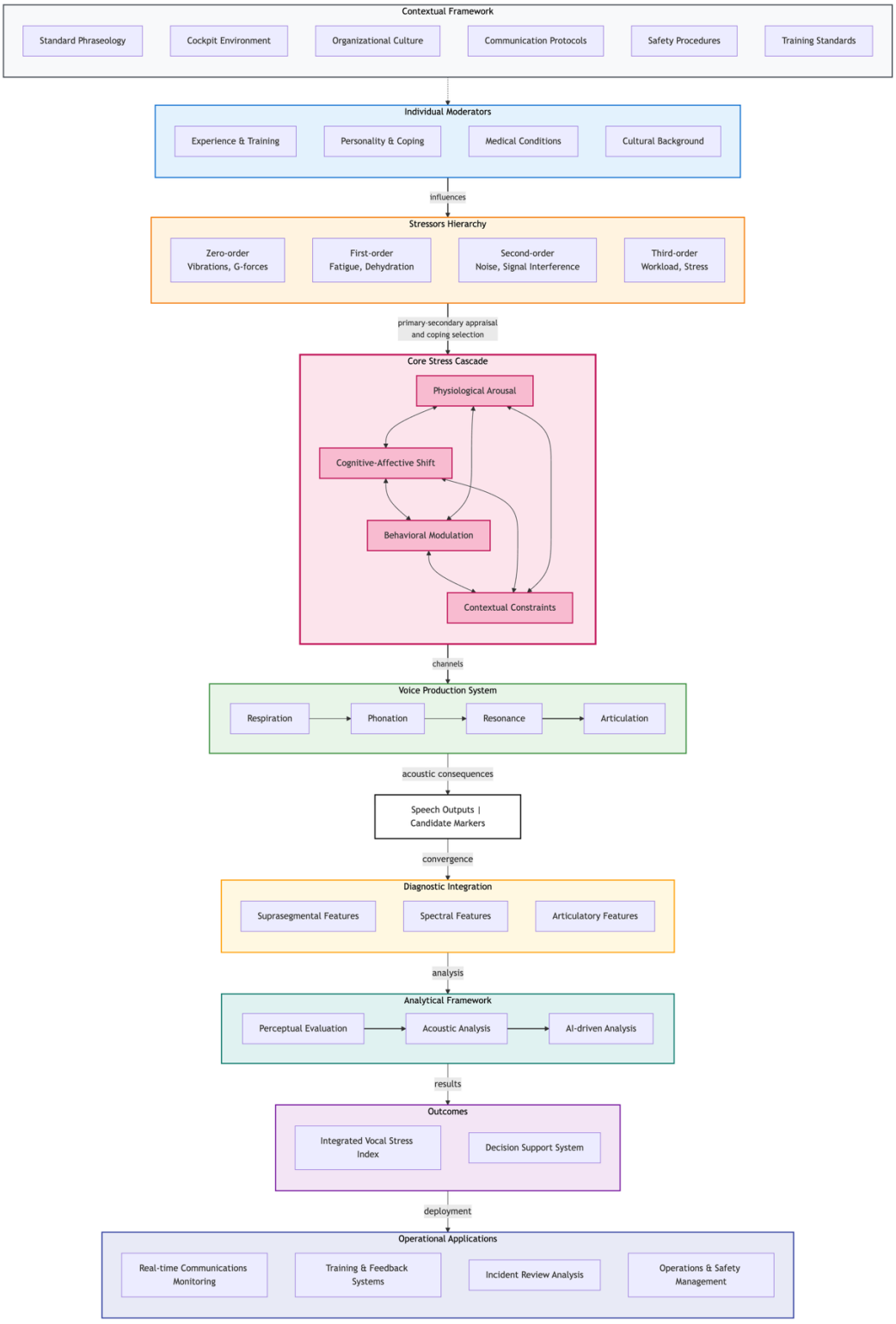
Exhaustion Stage: This occurs when exposure to stress becomes chronic or prolonged. Once this stage is reached, the stress response becomes ineffective in restoring physiological balance. This stage is particularly harmful, as prolonged stress exposure can significantly increase the risk of physical and psychological disorders. It is interesting to note that each of the three stages of the General Adaptation Syndrome can be associated with specific vocal markers, although such correspondence should not be considered absolute.

A study by Ruiz et al. (1996) analyzed the intonation profiles of a pilot and a co-pilot during the final phases of an aircraft accident, revealing two distinct reactions. The pilot showed a marked increase in F0 when he became aware of the technical failure (alarm stage), with an F0 of 150 Hz compared to 117 Hz under "normal" conditions and 144 Hz immediately before the crash—representing a 23% increase between phases 0 and 2. In contrast, the co-pilot did not exhibit a significant F0 increase during the initial discussion about the failure (153 Hz compared to 142 Hz under baseline stress condition 0), but displayed a sharp rise just before the crash, with an average F0 of 204 Hz, an increase

of 43.6% relative to baseline. In summary, for both speakers, a specific flight phase corresponded to a significant rise in pitch: the alarm stage for the pilot and the exhaustion stage for the co-pilot. We therefore propose that elevated F0 levels are associated with bottom-up sympathetic activation processes, that is, emotional and physiological reactions influencing voice unconsciously through the autonomic nervous system. Conversely, a narrowing of the F0 range or a reduction in its variability may reflect top-down regulation, indicating greater cognitive control exerted by the individual to manage the stressful situation. Thus, different F0 patterns in emergency contexts could be related to varying degrees of top-down regulation, shaped by role, position, personal attitude, and training. In summary, an increase in F0 range may correspond to a decrease in top-down control, reaching particularly high values when cognitive regulation is lost. Conversely, a reduction in F0 range may result from a high cognitive load, reflecting an intensified top-down control process. This interpretation aligns with Lazarus's transactional model, which emphasizes appraisal and coping: stress does not stem from the stressor per se or from the response alone, but from the individual's evaluation of the situation and the strategies mobilized to manage it (Lazarus & Folkman, 1984).

Figure 1

Conceptual model of vocal stress measurement



*Note.*Cultural and individual moderators (for example, those illustrated in the figure but not limited to them) filter the impact of stressors, which span zero order, first order, second order, and third order. Appraisal and coping mediate a cascade of physiological arousal, cognitive and affective shift, and behavioral modulation (Scherer & Moors, 2019) that acts on speech subsystems respiration, phonation, resonance, and articulation. The resulting acoustic patterns are integrated for diagnostic use and analyzed through perceptual evaluation, acoustic analysis, and artificial intelligence methods. Arrows indicate directional influence rather than causation, and pathways may operate in parallel.

CHAPTER 4

Vocal Markers in Aviation: A Systematic Review of Workload, Stress, Fatigue, and Sleepiness[4]

Abstract

Background: Stress, increased workload, sleepiness, and fatigue are prevalent in the aviation industry, impacting the performance and well-being of pilots and air traffic controllers (ATCOs). These psychophysiological states often manifest through alterations in vocal characteristics. This systematic review aimed to synthesize the current evidence on vocal changes associated with these states in pilots and ATCOs.

Methods: Following PRISMA guidelines, a comprehensive search was conducted across electronic databases, including Scopus, ScienceDirect, PsycINFO, and Web of Science. Studies were screened based on predefined inclusion criteria. Twenty studies met the inclusion criteria and were included in the review.

Results: Findings reveal that while each psychophysiological state has unique vocal markers, there is considerable overlap. Stress and workload were associated with increased vocal intensity and pitch, reflecting heightened sympathetic nervous system activation. Conversely, fatigue and sleepiness showed reduced vocal energy, slower speech rates, and increased pauses, indicating lowered central nervous system activity. MFCC parameters proved to be versatile indicators across all states. However, methodological inconsistencies, including measurement and study setting variations, limited comparability.

Conclusion: This review highlights the need for standardized definitions and protocols in vocal monitoring research to enhance the reliability and applicability of findings in aviation settings, enabling the development of real-time monitoring and intervention systems.

***Keywords**:* Vocal markers, acoustic analysis, aviation safety, workload, stress, fatigue, sleepiness

[4] Portions of this work have been published in Gnerre, M., & Biassoni, F. (2025). Stress, Fatigue, and Sleepiness: A Protocol Validation Study. *Safety Management and Human Factors*, 189, 182.

4.1 Introduction

Aviation pilots and ATCOs are two occupational groups that communicate extensively with each other and operate under highly demanding conditions, including frequent schedule changes, extended duty periods, shift work, and adverse environmental factors (Lee & Kim, 2018; Marqueze et al., 2017; Mélan & Cascino, 2022). These challenges can result in stress, fatigue, increased workload, and sleepiness, all of which are well-known contributors to human error along with miscommunication and decreased situational awareness (Alaminos-Torres et al., 2023; Kharoufah et al., 2018) and reduce overall well-being. Even minor details can have significant consequences in aviation, and human error remains the main critical challenge (Masi et al., 2023).

Recognizing and monitoring these states can help prevent dangerous situations; consequently, significant efforts have been made to develop solutions to mitigate them (Hu & Lodewijks, 2020). Consistent with such a purpose, in recent years, there has been growing interest in non-invasive, continuous monitoring methods to assess these states (Luzzani et al., 2024; Hu & Lodewijks, 2020). Among these, the analysis of vocal features has emerged as a promising approach, particularly because aviation often requires exclusive reliance on speech, sometimes in critical and high-stress situations (Li & He, 2024; Kuroda et al., 1976).

The voice production process is particularly sensitive to physiological and psychological changes, including those induced by stress (Patil et al., 2013), workload (Sandoval et al., 2022), sleepiness (Martin et al., 2021), and fatigue (Gao et al., 2022), so that variations in voice may effectively mirror the effects of such conditions. Although numerous studies have investigated the acoustic correlates of stress, workload, sleepiness, and fatigue, the findings remain inconsistent and fragmented (Van Puyvelde et al., 2018). Two main challenges emerge. The first is a theoretical issue: although these phenomena are interrelated, they are distinct, and there is currently no clear consensus on how to define and acoustically characterize them due to their complex, multidimensional nature.

The second concern is studying them within the aviation environment, given the potential implications for safety and performance. The specific challenges of the aviation environment—such as high-pressure situations, noise interference, and the need for constant communication—make it essential to synthesize existing evidence to provide clearer insights into the reliability and applicability of acoustic markers in this context. Therefore, it is impossible to study these phenomena as isolated (Masi et al., 2023).

This review aims to synthesize findings from various studies to identify the most consistent acoustic features associated with stress, workload, sleepiness, and fatigue. By consolidating evidence across contexts, the review seeks to determine which vocal markers reliably indicate each of these states,

highlighting consistently observed vocal features that can serve as effective indicators for monitoring physiological conditions in aviation setting.

4.1.1 A Definition of stress, fatigue, workload, and sleepiness

Stress, fatigue, workload, and sleepiness are interconnected yet distinct phenomena frequently observed in high-stakes fields like aviation, in relation to errors that can have serious consequences (MacDonald, 2003). Each represents a particular physiological and psychological state, uniquely influencing performance and safety (Alaminos-Torres et al., 2023). In aviation, these states are often the result of demanding 24-hour operations, irregular schedules, jet lag, layovers, disruptions to circadian rhythms, in-flight sleep quality, and the aircraft environment (Li & He, 2024). Together, these elements increase risk by diminishing performance and compromising safety. The interaction among these factors is complex; however, the literature on the subject often presents simple causal models that do not fully capture this complexity. Depending on the specific situation, each factor can serve as both cause and effect (Figure 1). For example, fatigue can lead to sleepiness (Lichstein et al., 1997), and the reverse is also true (Whitmore & Fisher, 1996).

This intricate, multidimensional interplay requires a flexible approach to understanding their interactions, rather than a straightforward causal model. However, these states have many similarities. Stress, fatigue, and workload can manifest in mental, physical, or emotional forms with similar symptoms (Appley et al., 2012; Lock et al., 2018; Meijman & Mulder, 2013). All four constructs can also accumulate over time, resulting in either acute (short-term) or chronic (long-term) effects (Apostolopoulos et al., 2010; Baqutayan, 2015; Bendak & Rashid, 2020). Numerous other factors, such as environmental conditions, nutrition, physical health, physical activity, and recovery periods, can further influence these states (Karl et al., 2018; MacDonald, 2003). These four states can be assessed using subjective, objective, and performance-based measures (Masi et al., 2023). Since these states are complex, synergistic responses that vary among individuals and situations, establishing a universally accepted conceptual framework remains challenging.

Workload

Workload, a central concept in human factors psychology, represents the difference between task demands and an individual's perceived capacity to handle them (Parasuraman et al., 2008). Based on cognitive load theory (Sweller, 1988), the human cognitive system functions as a limited-capacity information-processing mechanism. In high-stakes operational settings, such as aircraft cockpits and ATC towers, operators must rapidly process information to make critical decisions and uphold safety standards (Boyer et al., 2018). However, the resources available to human operators may not always be

sufficient to meet these demands effectively. For example, task complexity and time pressure combination lead to a significant mental load.

The workload is influenced by two main factors: exogenous demands, such as task difficulty, priority, and situational context, and endogenous resources, including attention and cognitive abilities required for perception, memory updating, planning, decision-making, and response execution (Vidulich & Tsang, 2012). There are various workloads; specifically, we refer to cognitive workload, which involves the mental resources needed to accomplish complex tasks. This type of workload becomes especially critical during high-demand phases of flight, such as takeoff and landing (Martins, 2016). In aviation, workload is influenced by factors like the complexity of processing information and limited time constraints (Wickens et al.,2023). Individual differences, such as skill level and expertise, also affect the available resource supply (Raby & Wickens, 1994). Research shows that greater skill and experience can reduce resource demands, allowing more capacity for executing additional tasks (Vidulich & Tsang, 2012). Practicing complex tasks can enhance performance while lowering brain activation, illustrating the efficiency that expertise provides. However, as workload increases, stress and fatigue are more likely, potentially impacting employees' well-being.

Stress

Stress represents a complex, multifaceted response involving neuroendocrine, autonomic, behavioral, psychological, emotional, and cognitive processes (Ghasemi et al., 2024; Selye, 1951). It is activated to aid adaptive coping when individuals face stimuli, termed stressors, that are perceived as demanding or challenging (Lazarus & Folkman, 1984). Frequently associated with high workloads, stress arises when there is a narrow margin between demands and coping capacities, leading to anxiety and frustration, especially if one perceives one's performance as insufficient (Baqutayan, 2015). Factors such as repetitive tasks, poor job design, and limited control can intensify stress (Masi et al., 2023). In essence, stress encompasses three core aspects: heightened arousal or excitability, a perception of negativity or adversity, and a sense of unpredictability or lack of control (Fink, 2016). Stress plays a critical role in maintaining internal stability when confronted with stressors, supporting long-term adaptation. These stressors are diverse, originating from internal or external sources and eliciting various responses.

The literature categorizes stressors into types such as physical (e.g., intense activity, sleep deprivation), environmental (e.g., noise, extreme temperatures), emotional (e.g., personal loss, social pressures), mental/task-related (e.g., cognitively demanding tasks), and chronic (e.g., financial difficulties, ongoing health conditions) (Masi et al., 2023). While stress often carries negative connotations, some authors argued that it can also motivate individuals to take action, think creatively, resolve issues, and consider others' perspectives (Kupriyanov & Zhdanov, 2014; Le Fevre et al., 2003). To clarify this dual

nature, Hans Selye introduced the terms “distress” (negative stress that harms health and performance) and “eustress” (positive stress that enhances motivation) (Selye, 1974).

Fatigue

Fatigue is the subjective experience of diminished physical and/or mental energy, leading to decreased motivation, alertness, situational awareness, and reaction time (Bendak & Rashid, 2020; Li & He, 2024). Some researchers define fatigue as a state between vigilance and sleepiness, with recovery only achievable through adequate sleep (Vagner et al., 2018). However, this perspective does not consider cases where individuals experience prolonged fatigue despite sufficient sleep, often due to chronic stress, heavy workloads, or psychological strain (Apostolopoulos et al., 2010; Phillips, 2015). Fatigue is characterized by a range of symptoms, including difficulties in concentration, heightened anxiety, a gradual loss of stamina that is disproportionate to energy expended, sleep disturbances, and increased sensitivity to light, sound, taste, and touch (Komaroff & Buchwald, 1991; Olson, 2007).

Fatigue is also the main symptom of chronic fatigue syndrome (CFS) and is also linked to various acute and chronic conditions, including rheumatoid arthritis, cancer, and multiple sclerosis (Shen et al., 2006). Desmond and Hancock (2000) distinguish between "active fatigue," caused by overload in high-demand situations such as heavy traffic, and "passive fatigue," resulting from underutilization, as seen when driving on open roads. Individuals experiencing fatigue often report that they "pushed through" these challenges. According to Olson (2007), a crucial aspect of fatigue is the additional effort required to manage its symptoms, especially when energy levels are already depleted. In this framework, fatigue typically arises from an inadequate response to initial feelings of tiredness. If an effective adaptive response is applied at this stage, the individual can revert to mere tiredness and eventually fully recover. However, if adaptation fails, it can gradually progress toward exhaustion (Olson, 2007). Fatigue encompasses complex psychological and physiological factors, whereas tiredness is generally a temporary condition that can be relieved by rest or sleep. Fatigue may persist despite short rest periods, often necessitating more comprehensive interventions, such as stress management techniques, lifestyle adjustments, or addressing underlying health issues (Olson, 2007).

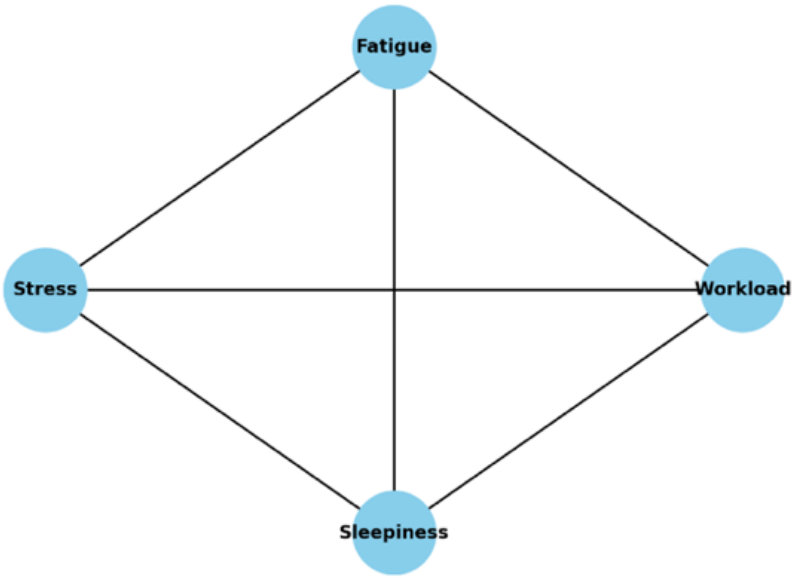
Sleepiness

Sleepiness is commonly understood as the natural tendency or inclination to fall asleep, as defined by Shen et al. (2006). According to dictionary definitions, sleepiness denotes a state in which one feels drowsy, struggles to stay awake, and is inclined to sleep (Onions, 1959). However, the term "sleepiness" has various interpretations across different fields. Kleitman (1963) describes it as a state of languor or inertia, while Dinges and Broughton (1989) view it as a subjective experience of needing sleep. In contrast, Aldrich (2000) characterizes it as a physiological drive resulting from sleep deprivation,

and Carskadon and Dement (1979) define it as a pronounced tendency to fall asleep. These perspectives underscore the complexity of sleepiness, which encompasses both subjective sensations and objective physiological states (Curcio et al., 2021). Historically, sleepiness was largely synonymous with drowsiness, but over recent decades, especially within sleep medicine, it has come to include more specific concepts such as "sleep propensity"—the ease or likelihood of falling asleep—and "sleep drive," a physiological urge that intensifies with sleep deprivation (Johns, 2010). This universal phenomenon is not only a symptom of various medical, psychiatric, and primary sleep disorders but also a routine physiological experience within any 24 hours. Sleepiness becomes pathological when it is persistently excessive, as in narcolepsy, or entirely absent, as in insomnia (Shen et al. 2006). It may also be considered abnormal when it occurs at inappropriate times or fails to arise when it would be expected, thus highlighting the nuanced nature of sleepiness in both health and disorder. This study refers not to "pathological sleepiness" but to "immediate sleepiness," a temporary and subjective condition observed in healthy individuals.

Figure 2

Conceptual framework depicting the interrelationships among fatigue, stress, workload, and sleepiness, highlighting their mutual influence without a single, straightforward causal direction



4.1.2 Voice analysis as assessment method for workload, stress, fatigue, and sleepiness

Assessment methods for workload, stress, fatigue, and sleepiness vary in invasiveness and in the level of interference with the subject's environment (Hagmüller et al., 2006). These methods include performance-based measures, subjective approaches and objective measures (Boyer et al., 2018; Hagmüller et al., 2006). Performance-based methods have certain limitations; for example, continuous monitoring may disrupt the operator's focus, and establishing objective performance criteria can be particularly challenging for complex tasks (Salas et al., 2017).

Subjective methods, primarily self-report questionnaires such as the NASA Task Load Index (Hart & Staveland, 1988), have the limitation of often being collected retrospectively rather than in real-time. On the other hand, objective methods, which rely on physiological or behavioral data (e.g., blood tests or brain electrode monitoring), are often invasive. Among non-invasive methods, we find contact-based options, such as wristbands for measuring heart rate, and contact-free techniques, like cameras (Hagmüller et al., 2006). Contact-free methods are preferred in real-world applications as they do not hinder the subject's tasks or induce additional stress. Within contact-free options, some are perceived as more intrusive (e.g., visible cameras), while others, such as voice analysis, are non-intrusive and ideal for tasks where voice is already used, such as by pilots and ATCOs. Thus, voice analysis is one of the most effective ways to detect workload, stress, fatigue, and sleepiness in this context (Boyer et al., 2018; Van Puyvelde et al., 2018; Rothkrantz et al., 2004). It can be easily captured in real-life settings using affordable, user-friendly equipment that provides real-time data (Van Puyvelde et al., 2018).

4.1.3 Effect of workload, stress, fatigue, and sleepiness on voice

Workload, stress, fatigue, and sleepiness can significantly impact our physical and mental states and influence how we produce language (Boyer et al., 2018; Hansen et al., 2000; Hagmüller et al., 2006; Van Puyvelde et al., 2018; Rothkrantz et al., 2004). Hansen et al. (2000) introduced an insightful model for understanding how psychological factors affect speech production. The model illustrates that, after ideation, the process involves selecting and organizing words, as well as activating motor programs and neuromuscular commands to produce sound. Although described as a sequence, these stages overlap, with feedback loops that enable adjustments. These factors can affect some or all specific stages of this process, ultimately altering speech production.

The production and processing of speech depend on the coordinated effort of around 100 muscles, activated by an intricate network of cranial and spinal nerves along with both subcortical and cortical regions of the brain and involve cardiorespiratory functions (Duffy, 2000). Consequently, speech quality commonly declines when performance is compromised or emotional regulation is disrupted (Hansen et al., 2000). This makes speech a psychophysiological process responsive to external

and internal stressors (Hansen and Patil, 2007). The resulting states, which in their turn induce cognitive and physiological changes—such as reduced muscle tension and alterations in body temperature—can indirectly affect different stages of speech production (Van Puyvelde et al., 2018). Researchers have focused on monitoring these states through speech signals, recognizing their potential to enhance performance and accuracy in human decision-making (Rothkrantz et al., 2004). Numerous studies across various contexts have investigated these phenomena to identify their respective acoustic markers, particularly in laboratory settings (Giddens et al., 2013).

Effects of workload on speech production

In relation to the cognitive workload, acoustic correlates refer to specific vocal features that change in response to mental or physical demands. Numerous studies have explored vocal production in laboratory settings, where participants performed tasks, such as reading a standardized passage displayed on a computer screen or completing Stroop tasks of varying difficulty levels (Meier et al., 2016; Yin et al., 2008; Rothkrantz et al., 2004; Van Segbroeck et al., 2014). Meier et al. (2016) identified changes in vocal tract characteristics and voice source parameters in participants' speech as primary indicators of cognitive workload during Stroop tasks of varying difficulty levels (low, medium, and high).

Voice source attributes, including F0, harmonics-to-noise ratio (HNR), cepstral peak prominence (CPP), and various glottal flow metrics, also showed variation with cognitive workload, although they were less effective for classification compared to vocal tract features (Meier et al., 2016). Speakers often demonstrate changes under increased workload, such as heightened F0 and F0 variations (Lively et al., 1993; Mendoza & Carballo, 1998; Rothkrantz et al., 2004; Scherer et al., 2002) elevated jitter and shimmer (Mendoza & Carballo, 1998), enhanced high-frequency harmonic energy and reduced spectral noise (Mendoza & Carballo, 1998; Van Segbroeck et al., 2014). Moreover, as cognitive workload increases, the duration of silence intervals typically shortens and becomes more variable (Van Segbroeck et al., 2014). Some speakers also exhibit increased amplitude and greater variability in amplitude (Pisoni et al., 1990; Van Segbroeck et al., 2014; Lively et al., 1993), faster-speaking rates, reduced spectral tilt, and decreased f_0 variability under workload conditions, while F1, F2, and F3 remain stable (Lively et al., 1993). These vocal adjustments suggest both laryngeal and sublaryngeal modifications that influence the timing of articulatory movements. Cepstral and Mel frequency cepstral coefficients (MFCCs) are also very useful in workload detection (Boril et al., 2011; Yap et al., 2015; Yin et al., 2008).

Additionally, a perceptual experiment confirmed that speakers who exhibited significant changes in speech production under workload also showed slight improvements in intelligibility, largely attributed to increased amplitude and variability (Lively et al., 1993). Finally, it appears that as a

controller's workload decreases, the average duration of their communications tends to increase (Uclés & García, 2014).

Effects of stress on speech production

The acoustic markers of stress have been extensively studied, particularly in the context of developing noninvasive methods for identifying potential deception in statements for law enforcement and forensic purposes (Hollien et al., 1987; Horvath, 1982; Patil et al., 2013). A systematic review by Giddens et al. (2013) highlighted that an increase in F0 is the most commonly observed effect of stress in controlled studies, likely resulting from heightened cricothyroid muscle tension and increased subglottal pressure. Although this effect is not universal (Hecker et al., 1968), most research suggests that f_0 related parameters are among the most distinctive markers of stress (Demenko & Jastrzębska, 2012; Hagmüller et al., 2006). Maximum F0 and F0 standard deviation also emerge as significant indicators, with f_0 standard deviation generally elevated under stress (Protopapas & Lieberman, 1997; Scherer et al., 1981; Scherer et al., 2002; Sondhi et al., 2015; Streeter et al., 1983), though one study found otherwise (Van den Broek, 2003). Additionally, formant frequencies F1 and F2 have been shown to increase under stress for most subjects (Protopapas & Lieberman, 1997; Sondhi et al., 2015). Increased loudness is also noted as a stress indicator (Hollien et al., 1980; Scherer et al., 2002; Streeter et al., 1983; Zhang et al., 2015). Furthermore, stress is associated with reduced vocal jitter and shimmer, indicating decreased vocal noise (Rothkrantz et al., 2004; Sondhi et al., 2015). Microtremors in the voice (e.g., measured using the Amplitude Tremor Intensity Index and Frequency Tremor Intensity Index) can also serve as indicators of stress, with significant reductions in tremor amplitude observed under conditions of high stress (Mendoza & Carballo, 1999; Sondhi et al., 2015). Moreover, highly stressed subjects typically exhibit a decrease in mean utterance duration and an increase in speech rate under high cognitive load (Scherer et al., 2002; Streeter et al., 1983). Stress-induced respiratory changes, such as increased or irregular breathing rates, contribute to shorter speech intervals, misplaced breaths, and altered speech timing and rate, often accompanied by inappropriate pauses and changes in articulation rate (Baker et al., 2008; Hansen & Patil, 2007; Pisanski & Sorokowski, 2021). Glottal waveform parameters (Godin & Hansen, 2008) and cepstral parameters are also valid stress indicators (Li et al., 2007).

Effects of fatigue on speech production

Our conceptualization of fatigue differs from that of Nanjundeswaran et al. (2015), who focus on localised throat tiredness and vocal weakness following prolonged vocal use. Fatigue, as defined earlier, affects the timing of sound articulation (Vollrath, 1994) and the intervals between sounds within

words (Krüger & Vollrath, 1996), leading to a decrease in F0 (Baykaner et al., 2015; Roelen & Stuut, 2016; Saito et al., 1980). Greeley et al. (2006) identified circadian patterns in MFCCs linked to sleep deprivation induced fatigue. In their study participants who recited a word list at six circadian-aligned intervals, they observed strong negative correlations between EEG-measured sleep onset latency and MFCCs, especially for the phonemes "p" (r = -0.89) and "t" (r = -0.67). Furthermore, the same participants exhibited changes in formant frequencies due to fatigue. Their findings demonstrated that quantifiable changes in MFCCs corresponded with both direct fatigue measures and time awake, with significant shifts in speech sounds requiring higher average airflow. Conversely, Gao et al. (2022) and Diepeveen et al. (2021) found no significant differences between baseline and fatigue states, noting only a non-significant decrease in energy, loudness, and f_0. In a study by Cho et al. (2011), analyses of acoustic features between high- and low-fatigue groups showed that, specifically in men, parameters related to vocal instability—such as shimmer and amplitude tremor—were lower in the high-fatigue group compared to the low-fatigue group. Additionally, noise-related measures, including HNR and signal-to-Noise Ratio (SNR), were higher in the high-fatigue group. In contrast, no significant acoustic markers of physical or mental fatigue were identified in women.

Effects of sleepiness on speech production

Sleepiness detection in speech, a relatively underexplored area, was the focus of three international challenges at the 2007, 2011, and 2019 Interspeech conferences (Schuller et al., 20211; Schuller et al., 2019). In general, research has shown that sleepiness is related to a reduction in f_0, intensity, articulatory precision, and articulation rate (Krajewski & Kröger, 2007; Schuller et al., 20211; Schuller et al., 2019). One study examined three types of sleepiness in a sample of individuals diagnosed with hypersomnia (Martins et al., 2016). "Average sleepiness," measured as the overall sleepiness level across the day, was found to correlate with an increase in the bandwidth of the fourth formant and a decrease in energy slope. This differs from subjective sleepiness perception, which was associated with an increase in the bandwidth of the second formant and a decrease in the third.

Additionally, the energy slope and fourth formant bandwidth coefficients were reversed compared to average sleepiness. Physiological sleepiness refers to sleepiness identified through reduced sleep latency. In individuals with hypersomnia, physiological sleepiness had a notable impact on reading quality rather than acoustic quality, resulting in reading pauses distributed in unnatural locations. Thus, findings indicate that while subjective sleepiness significantly affects acoustic properties of speech, physiological sleepiness primarily influences the structure of reading pauses: the greater the physiological sleepiness (i.e., lower sleep latency), the more frequent the placement of pauses in atypical positions. In another study, long-term sleepiness correlates with the HNR, often resulting in a less clear, noisier vocal quality (Martin et al., 2024). In this study, increased sleepiness can reduce the

duration and ratio of vocalic to non-vocalic parts in speech, indicating a slower pace and less precise articulation. The bandwidth of the first formant also widens with sleepiness, suggesting reduced articulatory precision and altered orofacial movement. Additionally, formant amplitude may decrease, reflecting lower vocal energy and intensity. Although not an acoustic feature per se, automatic speech recognition errors, such as insertions and substitutions, correlate with acoustic variations induced by sleepiness. Moreover, a lower energy slope is associated with subjective perceptions of sleepiness throughout the day (Martin et al., 2024).

4.1.4 The crucial role of the physical environment and behavioral environment

The physical environment (understood here as the set of physical and material characteristics of a work setting) and behavioral environment (understood as the broader circumstances and psychological demands surrounding a task) (Hafeez et al., 2019) are crucial elements to consider when dealing with acoustic data for at least two main reasons. The first reason is technical and relates to the physical environment: the operational environment where the voice is produced and recorded inevitably impacts the acoustic data (Deliyski et al., 2005). Both the aircraft and the control tower or radar facilities can influence the sound characteristics captured in voice recordings. The cockpit is characterized by high levels of background noise from the engine, ventilation, and various electronic systems, which can interfere with voice recording quality by masking or distorting certain sounds, making it challenging to separate voice from background noise (Lindgren et al., 2006). Additionally, constant vibrations and pressure variations can further affect acoustic quality, adding extra challenges for accurate voice data analysis. However, the workspace of ATCOs is a relatively quieter environment, but ongoing conversations, noise from communication tools, and interference from monitoring equipment can affect the clarity of voice recordings. The aviation behavioral environment introduces unique stressors inherent to these settings, which may be short-term or chronic (Masi et al., 2023). Stressors may arise from various sources, including environmental factors such as cockpit air dryness, dust, noise, and insufficient lighting (Lindgren et al., 2006); task-related challenges, such as managing high-density air traffic networks for ATCOs (Majumdar & Ochieng, 2002) or handling takeoffs and landings for pilots (Roscoe, 1978); interpersonal dynamics, such as communication difficulties between pilots and ATCOs (Wu et al., 2019); and psychological factors, such as depression or anxiety (DeHoff & Cusick, 2018). Given that different stressors uniquely affect vocal characteristics (Hagmüller et al., 2006; Van Puyvelde et al., 2018), analyzing how these factors alter voice properties is crucial.

4.2 Method

The searches, extraction, and reporting of data in the present review were conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher et al., 2009). The study protocol was registered with Open Science Framework (OSF); the registration code is **10.17605/OSF.IO/RW67S**.

4.2.1 Hypothesis

Based on the previously presented literature, two hypotheses are proposed:

H1: Workload, stress, fatigue, and sleepiness are interrelated yet distinct constructs with specific acoustic correlates. In essence, it is hypothesized that there are some similarities and some differences mirrored in the acoustic structure.

H2: Environmental factors, such as the experimental setting, military or civilian context, and the nature of the antecedent causing the phenomenon, significantly influence how workload, stress, fatigue, and sleepiness manifest in voice.

4.2.2 Objectives

The first aim of this systematic review is to examine the existing evidence regarding the effect of workload, stress, fatigue, and sleepiness on speech to find specific vocal markers of these phenomena. Finally, this review examines methodological variation (considering the software and parameters used, as well as the experimental and environmental conditions) in the collected studies.

4.2.3 Search strategy

Relevant literature was reviewed following systematic searches of library holdings and electronic databases, including Scopus, PsycINFO, ScienceDirect, and Web of Science. We conducted a search validation procedure using 4 key articles as benchmarks. These articles were chosen because they are highly relevant to our research question and should appear in the search results if the strategy is effective. After an initial search, we reviewed the results to ensure these benchmark articles were included. We adjusted search terms and filters as needed until the strategy consistently captured all 4 benchmark articles, confirming the comprehensiveness and accuracy of our search approach. The following keywords were used: ("Stress" OR "Workload" OR "Fatigue" OR "Sleepiness") AND ("Aircraft pilots" OR "Airline pilots" OR "Commercial pilots" OR "Military pilots" OR "Pilots" OR "Aircrew" OR "Flight crew" OR "Air traffic controllers" OR "ATC") AND ("Voice analysis" OR "Vocal analysis" OR "Speech analysis" OR "Acoustic analysis"). Database searching took place during October and November 2024. A cross-referencing search was also conducted using the references in the articles

identified. We employed the Ascendancy Approach in our search strategy by reviewing the reference lists of included studies. This allowed us to identify additional relevant sources cited within these articles, ensuring a more comprehensive coverage of key literature related to our research question.

4.2.4 Inclusion and exclusion criteria

This systematic review included studies written in English, available in full text, and published in peer-reviewed journals, with no restrictions on the publication date. The inclusion criteria followed the PICO (Population, Intervention, Comparison, Outcome) framework (Table 1). Eligible studies focused on aviation pilots and ATCOs, both civil and military, operating in high-stress environments, including real flights, simulators, and laboratory settings. Inclusion criteria required studies to use vocal monitoring to identify correlates of workload, stress, sleepiness, and fatigue. Studies had to include comparisons across varying operational conditions or examine voice alterations in high-stress versus lower-stress contexts. Book chapters, sections of books, letters, editorials, gray literature, reports, conference proceedings, and unpublished studies were excluded. Additionally, studies lacking clear descriptions of acoustic and statistical measures were not considered.

Table 1

Pico Framework

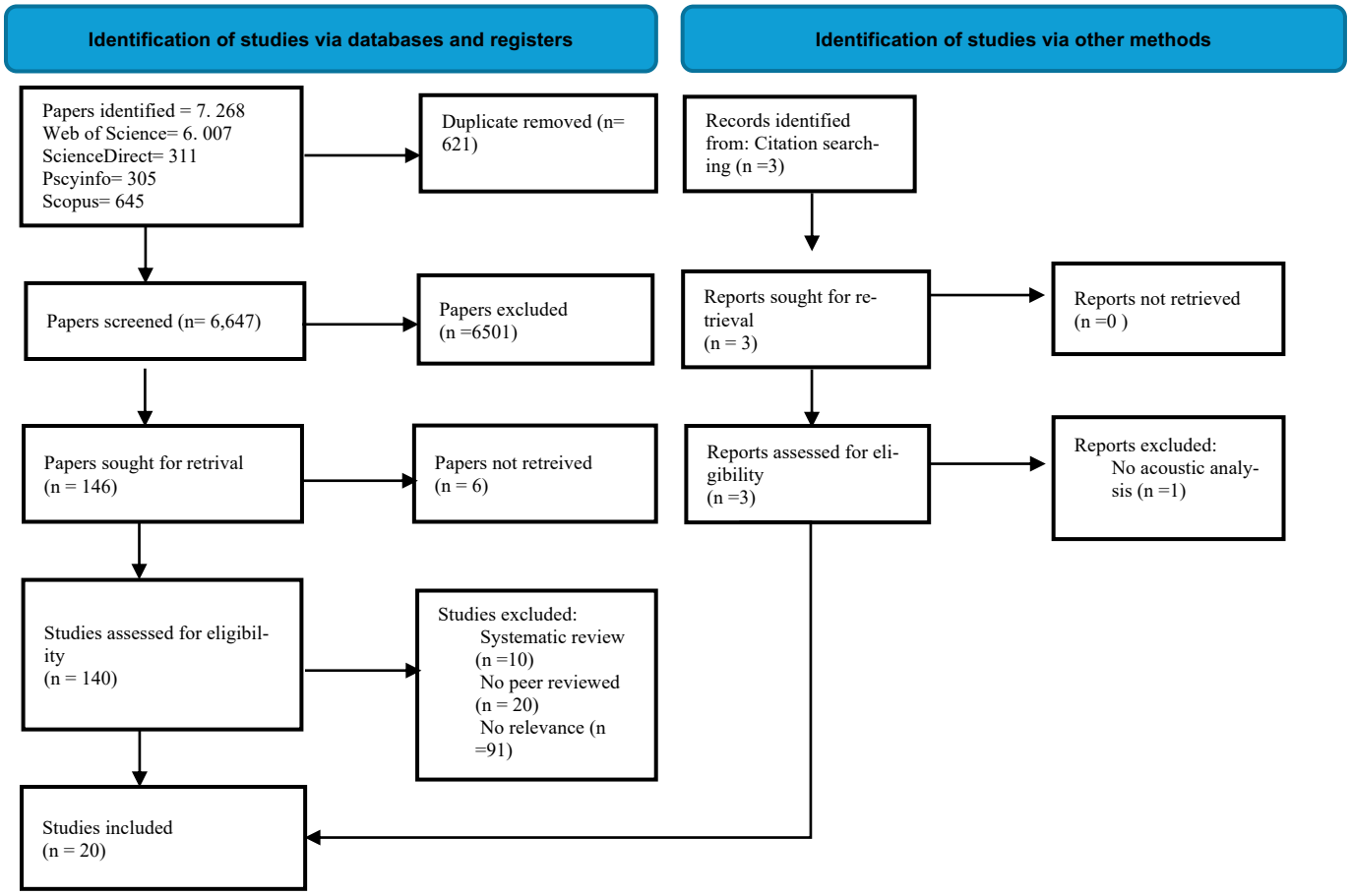
Pico Element	Description
P (Population)	Aviation pilots and ATCOs in civil or military contexts.
I (Intervention)	Extraction of vocal acoustic features to identify indicators of workload, stress, sleepiness, and fatigue in real-flight, laboratory, or simulator settings.
C (Comparison)	Comparison across different levels of conditions examining vocal variations in situations of high workload, stress, fatigue, or sleepiness compared to control conditions where these factors were reduced or absent.
O (Outcome)	Identification of consistent acoustic features reflecting workload, stress, workload, sleepiness, and fatigue.

4.2.5 Data screening and extraction

One author conducted the initial search. Then, two authors independently conducted the screening process, starting with abstracts and keywords. Subsequently, full-text reviews were performed for the articles selected during the initial screening. Consensus on article inclusion was reached by if necessary. 7.268 articles were in all the databases; 621 were duplicates, resulting in a total of 6647 unique articles. After reading the titles and abstracts, we found that 6501 articles did not meet the eligibility criteria and were excluded. Using the same criteria, the full text of the remaining 146 articles was assessed for eligibility, with 6 articles that were not retrieved. We contacted the authors to request the paper but did not receive a response within one month of sending the request. We obtained 18 articles that met the inclusion criteria chosen for analysis in the present work. A reverse search was also carried out in which 2 valid articles were identified and included from the lists of references of the articles that resulted from the original search. Finally, 20 articles were included in the review (see Figure 3). The extraction process involved systematically evaluating and categorizing studies according to predefined eligibility criteria, including study design and characteristics of the study, environmental conditions, acoustic parameters, software used, and phenomena assessed (e.g., stress, workload, fatigue). Studies were manually tabulated based on these characteristics to facilitate comparison across studies. To prepare the data for synthesis, any missing summaries were managed by consulting supplementary data when available.

Figure 3

PRISMA flow diagram for study selection



4.2.6 Quality assessment

The quality of all included articles was evaluated by two researchers independently using the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies (National Heart, Lung, and Blood Institute, 2019) (Table 2). Differences between research assistants were resolved through discussions until a mutual agreement was reached. Additionally, a qualitative approach was employed, ensuring that included studies adhered to accurate definitions of workload, stress, fatigue, and sleepiness as proposed in the introduction of this study. No studies were excluded based on methodological quality assessment.

Table 2

The Quality Assessment Rating

References	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Quality Rating
Alpert (1988)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Cañas (2022)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Congleton (1997)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Ćosić (2019)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
De Vasconcelos (2019)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	No	Yes	NR	NR	Yes	Fair
Huang (2024)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	NR	Fair
Huttunen (a) (2011)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Huttunen (b) (2011)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Kouba (2023)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Khan (2015)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Krajewski (2014)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Luig (2014)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Magnusdottir (2022)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Maina (2023)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Ruiz (2010)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Shao (2021)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Shen (2021)	Yes	Yes	NR	Yes	NR	Yes	Yes	NA	Yes	Yes	Yes	NR	NR	Yes	Good
Whitmore (1996)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Xu (2024)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good
Yang (2023)	Yes	Yes	NR	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	NR	NR	Yes	Good

Note. NA: not applicable; NR: not reported.

4.2.7 Data synthesis

The results of studies were tabulated, displaying key parameters (e.g., software used, acoustic features measured) to allow for clear visual comparison and facilitate identification of trends across studies. Due to the heterogeneity in study designs, methodologies, and acoustic parameters, data synthesis was conducted using a narrative approach.

4.3 Results

4.3.1 General characteristics of selected studies

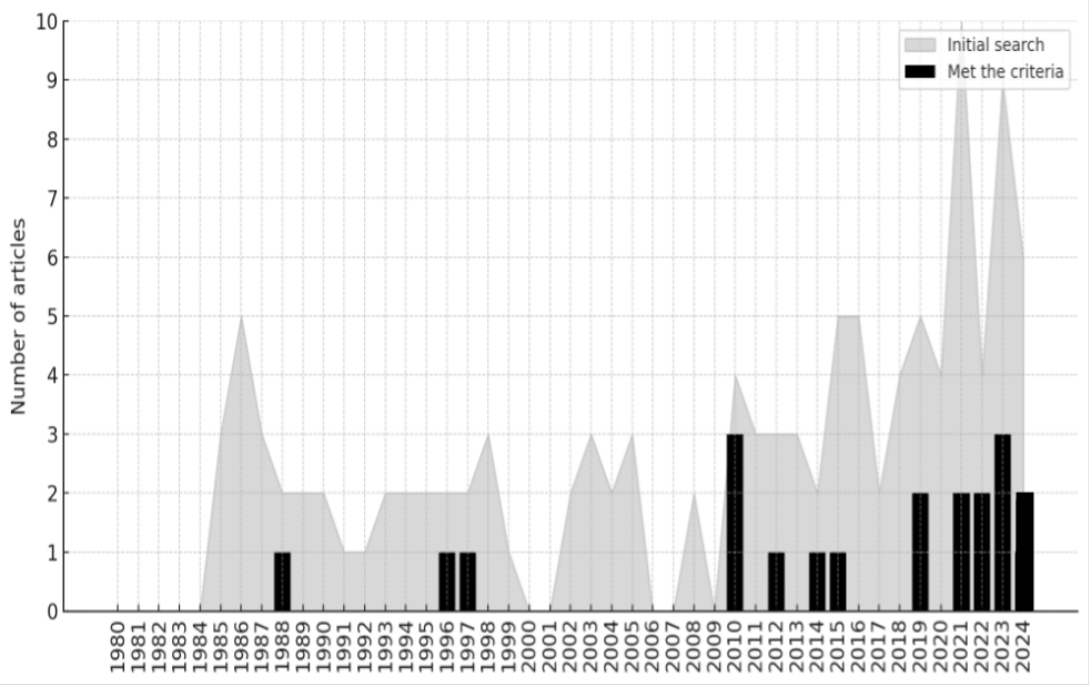
Most studies were published between 2020 and 2024 (n = 9) (Figure 4). China leads in the volume of research on these topics, with 6 studies, followed by the USA with 3 studies (Table 3). The research is conducted slightly more in civil contexts (14 studies) compared to military contexts (6 studies), with samples focusing on pilots (11 studies) and ATCOs (9 studies). The sample sizes are generally below 20 participants, except in three studies with larger samples: 40 participants in the studies by Ćosić et al. (2019) and Huang et al. (2024), and 57 participants in the study by Krajewski et al. (2014). The average sample size across all studies is approximately 16 participants.

Some studies included only male participants (5 studies), while another six did not specify the participants' gender. One study included males and females but did not specify the respective numbers (Huang et al., 2024). Among the studies that included female participants (8 studies), the proportion of females was generally lower than that of males, with two exceptions: one study (Krajewski et al., 2014) had a balanced gender representation, and another (Alpert & Schneider, 1988) included only two participants, both of whom were female.

The studies involved participants speaking various languages, with English being the most used. Some studies did not specify the language used (Cañas et al., 2022; Kouba et al., 2023; Khan et al., 2015; Shao et al., 2021; Shen et al., 2021; and Yang et al.,2023). Most studies used a within-subjects design, indicating participants experienced all conditions of the study (Alpert & Schneider, 1988; Cañas et al., 2022; Congleton et al., 1997; Ćosić et al., 2019; Huttunen et al., 2011a; Huttunen et al., 2011b; Kouba et al., 2023; Krajewski et al., 2014; Luig & Santocchi, 2014; Magnusdottir et al., 2022; Maina & Zhang, 2023; Ruiz et al., 2010; Shao et al., 2021; Whitmore & Fisher, 1996; Xu et al., 2024; Yang et al., 2023). There was also a mixed-design study (Khan et al., 2015) and a case study focused on real-world data (De Vasconcelos et al., 2019). The research was conducted in different settings, including simulators (Congleton et al., 1997; Huang et al., 2024; Huttunen et al., 2011a, 2010b; Kouba et al., 2023; Krajewski et al., 2014; Luig & Santocchi, 2014; Shao et al., 2021; Whitmore & Fisher, 1996; Xu et al., 2024; Yang et al., 2023), laboratories (Alpert & Schneider, 1988; Ćosić et al., 2019; Khan et al., 2015; Magnusdottir et al., 2022; Ruiz et al., 2010) and ecological (real-world) environments (Cañas, 2022; De Vasconcelos et al., 2019; Maina & Zhang, 2023; Shen et al., 2021).

Figure 4

Trends in the number of articles over time



Note. Trends in the number of articles over time: The shaded area represents the initial search results (including only peer-reviewed articles in scientific journals), while the black bars indicate the articles that met the criteria of this systematic review.

Table 3

General Characteristics of the Studies

Study (first author)	Year	Country	Context	Sample	N.	Sex	Language	Study design	Setting
Alpert	1988	USA	Military	Pilots	2	F	English	Within-subjects design	Laboratory
Cañas	2022	Spain	Civil	ATCOs	3	NS	NS	Within-subjects design	Ecological (real-world data)
Congleton	1997	USA	Military	Pilots	16	13 M; 3 F	English	Within-subjects design	AWACS simulator

Ćosić	2019	Croatia	Civil	ATCOs	40	35 M; 5 F	Croatian	Within-subjects design	Laboratory
De Vasconcelos	2019	Brazil	Civil	Pilots	1	NS	Brazilian Portuguese	Case-study	Ecological (real-world data)
Huang	2024	Cina	Civil	ATCOs	40	NS	NS	Within-subjects design	Simulator
Huttunen (a)	2010	Finland	Military	Pilots	13	M	Finnish	Within-subjects design	Flight simulator, F/A-18 Hornet
Huttunen (b)	2010	Finland	Military	Pilots	13	M	Finnish/English	Within-subjects design	Flight simulator, F/A-18 Hornet
Kouba	2023	Czech Republic	Civil	ATCOs	10	M	NS	Within-subjects design	Simulator IATCC
Khan	2015	India	Military	Pilots	18	M	NS	Mixed design	Laboratory
Krajewski	2014	German	Civil	ATCOs	57	28 M; 29 F	German/English	Within-subject design	Simulated ATC communications
Luig	2014	Austria	Civil	Pilots	8	M	English and German	Within-subject design	Flight simulator, Fokker F70/100
Magnusdottir	2022	Iceland	Civil	Pilots	20	18 M; 2F	Icelandic/English	Within-subject design	Laboratory
Maina	2023	China	Civil	Pilots	18	16 M; 2 F	Portuguese	Within-subject design	Ecological (real-world data)
Ruiz	2010	France	Civil	Pilots	3	NS	French	Within-subject design	Laboratory
Shao	2021	China	Civil	ATCOs	8	NS	NS	Within-subject design	Simulator

Shen	2021	China	Civil	ATCOs	NS	NS	NS	Algorithm development and evaluation study	Ecological (real-world data)
Whitmore	1996	USA	Military	Pilots	12	M	English	Within-subject design	Flight simulator B-1B
Xu	2024	China	Civil	ATCOs	14	7 M; 7 F	NS	Within-subject design	Simulator
Yang	2023	China	Civil	ATCOs	8	NS	NS	Within-subject design	Simulator

4.3.2 Types of investigated phenomena

Six studies investigate workload (Alpert & Schneider, 1988; Huttunen et al., 2011a; Huttunen et al., 2011b; Magnusdottir et al., 2022; Shao et al., 2021; Yang et al., 2023). Four studies explore stress (Congleton et al., 1997; Ćosić et al., 2019; Khan et al., 2015; Luig & Santocchi, 2014). Nine studies examine fatigue (Cañas, 2022; De Vasconcelos et al., 2019; Huang et al., 2024; Kouba et al., 2023; Maina & Zhang, 2023; Ruiz et al., 2010; Shen et al., 2021; Whitmore & Fisher, 1996; Xu et al., 2024). Two studies investigate sleepiness (Krajewski, 2012; De Vasconcelos et al., 2019). One study investigated both fatigue and sleepiness (De Vasconcelos et al., 2019). Of the 6 studies investigating workload, two focused exclusively on ATCOs (Shao et al., 2021; Yang et al., 2023), while the remaining four focused on pilots (Alpert & Schneider, 1988; Huttunen et al., 2011a; Huttunen, 2011b; Magnusdottir et al., 2022). Among the four studies examining stress, only one explored stress in ATCOs (Ćosić et al., 2019), whereas the others focused on pilots (Congleton et al., 1997; Khan et al., 2015; Luig & Santocchi, 2014). Regarding studies addressing fatigue, ATCOs were the focus in 5 works (Cañas, 2022; Huang et al., 2024; Kouba et al., 2023; Shen et al., 2021; Xu et al., 2024), while pilots were investigated in six (De Vasconcelos et al., 2019; Maina & Zhang, 2023; Ruiz et al., 2010; Shen et al., 2021; Whitmore & Fisher, 1996; Xu et al., 2024). Finally, sleepiness was investigated in one study on pilots (De Vasconcelos et al., 2019) and one on ATCOs (Krajewski, 2014). Overall, 11 studies focus on pilot samples, while 9 focus on ATCO samples. We have grouped the antecedents of the investigated phenomena into four main categories:

1. Complex Tasks: This category includes task that require high levels of attention and the capacity to manage multiple pieces of information simultaneously. Studies on cognitive load often involve complex tasks, such as the Stroop test (in laboratory settings) or air traffic management (in simulator or real-world settings).

2. Environmental Conditions: This category includes physical and environmental factors (e.g., low lighting, noise), which can impact physiological well-being and hinder the ability to maintain concentration.

3. Emergency Management: This category involves situations demanding rapid decision-making in emergency contexts. It is often associated with high-stress scenarios like managing critical events and handling high-decision-load situations.

4. Physiological Factors: This category includes circadian rhythms or homeostasis disruptions, which can significantly impact alertness and influence overall performance and cognitive functioning.

All except the second category belong to the behavioral environment, while the second belongs to the physical environment (see paragraph 1.4). Studies investigating workload consistently feature antecedents involving only high cognitive load (Alpert, 1988; Huttunen, et al., 2011a; Huttunen et al., 2011b; Magnusdottir et al., 2022; Shao et al., 2021; Yang et al., 2023). In studies examining stress, at least two antecedents belong to distinct categories (Congleton et al., 1997; Ćosić et al., 2019; Khan et al., 2015; Luig & Santocchi, 2014). Concerning fatigue, physiological factors (particularly sleepiness) are almost always present as antecedents (except in the study by Cañas et al., 2022), typically accompanied by other elements such as complex tasks. Similarly, the studies on sleepiness include antecedents spanning three out of the four categories in both studies (Krajewski, 2012; De Vasconcelos et al., 2019).

Most studies did not specify the duration of exposure to antecedents (Cañas et al., 2012; Congleton et al., 1997; De Vasconcelos et al., 2019; Huttunen et al., 2011b). On the other hand, some studies used antecedents with a duration of less than one hour (Alpert & Schneider, 1988; Ćosić et al., 2019; Huttunen et al., 2011a; Kouba et al., 2023; Main et al., 2023). Some studies, on the other hand, involved antecedents with a longer exposure duration (Khan et al., 2015; Krajewski et al., 2014; Luig & Santocchi, 2014; Magnusdottir et al., 2022; Ruiz et al., 2010; Xu et al., 2024; Yang et al., 2023).

Table 4

Antecedents and Categories of Tasks, Conditions, and Physiological Factors Associated With Workload, Fatigue, Stress, and Sleepiness

Study (first author)	Year	Antecedents	Categorization of Antecedents	Duration of exposure	Phenomenon
Alpert	1988	Presentation of numbers from 1 to 6 in a predetermined order, and the subject must press a button every time two consecutive numbers add up to seven.	Complex tasks	10 minutes	Workload
Cañas	2022	Mental workload (e.g manage air traffic flow)	Complex tasks	NS	Fatigue
Congleton	1997	Stimulating events, such as the number of enemy aircraft launched	Complex tasks, emergency management	NS	Stress
Ćosić	2019	Stressful emotional stimuli, like acoustic startle, airblast, semantically relevant aversive images and sounds, fear-potentiated startle, prepulse inhibition etc.; a variety of cognitive tasks with different workload intensity, like multiple versions of Stroop test.	Environmental conditions, complex Tasks	45 minutes	Stress
De Vasconcelos	2019	High mental workload, shift planning characteristics, limited cabin space, diverse maneuvers, alternating acceleration forces, poor airflow, low lighting, continuous background noise, and vibrations, self-define as tired and sleepy.	Environmental conditions and Complex Tasks; physiological factors	NS	Fatigue, Sleepiness
Huttunen (a)	2010	High-demand situational management (informational load, decision load and situational awareness).	Complex Tasks	5 minutes	Workload
Huttunen (b)	2010	High-demand situational management (informational load, decision load and situational awareness).	Complex Tasks	NS	Workload
Kouba	2023	Night shifts and factors such as high mental workload and prolonged operation duration.	Physiological Factors and complex Tasks	25 minutes	Fatigue
Khan	2015	Exposure to hypoxic conditions, both normobaric and hypobaric.	Environmental conditions, physiological Factors	The exposure lasted 4 hours for 4 consecutive days	Stress
Krajewski	2012	Long working hours, movement restrictions, low light levels, background noise, and a high workload.	Environmental Conditions, complex Tasks, Physiological Factors	8 days	Sleepiness

Luig	2014	Execution of complex tasks and simulated flight with malfunctions (critical situations)	Complex tasks, emergency management	from 8:00 PM to 4:00 AM	Stress
Magnusdottir	2022	Execution of complex cognitive tasks, such as the Stroop color-word test.	Complex Tasks	3.5 hours	Workload
Maina	2023	Insufficient sleep, workload, and circadian rhythm disturbances	Complex Tasks, physiological Factors	45 minutes	Fatigue
Ruiz	2010	Consecutive flight rotations, lack of time to rest, irregular working hours, sudden changes in daily rhythm	Physiological factors	18-20 hours	Fatigue
Shao	2021	Tasks at airports at different altitudes	Complex Tasks	NS	Workload
Shen	2021	High workload, biological and psychological factors, work shifts	Complex Tasks, physiological Factors	NS	Fatigue
Whitmore	1996	Sleepiness	Physiological factors	NS	Fatigue
Xu	2024	High workload, the need for high performance, and the physical and psychological effects associated with managing stressful situations	Environmental conditions, complex Tasks, Physiological Factors	36 hours	Fatigue
Yang	2023	Increase in air traffic flow	Complex tasks	six sets of control simulation tasks, each lasting 30 minutes	Workload

4.3.3 Acoustic analysis

The studies utilize diverse software tools, including PRAAT, OpenSMILE, OpenEAR, MATLAB, LabVIEW, and machine learning or deep learning models, to analyze specific acoustic parameters associated with vocal processes such as phonation and resonance (Table 5). Six studies employ

PRAAT as a primary tool for acoustic analysis (Cañas et al., 2022; De Vasconcelos et al., 2019; Huttunen et al., 2011(b); Khan et al., 2015Krajewski et al., 2014; Maina & Zhang, 2023). All studies use acoustic parameters focusing on phonation, resonance, and articulation, with none addressing respiration (Figure 5).

Table 5

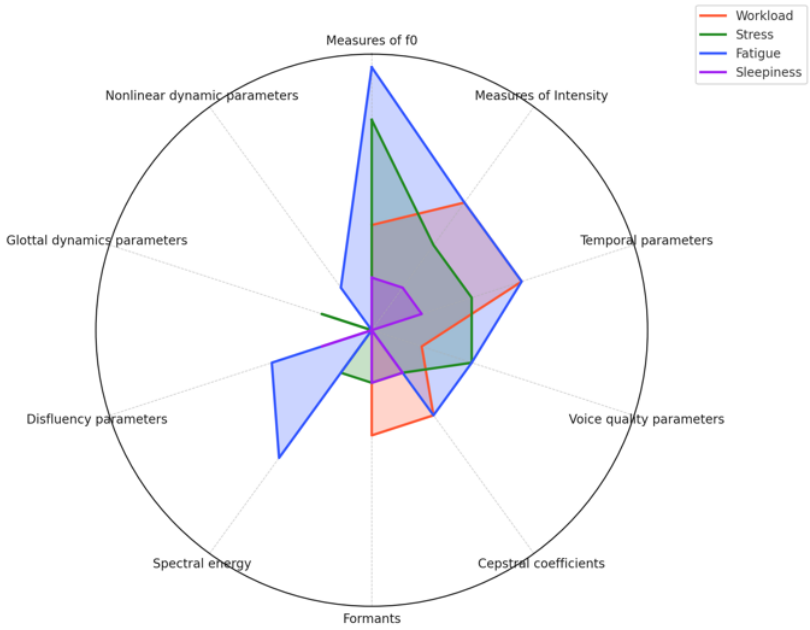
Overview of the software used for the acoustic parameters' extraction and of the analyzed speech processes

Study (first author)	Year	Phenomenon	Software	Acoustic Parameters	Speech Process
Alpert	1988	Workload	Hybrid analog/digital system	F_0, amplitude, the duration of utterances and pauses	Phonation, Articulation
Cañas	2022	Fatigue	PRAAT	F_0	Phonation
Congleton	1997	Stress	SWIFFT	F_0, amplitude, jitter, shimmer	Phonation
Ćosić	2019	Stress	OpenSMILE	F_0, amplitude, RMS, jitter, shimmer, formants, zero-crossing rate, cepstral coefficients, Spectral flux, speech rate, number of voiced segments per second and mean, voiced/unvoiced segment lengths in seconds.	Phonation, Resonance, Articulation
De Vasconcelos	2019	Fatigue, Sleepiness	PRAAT	Total articulation time, total pause time, elocution time, number of pauses, average pause duration, elocution rate, articulation rate, percentage of disfluency	Articolation
Huttunen (a)	2010	Workload	Algorithm using waveform matching	F_0, loudness	Phonation
Huttunen (b)	2010	Workload	PRAAT	formants, articulation rate	Resonance, Articulation
Kouba	2023	Fatigue	Layered Voice Analysis	F_0, the central point of the mass of the sampled spectrum, the uniformity of the distribution of the plateaus' lengths and the average length of plateaus on the speech wave form, contribution of the highest frequency area in the sampled frequency range, uniformity of the frequency, % contribution of the most significant frequency, contribution of the lowest frequency range, the highest absolute amplitude on the wave form, the presence of "thorns" (indicative of a local high frequencies superimposed on the wave), and the presence of plateaus in the voice sample (indicative of low frequencies superimposed on the wave)	Phonation

Khan	2015	Stress	PRAAT	F_0	Phonation
Krajewski	2012	Sleepiness	PRAAT	F_0, amplitude, formants, MFCC, ratio of frequency band	Phonation, Resonance
Luig	2014	Stress	PureData software	F_0, open quotient gradient, glottal opening gradient, skewness gradient, rate of closure gradient, incompleteness of closure, syllable rate, syllables per second	Phonation, Articulation
Magnusdottir	2022	Workload	KARMA algorithm	Formants	Resonance
Maina	2023	Fatigue	PRAAT	Total articulation time, total pause time, total disfluency time, elocution time, number of pauses, average pause duration, number of syllables, elocution rate, articulation rate, number of disfluencies and percentage of disfluency.	Phonation, Articulation
Ruiz	2010	Fatigue	MATLAB	F_0, coefficient of variation, jitter, shimmer, shimmer factor, digital amplitude length, spectral center of gravity, spectral balance frequency, energy balance frequency, spectral moments, formants, spectral distance, maximal lyapunov exponennt	Phonation
Shao	2021	Workload	OpenEAR	MFCCs	Phonation, Resonance
Shen	2021	Fatigue	Deep learning	Amplitude, f0, RMS, Sum of the RASTA-style filtered auditory spectrum, Energy in bands from 250 to 650 Hz and from 1 to 4 kHz, Spectral roll-off points of 25%, 50%, 75% and 90%, Spectral flux, Spectral entropy, Spectral variance, Spectral skewness Spectral kurtosis, Spectral slope, MFCC, RASTA-style auditory spectrum bands 1–26, Probability of voicing, Jitter, variation in jitter, shimmer	Phonation
Whitmore	1996	Fatigue	LabVIEW	F_0, word duration	Resonance
Xu	2024	Fatigue	Deep learning	Zero-crossing rate, MFCC, RMS, chromagram, Mel Spectrogram	Phonation, Resonance
Yang	2023	Workload	Deep learning	Speech speed, amplitude, zero crossing rate, Mel spectrogram	Phonation, Resonance

Figure 5

Radar Chart of acoustic parameters used in the selected studies



Workload

In studies examining workload, f_0 has been frequently identified as a primary indicator of vocal changes (Alpert & Schneider, 1988; Magnusdottir et al., 2022; Huttunen et al., 2011(a)). Moreover, there was also a possible trend for the variance of f_0 to be lower in the high workload condition (Alpert & Schneider, 1988; Huttunen et al., 2011(a)).

In the context of flight simulations, a mean increase of approximately 7 Hz in F0 has been reported under high cognitive load, with an increase up to 12 Hz during the most demanding flight phases (Huttunen et al., 2011(a)). An increase in vocal intensity of at least 1 dB has also been documented (Alpert & Schneider, 1988; Huttunen et al., 2011(a)), along with a reduction in intensity variance by 1 dB (Huttunen et al., 2011(a)). There is a potential trend for voice frequency and amplitude to be higher and for voice frequency variance to be lower under high workload conditions compared to low workload conditions (Alpert & Schneider, 1988; Huttunen et al., 2011(a)). Furthermore, voice under workload appears to exhibit faster reaction times (Alpert & Schneider, 1988). According to Yang et al. (2023), vocal markers of workload include changes in speech rhythm, stuttering (indicative of impaired command memory), and hesitations (suggesting unclear understanding of the air situation). The articulation rate may decrease as cognitive load increases, indicating a slower speech pace

(Huttunen et al., 2011(b)). Changes in formant frequencies, such as an increase in F1 and a decrease in F2 for front vowels, have been observed under high workload conditions (Huttunen, 2011(b); Magnusdottir et al., 2023). MFCCs also show promise as markers (Shao et al., 2021, p.1; Yang et al., 2023).

Studies on ATCOs use more complex parameters such as MFCC (Yang et al., 2023; Shao et al., 2021), making it difficult to compare with studies with pilots. In terms of setting, no differences were found in acoustic markers between studies that used a simulator setting (Huttunen et al., 2011a; Huttunen et al., 2011b) and those conducted in a laboratory setting (Alpert & Schneider, 1988; Magnusdottir et al., 2022), nor between civilian (Magnusdottir et al., 2022) and military contexts (Alpert & Schneider, 1988; Huttunen et al., 2011a; Huttunen et al., 2011b).

Stress

It has been observed that f_0 increases in response to stress (Khan et al., 2015; Congleton et al., 1996; Ćosić et al., 2019; Luig & Santocchi, 2014). Loudness appears to increase under stress (Ćosić et al., 2019), although one study reports no such effect (Congleton et al., 1996). Jitter tends to decrease during stress (Congleton et al., 1996), although one study has found it may increase (Ćosić et al., 2019). Shimmer measurements did not display significant results (Congleton et al., 1996; Ćosić et al., 2019). The amplitudes of formants and other measures related to spectral composition tend to shift toward higher values under stress conditions (Ćosić et al., 2019). Additionally, spectral characteristics and MFCCs exhibit increased variability and changes in their distributions in response to stressful situations, highlighting their utility as indicators of emotional and cognitive states (Ćosić et al., 2019). Under stress, glottal dynamic parameters can also serve as valid measures of stress (Luig & Santocchi, 2014).

There are no significant differences between studies conducted in military contexts (Khan et al., 2015; Congleton et al., 1996) and those in civilian settings (Ćosić et al., 2019; Luig & Santocchi, 2014), nor between studies focusing on ATCOs (Ćosić et al., 2019) and pilots (Khan et al., 2015; Congleton et al., 1996; Luig & Santocchi, 2014). Similarly, no notable differences were observed based on the setting, whether in a laboratory (Ćosić et al., 2019; Khan et al., 2015) or in a simulator (Congleton et al., 1996; Luig & Santocchi, 2014). Two key distinctions emerged: in Ćosić's study (civilian ATCOs in a laboratory setting), both jitter and loudness increased, whereas in Congleton's study (military pilots in a simulator), jitter decreased, and amplitude showed no significant relationship.

Fatigue

Studies have shown that F_0 can decrease with increased fatigue, reflecting a loss of muscle tone and motivation (Kouba et al., 2022; Whitmore & Fisher, 1996; Ruiz et al., 2010; Shen et al., 2021). Loudness also diminishes under fatigue (Kouba et al., 2023). Parameters such as jitter and shimmer can indicate fatigue due to increased variability, suggesting a loss of fine vocal control (Ruiz et al., 2010).

Word duration may increase with fatigue (Whitmore & Fisher, 1996; Maina & Zhang, 2023) while speech rate and articulation rate tend to decrease as fatigue levels rise (de Vasconcelos et al., 2019; Maina & Zhang, 2023). Increased fatigue also leads to a higher number and longer duration of pauses and disfluencies in speech, including silent pauses, hesitations, and repetitions (Maina & Zhang, 2023; de Vasconcelos et al., 2019). From a perceptual standpoint, one study noted that voices under fatigue exhibited vocalizations similar to "vocal fry" or creaky voice, which is associated with low subglottal pressure and low vocal intensity (de Vasconcelos et al., 2019). Furthermore, the Digital Amplitude Length (DAL) parameter has shown significant variations related to fatigue, suggesting it may be a useful indicator for monitoring changes in pilots' voices under fatigue conditions (Ruiz et al., 2010). Regarding the Maximum Lyapunov Exponent (λ), this parameter can reflect the presence or absence of chaos in the vocal signal. It has been observed that variations in λ values might correspond to states of fatigue in subjects, making it a potential fatigue indicator (Ruiz et al., 2010). Increased fatigue may also lead to greater instability in vocal production, reflected in changes in the ZCR and in the MFCC (Xu et al., 2024). There are no differences between studies involving ATCOs and pilots regarding military context, sample type (ATCOs vs. pilots), or study setting.

Sleepiness

Studies have observed that sleepiness leads to a decreased pitch rate and reduced variation in f_0, reflecting a monotonic vocal pattern (Krajewski et al., 2014). During states of sleepiness, articulation becomes more slurred, with smaller differences in formant positions, indicating that vocalizations become less clear and more indistinct (Krajewski, 2012). Specifically, changes in the positions of formants—particularly formant 1 (F1) and formant 3 (F3)—show a negative correlation with sleepiness scores. This suggests alterations in the shape of the vocal tract due to sleepiness, leading to less defined speech production (Krajewski et al., 2014). Additionally, a decrease in vocal tension quality is observed during sleepiness, accompanied by changes in the spectral profile of the voice, such as variations in the spectral density ratio across different frequency bands (Krajewski, 2012). Sleepiness also increases pauses and decreases fluent articulation time, leading to more frequent silent pauses, hesitations, and disruptions in speech flow (de Vasconcelos et al., 2019).

4.3.4 Machine learning approaches

Deep learning approaches are represented by studies such as Yang et al. (2023), Shen et al. (2021), and Huang et al. (2024). Yang et al. (2023) introduced a hybrid model combining Convolutional Neural Networks (CNN) and Transformer Encoders (SCNN-TransE). The model utilized Mel spectrograms to capture the spatiotemporal features of vocal signals, which were fed into the CNN component. With an accuracy of 97.48%, it outperformed traditional machine learning techniques such as K-Nearest

Neighbors (KNN), Support Vector Machines (SVM), Random Forests (RF), and AdaBoost. Shen et al. (2021) examined a range of neural network architectures for fatigue detection, including Autoencoders and Convolutional Autoencoders. Among these, the Densely Connected Convolutional Autoencoder achieved the best performance, with an accuracy of 98.35% in detecting fatigue among ATCOs. Huang et al. (2024) leveraged recurrent architectures, specifically Long Short-Term Memory (LSTM) networks in combination with Gated Recurrent Units (GRU). Their LSTM-based model demonstrated strong temporal modeling capabilities, achieving a peak accuracy of 95.12% for fatigue detection. In contrast to these single-modal approaches, Xu et al. (2024) proposed a multimodal framework. They introduced a dual-stream Convolutional Neural Network (CNN) for fatigue detection, integrating vocal and facial data. This multimodal model achieved an accuracy of 98.03%, significantly outperforming single-modal approaches such as radio telemetry, which alone reached only 62.88%. The findings underline the superiority of multimodal systems in improving detection reliability in operational contexts. Traditional machine learning approaches were explored in studies focusing on feature extraction and classification: Shao et al. (2021) investigated the impact of feature selection by comparing MFCC and Log-Mel spectrograms. Using AdaBoost and SVM, the model achieved an average accuracy of 69.96% during K-fold cross-validation. Log-Mel spectrograms were found to be moderately effective for classifying fatigue states. Krajewski et al., (2014) employed a regression-based approach to model drowsiness states using acoustic features extracted from speech. Utilizing a leave-one-sample-out cross-validation protocol, the study demonstrated the early potential of machine learning in fatigue detection, though it did not reach the performance levels of deep learning methods. Magnusdottir et al. (2022) presented a distinctive approach by utilizing formant track features for vocal signal analysis. The study applied machine learning classifiers, including SVM and RF, achieving misclassification rates of 61.75 ± 0.18% (SVM) and 53.50 ± 0.18% (RF) using vocal data alone. These results highlight the limitations of relying solely on vocal data and suggest the importance of integrating multiple modalities for enhanced performance.

4.4 Discussion

This systematic review explored acoustic markers linked to workload, stress, fatigue, and sleepiness within aviation contexts. The review highlights a non-invasive approach for monitoring these psychophysiological states by analyzing vocal characteristics of speech, which could finally be helpful in enhancing aviation safety and personnel well-being. Workload appears to be a relatively straightforward phenomenon compared to others, as its antecedent is exclusively task complexity. On the other hand, stress involves a broader range of antecedents and is the only phenomenon that includes emergencies as an antecedent. Fatigue and sleepiness almost always involve at least one physiological antecedent, highlighting that their origin is deeply rooted in the body's biological processes and is less

dependent on external cognitive demands (Matura et al., 2018). Although stress is more complex than workload regarding conceptualization and antecedents, and despite being distinct concepts rooted in different theoretical frameworks, the two phenomena display similar vocal patterns. Other studies have also shown that they are correlated as far as the physiological response is regarded (Alsuraykh et al., 2019; Gaillard, 1993). Both phenomena involve physiological responses driven by heightened activation of the sympathetic nervous system and suppression of the vagal system, leading to typical cardiovascular reactions such as increased blood pressure and heart rate (Mandrick et al., 2016), as well as vocal changes that progress in parallel (Abur et al., 2023). For instance, stress and workload are associated with increased F0 and vocal intensity, likely due to heightened sympathetic nervous system activity and the hypothalamic-pituitary-adrenal axis activation (Van Puyvelde et al., 2018). In contrast, fatigue and sleepiness are associated with reduced vocal energy, slower articulation rates, and increased pauses, reflecting diminished cognitive and physical readiness. Speech becomes less clear and dynamic, and the flow of speech fragments due to exhaustion. In this case, the vocal system struggles to maintain the required level of responsiveness for effective communication, revealing a decline in physical and cognitive capacities. These two phenomena appear to be linked to lower arousal levels and decreased central nervous system activity (Bakotić & Radošević-Vidaček, 2012; Diaz-Piedra et al., 2019).

In other words, stress and workload are conditions where the body is "charged" to meet high demands, both cognitive and emotional, which is reflected in greater intensity and variability in acoustical correlates of speech. In these cases, the vocal system is in a heightened state of responsiveness. Conversely, these parameters indicate that fatigue and sleepiness represent "weakness", where the body experiences a depletion of resources.

Finally, MFCC parameters are robust and versatile indicators for monitoring states such as workload, stress, fatigue, and sleepiness (Figure 6). MFCC parameters emerged as versatile indicators, demonstrating their promising role in capturing and differentiating these psychophysiological states. However, their effectiveness significantly improves when combined with other classic acoustic parameters (as demonstrated in a different domain by Verma et al., 2023 and Singh et al., 2012), emphasizing the importance of multimodal approaches for real-world applications. Thus, the first hypothesis was confirmed (H1: Workload, stress, fatigue, and sleepiness are interrelated yet distinct constructs with specific acoustic correlates).

Regarding the role of environmental factors, this study did not identify any vocal differences associated with environmental variations. This is partly due to the often incomparable methodologies or acoustic parameters employed. However, it may also stem from these states being strongly rooted in physiological processes, leading to consistency across different contexts. Similarly, workload appears to be a relatively simple phenomenon and less influenced by environmental variations. In contrast, stress showed notable differences between two studies. In Ćosić's study (civilian ATCOs in a laboratory

setting), both jitter and loudness increased, whereas in Congleton's study (military pilots in a simulator), jitter decreased, and amplitude showed no significant relationship. However, the results across studies appear broadly consistent, suggesting that contextual factors have minimal influence on vocal production.

This implies that variations in vocal parameters are primarily driven by intrinsic factors within the individual, such as physiological states and personal responses, rather than by the specific measurement environment. It could also be claimed that environmental aspects influence the extent or intensity of vocal changes rather than the general direction of the markers. The relative stability of vocal parameters observed across different contexts reinforces the idea that voice can be a robust indicator of psychophysiological states, regardless of environmental variations. This result gains further significance when considering that the literature highlighting vocal markers of these phenomena in other contexts (reviewed in the introduction) aligns with the findings observed in the aviation context analyzed here. Thus, the second hypothesis was not supported (H2: Environmental factors significantly influence how workload, stress, fatigue, and sleepiness manifest in voice in different ways). However, this is also due to the limited comparability of the studies: had the same acoustic measures been used, different results would likely have emerged. However, this limited comparability extends across various aspects. Some studies did not specify the language used (Cañas et al., 2022; Huang et al., 2024; Kouba et al., 2023; Khan et al., 2015; Shao et al., 2021; Shen et al., 2021; Yang et al., 2023), which poses a significant limitation. Language plays a crucial role in shaping vocal characteristics, and its omission prevents a proper evaluation of the generalizability of the results. Phonetic and prosodic variations across languages can influence speech analysis outcomes, making it unclear whether findings can be applied across different linguistic groups. Differences in how stress, workload, fatigue, and sleepiness are measured across studies further complicate comparisons and hinder the ability to draw definitive conclusions. The use of varied methodologies, different acoustic parameters, and different software points to a broader standardization challenge in this field. Few studies have been conducted in real-world settings (e.g., Cañas, 2022; De Vasconcelos, 2019; Maina & Zhang, 2023; Shen et al., 2021), and for workload specifically, no ecological studies exist in the aviation context. Addressing these limitations through increased methodological standardization, inclusion of linguistic considerations, and a greater focus on ecological validity will be essential for advancing the field and ensuring that findings are both robust and generalizable.

Moreover, the inconsistent use of terminology across disciplines such as cognitive neuroscience, physiology, psychology, medical sciences, and engineering creates communication barriers. This fragmentation limits the dissemination of findings beyond individual fields, often leading to the “reinventing-the-wheel” phenomenon, where similar discoveries or innovations are independently developed in isolation. To advance this area, cross-disciplinary collaboration is essential to establish standardized definitions and protocols for assessing psychophysiological states through vocal analysis.

Standardization would enable more effective comparisons across studies and facilitate the translation of research findings into practical tools to improve aviation safety and performance through voice monitoring.

4.4.1 Limitations and future directions

This study has certain limitations that should be acknowledged. First, the decision to include only peer-reviewed papers resulted in the exclusion of numerous studies, potentially narrowing the scope of the findings. Second, cultural differences and the languages spoken were not considered, as the limited number of available studies hindered robust comparisons. Despite these limitations, this review provides a valuable synthesis of evidence on acoustic markers linked to workload, stress, fatigue, and sleepiness, contributing to the growing interest in non-invasive, real-time monitoring systems. Importantly, it highlights the pressing need for standardized definitions, methodologies, and protocols, offering a roadmap for future research to enhance both the reliability and applicability of findings. Future research should prioritize the development of standardized frameworks to improve the comparability and consistency of results across studies. Additionally, a deeper focus on cultural and linguistic diversity is essential to expand the global understanding of acoustic markers associated with psychophysiological states, addressing a critical gap identified in the current review.

4.4.2 Practical Implications

The findings of this review underscore the potential for implementing real-time and non-invasive vocal monitoring systems of states affecting pilots' and ATCOs' performance and well-being within the aviation industry. Acoustic markers associated with workload, stress, fatigue, and sleepiness offer a promising avenue for monitoring psychophysiological states, enabling timely interventions to prevent dangerous situations and enhance performance.

4.5 Conclusions

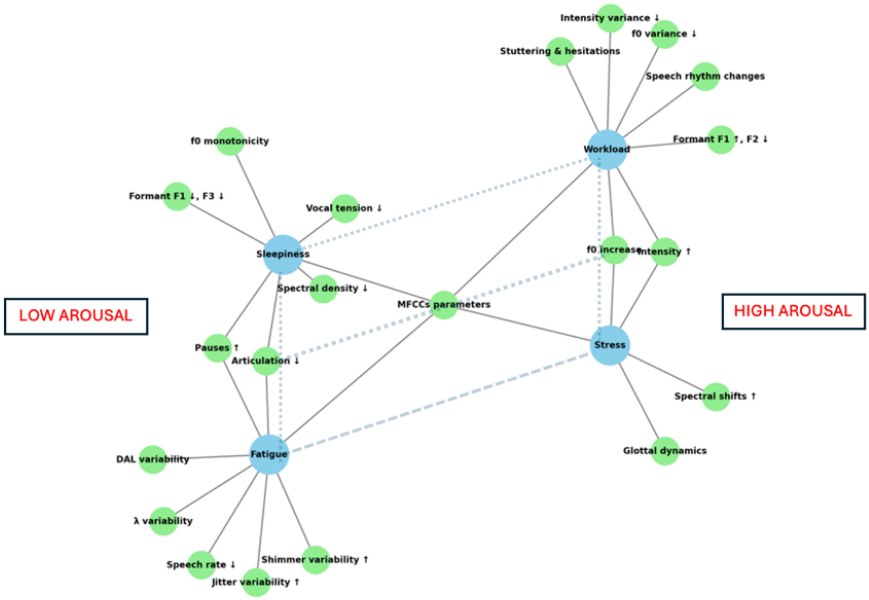
This systematic review highlights the potential of using acoustic markers as non-invasive indicators of workload, stress, fatigue, and sleepiness in aviation contexts. These findings emphasize the robustness of vocal analysis for real-time monitoring of psychophysiological states, offering a promising avenue to enhance safety and performance in the aviation industry. A key takeaway from this review is the interconnected yet distinct nature of these constructs, as evidenced by their specific vocal markers. Stress and workload are characterized by similar vocal patterns, such as heightened vocal intensity and pitch, which reflect increased activation of the sympathetic nervous system, related to both cognitive and affective responses to a challenge posed by the environment. In contrast, fatigue and sleepiness

share vocal patterns like reduced vocal energy, slower speech rates, and increased pauses, indicative of decreased central nervous system activity.

Despite these distinctions, the overlap in their physiological and cognitive processes underscores the complexity of these states, which vocal analysis can effectively capture. The role of contextual factors, as explored in this review, appears limited in influencing vocal production, with variations in acoustic markers primarily attributed to intrinsic physiological states rather than external environments. However, methodological variability across studies—such as differences in theoretical frameworks, acoustic measures, software, and contexts—limits comparability and highlights the urgent need for standardized protocols. Additionally, the lack of ecological studies hinders the translation of findings to real-world operational settings, further emphasizing the importance of validating research in practical aviation environments. In conclusion, vocal analysis represents a valuable tool for advancing aviation safety and performance. However, achieving its full potential will require overcoming current standardization, methodology, and interdisciplinary communication challenges. This review serves as a foundation for future research, providing a roadmap to enhance the reliability and applicability of vocal monitoring systems in aviation settings.

Figure 6

Network of associations between acoustic markers and psychophysiological states across arousal



Note. The diagram shows the connection between MFCC parameters and four states: workload, stress, fatigue, and sleepiness. Each state is linked to specific vocal changes, highlighting MFCCs' central role

as versatile indicators for psychophysiological monitoring. The dashed lines represent a conceptual framework illustrating the complex interrelationships among fatigue, stress, workload, and sleepiness, emphasizing their mutual influence without a clear, unidirectional causal pathway.

CHAPTER 5

Identifying Acoustic Markers of Stress in Aviation Emergencies: A Real-World Analysis of Pilot-ATC Communications[5]

Introduction: Acute stress is a major risk factor for human error in aviation, yet its real-world acoustic manifestations remain poorly understood. This study investigates real-world voice communication between pilots and air traffic controllers (ATCOs) during emergencies to identify acoustic indicators of situational stress.

Method: We analyzed 90 recordings from LiveATC.net, manually annotated for speaker (pilot vs ATCO) and flight phase (routine vs emergency). A total of 27 acoustic features were extracted using Parselmouth. Feature selection combined Welch ANOVAs with FDR correction, collinearity checks, and Boruta random forest filtering. Classification was performed using LDA, LASSO, Random Forest, and XGBoost, validated with leave-one-event-out cross-validation.

Results: ATCOs and pilots exhibited robust stress-related vocal changes. Ensemble models achieved good accuracy (for example, for RF = .96 ATCO, .97 Pilot). Shared stress markers included higher F0 mean and minimum and lower harmonic-to-noise ratio (HNR), reflecting F0 elevation and reduced voice quality. Divergences emerged by role: ATCOs showed increased intensity variability, F0 instability, and upward F3 shifts, indicating reduced vocal stability. Pilots demonstrated faster articulation, stronger F0 max increases, and higher mean intensity, consistent with accelerated, information-dense communication.

Conclusions: This study highlights robust, role-specific acoustic markers of acute stress in real-world aviation emergencies. The findings may inform real-time voice-based stress monitoring tools to enhance aviation safety.

***Keywords**:* Vocal markers, acoustic analysis, aviation safety, situational stress, Pilot-ATCO Communications, real-world data

[5] This work, authored by Gnerre and Biassoni, is currently under peer review at the *Aviation Psychology and Applied Human Factors Journal.*

5.1 Introduction

On January 15th, 2009, US Airways Flight 1549 suffered a dual engine failure shortly after takeoff, prompting the crew to execute an emergency water landing in the Hudson River, at the heart of Manhattan. All 155 individuals on board were successfully evacuated, an outcome largely attributed to the exceptional professionalism and composure of Captain Chesley Sullenberger. Despite the extreme psychological pressure, Sullenberger remained focused, maintained control of the aircraft, and carried out a complex maneuver under time-critical conditions. This event is frequently cited as a paradigmatic example of how the capacity to manage acute stress can critically influence outcomes in high-risk operational contexts. Had the pilot failed to maintain cognitive and emotional control, the consequences would likely have been catastrophic. Such cases highlight the pressing need to better understand the impact of stress on human performance, especially in safety-critical domains such as aviation. In this specific context, high stress levels have been consistently identified as a major contributing factor to pilot error (Causse et al., 2013; Fornette et al., 2012; Li et al., 2001). As technological innovation continues to reduce the incidence of mechanical failure, human error has emerged as the primary cause of aviation accidents (IATA, 2014). This shift has drawn increasing attention to the intrinsic limitations of human performance in complex socio-technical systems. As Reason (1995) argues, it is unrealistic to expect flawless behavior from individuals operating under intense cognitive and environmental demands. Even the most sophisticated risk mitigation systems cannot fully eliminate the potential for error. Accordingly, the emphasis has moved from the unrealistic goal of eradicating human fallibility to the more feasible objective of developing resilient systems—systems capable of anticipating, absorbing, and adapting to inevitable lapses in human performance (Kelly & Efthymiou, 2019).

Pilots and ATCOs operate within environments characterized by high physical and cognitive demands, including irregular work schedules, extended duty periods, and exposure to unpredictable or adverse environmental conditions (Lee & Kim, 2018; Mélan & Cascino, 2022). In these high-stakes operational settings, the human voice plays a dual function: it is not only a medium for transmitting mission-critical information, but also a psychophysiological signal that reflects the speaker's internal state (Cañas et al., 2022; Huttunen et al., 2011; Kuroda et al., 1976; Sondhi et al., 2015). Variations in vocal parameters can provide valuable insight into stress levels, cognitive workload, and emotional arousal, especially during time-sensitive and safety-critical situations (Van Puyvelde et al., 2018).

Despite growing interest in the diagnostic potential of vocal features, empirical investigations into how stress manifests in real-world aviation communications—particularly between pilots and ATCOs during actual emergencies—remain limited. Much of the existing literature consists of outdated case reports (e.g., Benson, 1995; Kuroda et al., 1976), lexical-level analyses of cockpit discourse (e.g., Huang et al., 2021; Prinzo, 1988), or experimental studies conducted under controlled laboratory conditions (Congleton et al., 1997; Huttunen et al., 2011). To date, no study has systematically examined

a corpus of real aviation emergencies—defined here as high-stress, time-critical scenarios requiring rapid decision-making and management of critical incidents—using acoustic analysis. This gap in the literature significantly limits our ability to understand how vocal stress markers operate precisely when early detection might be most valuable. Monitoring the psychophysiological state of aviation personnel during such events is therefore essential to preventing failures that could result in severe human and material losses.

5.1.1 The stress response

While stress has long been a central construct in psychological theory, its definition remains fragmented (Masi et al., 2023). Selye's General Adaptation Syndrome (1950) characterized stress as a nonspecific physiological reaction to any disruption of homeostasis. However, contemporary perspectives increasingly view stress as a highly individualized and context-sensitive process (Wheaton & Montazer, 2010). These individual differences are particularly relevant in high-stakes, safety-critical domains such as aviation, where the ability to appraise and regulate acute stress in real time is essential to performance and safety (Masi et al., 2023). In such environments, cognitive appraisal is influenced by several factors, including task-relevant expertise, personality traits, and preferred coping styles.

Moreover, in high-demand operational settings such as cockpits and towers, stress emerges from the combined effect of task-related and environmental factors. On one hand, the cognitive demands associated with monitoring complex systems, processing large volumes of information in real time, and making rapid, high-stakes decisions place a constant strain on attentional and executive resources (Silberstein & Dietrich, 2003). The physical environment itself can act as an additional source of stress. Environmental stressors (such as persistent noise, poor air quality, limited space, and visual or auditory overload) can further impair attention, decision-making, and communicative clarity (Hagmüller et al., 2006). These stressors do not simply form the backdrop of operations; they actively interact with cognitive load to shape human performance in safety-critical contexts.

5.1.2 Speech under stress

Under stress or elevated workload, speakers often exhibit systematic alterations in vocal production, including changes in pitch, intensity, speech rate, and prosodic timing (König et al., 2021). Empirical studies have documented such vocal shifts in real-world high-stress situations, such as aviation emergencies and large-scale disasters (Benson, 1995; Hausner, 1987; Kuroda et al., 1976; Ruiz et al., 1996). For example, Williams and Stevens (1972) analyzed vocal patterns during the Hindenburg disaster, while Streeter et al. (1983) examined acoustic markers of stress in professional exchanges during the 1977 New York blackout.

A substantial body of research on vocal behavior under stress consistently identifies a significant increase in F0 as one of the most robust acoustic correlates of high-stress situations (Giddens et al., 2013; Griffin & Williams, 1987; Kappen, Donckt, et al., 2022; Kappen, Hoorelbeke, et al., 2022; Rothkrantz et al., 2004; Van Puyvelde et al., 2018). This elevation in F0 is typically attributed to increased subglottal pressure, a physiological response triggered by stress (Giddens et al., 2013; Kurniawan et al., 2013). One proposed mechanism involves the activation of the sympathetic nervous system, which leads to elevated heart rate and bronchodilation, thereby influencing phonation indirectly (Eden & Inbar, 1978; Orlikoff, 1990).

In contrast to broader prosodic trends, microvariations in frequency and amplitude—specifically, jitter and shimmer—have shown inconsistent responses to stress. While some studies report an increase under stress (Postma-Nilsenová et al., 2016), others have observed a decrease (Giddens et al., 2013). Notably, in line with findings on F0 modulation, the effects on jitter and shimmer appear more pronounced during emotionally charged emergency situations (Brenner et al., 1985) than under conditions of purely cognitive load (Brenner et al., 1994). As for the HNR, findings remain mixed, with some evidence indicating an increase (Depestele, 2023; Gaur et al.,2024; Pisanski & Sorokowski, 2021) and others reporting a decrease (Kappen et al., 2022; Sondhi et al., 2015).

These patterns suggest that while stress induces generalized physiological arousal, it may also enhance neuromuscular control in certain contexts, particularly those requiring clear and precise communication. An illustrative example is provided by Benson (1995), who analyzed the vocal behavior of a pilot during a real-world emergency. The pilot's voice remained remarkably composed, with only two brief deviations from an otherwise steady vocal tension. Although signs of arousal were present, the speech output did not become disorganized or overtly distressed.

Notably, mean sentence length dropped from 5.6 words per utterance prior to the emergency to 3.7 words afterward, suggesting a shift toward concise and efficient verbal output under pressure. A systematic review by Giddens et al. (2013) reinforced the finding that F0 increase is the most consistent and robust vocal marker of stress. This acoustic change is believed to stem from physiological adjustments such as heightened tension in the cricothyroid muscle and elevated subglottal pressure, both of which influence vocal pitch. While not universal (Hecker et al., 1968), this effect has been widely replicated, and F0-related parameters are among the most distinctive indicators of stress (Demenko & Jastrzębska, 2012; Hagmüller et al., 2006). Additional acoustic features have also been associated with stress. These include maximum F0 and F0 standard deviation, the latter of which is often elevated in high-stress conditions (Protopapas & Lieberman, 1997; Scherer et al., 1981, 2002; Sondhi et al., 2015; Streeter et al., 1982), though some exceptions have been reported (Van den Broek, 2003). Formant frequencies F1 and F2 also tend to increase in most speakers under stress (Protopapas & Lieberman, 1997; Sondhi et al., 2015). Increased vocal intensity is another frequently reported marker (Hollien et al., 1980; Scherer et al., 2002; Streeter et al., 1982).

When it comes to time related parameters, highly stressed speakers often exhibit shorter utterance durations and an increased speech rate under cognitive load (Scherer et al., 2002; Streeter et al., 1982). Respiratory changes —such as increased or irregular breathing — can lead to shortened speech segments, misplaced breaths, and disrupted timing, often accompanied by inappropriate pauses and variation in articulation rate (Baker, 2008; Hansen & Patil, 2007; Pisanski & Sorokowski, 2021). Additionally, glottal waveform parameters (Godin & Hansen, 2008) and cepstral coefficients (Dahl & Stepp, 2023) have emerged as relevant non-invasive indicators of stress in speech.

5.1.3 The present study

In this study, in light of the theoretical and empirical background reviewed, we aim to examine how acute stress, elicited by real-world aviation emergencies, modulates vocal production in two critical operator roles: pilots and ATCOs. Drawing on a growing body of research that identifies human voice as a psychophysiological indicator of stress, we investigate whether and how acoustic features vary between routine and emergency flight phases, and whether these changes differ systematically with relation to role. This investigation is guided by the following research questions:

1) *Question one*: Which acoustic parameters significantly differ between routine and emergency speech in real-world pilot and ATCO communications?
2) *Question two*: Are there role-specific patterns in vocal stress markers, such that pilots and ATCOs exhibit distinct acoustic profiles under stress because of their role and – hence – of their different perspective in the emergency?
3) *Question three*: Can stress-related vocal changes in pilot and ATCO speech be reliably detected and classified using acoustic features alone, and with what level of accuracy?

Grounded in prior literature on vocal stress, psychophysiology, and aviation communication, we propose the following hypotheses:

1) *H1*. Emergency communications will be characterized by a consistent shift in vocal parameters indicative of increased physiological arousal and cognitive load. Specifically, we expect increases in F0 and related parameters, vocal intensity, and measures of vocal effort, along with reductions in harmonicity.

2) *H2*. Given their distinct task demands, communicative strategies and considering that pilots operate under conditions where their own safety is directly at risk (Morris & Leung, 2006; Taylor et al., 2005), pilots are expected to exhibit higher and more variable F0 parameters**,** greater vocal intensity**,** increased jitter and shimmer**,** lower HNR, slower articulation rate, more frequent pauses**,** and greater spectral tilt and centroid, reflecting heightened autonomic arousal. In contrast, ATCOs are expected to maintain a more stable vocal profile,

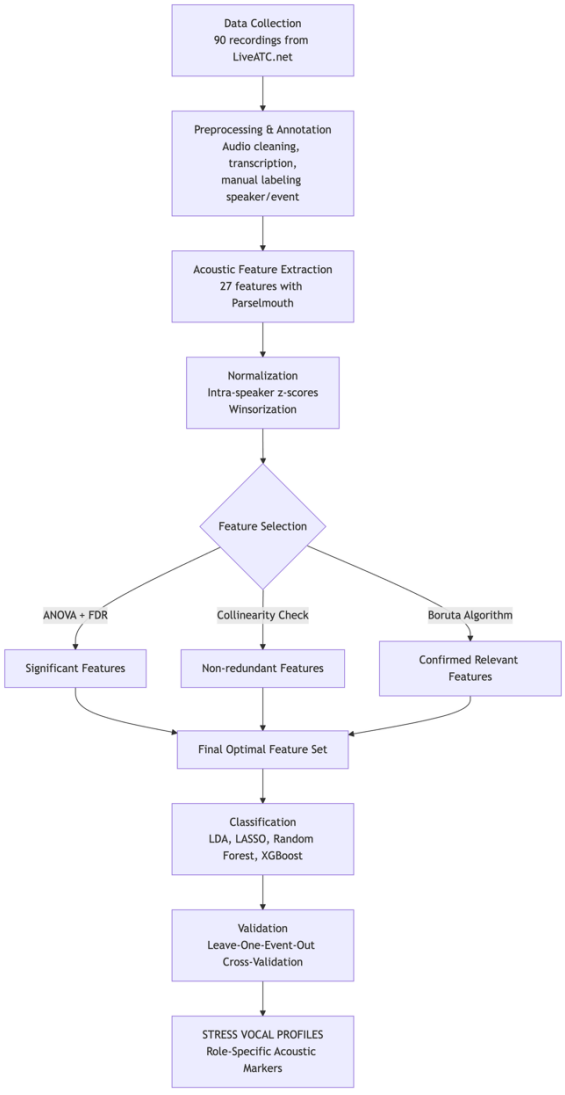
characterized by lower and more consistent F0 parameters, regular intensity**,** cleaner voice signals**,** faster speech rate**,** and a more compressed spectral distribution. Under high stress, however, they are expected to rely on automatized speech patterns, resulting in fewer and shorter pauses to meet the demands of real-time traffic management (Prinzo & Lieberman, 1998).

5.2 Method

In this section we detail the dataset construction, preprocessing pipeline, acoustic feature extraction, and statistical analysis designed to isolate stress-related vocal patterns (Figure 7). This analytic plan was preregistered on OSF.

Figure 7

Flowchart of the methodology



5.2.1 The database

The audio recordings analyzed in this study were sourced from the public database Live-ATC.net, a platform that streams and archives real-time radio communications between pilots and ATCOs. The selected data include authentic transmissions covering both routine operations and potentially critical events. Approximately 2,080 recordings were reviewed in total, of which 90 were included in the study. The total duration of the analyzed dataset amounts to 3,846.96 seconds, corresponding to 2 hours, 4 minutes, and 7 seconds of voice exchanges between pilots and ATCOs. Each event involves two primary speakers, one pilot and one ATCOs, resulting in 180 speaker instances across the dataset. Speaker gender was highly unbalanced, with only ten female speakers represented, including six controllers and four pilots. The distribution of emergencies revealed a predominance of smoke and fire events, together with engine failures or loss of thrust, which represented the most frequent scenarios across the dataset. Structural malfunctions, such as flap or hydraulic failures and anti-icing system issues, were also recurrent, particularly in the approach and landing phases. Landing gear problems accounted for another substantial portion of cases, confirming the criticality of final flight segments. Other emergencies included bird strikes, fuel shortages, and pressurization failures, which typically emerged during takeoff or cruise. Less frequent but highly severe incidents were documented as well, including medical emergencies during taxi, mid-air collisions on approach, runway excursions, bomb threats, and in-flight loss of control. Overall, the distribution highlighted a concentration of emergencies during the landing phase, followed by takeoff and cruise, suggesting that stress-inducing conditions are most likely to arise in the most operationally demanding phases of flight.

5.2.2 Inclusion and exclusion criteria

To empirically ground the temporal boundaries of the alarm phase, we first analyzed a subsample of 10 emergency. The average duration of communication following the standardized emergency declaration was approximately 90 seconds. Based on this observation, and to ensure analytical consistency across events, we operationally defined the communication of the emergency as a maximum 90-second window starting from the first emergency declaration. Within this time frame, communications typically include the pilot's formal declaration and description of the malfunction, followed by ATC acknowledgment, immediate vectors or instructions, and brief exchanges on critical details such as aircraft status and number of people onboard. This criterion helped standardize the comparison across events and allowed us to capture the period most likely to contain acute stress-related vocal modulations. Operationally, the emergency phase was defined as the segment beginning with the first standardized emergency expression (e.g., "Mayday", "Pan-Pan", "We're declaring an emergency") and continuing until one of the following conditions was met: a) a maximum duration of 90 seconds was reached; b) a significant shift in communicative content occurred.

To ensure the relevance and quality of the audio material, specific inclusion criteria were applied during the selection of recordings (Table 1). To ensure comparability and ecological validity, we exclusively selected recordings from commercial civil aviation involving jet aircraft operations and limited our selection to incidents occurring between 2005 and 2023. Recordings were required to be intelligible and of sufficient audio quality. Segments were retained only when they included two clearly identifiable communicative phases—typically, a routine phase followed by a recognizable emergency phase—and when adequate contextual information was available to reconstruct the associated incident. Recordings in which the speaker role (pilot vs. ATC) could not be confidently determined were excluded. Only recordings involving an explicit and standardized declaration of emergency were considered suitable for inclusion in the "emergency" condition. Conversely, vague references to technical issues or discomfort, in the absence of a formal emergency declaration, were excluded to maintain a consistent level of event severity. A final sample of 90 was selected.

Table 6

Summary of Inclusion and Exclusion Criteria for Emergency Communication Analysis

Category	Inclusion Criteria	Exclusion Criteria
Temporal Scope	Recordings from 2005–2023	Incidents outside this timeframe
Aviation Context	Commercial civil aviation (jet aircraft)	Military, search and rescue, non-jet aircraft
Communication Type	Direct pilot–ATCO exchanges with clear emergency declaration	Background chatter, ground personnel, unclear speaker roles
Emergency Clarity	Explicit declarations (e.g., "Mayday," "Pan-Pan," clear emergency statements)	Vague technical issues, no formal declaration, non-emergency distress
Audio Quality	Intelligible, minimal noise/clipping, reliable for acoustic analysis	Poor quality, severe interference, unidentifiable speech
Event Structure	Two distinct phases (routine + emergency), sufficient contextual information	Missing phases, incomplete incident data
Excluded Cases	—	Pilot impairment (intoxication, fatigue), non-standard emergencies

5.2.3 Data preparation

Audio files were first processed in Audacity to convert them into .wav format. Intrusive elements —such as background noise, unrelated speakers, and law enforcement communications — were manually removed. The output files of Audacity in .wav format (16-bit, 16 kHz, mono) were then imported in Praat (Version 6.4.09) (Boersma and van Heuven 2001). Audio noise reduction was performed

using the noisereduce Python library (run on Python version 3.10.9). All recordings were transcribed using the Open AI Whisper large-v3 automatic speech recognition model (Radford et al., 2022). Transcriptions were essential to support the comprehension of the context and the communicative flow, clarify the nature of the emergency, and accurately identify speaker roles (i.e., pilot vs. ATCO). Through a combination of attentive listening, and close reading of automatic transcriptions, it was possible to infer key contextual information for each event, including the type of aircraft, the year of the recording, the flight phase in which the event occurred, the type of emergency and**,** its corresponding ICAO emergency code**.** When such details were not discernible from the audio recordings alone, additional research was conducted using online sources to retrieve missing information.

5.2.4 Annotation

Each recording was also manually annotated in Praat, resulting in a dedicated TextGrid file for every audio sample. The unit of analysis is the speaking turn. Two annotation tiers were created: (1) speaker diarization – segments were labeled to indicate who was speaking (pilot or ATCO), based on auditory cues and contextual markers; (2) event diarization – segments were marked to indicate whether the communication occurred during routine operations or in the context of an emergency. To validate the speaker diarization, we also applied the *pyannote-audio* pipeline, a state-of-the-art automated speaker diarization system based on deep learning. The results obtained from the manual annotations were compared against the automated segmentation to ensure consistency and accuracy in speaker labelling (pilot vs. ATCO) (Bredin et al., 2017). To validate the event diarization, emergency segments were identified when either speaker explicitly used standard emergency phrases. To avoid any circularity between the analysed vocal parameters and the classification of flight phase (emergency vs. routine), emergency labels were assigned exclusively based on the semantic content of the transmission, rather than the non-verbal vocal behaviors. Specifically, utterances were marked as “emergency” only when pilots or ATCO explicitly used standardized emergency expressions or referred to concrete and verifiable critical events. This precaution ensured that the acoustic predictors used in subsequent analyses were independent of the labeling criteria, thus avoiding any risk of confirmation bias or tautological inference.

5.2.5 Acoustic parameter extraction

All 27 acoustic features were extracted using Python (v. 3.10.9) (Table 7). Features typically computed in Praat were obtained through the Parselmouth library (v. 0.4.3), which provides a Python interface to Praat. This program converted the audio into a Praat Sound object and extracted the features through a frame-based analysis, with 100 frames per second. All analyses of short-term spectra were performed using Praat’s default settings, which apply a 25 ms Hamming window with a 10 ms step size

(Gaussian smoothing of intensity), ensuring robust time–frequency resolution for spontaneous speech. To improve pitch tracking accuracy, we modified the default frequency range based on speaker sex: for male speakers, the range was set to 75–300 Hz, and for female speakers, to 100–500 Hz. Pauses were detected by analyzing the intensity values. If a value fell below 30 dB, a pause-duration counter began accumulating. When the intensity rose again, if the accumulated duration was at least 0.2 seconds, it was recorded as a valid pause. The number of these pauses and their average duration were saved. Jitter and Shimmer were calculated in Praat in two steps. First, the audio segment was converted into a "PointProcess" (a set of points marking the beginning of each vocal cycle). Second, this PointProcess and the original sound were passed to Praat's Get jitter (local) and Get shimmer (local) functions.

The results were then multiplied by 100 to be expressed as percentages. Formants (F1–F4) were estimated using Praat's to_formant_burg function, which applies Burg's LPC algorithm. For each segment, the mean frequency of each formant was calculated across its duration. It should be noted that, since the data consisted of spontaneous speech including different vowel contexts, these average values do not represent stable formant targets for specific vowels. Rather, they provide a descriptive proxy of overall vocal tract resonance patterns. An exploratory measure of pitch dynamics, termed F0 fluctuation was computed to capture the magnitude of short-term, cycle-to-cycle changes in F0. This metric is defined as the standard deviation of the first derivative of the F0 contour (np.std(np.diff(f0_values))) and is expressed in Hertz (Hz). It quantifies the absolute variability of instantaneous F0 jumps, providing a complementary view of pitch instability that focuses on the velocity of change rather than the overall dispersion of F0 values.

To ensure the robustness of this measure, several pre-processing steps were employed. The raw F0 track was first sanitized by removing any non-voiced frames and extreme outliers indicative of pitch tracking errors (e.g., octave jumps). Only segments with a sufficient number of voiced frames were retained for analysis. It is important to note that this is considered an exploratory metric. It is not a standard measure in voice analysis but is included to investigate the potential influence of fine-grained pitch dynamics on vocal expression under different conditions. Its interpretation is strictly focused on the absolute variability of F0 derivatives.

5.2.6 Normalization of the acoustic measures

To make meaningful comparisons between the speakers, we applied a normalization procedure to all acoustic parameters extracted from each speaking turn (as Chenausky et al., 2011; Lan et al., 2019; Rusz et al., 2021; Tawari et al., 2010). This procedure served two main purposes: (i) to control for baseline differences across speakers or utterances, and (ii) to ensure that differences observed between routine and emergency conditions reflected relative prosodic variation within each speaker, rather than absolute acoustic or phonetic values. The 1-sample Kolmogorov–Smirnov test did not indicate non-normally distributed acoustic features, which allows the proper application of zscore transformation.

To control for inter-individual variability in vocal baseline parameters and to eliminate differences from the recording environment, all acoustic features were normalized intra-speaker using the routine phase as a reference condition. Specifically, for each event and for each speaker involved, we computed the mean (μ_r) and standard deviation (σ_r) of each acoustic feature based solely on the speaker's speech segments recorded during the routine phase. These values were then used to compute z-scores for both routine and emergency segments, following the normalization procedure described by Disner (1980) and Rose (1987, 1991):

$$z = \frac{(x - \mu_r)}{\sigma_r}$$

where x is the observed value of a given feature in either routine or emergency speech segments, and μr and σr are the mean and standard deviation computed from the speaker's routine data only. This normalization procedure allows each speaker's emergency-related vocal change to be interpreted in terms of deviations from their own vocal baseline, effectively isolating condition-driven variability from individual-specific vocal characteristics. The resulting z-scores and averaged ratings were then Winsorized at the 10th–90th percentiles to account for the small sample size and reduce any effects of outliers.

Table 7

Acoustic parameters used in this study

Measure (Unit)	Description	Formula / Calculation Method	Primary Physiological Process
F0 Mean (Hz)	The average fundamental frequency of the voiced segments.	$\bar{F_0} = \frac{1}{N} \sum_{i=1}^{N} F_{0_i}$	Phonation (vocal fold vibration)
F0 Minimum (Hz)	The lowest fundamental frequency value.	$F_{0_{\min}} = \min({F_{0_i}}_{i=1}^{N})$	Phonation
F0 Maximum (Hz)	The highest fundamental frequency value.	$F_{0_{\max}} = \max({F_{0_i}}_{i=1}^{N})$	Phonation
F0 SD (Hz)	The standard deviation of the fundamental frequency, indicating F0 variability.	$\sigma_{F_0} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_{0_i} - \bar{F_0})^2}$	Phonation
Intensity Mean (dB)	The average sound pressure level of the segment.	$\bar{I} = \frac{1}{M} \sum_{j=1}^{M} I_j$	Respiration (subglottal pressure)

Measure (Unit)	Description	Formula / Calculation Method	Primary Physiological Process
Intensity Minimum (dB)	The minimum sound pressure level.	$I_{\min} = \min({I_j}_{j=1}^{M})$	Respiration
Intensity Maximum (dB)	The maximum sound pressure level.	$I_{\max} = \max({I_j}_{j=1}^{M})$	Respiration
Intensity SD (dB)	The standard deviation of the intensity, indicating loudness variability.	$\sigma_I = \sqrt{\frac{1}{M} \sum_{j=1}^{M} (I_j - \bar{I})^2}$	Respiration
Jitter (local, %)	The average absolute difference between consecutive periods, divided by the average period. A measure of frequency instability.	$Jitter = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{	Phonation (vocal fold biomechanics)
Shimmer (local, %)	The average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude. A measure of amplitude instability.	$Shimmer = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{	Phonation (vocal fold biomechanics)
HNR (dB)	Harmonic-to-Noise Ratio. The ratio of periodic (harmonic) to aperiodic (noise) energy in the voice.	$HNR = 10 \cdot \log_{10}\left(\frac{H}{N} \right)$	Phonation (vocal fold closure)
Duration (s)	The total duration of the speech segment (including pauses).	$T_{total} = t_{end} - t_{start}$	Prosody / Fluency
Speech Rate (syll/s)	The estimated number of syllables produced per second, including pause time.	$SR = \frac{N_{syll}}{T_{total}}$	Prosody / Articulation
Articulation Rate (syll/s)	The estimated number of syllables produced per second of actual phonation time (excluding pauses).	$AR = \frac{N_{syll}}{T_{total} - T_{pauses}}$	Articulation
Number of Pauses	The count of silent intervals within the segment that meet the duration and intensity threshold.	$N_{pauses} = \sum \mathbb{I}(I(t) < 30 \text{ dB} \cap \Delta t \geq 0.2 \text{ s})$	Fluency / Respiration
Mean Pause Duration (s)	The average length of the detected pauses.	$\bar{T}*{pause} = \frac{1}{N*{pauses}} \sum_{k=1}^{N_{pauses}} T_k$	Fluency / Respiration
F1 (Hz)	The mean frequency of the first formant, related to vowel height.	$F_1 = \frac{1}{T} \int_0^T F_1(t) dt$	Resonance (vocal tract)
F2 (Hz)	The mean frequency of the second formant, related to vowel backness.	$F_2 = \frac{1}{T} \int_0^T F_2(t) dt$	Resonance (vocal tract)

Measure (Unit)	Description	Formula / Calculation Method	Primary Physiological Process
F3 (Hz)	The mean frequency of the third formant, often related to rhoticity and voice quality.	$F_3 = \frac{1}{T} \int_0^T F_3(t) dt$	Resonance (vocal tract)
F4 (Hz)	The mean frequency of the fourth formant, related to voice quality and vocal tract length.	$F_4 = \frac{1}{T} \int_0^T F_4(t) dt$	Resonance (vocal tract)
VTL (cm)	Vocal Tract Length estimate, derived from the first formant.	$VTL = \frac{35000}{4 \cdot F_1}$	Resonance (vocal tract anatomy)
RMS Energy	Root Mean Square energy, a measure of the overall power of the acoustic signal.	$RMS = \sqrt{\frac{1}{L} \sum_{k=1}^{L} s[k]^2}$	Respiration / Phonation
Spectral Centroid (Hz)	The "center of gravity" of the spectrum; higher values indicate more high-frequency energy.	$\bar{\omega} = \frac{\sum_{m=1}^{M} \omega_m \cdot	Resonance (vocal tract filter)
Delta Formant (Hz)	The range between the highest and lowest measured formant frequency (F1-F4).	$\Delta F = \max({F_n}_{n=1}^4) - \min({F_n}_{n=1}^4)$	Resonance (vocal tract)
DI (Hz)	Formant Dispersion Index. The average spacing between consecutive formants. In the code, this is miscalculated as the standard deviation of the formant values. *See note below table.*	$DI = \sqrt{\frac{1}{4} \sum_{n=1}^{4} (F_n - \bar{F})^2}$	Resonance (vocal tract length)
F0 Fluctuation (Lambda)	The variability of the *change* in F0 (the derivative of F0). Measures pitch instability.	$\Lambda = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (\Delta F_{0_i} - \bar{\Delta F_0})^2}$	Phonation (laryngeal control)

Note. The formulas represent the mathematical concepts underlying the acoustic measures. Practical implementation follows standard algorithms in speech analysis software (Praat, Parselmouth). Formant measures (F1-F4) were obtained using Burg's algorithm with a 0.01s window. Pause detection used a threshold of 30 dB and minimum duration of 0.2s. Syllable counting was based on intensity peak detection with a 40 dB threshold.

5.2.7 Data analysis

Speech data collected often comprises a large number of acoustic parameters, which can increase computational complexity and lead to overfitting. Moreover, not all features contribute to the discrimination between neutral and stress-laden speech. Irrelevant or redundant variables can introduce noise, reduce model interpretability, and impair classification performance (Reem et al., 2024). To address these issues, we adopted a two-step analytical strategy involving feature selection followed by

dimensionality reduction. In the first step, we conducted a series of ANOVAs separately for ATCs and pilots to identify acoustic parameters that differed significantly between the routine and emergency phases. To control for the inflated risk of Type I error due to multiple comparisons, we applied an FDR correction using the Benjamini-Hochberg procedure. Only features with an adjusted p-value below or equal to 0.05 were retained for further analysis. Next, we assessed multicollinearity among the retained variables by computing pairwise Pearson correlations. In line with established practice, when the absolute correlation coefficient between two features exceeded .70, one of the two was excluded. The decision was based on the theoretical relevance of this vocal feature as a measure of stress. The feature selection process was conducted using a two-stage approach to ensure both robustness and efficiency. Subsequently, the Boruta algorithm was employed on the pre-filtered feature subset. To evaluate the generalizability of the identified acoustic markers across events, we implemented a set of classification algorithms that capture both linear and non-linear decision boundaries. Specifically: Linear Discriminant Analysis (LDA), penalized logistic regression with LASSO regularization, Random Forest (RF) and gradient boosting (XGBoost). Model performance was evaluated using LOO-CV.

5.2.8 Classifier choice

To evaluate the generalizability of the identified acoustic markers across events, we implemented a diverse set of classification algorithms that capture both linear and non-linear decision boundaries. Linear Discriminant Analysis (LDA) was included as a simple, interpretable baseline model that provides insights into linear separability of routine versus emergency speech. In parallel, we applied penalized logistic regression with LASSO regularization, which is well-suited for acoustic data and enables feature selection by shrinking irrelevant coefficients toward zero. To account for potential non-linear relationships between features, we incorporated ensemble-based methods such as Random Forest (RF) and gradient boosting (XGBoost). RF is robust to noise and provides measures of feature importance, while XGBoost leverages boosting to improve classification performance on imbalanced or heterogeneous data. This combination of classifiers was chosen to ensure complementary perspectives: interpretable linear models (LDA, LASSO) for transparency and validation of acoustic effects, and more flexible non-linear models (RF, XGBoost) to assess whether complex interactions among acoustic features contribute to stress detection. Model performance was evaluated using a Leave-One-Event-Out Cross-Validation (LOEO-CV) scheme, which tests the ability of classifiers to generalize to entirely unseen events, thus reflecting the real-world requirement that a reliable biomarker of stress should remain robust across speakers, communication channels, and other specific conditions.

5.3 Results

5.3.1 Features selection: *ANOVA's results*

Descriptive statistics with row data are presented in Table 8. Spectral tilt was initially extracted, but data inspection revealed the presence of extreme outlier values (~-4 dB to ~0 dB), attributable to the involuntary inclusion of non-vocalic segments in the automatic calculation. To avoid confounding the effect of phonetic composition with physiological effect, this parameter was excluded from subsequent analyses. CPP was initially estimated through a simplified cepstral peak proxy, because Praat's CPPS extraction did not provide stable outputs on the current dataset. However, this proxy was highly sensitive to channel noise and the presence of unvoiced material, and its absolute values were not directly comparable to CPPS values reported in the literature. For these reasons, CPP was also excluded from subsequent analyses.

Since the assumption tests indicated heteroscedasticity and non-normality, we used Welch ANOVA to compare Routine vs. Emergency. The p-values were corrected for FDR (FDR-adjusted $p < .05$). A total of 10 acoustic parameters significantly differed between the two phases for ATCOs (Table 9). For ATCO, the emergency phase was associated with greater mean intensity, $F(1, 551) = 9.92$, $p = .002$, $\eta p^2 = .018$, and increased variability of intensity, $F(1, 551) = 10.06$, $p = .002$, $\eta p^2 = .018$. Significant differences also emerged for HNR, $F(1, 551) = 9.18$, $p = .003$, $\eta p^2 = .016$, maximum F0, $F(1, 551) = 8.60$, $p = .004$, $\eta p^2 = .015$, speech rate, $F(1, 551) = 8.16$, $p = .005$, $\eta p^2 = .015$, and F3, $F(1, 551) = 7.60$, $p = .006$, $\eta p^2 = .014$. Further effects were found for F0 fluctuation, $F(1, 551) = 6.68$, $p = .010$, $\eta p^2 = .012$, F1, $F(1, 551) = 6.51$, $p = .011$, $\eta p^2 = .012$, F0 standard deviation, $F(1, 551) = 5.96$, $p = .015$, $\eta p^2 = .011$, and spectral dispersion (Df), $F(1, 551) = 5.35$, $p = .021$, $\eta p^2 = .010$. For pilots, the emergency condition produced even stronger acoustic changes. Jitter was significantly higher, $F(1, 514) = 23.15$, $p < .001$, $\eta p^2 = .043$, and significant effects were further observed for minimum F0, $F(1, 514) = 15.35$, $p < .001$, $\eta p^2 = .029$, maximum F0, $F(1, 514) = 15.81$, $p < .001$, $\eta p^2 = .030$, mean F0, $F(1, 514) = 13.20$, $p < .001$, $\eta p^2 = .025$, and F0 fluctuation, $F(1, 514) = 13.13$, $p < .001$, $\eta p^2 = .025$. In addition, significant decreases were found for HNR, $F(1, 514) = 10.02$, $p = .002$, $\eta p^2 = .019$, while articulation rate, $F(1, 514) = 9.27$, $p = .003$, $\eta p^2 = .018$, and speech rate, $F(1, 514) = 6.11$, $p = .014$, $\eta p^2 = .012$, both increased.

A significant difference was also observed for F3, $F(1, 514) = 5.80$, $p = .016$, $\eta p^2 = .011$. After computing correlation matrices separately for each speaker group, we identified pairs of highly correlated variables ($r > .80$) (Table 6). For ATCO, no pairs of variables exceeded the $|r| \geq .80$ cut-off, indicating no strong redundancy among predictors. In contrast, several highly correlated pairs emerged for pilots: ZW_Intensity_mean_dB and ZW_RMS_Energy ($r = .99$), from which we retained ZW_Intensity_mean_dB; ZW_Articulation_rate_syll_s and ZW_Speech_rate_syll_s ($r = .98$), for which we retained ZW_Articulation_rate_syll_s; and the negative correlation between ZW_F3_Hz and energy-related measures ($r = –.88/–.90$), where we retained ZW_F3_Hz.

Table 8

Descriptive Statistics for Acoustic Measures by Speaker and Phase

Measure	Speaker	Routine	Emergency
F0 Mean (Hz)	ATCO	135.24 (31.20)	145.11 (35.85)
	Pilot	124.51 (22.74)	142.24 (36.57)
F0 Minimum (Hz)	ATCO	95.76 (21.95)	100.24 (24.79)
	Pilot	91.28 (18.05)	100.09 (26.57)
F0 Maximum (Hz)	ATCO	193.18 (57.86)	208.11 (56.28)
	Pilot	171.66 (47.28)	202.41 (57.11)
F0 *SD* (Hz)	ATCO	19.06 (10.72)	21.61 (11.89)
	Pilot	17.93 (10.70)	22.52 (12.55)
Intensity Mean (dB)	ATCO	67.59 (6.35)	67.46 (7.13)
	Pilot	64.99 (7.21)	66.89 (7.40)
Intensity Minimum (dB)	ATCO	39.25 (14.27)	40.65 (14.19)
	Pilot	40.10 (14.01)	40.84 (14.55)
Intensity Maximum (dB)	ATCO	81.16 (5.37)	80.66 (6.19)
	Pilot	80.69 (5.79)	81.24 (6.07)
Intensity *SD* (dB)	ATCO	10.54 (4.12)	10.66 (3.97)
	Pilot	10.63 (4.65)	10.65 (4.13)
Jitter (local, %)	ATCO	1.95 (0.60)	1.87 (0.66)
	Pilot	1.97 (0.98)	1.83 (0.71)
Shimmer (local, %)	ATCO	12.16 (2.65)	12.44 (2.35)

	Pilot	12.23 (3.14)	12.40 (2.75)
HNR (dB)	ATCO	8.11 (3.70)	7.85 (2.91)
	Pilot	7.78 (3.58)	7.24 (2.98)
Duration (s)	ATCO	4.44 (2.59)	5.20 (3.84)
	Pilot	3.45 (1.92)	4.84 (3.32)
Speech Rate (syll/s)	ATCO	12.25 (2.70)	12.29 (2.57)
	Pilot	12.37 (3.82)	12.75 (2.61)
Articulation Rate (syll/s)	ATCO	12.50 (2.68)	12.46 (2.53)
	Pilot	12.35 (2.86)	12.98 (2.54)
Number of Pauses	ATCO	0.04 (0.20)	0.12 (0.42)
	Pilot	0.06 (0.25)	0.12 (0.39)
Mean Pause Duration (s)	ATCO	0.75 (0.93)	0.79(1.42)
	Pilot	0.39 (0.24)	1.12 (2.03)
F1 (Hz)	ATCO	699.50 (117.59)	708.74 (137.28)
	Pilot	740.65 (118.52)	742.21 (113.34)
F2 (Hz)	ATCO	1506.75 (166.56)	1531.98 (195.12)
	Pilot	1517.24 (168.39)	1494.60 (170.38)
F3 (Hz)	ATCO	2249.29 (231.14)	2290.94 (264.12)
	Pilot	2253.37 (245.62)	2234.26 (258.31)
F4 (Hz)	ATCO	2999.84 (305.37)	3047.32 (339.33)
	Pilot	3008.36 (334.97)	3001.77 (352.95)
VTL (cm)	ATCO	12.78 (1.78)	12.71 (1.88)

	Pilot	12.12 (1.73)	12.05 (1.59)
RMS Energy	ATCO	0.13 (0.07)	0.13 (0.08)
	Pilot	0.11 (0.07)	0.12 (0.07)
Spectral Centroid (Hz)	ATCO	1118.51 (167.94)	1158.08 (257.85)
	Pilot	1163.10 (230.03)	1173.35 (218.36)
Delta Formant (Hz)	ATCO	2296.35 (247.75)	2333.82 (273.28)
	Pilot	2264.67 (281.24)	2261.92 (297.92)
DI (Hz)	ATCO	854.55 (92.04)	868.54 (102.05)
	Pilot	843.15 (104.90)	841.58 (109.70)
F0 Fluctuation	ATCO	9.29 (6.07)	9.47 (5.85)
	Pilot	7.86 (6.31)	8.98 (6.59)

Note. Values are reported as Mean (Standard Deviation in parentheses)

Table 9

Results of ANOVA for Acoustic Variables for ATCO

Variable	***F (1,424)***	***p***	**η^2p**	***p*FDR**
ZW_Intensity_std_dB	10.055	.002	.018	**.012***
ZW_Intensity_mean_dB	9.920	.002	.018	**.012***
ZW_HNR_dB	9.179	.003	.016	**.015***
ZW_F0_max_Hz	8.597	.004	.015	**.016***
ZW_Speech_rate_syll_s	8.156	.005	.015	**.018***

Variable	*F (1,424)*	*p*	η^2p	*p*FDR
ZW_F3_Hz	7.602	.006	.014	**.021***
ZW_F0_fluctuation	6.680	.010	.012	**.031***
ZW_F1_Hz	6.512	.011	.012	**.031***
ZW_F0_std_Hz	5.962	.015	.011	**.038***
ZW_Df_Hz	5.346	.021	.010	**.050***
ZW_F0_min_Hz	4.631	.032	.008	.064 †
ZW_F4_Hz	4.525	.034	.008	.064 †
ZW_Intensity_max_dB	3.160	.076	.006	.133 †
ZW_F0_mean_Hz	1.954	.163	.004	.268
ZW_Shimmer_local_	1.752	.186	.003	.290
ZW_Spectral_Centroid_Hz	0.984	.322	.002	.474
ZW_RMS_Energy	0.864	.353	.002	.494
ZW_F2_Hz	0.759	.384	.001	.512
ZW_Articulation_rate_syll_s	0.580	.447	.001	.569
ZW_Jitter_local_	0.405	.525	< .001	.621
ZW_Mean duration pauses s	0.395	.533	< .001	.621
ZW_Pauses numbers	0.270	.605	< .001	.678
ZW_Delta_F_Hz	0.130	.719	< .001	.774
ZW_Intensity_min_dB	0.051	.822	< .001	.852
ZW_VTL_cm	0.005	.942	< .001	.942

Note. Significant *p*-values are highlighted in bold (**$p < .001$, *$p < .01$, $p < .05$). Trend-level effects are indicated with a dagger (†, $p < .10$).

Table 10

Results of ANOVA for Acoustic Variables for Pilots

Variable	*F*	*p*	η^2p	*p*FDR
ZW_Jitter_local_	23.154	< .001	.043	**< .001****
ZW_F0_max_Hz	15.813	< .001	.030	**< .001****
ZW_F0_min_Hz	15.352	< .001	.029	**< .001****
ZW_F0_mean_Hz	13.199	< .001	.025	**.001****
ZW_F0_fluctuation	13.130	< .001	.025	**.001****
ZW_HNR_dB	10.015	.002	.019	**.007****
ZW_Articulation_rate_syll_s	9.267	.002	.018	**.009****
ZW_Speech_rate_syll_s	6.110	.014	.012	**.043***
ZW_F3_Hz	5.803	.016	.011	**.046***
ZW_RMS_Energy	5.231	.023	.010	.057 †
ZW_Intensity_max_dB	5.111	.024	.010	.057 †
ZW_F4_Hz	4.683	.031	.009	.067 †
ZW_Intensity_mean_dB	3.946	.048	.008	.089 †
ZW_Spectral_Centroid_Hz	3.426	.065	.007	.104
ZW_F0_std_Hz	3.369	.067	.007	.104
ZW_Delta_F_Hz	3.551	.060	.007	.104

Variable	*F*	*p*	η²p	*p*FDR
ZW_Shimmer_local_	3.102	.079	.006	.116
ZW_VTL_cm	2.496	.115	.005	.160
ZW_Intensity_min_dB	2.433	.120	.005	.160
ZW_F1_Hz	2.340	.127	.005	.161
ZW_mean_duration_pauses_s	1.758	.189	.003	.231
ZW_Intensity_std_dB	0.248	.619	< .001	.693
ZW_Df_Hz	0.026	.873	< .001	.925
ZW_F2_Hz	0.019	.892	< .001	.925
ZW_Pauses number	0.008	.927	< .001	.927

Note. Significant p-values are highlighted in bold (**p < .001, *p < .01, p < .05). Trend-level effects are indicated with a dagger (†, p < .10).

Table 11

Highly correlated acoustic parameter pairs (r > .70) in pilot groups

Group	Variable 1	Variable 2	*r*
Pilot	ZW Intensity mean dB	ZW RMS Energy	.99
Pilot	ZW Articulation rate syll/s	ZW Speech rate syll/s	.98
Pilot	ZW F3 Hz	ZW Intensity mean dB	–.90
Pilot	ZW F3 Hz	ZW RMS Energy	–.88

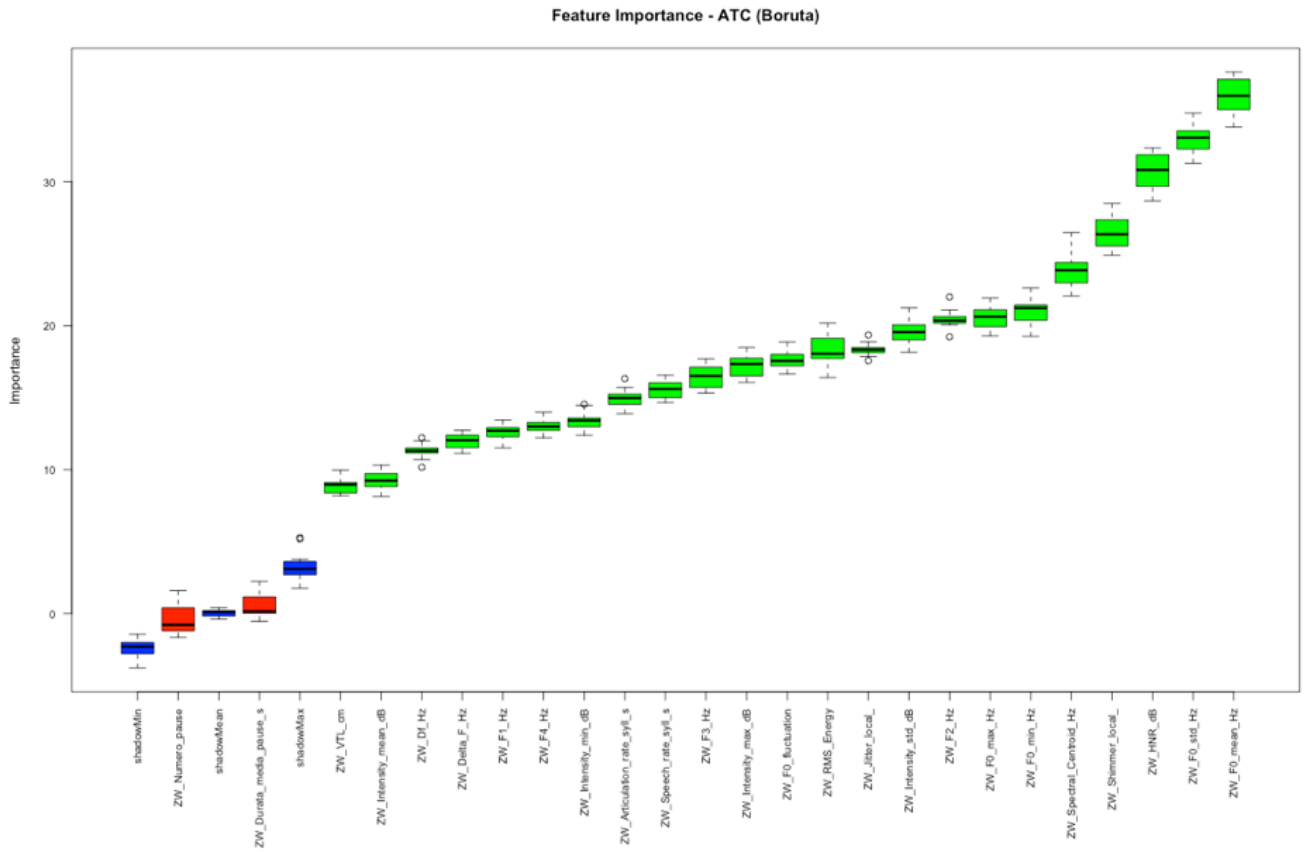
Note. r = Pearson correlation coefficient. Only pairs with |r| > .80 are reported.

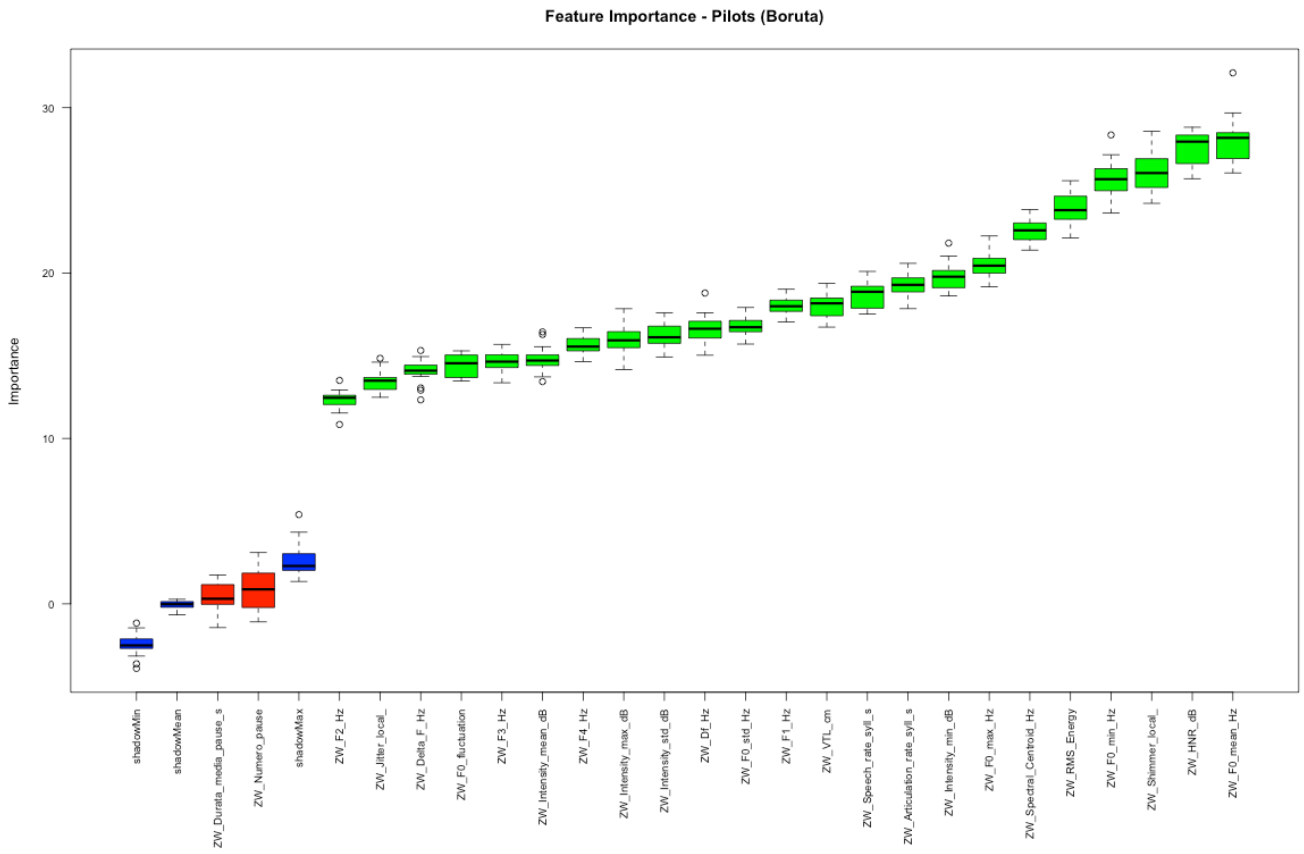
5.3.2 Features selection: BORUTA's results

To identify the most relevant acoustic predictors, we applied the Boruta feature selection algorithm (Kursa & Rudnicki, 2010) separately for ATCO and pilot data. The algorithm was run on the ZW-normalized features with Phase (Routine vs. Emergency) as the target. We used 100 iterations and the default Random Forest–based importance measure. Boruta confirmed 26 features as relevant for both ATCO and pilot datasets, while 2 features (Number of pauses, Mean pause duration) were consistently rejected. No features remained in the tentative zone after the iterations.

Figure 8

Feature importance results from the Boruta algorithm applied to ATC and pilot datasets





Note. Green boxplots represent features confirmed as important for distinguishing between routine and emergency speech, while red boxplots indicate features rejected by the algorithm. All acoustic features were confirmed as relevant, with the exception of ZW_ ZW_Pauses number and ZW_Mean duration pauses s, which were consistently rejected across both groups.

Table 11

Intersection between ANOVA and Boruta

Feature	ATCO ANOVA (FDR < .05)	ATCO Boruta	Pilot ANOVA (FDR < .05)	Pilot Boruta	Intersection
F0 mean (Hz)	–	✓	✓	✓	Pilot only
F0 min (Hz)	– († trend)	✓	✓	✓	✓
F0 max (Hz)	✓	✓	✓	✓	✓
F0 SD (Hz)	✓	✓	–	✓	ATCO only
F0 fluctuation	✓	✓	✓	✓	✓
Intensity mean (dB)	✓	✓	† trend	✓	✓
Intensity SD (dB)	✓	✓	–	✓	ATCO only
Intensity max (dB)	– († trend)	✓	–	✓	ATCO only

Feature	ATCO ANOVA (FDR < .05)	ATCO Boruta	Pilot ANOVA (FDR < .05)	Pilot Boruta	Intersection
Intensity min (dB)	–	✓	–	✓	–
HNR (dB)	✓	✓	✓	✓	✓
Jitter (%)	–	✓	✓	✓	Pilot only
Shimmer (%)	–	✓	–	✓	–
Speech rate (syll/s)	✓	✓	✓	✓	✓
Articulation rate (syll/s)	–	✓	✓	✓	Pilot only
Number of pauses	–	✗	–	✗	–
Mean pause duration	–	✗	–	✗	–
F1 (Hz)	✓	✓	–	✓	ATCO only
F2 (Hz)	–	✓	–	✓	–
F3 (Hz)	✓	✓	✓	✓	✓
F4 (Hz)	–	✓	–	✓	–
VTL (cm)	–	✓	–	✓	–
RMS Energy	–	✓	† trend	✓	–
Spectral Centroid (Hz)	–	✓	–	✓	–
Delta Formant (Hz)	–	✓	–	✓	–
Dispersion Index (Df)	✓	✓	–	✓	ATCO only

Note. confirmed/significant; – = not significant; † trend-level ($p < .10$); ✗ = rejected by Boruta. Intersection = variables confirmed by both ANOVA and Boruta in the same group.

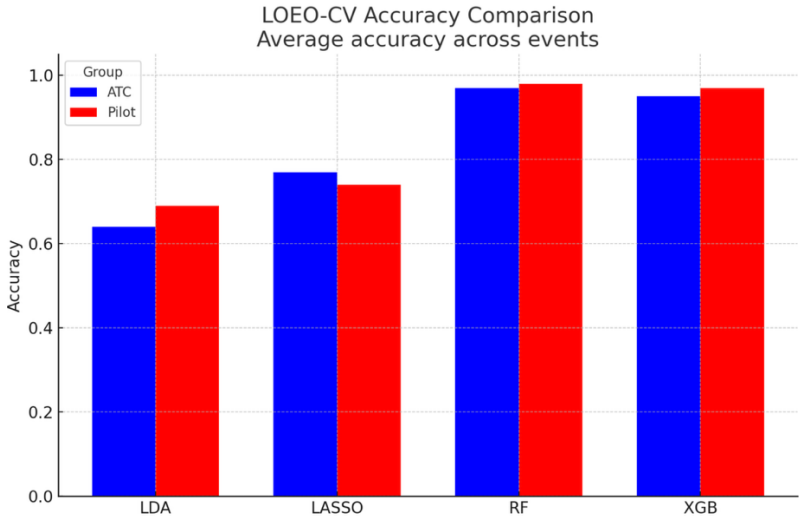
5.3.3 Classifier performance

To evaluate the generalizability of the acoustic markers, we compared the four classification algorithms using the feature set obtained from the intersection of ANOVA and Boruta results. For ATC speech, LDA achieved modest performance (Accuracy = .64, Balanced Accuracy = .64, F1 = .82), while LASSO yielded higher accuracy (.77) and perfect balanced accuracy (1.00), with an F1-score of 1.00, suggesting that a small subset of features consistently distinguished emergency from routine speech. Ensemble methods performed best: RF reached an accuracy of .97 (Balanced Accuracy = .97, F1 = .99), and XGB achieved .95 accuracy with perfect balanced accuracy (1.00) and F1 = 1.00. For pilot speech, LDA showed slightly higher performance compared to ATCs (Accuracy = .69, Balanced Accuracy = .69, F1 = .82). LASSO again improved performance (Accuracy = .74, Balanced Accuracy = .94, F1 =

1.00). RF and XGB achieved near-perfect discrimination, with RF yielding .987 accuracy (Balanced Accuracy = .97, F1 = .996) and XGB reaching .96 accuracy with perfect balanced accuracy (1.00) and F1 = 1.00. Overall, while linear models provided interpretable baselines, ensemble classifiers demonstrated almost perfect generalizability across unseen events, highlighting the robustness of the identified acoustic stress markers.

Figure 9

Classification accuracy of Random Forest and XGBoost models for ATCOs and pilots (Leave-One-Event-Out Cross-Validation)



LDA and Lasso

LASSO regression identified distinct sets of acoustic markers for ATCOs and pilots (Table 12). For ATCOs, the most stable predictors were jitter and HNR variability, as well as the variability of minimum F0 and intensity. Emergency speech was characterized by higher jitter median values, whereas routine speech showed greater variability in jitter, HNR, and pitch. For pilots, LASSO highlighted a broader range of features. Variability in F0, jitter, HNR, articulation rate, and intensity consistently predicted routine speech, while higher jitter mean and HNR median values were associated with emergency speech.

For ATCOs, LDA consistently assigned the highest weights to articulation rate and speech rate variability followed by intensity measures and micro-acoustic features such as jitter and HNR. For pilots, the strongest predictors were speech rate and articulation rate (median and mean values), together with intensity measures .Pitch variability and jitter contributed as secondary cues. Overall, ATCOs were mainly differentiated by rhythm, intensity, and vocal stability (jitter/HNR), whereas pilots were

distinguished primarily by prosodic and intensity-related features, with micro-acoustic measures playing a smaller role.

Table 12

Top Acoustic Features Selected by LASSO Regression and LDA for ATCOs and Pilots

Feature	ATCO – LASSO (Freq / Coef)	Pilot – LASSO (Freq / Coef)	ATCO – LDA (Mean Weight)	Pilot – LDA (Mean Weight)
ZW_Jitter_local__median	90 / 0.339	–	0.97	–
ZW_Jitter_local__iqr	90 / -0.332	89 / -0.301	0.72	–
ZW_HNR_dB_iqr	90 / -0.275	90 / -0.331	–	–
ZW_F0_min_Hz_sd	89 / -0.433	46 / -0.153	–	–
ZW_Intensity_mean_dB_sd	89 / -0.124	46 / -0.530	–	1.01
ZW_F0_fluctuation_mean	86 / -0.185	50 / -0.191	–	–
ZW_F0_mean_Hz_median	79 / 0.011	48 / -0.024	–	–
ZW_HNR_dB_mean	75 / 0.053	–	0.71	–
ZW_Intensity_std_dB_iqr	63 / -0.034	–	–	–
ZW_F0_min_Hz_iqr	25 / -0.056	–	–	–
ZW_HNR_dB_median	22 / 0.023	52 / 0.139	–	–
ZW_F0_fluctuation_median	16 / -0.049	–	–	–
ZW_Jitter_local__mean	5 / -0.078	73 / 0.180	–	–
ZW_Articulation_rate_syll_s_median	3 / -0.008	48 / -0.487	–	1.42
ZW_Intensity_mean_dB_iqr	2 / -0.030	74 / -0.177	–	–
ZW_F0_max_Hz_mean	–	79 / -0.067	–	–
ZW_Speech_rate_syll_s_sd	–	71 / -0.147	4.51	0.74
ZW_F1_Hz_iqr	–	65 / -0.021	–	–
ZW_Articulation_rate_syll_s_sd	–	–	4.65	–
ZW_Articulation_rate_syll_s_mean	–	–	2.12	1.36
ZW_Articulation_rate_syll_s_iqr	–	–	1.60	–
ZW_Speech_rate_syll_s_median	–	–	1.34	2.04
ZW_Speech_rate_syll_s_iqr	–	–	1.24	0.81
ZW_Intensity_std_dB_sd	–	–	1.15	0.84
ZW_Intensity_mean_dB_mean	–	–	1.13	0.94
ZW_F0_std_Hz_sd	–	–	0.94	–
ZW_Jitter_local__sd	–	–	0.86	0.91
ZW_F3_Hz_sd	–	–	0.78	–

Feature	ATCO – LASSO (Freq / Coef)	Pilot – LASSO (Freq / Coef)	ATCO – LDA (Mean Weight)	Pilot – LDA (Mean Weight)
ZW_F0_max_Hz_iqr	–	47 / -0.131	–	0.81
ZW_Df_Hz_sd	–	–	–	1.17
ZW_Intensity_mean_dB_median	–	–	0.83	1.50
ZW_Intensity_std_dB_mean	–	–	–	0.78
ZW_Df_Hz_iqr	–	–	–	0.79

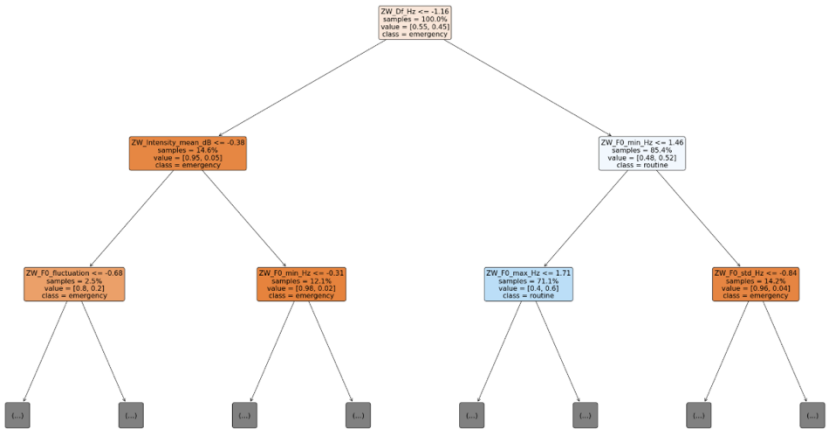
Note. LASSO: values represent selection frequency (out of 90 folds) and mean coefficient. LDA: values represent mean importance weight. Positive coefficients indicate higher feature values under emergency conditions; negative coefficients indicate lower values. sd = standard deviation; iqr = interquartile range.

Random Forest and XGBoost

Ensemble classifiers, namely Random Forest and XGBoost, provided the strongest performance across both ATCO and pilot speech, achieving near-perfect discrimination between emergency and routine communications. To better illustrate how these models operate, Figure 10 shows an example of a single decision tree extracted from the Random Forest. This visualization highlights how specific acoustic variables and their thresholds contributed to classification at the node level, although the overall RF prediction results from the aggregation of many such trees. The relative contribution of each acoustic feature was further quantified through feature importance analyses. Figure 11 presents the ranking of acoustic stress markers for both ATCO and pilot speech combined, as derived from RF and XGB models. These rankings emphasize the consistent role of prosodic, intensity-related, and micro-acoustic variables in distinguishing emergency from routine conditions.

Figure 10

Example of a decision tree extracted from the Random Forest (maximum depth = 3)



Note. The tree illustrates how specific acoustic variables contribute to distinguishing between routine and emergency communications. Each node reports the decision threshold, the proportion of samples in each class (routine vs. emergency), and the majority class at that node. This tree represents a single classifier among the many that constitute the Random Forest. The overall model output is based on the aggregation of dozens of trees, so the rules shown here should not be interpreted as rigid diagnostic criteria, but rather as an illustrative example of the model's internal logic.

Figure 11

Random Forest and XGBoost feature importance ranking of acoustic stress markers (ATCO and Pilot speech combined)

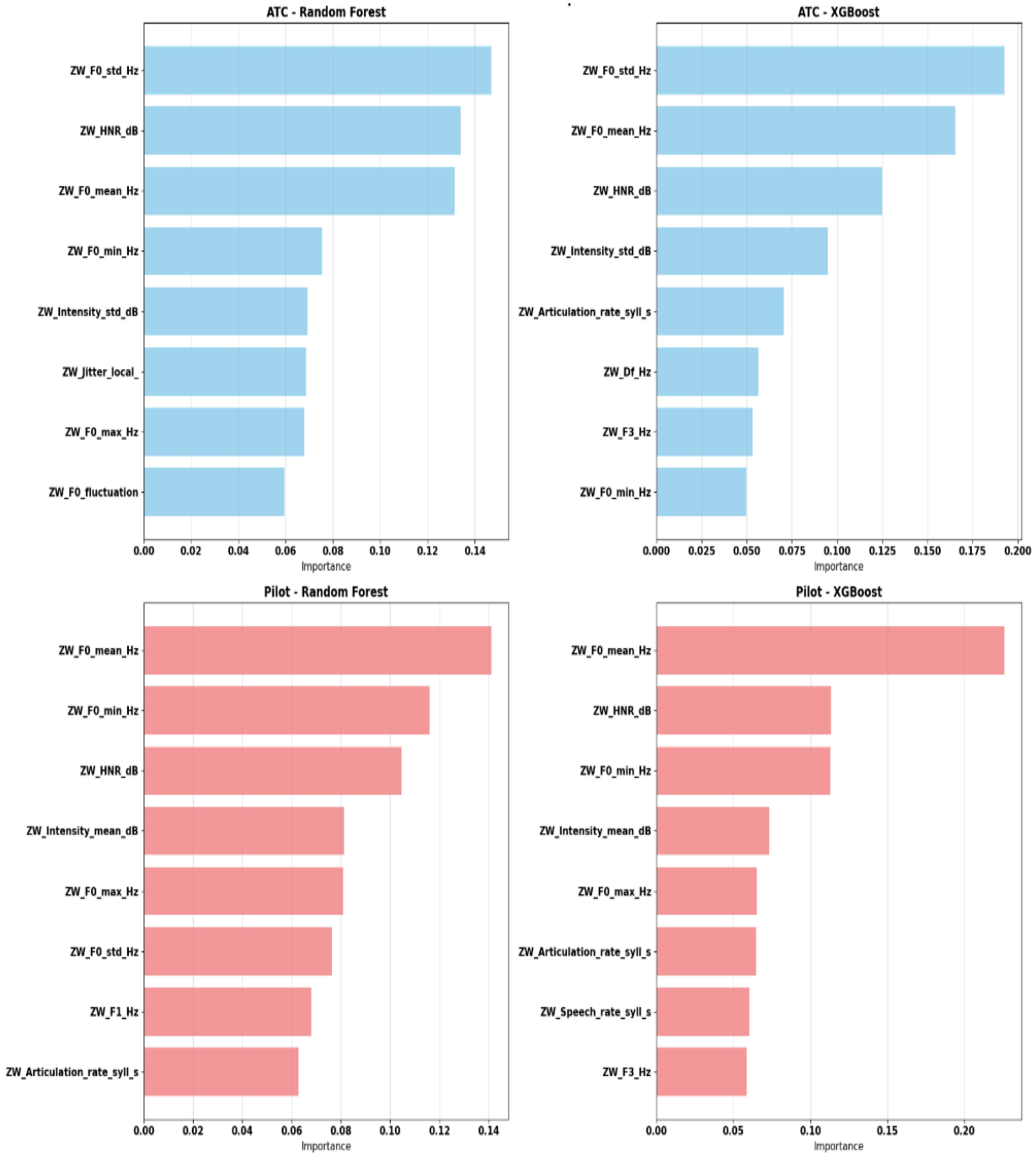


Table 13

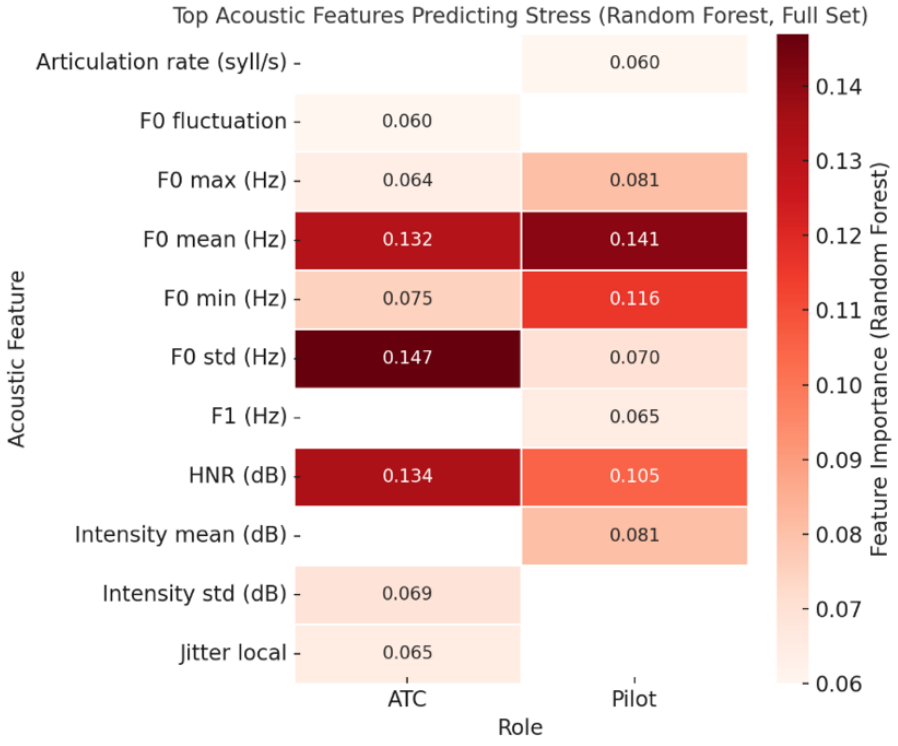
Most important features for ATCO and Pilot speech across Random Forest and XGBoost classifiers

Feature	ATCO – Random Forest	ATCO – XGBoost	Pilot – Random Forest	Pilot – XGBoost
ZW_F0_std_Hz	.1470	.1925	–	–
ZW_HNR_dB	.1339	.1250	.1047	.1133
ZW_F0_mean_Hz	.1315	.1655	.1411	.2260
ZW_F0_min_Hz	.0754	–	.1161	.1130
ZW_Intensity_std_dB	.0692	.0947	–	–
ZW_Articulation_rate_syll_s	–	.0705	–	–
ZW_Intensity_mean_dB	–	–	.0813	.0732
ZW_F0_max_Hz	–	–	.0808	.0651

Note. Values represent feature importance scores for distinguishing emergency vs. routine speech. Dashes indicate features not among the top five for the given classifier.

Figure 12

Heatmap of Random Forest coefficients of acoustic parameters by speaker role



Note. Heatmap of coefficients from Random Forest, showing the main contribution of the most important acoustic parameter to the classification between routine and emergency speech, separately for ATCs and pilots. Red indicates higher values in emergency.

5.4 Discussion

The present study is, to our knowledge, the first systematic attempt to identify acoustic stress markers in real-world aviation emergencies using a sample of pilot–ATC communications. By combining univariate statistical tests, feature selection algorithms, and cross-validated machine learning classifiers, we demonstrated that stress-related vocal changes can be reliably detected in authentic operational contexts. These findings extend prior laboratory-based research by showing that stress is mirrored by measurable imprints across multiple acoustic domains, even within the highly constrained and standardized register of aviation radiotelephony.

Across roles, two parameters emerged as core stress markers: elevation of F0 and reduction of HNR. The consistent rise in F0 under stress corroborates a long-standing body of evidence (Congleton et al., 1997; Giddens et al., 2013; Hall et al., 2021; Ruiz et al., 1996; Van Puyvelde et al., 2018), and is plausibly linked to autonomic and HPA axis activation (Hall et al., 2021). This supports the interpretation of F0 increase as a general arousal response, consistent with Selye's definition of stress as a non-specific reaction of the organism. In parallel, the observed decline in HNR suggests increased turbulence and reduced vocal fold periodicity, compatible with sympathetic activation (Giddens et al., 2013; Mendoza & Carballo, 1998). However, given mixed evidence in prior studies (Depestele, 2023), HNR should be considered a promising but context-dependent marker. Moreover, also formant measures, though not primary stress markers, captured relevant role-specific adjustments. In our study, F1 and F3 emerged as significant. F1 increased in both groups, consistent with previous observations (Hansen & Patil, 2007; Sigmund, 2006).

Beyond this shared acoustic core, our analyses revealed role-specific adaptations. Pilot speech under stress was characterized by accelerated temporal dynamics (increased speech and articulation rates), elevated F0, and reduced HNR. Together, these changes produced a higher-pitched, faster, and noisier vocal profile. Spectral measures provided additional evidence of role differentiation. In pilots, reductions in F3 were observed, consistent with previous reports linking downward shifts in F3 to a top-down self-regulation or cognitive control (Ruiz et al., 1996; vPuyvelde et al., 2018). By contrast, controllers showed upward shifts in F3. These opposing trends suggest that pilots and controllers may engage distinct regulatory strategies under stress. Importantly, changes in F3 have been documented as a primary effect of stress, sometimes increasing and sometimes decreasing, reinforcing its status as a sensitive, context-dependent marker (Gopalan, 2021; Hansen & Patil, 2007; Ruiz et al., 1996; Sondhi et al., 2015; Waaramaa, 2006). Complementing this, the Dispersion Index (Df) increased significantly

in controllers but not in pilots, indicating selective spectral adjustments among ATCOs aimed at maintaining communicative clarity under load. For controllers, stress effects followed a different pattern. Rather than a uniform rise in F0, controllers displayed greater F0 variability, along with significant increases in mean intensity and intensity variability. These findings suggest a more unstable prosodic profile, with modulation of vocal energy serving as a compensatory strategy under cognitive load. Spectral changes also differentiated controllers: F3 shifted upward, and dispersion indices increased, consistent with greater regulatory effort to probably maintain communicative clarity. Temporal dynamics were less uniform than in pilots: speech rate increased significantly, but articulation rate effects remained inconsistent.

Moreover, perturbation measures added further nuance. Jitter was significant in ANOVA for pilots and in Random Forest models for controllers, suggesting that stress-related irregularities in cycle-to-cycle frequency perturbations may manifest differently by role. Shimmer, by contrast, showed no systematic modulation, corroborating prior null findings (Kappen et al., 2024; Pisanski et al., 2021; Tavi, 2017). Pause-related features did not survive feature selection, although descriptive tendencies toward longer pauses replicate earlier evidence that hesitation may increase under stress (Buchanan et al., 2014). The observed role dependence is theoretically consistent with evidence that pilots and controllers operate under different task demands and stressors, which shape distinct vocal adaptations through differential cognitive workload, communicative goals, and regulatory strategies (Morris & Leung, 2006; Prinzo & Britton, 1993; Van Puyvelde et al., 2018).

In relation to our hypotheses, the evidence can be summarized as follows.

- H1 was partially supported: F0 elevation and HNR reduction confirmed the expected stress-related shifts, but intensity and perturbation measures were less consistent than predicted.
- H2 was partially supported: role-specific differences were evident, but not in the anticipated direction. Pilots accelerated rather than slowed articulation, and did not increase pausing, while controllers displayed prosodic instability and spectral modulation rather than the hypothesized stable, automatized profile. These unexpected adaptations suggest more dynamic regulatory strategies than originally hypothesized.

5.4.1 Linear vs. Non- Models

A central methodological contribution of this study lies in the comparative evaluation of linear and non-linear classifiers for detecting stress-related vocal changes. Linear approaches (ANOVA, LASSO, LDA) provided interpretable baselines and consistently highlighted local perturbations and temporal dynamics as discriminative cues. These findings could suggest that linear techniques are particularly sensitive to micro-instabilities in voice production and timing short-term perturbations that

align well with linear separability but may not capture the full physiological signature of stress. By contrast, ensemble-based models (Random Forest, XGBoost) consistently achieved high discrimination across unseen events, underscoring the added value of capturing non-linear interactions among features. In these models, the most systematic and physiologically grounded stress markers, F0 elevation (mean, max, min, and variability) and reductions in HNR, emerged as top predictors across both roles. Such findings align with prior evidence that non-linear models are better suited to capture complex and dynamic relationships in speech (Barker & Berthommier, 1999; Holambe & Deshpande, 2012), and underscore the importance of hybrid modeling strategies in applied stress monitoring.

5.4.2 Limitations and strengths

As highlighted by recent work (Kappen et al., 2024), research should further test the robustness of these markers in increasingly naturalistic contexts, including freely spoken speech and noisy environments. In the present study, we embraced this challenge with all its advantages and limitations. A key strength of this study lies in its ecological validity: unlike the majority of prior research conducted in laboratory settings, our data capture stress in genuine operational emergencies, where the stakes are real. At the same time, this ecological approach introduces limitations. The quality of the recordings was not optimal, and the material was not phonetically balanced. Several parameters showed extreme outliers and could not be retained. Moreover, while we correlated the emergency event (the stressor) with acoustic changes, we cannot directly and objectively measure “stress”. No physiological or self-report measures of stress were available; rather, stress was inferred from the situational antecedent. We also considered different types of stress collectively. Some of the observed changes may therefore reflect heightened communicative workload or emotional arousal rather than stress per se. In this study, stress was therefore operationalized broadly, encompassing cognitive, emotional, and workload-related demands. Importantly, the contextual and situational information available for each event was partial and limited to what could be retrieved from the recordings and public sources. As a result, several potentially relevant confounds, such as finer grained phase of flight characteristics, traffic density, weather, concurrent cockpit workload, and other operational constraints, could not be systematically modeled. Future studies, where feasible, should integrate richer contextual and situational metadata to better isolate stress related vocal change from co occurring operational demands and recording conditions.

5.4.3 Applied and theoretical implications

These findings carry both applied and theoretical implications. From an applied perspective, the identification of robust and role-specific acoustic stress markers opens the possibility of developing real-time monitoring systems capable of detecting acute stress in pilots and ATCOs during operations.

Such systems could be integrated within cockpits or control towers, where vocal signals are already continuously available, and combined with other non-invasive indicators such as heart rate variability or eye-tracking to provide multimodal, context-sensitive assessments of operator state. From a theoretical standpoint, the consistent emergence of F0 elevation and HNR reduction reinforces their role as core physiological stress responses, reflecting autonomic activation under acute load. At the same time, the role-dependent patterns highlight how the voice is not merely a passive physiological signal, but also a medium shaped by cognitive strategies and communicative demands. This dual perspective underscores the need to conceptualize vocal stress markers as situated at the intersection of physiology and pragmatics, where universal arousal responses are dynamically modulated by a lot of factors, as personality traits, task-specific roles and communicative goals.

5.5 Conclusion

Taken together, our findings indicate that while some markers (F0-related features and HNR) emerge as robust and shared across roles, others display role-specific patterns, underscoring the importance of considering operational context when identifying vocal biomarkers of stress. Future research should acknowledge individual variability. Not all speakers exposed to the same stressor show identical physiological or vocal responses. Speech alterations may derive from both involuntary physiological reactions and deliberate regulatory strategies. Factors such as coping style, professional role, position, and training shape how stress is vocally manifested.

Moreover, speakers may intentionally mask or suppress emotional states, further complicating the detection of stress in real-world contexts (Schewski et al., 2025). From an applied perspective, this convergence is not unequivocally encouraging. F0, while highly sensitive, is also a non-specific index, influenced by multiple physiological and contextual factors, and therefore cannot provide a fully reliable or disentangled measure of stress. HNR, though robust in our dataset, has shown inconsistent patterns in the literature, with both increases and decreases depending on context.

We argue that speech parameters should not be interpreted in isolation. Rather, they interact in a networked manner, shaping complex and multidimensional stress signatures. In this regard, the pronounced role of F0 in our results, while compelling, is not completely sufficient on its own; more integrative methodological approaches may be required to capture the full complexity of vocal responses under pressure. In this sense, the value of our findings lies less in the identification of single “key” markers than in what they reveal about the limits of reductionism itself: stress leaves traces in the voice, but these traces are emergent properties of a dynamic systems emergent properties of a dynamic system that remains inherently difficult to quantify. Recognizing this complexity is not a weakness of the model, it is its next frontier.

CHAPTER 6

Shape Matters: A Morphometric Approach to Speech Under Stress in Aviation Emergencies

Abstract

Introduction: Geometric morphometrics (GM) quantifies the shape of biological and behavioral structures yet is rarely applied to human speech. This case study tests whether GM captures stress-linked deformation in aviation communications.

Method: We analyzed four tokens of the utterance "American 1779" produced by a pilot and an air-traffic controller (ATCO) in emergency and routine phases. Speech was converted to mono, resampled, duration-normalized, and converted to spectrograms. In a full-surface GM pipeline, spectrograms were resampled to a semi-landmark grid and submitted to PCA. A ridge-only pipeline modelled time-normalized F0 contours as pseudo-landmarks. For comparison, we ran a PCA of conventional features (mean/SD F0, spectral centroid, RMS, intensity mean/SD). Moreover, perceptual data were collected from 57 listeners who rated the perceived level of stress for the four tokens on a 0–100 VAS.

Results: Full-surface GM separated tokens by speaker and phase: PC1 indexed global spectral height/energy shifts, PC2 captured changes in spectral prominence; the pilot's emergency token showed extreme positive PC1/PC2 scores. Ridge-only GM revealed marked flattening and reduced modulation of the pilot's emergency F0, with smaller phase shifts for the ATCO. The conventional-feature PCA primarily segregated speakers and only weakly reflected phase. In the perceptual task, 57 listeners consistently judged the ATCO as more stressed than the pilot; notably, the pilot in emergency was perceived as even less stressed than in routine, plausibly reflecting top-down regulatory control. Full-surface GM showed the strongest correspondence with perceptual distances, standard features exhibited substantial yet incomplete alignment, and ridge-only GM inverted perceived proximities. Treating speech as a geometric object preserves spectro-temporal structure and exposes stress-related deformations that scalar features obscure.

Conclusions: Although limited to a single phrase and few tokens, the results indicate that GM is a feasible and interpretable complement to standard acoustics for modeling speech under operational stress in aviation.

Keywords: Morphometrics; Vocal Markers; Vocal Analysis; Aviation; Stress

6.1 Introduction

The analysis of morphological variability has a long tradition in anthropology and evolutionary and developmental biology (Adams et al., 2004; Rohlf, 1990). Within this tradition, geometric morphometrics (GM) provides a rigorous framework for quantifying and comparing forms while preserving their spatial structure (Adams et al., 2004; Slice, 2007), with well-established applications to anatomical structures such as skulls, wings, and bones. More recently, GM principles have been extended to acoustic phenomena—particularly in non-human species—by representing signals as geometric surfaces (e.g., spectrograms) or trajectories (e.g., pitch contours) amenable to morphometric analysis (MacLeod et al., 2013). The core idea is to construe sound as a structured "shape," enabling quantitative, visually interpretable comparisons across species and ecological contexts.

Despite these advances, applications of GM to human vocalizations remain scarce, with only two studies to date applying GM only to prosodic contours (Knoll & Costall, 2015; van Rijn et al., 2023). Research on speech and prosody typically emphasizes isolated acoustic parameters (e.g., mean F0, intensity, spectral centroid) rather than the global configuration of the signal, leaving a methodological gap whereby systematic deformations of vocal signals—such as those elicited by stress—may be missed by parameter-by-parameter approaches.

A principal obstacle to importing GM into speech research is the absence of stable landmarks in vocal signals. Whereas biological forms often exhibit clear points of correspondence (e.g., sutures, process tips, joints), pitch contours are continuous, fluid, and highly variable across speakers, lexical items, and contexts.

To address this challenge, van Rijn et al. (2023) proposed pseudo-landmarks: each contour is resampled into a fixed number of equidistant, time-normalized points, permitting geometric alignment and comparison across tokens of different durations. Moreover, Rocha and Romano (2021) introduced SoundShape, an R package that adapts the eigensound protocol (in line with MacLeod et al., 2013) to transform sounds into analyzable shape surfaces, extending full-surface GM to bioacoustics. Building on these advances, the present study applies GM to stress-induced speech in real-world aviation emergencies. Motivated by Scherer's (1986, p. 143) observation that listeners readily perceive emotions while consistent acoustic markers are elusive, instead of relying solely on discrete acoustic parameters (e.g., mean F0, jitter, intensity), we treat the entire spectrogram as a dynamic surface and analyze its geometry using GM. This surface-based approach may capture stress-linked deformations in the global organisation of the vocal signal that conventional feature sets may overlook.

6.1.1 Understanding geometric morphometrics

In the 1990s, a major methodological shift, often dubbed the "morphometric revolution", reframed how biological form is studied. Here, 'form' refers not only to size but to the geometry of

structures, the relative spatial arrangement of parts and their variation across individuals or species (Bookstein, 1982). Rather than summarizing structures with a handful of linear distances, angles, or ratios, researchers began representing entire configurations with geometric coordinates (Adams et al., 2004). This approach, known as GM, preserves the full spatial arrangement of anatomical elements throughout data acquisition, analysis, and visualization, yielding results that are both statistically rigorous and visually interpretable (Slice, 2007; Slice 2005).

Conceptually, GM can be defined as a family of statistical techniques for analyzing shape—the geometric information that remains once differences in size, position, and orientation have been removed—in a spatially explicit manner. In other words, GM refers to the integrated suite of techniques for acquiring, processing, analyzing, and visualizing shape. The term, commonly credited to Les Marcus and introduced in print by Rohlf (1993), is used specifically for methods that rigorously pursue a complete, formally defined characterization of shape information from data collection through analysis.

Traditional morphometrics was based on distances and angles and suffers from well-known drawbacks: as the number of anatomical points increase, the number of pairwise measures grows combinatorially, yet still fails to capture the complete configuration from which those measures are derived. By contrast, Cartesian landmark coordinates constitute a compact and comprehensive representation, because they inherently encode all information contained in any subset of distances or angles among those points. It is this preservation of geometric structure that motivated the term GM (Corti, 1993; Perez et al., 2006). Generalized Procrustes Analysis (GPA) aligns landmark configurations by translating, scaling, and rotating them to minimize squared deviations from a consensus (mean) shape, thereby isolating variation in shape from nuisance variation in size, orientation, and position. The resulting Procrustes-aligned coordinates (shape variables) can then be explored with standard multivariate tools, including principal component analysis, regression, and discriminant methods. Deformation grids and shape-change vectors provide intuitive visualizations of how one shape differs from another.

Recent advances have expanded the framework beyond discrete, easily homologized points (Perez et al., 2006). Semi-landmarks—points allowed to “slide” along curves or surfaces—enable researchers to capture continuous morphological features where true landmarks are sparse or absent, improving fit to complex outlines and surfaces while maintaining statistical tractability. Together, landmarks and semi-landmarks make GM a flexible, robust toolkit for quantifying shape variation across an increasingly wide range of biological (and even non-biological) systems.

6.1.2 Geometric morphometrics applied to human voice

Traditional acoustic analyses of vocal expression—whether neutral, emotional, or pathological—have typically relied on discrete scalar features such as fundamental frequency (F0), jitter, shimmer, and intensity (Gnerre et al., 2023; Scherer, 1986). While effective in capturing localized acoustic

events, these measures fall short in representing the global structural dynamics that characterize vocal phenomena as temporally extended and multidimensional processes. Stylization of the pitch contour is more informative than simply using mean F0 as it preserves the dynamic shape of intonation (Knoll, & Costall, 2015). However, it is not sufficient on its own to fully capture the structural complexity of vocal expression (Rodero, 2011).

GM goes further by treating the entire contour as a shape within a spatial framework. This allows for more precise and holistic comparisons across vocal patterns. In their study on emotion classification from vocal expression, Van Rijn and colleagues (2023) apply GM to emotional human pitch in an innovative way, treating pitch contours as dynamic two-dimensional shapes (time × frequency) rather than relying on summary statistics. Their work demonstrates that morphometric descriptors of F0 contours—obtained through both landmark-based PCA and outline-based Fourier analysis—effectively capture dynamic features of vocal expression relevant to emotion recognition. By treating pitch trajectories as shapes and extracting pseudo-landmarks or harmonic components, they were able to derive low-dimensional representations that significantly outperformed standard summary statistics (e.g., mean F0, jitter) in multi-corpus classification tasks. Notably, their findings show that scrambling the temporal sequence of F0 points results in a marked drop in classification accuracy, underscoring the fact that morphometric descriptors retain meaningful temporal and structural information. Knoll and Costall (2015) combined qualitative inspection with quantitative Eigenshape analysis to capture holistic differences in F0 contour shapes across infant, foreigner, and adult-directed speech in natural versus simulated interactions.

In contrast, Rocha et al. (2021) adapt this approach to full acoustic spectrograms—primarily applied to animal vocalizations—treating sound as a three-dimensional shape defined by time, frequency, and amplitude. GM offers a mathematically coherent and conceptually rich framework for analyzing complex forms as configurations of homologous points within a structured shape space. Applied to voice, this framework invites us to treat spectrographic or pitch-based representations as morphable surfaces embedded in a three-dimensional domain defined by time, frequency, and amplitude.

Following the eigensound analysis approach introduced by MacLeod and colleagues (2013)—where Hanning-windowed spectrograms are sampled using regular semilandmark grids and projected into Kendall's shape space—vocal utterances can be transformed into high-dimensional geometric data, where dimensionality reflects the number of semilandmarks used to capture the spectro-temporal shape. Once normalized for duration and amplitude, these shapes can be subjected to multivariate statistical techniques such as Principal Component Analysis (PCA), which identifies dominant axes of variation across speakers, emotional states, or clinical populations. Canonical Variates Analysis (CVA) can then be employed to maximize discrimination between predefined groups, effectively modeling voice as a morphometric continuum (Rocha et al., 2021). In sum, the application of GM to human voice signals represents a novel and analytically powerful frontier in bioacoustics. By shifting focus from isolated

parameters to structured acoustic forms, this approach aligns with the embodied, relational, and temporally dynamic nature of vocal communication—offering new opportunities for investigating how psychological, physiological, and situational variables become inscribed in the acoustic fabric of the human voice.

6.1.3 The present study

This case study pioneers the application of GM methods to the analysis of speech produced under real-world stress conditions during an aviation emergency. While two previous studies have applied morphometric techniques to human pitch contours (Knoll & Costall, 2015; Van Rijn et al., 2023), no study to date has extended these methods to the full acoustic spectrum, where multiple dimensions of vocal production may interact to convey the impact of stress on speech. In this case study, we applied GM methods to speech produced under real-world stress conditions during an aviation emergency, modelling spectrograms as three-dimensional shapes (time × frequency × amplitude) to capture holistic, stress-related deformations in vocal output.

Following eigensound-inspired pipelines (Rocha & Romano, 2021), the full spectrographic surface was retained (preserving both harmonic and aperiodic components) allowing the quantification of global time–frequency–amplitude patterns beyond isolated acoustic parameters. For comparison, a complementary "ridge-only" analysis was conducted, in which the F0 contour of each utterance was extracted, aligned, and represented through a sequence of semi-landmarks.This approach reduces the acoustic signal to its f0 trajectory, which was then submitted to PCA. In doing so, it isolates pitch modulation from the broader acoustic spectrum, allowing us to examine whether stress-related vocal changes can be detected specifically through systematic deformations in the F0 contour. We focused on a single, phonetically matched phrase "American 1779" spoken by both a pilot and ATCO during an actual in-flight emergency, as well as in routine communication. This controlled lexical anchor allows for direct, shape-based comparison of vocal output across speakers and emergency vs routine conditions.

We used a short unit consistently with previous morphometric analyses that were done on single-word contours (Knoll & Costall, 2015). By resampling spectrograms onto a standardized time–frequency grid and analyzing their geometry via PCA, we aim to identify dominant modes of variation in spectral shape that may be associated with stress. Building on these morphometric analyses, a subsequent perceptual validation phase was designed to examine how the geometric deformations captured by GM relate to human listeners' perception of stress. Specifically, we aimed to test whether the morphometric space (as defined by PC1 and PC2) aligns with the perceptual space derived from listeners' stress ratings. If perceptual judgments mirror the GM patterns, this would suggest that morphometric deformation encodes perceptually salient information; conversely, a divergence between the two would

indicate that GM captures sub-perceptual acoustic structures, potentially valuable for automatic stress detection systems. Through this approach, we explore whether GM techniques can capture subtle, systematic deformations in vocal output linked to stress states. Given the limited number of tokens, this work should be regarded as a methodological case study, with the primary aim of demonstrating the feasibility and sensitivity of the GM approach to stressed speech, rather than providing robust estimates of effect sizes.

6.2 Method

Two complementary GM pipelines were used: a full-surface spectrographic analysis preserving the entire time–frequency–amplitude structure, and a ridge-only analysis based on normalised F0 contours capturing prosodic modulation. In both, utterances were pre-processed, resampled to homologous landmarks, and analyzed via PCA to visualize routine–emergency shifts for each speaker. To contextualize GM results, we also extracted conventional acoustic features (mean/SD F0, RMS energy, intensity, spectral centroid) from the same tokens and submitted them to PCA. The following subsections describe the source episode, units of analysis, analytical pipelines, and statistical procedures. All analyses described were conducted in R (version 4.5.1) using the packages tuneR, seewave, prcomp() and custom scripts implementing an eigensound-style semi-landmark resampling pipeline.

6.2.1 Episode description

The target utterance “American 1779” was drawn from an authentic aviation communication that began in a routine phase and later included an emergency exchange. This critical situation occurred on 21 May 2023, when an American Airlines Boeing 737-800 (registration N841NN), operating flight AAL1779 from New Orleans International Airport (KMSY) to Miami International Airport (KMIA), declared an emergency during climb-out from New Orleans after reporting a fuel leak. The crew subsequently requested and executed a return to New Orleans. Upon perceptual (auditory) assessment, no appreciable differences between the two phases (routine vs. emergency) are observed, which is perfectly consistent with the standardization and procedural control inherent to the aeronautical system. In the publicly available video, the aircraft’s flight path is also visible alongside the ATCO communication (https://www.youtube.com/watch?v=mAEm73iB2BU). We selected this cockpit-ATCO recording of American Airlines Flight for morphometric analysis because it captures a controlled transition from routine operations to a declared fuel-leak emergency. Additionally, unlike many aviation incident recordings, this audio offers exceptional clarity, minimal background noise and no overlapping transmissions, enabling precise extraction of vocal features. Critically, the emergency stemmed from a single failure (isolating stress-induced vocal modulation) while crews maintained strict procedural

communication throughout. This allowed us to detect sub-perceptual stress biomarkers—inaudible to human listeners but quantifiable morphometrically—within a real high-stakes scenario.

6.2.2 Unit of analysis

In the available recordings, both the pilot and ATCO uttered the call sign "American 1779" during the emergency. This served as a phonetically identical linguistic anchor across speakers and conditions, enabling direct comparison of spectral and prosodic characteristics within an identical phonetic context under control of any possible effect of the phonemic level. Both speakers were male. For the present analysis, we extracted the first two occurrences of the utterance produced by each speaker (pilot and ATCO) during the routine phase (0:37, 0:54 seconds), and the first two occurrences during the emergency phase (1:14, 1:24). Each acoustic unit was converted in a WAV file.

6.2.3 Full-surface geometric morphometric pipeline

Pre-Processing

Audio recordings were downsampled to 16 kHz to standardize the sampling rate and reduce computational load without compromising the frequency range relevant to speech. Stereo files were converted to mono by selecting the left channel to ensure consistency across tokens and to avoid artifacts introduced by averaging channels, as the recordings did not contain meaningful stereo information. Signals were amplitude-normalized to the ±16-bit PCM range to standardize intensity values across recordings. Signals were band-pass filtered between 300 and 3400 Hz using a fourth-order Butterworth filter to approximate telephony bandwidth and remove out-of-band noise. A second amplitude normalization compensated for minor level changes was introduced by filtering. Each utterance was cropped to match the duration of the shortest token in the dataset, ensuring strict temporal comparability for subsequent resampling.

Spectrogram computation

Spectrograms were computed using seewave::spectro() with a Hanning window length of 512 samples and 90% overlap, yielding high-resolution time–frequency representations. Spectral amplitude values were expressed in decibels and thresholded at (max – 25 dB) to suppress low-energy background noise.

Morphometric sampling

Following the eigensound protocol (MacLeod et al., 2013) adapted by Rocha and Romano (2021), each spectrogram was resampled onto a uniform semi-landmark grid of 47 frequency bands × 70 time bins. Linear interpolation along temporal and frequency dimensions generated pseudo-homologous points across all tokens. Amplitude values were row-centered (per frequency band) to remove absolute level offsets while preserving the relative distribution of energy. The complete spectrographic surface—rather than only the F0 ridge—was retained, allowing both harmonic and aperiodic components to contribute to the analysis.

6.2.4 Ridge-Only geometric morphometric pipeline

F0 Extraction

For the ridge-only analysis, the fundamental frequency (F0) contour of each token was extracted from the pre-processed waveform using seewave::fund(), an autocorrelation-based method. The pitch floor and ceiling were set to 75 Hz and 350 Hz, respectively, to accommodate male voices while avoiding octave errors. Analysis was performed with a 512-point Hanning window and 90% overlap. Octave jumps and spurious values were removed by discarding non-finite points and applying linear interpolation. Contours were expressed in semitones relative to each speaker's median F0 to control for inter-speaker differences in pitch range.

Morphometric Sampling

Each normalized F0 contour was time-normalized to the duration of the shortest token, then interpolated to 70 equally spaced pseudo-landmarks along the temporal axis, yielding homologous points across all tokens. This preserved modulation patterns while eliminating variation due to original duration differences.

6.2.5 Perceptual phase

Fifty-seven participants (with self-reported normal hearing) took part in the perceptual validation experiment. Participants ranged in age from 28 to 70 years and were recruited from the general population. The sample was approximately balanced by gender (30 males and 27 females). None reported formal training in phonetics or aviation communication, ensuring that all listeners were naïve to the experimental aims and to the specific nature of the stimuli. Each participant completed the task in a sound-attenuated laboratory room. The stimuli consisted of four audio tokens corresponding to the same recordings analyzed in the morphometric study: *American 1779* spoken by a pilot and the ATCO during routine and emergency phases. Each token was presented three times per trial, with

presentation order randomized across participants to prevent sequence effects. The experiment was conducted in a quiet, sound-treated environment. Participants were seated approximately 50 cm from a computer screen and wore circumaural headphones calibrated to a comfortable listening level. After hearing each stimulus, participants rated the perceived level of stress on a continuous Visual Analogue Scale (VAS) ranging from 0 ("Not at all stressed") to 100 ("Extremely stressed"). Each participant provided one rating per stimulus, for a total of four trials. The overall duration of the session was approximately nine minutes.

6.2.6 Statistical analysis

Morphometric and acoustic analysis

For both the full-surface and ridge-only pipelines, resampled data were assembled into a design matrix with tokens as rows and either spectrographic grid points (full-surface: 47 × 70 = 3,290 elements) or F0 pseudo-landmarks (ridge-only: 70 elements) as columns. All variables were mean-centered without scaling to preserve the original variance structure. PCA was performed using prcomp() in R. The proportion of variance explained by each component was computed to guide interpretation. Using the full-surface pipeline, we reconstructed mean spectrograms and the ±2 SD deformations along PC1 and PC2 on the original 47 × 70 grid and rendered them as 3D time–frequency–amplitude surfaces with a common z-axis scale, allowing direct visual comparison of spectral energy redistribution across speakers and conditions. In the ridge-only pipeline, we reconstructed F0 contours at ±2 SD along PC1 and PC2 to provide a geometric interpretation of pitch-shape variation in relation to prosodic modulation. In both pipelines, the transitions from routine to emergency for each speaker were plotted in the PC1–PC2 space as arrow vectors to illustrate condition-related shifts. Given the small sample size (n = 4), PCA was used solely as a descriptive tool to summarize and visualize variation, without inferential statistical testing.

To contextualize the morphometric findings, a third analysis was performed using a conventional acoustic feature set. For each token, mean and standard deviation of F0 (Hz), root-mean-square (RMS) energy, mean and standard deviation of intensity (dB), and spectral centroid (Hz) were extracted from the pre-processed waveforms. All features were z-scored within the dataset to remove scale differences.

The resulting feature matrix (rows = tokens; columns = features) was submitted to PCA using prcomp() on centered, unscaled data. The transitions from routine to emergency were plotted in the PC1–PC2 space to enable qualitative comparison of separation patterns between the morphometric pipelines and standard acoustic measures. For each PCA, the first two principal components were retained, and the percentage of variance explained by each is reported in the results section to allow direct comparison across analyses. Absolute PC scores are not directly comparable across pipelines due to differences in

scaling, number of landmarks, and variance structure; interpretations are based on relative positioning within each morphospace.

6.2.7 Perceptual analysis

A 2 × 2 within-subjects repeated-measures ANOVA with Speaker (pilot, ATCO) and Condition (Routine, Emergency) on VAS stress (0–100) was conducted. Moreover, we converted mean VAS stress ratings for the four tokens into pairwise "psychological distances". These distances were embedded via classical multidimensional scaling into a 2-D perceptual map (near points = similar; far points = dissimilar). From the same four tokens we derived three alternative 2-D spaces: (i) Full-surface GM (spectrogram semilandmarks → PCA scores), (ii) Ridge-only GM (time-normalized F0 pseudo-landmarks → PCA scores), and (iii) Standard Acoustics (conventional features → PCA scores). Together with the multidimensional scaling solution, this yielded four maps of the same items. Similarity between each analytic map and the perceptual map was quantified in two ways Distance–distance correlation**:** we computed inter-item distance matrices within each space and correlated them with the perceptual distance matrix (Spearman coefficient). High correlations indicate that items that are close/far perceptually are also close/far in the analytic space.

6.3 Results

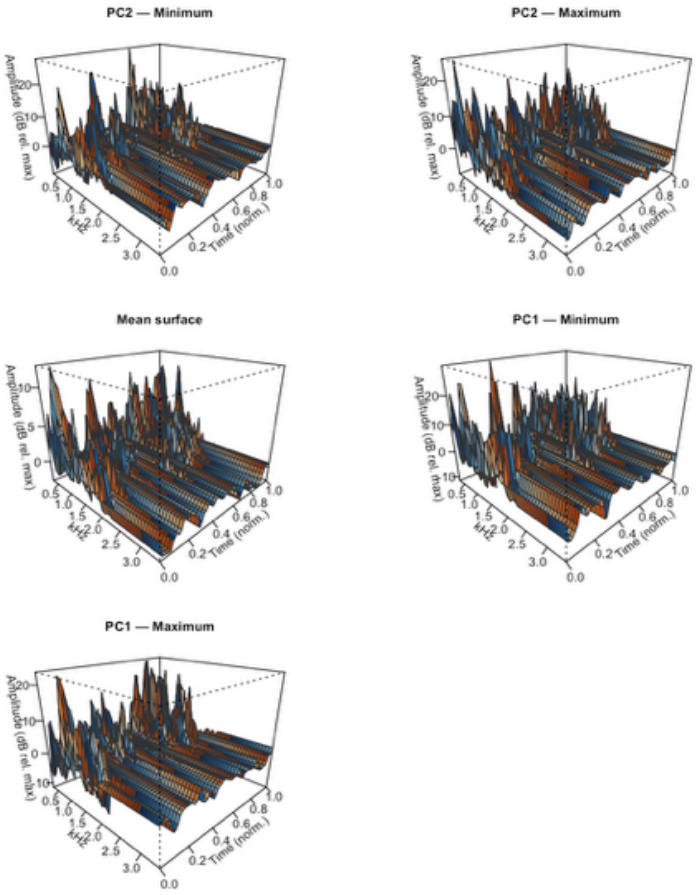
6.3.1 Full-surface spectrogram analysis

PCA of the full spectrographic surfaces revealed that the first two components accounted for 77.0% of the total variance, with PC1 explaining 50.5% and PC2 explaining 26.5%. PC1 primarily represented large-scale shifts in spectral height and overall energy distribution, whereas PC2 captured changes in spectral prominence and emphasis patterns across the time–frequency plane. In the PC1–PC2 space (Figure 13), tokens clustered distinctly according to both speaker role and communicative context. The pilot's emergency production of "American 1779" exhibited a strong positive score on both PC1 (84.92) and PC2 (76.08), reflecting increased overall spectral height and greater prominence variation in higher frequency bands compared to other conditions. The pilot's routine production also showed a positive PC1 score (72.18) but a large negative PC2 score (–66.16), suggesting similarly elevated spectral height but reduced prominence variability. By contrast, ATCO productions were located in the negative PC1 range, indicating lower overall spectral height. The ATCO emergency token (–107.36, 37.82) was characterized by reduced spectral height but increased mid-frequency prominence, whereas the ATCO routine token (–49.74, –47.74) combined both lower spectral height and reduced prominence variation. Morphometric reconstructions of ±2 SD deformations (Figure 13, right) illustrated that PC1 maxima corresponded to an upward shift of the entire spectral energy distribution,

with enhanced emphasis in higher frequency regions, while PC1 minima concentrated energy in lower frequency bands. PC2 maxima were associated with localized increases in mid-to-high frequency prominence, whereas PC2 minima displayed a flatter spectral prominence profile. Overall, both speaker role (pilot vs. ATCO) and communicative context (emergency vs. routine) influenced the global spectro-temporal configuration of the utterance. Emergency speech tended to exhibit more extreme shifts in energy distribution and prominence allocation than routine speech, particularly for the pilot. As a complementary visualization, absolute differential spectrograms (Figure 14) highlight where the energy distribution differs between emergency and routine tokens for each speaker. The pilot exhibits broader and stronger changes, with a band of enhanced differences between ≈300–800 Hz spanning much of the utterance. By contrast, the ATCO map shows sparser, lower-magnitude differences concentrated below ≈2 kHz. Because these maps plot |Emergency − Routine| in dB, they convey the magnitude (not the direction) of change and are consistent with the PC1/PC2 pattern indicating larger emergency-related reallocations of energy for the pilot.

Figure 13

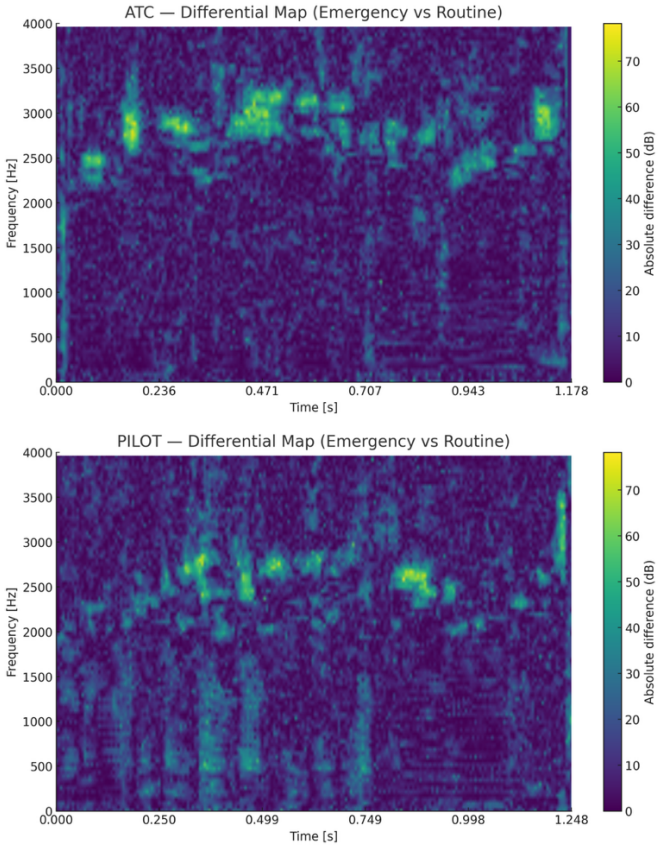
Morphometric reconstructions of full-surface spectrogram variation along the first two principal components (PCs)



Note. The central panel shows the mean spectrographic surface of the utterance "American 1779." Panels labeled "PC1 – Minimum" and "PC1 – Maximum" illustrate the ±2 SD deformations along PC1 (50.5% of variance), which primarily represent large-scale vertical shifts in spectral energy, with maxima emphasizing higher-frequency regions and minima concentrating energy in lower-frequency bands. Panels labeled "PC2 – Minimum" and "PC2 – Maximum" show the ±2 SD deformations along PC2 (26.5% of variance), which capture changes in spectral prominence and emphasis patterns, particularly in mid-to-high frequency bands.

Figure 14

Differential spectrograms



Note. Absolute differential spectrograms (|Emergency − Routine|, dB) for the ATCO (top) and pilot (bottom) productions of "American 1779." Audio was converted to mono and analyzed at the native sampling rate using Hann-windowed STFTs (1024-sample window, 75% overlap). Maps display the 0–4000 Hz range; color scales are shared across panels and reflect absolute dB differences (0 = no change). In these tokens, the pilot shows larger low-frequency (≈300–800 Hz) differences, whereas ATCO changes are sparser and mostly below ≈2 kHz.

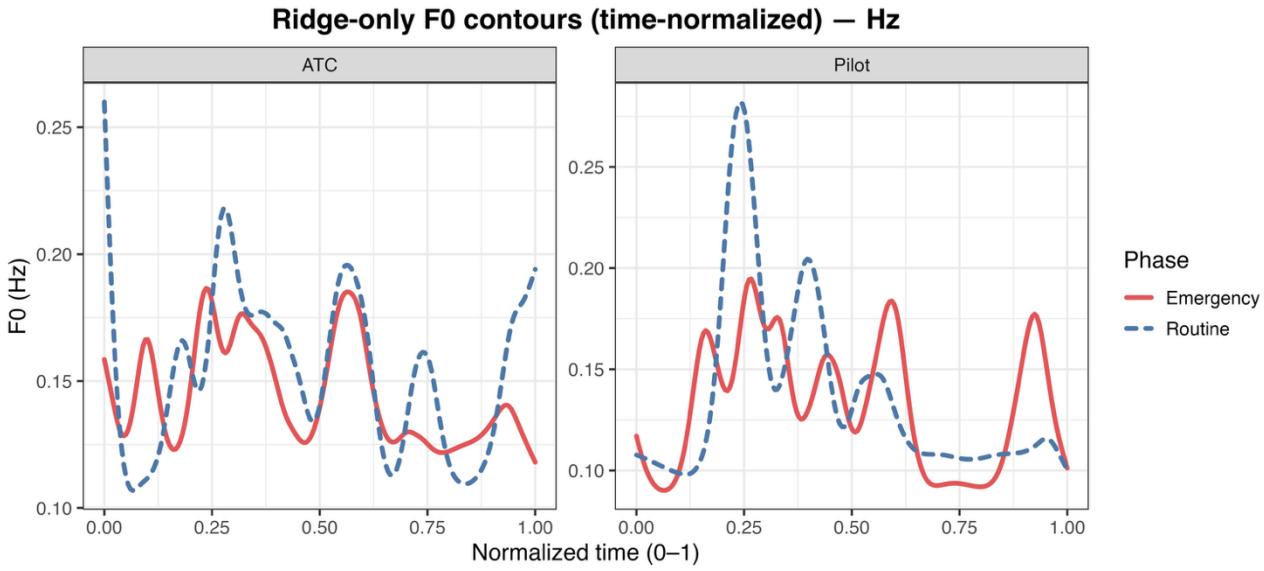
6.3.2 Ridge-only F0 contour analysis

The first two principal components accounted for 49.5% and 28.0% of the total variance, respectively. As shown in Figure 15, the pilot's F0 contour exhibited a marked shift from routine to emergency conditions along both PC1 and PC2, indicating a flatter and less modulated pitch profile under emergency. In contrast, the ATCO's F0 contour showed a smaller displacement between phases, suggesting a relatively stable pitch modulation pattern. These differences imply that, for the pilot, the shape of the F0 contour alone captures a substantial portion of the stress-induced variation observed in the

full-surface spectrogram analysis, whereas for the ATCO the ridge-only and full-surface morphospaces are more similar.

Figure 15

Morphometric reconstructions of pitch contours



Note. Time-normalised F0 contours) for routine (blue dashed line) and emergency (red solid line) productions of "American 1779" by the ATC O(left) and the pilot (right). For the pilot, the emergency contour shows reduced pitch modulation and a flatter overall profile compared to the routine condition. In contrast, the ATCO's contours display smaller differences between phases, indicating a more stable modulation pattern.

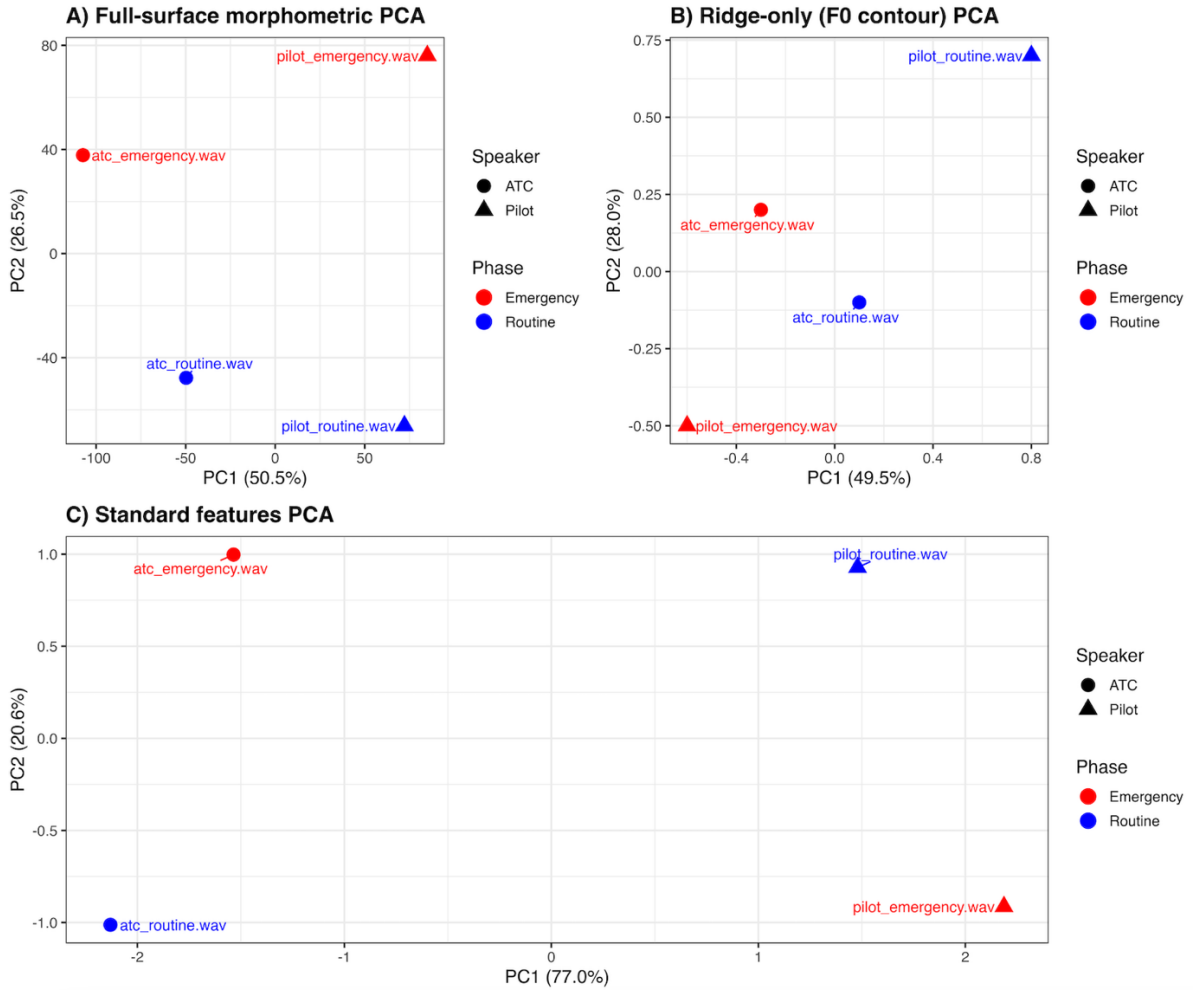
6.3.3 Comparison with standard acoustic features

A PCA based on the subset of complete standard acoustic features (F0 mean and standard deviation, spectral centroid, RMS energy, mean and standard deviation of intensity) revealed that the first two components explained 77.0% and 20.6% of the variance, respectively. In the PC1–PC2 space, separation was primarily driven by speaker role (pilot vs. ATCO) along PC1, with only partial discrimination between emergency and routine tokens. The pilot's emergency utterance (PC1 = 2.19, PC2 = −0.91) was more distant from the pilot's routine utterance (PC1 = 1.48, PC2 = 0.93) than were the ATCO tokens (routine: PC1 = −2.13, PC2 = −1.01; emergency: PC1 = −1.54, PC2 = 1.00). Compared to the morphometric analysis, the standard feature PCA yielded a less distinct clustering by communicative

context, suggesting that GM captures additional information beyond that available from conventional summary parameters. For a side-by-side view of the three PCA analyses, see Figure 16.

Figure 16

Comparison of morphometric and conventional acoustic feature analyses



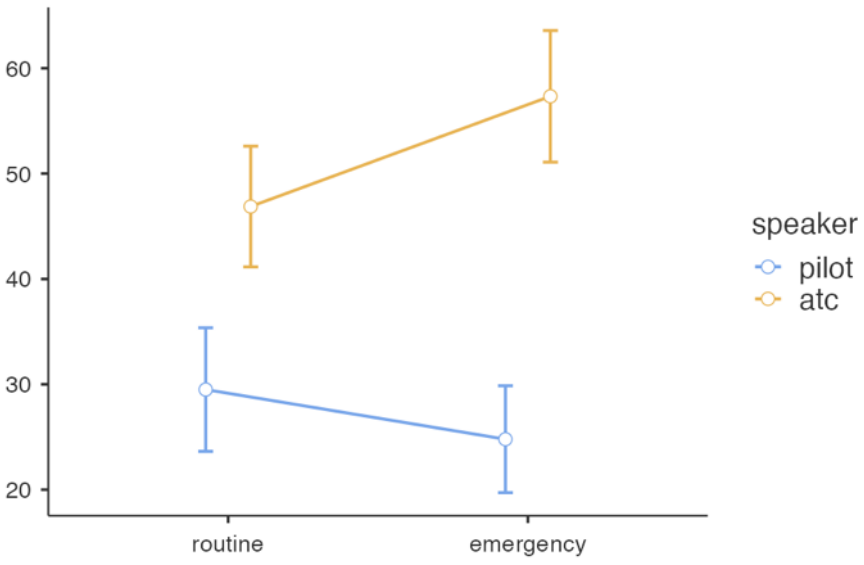
Notes. A) PCA of full-surface spectrographic shapes (morphometric approach), explaining 50.5% (PC1) and 26.5% (PC2) of the variance. Tokens cluster clearly by both speaker role and context, with a marked separation between emergency and routine conditions. (B) PCA of ridge-only (F0 contour) shapes, explaining 49.5% (PC1) and 28.0% (PC2) of the variance, showing a pronounced shift in the pilot's pitch contour between phases and a smaller displacement for the ATCO. (C) PCA of standard acoustic features (F0 mean and SD, spectral centroid, RMS energy, intensity mean and SD), explaining 77.0% (PC1) and 20.6% (PC2) of the variance. Separation is mainly driven by speaker role, with less distinct clustering context compared to the morphometric approaches.

6.3.4 Perceptual analysis

A 2 × 2 within-subjects repeated-measures ANOVA with speaker (pilot, ATCO) and condition (Routine, Emergency) on VAS stress ratings (0–100) showed a main effect of speaker, $F(1,56) = 49.76$, $p = .001$, $\eta p^2 = .47$, no main effect of condition, $F(1, 56) = 3.09$, $p = .08$, $\eta p^2 = .05$, and a significant Speaker × Condition interaction, $F(1, 56) = 15.53$, $p = .001$, $\eta p^2 = .22$. In Tukey HSD adjusted pairwise comparisons for the speaker × condition interaction, the contrast between pilot in routine condition and pilot in emergency condition was not significant ($t(56) = 1.66$, $p = .356$). ATCO in routine condition scored higher than pilot in the same condition by 17.38 units in the perceptual ratings, SE = 3.97, $t(56) = -4.37$, $p < .001$, and higher than pilot in emergency condition by 22.09 units, SE = 3.50, $t(56) = -6.31$, $p < .001$. ATCO in emergency condition scored higher than pilot in routine condition by 27.84 units, SE = 4.26, $t(56) = -6.54$, $p < .001$, and higher than pilot in emergency condition by 32.55 units, SE = 4.08, $t(56) = -7.97$, $p < .001$. Within ATCO, emergency was higher than routine by 10.46 units, SE = 2.16, $t(56) = -4.84$, $p < .001$.

Figure 17

Perceived stress by speaker and condition



6.3.5 Relationship between morphometric and perceptual spaces

To relate the acoustic representations to listeners' judgments, mean VAS stress ratings were transformed into a perceptual dissimilarity matrix (Table 14). This matrix was embedded via classical

multidimensional scaling (MDS), yielding a two-dimensional perceptual space for the four tokens. Concordance between each analytic space and this perceptual solution was indexed by the Spearman rank correlation (ρ) computed on inter-item distances.

The results show a marked divergence across approaches. Full-Surface GM displayed the strongest alignment with perception (ρ = .771), suggesting that the global spectrotemporal "shape" of the signal preserves information that listeners find salient. Standard acoustics also correlated positively with perception (ρ = .714), albeit more weakly, indicating that conventional summary features capture a substantial—though not exhaustive—portion of the perceptual structure. By contrast, Ridge-Only GM yielded a weak negative association (ρ = −.257), implying an inversion of proximity: pairs judged similar by listeners tend to be mapped farther apart in the F0-based morphometric space, and vice versa. Collectively, these patterns point to the primacy of full spectrotemporal geometry over pitch-only contours in mirroring perceived stress differences.

Table 14

Perceptual Dissimilarity Matrix Based on Stress Ratings

Token	Pilot-Routine	Pilot-Emergency	ATCO-Routine	ATCO-Emergency
Pilot-Routine	—	16.82	30.73	37.66
Pilot-Emergency	16.82	—	30.09	40.05
ATCO-Routine	30.73	30.09	—	15.50
ATCO-Emergency	37.66	40.05	15.50	—

Note.Values represent mean absolute differences in stress ratings on a 0–100 Visual Analogue Scale b etween all pairs of trials across 57 participants. Higher values indicate greater perceptual dissimilarity.

6.4 Discussion

The findings indicate that GM is a sensitive and interpretable approach to human speech. In the full-surface analysis, tokens separated by both speaker role (pilot vs. ATCO) and context (emergency vs. routine) along the first two principal components. Functionally, PC1 indexed global shifts in spectral height and overall energy allocation, whereas PC2 captured changes in spectral prominence across the time–frequency plane. The ridge-only analysis of the F0 contour recovered a pronounced displacement from routine to emergency for the pilot, in comparison to a comparatively modest shift for the ATCO, indicating flatter and less modulated intonation under stress for the former. By contrast, a PCA of conventional summary features primarily segregated speakers and offered only partial phase discrimination. Treating the signal as a geometric object thus preserves global spectro-temporal organization and

reveals stress-linked deformations that scalar descriptors tend to obscure. A parsimonious interpretation is that stress perturbs coupled source–filter dynamics (e.g., subglottal pressure, laryngeal configuration, spectral tilt, aperiodicity) rather than any single acoustic parameter. The stronger separation in the full-surface space aligns with the expectation that stress modulates not only F0 but also timbral and noise components. In the ridge-only space, the pilot-specific flattening coheres with accounts of increased physiological arousal and task load compressing F0 excursions, whereas the ATCO's smaller displacement is consistent with training that stabilizes prosody during emergencies. In simulator studies of aviation communication, rising cognitive load reliably elevates mean F0 while narrowing its range—an intonational "compression" consistent with reduced modulation under controlled effort (e.g., Huttunen et al., 2011) and with reports that F0 variability contracts at tolerable workload levels (Johannes et al., 2007). The controller's comparatively modest shift aligns with doctrine and training that institutionalize neutral tone and standardized phraseology in abnormal events (e.g., ICAO/CAP/FAA guidance), fostering prosodic stability despite heightened task demands. Role asymmetries of this kind are also noted in field analyses, where operators with explicit regulatory duties show smaller prosodic deviations than actors directly engaged in time-critical actions (e.g., Ruiz et al., 1996). Taken together, these converging observations suggest that F0-shape metrics are sensitive not merely to "pitch height," but to the dynamic balance between arousal and proceduralized control, yielding a mechanistic account of why pilots may exhibit contour flattening under stress while trained controllers remain comparatively stable.

Moreover, the perceptual pattern is unambiguous: listeners consistently judge the ATCO as more stressed than the pilot; moreover, the pilot in emergency condition is perceived as even less stressed than in routine condition. A plausible account is that, in critical phases, the pilot engages top-down regulatory mechanisms that dampen audible stress markers, a functional strategy for managing cognitive load and preserving communicative clarity (cf. Ruiz et al., 1996).

Turning to the correspondence between acoustic representations and perceptual judgments, a clear hierarchy emerges. Full-Surface GM shows the strongest alignment with perception, indicating that the complete spectro-temporal geometry of the signal preserves the information most salient for stress appraisal. By contrast, Ridge-Only GM exhibits a weak negative association with perceptual structure: it not only fails to capture it but tends to invert it (pairs that sound similar are mapped far apart in the F0 space, and vice versa). Notably, Ridge-Only GM sharply separates the two pilot tokens, whereas this separation is absent perceptually suggesting that pitch-contour shape, by itself, indexes differences in prosodic control that listeners do not weight as "stress" in this task/context. Standard acoustic features display substantial yet incomplete correspondence: they explain a sizeable share of perceptual variance but fall short of the comprehensive coverage achieved by full-spectrum morphometry. Overall, the pattern implies that perception favors multi-cue integration and spectro-temporal configuration.

Methodologically, the results underscore a limitation of conventional parameters (e.g., mean/SD F0, RMS, intensity): by aggregating over time, they attenuate primarily configurational differences, such as the geometry of rising, falling movements or the relocation of spectral emphasis. GM, in contrast, encodes the geometry of trajectories and surfaces and yields low-dimensional summaries whose ±2 SD reconstructions afford direct interpretability of the underlying deformations. The proposed pipeline—duration normalization and (semi/pseudo-) landmark resampling, PCA in shape space, and visualization of shape reconstructions—combines statistical compression with transparent, mechanism-oriented visualization.

6.4.1 Limitations

This case study is constrained by a very small dataset (four tokens of a single phrase from two speakers) and telephony-bandwidth audio, limiting generalizability to broadband speech. Time normalization may conflate duration with "shape," and ridge-only results depend on potentially error-prone F0 tracking. All the analysis presented was used only descriptively.

6.4.2 Implications and future work

Treating speech as a geometric object yields shape-based descriptors that capture stress-linked spectro-temporal deformations beyond conventional scalar features while remaining interpretable via PC reconstructions. This suggests practical value for monitoring and decision-support in operational settings, provided models include speaker normalization, rigorous validation, and calibrated false-alarm control. Future work should: (i) scale to larger, balanced corpora spanning multiple phrases, speakers, and languages; (ii) benchmark full-surface vs. ridge-only GM under controlled channel/noise manipulations and alternative preprocessing choices; (iii) treat single acoustic parameters (e.g., CPP, HNR) as covariates rather than replacements. In our view, GM could also be applied to the study of aviation accident reports and holds broad potential in other domains of acoustic analysis, such as the diagnosis of speech and voice disorders or the diagnosis of neurological diseases from speech and voice variations.

6.5 Conclusions

Within the constraints of an N-of-few design, GM provided sensitive and interpretable summaries of stressed speech in an authentic aviation context. Full-surface GM captured global reallocations of spectral energy, while ridge-only GM revealed pilot-specific flattening of the F0 trajectory. Standard acoustic features were comparatively less informative for phase discrimination. GM should therefore be considered a complementary tool to established acoustic analytics, with clear potential for

validation and extension on larger datasets. This perspective bridges phonological theory, signal processing, and shape analysis to reconceptualize the voice as a morphable, expressive surface.

CHAPTER 7

General Conclusions

This dissertation is positioned within a long-standing scholarly dialogue. The claim that vocal expression discloses internal states is not novel, and the empirical investigation of the effect of stress on voice started at the beginning of the 20th century, with psychiatrists trying to diagnose emotional disturbances through the newly developed methods of electroacoustic analysis (e.g., Isserlin, 1925; Skinner, 1935).

The enduring interest in this relationship is powerfully exemplified in aviation, a domain where vocal behavior is actively investigated as a biomarker for human performance in complex, safety-critical operations. Yet, despite this enduring interest, substantial challenges remain.

The unresolved challenge does not concern the existence of informative vocal variation, but rather the establishment of a methodological and conceptual consensus that permits measurement to be valid, comparable, and transferable. The fundamental questions remain open: what to measure, how to measure it, and under which conditions. Addressing this challenge, which is epistemic before it is technical, the three studies presented here form a deliberate and coherent trajectory. The argument proceeds from conceptual clarification, through empirical validation on real communications, and culminates in a reframing of the measurement.

The systematic review (Chapter 4) establishes the conceptual foundation for this dissertation by mapping the extant literature and diagnosing its central impediment: a profound fragmentation of methods and definitions that obstructs the identification of stable vocal markers for critical psychophysiological states, such as stress. It thereby not only provides the rationale for the subsequent empirical investigation but precisely pinpoints the critical knowledge gap: the absence of ecological, role-sensitive analyses of how stress manifests vocally in authentic emergency communications.

The ecological study (Chapter 5) occupies precisely this gap. The reported analysis of real pilots–ATCOs exchanges during emergencies fulfils the review's call to move from laboratory simulations to field data. The demonstration of robust, role-specific acoustic signatures extends the synthesis of the review, moving beyond generic correlates to show how operational roles shape the vocal expression of stress. While this study achieves notable discriminative performance with advanced yet conventional acoustic indices, it also intimates the limits of a purely parametric approach.

The morfometric study (chapter 6) addresses those limits directly by proposing a shift in analytic perspective. The morphometric study challenges the prevailing framework, challenging (o questioning) that discrete scalar parameters are insufficient to capture the holistic deformation of the

vocal signal under stress. By reconceptualizing vocal signal as a dynamic geometric form, it reframes the problem of measurement and opens a complementary path for modelling the global shape of vocal change.

As set out in the introduction (chapters 3 and 4), the entire enterprise is situated within a horizon that seeks to connect meaning, measurement, and context, moving explicitly "Beyond the Black Box." Measurement of voice is not a mere search for numerical residue of physiological processes; it is the pursuit for an embodied, embedded and situated signal in which physiology, cognition, emotion, linguistic component and social roles are inextricably intertwined. The general discussion that follows, therefore, transcends a chapter-by-chapter commentary to offer a meta-reflection of the entire work, evaluating its collective theoretical contributions, practical implications, limitations, and future directions.

7.1 Summary of results

The systematic review corroborates a set of recurrent directional shifts in vocal behavior associated with states that are operationally salient in aviation. It identifies a fundamental dichotomy: stress and workload typically elevate arousal-mediated parameters (e.g., F0), whereas fatigue and sleepiness manifest as a reduction in vocal energy. These general trends, however, are insufficient to define a universal vocal marker of stress. The primary impediment is profound methodological heterogeneity—in definitions, protocols, and measurement techniques—that severely limits both the comparability of findings and their operational applicability. This is not merely a technical limitation but a conceptual one, stemming from the treatment of overlapping, interdependent constructs as discrete categories. Consequently, the review underscores an urgent need for standardized frameworks to define and measure stress and, crucially, for more ecologically valid research.

This corpus study of authentic pilot-ATCO exchanges addresses a critical gap, demonstrating that a shared core of acoustic stress markers emerges across different operational roles, most notably an elevation in fundamental frequency consistent with acute sympathetic arousal. In contrast to the review corpus, this study also documents a reduction in HNR, further signaling heightened vocal effort in emergencies. The convergence of elevated F0 and reduced HNR strongly aligns with a physiological profile of acute sympathetic arousal. This finding constitutes the primary contribution of the thesis: the first systematic identification of a robust acoustic profile of acute stress in genuine emergency communications. In addition, this study demonstrates a fundamental role-based divergence in vocal expression. These distinct profiles are interpreted as emergent adaptations to the unique cognitive demands and operational responsibilities inherent to each role.

The morphometric case study contributes a methodological advance rather than definitive population-level estimates. It explores the application of Geometric Morphometrics—a technique for quantifying shape, borrowed from evolutionary biology—to the analysis of voice under stress. Its

primary value lies not in the statistical generalizability of its results, given its proof-of-concept design, but in its demonstration of feasibility and its profound heuristic potential. The analysis reveals that a "full-surface" morphometric approach can differentiate vocalizations with regard to the speaker and the operational phase within a geometric shape space, capturing spectro-temporal deformations induced by stress that conventional acoustic metrics partially obscure. Crucially, the inclusion of a perceptual validation stage relates these geometric distances to the listeners' stress judgments, suggesting that the morphometric representation preserves a perceptual gestalt of the signal. Thus, this study proposes a new analytical paradigm that reconceptualizes vocal stress not as a set of discrete parameter changes, but as a systemic deformation of the whole vocal signal.

7.2 Main contributions of the dissertation

This dissertation makes three interconnected contributions that advance the field of vocal analysis, moving from a theoretical reconceptualization to a methodological innovation.

- A situated and pragmatic view of voice

  Rather than treating vocal output as either a purely physiological reflex or a purely communicative act, I argue that it is a hybrid signal living at the crossroads of biology and pragmatics of communication. The consequence is a paradigm shift: away from the search for universal, invariant markers and toward adaptive profiles that are sensitive to role, task and context. At the same time, this thesis acknowledges a thin, quasi-universal substrate of vocal response, most consistently, elevations in F0 under acute stress. These first-order regularities are genuine, but they are too coarse-grained: an elevated F0 is a broad marker of heightened arousal, for example, equally consistent with stress, anger, or emphasis, and thus lacks the specificity needed to discriminate neighboring states on its own. In this light, such tendencies are better treated as priors to be refined by adaptive, context-sensitive profiles rather than as diagnostic signatures in their own right. This shift carries methodological implications: it prioritizes longitudinal designs, individualized baselines, and explicit context control, and it cautions against the perennial "holy grail" of a single, one-size-fits-all biomarker. The more promising path is toward person-specific ensembles of markers whose meaning is anchored to the speaker and the situation.

- An ecological validation in a high-stakes domain

  The empirical contribution of this thesis lies in its ecological grounding. To our knowledge, this is the first study to subject a large corpus of real aviation emergencies to systematic vocal analysis, moving beyond isolated case studies or laboratory simulations. The scale and fidelity

of this dataset—featuring standard phraseology, operational noise, time pressure, and genuine consequences—provide a unique testbed. It allows for the examination of vocal stress not as a laboratory artifact, but as it unfolds in the very environments where its accurate interpretation is most critical for safety and performance.

- A novel morphometric paradigm for vocal measurement

  Robust perceptual evidence shows that listeners can differentiate emotions and states such as stress from the voice (Ciceri & Anolli, 2000). Attempts to map these categories onto discrete, one-to-one acoustic markers have had only partial success—not necessarily because such patterns do not exist, but because conventional scalar features capture them imperfectly (Banse & Scherer, 1996; Juslin & Laukka, 2003; Juslin & Scherer, 2005; Patel et al., 2011). Here, Geometric Morphometrics is advanced as a promising complement: a shape-based representation of acoustic structure (e.g., contours and spectral/formant configurations) that more closely tracks perceptual organization. In this sense, morphometrics functions as a potential bridge between what listeners hear (in a gestaltic sense) and what algorithms measure, filling a key methodological gap.

7.3 Practical implication

The findings of this dissertation translate into at least two concrete applications along the safety management continuum: from post-incident investigation to proactive risk mitigation.

- Post-occurrence applications

  In the aftermath of incidents and serious occurrences, structured analysis of cockpit voice recordings can reconstruct crew cognitive and affective states with greater precision than narrative accounts alone. By time-aligning speech segments with flight telemetry and ATCO logs, and by quantifying interpretable markers of vocal behavior, investigators obtain an independent and auditable line of evidence about workload, stress, fatigue, coordination, and communication breakdowns. This triangulated approach constrains post hoc speculation, clarifies contributory human-factors mechanisms, and grounds safety recommendations and training interventions in measurable signal properties rather than anecdotal impressions.

- Prospective monitoring in operations

  Beyond retrospective use, voice can serve as an early, non-invasive, low-burden sensor of psychophysiological state for pilots and ATCOs in real flight conditions. When referenced to

intra-speaker baselines and modeled with contextual information about role, phraseology, and phase of flight, vocal indices can yield timely, interpretable indications of rising activation or deteriorating resources. Delivered as adaptive, graduated cues rather than binary alarms, such information can support crew resource management, workload regulation, and fatigue-risk programs without displacing human judgment. Crucially, operational deployment requires robust signal conditioning for noisy channels, transparent uncertainty estimates, and privacy-preserving governance to ensure that monitoring enhances safety while respecting professional autonomy.

7.3.1 An industrial application case

The methodological infrastructure developed in this thesis provides the practical basis for operational deployment in aviation. On the strength of this framework, a collaboration with an aviation accident-investigation authority[6] has been initiated. In practical terms, the project turns the entire analysis pipeline into a single software application and aims to test it on real cockpit voice recorder data. Within this framework, each audio segment is time-synchronized with flight telemetry during critical phases, so we can examine how changes in the voice co-vary with the aircraft's state. The objective is to derive and qualify context-sensitive vocal biomarkers of situational stress and related functional states.

This is achieved by synchronizing crew communications with the evolving flight profile, performing a diarization for speaker and flight phases with high precision, and estimating indices against individual baselines so that deviations can be meaningfully interpreted within the operational context. Anticipated uses span investigations, training, and live operations. In investigations, objective quantification of vocal markers offers an independent line of evidence for assessing crew state and its contributory role, thereby reducing interpretive ambiguity and complementing established sources. In pilot training, synthetic and interpretable indices returned within the simulator increase state awareness, support workload regulation, and enhance communicative clarity, enabling debriefings tailored both to individual vocal profiles and to situational departures from baseline. In live operations, near-real-time monitoring with adaptive thresholds and graduated alerts aims to curb unsafe behaviors while preserving the primacy of human judgment and the principles of crew resource management.

Several conditions require disciplined control to preserve validity and operational usefulness. Characteristics of the CVR channel and ambient noise must be managed through explicit signal-conditioning protocols, uncertainty estimation, and robustness testing. Linguistic variability, standardized phraseology, and role differentiation systematically shape

[6] Under the terms of a confidentiality agreement, the partner's identity and project specifics cannot be disclosed.

vocal expression and must be modeled to avoid false positives driven by style or task structure. The mapping between vocal markers and psychological constructs cannot rely on any single parameter; it must be anchored in triangulation with contextual information and aircraft telemetry, privileging composite indices with clear theoretical grounding and demonstrated operational utility. Ethical governance of voice data further demands attention to privacy and purpose limitation, supported by transparent oversight and periodic audits.

7.4 Principal limitations

Beyond the individual studies, this thesis is subject to several inherent limitations, arising from its commitment to ecological validity and its exploratory methodological scope.

- A primary limitation concerns construct validity

  In the analysis of real cockpit–ATCO conversations it is not feasible to cleanly dissociate neighboring constructs such as stress, emotional arousal and communicative workload. In practice, we correlated the occurrence of the emergency event (the stressor) with acoustic variation, but we were unable to directly measure objectively stress, quantify it, or distinguish among different sources of stress. No physiological recordings or self-report measures were available; stress was therefore inferred from the situational antecedent (the emergency). As a result, some observed changes may index heightened workload or arousal rather than stress per se. Throughout, “stress” was operationalized broadly, encompassing cognitive, emotional and workload-related demands that co-occur in critical phases.

- The trade-off of ecological validity

  This thesis privileges ecological validity by analyzing spontaneous, real-world communications; such choice strengthens external validity while constraining experimental control. Several consequences follow. First, phonetic content is not balanced, and some acoustic indices were excluded because of outliers; both factors may have attenuated or distorted effects that would be clearer under controlled elicitation. Second, while the systematic review highlighted the relevance of MFCC-based representations, these features were not included in the subsequent empirical analyses. The resulting misalignment between evidence synthesis and empirical implementation limits comparability across studies and leaves open whether MFCCs would have improved sensitivity or robustness in the present datasets. All acoustic features were standardized per speaker using z-score normalization (mean 0, unit variance), computed with respect to each speaker’s routine-phase baseline. This

speaker-intrinsic scaling is a conventional way to render heterogeneous measurements comparable across speakers and recording conditions (Bauer et al., 2024; Rose, 1987). We adopted it for two reasons: to suppress anatomy-driven dispersion that can otherwise swamp linguistic contrasts and fine phonetic detail (Rose, 1987), and to temper the uncontrolled segmental and phraseological variability inherent to spontaneous operational speech. Framed this way, estimated effects are read as deviations from each speaker's own baseline, which improves the interpretability of situational changes and facilitates cautious cross-speaker comparison. This normalization method, however, is a deliberate transformation. By anchoring features to speaker-specific baselines, we intentionally remove between-speaker variation to focus the analysis on within-speaker, situational changes.

- The proof-of-concept nature of the morphometric approach

  Finally, the geometric-morphometric component was intentionally framed as a proof-of-concept based on a case study, without claims of immediate generalizability. Its purpose here is demonstrative: to show that spectro-temporal shape captures perceptually relevant structure. Establishing population-level stability and operational thresholds for these representations requires larger samples, controlled replications and convergence with external criteria.

7.5 Future directions

This dissertation opens several critical pathways for future research, extending from methodological refinement to theoretical integration.

- The psychophysiology of vocal markers

  A critical next step is to move beyond establishing correlation and toward explaining the underlying psychophysiology. We must clarify how individual differences in coping resources and top-down regulatory strategies directly shape the acoustic structure of the voice during high-stakes events. The central question is not just if stress affects the voice, but through what specific cognitive and physiological pathways these internal states become acoustically manifest. Elucidating these mechanisms is paramount for transforming vocal analysis from a descriptive tool into an interpretative framework for an operator's cognitive-emotional state, with profound implications for real-time monitoring and operational diagnostics in safety-critical domains.

- A multimodal framework

  Voice is one channel within a larger communicative system. Progress therefore requires integrating acoustic analysis with other meaning-making streams—verbal–linguistic structure (not only lexical choice, but also syntax, morphology, pragmatics and discourse), facial and bodily kinematics, and physiological indices of arousal. A multimodal framework should temporally align these channels and model when they provide complementary information, when they are redundant, and when they diverge or co-regulate under stress. By testing patterns of convergence and contradiction across streams—and doing so with interpretable fusion methods—it becomes possible to recover a more complete and calibrated picture of human performance, one that links what is said, how it is said, and how the body and physiology jointly sustain or disturb communication in demanding conditions.

- The dyadic lens: interpersonal tuning and miscommunication

  Pilot and controller do not simply alternate turns; they co-construct a shared communicative state. Future work must therefore model interpersonal vocal tuning, asking how one agent's vocal markers anticipate, amplify, or dampen those of the other. Analyses of coupling—such as temporal coordination, prosodic convergence, and latency of responses—can reveal where coordination holds under pressure and where it imperceptibly frays. This perspective also reframes miscommunication. In aviation, miscommunication is too often attributed solely to deficits in English proficiency, in line with Shannon and Weaver's mathematical model of communication (1949). Proficiency matters, but it is only one piece of a larger system in which pilots and ATCOs co-construct a shared communicative state. A more productive agenda models interpersonal tuning directly, asking how one agent's vocal behavior anticipates, amplifies, or dampens the other's responses, and whether impending breakdowns are foreshadowed by small but measurable losses of alignment in timing, prosody, and repair. Viewed through this lens, miscommunication is not "noise" in the channel or merely a matter of language proficiency; it is a measurable failure of alignment across acoustic form, lexical content, pragmatic intent, and the shared situational model.

- Scaling the morphometric paradigm

  The promise of Geometric Morphometrics must be scaled from a proof-of-concept to a robust analytical tool. This requires expanding its application from a few tokens to large-scale corpora to firmly establish its generalizability. A key empirical question is

to systematically verify that the morphometric signal is indeed the measurable counterpart of perceptual gestalt, outperforming discrete acoustic parameters. Furthermore, morphometric studies should become longitudinal and person-specific. Establishing an individual's baseline "vocal geometry" in low-stress conditions and tracking deviations over time would allow for the estimation of individualized stress thresholds, recovery patterns, and detectable change.

- Broadening beyond emergencies, conceptual and methodological generalization of situational stress

  Future work should extend beyond declared emergencies and sample a broader range of operationally challenging yet non emergency conditions. Aviation routinely exposes operators to acute demands that are safety relevant without involving an explicit threat to life or aircraft integrity, such as high workload phases, time pressure, complex or rapidly changing clearances, dense traffic, degraded visibility, abnormal but controlled procedures, and minor technical irregularities. Incorporating these contexts would allow situational stress to be modeled as a graded phenomenon, where threat, uncertainty, controllability, and cognitive load vary continuously. In practical terms, sampling a wider set of operational stressors would strengthen external validity and help derive models that perform reliably across routine, abnormal, and emergency conditions.

7.6 A final comment

The cockpit offers a particularly revealing setting for studying human communication: a highly regulated environment, governed by standardized phraseology and tightly timed, where the human element, with its unpredictable complexity, remains both the most critical and the least understood factor. Within this tension between protocol and person, the voice emerges as the natural object of inquiry: at once the technical instrument that executes regulated exchange and the porous channel that, despite itself, lets inner states, intentions, and cognitive resources show through.

I conclude this work with the hope of having contributed an informed perspective—one piece in a scientific conversation that must, by its very nature, remain plural and collaborative. The challenge before us is fundamentally interdisciplinary. Engineering, linguistics, human factors, and psychology are called to construct a shared lexicon, unified procedures, and transparent validation criteria. Without this common foundation, we risk dissipating energy by perpetually *"reinventing the wheel"*.

At the heart of this collective endeavor lies a question both simple and profound: What does a voice truly reveal about its speaker, and how can we capture its most authentic meaning? This thesis represents my first deeply felt contribution to that inexhaustible question.

REFERENCES

Abbas, A., Hansen, B. J., Koesmahargyo, V., Yadav, V., Rosenfield, P. J., Patil, O., ... & Galatzer-Levy, I. R. (2022). Facial and vocal markers of schizophrenia measured using remote smartphone assessments: observational study. *JMIR formative research, 6*(1), e26276.

Abbas, A., Sauder, C., Yadav, V., Koesmahargyo, V., Aghjayan, A., Marecki, S., ... & Galatzer-Levy, I. R. (2021). Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. *Frontiers in digital health, 3*, 610006.

Abercrombie, D. (1967). Elements of general phonetics. *Edinburgh University Press*, *2*, 257–279.

Abur, D., MacPherson, M. K., Shembel, A. C., & Stepp, C. E. (2023). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load in older adults. *Journal of Voice*, *37*(2), 194–202.

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology, 71*(1), 5–16.

Al Ismail, M., Deshmukh, S., & Singh, R. (2021, June). Detection of COVID-19 through the analysis of vocal fold oscillations. In ICASSP 2021-2021 *IEEE International Conference on Acoustics*, Speech and Signal Processing (ICASSP) (pp. 1035-1039). IEEE.

Alaminos-Torres, A., Martínez-Álvarez, J. R., Martínez-Lorca, M., López-Ejeda, N., & Marrodán Serrano, M. D. (2023). Fatigue, work overload, and sleepiness in a sample of Spanish commercial airline pilots. *Behavioral Sciences, 13*(4), 300.

Albano Leoni, F. (2024). *Voce: il corpo del linguaggio*. Carocci.

Alderson, J. C. (2009). Air safety, language assessment policy, and policy implementation: The case of aviation English. *Annual Review of Applied Linguistics*, *29*, 168–187.

Aldrich, M. S. (2000). Parkinsonism. *Principles and practice of sleep medicine*, *3*, 1051–1057.

Alketbi, F., & Sipos, A. (2025). Enhancing Aviation Safety through Effective English Language Communication under the ICAO Requirements: Regulatory Challenges. *Journal of East Asia & International Law*, *18*(1).

Allred, K. D., & Smith, T. W. (1989). The hardy personality: Cognitive and physiological responses to evaluative threat. Journal of Personality and Social Psychology, 56(2), 257.

Alpert, M., & Schneider, S. J. (1988). Voice-stress measure of mental workload. *NASA. Langley Research Center, Mental-State Estimation, 1987*.

Alsuraykh, N. H., Wilson, M. L., Tennent, P., & Sharples, S. (2019, May). How stress and mental workload are connected. In *Proceedings of the 13th EAI international conference on pervasive computing technologies for healthcare* (pp. 371–376).

Amir, O., Wolf, M., & Amir, N. (2009). A clinical comparison between two acoustic analysis softwares: MDVP and Praat. *Biomedical Signal Processing and Control*, *4*(3), 202–205.

Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *Quarterly Journal of Experimental Psychology*, *71*(3), 622-641.

Apostolopoulos, Y., Sönmez, S., Shattell, M., & Belzer, M. H. (2010). Worksite-induced morbidities of truck drivers in the United States. *AAOHN Journal, 58*(7), 285–296.

Appley, M. H., & Trumbull, R. A. (Eds.). (2012). *Dynamics of stress: Physiological, psychological and social perspectives*. Springer Science & Business Media.

Baker, S. E., Hipp, J., & Alessio, H. (2008). Ventilation and speech characteristics during submaximal aerobic exercise. *Journal of Speech, Language, and Hearing Research*, *51*(5), 1203–1214.

Bakotić, M., & Radošević-Vidaček, B. (2012). Regulation of sleepiness: the role of the arousal system. Arhiv za higijenu rada i toksikologiju, 63(Supplement 1), 23–33.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614.

Baqutayan, S. M. S. (2015). Stress and coping mechanisms: A historical overview. *Mediterranean Journal of Social Sciences, 6*(2), 479–488.

Barker, J. P., & Berthommier, F. (1999, August). Estimation of speech acoustics from visual speech features: A comparison of linear and non-linear models. In AVSP (p. 19).

Bauer, J., Zalkow, F., Müller, M., & Dittmar, C. (2024). Evaluating the impact of prosody feature normalization on the controllability of pitch in speech synthesis. In *Elektronische Sprachsignalverarbeitung 2024, Tagungsband der 35. Konferenz, Regensburg, 6.-8. März 2024* (pp. 188–195). TUDpress.

Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics, 76,* 102928.

Benson, P. (1995). Analysis of the acoustic correlates of stress from an operational aviation emergency. In *Proc. of the ESCA-NATO tutorial and research workshop on speech under stress*, Lisbon, Portugal, September (pp. 14–15).

Biassoni, F., Gnerre, M., Malaspina, E., Di Tella, S., Anzuino, I., Baglio, F., & Silveri, M. C. (2022). How does prosodic deficit impact naïve listeners recognition of emotion? An analysis with speakers affected by Parkinson's disease. *Psychology of Language and Communication*, *26*(1), 102–125.

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., & Berke, G. S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech, Language, and Hearing Research*, *39*(1), 126–134.

Boersma, P., and V. van Heuven. 2001. “Speak and unSpeak With PRAAT.” *Glot International* 5, no. 9/10: 341–347.

Bookstein, F. L. (1982). Foundations of morphometrics. *Annual Review of Ecology and Systematics*, *13*, 451–470.

Boril, H., Grézl, F., & Hansen, J. H. (2011, August). Front-End Compensation Methods for LVCSR Under Lombard Effect. In *INTERSPEECH* (pp. 1257–1260).

Boyer, S., Paubel, P. V., Ruiz, R., El Yagoubi, R., & Daurat, A. (2018). Human voice as a measure of mental load level. *Journal of Speech, Language, and Hearing Research*, *61*(11), 2722–2734.

Bredin, H. (2017, August). pyannote. metrics: A Toolkit for Reproducible Evaluation, Diagnostic, and Error Analysis of Speaker Diarization Systems. In INTERSPEECH (pp. 3587–3591).

Brenner, M., Shipp, T., Doherty, E. T., and Morrissey, P. (1985). “Voice measures of psychological stress: laboratory and field data,” in Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control, eds J. R. Titze and R. C. Scherer (Denver, CO: The Denver Center for the Performing Arts),239–248.

Brenner, M., & Shipp, T. (1988). Voice stress analysis. *NASA. Langley Research Center, Mental-State Estimation, 1987*.

Brenner, M., Doherty, E. T., and Shipp, T. (1994). Speech measures indicating workload demand. Aviat. Space Environ. Med. 65, 21–26.

Brunswik E. 1952. *The Conceptual Framework of Psychology*. University of Chicago Press.

Buchanan, T. W., Laures-Gore, J. S., & Duff, M. C. (2014). Acute stress reduces speech fluency. *Biological psychology*, *97*, 60-66.

Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Gustav Fischer.

Cahill, J., Cullen, P., & Gaynor, K. (2020). Interventions to support the management of work-related stress (WRS) and wellbeing/mental health issues for commercial pilots. *Cognition, Technology & Work*, *22*(3), 517–547.

Cañas, J. J., Muñoz-de-Escalona, E., Frutos, P. L. D., Rodríguez, R., & Celorrio, F. (2022). Estimation of air traffic controllers' fatigue based on the analysis of the human voice's fundamental frequency. *International journal of human factors and ergonomics*, *9*(4), 311–327.

Carding, P. N., Wilson, J. A., MacKenzie, K., & Deary, I. J. (2009). Measuring voice outcomes: state of the science review. *The Journal of Laryngology & Otology*, *123*(8), 823–829.

Carskadon, M. A., & Dement, W. C. (1979). Effects of total sleep loss on sleep tendency. *Perceptual and motor skills*, *48*(2), 495–506.

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality & Social Psychology, 56*,267–283.

Casper, J. K., & Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins.

Causse, M., Deniel, J., Schwartz, F., Duchevet, A., Matton, N., Imbert, J. P., & Cegarra, J. (2025). Cognitive incapacitation in aviation: a narrative review. *Theoretical Issues in Ergonomics Science*, 1–19.

Chenausky, K., MacAuslan, J., & Goldhor, R. (2011). Acoustic analysis of PD speech. *Parkinson's Disease*, *2011*(1), 435232.

Cho, S. W., Yin, C. S., Park, Y. B., & Park, Y. J. (2011). Differences in self-rated, perceived, and acoustic voice qualities between high-and low-fatigue groups. *Journal of Voice*, *25*(5), 544–552.

Ciceri, M. R., & Anolli, L. M. (2000). *La voce delle emozioni. Verso una semiosi della comunicazione vocale non-verbale delle emozioni*. Franco Angeli.

Congleton, J. J., Jones, W. A., Shiflett, S. G., Mcsweeney, K. P., & Huchingson, R. D. (1997). An evaluation of voice stress analysis techniques in a simulated AWACS environment. *International Journal of Speech Technology*, *2*(1), 61-69.

Cooper, C. L., & Payne, R. (1980). *Current concerns in occupational stress*. John Wiley & Sons Ltd.

Corti, M. (1993). Geometric morphometrics: an extension of the revolution. *Trends in Ecology & Evolution, 8*, 302–303.

Ćosić, K., Popović, S., Šarlija, M., Mijić, I., Kokot, M., Kesedžić, I., ... & Zhang, Q. (2019). New tools and methods in selection of air traffic controllers based on multimodal psychophysiological measurements. *Ieee Access*, *7*, 174873–174888.

Cuevas, H. M. (2003, October). The pilot personality and individual differences in the stress response. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 47, No. 9, pp. 1092–1096). Sage CA: Los Angeles, CA: SAGE Publications.

Curcio, G., Casagrande, M., & Bertini, M. (2001). Sleepiness: Evaluating and quantifying methods. *International Journal of Psychophysiology, 41*(3), 251–263.

Dahl, K. L., & Stepp, C. E. (2023). Effects of cognitive stress on voice acoustics in individuals with hyperfunctional voice disorders. *American Journal of Speech-Language Pathology*, *32*(1), 264–274.

Darwin, C. (1955). *Expression of the emotions in man and animals*, Philosophical Library.

De Mauro, T. (1994). *Capire le parole*. Laterza.

de Vasconcelos, C. A., Vieira, M. N., Kecklund, G., & Yehia, H. C. (2019). Speech analysis for fatigue and sleepiness detection of a pilot. *Aerospace medicine and human performance*, *90*(4), 415–418.

DeHoff, M. C., & Cusick, S. K. (2018). Mental health in commercial aviation-depression & anxiety of pilots. *International Journal of Aviation, Aeronautics, and Aerospace*, *5*(5), 5.

Deliyski, D. D., Evans, M. K., & Shaw, H. S. (2005). Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice*, *19*(2), 176–186.

Demenko, G., & Jastrzębska, M. (2012). Analysis of natural speech under stress. *Acta Physica Polonica A*, *121*(1A).

Depestele, F. (2023). *"You sound stressed!" The Impact of Psychosocial Stress on Speech Parameters* (Doctoral dissertation, Ghent University).

Desmond, P. A., & Hancock, P. A. (2000). Active and passive fatigue states. In *Stress, workload, and fatigue* (pp. 455–465). CRC Press.

Di Salle, F., Esposito, F., Scarabino, T., Formisano, E., Marciano, E., Saulino, C., ... & Seifritz, E. (2003). fMRI of the auditory system: understanding the neural basis of auditory gestalt. *Magnetic resonance imaging*, *21*(10), 1213–1224.

Diaz-Piedra, C., Gomez-Milan, E., & Di Stasi, L. L. (2019). Nasal skin temperature reveals changes in arousal levels due to time on task: An experimental thermal infrared imaging study. Applied Ergonomics, 81, 102870.

Diepeveen, H., van Miltenburg, M., van Drongelen, A., van den Oever, F., & van Dijk, H. (2021, May). Fatigue-Indicator in Operational Settings: Vocal Changes. In *Congress of the International Ergonomics Association* (pp. 128–135). Cham: Springer International Publishing.

Dinges, D. F., & Broughton, R. J. (1989). Sleep attacks, naps, and sleepiness in medical sleep disorders. Sleep and Alertness: Chronobiological, Behavioral and Medical Aspects of Napping. Raven Press.

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America*, 67(1), 253–261.

Drayton, J., & Coxhead, A. (2023). The development, evaluation and application of an aviation radiotelephony specialised technical vocabulary list. *English for Specific Purposes*, *69*, 51–66.

Duffy, J. R. (2000). Motor speech disorders: Clues to neurologic diagnosis. In Parkinson's Disease and Movement Disorders: Diagnosis and Treatment Guidelines for the Practicing Physician (pp. 35-53). Humana Press.

Eden, G., & Inbar, G. F. (1978). Physiological model analysis of involuntary human-voice tremor. *Biological cybernetics*, *30*(3), 179–185.

Engeström, R. (1995). Voice as communicative action. *Mind, culture, and activity*, *2*(3), 192-215.

Ekman, P. (1992). Facial expressions of emotion: an old controversy and new findings. *Philosophical transactions of the royal society of London. Series B: Biological Sciences*, *335*(1273), 63–69.

Estival, D., Prado, M., & Ishihara, N. (2023). Not using standard phraseology: Misunderstandings and delays. *Applied Linguistics Papers*, *27*(2), 4–28.

Farris, C., & Molesworth, B. (2016). Communications between air traffic control and pilots. In *Aviation English* (pp. 92–110). Routledge.

Finch, M. I., & Stedmon, A. W. (1998). The complexities of stress in the operational military environment. *Contemporary Ergonomics*, 388–392.

Fink, G. (Ed.). (2016). *Stress: Concepts, Cognition, Emotion, and Behavior: Handbook of Stress Series, Volume 1* (Vol. 1). Academic Press.

Fornette, M. P., Bardel, M. H., Lefrançois, C., Fradin, J., Massioui, F. E., & Amalberti, R. (2012). Cognitive-adaptation training for improving performance and stress management of air force pilots. *The International Journal of Aviation Psychology*, *22*(3), 203–223.

Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*(1), 25-50.

Gaillard, A. W. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, *36*(9), 991–1005.

Gangl, E. C. (2006). Evolution from analog to digital integration in aircraft avionics-a time of transition. *IEEE Transactions on Aerospace and Electronic Systems*, *42*(3), 1163–1170.

Gao, X., Ma, K., Yang, H., Wang, K., Fu, B., Zhu, Y., ... & Cui, B. (2022). A rapid, non-invasive method for fatigue detection based on voice information. Frontiers in Cell and Developmental Biology, 10, 994001.

Gaur, S., Kalani, P., & Mohan, M. (2024). Harmonic-to-noise ratio as speech biomarker for fatigue: K-nearest neighbour machine learning algorithm. *Medical Journal Armed Forces India*, 80, S120-S126.

Geacăr, C. M. (2010, September). *Reducing pilot/ATC communication errors using voice recognition*. In *Proceedings of ICAS* (Vol. 2010).

Ghasemi, F., Beversdorf, D. Q., & Herman, K. C. (2024). Stress and stress responses: A narrative literature review from physiological mechanisms to intervention approaches. *Journal of Pacific Rim Psychology, 18*, 18344909241289222.

Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: a review. *Journal of voice*, *27*(3), 390-e21.

Gnerre, M., Malaspina, E., Di Tella, S., Anzuino, I., Baglio, F., Silveri, M. C., & Biassoni, F. (2023). Vocal emotional expression in Parkinson's disease: roles of sex and emotions. *Societies*, *13*(7), 157.

Gnerre, M., & Biassoni, F. (2024). Marcatori vocali per la detezione di stati di stress nel contesto dell'aviazione: Verso un nuovo framework di analisi. In F. Biassoni (Ed.), *Il fattore umano in aviazione: Sfide e frontiere in una prospettiva interdisciplinare* (pp. 71–91). EDUCatt. https://hdl.handle.net/10807/314099

Godin, K. W., & Hansen, J. H. (2008, September). Analysis and perception of speech under physical task stress. In *INTERSPEECH* (pp. 1674–1677).

Goldstein, D. S., & Kopin, I. J. (2007). Evolution of concepts of stress. *Stress, 10*(2), 109-120.

Gopalan, K. (2001, May). On the effect of stress on certain modulation parameters of speech. In 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221) (Vol. 1, pp. 101-104). IEEE.

Greeley, H. P., Friets, E., Wilson, J. P., Raghavan, S., Picone, J., & Berg, J. (2006, August). Detecting fatigue from voice using speech recognition. In *2006 IEEE International Symposium on signal processing and information technology* (pp. 567–571). IEEE.

Griffin, G. R., & Williams, C. E. (1987). The effects of different levels of task complexity on three vocal measures. *Aviation, space, and environmental medicine*, *58*(12), 1165-1170.

Hafeez, I., Yingjun, Z., Hafeez, S., Mansoor, R., & Rehman, K. U. (2019). Impact of workplace environment on employee performance: mediating role of employee health. *Business, Management and Economics Engineering*, *17*(2), 173–193.

Hagmüller, M., Rank, E., & Kubin, G. (2006). Evaluation of the human voice for indications of workload-induced stress in the aviation environment. *EEC Note*, *18*(06).

Hall, A., Kawai, K., Graber, K., Spencer, G., Roussin, C., Weinstock, P., & Volk, M. S. (2021). Acoustic analysis of surgeons' voices to assess change in the stress response during surgical in situ simulation. *BMJ Simulation & Technology Enhanced Learning, 7*(6), 471.

Hansen, J. H., Swail, C., South, A. J., Moore, R. K., Steeneken, H., Cupples, E. J., ... & Verlinde, P. (2000). The impact of speech under 'stress' on military speech technology. *NATO Project Report*.

Hansen, J. H., & Patil, S. (2007). Speech under stress: Analysis, modeling and recognition. In *Speaker classification I: Fundamentals, features, and methods* (pp. 108-137). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX Task Load Index: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati Eds., *Human mental workload Advances in Psychology, Vol. 52, pp. 139–183*.

Hecker, M. H., Stevens, K. N., von Bismarck, G., & Williams, C. E. (1968). Manifestations of task-induced stress in the acoustic speech signal. The Journal of the Acoustical Society of America, 44(4), 993-1001.

Hirose, C., Akamatsu, N., & Domen, K. (1992). Formulas for the analysis of the surface SFG spectrum and transformation coefficients of cartesian SFG tensor components. *Applied Spectroscopy*, *46*(6), 1051–1072.

Holambe, R. S., & Deshpande, M. S. (2012). *Advances in non-linear modeling for speech processing.* Springer Science & Business Media.

Hollien, H. (1980). Vocal indicators of psychological stress. *Annals of the New York Academy of Sciences*, *347*(1), 47–72.

Hollien, H., Geison, L., & Hicks Jr, J. W. (1987). Voice stress evaluators and lie detection. *Journal of Forensic Sciences, 32*(2), 405–418.

Horvath, F. (1982). Detecting deception: The promise and the reality of voice stress analysis. *Journal of Forensic Sciences, 27*(2), 340–351.

Hu, X., & Lodewijks, G. (2020). Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of Safety Research, 72,* 173–187.

Huang, Y., & Prahl, A. (2022, August). Linguistic Indicators of Success in Aviation Emergencies: A Cockpit Voice Recorder (CVR) Investigation. In *IRC-SET 2021: Proceedings of the 7th IRC Conference on Science, Engineering and Technology, August 2021, Singapore* (pp. 307–317). Singapore: Springer Nature Singapore.

Huang, Z., Tang, W., Tian, Q., Huang, T., & Li, J. (2024). Air traffic controller fatigue detection based on facial and vocal features using long short-term memory. *IEEE Access*, *12*, 56663-56682.

Huttunen, K. H., Keränen, H. I., Pääkkönen, R. J., Päivikki Eskelinen-Rönkä, R., & Leino, T. K. (2011) (a). Effect of cognitive load on articulation rate and formant frequencies during simulator flights. *The Journal of the Acoustical Society of America*, *129*(3), 1580–1593.

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011) (b). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. *Applied ergonomics*, *42*(2), 348–357.

IATA, I. (2018). *Controlled Flight into Terrain Accident Analysis Report—2008–2017 Data*.

ICAO (2004). *Manual on the implementation of ICAO language proficiency requirements* (ICAO Doc 9835, 1st edn). Chicago, International Civil Aviation Organization.

ICAO (2010). *Manual on the implementation of ICAO language proficiency requirements* (ICAO Doc 9835, 2nd edn). Chicago, International Civil Aviation Organization.

Isserlin, M. (1925). Psychologisch-phonetische untersuchungen. II. Mitteilung. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, *94*(1), 437–448.

Jang, R., Molesworth, B. R., Burgess, M., & Estival, D. (2014). Improving communication in general aviation through the use of noise cancelling headphones. *Safety science*, *62*, 499–504.

Johannes, B., Salnitski, V. P., Gunga, H. C., & Kirsch, K. (2000). Voice stress monitoring in space-possibilities and limits. *Aviation Space and Environmental Medicine*, *71*(9; PART 2), A58–A65.

Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C. A., Thomas, J. L., ... & Gerzer, R. (2007). Non-linear function model of voice pitch dependency on physical and mental load. *European journal of applied physiology*, *101*(3), 267–276.

Johns, M. W. (2010). A new perspective on sleepiness. *Sleep and biological rhythms*, *8*(3), 170–179.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological bulletin*, *129*(5), 770.

Juslin, P. N., & Scherer, K. R. (2008). Speech emotion analysis. *Scholarpedia*, *3*(10), 4240.

Kalia, A., Boyer, M., Fagherazzi, G., Bélisle-Pipon, J. C., & Bensoussan, Y. (2025). Master protocols in vocal biomarker development to reduce variability and advance clinical precision: a narrative review. *Frontiers in Digital Health*, *7*, 1619183.

Kappen, M., Van Der Donckt, J., Vanhollebeke, G., Allaert, J., Degraeve, V., Madhu, N., ... & Vanderhasselt, M. A. (2022). Acoustic speech features in social comparison: how stress impacts the way you sound. *Scientific Reports*, *12*(1), 22022.

Karl, J. P., Hatch, A. M., Arcidiacono, S. M., Pearce, S. C., Pantoja-Feliciano, I. G., & Soares, J. W. (2018). Effects of psychological, environmental and physical stressors on the gut microbiota. *Frontiers in microbiology*, *9*, 372026.

Kelly, D., & Efthymiou, M. (2019). An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. *Journal of safety research*, *69*, 155–165.

Keränen, H., Väyrynen, E., Pääkkönen, R., Leino, T., Kuronen, P., Toivanen, J., & Seppänen, T. (2004). Prosodic features of speech produced by military pilots during demanding tasks. In *Proceedings of the Fonetiikan Päivät Conference* (pp. 88–91).

Khan, M., Sondhi, S., Vijay, R., Sharma, S. K., & Salhan, A. K. (2015). Fundamental Frequency of Voice under Normobaric and Hypobaric Hypoxia. *Indian Journal of Aerospace Medicine*, *59*(2), 37–43.

Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of human factors causations in commercial air transport accidents and incidents: From 2000–2016. *Progress in Aerospace Sciences, 99*, 1–13.

Kikuchi, H., & Ogawa, T. (2018). Voice analysis researches and application in the JASDF. *Aeromedical Laboratory Reports, 58*(2), 9–15.

Kleitman, N. (1963). *Sleep and Wakefulness*. University of Chicago Press.

Knoll, M. A., & Costall, A. (2015). Characterising F (0) contour shape in infant-and foreigner-directed speech. *Speech Communication, 66*, 231–243.

Komaroff, A. L., & Buchwald, D. (1991). Symptoms and signs of chronic fatigue syndrome. *Reviews of Infectious Diseases*, *13*(Supplement_1), S8–S11.

König, A., Riviere, K., Linz, N., Lindsay, H., Elbaum, J., Fabre, R., ... & Robert, P. (2021). Measuring stress in health professionals over the phone using automatic speech analysis during the COVID-19 pandemic: observational pilot study. *Journal of medical Internet research*, *23*(4), e24191.

Kouba, P., Šmotek, M., Tichý, T., & Kopřivová, J. (2023). Detection of air traffic controllers' fatigue using voice analysis-An EEG validation study. *International Journal of Industrial Ergonomics*, *95*, 103442.

Krajewski, J., & Kröger, B. J. (2007, August). Using prosodic and spectral characteristics for sleepiness detection. In *Proceedings of Interspeech 2007: 8th Annual Conference of the International Speech Communication Association* pp. 27–31. ISCA.

Krajewski, J., Schnupp, T., Heinze, C., Schnieder, S., Laufenberg, T., Sommer, D., & Golz, M. (2014). A phonetic approach for detecting sleepiness from speech in simulated Air Traffic Controller-communication. *D. d. Waard et al.(Hg.): Human Factors of Systems and Technology. Maastricht*, 147–155.

Krüger, H.-P., Schulz, E., Magerl, H., Hein, P. M., Hilsenbeck, T., & Vollrath, M. (1996). *Medikamenten- und Drogennachweis bei verkehrsunauffälligen Fahrern* (Berichte der Bundesanstalt für Straßenwesen, Reihe M: Mensch und Sicherheit, Heft M 60). Bundesanstalt für Straßenwesen.

Kupriyanov, R., & Zhdanov, R. (2014). The eustress concept: problems and outlooks. *World Journal of Medical Sciences*, *11*(2), 179–185.

Kurniawan, H., Maslov, A. V., & Pechenizkiy, M. (2013, June). Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems CBMS 2013* pp. 209–214. IEEE.

Kuroda, I., Fujiwara, O., Okamura, N., & Utsuki, N. (1976). Method for determining pilot stress through analysis of voice communication. *Aviation, Space, and Environmental Medicine, 47*(5), 528–533.

Lan, C., Hui, P. L., Xu, W., & Mok, P. (2019). Revisiting acoustic markers of sarcasm in Cantonese. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)* (pp. 77–81).

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143-R145.

Laver, J. (1975). Individual Features in Voice Quality. Ph.D. thesis, University of Edinburgh.

Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. *Social markers in speech*, *1*, 32.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Publishing Company.

Le Fevre, M., Matheny, J., & Kolt, G. S. (2003). Eustress, distress, and interpretation in occupational stress. *Journal of Managerial Psychology*, *18*(7), 726–744.

Lee, S., & Kim, J. K. (2018). Factors contributing to the risk of airline pilot fatigue. *Journal of Air Transport Management*, *67*, 197–207.

Lempereur, I., & Lauri, M. A. (2006). The psychological effects of constant evaluation on air line pilots: An exploratory study. *The International Journal of Aviation Psychology*, *16*(1), 113–133.

Leoni, F. A. (2001). Il ruolo dell'udito nella comunicazione linguistica. Il caso della prosodia. *Italian Journal of Linguistics*, *13*, 45–68.

Li, G., Baker, S. P., Grabowski, J. G., & Rebok, G. W. (2001). Factors associated with pilot error in aviation crashes. *Aviation, Space, and Environmental medicine*, *72*(1), 52–58.

Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A., Leong, K. M., & Newman, J. D. (2007, April). Stress and emotion classification using jitter and shimmer features. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07* (Vol. 4, pp. IV-1081). IEEE.

Li, Y., & He, J. (2024). A Review of Strategies to Detect Fatigue and Sleep Problems in Aviation: Insights from Artificial Intelligence. *Archives of Computational Methods in Engineering*, 1-18.

Lichstein, K. L., Means, M. K., Noe, S. L., & Aguillard, R. N. (1997). Fatigue and sleep disorders. *Behaviour Research and Therapy*, *35*(8), 733–740.

Lindgren, T., Andersson, K., & Norbäck, D. (2006). Perception of cockpit environment among pilots on commercial aircraft. *Aviation, Space, and Environmental Medicine*, *77*(8), 832–837.

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. Routledge.

Liu, H., & Guo, W. (2025). Effectiveness of AI-Driven Vocal Art Tools in Enhancing Student Performance and Creativity. *European Journal of Education*, *60*(1), e70037.

Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *The Journal of the Acoustical Society of America*, *93*(5), 2962–2973.

Lock, A. M., Bonetti, D. L., & Campbell, A. D. K. (2018). The psychological and physiological health effects of fatigue. *Occupational medicine*, *68*(8), 502–511.

Lu, S., Wei, F., & Li, G. (2021). The evolution of the concept of stress and the framework of the stress system. *Cell stress*, *5*(6), 76.

Luig, J., & Sontacchi, A. (2014). A speech database for stress monitoring in the cockpit. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *228*(2), 284–296.

Luzzani, G., Buraioli, I., Demarchi, D., & Guglieri, G. (2024). A review of physiological measures for mental workload assessment in aviation: A state-of-the-art review of mental workload physiological assessment methods in human-machine interaction analysis. *The Aeronautical Journal, 128*(1323), 928–949.

Lyons, J. (1977) *Semantics*. 2 vols., Cambridge University Press.

Lyssakov, N., & Lyssakova, E. (2019, June). Human factor as a cause of aircraft accidents. In *II International Scientific-Practical Conference" Psychology of Extreme Professions"(ISPCPEP 2019)* (pp. 130–132). Atlantis Press.

MacDonald, W. (2003). The impact of job demands and workload on stress and fatigue. *Australian Psychologist, 38*(2), 102–117.

MacLeod, N., Krieger, J., & Jones, K. E. (2013). Geometric morphometric approaches to acoustic signal analysis in mammalian biology. *Hystrix, 24*(1), 110.

Magnani, S., & Fussi, F. (2021). *Ascoltare la voce: Itinerario percettivo alla scoperta delle qualità della voce*. Franco Angeli.

Magnusdottir, E. H., Johannsdottir, K. R., Majumdar, A., & Gudnason, J. (2022). Assessing cognitive workload using Cardiovascular measures and voice. *Sensors*, *22*(18), 6894.

Mahmoud, M. S. B., Guerber, C., Larrieu, N., Pirovano, A., & Radzik, J. (2014). *Aeronautical air-ground data link communications*. John Wiley & Sons.

Maina, P. A. W., & Zhang, S. (2023). Pilot fatigue detection via speech analysis, electrocardiogram and photoplethysmogram. *United International Journal of Engineering and Sciences*, *4*(1), 1–23.

Majumdar, A., & Ochieng, W. Y. (2002). Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload. *Transportation Research Record*, *1788*(1), 58–69.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological psychology*, *121*, 62–73.

Marqueze, E. C., Nicola, A. C. B., Diniz, D. H. M., & Fischer, F. M. (2017). Working hours associated with unintentional sleep at work among airline pilots. *Revista de Saúde Pública, 51*, 61.

Martin, V. P., Rouas, J. L., & Philip, P. (2024). Automatic detection of sleepiness-related symptoms and syndromes using voice and speech biomarkers. *Biomedical Signal Processing and Control*, *91*, 105989.

Martins, A. P. (2016). A review of important cognitive concepts in aviation. *Aviation, 20*(2), 65-84.

Martin, V. P., Rouas, J. L., Micoulaud-Franchi, J. A., Philip, P., & Krajewski, J. (2021). How to design a relevant corpus for sleepiness detection through voice?. *Frontiers in Digital Health, 3*, 686068.

Masi, G., Amprimo, G., Ferraris, C., & Priano, L. (2023). Stress and workload assessment in aviation—A narrative review. *Sensors*, *23*(7), 3556.

Mason, J. W. (1975). A historical view of the stress field. *Journal of human stress*, *1*(2), 22-36.

Matura, L. A., Malone, S., Jaime-Lara, R., & Riegel, B. (2018). A systematic review of biological mechanisms of fatigue in chronic illness. *Biological research for nursing*, *20*(4), 410-421.

Meier, M., Borsky, M., Magnusdottir, E. H., Johannsdottir, K. R., & Gudnason, J. (2016, October). Vocal tract and voice source features for monitoring cognitive workload. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 000097-000102). IEEE.

Meijman, T. F., & Mulder, G. (2013). Psychological aspects of workload. In P. J. D. Drenth, H. Thierry, & C. J. de Wolff (Eds.), *Handbook of work and organizational psychology* (pp. 5–33). Psychology Press.

Mélan, C., & Cascino, N. (2022). Effects of a modified shift work organization and traffic load on air traffic controllers' sleep and alertness during work and non-work activities. *Applied Ergonomics, 98*, 103596.

Mendoza, E., & Carballo, G. (1998). Acoustic analysis of induced vocal stressby means of cognitive workload tasks. *Journal of Voice*, *12*(3), 263-273.

Menne, F., Lindsay, H., Tröger, J., Paulmann, S., König, A., Steinbach, N., ... & Schmidt-Kassow, M. (2025). Voice as Objective Biomarker of Stress: Association of Speech Features and Cortisol. *Acta Neuropsychiatrica*, 1−39.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & TP Group. (2009). Linee guida per il reporting di revisioni sistematiche e meta-analisi: il PRISMA Statement. *PLoS Med*, *6*(7), e1000097.

Molesworth, B. R., & Estival, D. (2015). Miscommunication in general aviation: The influence of external factors on communication errors. *Safety science*, *73*, 73−79.

Morris, C. H., & Leung, Y. K. (2006). Pilot mental workload: how well do pilots really perform? *Ergonomics*, *49*(15), 1581−1596.

Mosier, K. L., Rettenmaier, P., McDearmid, M., Wilson, J., Mak, S., Raj, L., & Orasanu, J. (2013). Pilot–ATC communication conflicts: Implications for NextGen. *The International Journal of Aviation Psychology*, *23*(3), 213−226.

Murray, I. R., Baber, C., & South, A. (1996). Towards a definition and working model of stress and its effects on speech. *Speech Communication*, *20*(1-2), 3-12.

Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Abbott, K. V. (2015). Vocal Fatigue Index (VFI): development and validation. *Journal of Voice*, *29*(4), 433-440.

Nguyen, D. D., & Madill, C. (2023). Auditory-perceptual parameters as predictors of voice acoustic measures. *Journal of Voice*.

NoiseReduce, Ver. 2.0.1, Timothy Roberts, [Online]. Available: https://pypi.org/project/noisereduce/. [Accessed: 16-May2025].

O'Shaughnessy, D. (2025). Review of Automatic Estimation of Emotions in Speech. *Applied Sciences, 15*(10), 5731.

Ogden, C. K., & Richards, I. A. (1923). *The Meaning of Meaning.* Harcourt Brace Jovanovich.

Olson, K. (2007). A new way of thinking about fatigue: A reconceptualization. *Oncology Nursing Forum*, 34(1), 93–99.

Onions, C. T. (1959). *The Shorter Oxford English Dictionary.* Oxford University Press.

Orlikoff, R. F., & Baken, R. J. (1990). Consideration of the relationship between the fundamental frequency of phonation and vocal jitter. *Folia Phoniatrica et Logopaedica*, *42*(1), 31-40.

Özmen, S., Hamzaoui, R., & Chen, F. (2024). Survey of IP-based air-to-ground data link communication technologies. *Journal of Air Transport Management*, *116*, 102579.

Pacak, K., Palkovits, M., Yadid, G., Kvetnansky, R., Kopin, I. J., & Goldstein, D. S. (1998). Heterogeneous neurochemical responses to different stressors: a test of Selye's doctrine of nonspecificity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *275*(4), R1247-R1255.

Parasuraman, R., Sheridan, T., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Cognitive Engineering and Decision Making, 2,* 140–160.

Parsa, V., Jamieson, D. G., & Pretty, B. R. (2001). Effects of microphone type on acoustic measures of voice. *Journal of Voice*, *15*(3), 331–343.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98.

Patil, V. P., Nayak, K. K., & Saxena, M. (2013). Voice stress detection. *International Journal of Electrical, Electronics and Computer Engineering, 2*(2), 148-154.

Peirce, C. P. (1965). Basic concepts of Peircean sign theory. *Semiotics*, *1*–10.

Perdrizet, G. A. (1997). Hans Selye and beyond: responses to stress. *Cell stress & chaperones*, *2*(4), 214.

Perez, S. I., Bernal, V., & Gonzalez, P. N. (2006). Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of anatomy*, *208*(6), 769–784.

Pesina, S., & Solonchak, T. (2014). The sign in the communication process. In *International Science Conference: International Conference on Language and Technology* (June 19–20) (International Science Index, Vol. 8, No. 6, Part XI, pp. 1021–1029). World Academy of Science, Engineering and Technology.

Petrie, K. J., & Dawson, A. G. (1997). Symptoms of fatigue and coping strategies in international pilots. *International Journal of Aviation Psychology*, 7(3), 251–258.

Phillips, R. O. (2015). A review of definitions of fatigue–And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour, 29*, 48–56.

Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. *The Oxford handbook of voice studies*, 269-300.

Pisanski, K., & Sorokowski, P. (2021). Human stress detection: cortisol levels in stressed speakers predict voice-based judgments of stress. *Perception*, *50*(1), 80–87.

Pisoni, D. B. (1990). Speech perception and production in severe environments. *Indiana University*.

Postma-Nilsenová, M., Holt, E., Heyn, L., Groeneveld, K., & Finset, A. (2016). A case study of vocal features associated with galvanic skin response to stressors in a clinical interaction. *Patient Education and Counseling*, *99*(8), 1349–1354.

Prinzo, O. V. (1998). *An analysis of voice communication in a simulated approach control environment* (No. DOT/FAA/AM-97/17). Civil Aeromedical Institute.

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications: A survey of the literature (*Final report). Federal Aviation Administration.

Prinzo, O. V., & Lieberman, P. (1998). *An Acoustic Analysis of ATC Communication* (No. DOTFAAAM9820). Federal Aviation Administration.

Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *The Journal of the Acoustical Society of America*, *101*(4), 2267–2277.

Rabinov, C. R., Kreiman, J., Gerratt, B. R., & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech, Language, and Hearing Research*, *38*(1), 26–32.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology*, *4*(3), 211–240.

Radford, A., Gao, L., Connor, J., & Wu, J. (2022). *Whisper: Robust speech recognition via large-scale weak supervision* [Computer software]. OpenAI. https://github.com/openai/whisper

Ramalingam, C. S. (1995). *Analysis of non-stationary, multi-component signals with applications to speech*. University of Rhode Island

Reason, J. (1995). Understanding adverse events: human factors. *BMJ Quality & Safety*, *4*(2), 80-89.

Reem, M., Rayhan, S., & Wafa, I. (2024). Impact of Dimension Reduction Techniques on the Accuracy of Speech Emotion Recognition. *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, 454–468.

Rocha, P. C., & Romano, P. S. (2021). The shape of sound: A new R package that crosses the bridge between Bioacoustics and Geometric Morphometrics. *Methods in Ecology and Evolution*, *12*(6), 1115–1121.

Rochette, L., Dogon, G., & Vergely, C. (2023). Stress: eight decades after its definition by hans selye:“stress is the spice of life”. *Brain Sciences*, *13*(2), 310.

Rodero, E. (2011). Intonation and emotion: Influence of pitch levels and contour type on creating emotions. *Journal of Voice, 25*(1), e25–e34.

Roelen, A. L., & Stuut, R. (2016, April). Association of sleep deprivation with speech volume and pitch. In *Ergonomics & Human Factors 2016*.

Rohlf, F. J. (1990). Morphometrics. *Annual Review of ecology and Systematics*, 299–316.

Rohlf, F. J. (1993). Relative warp analysis and an example of its application to mosquito wings. *Contributions to morphometrics*, *8*, 131–159.

Roscoe, A. H. (1978). Stress and workload in pilots. *Aviation, Space, and Environmental Medicine, 49*(4), 630–633.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech communication*, 6(4), 343–352.

Rose, P. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency?. *Speech Communication*, 10(3), 229–247.

Rothkrantz, L. J., Wiggers, P., Van Wees, J. W. A., & van Vark, R. J. (2004, September). Voice stress analysis. In *Text, speech and dialogue* (pp. 449–456). Springer.

Ruiz, R., Clouer, L., & Gunn, A. (1990). Voice analysis to predict the psychological or physical state of a speaker. *Aviation, Space, and Environmental Medicine*, 61(7), 675–676.

Ruiz, R., Absil, E., Harmegnies, B., Legros, C., & Poch, D. (1996). Time-and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication*, *20*(1-2), 111-129.

Ruiz, R., de Hugues, P. P., & Legros, C. (2010). Advanced voice analysis of pilots to detect fatigue and sleep inertia. *Acta Acustica united with Acustica*, *96*(3), 567-579.

Rusz, J., Hlavnička, J., Novotný, M., Tykalová, T., Pelletier, A., Montplaisir, J., ... & Šonka, K. (2021). Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Annals of Neurology*, *90*(1), 62-75.

Rybner, A., Jessen, E. T., Mortensen, M. D., Larsen, S. N., Grossman, R., Bilenberg, N., ... & Fusaroli, R. (2022). Vocal markers of autism: Assessing the generalizability of machine learning models. *Autism Research, 15*(6), 1018–1030.

Saito, I., Fujiwara, O., Utsuki, N., Mizumoto, C., & Arimori, T. (1980). Hypoxia-induced fatal aircraft accident revealed by voice analysis. *Aviation, Space, and Environmental medicine*, *51*(4), 402–406.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2017). Performance measurement in simulation-based training: A review and best practices. *Simulation in Aviation Training*, 393–441.

Sandoval, C., Stolar, M. N., Hosking, S. G., Jia, D., & Lech, M. (2022). Real-time team performance and workload prediction from voice communications. *IEEE Access, 10,* 78484–78492.

Scherer, K. R. (1981). Vocal indicators of stress. In J. K. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 171–187). Grune & Stratton.

Scherer, K. R. (1985). Vocal affect signaling: A comparative approach. In J. S. Rosenblatt, C. Beer, M.-C. Busnel, & P. J. B. Slater (Eds.), *Advances in the study of behavior* (Vol. 15, pp. 189–244). Academic Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165.

Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. *Geneva Studies in Emotion and Communication*, *1*, 1–98.

Scherer, K. R. (1988). Criteria for emotion-antecedent appraisal: A review. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation* (pp. 89–126). Kluwer Academic.

Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., & Bänziger, T. (2002). Acoustic correlates of task load and stress. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)* (pp. 2017–2020).

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113–130.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*(7), 1307–1351.

Scherer, K. R. (2013a). Emotion in action, interaction, music, and speech. In M. A. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship* (pp. 107–139). MIT Press.

Scherer, K. R. (2013b). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language*, *27*(1), 40–58.

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745.

Schewski, L., Doss, M. M., Beldi, G., & Keller, S. (2025). Measuring negative emotions and stress through acoustic correlates in speech: A systematic review. *PLoS One, 20*(7), e0328833.

Schuller, B., Steidl, S., Batliner, A., Schiel, F., & Krajewski, J. (2011). The INTERSPEECH 2011 speaker state challenge. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 2011)* (pp. 3201–3204).

Schuller, B., Batliner, A., Bergler, C., Pokorny, F. B., Krajewski, J., Cychosz, M., ... Schmitt, M. (2019). The INTERSPEECH 2019 computational paralinguistics challenge: Styrian dialects, continuous sleepiness, baby sounds & orca activity. In *Proceedings of the 20th International Conference on Speech and Communication (INTERSPEECH 2019)* (pp. 2378–2382).

Schwarzer, R., Van der Ploeg, H. M., & Spielberger, C. D. (1982). Test anxiety: An overview of theory and research. *Advances in Test Anxiety Research, 1*, 3-9.

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature*, *138*(3479), 32-32.

Selye, H. (1950). Stress and the general adaptation syndrome. *British Medical Journal*, *1*(4667), 1383.

Selye, H. (1951). The general-adaptation-syndrome. *Annual Review of Medicine*, *2*(1), 327-342.

Selye, H. (1974). *Stress without distress.* McClelland Stewart.

Selye, H. (1976). The stress concept. *Canadian Medical Association Journal*, *115*(8), 718.

Shannon, C. E., & Weaver, W. (1949). A mathematical model of communication. *University of Illinois Press*, *11*, 11-20.

Shao, Q., Dong, M., Shen, Z., Yang, R., & Wang, H. (2021). Integrating Ergonomics into Safety Management: A conceptual risk assessment model for tower controllers at multiple altitudes. *IEEE Access*, *9*, 93364-93383.

Shappell, S. and Wiegmann, D. (2001). Applying Reason: The human factors analysis and classification system. *Human Factors and Aerospace Safety*, *1,* 59-86.

Shappell, S. A., Detwiler, C. A., Holcomb, K. A., Hackworth, C. A., Boquet, A. J., & Wiegmann, D. A. (2006). *Human error and commercial aviation accidents: A comprehensive, fine-grained analysis using HFACS* (No. DOTFAAAM0618).

Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep medicine reviews*, *10*(1), 63-76.

Shen, Z., & Wei, Y. (2021). A high-precision feature extraction network of fatigue speech from air traffic controller radiotelephony based on improved deep learning. *ICT Express*, *7*(4), 403-413.

Sigmund, M. (2006). "Introducing the database ExamStress for speech under stress," in Proceedings of the 7th Nordic Signal Processing Symposium, 2006. NORSIG 2006, (Reykjavik: IEEE), 290–293.

Silberstein, D., & Dietrich, R. (2003). Cockpit communication under high cognitive workload. *Communication in high risk environments*, *12*, 9.

Singh, N., Khan, R. A., & Shree, R. (2012). MFCC and prosodic feature extraction techniques: a comparative study. *International Journal of Computer Applications*, *54*(1), 9–13.

Skinner, E. R. (1935). A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft, and loud tones. *Communications Monographs*, *2*(1), 81–137.

Skodda, S., Grönheit, W., & Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease*. PloS one, 7*(2), e32132.

Slice, D. E. (2005). Modern morphometrics. In *Modern morphometrics in physical anthropology* (pp. 1-45). Springer US.

Slice, D. E. (2007). Geometric morphometrics. *Annual Review of Anthropology, 36*(1), 261–281.

Sondhi, S., Khan, M., Vijay, R., Salhan, A. K., & Chouhan, S. (2015). Acoustic analysis of speech under stress. *International Journal of Bioinformatics Research and Applications*, *11*(5), 417–432.

Steptoe, A. (1991). Psychological coping, individual differences, and physiological stress responses. In C. L. Cooper & R. Payne (Eds.), Personality and stress: Individual differences in the stress process (pp. 205–233). Chichester, England: John Wiley & Sons.

Stokes, A. & Kite, K. (1994). *Flight stress: Stress, fatigue, and performance in aviation*. Aldershot, England: Avebury Aviation.

Stokes, A. F., & Kite, K. (2017). *Flight stress: Stress, fatigue and performance in aviation*. Routledge.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *The Journal of the Acoustical Society of America*, *73*(4), 1354–1360.

Strelau, J. (1989). Individual differences in tolerance to stress: The role of reactivity. In C. D. Spielberger, I. G. Sarason, & J. Strelau (Eds.), Stress and anxiety (Vol 12) (pp. 155–166). Hemisphere.

Suppa, A., Costantini, G., Asci, F., Di Leo, P., Al-Wardat, M. S., Di Lazzaro, G., ... & Saggio, G. (2022). Voice in Parkinson's disease: a machine learning study. *Frontiers in Neurology, 13,* 831428.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes*, *23*(3), 451–470.

Tavi, L. (2017). Acoustic correlates of female speech under stress based on/i/-vowel measurements. The International Journal of Speech, Language and the Law, 24(2), 227–241.

Tawari, A., & Trivedi, M. M. (2010). Speech emotion analysis: Exploring the role of context. *IEEE Transactions on multimedia*, *12*(6), 502-509.

Taylor, J. L., O'Hara, R., Mumenthaler, M. S., Rosen, A. C., & Yesavage, J. A. (2005). Cognitive ability, expertise, and age differences in following air-traffic control instructions. *Psychology and Aging*, 20(1), 117.

Tenney, J., & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, *24*(2), 205–241.

Terenzi, M., Tempestini, G., & Di Nocera, F. (2024). Air Traffic Controllers' Rostering: Sleep Quality, Vigilance, Mental Workload, and Boredom: A Report of Two Case Studies. *Aerospace*, *11*(6), 495.

Uclés, N. R., & García, J. M. C. (2014, May). Relationship between workload and duration of ATC voice communications. In *6th International Conference on Research in Air Transportation, Istanbul, Turkey*.

United States National Transportation Safety Board. (2001). *Special Investigation Report: Vehicle-and Infrastructure-based Technology for the Prevention of Rear-end Collisions*. National Transportation Safety Board.

Vagner, J., Ćekanova, A., Szabo, S., & Rozenberg, R. (2018). Fatigue and stress factors among aviation personnel. *Acta Avionica*, *20*(2), 39.

van den Broek, E. L. (2003). A stress marker in speech. *Toegepaste Taalwetenschap in Artikelen*, *69*(1), 143-153.

Van Mersbergen, M., & Lanza, E. (2019). Modulation of relative fundamental frequency during transient emotional states. *Journal of Voice*, *33*(6), 894-899.

Van Puyvelde, M., Neyt, X., McGlone, F., & Pattyn, N. (2018). Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in Psychology*, *9*, 1994.

van Rijn, P., Poeppel, D., & Larrouy-Maestri, P. (2023). Contribution of pitch measures over time to emotion classification accuracy. Preprint.

Van Segbroeck, M., Travadi, R., Vaz, C., Kim, J., Black, M. P., Potamianos, A., & Narayanan, S. S. (2014, September). Classification of cognitive load from speech using an i-vector framework. In *Interspeech* (pp. 751-755).

Verma, V., Benjwal, A., Chhabra, A., Singh, S. K., Kumar, S., Gupta, B. B., ... & Chui, K. T. (2023). A novel hybrid model integrating MFCC and acoustic parameters for voice disorder detection. *Scientific Reports*, *13*(1), 22719.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics*, 243-273.

Vollrath, M. (1994). Automatic measurement of aspects of speech reflecting motor coordination. *Behavior Research Methods, Instruments, & Computers*, *26*(1), 35-40.

Waaramaa, T., Alku, P., & Laukkanen, A. M. (2006). The role of F3 in the vocal expression of emotions. *Logopedics Phoniatrics Vocology, 31*(4), 153-156.

Whitmore, J., & Fisher, S. (1996). Speech during sustained operations. *Speech Communication*, *20*(1-2), 55–70.

Wickens, C. D., Vidulich, M. A., & Tsang, P. S. (2023). Information processing in aviation. In *Human Factors in Aviation and Aerospace* (pp. 89–139). Academic Press.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The journal of the acoustical society of America*, *52*(4B), 1238–1250.

Wu, Q., Molesworth, B. R., & Estival, D. (2019). An investigation into the factors that affect miscommunication between pilots and air traffic controllers in commercial aviation. *The international journal of aerospace psychology*, *29*(1-2), 53–63.

Xu, L., Ma, S., Shen, Z., & Nan, Y. (2024). Air Traffic Controller Fatigue Detection by Applying a Dual-Stream Convolutional Neural Network to the Fusion of Radiotelephony and Facial Data. *Aerospace*, *11*(2), 164.

Yang, J., Yang, H., Wu, Z., & Wu, X. (2023). Cognitive load assessment of air traffic controller based on SCNN-TransE network using speech data. *Aerospace*, *10*(7), 584.

Yap, T. F., Epps, J., Ambikairajah, E., & Choi, E. H. (2015). Voice source under cognitive load: Effects and classification. *Speech Communication*, *72*, 74–95.

Yin, B., Chen, F., Ruiz, N., & Ambikairajah, E. (2008, March). Speech-based cognitive load monitoring system. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2041–2044). IEEE.

Zhang, Z. (2015). Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control. *The Journal of the Acoustical Society of America*, *137*(2), 898–910.

Zhang, Z., Cummins, N., & Schuller, B. (2017). Advanced data exploitation in speech analysis: An overview. IEEE Signal Processing Magazine, 34(4), 107-129.