Applied Acoustics 89 (2015) 188–198

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Influence of sound source characteristics in determining objective speech intelligibility metrics

Peisheng Zhu [a], Fangshuo Mo [b,*], Jian Kang [c]

[a] School of Architecture and Fine Art, Dalian University of Technology, Dalian 116023, China
[b] Institute of Acoustics, Tongji University, Shanghai 200092, China
[c] School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

ARTICLE INFO

Article history:
Received 17 December 2013
Received in revised form 21 September 2014
Accepted 25 September 2014

Keywords:
Sound source characteristics
Directivity
Frequency response
Speech intelligibility
Objective speech intelligibility metric

ABSTRACT

Sound source characteristics may be one of the main causes of objective speech intelligibility metric inaccuracy. In this study, the influences of the sound source directivity and frequency response were investigated using three typical sound sources: an artificial mouth, a monitor speaker, and a dodecahedral sound source. The results show that, the simultaneous influences of directivity and frequency response on the objective speech intelligibility metric are significant, typically with a variation of 0.147 in speech transmission index (STI); sound source directivity may also result in a noticeable difference in the objective speech intelligibility metric, typically with a variation of 0.123 in STI. In comparison with sound sources with a high directivity index (DI), the measurement results for sound sources with a relatively low DI may be higher when background noise is high, and may be lower when background noise is low. The influence of sound source directivity may also depend on the room acoustic conditions, and at receiver position where reflections are abundant, the influence of sound source directivity may be more significant. Not applying frequency response equalisation resulted in large errors in the values being measured, which deviate from the real values of STI by up to 0.172, depending on the original frequency response characteristics of the sound sources that are used.

1. Introduction

Speech intelligibility is an important metric and can be used to evaluate the sound transmission quality of auditorium, public address system and communication system. Attempts for objective evaluation of speech intelligibility began in the 1920s, and the first objective evaluation metric, the articulation index (AI) [1], that was developed into SII (speech intelligibility index) recently [2], was put forward subsequently in the 1940s to 1950s. At the end of the 1970s, other metrics, such as speech transmission index (STI) [3] and articulation loss of consonants (%ALcons) [4], were created. STI and SII are currently the two most commonly used objective evaluation metrics for speech intelligibility, corresponding to two current standards: IEC 60268-16 [5] and ANSI S3.5 [2], respectively.

There are many factors affecting STI and SII measurement results, and one of them is the characteristics of the sound source. For sound sources, directivity and frequency response are important characteristics that can influence the results of STI and SII measurements [6–9]. In IEC 60268-16 [5], it is specified clearly that STI measurement shall be conducted using a sound source with directivity and radiation pattern similar to those of the average human mouth and an omnidirectional microphone, and there are also corresponding specifications for the frequency response of sound sources. In ANSI S3.5 [2], there is no clear specification for the directivity of the sound sources, and both directional sound sources and omnidirectional sound sources can be used. However, there are some specifications for the frequency response of the sound sources.

Bozzoli and Farina [6] conducted a study on the influence of the directivity of three artificial mouths on STI measurements in different acoustic environments. The study shows that the measurement of STI is not strongly influenced by the directivity of the artificial mouths, for room acoustics applications because of the substantial distance between speaker and receiver, and the presence of numerous reflections. However, in their study the influence of frequency response was not investigated. Another study conducted by Mapp [7] shows that equalisation can significantly affect the intelligibility. However, the study focuses particularly on the influence of sound system equalisation on speech intelligibility, and the influence of directivity was not investigated. Petra and Hongistob

* Corresponding author.
E-mail address: mfs@tongji.edu.cn (F. Mo).

http://dx.doi.org/10.1016/j.apacoust.2014.09.023
0003-682X/

[8] compared the STI and SII measurement methods and suggested that loudspeakers possibly have a considerable influence on the STI and SII measurements. Peng et al. [9] evaluated subjective Chinese speech intelligibility using three sources with different directional patterns: an omnidirectional source, a source with directivity similar to a human speaker, and a human speaker in both real and virtual rooms with different reverberation times. The results show that speech intelligibility scores obtained using an omnidirectional source are lower than those obtained using the other two sources. However, the influence of frequency response was not considered. Overall, there is still a lack of study of the systematic and simultaneous influence of frequency response and directivity of sound sources on the intelligibility.

The aim of this study is therefore to systematically investigate the influence of sound source characteristics, including frequency response and directivity in determining objective speech intelligibility metrics. This paper starts with selecting three typical sound sources with different directivities and frequency responses; then the full STI of eight receiver positions, and a total of 32 sound environments in three rooms were measured. Finally, analyses of the results are presented.

2. Methods

2.1. Sound sources

In this study, three typical sound sources were used: (1) An artificial mouth GRAS 44AA with a directivity and radiation pattern similar to those of the average human mouth, which is the standard sound source for measuring STI recommended in IEC 60268-16 [5]; (2) a monitor loudspeaker GENELEC 8020B (cone diameter is 4 in.), which is the alternative sound source for measuring STI recommended in IEC 60268-16 [5]; and (3) a dodecahedral sound source B&K 4292L, which is one of the sound sources that can be used for measuring SII in ANSI S3.5 [2]. Although the three sound sources are recommended or allowed to be used in the standards, they are rather different in acoustic characteristics, having completely different frequency response and directivity patterns.

The relative amplitude in relation to 1000 Hz of the three sources was measured in an anechoic chamber using impulse responses. For the dodecahedral sound source 4292L, because there was no main radiation and the directivity changed with orientations, an average of the 20 frequency responses from the 20 measurement points with solid angles covering the measurement sphere based on ISO 3745 [10] was used as the equivalent frequency response. The results are shown in Fig. 1. It can be seen that the frequency response of the monitor loudspeaker 8020B is the best of the three sound sources, and the response is almost flat in the entire frequency range, which also meet the specification in IEC 60268-16 [5] that "over the range 88 Hz to 11300 Hz, the 1/3 octave frequency response of the test signal source is within ±1 dB when measured in a free field" without frequency response equalisation. The frequency response of the artificial mouth 44AA is poor, which is strongest at 1000 Hz but decreases greatly at other frequencies, especially at 125 Hz and as the frequency exceeds 2000 Hz. The frequency response of the dodecahedral sound source 4292L is not good either, which is relatively strong at 125–250 Hz or so, but decreased as the frequency increases once the frequency exceeds 125 Hz.

The relative amplitude in relation to 1000 Hz of the artificial mouth 44AA and the dodecahedral sound source 4292L, after being equalised by inverse filtering of the frequency response, was measured in an anechoic chamber with the same layout for the artificial mouth 44AA and the layout rotating by 180° about the *z*-axis for the dodecahedral sound source 4292L, and the results are shown in Fig. 2. It can be seen that the frequency responses for both the artificial mouth 44AA and the dodecahedral sound source 4292L are almost flat in the entire frequency range, both meet the specification in IEC 60268-16 [5].

In Fig. 2, the inverse filters were generated from the sound source frequency response measured in an anechoic chamber by employing the Kirkeby method [11,12]. For the dodecahedral sound source 4292L, the equivalent frequency response $\bar{H}(f)$ that was used to generate the inverse filter was the average of 20 frequency responses $H_n(f)$ from 20 measurement points with solid angles covering the measurement sphere, which was based on ISO 3745 [10]. $\bar{H}(f)$ can be calculated using Eq. (1):

$$\bar{H}(f) = |\bar{H}(f)|e^{j\angle\bar{H}(f)} \tag{1}$$

where $|\bar{H}(f)|$ is the equivalent amplitude response and $\angle\bar{H}(f)$ is the equivalent phrase response. The equivalent amplitude response $|\bar{H}(f)|$ can be calculated using Eq. (2):

$$|\bar{H}(f)| = \sqrt{\frac{1}{4\pi}\sum\nolimits_{n=1}^{N}\Omega_n|H_n(f)|^2} \tag{2}$$

where $N = 20$ and $\Omega_n = \frac{\pi}{5}$ for equal solid angles ($n = 1$ to 20) in the measurement. The equivalent phrase response $\angle\bar{H}(f)$ can be calculated by integrating the energy-weighted average group delay $\bar{\tau}(f)$, which can be calculated using Eq. (3):

$$\bar{\tau}(f) = -\frac{d\angle\bar{H}(f)}{2\pi\cdot df} = -\frac{1}{4\pi|\bar{H}(f)|^2}\sum_{n=1}^{N}\Omega_n|H_n(f)|^2\frac{d\angle H_n(f)}{2\pi\cdot df} \tag{3}$$

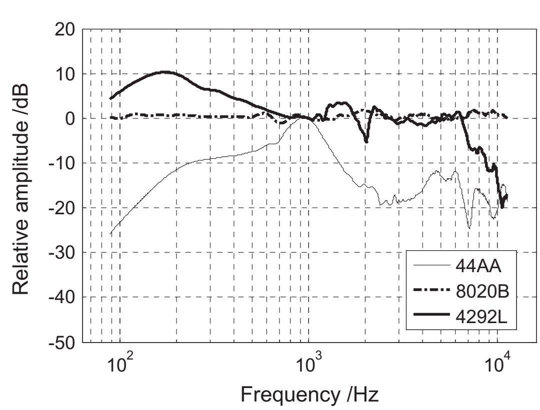

Fig. 1. The relative amplitude in relation to 1000 Hz of the three sound sources, measured in an anechoic chamber.

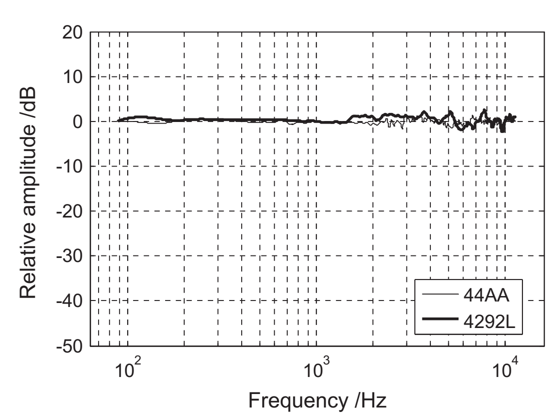

Fig. 2. The relative amplitude in relation to 1000 Hz of the artificial mouth 44AA and the dodecahedral sound source 4292L after being equalised by inverse filtering of the frequency response, measured in an anechoic chamber.