Clinical Images/Videos of Neurodegenerative and Genetic Disorders Combined with Voice-Related Biomarkers

Mette Pedersen[1,2]*

[1]*Ear-Nose-Throat specialist, Head and Neck Surgeon Consultant, the Medical Centre, Østergade 18, DK 1000 Copenhagen, Denmark.*

[2]*Fellow of the Royal Society of Medicine, UK.*

*Corresponding Author: Mette Pedersen, MD, PhD

Email: m.f.pedersen@dadlnet.dk

Abstract

Camera imaging of patients has been used clinically for some years. Based on studies of voice-related biomarkers in a new book by Springer publishers a combination of video images and voice-related biomarkers is suggested in neurogenerative and genetic disorders. The setup includes a 16-channel video mixer (ATEM Mini Pro, 2018) that gets live feed from two high-definition cameras seen on a Sony screen. It is supplemented with an 18-channel Allen & Heath CQ-18T audio mixer for online evaluation and storage. The cameras and acoustical setup are manually handled and capture both facial and full-body motor views in real time with transitions after choice. A MacBook 16, (2023) is used to coordinate to acoustic measures of Fundamental frequency, Jitter, Shimmer, and Harmonics to Noise Ratio. Integrated are Voice Handicap Index, GRBAS perception test, and Maximum Phonation Time. Overall, the workflow functions as a new tool for clinical diagnostics and treatment. Future aspects of AI and of voice-related genetic research are discussed.

Introduction

Camera imaging of patients with neurogenerative disorders has been used clinically for some years [1]. In Parkinson's disease, the type of walking has been in focus [2]. It is not till recently that voice analysis has been included in neurogenerative and genetic disorders. There is a great need for help with verbal rehabilitation in patients with all kinds of neurological disorders, where well-defined voice-related biomarkers may be a well-defined basis for better verbal rehabilitation.

The use of voice-related biomarkers is presented in a recent book [3], where neurogenerative and genetic aspects are also discussed in two chapters related to verbal rehabilitation. The voice-related biomarkers include Voice Handicap Index (VHI), GRBAS perception test, Maximum Phonation Time (MPT) as an airflow measure, and basic acoustic measures (Fundamental frequency (F0), Jitter, Shimmer, and Harmonics to Noise Ratio (HNR)).

The focus on voice is also related to new aspects of genetic understanding of voice regulation, especially the fundamental frequency. The results can have interesting consequences for gene regulation in pathology. Voice pitch and vowel acoustics had a heritable component and correlated with common variants in ABCC8 that associate with voice pitch – found by Gisladottir RS et al. [4] and documented by Di Y et al. [5].

It has been shown that only a few patients with genetic disorders have had their voice analysed [3]. And the analyses are spread over a great many diagnoses; therapy is mostly not considered. Many multi-handicapped syndromes are genetic with developmental voice-related aspects [6].

In this connection, image/video techniques developed with advanced camera switchers and connected with voice analyses are of clinical interest for diagnostics and treatment. Once an image/ video has been correctly made in a dataset with defined camera positions, a combination with, at best, complete anal-

ysis of voice-related biomarkers can be made for better measures in clinical diagnostics. It is also usable as a basis in genetic research [4]. And for better clinical treatment, there are valuable possibilities that include updated AI models.

The voice-related biomarkers were made with reference to a Delphi questionnaire of the best clinical voice measures, made by the European Union of Phoniatricians (UEP) and the European Laryngology Society (ELS) [3].

Aim: This article aims to present an instrumental setup that includes updated video images combined with measures of voice-related biomarkers.

Materials and methods

The setup was managed by a Clinical Technician (CW) using readily available, off-the-shelf hardware to illustrate core principles rather than a custom-built system.

A compact video mixer with 16 channels (ATEM Mini Pro, 2018) ingests live feeds from two high-definition cameras, capturing frontal views of facial and laryngeal movement, and recording full-body posture and gait dynamics, and composites them into real-time monitoring displays on a Sony screen. A MacBook 16, [2023] served as the control workstation, running capture software that handled live annotations, embedded timecode synchronization, and automated archiving of the merged audio–video streams for online and later analysis.

On the audio side, signals from a headset microphone and two ambient room mics are routed through an 18-channel Allen & Heath CQ-18T mixer, with four dedicated channels chosen for the study recordings. These channels passed through the mixer's onboard EQ and compression before being recorded, then were subjected online to standard acoustic analyses—extracting fundamental frequency (F_0), jitter, shimmer, and Harmonics-to-Noise Ratio (HNR) as voice-related biomarkers. During the demonstration, CW manually executes camera adjustments to highlight notable shifts in facial expression, articulatory movement, or overall motor patterns, while simultaneously monitoring the live audio metrics on-screen. The system framework also accommodates voice assessments of self-report questionnaires, perceptual voice ratings, and phonation time measures, showing how synchronized visual and acoustic data streams can be integrated into routine clinical workflows.

Results

A switcher-based setup with four camera angles captures both facial and full-body motor views in real time, with transitions after choice, depending on various interesting parts. This highlights expression, articulatory movement, and gait. The MacBook workstation handled live annotations and timecode synchronization without dropped frames, and all audio–video streams were archived correctly for online and post-session review. The point is that it is possible to evaluate the function of the voice, and hence provide better diagnostics and planning of treatment.

Audio signals routed through the CQ-18T mixer were recorded cleanly, and the MacBook-hosted software immediately extracted standard acoustic parameters, F_0, jitter, shimmer, and HNR, with minimal latency. By aligning the visual cuts with instantaneous acoustic readouts, observers could observe, for example, moments when increased shimmer coincided with visible tremor of the lower facial musculature. Overall, the inte-grated workflow functioned as intended, demonstrating a practical, off-the-shelf approach for combining clinical imaging and voice-related biomarkers.

Discussion

Several new aspects of voice diagnostics have been enlightened. There has for a long time been too little focus on verbal rehabilitation in neurological disorders, which include many neurodegenerative disorders, genetic disorders, and others [3]. There is no evidence in the literature of treatment effects related to voice diagnostics, with Randomized Controlled Trials (RCTs). The patient's clinical voice diagnostics are mostly not compatible between centres either. With the voice-related biomarkers, which were based on a consensus using the Delphi questionnaire method, we now have a tool to coordinate voice diagnostics results. This is very much necessary in the new area of genetic research of the voice, it is a breakthrough that a genetic reference to the fundamental frequency has been found.

Many neurologists will not find it easy to use the voice-related biomarkers alone for clinical voice diagnostics, but in combination with video images of the patients, evaluation and treatment planning can be optimized. This will make comparisons of voice diagnostics from various clinical centres possible as a better basis for genetic research in the future. The use of Delphi questionnaires for validation of the combined setups can help with decision-making to some extent. RCTs of treatment based on voice diagnostics are a perspective. Video images can be evaluated with scores, but in the long term, analyzed with AI, depending on well-defined datasets [3,7].

Conclusion

In neurological disorders, a combination of video images and voice-related biomarkers is presented. The voice-related biomarkers are based on a Delphi questionnaire with a consensus of the UEP/ELS. A supplemental new switch-based video image camera set up with a coordinated sound analysis is a help for better clinical voice diagnostics and therapy of neurological disorders, and a basis for well-defined genetic research.

References

1. Vun DSY, et al. Vision-Based Motion Capture for the GAIT Analysis of Neurodegenerative Diseases: A Review. Gait & Posture. 2024; 112: 95–107.

2. Van Kersbergen J, et al. Camera-based objective measures of Parkinson's disease gait features. BMC Research Notes. 2021: 14.

3. Pedersen M, Camesasca V, Nashaat NH, Hernández-Villoria R, Das S (2025). In print. Voice Related Biomarkers. Springer Int Publishers. Berlin. Heidelberg. New York. 2025.

4. Gisladottir RS, et al. Sequence variants affecting voice pitch in humans. Science Advances. 2023: 9.

5. Di Y, et al. Genetic association analysis of human median voice pitch identifies a common locus for tonal and non-tonal languages. Communications Biology. 2024: 7.

6. Pedersen M. Normal development of voice in Childhood, 2nd Ed. Springer Publishers. Berlin. Heidelberg. New York. 2024.

7. Pedersen M. Voice-related Biomarkers in Singing: Suggested Clinical Measurements Standards of Voice Complaints, Listeners' Evaluation, Acoustical Measurement, and Airflow. Journal of Clinical Medical Research. 2025; 6: 1-6.