

ACOUSTICAL LETTER

Signal to overlap-masking ratio of the broadcasted speech and its listening difficulty: An application for an evaluation tool of sound system tuning

Kiyohiro Kurisu[1,*], Susumu Nakamura[2], Keiichi Yasu[3] and Takayuki Arai[3]

[1]*TOA Corporation, 2–1 Takamatsu-cho, Takarazuka, 665–0043 Japan*
[2]*Graduate School of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan*
[3]*Faculty of Science and Technology, Sophia University, 7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan*

(*Received 22 February 2013, Accepted for publication 9 April 2013*)

Keywords: Speech intelligibility, Signal to overlap-masking ratio, Listening difficulty rating, Public address system
PACS number: 43.55.Hy, 43.71.Gv, 43.55.Jz [doi:10.1250/ast.34.354]

1. Introduction

STI (Speech Transmission Index) [1] is an index of speech clarity and is widely utilized for an evaluation scale of public address systems. The numerical value of the *STI* is, however, mostly dominated and determined by the number and the characteristics of loudspeakers as well as the physical conditions of the room's interior surfaces. It can, therefore, hardly be improved on in spite of all the efforts made by engineers, for example, tuning the sound system after installation. This theoretical restriction has existed and been noticed by researchers [2].

To address such a problem, an evaluation for the adjustment of the sound system by means of the improvement of *LDR* (Listening Difficulty Rating) [3] was conducted with the investigation of the relation between broadcasted speeches and their measured *SOR* (Signal to Overlap-masking Ratio) [4]. This paper gives a review of the previous paper [5] and discusses the difference between *SOR* and C_{50}.

The *LDR* in this paper was investigated using a specific short speech pattern instead of the phonetically balanced word set with high word familiarities used previously [3]. The speech signal transmission path, i.e., the pathware [6], is not only the acoustical path in the room but also includes the electrical public address system (Fig. 1).

2. SOR

A short speech pattern was used for the investigation with *SOR* in this paper, while on the contrary *SOR* [4] was originally proposed for an index of the improvement of monosyllable articulation by the Steady State Suppression method [7,8]. Therefore, *SOR* for a speech pattern s broadcasted through a pathware with an impulse response h was defined as follows:

$$SOR_i = 10\log_{10}\frac{E_{\mathrm{d},i}}{E_{\mathrm{r},i}}\ \mathrm{dB}, \tag{1}$$

$$SOR = \frac{1}{N}\sum_{i}^{N} SOR_i\ \mathrm{dB}, \tag{2}$$

where, E_d was the energy of the direct sound component: $p_\mathrm{d} = s * h_{50}$ ($*$: convolution, h_{50}: early ($< 50\,\mathrm{ms}$) impulse respons of h), E_r was the residual energy which remained after E_d was subtracted from the total sound energy E at the listening point, and subscript i described the i-th time frame (the number of the frame was N).

3. Experiment

Test stimuli S_j ($j = 1$ to 5) on the assumption that a speech pattern s was broadcasted in rooms with no prominent resonance were derived by convolutions of a male speech source s (an emergency evacuation message with the speech rate of 6.6 mora/s) and five acoustic paths h_i ($j = 1$ to 5, $T_{60} = 0.2\,\mathrm{s}$, $0.5\,\mathrm{s}$, $0.9\,\mathrm{s}$, $1.4\,\mathrm{s}$, $2.6\,\mathrm{s}$), and the other five test stimuli R_j ($j = 1$ to 5) as in the rooms with resonance were simulated by convolving S_j additionally with a filter g which was gained 14.5 dB at around 170 Hz and 4 dB at around 620 Hz.

Thirty-four Japanese native listeners: 31 males and 3 females between their twenties and sixties (average 38.4 y/o) were directed to evaluate the stimuli provided diotically through a circumaural headphone into the four listening difficulty categories: 1. Not difficult, 2. A little difficult, 3. Fairy difficult, and 4. Extremely difficult. The sum of the evaluation 2 to 4 for each test stimulus was regarded as the status of "Listening Difficulty," and *LDR* (Listening Difficulty Rating) was defined as the ratio of the sum (2 to 4) to the total (1 to 4) [3]. Figure 2 shows the relation between *SOR* and Z_{LDR} (Z-score of *LDR*) when three stimuli out of ten were omitted because their $Z_{LDR} = \pm\infty$ were not appropriate to the linear regression. Then, the regression line:

$$Z_{LDR} = -0.33SOR - 1.4 \tag{3}$$

was obtained with a high coefficient of determination, then we got

$$LDR = \frac{1 \pm \sqrt{1 - \exp\left(-\frac{2}{\pi}Z_{LDR}^2\right)}}{2} \times 100\%, \tag{4}$$

where the negative sign would be applied when $Z_{LDR} \leq 0.0$.

4. An evaluation tool for PA system

Table 1 shows the measured *STI* and the estimated *LDR* derived from the measured *SOR* using Eq. (4) for the broadcasted speech pattern in four venues where the public

*e-mail: kurisu_kiyohiro@toa.co.jp

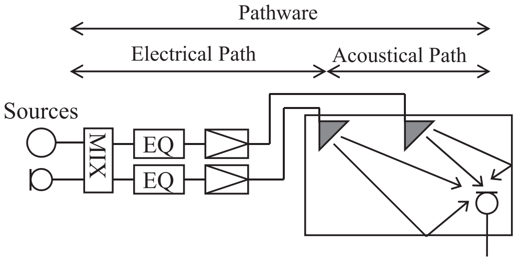

Fig. 1 Pathware = Electrical path + Acoustical path.

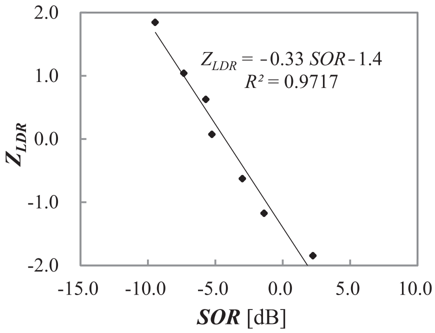

Fig. 2 *LDR* (Listening Difficulty Rating) vs. *SOR*.

Table 1 Variations of *STI*, *SOR* and the estimated *LDR*.

Venues (floor space)		Pre-tuning	Post-tuning	Amount of change
a small room ($30.9\,m^2$)	*STI*	0.70	0.70	0.00
	SOR [dB]	−7.2	−5.7	1.5
	estimated *LDR* [%]	83.0	67.6	−15.4
a hotel lobby ($300\,m^2$)	*STI*	0.68	0.69	0.01
	SOR [dB]	−5.0	−1.2	3.8
	estimated *LDR* [%]	59.1	15.4	−43.7
a concert hall ($313\,m^2$)	*STI*	0.66	0.65	−0.01
	SOR [dB]	−1.4	1.2	2.6
	estimated *LDR* [%]	17.0	3.3	−13.7
a gymnasium ($1{,}600\,m^2$)	*STI*	0.49	0.47	−0.02
	SOR [dB]	−6.6	−6.0	0.6
	estimated *LDR* [%]	77.5	71.1	−6.4

address system has been installed. *SOR* and its consequent *LDR* vary their values between before and after system tuning by engineers, while *STI* hardly changes on the contrary. The effect of the system tuning can be evaluated numerically by the improvement of the *LDR* based on *SOR*.

5. Discussion

Though *SOR* is a power ratio of the direct sound components ranging over the 50 ms interval from the initial direct sound to the other components and it is similar to C_{50}, there are differences as follows.

The frequency band required for the evaluation

SOR method utilizes a speech pattern as a test signal and measures the power ratio concerning the speech frequency band. On the contrary, C_{50} is not always specialized for the speech evaluation because it comes from an impulse response containing the characteristics of the needless frequency band.

Evaluation of a non-linear public address system

It is impossible to evaluate the non-linearity of the pathware by use of the impulse response based C_{50} because the non-linearity does not appear in the impulse response. On the contrary, *SOR* is calculated from an actual recorded speech pattern at a listening point which contains all the signal modifications within the pathware, e.g., the non-linearity of the public address system and the reverberation of the acoustical path, therefore it might be able to correspond to the non-linearity. This paper was concerned with the variety of *LDR* when the pathwares were expressed as linear filters (impulse responses) only. It is also necessary to reveal the relationship between *LDR* and *SOR* when the public address system uses a non-linear signal processing in view of the possibility that the steady state suppression method and the hyper compression processing [9] might be installed in public address systems in near future.

6. Conclusion

This paper showed the possibility that the effect of the public address system adjustment could be evaluated by *SOR* or the estimated *LDR*, which had been difficult by *STI*.

References

[1] IEC60268-16, Sound system equipment Part 16: Objective rating of speech intelligibility by speech transmission index, Edition 4.0 (2011).

[2] P. Mapp, "Is STI a robust measure of sound system speech intelligibility performance?," *First Pan-American/Iberian Meeting on Acoustics* (2002).

[3] M. Morimoto, H. Sato and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," *J. Acoust. Soc. Am.*, **116**, 1607–1613 (2004).

[4] T. Arai, Y. Murakami, N. Hayashi, N. Hodoshima and K. Kurisu, "Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking," *Acoust. Sci. & Tech.*, **28**, 438–441 (2007).

[5] K. Kurisu, S. Nakamura, K. Yasu and T. Arai, "Listening difficulty of the transmitted speech and its SOR, An evaluation tool for acoustic system tuning," *Tech. Rep. Archit. Acoust., Acoust. Soc. Jpn.*, AA2011-47 (2011) (in Japanese).

[6] D. Campbell, "STI: Where did it come from and what does it do?," *Syn-Aud-Con Newsl.*, **35**, 10–17 (2007).

[7] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 449–450 (2001) (in Japanese).

[8] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," *Acoust. Sci. & Tech.*, **23**, 229–232 (2002).

[9] E. Vickers, "The loudness war," *J. Audio Eng. Soc.*, **59**, 346–351 (2011).