Evaluation of speech intelligibility of sound fields in underground stations

Ryota Shimokura and Yoshiharu Soeta*

Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1–8–31 Midorigaoka, Ikeda, 563–8577 Japan
(Received 30 March 2010, Accepted for publication 6 September 2010)

Keywords: Speech intelligibility, Public address system, Underground station, Impulse response measurement, Sound field

PACS number: 43.28.Lv [doi:10.1250/ast.32.73]

1. Introduction

The speech transmission quality of public address (PA) systems in railway stations is not usually very high [1]. In underground stations, the speech intelligibility of the PA system may be degraded by reflection and reverberation. In a reverberant sound field, hearing the content of PA announcements can be particularly difficult for elderly people [2], and poorly heard emergency announcements may cause increased safety risks when instructing passengers to exit the station [3].

Underground railway stations are normally long enclosures in which sound does not diffuse radially. Thus, the modeling of the sound field in such a long enclosure is different from that of a diffused field [4–9]. Although models can simulate the sound fields in a long enclosure, there have not been any studies on evaluating the sound field in underground railway stations with respect to speech intelligibility.

The aim of this study is to evaluate the speech intelligibility of the sound field in underground railway stations. Impulse responses were measured in two underground stations by positioning a sound source and receiver on the ceiling and the platform, simulating the loudspeaker of a PA system and a passenger, respectively. From the impulse response, the acoustical parameters were calculated to evaluate the speech intelligibility of the PA system.

2. Impulse response measurements

2.1. Measurement stations

The stations (denoted as “A” and “B”), at which measurements were taken, are on the same underground line and have island-type platforms (one platform at the center and railway tracks on either side). The columns and ceiling are covered by metallic panels, and the platforms and railway floors are covered by stone panels and concrete, respectively. However, the lateral walls of the platforms are covered by metallic panels in station A and wooden panels in station B. The platform in station B (length: 332 m, maximum width: 8 m) is longer than that in station A (length: 162 m, maximum width: 7.3 m). Since the underground line had not yet begun operations at that time, there were no passengers or trains in the station. Therefore, the background noise was small (46 dB).

2.2. Sound source and receiver

One sound source and four receiver positions (r1 to r4) were located at the center of the platform (Fig. 1(a)). To represent the various PA loudspeakers in use in railway stations, omnidirectional sources were used to measure sound fields of the underground station [7] and train compartment [10], because the speech intelligibility and loudspeaker’s directivity were not directly related to each other [11]. The sound source we used was an omnidirectional loudspeaker (Type 4292, B&K), as shown in Fig. 2(a). It was located close to the loudspeaker of the PA system on the ceiling of the stations (Fig. 1(b)). The receiver was a dummy head microphone (KU100, Neumann), as shown in Fig. 2(b). The dummy head microphone is a model of a human head with two microphones inserted at the locations of the ear canals. It can simulate the frequency-dependent distortions of phase and amplitude in sound reaching the left and right ears. The dummy head microphone was arranged facing the sound source, and the height of the microphones in the dummy head was 1.6 m from the floor. In the stations, there are columns at 7.5 m intervals. Because of the locations of PA system loudspeakers, the sound source was close to the column (distance of 0.3 m) in station B (Fig. 2(b)). The distance was 1.2 m in station A (Fig. 2(a)).

2.3. Procedure

A sinusoidal signal with an exponentially varying frequency swept from 40 Hz to 20 kHz over a period of 18 s was used to measure the impulse responses. The responses of the receivers were deconvoluted so that the impulse responses could be obtained [12]. A laptop computer (Let’s Note, Panasonic) generated the signals and saved the responses via an AD/DA converter (AudioFire8, Echo Digital Audio) at a sampling rate of 48 kHz and a sampling resolution of 32 bits.

From the impulse responses of the dummy head microphone, the strength (G), early decay time (EDT), clarity (C50) [13], speech transmission index (STI) [14,15], and interaural cross-correlation coefficient (IACC) [15,16] were calculated to evaluate the speech intelligibility in the sound field.

3. Results

Figure 3 shows the acoustical parameters obtained in stations A and B. The values of G, EDT, and C50 were the mean of the values calculated from the impulse responses at the left and right ear positions. The G was attenuated more than 4 kHz and less than 250 Hz, respectively. Peaks of EDT were observed in the 500 Hz octave band. The C50 began to...
increase at the 500-Hz octave band. The \( IACC \) decreased up to the 500-Hz octave band, and remained constant or increased beyond the band. The \( G \), \( C_{50} \), and \( IACC \) decreased, and the \( EDT \) increased as the distance between the source and receivers became greater. The \( EDT \) in station A was longer than that in station B for \( r_3 \) and \( r_4 \).

The speech intelligibility is ranked according to the speech-weighted \( C_{50} \) and STI [17,18]. Figure 4 shows the speech-weighted \( C_{50} \) and STI as a function of the distance from the source in stations A and B. The distance was more than 15 m. Also, the speech-weighted \( C_{50} \) and STI in station A were smaller than those in station B.

4. Discussion

The acoustical parameters as a function of the distance from the sound source to the receiver, except for \( G \) and \( IACC \), were different in the two stations. At receiver positions \( r_3 \) and \( r_4 \) far from the sound source, the \( EDT \) was longer and the speech-weighted \( C_{50} \) and STI were smaller in station A. One possible reason is the difference in the interior materials at the lateral walls (metallic and wooden panels in stations A and B, respectively). Therefore, the average absorption coefficients of the stations were calculated according to the image source model for a long rectangular enclosure, because Sabine’s reverberation equation is not appropriate for a nondiffused sound field such as an underground station [5]. When the geometrical data and reverberation times at \( r_4 \) were assigned in the image source model, the average sound absorption coefficients were similar for stations A (0.17) and B (0.18).
loudspeaker in an underground station with a platform screen to investigate the effect of early reflection on the speech intelligibility.

Acknowledgments
This work was supported by a Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (18680025), a grant from Ono Acoustics Research Fund, and a Sasagawa Scientific Research Grant.

References