

Title	音声伝送指標を基準としたスピーチプライバシー保護の研究
Author(s)	柏原, 佑太
Citation	
Issue Date	2017-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/14139
Rights	
Description	Supervisor: 鶴木 祐史, 情報科学研究科, 修士

Study on privacy protection for speech based on speech transmission index

Yuta Kashihara (1510010)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2016

Keywords: speech transmission index (STI), room impulse response (RIR), modulation transfer function (MTF), privacy protection for speech, listening comprehension.

The speech transmission index (STI) is an objective measurement for evaluating the speech quality in room acoustics. Thus, the STI is usually used to predict speech intelligibility by listeners in the room. STI is calculated from the room impulse response (RIR), based on the concept of modulation transfer function (MTF). The correspondence between STI and its effectiveness in evaluating the quality of speech transmission in room acoustics is as follows : bad ranging from 0.0 to 0.3, poor ranging from 0.3 to 0.45, fair ranging from 0.45 to 0.6, good ranging from 0.6 to 0.75, and excellent ranging from 0.75 to 1.0. It is known that the STI has high correlation with listening difficulty. Therefore, STI and RIR can be used to predict listening comprehension in the room. In this thesis, we aim to protect speech privacy by manipulating STI based on this STI, making it difficult to listen to leaked speech. This thesis also aims to investigate a possibility of controlling listening comprehension in the room acoustics by manipulating parameters of the extended RIR model related to late reverberation.

Due to inconvenience for directly controlling the room acoustics in real communication, it is difficult to control the STI in the room. Fortunately, since presentation of speech and additive delayed-manipulated speech can

be regarded as convolution of speech with late reverberation, so we predicted that the STI can be controlled by convolving speech with the simulated RIR that we can manipulate. Hence, it is predicted that our listening comprehension of speech in the room can be controlled by manipulating the parameters of RIR model related to late reverberation and then manipulating the derived STI. We operate the RIR of the room by the reverberant speech. The STI is reduced by the reverberant speech convoluted with the RIR and the target speech. Thus, the speech privacy protection is done by decreasing the STI near the third party. The proposed method of this research can create a safe area where voice communication can be performed safely. In this thesis, we investigate the protection for speech by manipulating the STI by manipulating the parameters using RIR composed of direct sound and late reverberation, controlling listening comprehension of speech in the room.

In order to predict the listening comprehension, we need to know the STI and the RIR in the room. RIR consists of three parts: direct sound, early reverberation, and late reverberation. From our previous studies, it is known that the STI is especially dominated by late reverberation. It is also known that the STI can be predicted by using the stochastic RIR model consisting of direct sound and late reverberation. Thus, we predicted that the STI can be directly controlled by manipulating the model parameters related to late reverberation. The STI can be controlled by manipulating model parameters constituting the late reverberation. We can also control listening to speech and can protect speech privacy. At first, the late reverberation model with extended RIR as late reverberation was defined. The MTF was calculated from this model. Then, The MTF were seen low pass characteristics. As the results, it was found that two parameters of the extended RIR model, T_h and T_t , can control STI in the range from 0.2 to 1.0. The late reverberation model with extended RIR as late reverberation can control the parameter T_h is a parameter of the growth of RIR, T_t is a parameter of the decay of RIR, a is gain factor of late reverberation. This method is called the proposed method.

Next, We confirmed whether listening comprehension can be controlled by manipulating STI. Thus, three experiments: word intelligibility, listening difficulty, and annoyance tests were carried out in this paper to evaluate

the proposed approach. The test stimuli were chosen from the Familiarity-controlled Word-lists 2007 (FW07). The words were composed of four mora. Stimuli were spoken by a male speaker (mya) and word familiarity of 1.0 to 7.0. Twenty lists were selected for each listener. Five RIRs were generated by using the proposed method and then the corresponded STI were 0.875, 0.675, 0.525, 0.375, and 0.230. Test stimuli were generated by convolving speech signals in FW07 with RIRs. The total number of test stimuli was 1200 (five STI conditions \times four WF conditions \times twenty words \times three experiments). All speech signals had a sampling frequency of 48 kHz. In the word intelligibility test and the listening difficulty test, six male listeners aged from 23 to 25 participated in the experiments. In the annoyance test, three male listeners aged from 23 to 25 participated in the experiments. All listeners had normal hearing. They were native speakers of Japanese. In the word intelligibility test, the intelligibility was regarded as the correct answer rate of the word answered by listeners. In listening difficulty test, listener was asked to choose one from the four categories: (I) Not difficult, (II) A little difficult, (III) Fairly difficulty, and (IV) Extremely difficulty. Listening difficulty rate (LDR) is calculated by counting the number of (I). In the annoyance test, listener was asked to choose one from the four categories: (i) Not at all annoying, (ii) Slightly annoying, (iii) Moderately annoying, and (iv) Extremely annoying. Annoyance rate (ANR) is calculated by counting the number of (i). In word intelligibility test, listener was asked to repeat words by typing characters as they listened. Word intelligibility was calculated by counting the number of words correctly heard and then rating them. As a result, it was found that the word intelligibility can be controlled by operating the STI although it depends on word familiarity. Moreover, it was found that the listening difficulty and the annoyance are controlled by operating the STI irrespective of word familiarity. Thus, we can control the word intelligibility, the listening difficulty and annoyance by using the proposed method. We investigated whether listening comprehension in the room can be controlled by manipulating the extended RIR model relate to late reverberation and by manipulating the simulated STI. Three experiments: word intelligibility, listening difficulty, and annoyance tests, were carried out to evaluate our proposed approach. As the results, it was found that word intelligibil-

ity is decreased by manipulating RIR and STI to be reduced. Moreover, it was shown that listening difficulty and annoyance are increased by manipulating RIR and STI to be reduced. From these results, it was found that listening comprehension can be controlled by manipulating STI.

Finally, a comparative experiment was conducted to examine whether the proposed method is superior to other methods. The total number of test stimuli was 360 (three methods \times two WF conditions (low and high) \times twenty words \times three experiments). Eight male listeners aged from 23 to 31 participated in the experiments. We conducted a listening experiment on word intelligibility, listening difficulty, and the annoyance to compare the proposed method with the reverberation speech and pink noise when the STI is 0.23. As a result, the proposed method was able to demonstrate almost the same performance as the other two methods. Moreover, since the proposed method has lower signal to noise ratio (SNR) than the other two methods, it can be said that STI could be controlled efficiently. This proved that it is easier to protect speech privacy than other two speech privacy protection methods.

In this thesis, we clarified the following three. (1) The parameters T_h and T_t that control the growth and decay of the extended RIR have a large influence on the STI. (2) Speech privacy can be protected by manipulating STI. (3) It is easier to protect speech privacy than previous speech privacy protection technology.