

Estimating the quality of sound spaces from observed speech

November 18 2021



Photo: Blind estimation of room acoustic parameters (i.e., T60, EDT, C80, D50, TS) and speech transmission index (STI) in noisy and reverberant environments: a scenario for improving an emergency announcement. Credit: Japan Advanced Institute of Science and Technology

In the future, smartphones, which almost everyone has, and smart speakers, 3.7 million installed in Japanese households, might save your life. Apart from daily-use features, these devices can read emergency messages aloud to inform us of the current situation of an earthquake and how to evacuate. However, we might lose such crucial information due to difficulty listening in some circumstances. The intelligibility of speech is dramatically degraded by noise such as conversations and vacuum cleaners and reverberation as in poor auditoriums or subways.

On the other hand, on a typical day, have you ever been curious about why you enjoy watching movies in theaters more than in your [living room](#)? A bigger screen and a better sound system? Yes, of course, but there is one more factor that is "the well-designed room acoustics."

In the field of architectural acoustics, the speech intelligibility and sound quality of a sound field can be described by measuring speech transmission index (STI) and room acoustic parameters, such as reverberation time (T60), early decay time (EDT), and clarity index (C80). It is also known that the measured acoustical parameters and STI vary from changes in environments such as the number of people, new furniture, or new decorations. Hence, some techniques for estimating these room-acoustic parameters from only a speech have been studied without special instruments and settings. However, assessing various parameters for different purposes of sound spaces in almost real-time remains to be uninvestigated.

In a new study published in *Applied Acoustics*, a team of scientists from the Japan Advanced Institute of Science and Technology (JAIST) has invented a blind estimating method of five-room acoustic parameters and STI simultaneously from a few seconds of speech. Professor Masashi Unoki, a team leader, outlines their approach, "We assumed a speech transmitted in an enclosure is distorted by reverberation and noise associated with the concept modulation transfer function or MTF. The MTF can explain the characteristics of a transmission channel or room acoustics from the modulation ratio between the input and output signal. Based on this assumption, we focused on extracting this relationship from only the [output signal](#) into the previously proposed room-impulse-response (RIR) model, namely, the extended RIR model."

In simulations, the observed signals were synthesized by the convolution of speech signal uttered by five males and five females and 43 realistic RIRs measured from different spaces and configurations. Then, the proposed method estimates room acoustic parameters, including T60, EDT, C80, D50, Ts, and STI, from a short period (five seconds) of these reverberant speech signals. The team found that: (1) the envelope of a reverberant speech signal provides underlying information of room-acoustic characteristics, (2) reverberation and noise affect speech signals in octave bands differently, and (3) a more reasonable stochastic RIR model can accurately approximate a realistic RIR. Therefore, applying the [convolutional neural networks](#) for mapping the envelopes extracted from an observed [speech](#) signal can approximate an unknown RIR. Then, this approach can estimate the STI and various room-acoustic parameters from the approximated RIR.

Based on this finding, architects and acousticians might be able to monitor and diagnose an auditorium during the live performance of concerts with attendees. In the future, our smartphones or smart speakers in a kitchen might save our life one day—our lives tend to be safer, easier, and happier from this technology.

More information: Suradej Duangpummet et al, Blind estimation of speech transmission index and room acoustic parameters based on the extended model of room impulse response, *Applied Acoustics* (2021).
[DOI: 10.1016/j.apacoust.2021.108372](https://doi.org/10.1016/j.apacoust.2021.108372)

Provided by Japan Advanced Institute of Science and Technology

Citation: Estimating the quality of sound spaces from observed speech (2021, November 18)
retrieved 20 June 2024 from <https://phys.org/news/2021-11-quality-spaces-speech.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.