

Characterization of speech understanding in various types of noise

Lena L. N. Wong^{a)} and Elaine H. N. Ng
University of Hong Kong, Pokfulam Road, Hong Kong

Sigfrid D. Soli
House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 1 October 2010; revised 9 December 2011; accepted 27 August 2012)

This study examined (1) the effects of noise on speech understanding and (2) whether performance in real-life noises could be predicted based on performance in steady-state speech-spectrum-shaped noise. The noise conditions included a steady-state speech-spectrum-shaped noise and six types of real-life noise. Thirty normal-hearing adults were tested using sentence materials from the Cantonese Hearing In Noise Test (CHINT). To achieve the first aim, the performance–intensity function slopes in these noise conditions were estimated and compared. Variations in performance–intensity function slopes were attributed to differences in the amount of amplitude fluctuations and the presence of competing background speech. How well the data obtained in real-life noises fit the performance–intensity functions obtained in steady-state speech-spectrum-shaped noises was examined for the second aim of the study. Four out of six types of noise yielded performance–intensity function slopes similar to that in steady-state speech-spectrum-shaped noise. After accounting for individual differences in sentence reception threshold (SRT) and the offset between the signal-to-noise ratio for 50% intelligibility across different types of noise, performance in steady-state speech-spectrum-shaped noise was found to predict well the performance in most of the real-life noise conditions. © 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4751538>]

PACS number(s): 43.71.An, 43.71.Gv, 43.71.Es [AA]

Pages: 2642–2651

I. INTRODUCTION

Having a conversation in noise is more difficult than in quiet, particularly for those with a hearing impairment (Wilson *et al.*, 1990). Although competing human speech is the most common interference in conversation, environmental noise, such as traffic or industrial noise, is also common in daily encounters (Plomp, 1978). The characteristics of background noise differ from place to place in real-life situations and it is uncertain whether speech recognition ability measured in steady-state speech-spectrum-shaped noise, which is commonly used in clinical speech audiometric procedures, would predict one's speech understanding in real-life situations.

Only a few published studies have investigated the effects of environmental noise on speech intelligibility. For example, Prosser *et al.* (1991) examined speech discrimination using sentences in four types of noise: speech noise, cocktail party noise, traffic noise, and continuous discourse. At unfavorable signal-to-noise ratios (SNRs), i.e., at 0 dB SNR or below, performance in traffic noise and continuous discourse was generally better across all subjects compared to that in speech noise and cocktail party noise. The authors attributed the differences to lower energy at high frequencies, thus favoring the perception of speech cues; and more amplitude fluctuations in traffic noise and continuous discourse, compared to speech noise and cocktail party noise. However, the characteristics of the test materials and noises

used in the study were not documented, making comparison to other studies difficult.

Danhauer and Leppler (1979) studied speech understanding ability in normal-hearing listeners using the California Consonant Test (Owens and Shubert, 1977) presented in four types of noise: Four-talker competitor, nine-talker competitor, cocktail party noise, and white noise. At low SNRs (0 and +5 dB), speech understanding in cocktail party noise and white noise was better than performance in four- or nine-talker competitors. The authors suggested that the additional effect of informational masking of the multi-talker competitors could account for the findings. Rhebergen and colleagues (2008) compared speech intelligibility in 12 real-life noises using Dutch sentences. They concluded that for individuals with normal hearing, masking was more efficient in stationary-like noises, such as crowd, car, and construction noises, than fluctuating or interrupted noises, such as machine gun and bird noises, when the spectra between noise and speech spectra were similar.

Previous research also demonstrated that background noise with different characteristics would yield different masking effects on speech understanding. For instance, amplitude modulations or fluctuations (Gustafsson and Arlinger, 1994); energy allocation in the frequency spectrum of noise (Kaplan and Pickett, 1982; Prosser *et al.*, 1991); and the existence of linguistic information in background noise (Bacon *et al.*, 1998) may have differential effects on speech intelligibility. However, comparison of results is somewhat difficult because these studies differed in the types of background noise, scoring method, and test materials used;

^{a)}Author to whom correspondence should be addressed. Electronic mail: llnwong@hku.hk

syllables, words, and sentences that differ in the amount of contextual cues were used across studies.

Speech recognition or speech intelligibility test is one of the most common clinical tools to measure functional hearing ability. The slope of speech recognition performance–intensity function was found to be influenced by a number of factors. First, the use of different speech materials, such as sentences, words, or nonsense syllables, affects the slope of the function (McArdle *et al.*, 2005; Studebaker *et al.*, 1994). Sentences, which contain contextual cues, produce a steeper function (Studebaker *et al.*, 1987). Second, if the frequency spectrum of the background noise matches that of the talker’s voice, the slope of the performance–intensity function would become steeper, compared to when the background noise and target speech exhibit different spectra. Finally, noise with more temporal fluctuations also tends to cause a flatter slope in the performance–intensity function in speech recognition tasks compared to a steady-state speech-shaped noise. Temporal gaps in noise allow the listener to hear some speech cues even in a low or very difficult SNR (Eisenberg *et al.*, 1995). Speech understanding or recognition tasks in unmodulated or steady-state background noise would yield a steeper slope of the performance–intensity function than modulated noise in normal-hearing population, whether sentences (Takahashi and Bacon, 1992), words (Stuart *et al.*, 1995), or nonsense syllables (Eisenberg *et al.*, 1995) are used as target stimuli.

In clinical situations, steady-state speech-spectrum-shaped noise is often used to assess speech intelligibility and the performance does not necessarily reflect speech understanding ability in real-life situations. A study to relate performance in various types of noise situation is thus important. Laroche *et al.* (2003) has found performance obtained in the Canadian French version of the hearing in noise test (HINT) could be used to predict functional hearing abilities in hearing-critical jobs such as the Canadian Coast Guards. With all language versions of the HINT being developed using the same paradigm (Soli and Wong, 2008), the Cantonese version of the hearing in noise test (CHINT) also has the potential to be used to predict speech understanding in other real-life situations.

The present study was aimed at (1) investigating the effect of steady-state speech-spectrum-shaped noise and six types of real-life noise on speech understanding and (2) examining whether performance in these situations could be predicted based on performance in a background of steady-state speech-spectrum-shaped noise. The masking effect of different types of noise on speech understanding was investigated based on the slope of the performance–intensity functions and the estimated SNR to yield 50% intelligibility. Results obtained in steady-state speech-spectrum-shaped noise were than compared to those obtained in other types of real-life noise.

II. METHOD

A. Participants

Two different groups of participants were recruited and all of them have mainly resided in Hong Kong and speak Cantonese as their first language. The first group of 10 participants rated the naturalness of recordings of environmental noises to

identify noise samples for use in the study. To obtain data to plot the performance–intensity function slopes, a second group of 30 participants was recruited. These were 15 males and 15 females, at 18–35 years of age [$M = 22.90$, standard deviation (SD) = 4.33]. They had audiometric pure-tone air-conduction thresholds not worse than 15 dB hearing level across the octave frequencies from 250 to 8000 Hz in both ears and reported no history of otological problems or difficulties in speech understanding. These participants presented with otoscopic and normal tympanometric findings (Wan and Wong, 2002) that are consistent with the absence of outer and middle ear disorders.

B. Recording of environmental noises

Six environmental noises were recorded in the upper deck of a bus, in the lower deck of a bus, in a café, in a Chinese restaurant, in a subway train, and on a busy street with a reasonable amount of traffic noise. The noises were recorded using a microphone with mounting brackets connected via a sound mixer Audio-Technica AT 877 (Tokyo, Japan) to a laptop computer. COOL EDIT version 2.0 was used to record the noises on the laptop computer.

Continuous samples of each environmental noise were recorded for at least 10 min and their sound pressure levels were noted. The six real-life noises were all recorded at times that were representative of listening in that particular situation. Three 1 min sound segments of each noise were chosen and the naturalness of these noise samples was rated. Naturalness refers to how well the noise samples resemble the typical background noise in the corresponding environments and was rated using a seven-point scale, with 1 representing the least natural and seven presenting the most natural. The sample with the highest score in each condition was chosen and used in the study and their naturalness ratings were all above 5, out of 7 points. The real-life noises were played back continuously at an average sound pressure level in the corresponding real-life situation (see Table I). The above-mentioned procedures were used to make sure that the real-life noises used in the present study are representative of the actual environment.

These noises vary in the amount of amplitude fluctuations, frequency spectra, and level. Figure 1 shows frequency spectra of the six noises at 200–6400 Hz, as compared to the steady-state speech-spectrum-shaped noise. According to Fig. 1, the long-term average spectra of noises recorded on the street and on the lower deck of the bus were similar to that of the steady-state speech-spectrum-shaped noise (the reference noise in the study). Noise recorded on the upper deck of a bus exhibited less low-frequency energy than the reference noise; whereas noise recorded in a subway train, a café, and a restaurant contained greater low-frequency energy than the reference noise.

Figure 2 shows 1 min samples of the waveforms and Fig. 3 shows the time–frequency analyses of the noises, arranged in increasing amounts of amplitude fluctuations. Time–frequency analysis for each noise was performed at 1 s intervals and in 100 Hz intervals from 150 to 6450 Hz. There are virtually no amplitude fluctuations in steady-state speech-spectrum noise. Speech signals are noted in most of

TABLE I. Noise levels used in each noise condition and assigned SNRs that were expected to yield approximately 40%, 60%, and 80% intelligibility.

Type of noise	Noise level (dB A)	SNR (dB) to achieve the desired intelligibility		
		80%	60%	40%
Steady-state speech-spectrum-shaped	65	-5	-7	-8
Bus lower deck	72	-1	-3	-5
Bus upper deck	67	0	-2	-4
Café	69	-1	-3	-5
Chinese restaurant	72	-4	-6	-7
Street	73	-2	-4	-6
Subway train	73	-4	-5	-7

the environmental noises, except street noise. Noises recorded in a café and a Chinese restaurant exhibit the greatest amounts of amplitude fluctuations, due to conversation and clattering of dishes. In other words, these types of noise contain the greatest amounts of amplitude fluctuation among all noises used in the study, as well as informational masking. Periodic amplitude modulations are also noted in the noise recorded on the lower deck of a bus due to door opening and closing and in street noise as cars drove pass the recording site. The upper deck of a bus is often quieter than the lower deck where there was more people flow.

C. Sentence materials

Sentences from the 20-sentence list version of the CHINT (Wong and Soli, 2005) were used. There are 12 lists and each sentence contains 10 characters and is simple and conversational in style. Frequencies of distribution of the phonemes in the sentence lists reflect those in general conversational Cantonese spoken in Hong Kong (Leung *et al.*, 2004). Syntax of

the sentences was not complicated and the difficulty of the sentences has been equated across all lists.

D. Procedures

Sentences were presented to the second group of participants in the sound field via a loudspeaker placed 1 m in front of them in a sound-treated booth (2.7 m × 2.5 m) with reverberation time of 0.11 s at 500 Hz, which met ANSI standards for ambient noise levels for sound field testing (ANSI, 1999). Participants were instructed to verbally repeat the sentences as accurately as possible. Two practice lists were administered, followed by a measurement of sentence reception threshold (SRT) using one 20 sentence list. The SRT, defined as the SNR in decibels (dB) in which one could repeat half of the sentences (Wong and Soli, 2005), was then obtained with speech and noise presented from the loudspeaker situated at 0° azimuth using the standard adaptive procedure in the CHINT. In the standard adaptive procedure, both speech and noise were first presented at 65 dB A-weighted. The presentation levels of the speech varied according to the participant's response. For the first four sentences, when a sentence was repeated correctly, the level of the following sentence would be decreased by 4 dB; otherwise, the level would be increased by 4 dB. The presentation level of the fifth sentence was set as the average presentation level for the first four sentences. The step size of the adaptive procedure was refined to 2 dB from the fifth sentence onward.

After obtaining the SRT, each individual was tested in three of the seven noise conditions using the other eleven 20 sentence lists. The test conditions for each participant and lists to be used in each test condition were counterbalanced across all participants. The order of test conditions and SNRs used in each condition was randomized.

In each noise condition, participants were tested using three SNRs, which were derived from a pilot study using the 10 sentence list version of the CHINT, to yield approximately 40%, 60% and 80% speech intelligibility. The corresponding SNRs in the seven noise conditions are shown in Table I, and the sentences were presented at levels based on these SNRs. The steady-state speech-spectrum-shaped noise was used in the original CHINT program; its properties and recording procedures are described in Wong and Soli (2005). The presentation levels of the six environmental noises and the steady-state speech-spectrum-shaped noise are also shown in the same table. Participants were given the same

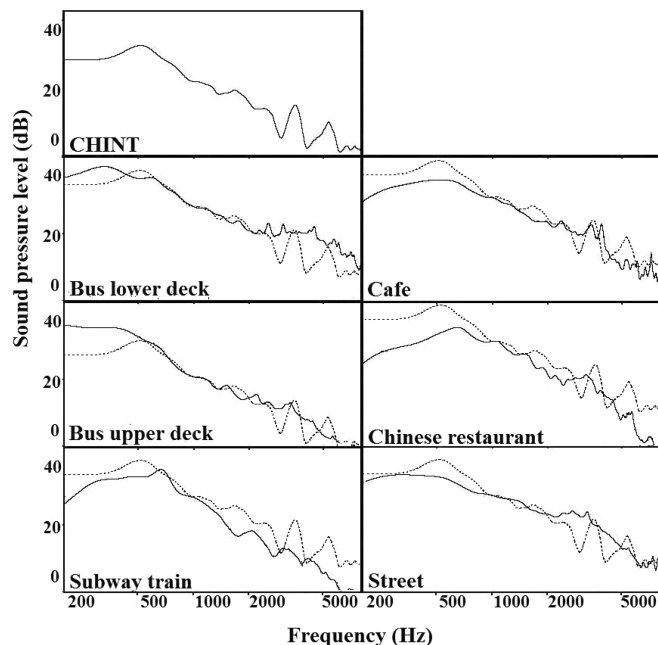


FIG. 1. Frequency spectra of the six real-life noises (solid line) in comparison to the frequency spectrum of the steady-state speech-spectrum-shaped noise (dotted line) from 200 to 6400 Hz. The noises were aligned at 1000 Hz with the steady-state speech-spectrum-shaped noise, despite differences in sound pressure level.

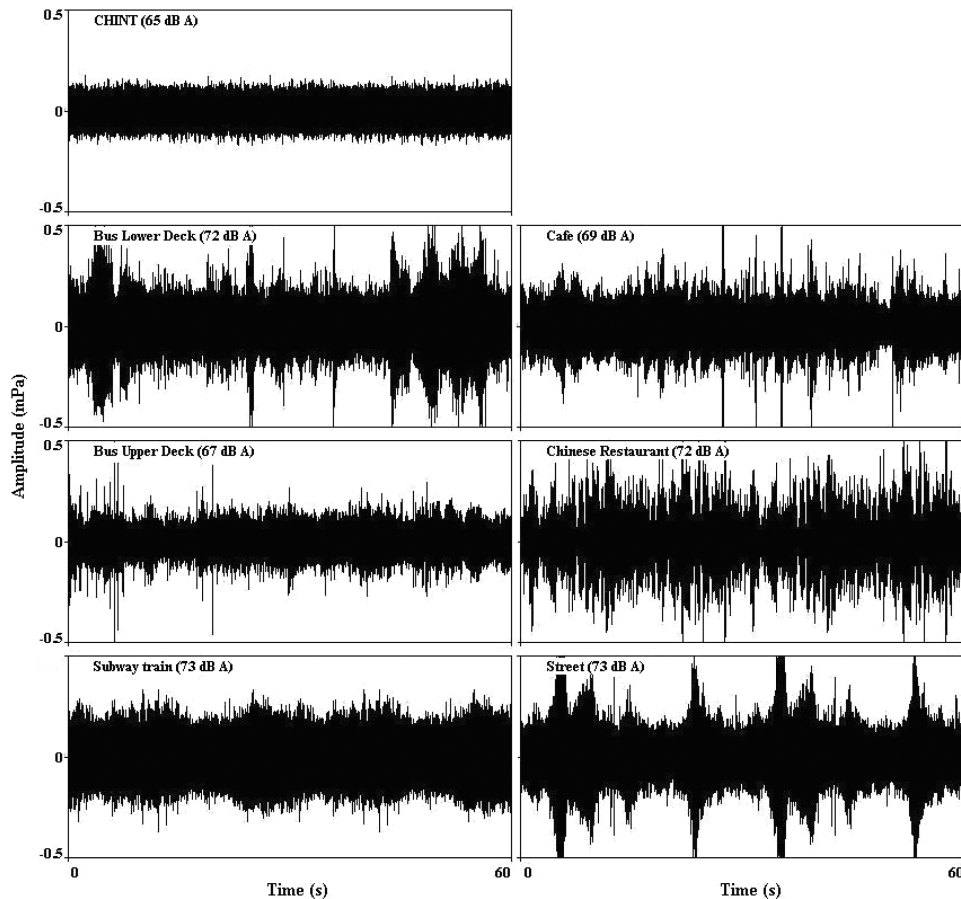


FIG. 2. One minute samples of the waveforms of the seven noises.

instructions as above and scoring was based on the number of words correctly repeated in each sentence for these 11 lists.

The sentence lists and steady-state speech-spectrum-shaped noise were presented using the HINT program (version 5.0.5) on a computer, which is connected to a Madsen OB822 (Taastrup, Denmark) clinical audiometer. Other background noises were played back using a Sony (Tokyo, Japan) compact disk player CDP-XE200. All speech and noise stimuli were mixed using an Allen and Heath GR1 sound mixer (Cornwall, UK) before they were played back via a set of Westra Lab-501 (Meitingen, Germany) loudspeakers in the sound field. Speech and noise levels were calibrated to the desired levels using a Bruel and Kjaer Type 4144 microphone (Nærum, Denmark) connected to a sound level meter, with the microphone situated ~ 1 m away from the loudspeaker. Calibration of test stimuli was performed prior to subject testing.

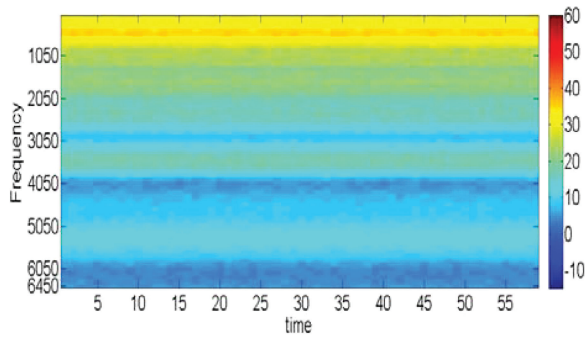
We aimed to obtain three pairs of SNR-intelligibility data points in each noise condition from each participant. These data points should lie within the range of 30%–90% (i.e., $40 \pm 10\%$ – $80 \pm 10\%$) intelligibility, which represents the linear portion of the performance–intensity function for sentences from the CHINT. The assigned SNRs were also adjusted according to the individual SRT obtained. That is, if a participant had a SRT 1 SD above the published mean SRT (see Wong and Soli, 2005), the assigned SNRs would be elevated by 1 dB in all test conditions. Theoretically speaking, each subject was tested using three 20 sentence lists in each of three noise conditions, therefore nine pairs of SNR-intelligibility score data points were obtained. However, if any of these

scores lie outside the range of 30%–90%, the data point would be discarded. The participant would then be re-tested using a different SNR with one of the two remaining 20 sentence lists. This procedure was done to ensure that all intelligibility scores obtained were within the range of 30%–90% correct in each condition for each subject. Hence, not all the subjects were tested at the same SNRs. In addition, when the scores did not cover the lower range of 30%–50% or the upper range of 70%–90%, extra data points would be obtained until all lists had been exhausted. For instance, if the three scores obtained using the three assigned SNRs were all between 30% and 60% speech intelligibility, an extra list targeting 80% speech intelligibility was tested. Therefore, different numbers of data points were obtained on each participant.

Ethical approval was obtained at the University of Hong Kong. Prior to testing, informed consent was obtained.

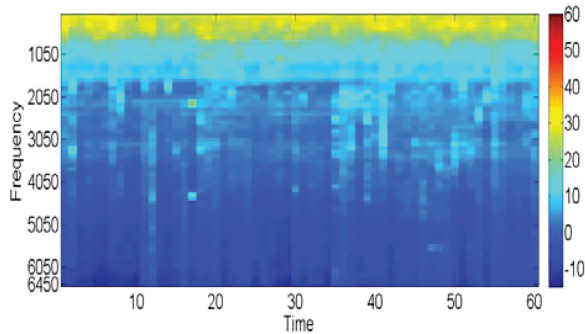
E. Data analysis

A total of 237 pairs of SNR-intelligibility data points within the 30%–90% range were obtained. The performance–intensity functions were generated by pooling all the data points for each of the seven noise conditions to determine the best fit linear regression equation for the data using spss version 13.0 for Windows (New York). The slope of the performance–intensity function generated using this technique would be the same as the one generated by averaging the slope of the performance–intensity function obtained for each subject (Studebaker *et al.*, 1994). The slope of the equation obtained describes the improvement in speech understanding as a



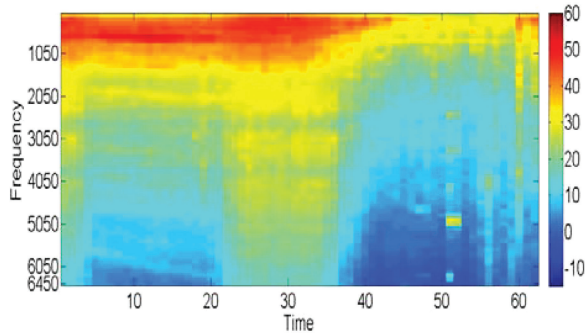
Steady-state speech spectrum
shape (CHINT) noise

(10.91%/dB)



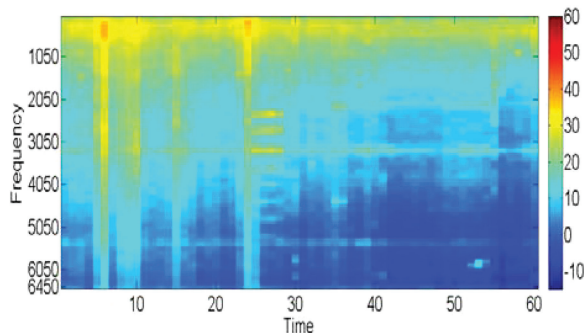
Upper Deck of a Bus

(9.55%/dB)



Subway train

(9.17%/dB)



Lower deck of a bus

(8.6%/dB)

FIG. 3. Time–frequency analysis for each type of noise used in the study arranged in ascending order in the amounts of amplitude fluctuation. The noises were sampled every 1 s in 100 Hz intervals from 150 to 6450 Hz. The slope values of the performance–intensity functions are listed as well.

function of SNR in each noise condition. The SNR for 50% intelligibility in each noise condition was calculated using these regression equations.

III. RESULTS

The mean SRT in noise in the sound field was -4.76 dB SNR (SD = 0.89 dB SNR), which is also within 1 SD of the published norm (see Wong and Soli, 2005). How well performance can be predicted in various real-life noises is discussed in the following.

First, the slopes of the performance–intensity functions were compared and the estimated SNR for 50% intelligibility

was calculated for each noise condition in order to determine whether performance in other types of noise could be predicted from performance in steady-state speech-spectrum-shaped noise. Scores that ranged from 30% to 90% at various SNRs were pooled according to noise conditions. Linear regression functions were fitted. Table II shows the r^2 and the slope values of the regression equations and the estimated SNRs to yield 50% intelligibility in all seven conditions. The r^2 values obtained ranged from 0.72 (in upper deck of a bus) to 0.87 (in café noise). The slopes of the regression equations describing the performance–intensity functions varied from 6.42%/dB (in street noise) to 10.91%/dB (in steady-state speech-spectrum-shaped noise). Four noise conditions

TABLE II. r^2 , performance–intensity function slopes, and the estimated SNRs for 50% intelligibility in each noise condition; the conditions have been arranged in ascending order of estimated SNR for 50% intelligibility.

Type of noise	r^2	Slope (%/dB)			dB SNR for 50% correct		
		Mean	Range of 95% CI		Value	Range of 95% CI	
Steady-state speech-spectrum-shaped	0.86	10.91	9.24	12.58	-8.21	-7.14	-9.27
Chinese restaurant	0.85	10.82	9.24	12.39	-6.72	-5.86	-7.58
Café	0.87	10.62	9.13	12.11	-4.43	-3.95	-4.91
Bus upper deck	0.72	9.53	7.34	11.73	-3.25	-2.66	-3.83
Subway train	0.74	9.17	7.36	10.99	-6.35	-5.25	-7.45
Bus lower deck	0.80	8.60	7.06	10.18	-4.32	-3.72	-4.93
Street	0.75	6.42	5.07	7.77	-6.07	-5.05	-7.09

produced performance–intensity function slope values that are within the 95% CI range of slope value obtained in steady-state speech-spectrum-shaped noise. These conditions are upper deck of a bus, café, subway train, and Chinese restaurant. Slopes of performance–intensity functions obtained in noise recorded in the lower deck of a bus and on the street are more gentle than those obtained in other noise conditions.

The SNRs for 50% intelligibility were estimated using the corresponding regression equations. Table II shows that the best SNR for 50% speech intelligibility was obtained in steady-state speech-spectrum-shaped noise (-8.21 dB SNR). The effects on intelligibility in different noises could roughly be divided into two categories. The first category includes listening in three noises (upper and lower deck of a bus and in a café) that resulted in SNRs ranging from -4.43 to -3.25 dB for 50% intelligibility. Compared to listening in steady-state speech-spectrum-shaped noise, these SNRs were elevated by ~4 dB. The second category includes listening in three other noises (street, subway train, and Chinese restaurant) that resulted in SNRs from -6.72 to -6.07 dB. Performance was elevated by ~2 dB, compared to results obtained in steady-state speech-spectrum-shaped noise.

In order to examine whether speech intelligibility in steady-state speech-spectrum-shaped noise predicts that in real-life noises, the SNR-intelligibility data points obtained in the six real-life noises were plotted against the 95% confidence interval (CI) of the slope value obtained in steady-state speech-spectrum-shaped noise (Fig. 4). Overall, 46%, 12%, and 12% of the data points obtained in noises recorded in a Chinese restaurant, in a subway train, and on a street lie in the 95% CI, respectively. None of the data points in noises recorded on a bus and a café fit into these CIs. These noises also yielded comparatively more elevated SNRs for 50% intelligibility.

To further explore how well the steady-state speech-spectrum-shaped noise predicts performance in the real-life noises, the SNRs used for each noise condition were then normalized against the SNR that yielded 50% intelligibility in steady-state speech-spectrum-shaped noise. That is, SNR used in each noise condition was corrected by the amount of the offset in SNRs for 50% intelligibility obtained in that noise condition and in steady-state speech-spectrum-shaped noise. The percentage of sets of corrected SNR-intelligibility data points that lie within 95% CIs of the regression line derived from testing in steady-state speech-spectrum-shaped noise was identified.

Figure 5 shows the corrected SNR-speech intelligibility data and their relationships with the 95% CI of the performance–intensity function slope derived from testing in steady-state speech-spectrum-shaped noise. Overall, 88% of sets of the corrected SNR-speech intelligibility data (176 of 200 data points) fit well into the 95% CI of the performance–intensity function derived from performance in steady-state speech-spectrum-shaped noise. Figure 5 also shows that 5 sets of data obtained in noise recorded on the upper deck of a bus, 3 sets of data obtained in noise recorded on the lower deck of a bus, 1 set obtained in café noise, 3 sets obtained in subway train noise, and 12 sets obtained in street noise did not fit into this 95% CI range. However, these sets of data points, except those in the street noise condition, are only marginally outside the 95% CI range.

In addition, root mean square (rms) errors associated with using the performance–intensity function obtained in steady-state speech-spectrum-shaped noise to predict performance in real-life noises were examined. The highest rms errors were observed when predicting performance in noise from the upper deck of a bus (11.82%), the lower deck of a bus (11.28%), and a street (10.90%). rms errors were 6.28% when performance was predicted in a café, 6.23% in subway train, and 6.01% in Chinese restaurant. These are comparable to the rms error obtained (5.97%) when the same performance–intensity function was used to predict performance in steady-state speech-spectrum-shaped noise.

IV. DISCUSSION

Overall, various real-life noises exhibited differential effects on speech understanding. Using the performance–intensity function derived from testing in steady-state speech-spectrum-shaped noise, performance in at least four types of real-life noise (upper deck of a bus, café, subway train, and Chinese restaurant) could be predicted in the present study. Informational masking and amount of amplitude fluctuation seemed to have impacts on the performance–intensity function slopes in various types of noise.

A. Prediction of performance in various types of noise

The linear regression equations describe well the relationships between SNRs and intelligibility scores, as much of the variance in the data has been accounted for. Steady-state speech-spectrum-shaped noise in the present study

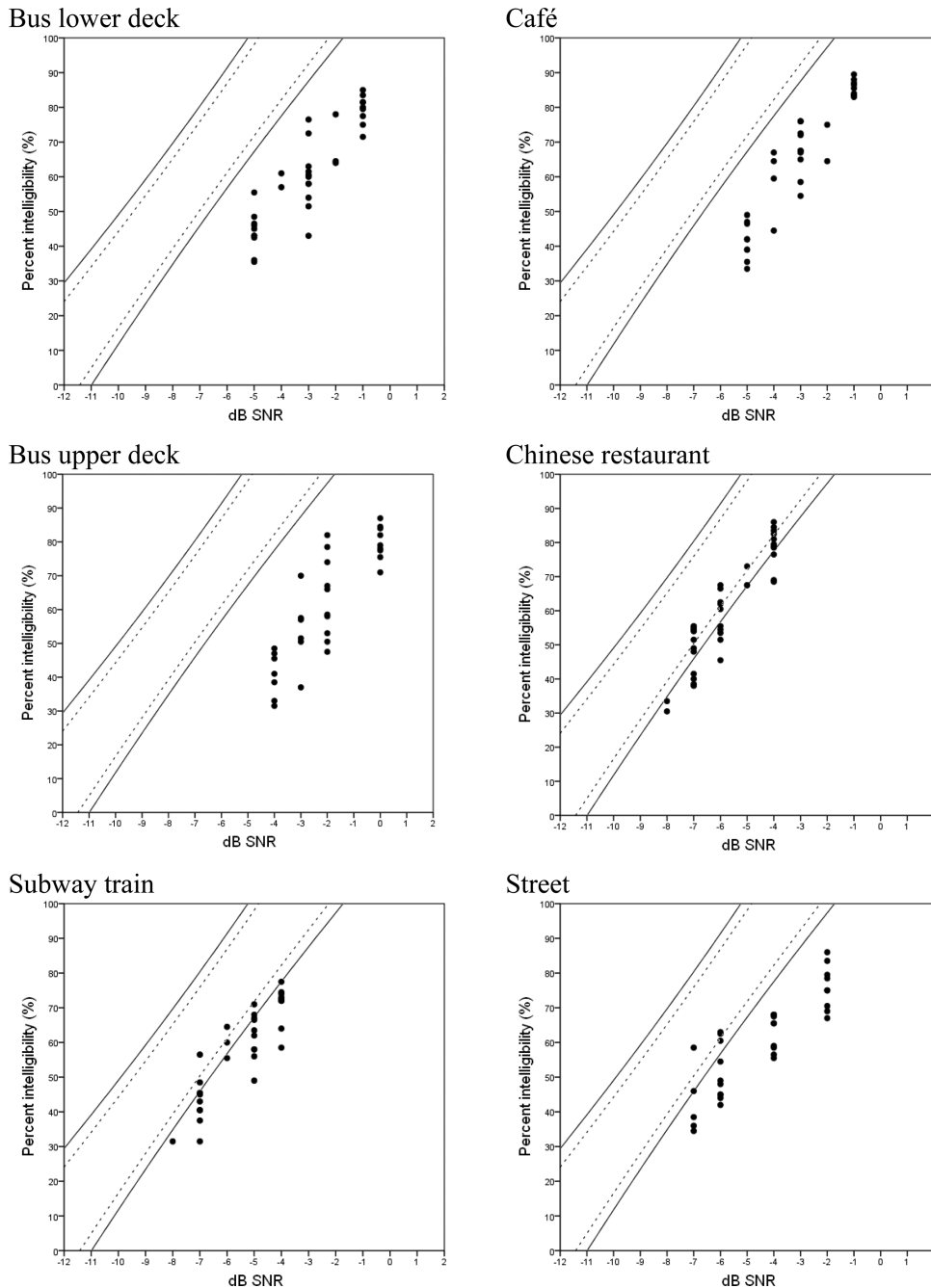


FIG. 4. Scatter plots of SNR-intelligibility score data points. The dashed lines and solid lines represent the 95% and 99% CI of the slope obtained from testing in steady-state speech-spectrum-shaped noise.

produced the steepest performance–intensity function slope, because this type of noise matches the frequency spectrum of the talker’s voice (Studebaker *et al.*, 1987) and contains few temporal fluctuations (Takahasi and Bacon, 1992). The performance–intensity function slope generated using this steady-state speech-spectrum-shaped noise is in agreement with that reported by Wong and Soli (2005) and other similar materials (Soli and Wong, 2008; Vaillancourt *et al.*, 2005).

In addition to the properties of the performance–intensity functions of the real-life noises, whether performance in steady-state speech-spectrum-shaped noise predicts that in the real-life noise conditions was also examined. In the present study, the slopes of the performance–intensity functions obtained in four real-life noise conditions (upper deck of a bus, café, subway train, and Chinese restaurant) are highly similar to the one obtained in steady-state speech-spectrum-

shaped noise. Based on the results shown in Fig. 5 and the rms errors associated with using the performance–intensity function obtained in steady-state speech-spectrum-shaped noise to predict performance in real-life noises, we can conclude that performance in these four listening situations could be predicted. That is, once variations in individual SRT and the offset in SNR to achieve 50% intelligibility in these noise conditions have been accounted for, performance in four types of real-life noises could be predicted using the performance–intensity function established in steady-state speech-spectrum-shaped noise. In clinical situations where there is limited time, being able to predict performance in various types of noise is desirable.

This model of prediction, which was established based on results obtained in normal hearing listeners, will require validation on those with various degrees of hearing

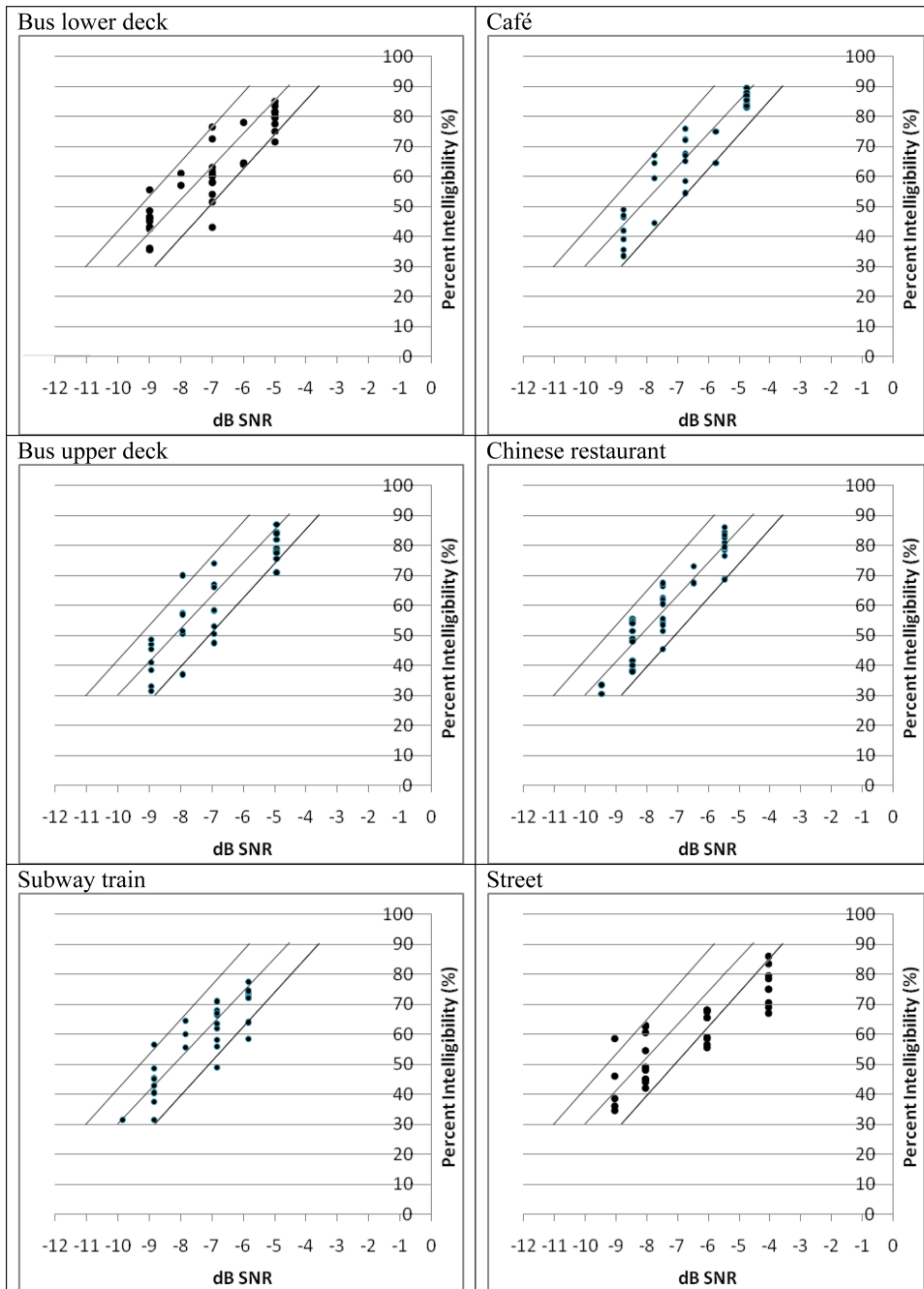


FIG. 5. (Color online) Scatter plots of pairs of corrected SNR-intelligibility score data points. The performance-intensity function (middle line) and the 95% CI of the slope (the other two lines) obtained from testing in steady-state speech-spectrum-shaped noise.

impairment. Although some research (e.g., [Studebaker et al., 1993](#)) demonstrated that hearing-impaired listeners would yield the same performance-intensity function slope values and SNR for 50% speech recognition score in noise as normal-hearing listeners, other research shows that hearing-impaired listeners may produce a flatter performance-intensity function than the normal-hearing group (e.g., [Eisenberg et al., 1995](#)). How the presentation levels of test stimuli should be adjusted to compensate for the degree of hearing impairment should also be examined.

B. Effects of frequency spectra, informational masking and temporal fluctuation

Three factors could have affected the performance-intensity function slope values obtained in various types of

noise. These include variations in frequency spectra, informational masking, and temporal fluctuations of the noise.

At first glance, the frequency spectra of the noises recorded in the lower deck of a bus and on a street are similar to that of the steady-state speech-spectrum-shaped noise (see Fig. 2), so that these background noises should have similar effects on the target speech stimuli across frequencies ([Studebaker et al., 1987](#)). In other words, as the noise level increases, the noise would exhibit similar effects on all bands of target speech. This type of masking is supposedly more efficient and thus the rate of change in performance is also faster, making the slope of the performance-intensity function steeper ([Studebaker et al., 1987](#)). In addition, Cantonese is a tonal language and low-frequency information carries more weight for speech understanding than English; it is likely that noise energy at low-frequency regions could

have more detrimental effects on Cantonese speech intelligibility than English (Wong *et al.*, 2007). To the contrary, noises recorded in the lower deck of a bus and on a street produced performance–intensity function slopes that are flatter than that produced when listening to speech-spectrum-shaped noise. Other types of noise that exhibit slightly less or greater low-frequency energy produced performance–intensity function slopes that approximated that of speech-spectrum-shaped noise. Thus, frequency spectra alone do not seem to determine the performance–intensity function slope values in the present study.

Next, the effects of informational masking and amplitude fluctuation were examined. Amplitude fluctuations could lead to non-matching temporal spectra and therefore produce flatter slopes of performance–intensity functions (Takahashi and Bacon, 1992). As discussed in the results section, noises recorded in the upper deck of a bus and subway train produced performance–intensity function slope values (9.53%/dB and 9.17%/dB, respectively) that are similar but slightly flatter than that obtained in steady-state speech-spectrum-shaped noise (10.91%/dB). These types of noise exhibited more amplitude fluctuations, compared to the steady-state speech-spectrum-shaped noise. As the amount of amplitude fluctuation increased, the values of the performance–intensity function slope dropped to 8.60%/dB for listening in noise recorded on the lower deck of a bus and 6.42%/dB in street noise.

Although noise recorded in a café and a Chinese restaurant exhibited greater amounts of amplitude fluctuation than other types of noise used in the present study, the performance–intensity function values thus generated did not reduce, as would be expected. Instead, they (10.62%/dB and 10.82%/dB, respectively) approximated the values obtained in steady-state speech-spectrum-shaped noise. Because in these types of noise, clear speech was present, informational masking seemed to have a greater impact on the performance–intensity function slope than the amount of amplitude fluctuations. To verify whether informational masking could indeed have effects as described here, future research could examine whether speech superimposed on noises that produced flatter slopes would result in performance–intensity functions that approximate that of the steady-state speech-spectrum-shaped noise.

The above-presented findings also suggest that variations in performance–intensity function slope values could have an impact on how performance obtained in various types of noise could be compared. For example, Prosser *et al.* (1991) found that performance in cocktail party noise is poorer than in traffic noise. Cocktail party noise is probably quite similar to noise in a Chinese restaurant because both types of noise are characterized by crowds of people around the listener and there is noise from clattering of dishes. Table II shows that performance in noises recorded in a Chinese restaurant and on the street was comparable when measured as a function of SNRs for 50% intelligibility (−6.72 dB SNR and −6.06 dB SNR, respectively). However, due to variations in performance–intensity function slopes, different conclusions may be made if performance is measured at other SNRs. That is, if measured at a lower SNR,

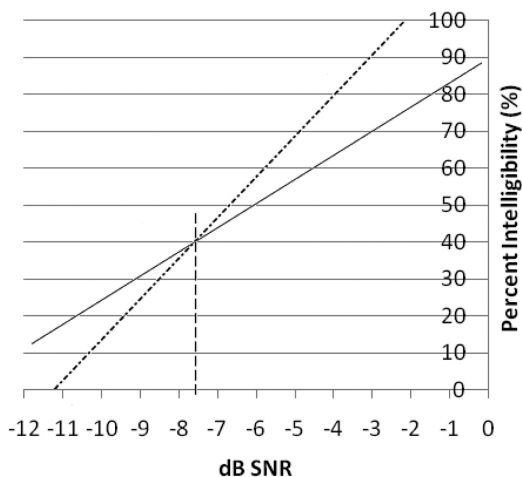


FIG. 6. P - I functions obtained from testing in Chinese restaurant (dashed line) and in street (solid line) noise. Although the SNRs for 50% intelligibility measured in these two conditions are comparable (−6.72 and −6.06 dB SNR, respectively), performance at a low SNR, which yields less than 50% intelligibility (to the left of the vertical dotted line), is expected to be better in street noise. However, performance at higher SNRs (to the right of the vertical dotted line) will result in better performance in Chinese restaurant noise.

performance in restaurant noise is expected to be poorer than in traffic noise; whereas performance in restaurant noise is expected to be better when measured at a higher SNR (see Fig. 6). Therefore, when the performance–intensity function slopes are different, results across test conditions must be compared with caution. The conclusion on whether one noise is more efficient than another as a masker would depend on the performance–intensity function slopes associated with the noise conditions.

The SNRs to achieve 50% intelligibility were compared across noises that yielded similar performance–intensity function slopes. Mean SNRs to yield 50% intelligibility in the Chinese restaurant and subway train conditions were on average 3 dB higher than those in the upper deck of a bus and café condition, which were on average 4 dB higher than that in steady-state speech-spectrum-shaped noise. Although a number of studies showed that higher-than-normal speech and noise levels would have a deleterious effect on speech intelligibility even when the SNR was kept constant (see, e.g., Dubno *et al.*, 2005 and Studebaker *et al.*, 1999), some found no significant difference in speech intelligibility even when both speech and noise levels were raised simultaneously (Hagerman, 1982; Wagener and Brand, 2005). In the present study, speech intelligibility was most disturbed in the upper deck of a bus and café listening conditions, and the masking effects were comparatively less in noise conditions with the highest presentation levels of noise. Therefore, possible deleterious effect of presentation level seemed to be minimal here.

V. SUMMARY AND CONCLUSION

Performance–intensity slope values are similar in most of the real-life noise conditions (Chinese restaurant, café, subway train, and upper deck of a bus) as in the steady-state speech-spectrum-shaped noise. Performance in these types of real-life noise could be predicted based on the performance–intensity

function obtained in steady-state speech-spectrum-shaped noise when the offset in SNR to achieve 50% intelligibility in these noise conditions has been accounted for. These offsets range from 2 to 4 dB. Performance in the lower deck of a bus and street could not be predicted as well as in the other types of real-life noise. Further research can focus on validation of this model in individuals with various degrees of hearing impairment.

Previous studies have found that speech intelligibility in various types of noise are affected by three factors, namely the existence of linguistic information, the amount of noise at low frequency region and the presence of amplitude fluctuations in the temporal spectra. In the present study, the real-life noises and steady-state speech-spectrum-shaped noise have comparable spectra, although slight differences at low frequency are noted. However, the slight differences did not seem to affect the performance–intensity function slopes. Informational masking and amplitude fluctuations were found to impact the slopes of the performance–intensity functions that describe the relationships between SNR and speech intelligibility.

ACKNOWLEDGMENTS

We would like to express our gratitude to Raymond Wu for his technical assistance, Nan Yan for preparing the time-frequency analysis, and all participants in this study. We also appreciate Fiona Yip and Jenny Chan for their help in editing the manuscript. This research was reported in a M.Sc. dissertation submitted by E.H.N.N. to the University of Hong Kong. This research was supported by the General Research Fund No. HKU778707M, the University Research Grants Council, Hong Kong.

ANSI. (1999). *S3.1, Maximum Permissible Ambient Noise for Audiometric Test Rooms* (American National Standards Institute, New York).

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). "The effect of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," *J. Speech Lang. Hear. Res.* **41**, 549–563.

Danhauer, J. L., and Leppler, J. G. (1979). "Effects of four noise competitors on the California Consonant Test," *J. Speech Hear. Disord.* **44**, 354–362.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2005). "Word recognition in noise at higher-than-normal levels: decreases in scores and increases in masking," *J. Acoust. Soc. Am.* **118**, 914–922.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (1995). "Speech recognition in amplitude-modulated noise of listeners with impaired hearing," *J. Speech Hear. Res.* **38**, 222–233.

Gustafsson, H. Å., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**, 518–529.

Hagerman, B. (1982). "Sentences for testing speech intelligibility in noise," *Scand. Audiol.* **11**, 79–87.

Kaplan, H., and Pickett, J. M. (1982). "Differences in speech discrimination in the elderly as a function of type of competing noise: Speech-babble or cafeteria," *Audiology* **21**, 325–333.

Laroche, C., Soli, S., Giguere, C., Lagace, J., Vaillancourt, V., and Fortin, M. (2003). "An approach to the development of hearing standards for hearing-critical jobs," *Noise Health* **6**, 17–37.

Leung, M. T., Law, S. P., and Fung, R. (2004). "Type and token frequencies of phonological units in Hong Kong Cantonese," *Behav. Res. Methods Instrum. Comput.* **36**, 500–505.

McArdle, R. A., Wilson, R. H., and Burks, C. A. (2005). "Speech recognition in multi-talker babble using digits, words, and sentences," *J. Am. Acad. Audiol.* **16**, 726–739.

Owens, E., and Schubert, E. D. (1977). "Development of the California Consonant Test," *J. Speech Hear. Res.* **20**, 463–474.

Plomp, R. (1978). "Auditory handicap of hearing-impaired and the limited benefit of hearing aids," *J. Acoust. Soc. Am.* **63**, 533–549.

Prosser, S., Turrini, M., and Arslan, E. (1991). "Effects of different noises on speech discrimination by the elderly," *Acta Oto-Laryngol. Suppl.* **476**, 136–142.

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2008). "Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing," *Ear. Hear.* **29**, 169–175.

Soli, S. D., and Wong, L. L. N. (2008). "Assessment of speech intelligibility in noise with the hearing in noise test," *Int. J. Audiol.* **47**, 256–361.

Stuart, A., Philips, D. P., and Green, W. B. (1995). "Word recognition performance in continuous and interrupted broad-band noise by normal-hearing and simulated hearing-impaired listeners," *Am. J. Otol.* **16**, 658–663.

Studebaker, G. A., Gilmore, C., and Sherbecoe, R. L. (1993). "Performance–intensity functions at absolute and masked thresholds," *J. Acoust. Soc. Am.* **93**, 3418–3421.

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (1987). "A frequency importance function for continuous discourse," *J. Acoust. Soc. Am.* **81**, 1130–1138.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* **105**, 2431–2444.

Studebaker, G. A., Taylor, R., and Sherbecoe, R. L. (1994). "The effect of noise spectrum on speech recognition performance–intensity functions," *J. Speech Hear. Res.* **37**, 439–448.

Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Hear. Res.* **35**, 1410–1421.

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., Soli, S. D., and Giguère, C. (2005). "Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations," *Int. J. Audiol.* **44**, 358–369.

Wagener, K. C., and Brand, T. (2005). "Sentence intelligibility in noise for listeners with normal-hearing and hearing-impairment: Influence of measurement procedure and masking parameters," *Int. J. Audiol.* **44**, 145–156.

Wan, I. K. K., and Wong, L. L. N. (2002). "Tympanometric norms for Chinese adults," *Ear. Hear.* **5**, 416–421.

Wilson, R. H., Zizz, C. A., Shanks, J. E., and Causey, G. D. (1990). "Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker," *J. Speech Hear. Disord.* **55**, 771–778.

Wong, L., Ho, A., Chua, E., and Soli, S. (2007). "Cantonese speech intelligibility index," *J. Acoust. Soc. Am.* **121**, 2350–2361.

Wong, L. L. N., and Soli, S. D. (2005). "Development of the Cantonese hearing in noise test (CHINT)," *Ear Hear.* **26**, 276–289.