Chapter 18

Machine Audition of Acoustics: Acoustic Channel Modeling and Room Acoustic Parameter Estimation

Francis F. Li
The University of Salford, UK

Paul Kendrick
The University of Salford, UK

Trevor J. Cox
The University of Salford, UK

ABSTRACT

Propagation of sound from a source to a receiver in an enclosure can be modeled as an acoustic transmission channel. Objective room acoustic parameters are routinely used to quantify properties of such channels in the design and assessment of acoustically critical spaces such as concert halls, theatres and recording studios. Traditionally, room acoustic parameters are measured using artificial probe stimuli such as pseudo random sequences, white noise or sine sweeps. The noisy test signal hinders occupied in-situ measurements. On the other hand, virtually all audio signals acquired by a microphone have undergone a process of acoustic transmission in the first place. Properties of acoustic transmission channels are essential for the design of suitable equalizers to facilitate subsequent machine audition. Motivated by these needs, a number of new methods and algorithms have been developed recently to determine room acoustic parameters using machine audition of naturally occurring sound sources, i.e. speech and music. In particular, reverberation time, early decay time and speech transmission index can be estimated from received speech or music signals using statistical machine learning or maximum likelihood estimation in a semi-blind or blind fashion. Some of these estimation methods can achieve accuracies similar to those of traditional instrument measurements.

DOI: 10.4018/978-1-61520-919-4.ch018
INTRODUCTION

Propagation of sound from a source to a receiver (a listener or a microphone) in an enclosed or semi-enclosed space is subject to multiple reflections and reverberation. This can be viewed as an acoustic transmission channel of sound. Characteristics of acoustic transmission channels are often a part of music enjoyment and a feature of perceived sound. Recordings made in relatively “dry” studios are often processed using artificial reverberation algorithms to achieve preferred artistic effects. Different concert halls have different acoustics; different lecture theatres may show different speech intelligibility. Research into room acoustics over 100 years has accumulated a large knowledgebase of preferred or suitable acoustics of spaces for various purposes in terms of objective acoustic parameters, a set of physical measures of acoustic transmission channels. These parameters are routinely used to quantify acoustics in the design and assessment of acoustically critical spaces such as concert halls, where music needs to be good sounding, lecture theatres, in which speech communication is vital, and transportation hubs where public address broadcast needs to be clearly heard. Objective parameters are traditionally measured using high sound pressure level artificial test signals, which are unacceptable to audiences. For this reason, occupied in-use measurements are rarely carried out even though it is well established that occupancy significantly affects absorption and hence acoustic parameters. It is therefore suggested that the obstacle of obtaining the much sought after occupied data can be circumvented by extracting these parameters from naturally occurring sound sources such as music or speech when the spaces are in use.

On the other hand, machine audition involves predominately digital processing, feature extraction and pattern recognition of acquired acoustic signals. Sources may include but not limited to speech, music and other event sounds. Audio signals are typically picked up by microphones in an enclosure, for example, a room, concert hall, auditorium or recording studio. Sound transmitted from the source to the microphones endures a multiple reflection and reverberation process, i.e. an acoustic transmission channel. Such transmission channels are complicated and may have significant impact on the effectiveness of signal processing algorithms for machine audition purposes. But modeling of the acoustic transmission channel is often neglected, resulting in some algorithms working only under controlled conditions. For example, some automatic speech recognition systems cannot handle speech signals picked up in noisy and reverberant spaces, especially when the microphones are placed at a remote distance from source. An equalization pre-processor would be beneficial. Impulse responses of room acoustic transmission channels are extremely long (hundreds and thousands taps) and non-minimum phase. An exact inverse does not exist. Room acoustic parameters as descriptors of an acoustic transmission channel can therefore provide useful information for the design of equalization algorithms.

Motivated by these needs, the use of machine audition of speech or music to determine acoustic parameters of the channel through which the speech and music have been transmitted has been studied, leading to the development of a number of new algorithms and methods. Reverberation Time (RT) is the most common acoustic parameter used for the assessment in room acoustics. Early Decay Time (EDT) was reported to better correlate with human perception of the reverberation effect and hence has become an ever-popular acoustic parameter. Speech Transmission Index (STI) is an IEC standardized objective parameter for speech intelligibility assessment in a space. These parameters are considered in this chapter. Short-time RMS values of received discrete speech utterances, in particular pronounces digits can be used to determine RT and EDT by statistical machine learning with an Artificial Neural Network.