A multivariate model incorporating subharmonic measurements for evaluating vocal roughness

Department of Otorhinolaryngology–Head and Neck Surgery, Graduate School of Medicine

Visiting Academic Staff Itsuki Kitayama Researchmap https://researchmap.jp/itkktym?lang=en
Associate Professor (Lecturer) Kiyohito Hosokawa Researchmap https://researchmap.jp/drink?lang=en
Professor Hidenori Inohara Researchmap https://researchmap.jp/read0046473?lang=en



Abstract

Vocal roughness arises from aperiodic vocal-fold vibration. We automatically detected acoustic hallmarks of rough voices—subharmonics and chaotic noise—and quantified their subtypes and temporal proportions. These features, integrated with conventional acoustic indices, yielded a multivariate model, the Acoustic Roughness Index (ARI), expressing roughness on a 0–10 scale. ARI, computed from concatenated reading plus sustained-vowel samples, showed high concordance with auditory–perceptual ratings and strong discrimination of rough voices.

Background & Results

The perceptual salience of roughness increases when the periodicity of vocal-fold vibration breaks down. Physical correlates include frequency components corresponding to fractional multiples of the fundamental frequency (fo) – such as 1/2 and 1/3 subharmonics – as well as chaotic noise lacking regular spectral structure. In real connected speech, however, these phenomena intermingle temporally and may further overlap with left–right desynchronization of vocal folds (biphonation) and higher-order subharmonics. Consequently, stable detection and quantification of subharmonics have been challenging. The principal bottleneck lies in the fact that fo estimation itself becomes unstable when aperiodicity is prominent.

In this study, we first developed a robust fo estimation procedure as a preprocessing step and then performed frame-based analysis of the fo–2fo region. Each frame was automatically classified into four categories—1/2 subharmonics, 1/3 subharmonics, higher-order subharmonics, and chaotic noise—and their temporal proportions (and relative intensities when applicable) were quantified. Although single parameters showed only limited correlations with auditory-perceptual roughness, the total subharmonic proportion (SubSUM) and the chaotic component indicator (ChaoN) demonstrated moderate-to-strong associations, improving explanatory power. These newly derived features were then integrated with conventional acoustic measures, and a multivariate regression model, ARI, was established.

In validation, ARI values obtained from concatenated speech (reading + sustained vowels) agreed well with auditory-perceptual roughness ratings (Spearman's rank correlation coefficient = 0.807) and showed high discriminative ability (area under the ROC curve = 0.916). Rather than replacing subjective judgment, the model functions as an auxiliary metric that reduces inter-rater variability and enables comparison of pre-/post-treatment as well as longitudinal trajectories on a common scale. Future development will include extension to non-standard phonation (multilingual, emotional, singing) and optimization for real-world deployment in clinical and at-home environments (recording conditions, noise robustness), establishing a foundation for objective voice-health assessment independent of distance or setting.

Significance of the research and Future perspective

ARI objectifies roughness severity, enabling quantitative tracking of treatment effects and longitudinal change. It is suitable as a clinical research outcome and may foster diagnostic standardization and cross-site comparability. Future work includes extension to non-standard voice tasks (multilingual, emotional, singing), deployment in telemedicine and home monitoring, and integration with AI-based decision support toward a practical, environment-independent voice-health infrastructure.

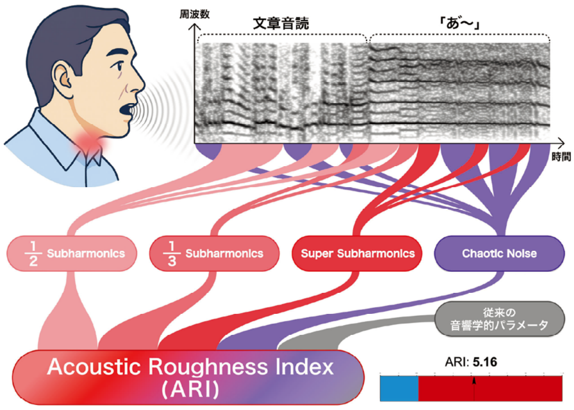

Figure 1: Overview of ARI

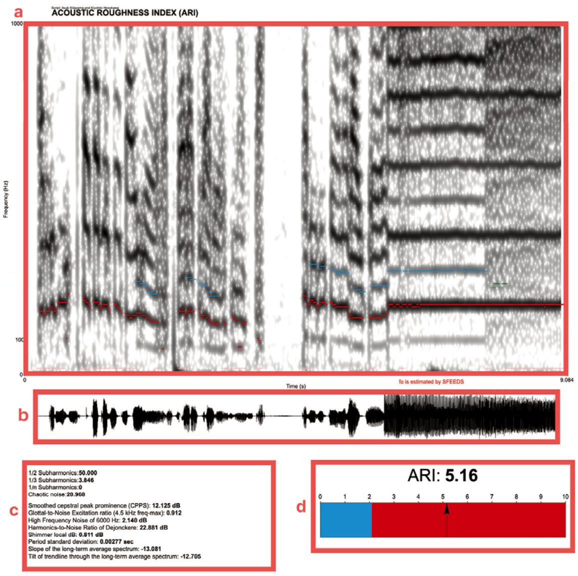

Figure 2: ARI output results

Patent Japanese Unexamined Patent Publication No.2025-180166

Treatise Kitayama, Itsuki; Hosokawa, Kiyohito; Iwaki, Shinobu et al. A multivariate model incorporating subharmonic measurements for evaluating vocal roughness. NPJ Digit Med. 2025, 8(1), 295. doi: 10.1038/s41746-025-01702-2
Kitayama, Itsuki; Hosokawa, Kiyohito; Iwaki, Shinobu et al. Robust fundamental frequency-detection algorithm unaffected by the presence of hoarseness in human voice. J Acoust Soc Am. 2024, 156(6), 4217-4228. doi: 10.1121/10.0034624

URL https://resou.osaka-u.ac.jp/en/research/2025/20250701_1?lang=en

Keyword acoustic analysis, hoarseness, roughness, subharmonics, acoustic model