

PRESS RELEASE

Source: Toyohashi University of Technology, Japan, Committee for Public Relations

Release Title: Diagnosis of Voice Condition from Call Audio

Release Subtitle: Toward the Development of a Dysarthria Diagnostic System in Telemedicine

Overview

Assistant Professor Yuya Hosoda of the Center for IT-Based Education (CITE), Toyohashi University of Technology developed a method for estimating the pitch of vocal cord vibrations of humans from call audio. In this method, the pitch is estimated by integrating the feature quantities extracted from the amplitude and phase spectra of speech on the complex plane. Through experiments, we have demonstrated that the proposed method is not only efficient for call audios whose frequency band is restricted by communication standards, but also works robustly in an environment with background noise.

Details

To prevent the aggravation of neurodegenerative diseases such as Parkinson's disease, the early diagnosis of dysarthria, which is an early symptom, is desirable. Dysarthria is characterized by tremors in voice and disturbed breathing. Although clinical tests diagnose symptoms from the patient's voice, they are time consuming and labor intensive. Additionally, conducting face-to-face interviews in remote locations such as mountainous areas is difficult. Therefore, in this research, we aim to develop a system that automatically diagnoses dysarthria through telemedicine by performing ward rounds via communication devices.

In patients with dysarthria, abnormalities occur during vocalization wherein voice is produced by vocal cord vibrations generated by air released from the lungs in the throat and oral cavity. In this study, our purpose is to estimate the vibration period (pitch) to diagnose the condition of these vocal cord vibrations. Until now, a pitch measurement method that is robust against background noise has been devised based on the feature quantities of the amplitude spectrum obtained via the frequency analysis of speech. However, due to communication standards, call audio via telemedicine lacks some of the desired amplitude spectrum. Thus, extracting feature quantities from an amplitude spectrum with reduced information can lead to errors in pitch estimation.

In this research, we propose a method to extract additional feature quantities from the phase spectrum, a by-product of frequency analysis, in addition to the amplitude spectrum. Deriving a relational equation between the phase shift and pitch in the time and frequency directions, we have verified that pitch can be

estimated by applying the observed phase shift to the relational equation. Based on this finding, we extracted new feature quantities from the phase spectrum to quantitatively evaluate the degree of fit to the relational equation. Finally, by integrating the feature quantities extracted from the amplitude spectrum on the complex plane, we compensated for the lack of feature quantities occurring in the pitch estimation of call audio while maintaining robustness against background noise.

The figure shows the pitch estimation results for a call audio. In previous studies that used only the amplitude spectrum, since the amount of information was reduced by band limitation, the pitch was estimated to be higher than the original value. However, in the proposed method, the pitch is accurately estimated from call audio using the feature quantities related to the amplitude and phase spectra. Further, the gross pitch error (GPE), an evaluation index that indicates the percentage of segments where errors occurred, improved to 9.5% in the proposed method, compared to 42.2% in the previous study. In addition, even for call audio with background noise, this method achieved a GPE of 15.2%, demonstrating robustness.

Future Outlook

Although this study focused on pitch estimation to detect abnormalities in vocal cord vibrations, respiratory and oral abnormalities also cause dysarthria. To detect these symptoms, methods that extract feature quantities from the amplitude spectrum have been devised. However, the use of the phase spectrum has not been sufficiently validated. In the future, we will work on extracting relevant feature quantities from the phase spectra for the other cases as well. Further, by comprehensively analyzing these feature quantities, we aim to develop a dysarthria diagnostic system that can function effectively with telemedicine.

This research was supported by Ono Charitable Trust for Acoustics.

Reference

Yuya Hosoda, Arata Kawamura and Youji Iiguni, "Complex-Domain Pitch Estimation Algorithm for Narrowband Speech Signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2067-2078, 2023,
doi: 10.1109/TASLP.2023.3278488.

Further information

Toyohashi University of Technology

1-1 Hibarigaoka, Tempaku, Toyohashi, Aichi Prefecture, 441-8580, JAPAN

Inquiries: Committee for Public Relations

E-mail: press@office.tut.ac.jp

Toyohashi University of Technology founded in 1976 as a National University of Japan is a research institute in the fields of mechanical engineering, advanced electronics, information sciences, life sciences, and architecture.

Website: <https://www.tut.ac.jp/english/>

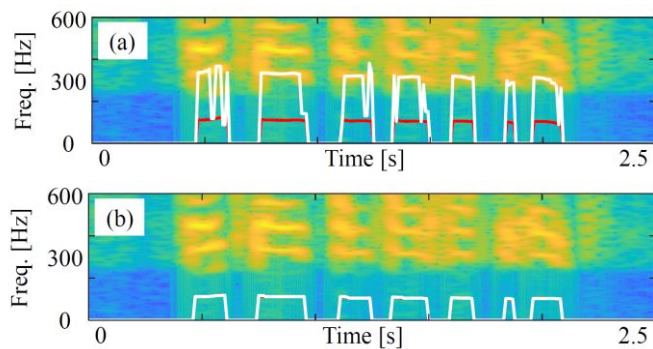


Figure: Pitch estimation results for call audio

The red and white lines indicate the measured and estimated values, respectively.

(a) Prior studies (b) Proposed method

Keywords: Speech development, Voice