

Evaluation of speech intelligibility for classroom-to-classroom collaborative learning via multimedia network

Manabu Fukushima, Naoto Nakamura¹ and Hifofumi Yanagawa¹

Fukuoka Institute of Technology

Chiba Institute of Technology¹

fukusima@fit.ac.jp, nakamura@net.it-chiba.ac.jp,

yanagawa@net.it-chiba.ac.jp

Abstract

This paper describes the analysis for the decrease of speech intelligibility when using microphones that are placed in a classroom for classroom-to-classroom collaborative learning via multimedia network. The most of it is caused when recording the students' voices by distant location of the microphone such as being mounted at the wall but not in front of the students. The speech intelligibility in classroom is already investigated and some indexes of room acoustics are proposed. However these indexes are not focused on distance learning. In this paper, firstly we explain how the reverberation of the classroom affects the speech intelligibility of recorded voice using microphones. Secondly, we pointed out that the decrease occurs when mixing the voice recorded by several microphones, since arriving times of the voice to each microphone are different. These two causes were verified theoretically in the problem of the voice obscurement of the distance learning between an elementary school and a school class in the hospital under the project of the Ministry of Education, Culture, Sports, Science and Technology. As the result, the problem in the configuration of the microphone in the elementary school was clarified quantitatively.

1. Introduction

Recently, a network of ISDN or ATM can be used with a sufficient bandwidth to transmit images and voices clearly for distance learning¹⁾. However, an issue is raised that deterioration of the transmitted voice clearness by the classroom acoustics and that by capturing a voice using a microphone not closely located to a speaker must be analyzed and some measures are needed²⁾. Acoustics of the classroom and the speech intelligibility in that has been studied already. The speech intelligibility from a teacher to students in the classroom was investigated quantitatively³⁾. However, in the study an ordinary style of learning was assumed rather than the distance learning by the network. Though, the technique of the digital signal

processing can capture a voice selectively by the directivity control of the microphone array⁴⁾, there still exist problems as follows. 1) Some students speak simultaneously in a short duration so that the directivity control is difficult to trace them. 2) The weak voice may be missed to capture when another student speaks loud. In a case of classroom-to-classroom collaborative learning by using a multimedia network, the capturing and reproducing voices in both classrooms should be investigated where the acoustics are different each other. The investigation includes not only a quantitative evaluation of the speech intelligibility in the classroom but also a quantitative evaluation of the effects for the intelligibility by a mixing of the outputs of the microphones arranged in the classroom.

In this paper, a factor worsen the speech intelligibility is mentioned. Then, the room acoustic parameters to estimate the speech intelligibility and measured result of them in the 34 classrooms are described. Lastly, the speech intelligibility is evaluated by the room acoustic parameters and by the listening test under the condition when classroom-to-classroom collaborative learning is being operated.

2. The speech intelligibility and its evaluation

2.1. Deterioration of speech intelligibility

The reverberation in a room worsen the speech intelligibility. Generally, the longer reverberation, the worse speech intelligibility. The reverberation time is expressed with Tr in room acoustics. The length of the reverberation time Tr is defined as the time(t) that satisfy

$$-60 = 20 \log_{10}(e^{-6.91t}) = 20 \log_{10}(e^{-\alpha t}) \quad (1)$$

where, α is the time constant.

Equation (1) shows the Tr is the time when the acoustic energy decay as 1/1000,000 or -60dB. An impulse response includes all acoustic characteristics from a sound source point to an observation point in a room. The reverberation time Tr can be obtained from the decay curve of the impulse response.

To investigate the decrease of speech intelligibility by the reverberation, the impulse response can be modeled with equation(1) as

$$h(t) = e^{-6.91t/Tr} c(t) \quad (2)$$

where $e^{-6.91t/Tr}$ is the envelope, and $c(t)$ is the carrier signal composed of randomly spaced pulse train.

The model of a speech sound as sinusoidal amplitude modified white noise $s(t)$ expressed as

$$x(t) = \frac{1 - \cos(\omega t)}{2} s(t) \equiv d_x(t) s(t). \quad (3)$$

The obtained reverberant speech signal $y(t)$ can be expressed with equation (2) and (3) as

$$\begin{aligned} y(t) &= \int_{\tau=0}^{\tau=Tr} x(t-\tau) h(\tau) d\tau \\ &= \int_{\tau=0}^{\tau=Tr} d(t-\tau) s(t-\tau) e^{-\frac{6.91\tau}{Tr}} c(\tau) d\tau. \end{aligned} \quad (4)$$

The power ensemble of the $s(t)$ and $c(t)$ becomes zero, since they are incoherent. Therefore, the power ensemble average of equation (4) can be expressed as

$$\begin{aligned} \overline{y(t)^2} &= \overline{\left\{ \int_{\tau=0}^{\tau=Tr} d(t-\tau) s(t-\tau) e^{-\frac{6.91\tau}{Tr}} c(\tau) d\tau \right\}^2} \\ &= \overline{\left\{ \int_{\tau=0}^{\tau=Tr} d(t-\tau) e^{-\frac{6.91\tau}{Tr}} d\tau \right\}^2} \end{aligned} \quad (5)$$

Equation (5) shows that the envelope of modeled speech sound is averaged with the envelope of the impulse response. That means the speech intelligibility decreased according to the envelope of the impulse response.

Figure 1 shows the envelope of modeled speech sound and the envelope of obtained reverberant speech sound. Each amplitude of the waveforms is normalized respectively. The horizontal axis shows the time and the vertical axis shows the amplitude.

2.2. Estimation method for speech intelligibility

MTF (Modulated Transfer Function) is one of the measurement methods for speech intelligibility that shows the ratio of modulation⁵⁾. The MTF $m(F)$ is obtained from the impulse response as

$$m(F) = \frac{\int_{t=0}^{t=Tr} h^2(t) \exp(-j2\pi t F) dt}{\int_{t=0}^{t=Tr} h^2(t) dt} \quad (6)$$

where F is the modulation frequency.

From the figure 1, MTF can be obtained as $m(2) = 1 - 0.75 = 0.25$. STI (Speech Transmission Index)⁵⁾ is calculated from the MTF as

$$STI = \sum_{k=1}^7 w_k \left\{ \frac{1}{14} \sum_{i=1}^{14} 10 \log_{10} \left\{ \frac{m_k(i)}{(1 - m_k(i))} \right\} \right\} \quad (7)$$

where, $w_1 = 0.129$, $w_2 = 0.143$, $w_3 = w_4 = 0.114$, $w_5 = 0.186$, $w_6 = 0.171$, $w_7 = 0.143$, k means number of octave band, when $k=1$ center frequency of octave band is 125Hz.

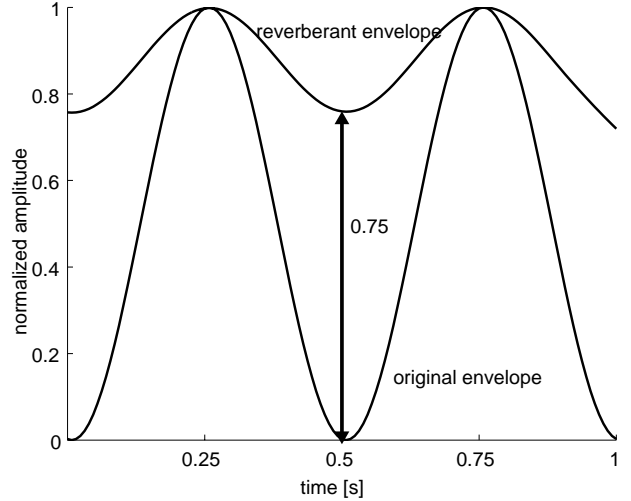


Figure 1. The envelope of original modeled speech sound and the envelope of reverberant speech sound in case of the reverberant time is 1 second, and the modulation frequency is 2Hz

2.3. Physical indexes of room acoustics

Physical indexes D, C and R are proposed for evaluating speech intelligibility in room acoustics. These indexes are defined as follows:

$$D \equiv \frac{\int_{t=0}^{t=50ms} h(t)^2 dt}{\int_{t=0}^{t=Tr} h(t)^2 dt} \quad (8)$$

$$C \equiv 10 \log_{10} \left(\frac{\int_{t=0}^{t=80ms} h(t)^2 dt}{\int_{t=80ms}^{t=Tr} h(t)^2 dt} \right) \quad (9)$$

$$R \equiv 10 \log_{10} \left(\frac{\int_{t=0}^{t=50ms} h(t)^2 dt}{\int_{t=0}^{t=Tr} h(t)^2 dt} \right) \quad (10)$$

For the evaluation of initial part of an impulse response, EDT(Early Decay Time) and T_s (Center Time) are proposed. EDT is calculated with the time when the acoustic energy decays as 1/10 or -10dB. And the T_s is defined as

$$T_s \equiv \int_0^{Tr} t h(t)^2 dt / \int_0^{Tr} h(t)^2 dt. \quad (11)$$

3. The acoustics of classroom

Impulse responses are measured in 34 classrooms. In case of the sound source is placed near the blackboard and the observation points are located at student's position. The physical indexes and STI is calculated from the measured impulse responses. The results are shown in table1. And also they are figured out in figure 2 and 3. The horizontal axis shows the room volume and the vertical axis shows the physical indexes calculated from measured impulse responses. The rank of STI are, over 0.8: Excellent, 0.8 to 0.6: Good, 0.6 to 0.4: Fair, 0.4 to 0.2: Poor. Figure 3 shows all classrooms are 'Excellent'.

Table 1. The physical indexes and STI of classrooms

no.	vol	D(%)	C(dB)	R(dB)	Ts(s)	EDT(s)	Tr(s)	STI
1	62	86	13	8	0.028	0.310	0.433	0.94
2	63	88	13	9	0.021	0.393	0.653	0.94
3	63	88	13	8	0.021	0.393	0.584	0.93
4	63	75	10	5	0.035	0.488	0.552	0.90
5	63	78	9	5	0.034	0.487	0.546	0.87
6	63	89	14	9	0.020	0.363	0.565	0.87
7	63	77	10	5	0.036	0.424	0.546	0.90
8	63	84	12	7	0.029	0.362	0.554	0.93
9	63	90	14	10	0.021	0.294	0.626	0.95
10	63	79	10	6	0.035	0.409	0.567	0.91
11	63	67	8	3	0.046	0.517	0.559	0.88
12	68	70	8	4	0.047	0.498	0.510	0.96
13	68	70	9	4	0.046	0.454	0.502	0.96
14	68	64	8	3	0.048	0.528	0.521	0.96
15	68	81	11	6	0.032	0.391	0.498	0.97
16	68	89	13	9	0.020	0.336	0.493	0.98
17	68	71	9	4	0.041	0.522	0.513	0.96
18	68	70	9	4	0.043	0.461	0.495	0.96
19	68	69	8	3	0.047	0.459	0.504	0.96
20	68	70	9	4	0.042	0.488	0.485	0.96
21	68	87	13	8	0.024	0.370	0.447	0.98
22	68	92	15	10	0.018	0.279	0.454	0.98
23	68	80	11	6	0.033	0.413	0.456	0.97
24	163	87	13	8	0.036	0.283	1.290	0.93
25	264	68	7	3	0.053	0.566	1.227	0.87
26	311	89	13	9	0.027	0.263	0.637	0.94
27	311	89	13	9	0.027	0.264	0.564	0.94
28	311	90	13	9	0.027	0.246	0.565	0.94
29	311	73	9	4	0.041	0.454	0.533	0.90
30	311	94	19	12	0.024	0.204	0.315	0.95
31	556	64	8	3	0.056	0.480	1.441	0.87
32	574	54	5	1	0.070	0.701	1.509	0.83
33	1911	28	2	-4	0.096	0.873	1.357	0.77
34	1000	78	8	5	0.046	0.522	1.697	0.89

4. The evaluation for distance learning

We focused on a classroom where is used for distance learning between an elementary school and a school class in the hospital under the project of the Ministry of Education, Culture, Sports, Science and Technology. The shape of the classroom in elementary school is rectangle and the room volume is 68m³. The classrooms are connected using the network of ISDN1500. Figure 4 shows the classroom in the elementary school.

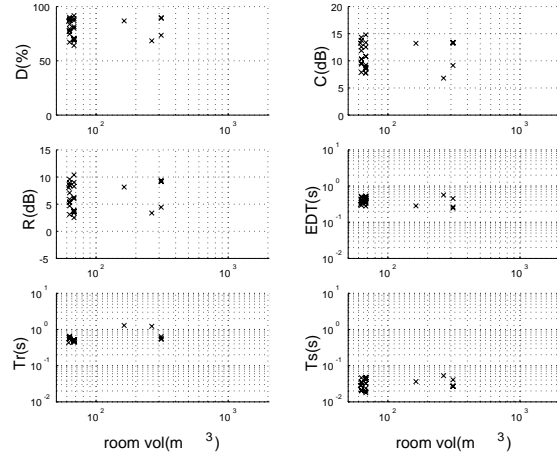


Figure 2. The relation between room volume and physical indexes

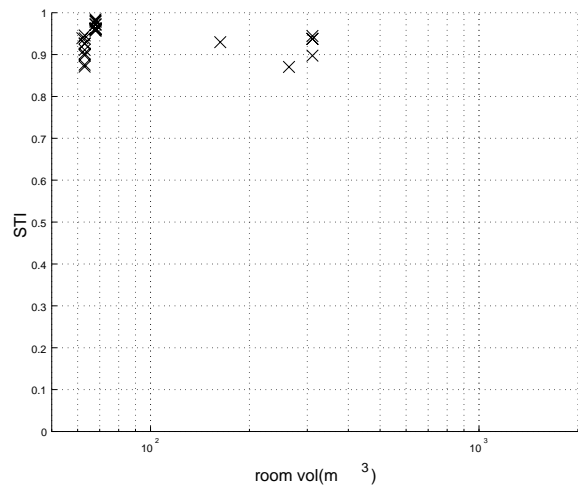


Figure 3. The relation between room volume and STI



Figure 4 The classroom in elementary school

The sight of the classroom and the position of sound source (S1, S2, S3), microphone (M1, M2, M3, M4) and dummy head microphone (DH1, DH2) are shown in figure 5. An example of measured impulse response is shown in figure 6. In figure 6, the horizontal axis shows the time and the vertical axis shows the amplitude.

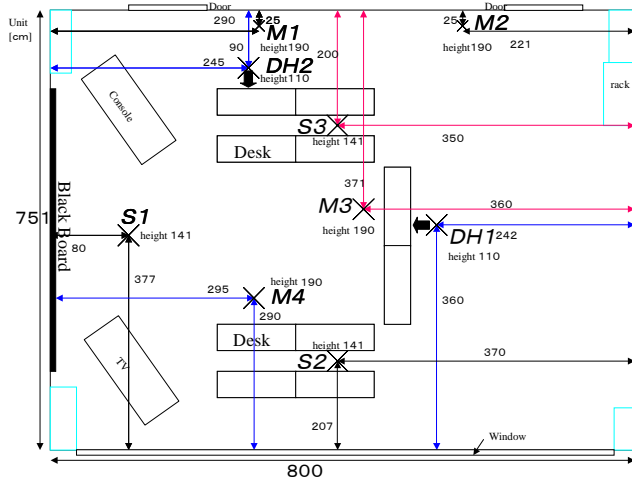


Figure 5. The shape of classroom in elementary school and the position of sound source (S1, S2, S3), microphone(M1, M2, M3, M4) and dummy head microphone(DH1, DH2)

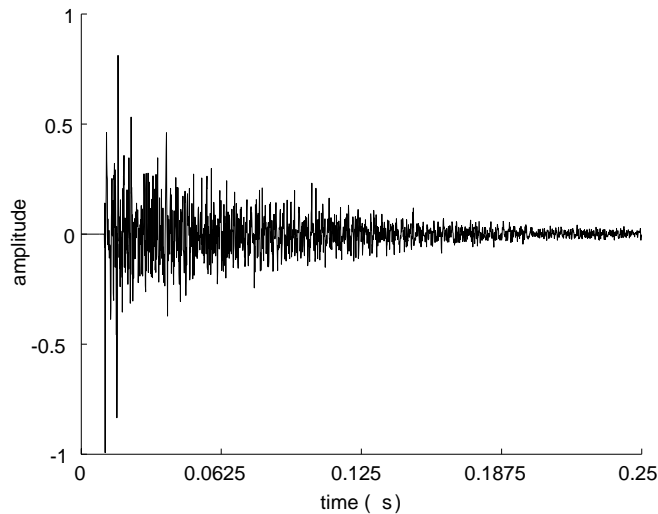


Figure 6. An example of measured impulse response (vertical axis: amplitude, horizontal axis: time)

The physical indexes and STI calculated from measured impulse responses are shown in table 2. In the table 2, notation S*_MIX means console output signal that is a mixed signal of the outputs from 4 microphones which are mounted on the wall (regular microphone-setting configuration for distance learning). They are figured out in figure 6 and 7. The horizontal axis shows the room volume and the vertical axis shows the physical indexes calculated from measured impulse responses.

Table 2. The physical indexes and STI calculated from measured impulse responses

	D(%)	C(dB)	R(dB)	Ts(s)	EDT(s)	Tr(s)	STI
S1_M1	76	11	5	0.036	0.418	0.490	0.97
S1_M2	70	9	4	0.046	0.442	0.503	0.97
S1_M3	75	9	5	0.041	0.438	0.500	0.97
S1_M4	78	10	6	0.035	0.450	0.504	0.97
S2_M1	74	9	5	0.045	0.435	0.514	0.96
S2_M2	71	9	4	0.045	0.433	0.492	0.96
S2_M3	84	11	7	0.027	0.413	0.490	0.97
S2_M4	85	11	8	0.027	0.372	0.495	0.97
S3_M1	80	11	6	0.032	0.427	0.483	0.97
S3_M2	84	11	7	0.027	0.414	0.474	0.97
S3_M3	85	12	7	0.026	0.374	0.458	0.98
S3_M4	74	10	5	0.040	0.435	0.467	0.97
S1_MIX	77	10	5	0.038	0.447	0.493	0.75
S2_MIX	76	9	5	0.034	0.513	0.514	0.73
S3_MIX	81	10	6	0.031	0.441	0.496	0.79
S1_DH1	70	9	4	0.046	0.454	0.502	0.96
S1_DH2	81	11	6	0.032	0.391	0.498	0.97
S2_DH1	71	9	4	0.041	0.522	0.513	0.96
S2_DH2	69	8	3	0.047	0.459	0.504	0.96
S3_DH1	87	13	8	0.024	0.370	0.447	0.98
S3_DH2	80	11	6	0.033	0.413	0.456	0.97

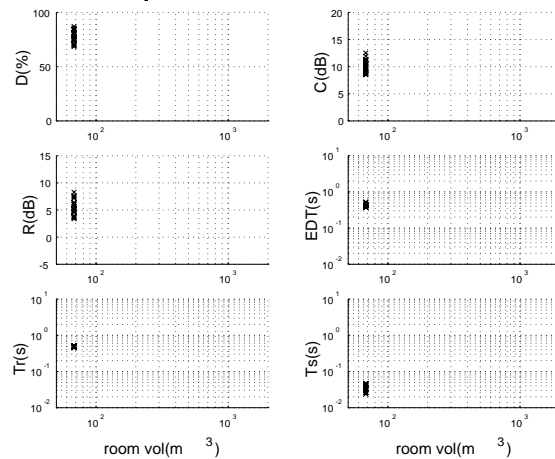


Figure 7. The physical indexes calculated from measured impulse responses (microphone, mixing console output, dummy head)

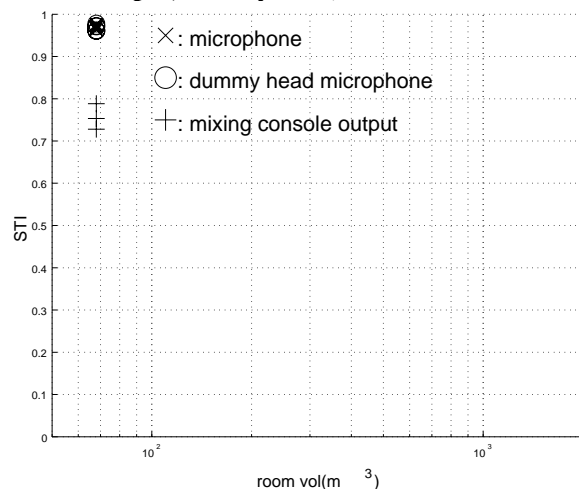


Figure 8. The STI calculated from measured impulse responses (microphone, mixing console output, dummy head)

Figure 8 shows the STI from impulse responses measured by using microphone and dummy head are 'Excellent', but the mixing console output is 'Good'.

To confirm the decreases, the syllable articulation test⁶⁾ was done. It was listening test to obtain percentage of correct answer when listen to nonsense words of three successive syllables. The percentage of correct answer, that is the articulation score, is a direct measure of the speech intelligibility. The result is shown in table 3. In the table 3, the result of listening test and STI from a classroom in a hospital is also shown. The shape of the classroom in a hospital and measured points are shown in figure 9.

Table 3. Articulation score and STI
(Upper 3 lines: Articulation score,
Lower 3 lines: STI)

	M1	M2	M3	M4	MIX	Hospital
S1	60%	56%	62%	58%	40%	40%
S2	56%	56%	60%	60%	44%	44%
S3	62%	64%	64%	64%	44%	44%
S1	0.97	0.97	0.97	0.97	0.75	0.72
S2	0.96	0.96	0.97	0.97	0.73	0.71
S3	0.97	0.97	0.98	0.97	0.79	0.77

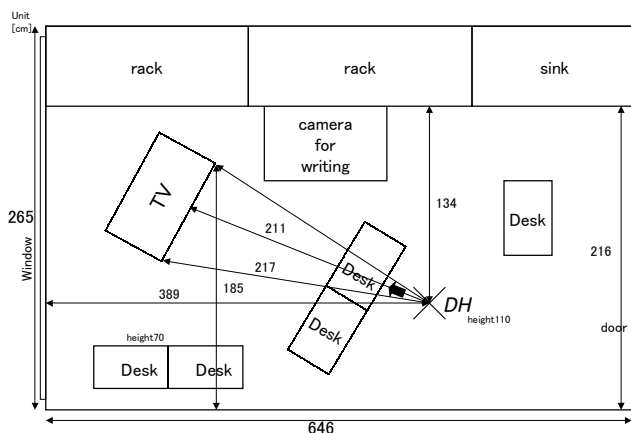


Figure 9. The shape of a classroom in a hospital and the position of dummy head microphone (DH) (the sound source is placed in the classroom of the elementary school and the transmitted signal is reproduced with a loud-speaker of TV)

Table 3 shows that the articulation score decreases about 20% and also STI decreases 0.2. The score of listening test with mixing console output and a classroom in a hospital are the same. It means the decreases are occurred when mixing speech signals.

5. Conclusion

This paper describes the analysis for the decrease of speech intelligibility when using microphones that are placed in a classroom for classroom-to-classroom collaborative leaning via multimedia network.

A factor worsen the speech intelligibility is mentioned. Then, the room acoustic parameters to estimate the speech intelligibility and measured result of them in the 34 classrooms are described. As a result all classrooms are categorized in 'Excellent' by STI.

Then, we evaluate the speech intelligibility by the room acoustic parameters and by the articulation test under the same conditions as classroom-to-classroom collaborative learning is being operated. We pointed out that the decrease occurs by 1) reverberation in a classroom, 2) mixing the voice from the outputs of several microphones, since arriving times of the voice to each microphone are different. These two causes were confirmed theoretically in the case of the voice obscureness of the distance learning between an elementary school and a school class in the hospital. As the result, the problem in the configuration of the microphone in the elementary school was clarified quantitatively. The speech intelligibility decreases about 20% in articulation test and STI decreases 0.2. The STI shows the mixing console output is categorized in 'Good' even the classroom is categorized in 'Excellent'.

6. References

- [1]Maeda, et. al., "Multimedia Communication Environment for Distance Learning", IEICE, J80-B-1, No.6, pp.348-354, 1997 (in Japanese)
- [2]Kimura et. al., "A Study on Structure and Quality of Multimedia Network Systems for Education with Satellite Interactive Communication", IEICE, J80-B-1, No.6, pp.355-365, 1997 (in Japanese)
- [3]Kobayashi, and Ishii, "Articulation Tests in Lecture Rooms", ASJ, vol.7, no.2, pp.59-63, 1951 (in Japanese)
- [4]Kaneda, "Directivity control using microphone array", ASJ, vol.5, no.5, pp.390-394, 1995 (in Japanese)
- [5]H.J.M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality", J. Acoust. Soc. Am. 67, pp.318-326, 1980
- [6]Iida, "On the articulation test", ASJ, vol.43, no.7, pp.532-536, 1987 (in Japanese)