ESTIMATING SYMPTOMS AND CLINICAL SIGNS INSTEAD OF DISORDERS: THE PATH TOWARD THE CLINICAL USE OF VOICE AND SPEECH BIOMARKERS IN PSYCHIATRY

Vincent P. Martin[1,2,3], *Jean-Luc Rouas*[2]

[1] DDP Research Unit, Department of Precision Health, LIH, Strassen, Luxembourg
[2] Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
[3] Univ. Bordeaux, CNRS, SANPSY, UMR 6033, CHU Pellegrin, F-33076 Bordeaux, France

ABSTRACT

Despite the continuous innovation in voice biomarkers domain for more than a decade and the apparent need for clinicians to have objective diagnostic tools, no device has yet been implemented in real clinical settings or widely adopted by clinicians. After giving a short overview of the literature, we argue that in addition to the factors usually mentioned in the literature (low performance, database sizes, transparency, etc.), an underestimated but crucial factor preventing the use of such systems is the therapeutic relationship. We also discuss the "objectivity" of such systems, and the place of diagnosis in clinical practice and its conceptual limitations. In order to shape useful and relevant voice biomarkers, we propose to estimate symptoms instead of diagnosis, and draw perspectives related to this paradigm, which will require databases annotated with patients' symptoms rather than only their pathological status.

***Index Terms*—** Voice biomarkers, Psychiatry, Clinical usefulness, Diagnosis criteria, Epistemology

1. INTRODUCTION

Today's gospel when estimating psychiatric disorders using voice is that "Gold-standard diagnostic and assessment tools for depression and suicidality remain rooted, almost exclusively, on the opinion of individual clinicians risking a range of subjective biases." [1] or that "There is an urgency to objectively diagnose, monitor over time, and provide evidence-based interventions for individuals with mental illnesses" [2]. This idea of replacing clinicians' judgment with "objective" algorithmic approaches is not new, particularly with the aim of obtaining a diagnosis independent of the human factors affecting the psychiatrist [3, 4]. A 2007 survey even showed that a majority of psychiatrists (87%) consider their own diagnosis unreliable: factors related to clinicians themselves (e.g., education, bias, and interview style) were the most given explanations for discrepancies between psychiatrists (63.5%), rather than patient characteristics (21.6%) or disorders definitions) (14.9%) [5]. In other words, the main factor determining the diagnosis that is assigned to a patient would not be the patient's complaint but the clinicians themselves.

The need for objective diagnosis in psychiatry thus seems legitimate. Among digital tools, voice analysis is a good candidates to do so: it is a minimally invasive measurement tool, that can collect data passively (e.g. when interacting with a connected device), and is implemented in all smartphones. Moreover, since it is linked to many neuromotor [6] and neurolinguistic [7] mechanisms, voice is sensible to numerous disorders [8]. All these advantages have made it a well-studied tool to detect many psychiatric disorders [2]. However, despite these advantages, voice biomarkers are not used in clinical practice. Why?

To answer this question, we first describe in Section 2 the state of the art of psychiatric disorders detection using voice and speech descriptors based on a systematic review published in 2020, based on a systematic review published in 2020 [2]. Some metrics extracted from the supplementary data of this review are used to support the arguments developed in other sections. In Section 3, we investigate the causes of the absence of such devices in clinical practice, and identify *therapeutic relationship* as an underestimated factor. In Section 4, we discuss the role of diagnosis in clinical practice and the very existence of *objective* systems to reconsider the initial hypothesis that we need systems to *objectively* diagnose psychiatric disorders. Finally, we propose to estimate symptoms instead of diagnosis in Section 5, and draw perspectives related to this new paradigm in Section 6, in order to shape useful and relevant voice biomarkers in digital psychiatry. Finally, we conclude in Section 7.

2. STATE OF THE ART

In a systematic review conducted in 2020, Low et al. have inventoried 127 studies published between 2009 and 2019 focusing on the detection of psychiatric disorders using voice and speech [2]. They identified 8 different pathologies, and among the 127 included studies, 63 (49.8%) were focused on the estimation of Major Depressive Disorder (MDD), 23

This research is funded by the French Research Agency ANR through the axis "Autonom-Health" of the PEPR "Santé Numérique", Grant agreement n°ANR-22-PESN-000X

(18.1%) were about schizophrenia and 21 (16.5%) about Bipolar Disorder (BD), which are among the most prevalent and harmful psychiatric disorders [9]. In the same vein, some recent works have also been focusing on depression [10], schizophrenia [11], bipolar disorders [12] but also borderline trouble [13], anxiety [14], or ADHD [15].

Diagnostic estimation is the most widely studied task in the literature: of the 85 studies that used *machine-learning* approaches (in contrast to studies that only perform statistical tests), 61 (71.8%) focused on pathology detection (binary classification). To do so, databases are annotated in two different ways. On the one hand, the data of 45 (35.4%) of the 127 studies included by Low et al. rely on the diagnosis made during clinical interviews, established on the basis of international classification criteria such as those of the Diagnostic and Statistical Manual of mental disorders (DSM); or hetero-questionnaires (i.e. filled by the clinician, e.g. the MADRS). On the other hand, 78 (61.4%) of the databases are annotated by the score to self-reported medical questionnaires (i.e. filled by patients themselves, e.g. the Patient Health Questionnaire – PHQ).

Given the diversity of studies, databases, and pathologies studied, it would seem that the use of voice in this field is quite successful. So why is it still not used in clinical practice?

3. WHY ARE VOICE AND SPEECH BIOMARKERS NOT WIDELY USED IN CLINICAL SETTINGS?

3.1. Commonly reported explanations

To explain the absence of these devices from clinical usefulness, three factors are commonly reported in the literature:

(i) The performance of systems is not sufficient. Regarding the results obtained in recent years, the absence of voice-based systems in clinical practice does not seem to be a matter of performances that are above clinically useful thresholds in numerous tasks [16];

(ii) Databases are too small. The amount of data in most databases is small due to the cost of collecting it [16], making the model trained on them lack generalization. However, the collection of crowd-recorded databases using smartphones has led to big new databases that open the way to robust and reliable machine learning systems (e.g. n = 9920 in [17]);

(iii) Systems lack transparency. It is sometimes argued that the major obstacle to this clinical implementation is the lack of transparency of the decision made by the classification systems, interfering with the clinicians' confidence in these devices [18]. However, clinicians use everyday tools whose inner workings they do not understand (e.g., electronic thermometers), which they trust: transparency is not the only factor at play [19].

3.2. Therapeutic relationship

In addition to the previous factors, we have identified one crucial but nevertheless very little mentioned in the literature: *therapeutic relationship*. Indeed, in their own words, psychiatrists fear that the use of AI in their practice to deteriorate the *therapeutic relationship* between them and their patients [20]. Indeed, during a clinical interview, clinicians not only identify and categorize the complaints expressed by the patient in order to make a diagnosis, but they establish a real human relationship, during which the trust granted by the patient to the clinician is crucial for the progress and outcome of the therapeutic process [21].

4. A REAL NEED FOR OBJECTIVE DIAGNOSIS?

4.1. A need for *automatic* and *objective* diagnosis?

From a clinical aspect, given the importance of the therapeutic relationship, the usefulness of system output such as "You may have depression", "You have a probability p of having schizophrenia" or "You have bipolar disorder" is questionable, both for clinicians and patients. For clinicians, the only information of the diagnostic is useless (this point is made in a subsequent paragraph). For patients, the delivering of the diagnosis is a delicate step and a key moment in the care of a patient suffering from a mental disorder [22], that should not be delivered by a smartphone application but by a trained professional.

Moreover, one of the main arguments in favor of estimating psychiatric disorders using voice and speech biomarkers is the alleged *objectivity* of such systems, by contrast with the subjectivity of clinicians. However, the very existence of such an *objective* system is questionable: not only do the inherent biases of the databases not allow a universal generalization of the targeted concepts and indirectly reflect the biases of their designers (e.g. [23]), but the very design of these systems relies on engineers/coders, themselves human and subject to biases and constraints [24] – among which the belief that clinicians need objective and automatic diagnostic systems.

4.2. Conceptual limitations of diagnosis annotations

A third and final argument about the limitations of automatic diagnosis estimation using speech biomarkers relies on the limitations of the database annotations themselves.

4.2.1. *Autoquestionnaires*

Regarding autoquestionnaires, they cost nothing, are non-invasive, and measurable anywhere and anytime, which makes them very suitable for data collection [25]. In addition, they do not require clinical supervision during filling, so they are used to collect large databases under ecological conditions [17]. However, these questionnaires suffer from three limitations:

(i) While they could have sufficient accuracy to be implemented in clinical settings, the combination of questionnaire measurement errors and classification model errors may lead to lower than expected disease detection performance [26];

(ii) In their standard clinical practice, clinicians rarely use questionnaires, because of lack of time but also because of a lack of training in the use of these tools [27];

(iii) These questionnaires are medically validated on the basis of their ability to discriminate patients belonging to different diagnostic criteria, and thus have the same limitations.

4.2.2. *Diagnostic criteria*

Similarly, diagnostic criteria such as those used in the Diagnostic and Statistical Manual of Mental Disorders (DSM) have significant limitations.

(i) Symptomatic profiles in diagnostic criteria are very heterogeneous, and collected symptoms are often collected through the scope of the disorders. In a study that measured complete symptomatic profiles (47 symptoms) of 107,349 adults with the ten most common psychiatric disorders diagnosed using DSM-5 criteria, it has been concluded that 'DSM-5 disorder criteria do not separate individuals from random when the complete mental health symptom profile of an individual is considered.' [9]. This heterogeneity questions the concept generalized by models trained on small (databases containing < 2000 patients), in which inter-speaker differences could not be differentiated from intergroup variations.

(ii) Additionally, diagnostic criteria are both unstable through cultures (e.g. the *hikikomori* diagnosis that is specific to the Nippon culture [28]) and unstable through time: with the advancement of scientific knowledge, updated versions of classification reference manuals are regularly published (e.g. DSM-IV in 1994, DSM-IV-TR in 2000, DSM-5 in 2013, DSM-5-TR in 2022). Since collecting this type of data requires both important human and financial investments, annotating databases with diagnostic criteria that are dedicated to some specific populations and/or that which can change in the following years does not seem to be the most sustainable investment.

4.3. But then, what is the purpose of the diagnosis?

From an epistemological point of view, "one of its most important goal is to facilitate communication among clinicians, researchers, administrators and patients [...] by establishing a common language."[29] The diagnosis is thus an object of communication between the different parties involved in the pathology (clinicians, patient, patient's entourage, ...) but also an important element of the recognition of the patient's complaints by a health professional and by society. It is for example necessary for health insurance procedures, and the criteria have sometimes been widened to facilitate these procedures for patients who could present subclinical complaints, to the point that some epidemiological studies have estimated that up to 50% of the world population fulfill mental disorder criteria [30]. Moreover, unlike other medical disciplines, treatment in psychiatry is not formulated on the basis of the diagnosis but on the basis of the patient's symptoms and clinical signs in a transdiagnostic manner [31]. The automatic estimation of diagnoses is therefore of no real clinical use.

5. ESTIMATING SYMPTOMS IS THE KEY

(i) Clinical usefulness of symptoms. We argue that contrary to the approaches based on diagnostic criteria in the literature a *semiological* approach (i.e. based on symptoms and clinical signs) of voice and speech biomarkers in digital psychiatry would give innovative tools to clinicians, who can better measure the different symptomatic dimensions of their patients in ecological conditions, measure the response to the treatments they have provided or estimate relapses early.

(ii) Time and cultural stability. Moreover, while some symptoms are not still diagnosed (e.g. hysteria [32]) and some appear with the evolution of society (e.g. misuse of smartphones [33]), their fundamental nature makes them constant units across cultures and time (e.g. headaches during early antiquity [34]). Consequently, annotating a database with symptoms seems to be a more sustainable approach than annotating them with variable diagnostic criteria.

(iii) Therapeutic relationship and ethical issues. Estimating the signs and symptoms of the studied disorders has the advantage of preserving the therapeutic relationship between patients and clinicians, the latter having the choice to use or not the information given by the digital system but also of evacuating the ethical problems on the stakes of the formulation of diagnosis by machine learning systems [35], since it is the clinician who remains the decision maker.

(iv) Reduction of biases and epistemic injustices In addition to the benefits of the semiological approach, the estimation of signs and symptoms from vocal biomarkers can reduce both clinician and patient biases (e.g. illusory correlation or retrospective prejudice)[36] by providing behavioral data to clinicians, allowing them to refine their decisions based on both the patient's narrative and objective behavioral data. This complementarity between patient narrative and behavioral measures can be a vector for reducing epistemic injustices in psychiatry [37, 38].

6. PERSPECTIVES

Designing symptom biomarkers instead of diagnostic criteria offers many opportunities, both in speech signal processing and in digital psychiatry.

(i) New tasks. Since "classification in itself is less important than often supposed to be, and less important than other tasks" [29], we have identified new tasks that could be useful to clinicians. One of them is the estimation of symptom

severity. While previous work has looked at the severity of the condition (e.g. estimating the PHQ score for MDD [39, 40]), no work has looked at estimating the severity of symptoms: increasing in the severity of the condition is of no use to clinicians without identifying the symptom(s) that cause it. Another one is *differential diagnosis*: when two disorders are similar, clinicians sometimes have difficulty estimating the patient's diagnosis. While almost all the studies focus only on one disorder (e.g. MDD vs. Healthy Control), more recent work tends to estimate multiple disorders at the same time, e.g. [13, 41, 42, 43]. However, this problem has never been formulated in terms of symptoms. Finally, tasks related to *prognosis* have already been proposed in an Interspeech challenge in 2021 for Alzheimer's disease [44]. However, no article to our knowledge has addressed the estimation of the progression of mental disorders using voice or speech.

(ii) Stratified medicine. Estimating symptoms using vocal biomarkers on a large scale will in turn make it possible to refine the symptomatic profiles related to mental disorders, and thus refine the diagnostic criteria.

(iii) Symptom networks. Estimating the symptoms allows both to reproduce the diagnostic criteria (by estimating all the symptoms of the checklist) but also to propose new models of psychopathology. In particular, symptom networks focus on the link between symptoms in the establishment of a disease state [45]. The conjunction of symptom networks with speech biomarkers would support the spreading of this new modeling in computational psychiatry. but also improve the estimation of classification performance, by taking into account the link with other symptoms when estimating a particular symptom or a disorder [46].

7. CONCLUSION

In conclusion, due to the associated conceptual limitations and the uselessness in clinical practice of systems estimating mental disorders from speech biomarkers, we encourage the community to refocus on symptom estimation, which opens many perspectives. While there are already some initiatives focused on symptoms, such as stress [47] or sleepiness [48, 49], databases annotated with symptoms are lacking. The collection of annotated databases with patients' symptoms rather than their unique pathological status thus seems to us to be a priority for the establishment of voice biomarkers in digital psychiatry that are relevant and useful to medical practice.

8. REFERENCES

[1] Nicholas Cummins, et al., "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[2] Daniel M. Low, Kate H. Bentley, and Satrajit S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," *Laryngoscope Investigative Otolaryngology*, vol. 5, no. 1, pp. 96–116, 2020.

[3] Paul Meehl, "What Can the Clinician Do Well?," *Problems in human assessment*, pp. 594–599, 1967.

[4] Robert L. Spitzer, "Psychiatric diagnosis: Are clinicians still necessary?," *Comprehensive Psychiatry*, vol. 24, no. 5, pp. 399–411, 1983.

[5] Ahmed Aboraya, "Clinicians' opinions on the reliability of psychiatric diagnoses in clinical settings," *Psychiatry*, vol. 4, no. 11, pp. 31–33, 2007.

[6] Peter B. Denes and Eliott N. Pinson, *The Speech Chain: The Physics and Biology of Spoken language*, Bell telephone laboratories, edition, 1963.

[7] Bernd J. Kröger, et al., "Hierarchical Sequencing and Feedforward and Feedback Control Mechanisms in Speech Production: A Preliminary Approach for Modeling Normal and Disordered Speech," *Front. Hum. Neurosci.*, vol. 14, no. 573554, 2020.

[8] Guy Fagherazzi, et al., "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice," *Digital Biomarkers*, pp. 78–88, 2021.

[9] Jennifer Jane Newson, Vladyslav Pastukh, and Tara C. Thiagarajan, "Poor Separation of Clinical Symptom Profiles by DSM-5 Disorder Criteria," *Front. Psychiatry*, vol. 12, pp. 775762, 2021.

[10] Edward L. Campbell, et al., "Speech and Text Processing for Major Depressive Disorder Detection," in *IberSPEECH 2022*, 2022, pp. 86–90.

[11] Alberto Parola, et al., "Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis," *Schizophrenia Research*, vol. 216, pp. 24–40, 2020.

[12] Mireia Farrús, Joan Codina-Filbà, and Joan Escudero, "Acoustic and prosodic information for home monitoring of bipolar disorder," *Health Informatics Journal*, vol. 27, no. 1, pp. 1460458220972755, 2021.

[13] Gábor Gosztolya, et al., "Making a Distinction Between Schizophrenia and Bipolar Disorder Based on Temporal Parameters in Spontaneous Speech," in *Interspeech 2020*, 2020, pp. 4566–4570.

[14] Alice Baird, et al., "An Evaluation of the Effect of Anxiety on Speech — Computational Prediction of Anxiety from Sustained Vowels," in *Interspeech 2020*, 2020, pp. 4951–4955.

[15] Nicole M. Etter, et al., "Speech motor control and orofacial point pressure sensation in adults with ADHD," *Neuroscience Letters*, vol. 744, pp. 135592, 2021.

[16] Manuel Milling, et al., "Is Speech the New Blood? Recent Progress in AI-Based Disease Detection From Audio in a Nutshell," *Front. Digit. Health*, vol. 4, pp. 886615, 2022.

[17] Tomasz Rutowski, et al., "Toward Corpus Size Requirements for Training and Evaluating Depression Risk Models Using Spoken lollanguage," in *Interspeech 2022*, 2022, pp. 3343–3347.

[18] Chelsea Chandler, Peter W Foltz, and Brita Elvevåg, "Using Machine Learning in Psychiatry: The Need to Establish a Framework That Nurtures Trustworthiness," *Schizophrenia Bulletin*, p. sbz105, 2019.

[19] Lena Kastner, et al., "On the Relation of Trust and Explainability: Why to Engineer for Trustworthiness," in *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, 2021, pp. 169–175.

[20] Alexis Bourla, et al., "Psychiatrists' Attitudes Toward Disruptive New Technologies: Mixed-Methods Study," *JMIR Mental Health*, vol. 5, no. 4, pp. e10240, 2018.

[21] Daniel J. Martin, John P. Garske, and M. Katherine Davis, "Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review.," *Journal of Consulting and Clinical Psychology*, vol. 68, no. 3, pp. 438–450, 2000.

[22] Michelle Cleary, Glenn E. Hunt, and Jan Horsfall, "Delivering Difficult News in Psychiatric Settings," *Harvard Review of Psychiatry*, vol. 17, no. 5, pp. 315–321, 2009.

[23] Ravi B. Parikh, Stephanie Teeple, and Amol S. Navathe, "loladdressing Bias in Artificial Intelligence in Health Care," *JAMA*, vol. 322, no. 24, pp. 2377–2378, 2019.

[24] Vincent P. Martin, et al., "How Does Comparison With Artificial Intelligence Shed Light on the Way Clinicians Reason? A Cross-Talk Perspective," *Front. Psychiatry*, vol. 13, 2022.

[25] Kun Qian, et al., "Computer Audition for Healthcare: Opportunities and Challenges," *Front. Digit. Health*, vol. 2, pp. 5, 2020.

[26] Simon Gilbody, et al., "Screening for Depression in Medical Settings with the Patient Health Questionnaire (PHQ): A Diagnostic Meta-Analysis," *Journal of General Internal Medicine*, vol. 22, no. 11, pp. 1596–1602, 2007.

[27] Mark Zimmerman and Joseph B. McGlinchey, "Why don't psychiatrists use scales to measure outcome when treating depressed patients?," *The Journal of Clinical Psychiatry*, vol. 69, no. 12, pp. 1916–1919, 2008.

[28] Alan R. Teo, "A New Form of Social Withdrawal in Japan: a Review of Hikikomori," *International Journal of Social Psychiatry*, vol. 56, no. 2, pp. 178–185, 2010.

[29] Derek Bolton, "Classification and causal mechanisms: a deflationary approach to the classification problem," in *Philosophical Issues in Psychiatry II - Nosology*, pp. 6–11. Oxford university press edition, 2012.

[30] Allen Frances, "Saving normal: An insider's look at what caused the epidemic of mental illness and how to cure it," *New York, NY: William Morrow*, 2013.

[31] Monika A. Waszczuk, et al., "What do clinicians treat: Diagnoses or symptoms? The incremental validity of a symptom-based, dimensional characterization of emotional disorders in predicting medication prescription patterns," *Comprehensive Psychiatry*, vol. 79, pp. 80–88, 2017.

[32] Jon Stone, et al., "The 'disappearance' of hysteria: Historical mystery or illusion?," *Journal of the Royal Society of Medicine*, vol. 101, no. 1, pp. 12–18, 2008.

[33] Mi Jung Rho, et al., "Types of problematic smartphone use based on psychiatric symptoms," *Psychiatry Research*, vol. 275, pp. 46–52, 2019.

[34] Emmanouil Magiorkinis, et al., "Headaches in antiquity and during the early scientific era," *Journal of Neurology*, vol. 256, no. 8, pp. 1215–1220, 2009.

[35] Georg Starke, et al., "Computing schizophrenia: ethical challenges for machine learning in psychiatry," *Psychological Medicine*, pp. 1–7, 2020.

[36] Stéphane Mouchabac, et al., "Improving clinical decision-making in psychiatry: implementation of digital phenotyping could mitigate the influence of patient's and practitioner's individual cognitive biases," *Dialogues in Clinical Neuroscience*, vol. 23, no. 1, pp. 52–61, 2021.

[37] B. Bogaert, "L'application du concept d'injustice épistémique dans le soin : conceptualisation, limites, et perspectives," *Éthique & Santé*, vol. 18, no. 2, pp. 127–133, 2021.

[38] Anke Bueter, "Epistemic Injustice and Psychiatric Classification," *Philosophy of Science*, vol. 86, no. 5, pp. 1064–1074, 2019.

[39] Emna Rejaibi, et al., "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech," *Biomed. Signal Process. Control*, vol. 71, pp. 103107, 2022.

[40] Nicholas Cummins, et al., "Generalized Two-Stage Rank Regression Framework for Depression Score Prediction from Speech," *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 272–283, 2020.

[41] Maria Faurholt-Jepsen, et al., "Discriminating between patients with unipolar disorder, bipolar disorder, and healthy control individuals based on voice features collected from naturalistic smartphone calls," *Acta Psychiatrica Scandinavica*, vol. 145, no. 3, pp. 255–267, 2022.

[42] Wei Pan, et al., "The Discriminatory Power of Vocal Features in Detecting Mental Illnesses Under Complex Context," preprint, In Review, 2021.

[43] Jinfang Wang, et al., "Automatic Assessment for Severe Self-Reported Depressive Symptoms Using Speech Cues," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 875–884, 2021.

[44] Saturnino Luz, et al., "Detecting cognitive decline using speech only: The ADReSSO Challenge," Tech. Rep., 2021.

[45] Denny Borsboom and Angélique O.J. Cramer, "Network Analysis: An Integrative Approach to the Structure of Psychopathology," *Annual Review of Clinical Psychology*, vol. 9, no. 1, pp. 91–121, 2013.

[46] Salvatore Fara, et al., "Bayesian Networks for the robust and unbiased prediction of depression and its symptoms utilizing speech and multimodal data," in *Interspeech 2023*. Aug. 2023, pp. 1728–1732, ISCA.

[47] Mitchel Kappen, et al., "Acoustic speech features in social comparison: how stress impacts the way you sound," *Scientific Reports*, vol. 12, no. 1, pp. 22022, 2022.

[48] Vincent P. Martin, et al., "Automatic Speech Recognition systems errors for objective sleepiness detection through voice," in *Interspeech 2021*, 2021, pp. 2476–2480.

[49] Vincent P. Martin, et al., "Automatic Speech Recognition system errors for accident-prone sleepiness detection through voice," in *EUSIPCO 2021*, 2021, pp. 541–545.