Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions

Adam Weisser and Jörg M. Buchholz[a)]
Department of Linguistics—Audiology Section, Macquarie University, Australian Hearing Hub—Level 3, 16 University Avenue, New South Wales 2109, Australia

(Received 24 October 2018; revised 19 December 2018; accepted 21 December 2018; published online 23 January 2019)

Estimating the basic acoustic parameters of conversational speech in noisy real-world conditions has been an elusive task in hearing research. Nevertheless, these data are essential ingredients for speech intelligibility tests and fitting rules for hearing aids. Previous surveys did not provide clear methodology for their acoustic measurements and setups, were opaque about their samples, or did not control for distance between the talker and listener, even though people are known to adapt their distance in noisy conversations. In the present study, conversations were elicited between pairs of people by asking them to play a collaborative game that required them to communicate. While performing this task, the subjects listened to binaural recordings of different everyday scenes, which were presented to them at their original sound pressure level (SPL) via highly open headphones. Their voices were recorded separately using calibrated headset microphones. The subjects were seated inside an anechoic chamber at 1 and 0.5 m distances. Precise estimates of realistic speech levels and signal-to-noise ratios (SNRs) were obtained for the different acoustic scenes, at broadband and third octave levels. It is shown that with acoustic background noise at above approximately 69 dB SPL at 1 m distance, or 75 dB SPL at 0.5 m, the average SNR can become negative. It is shown through interpolation of the two conditions that if the conversation partners would have been allowed to optimize their positions by moving closer to each other, then positive SNRs should be only observed above 75 dB SPL. The implications of the results on speech tests and hearing aid fitting rules are discussed. *© 2019 Acoustical Society of America.*
https://doi.org/10.1121/1.5087567

[GCS] Pages: 349–360

I. INTRODUCTION

Acoustic communication is an adaptive process in which talkers modify their vocal effort to be able to hear themselves and one another, despite uncontrolled environmental conditions such as variable noise level. For the listener, being able to follow the conversation requires a favorable signal-to-noise ratio (SNR) between the talker's speech and the background noise. Talkers raise their voice level in noise in what is referred to as the Lombard effect (Lombard *et al.*, 1911), which helps maintain high SNR despite the noise (Lane and Tranel, 1971), but also includes a host of additional acoustic, linguistic, and conversational speech modifications that go beyond level (Beechey *et al.*, 2018; Junqua, 1993). The effective SNR depends also on the distance between the interlocutors, as the direct sound pressure level (SPL) from the acoustic source drops by about 6 dB for every doubling of distance between the mouth of the talker and the listener's ear. Therefore, during conversation people appear to naturally come closer in noisier environments to improve communication (Pearsons *et al.*, 1977), as talkers want to be heard and listeners want to be able to hear and respond. Alternatively, listeners can turn their heads to give an advantage for one ear over another by reducing the head shadow effect for that ear, and thereby increasing the target level (Brimijoin *et al.*, 2012; Grange and Culling, 2016). It is suggested that in some situations, head turns are traded off with the ability to read the talker's lip movements (Brimijoin *et al.*, 2012), which improves speech reception, in particular at poor SNRs, by providing a parallel channel of visual information (Sumby and Pollack, 1954).

The adaptation of the talker's vocal effort, the listener's head angle, the amount of lip-reading, and the distance between the conversational partners is constrained by physical, psychological, and social factors. In an experiment that had interlocutors sitting at a distance of 1.5 m, while listening to speech-shaped noise on headphones that varied in level every 15–25 s, the subjects tended to decrease their distance by only up to 10 cm by leaning forward as a function of noise level, which led to an almost negligible increase in SNR (<1 dB; Brimijoin *et al.*, 2017). At the same time, talkers are known to be able to judge the distance to a listener and adapt their voice level accordingly (Warren, 1968; Zahorik and Kelly, 2007) while factoring in the room acoustics (reverberation, volume, amplification), which entails smaller power loss over distance compared to free field (Pelegrín-García *et al.*, 2011). However, even if vocal effort can be raised to the point of shouting, it is likely to be both fatiguing and unpleasant for the interlocutors over time. For example, while it is theoretically possible to have a conversation at a noisy party from a distance of 1 m, it requires

[a)]Electronic mail: jorg.buchholz@mq.edu.au

significant vocal effort that is loud enough to overcome the background babble and music, which can be very uncomfortable for both talker and listener. Similarly, cultural norms dictate certain distances that are socially acceptable, which depend, among others, on gender, language, age, and familiarity (Sommer, 2002), as well as the listener's hearing status. In situations where people are confined to fixed seating and cannot get closer, communication can become frustratingly difficult, and higher reliance on visual speech becomes valuable as the noise level increases (Brimijoin *et al.*, 2017). In the noisy party example, the smallest distance is achieved by talking directly into somebody's ear, without eye contact. Resorting to this extreme strategy, however, is undesirable in many other formal daily situations. In contrast, at low background noise levels, speech tends to remain approximately constant at a comfortable mean level of 55 dBA (Pearsons *et al.*, 1977). Therefore, adapting distance and vocal effort constitute, in effect, an optimization problem that has to be solved in practice by the interlocutors.

Contrary to the variable and adaptive nature of realistic conversations, laboratory-based speech intelligibility tests are designed to provide stable listening conditions. Inasmuch as adaptation is incorporated in speech tests, it only functions procedurally to control the linear amplification of speech and/or noise levels according to the subject's performance, which constitutes a predefined point on the psychometric function (Bode and Carhart, 1973; Levitt and Rabiner, 1967). The relevance of the speech test scores is inferred by drawing an analogy between the test conditions to common speech-in-noise situations encountered by listeners (Carhart and Tillman, 1970), which are characterized by certain broadband background noise levels and SNRs (Plomp, 1986). To maintain control of the signal, the speech material in standardized tests is typically recorded in sound-treated rooms, lacks any conversational context, is well articulated by the reader, and is generated at normal vocal effort independently of the rather artificial noise (e.g., speech babble or steady-state noise) that is used as masker (e.g., Cox *et al.*, 1987a,b; Hagerman, 1982, 1984; Kalikow *et al.*, 1977; Nielsen and Dau, 2009; Nilsson *et al.*, 1994). These tests were originally administered using headphones for unaided listeners, but variations exist using fixed-distance loudspeakers for aided and unaided listeners so that speech and noise levels at the ear-level are both known (e.g., Ching *et al.*, 2004; Hanks and Johnson, 1998; Lunner *et al.*, 2016; Ricketts and Hornsby, 2005). Visual channels are mostly unavailable in the standardized tests, but exceptions do exist (e.g., MacLeod and Summerfield, 1990). Therefore, the speech test conditions are designed in a manner that effectively removes all of the adaptive aspects of real conversations that may affect their acoustic SNR.

It is unclear how well the combination of speech and noise levels applied in common speech tests represents the ones experienced by a person participating in a real conversation in which the conversation partners have already adapted their vocal effort, as well as their physical behavior, to the given environment. Since no single typical conversation scene exists in real life, it is possible that the fixed conditions of speech tests are still representative of typical speech and noise levels encountered in an ecologically relevant range of daily conversations, despite some concern of the contrary (Naylor, 2016).

Pearsons *et al.* (1977) recorded the speech levels, as well as the ambient noise levels, on test subjects wearing eye-level head-worn microphones in homes, hospitals, department stores, schools, inside trains, and inside airplanes, and talking to each other at 1 and 2 m distances (Pearsons *et al.*, 1977, Table II). The data suggest that in the noisiest environments (trains and airplanes) the mean conversational SNR is negative. However, the same report suggests that interlocutors tend to stand much closer in these two environments (Pearsons *et al.*, 1977, Fig. 22), so that the SNR may have not been negative after all, as the speech level increases while the ambient noise remains constant at closer distances. It is unreported, though, how these distances were estimated, especially since the interiors of airplanes and trains may constrain people to sit much more closely than is typical in the quieter places surveyed.

The acoustics of conversations in everyday environments was the focus of two more recent surveys, but without controlling for distance between talkers (Smeds *et al.*, 2015; Wu *et al.*, 2018). The studies were based on daily environments experienced by hearing aid users who carried chest-level (Wu *et al.*, 2018) and ear-level (Smeds *et al.*, 2015) recording devices, which provided data on the person's ear with the better SNR or, in the latter case, on both ears. Even though listeners in realistic and simulated restaurant scenes had speech reception threshold (SRT) performance of −5 dB SNR and below (Culling, 2016), the majority of field-surveyed SNRs have been positive (Smeds *et al.*, 2015, better-ear: 95.8%, worse-ear: 86.3% of the samples; Wu *et al.*, 2018, chest-level: 92.5%) and higher than the data reported by Pearsons *et al.* (1977; 84.5%). The authors suggested that the discrepancy between the observations may have occurred because noisy situations were undersampled due to differences in recording methods used to acquire the signals. However, all measurements were done in conditions with unknown distance to the recorded interlocutor, which may have been optimized by the interlocutors to achieve favorable SNRs. The SNR may have also increased by the fact that the microphones in the more recent studies were attached to older hearing-impaired subjects, which may have led to an adaptive speech level increase by interlocutors to partly compensate for the hearing loss and associated reduction in speech perception. Moreover, as acknowledged by the authors, the SNR values were estimated by manually separating speech-plus-noise and noise-only portions from the recordings, a procedure that is particularly susceptible to imprecision at poor SNR conditions, where speech energy is low compared to noise. Thus, none of the available surveys provides fully transparent data that enables assessing whether the speech and noise level ranges used in common speech tests are representative for everyday communication, in particular when taking into account the adaptive nature of real conversations. This may also explain the very large variances seen in the previous studies as measured SNRs vary by more than 20 dB at several given noise levels (Smeds *et al.*, 2015).

In the present study, a well-controlled laboratory experiment was carried out to measure conversational speech levels in noisy environments at realistic distances between two communication partners. The experiment allowed to separately investigate the possible effects of distance-dependent adaptation and the talkers' vocal effort. The measurements also generated new SNR distributions for a set of reproducible virtual realistic environments that are publicly available (Weisser, 2018), using a robust methods for estimating both the realistic speech and noise levels. The results have direct implications for the design of speech tests for assessing individual hearing ability and hearing device benefit, as well as for the design of hearing aid fitting rules.

II. METHODS

A. Subjects

Ten male and 10 female subjects (ages 18–52 yr, median 27.5 yr for male and 23 yr for female subjects) participated in the test and were compensated for their time. Subjects had all normal hearing with pure tone thresholds below <20 dB hearing level (HL) at 500, 1000, 2000, and 4000 kHz. One female subject had to be removed from the data set because of a technical fault in the recording. Treatment of subjects was approved by the Australian Hearing as well as Macquarie University ethics committee and conformed in all respects to the Australian Government's National Statement on Ethical Conduct in Human Research.

B. Procedure

Two people were seated 1 m apart (mouth-to-ear distance) in front of each other and had to complete a puzzle together through conversation (Beechey *et al.*, 2018). In this highly motivating task, subjects received complementary halves of the puzzle printed on paper—containing missing information that had to be gathered from the other player. The only way to solve the puzzle was by verbally describing the parts of the puzzle that were invisible to the other player and making decisions jointly. Subjects wore highly open headphones and boom microphones that allowed them to converse freely, while 13 different noisy environments were played to them for 2 min per scene, at a random order. In between the scenes there was a short break during which the subjects were requested to stop talking. Thereafter, the test was repeated for a distance of 0.5 m. Because this distance would have been too close to comfortably accommodate the subjects while sitting facing each other, the arms of the two chairs were put together, but one chair was turned 180°. This way the subjects could be close, but had to rotate their heads slightly to look at each other. This seating allowed for an easy acoustic communication and removed any awkwardness that would have been experienced by the two subjects, who are not intimate, if they had sat face-to-face at such a close distance. Note, however, that the binaural display of the scenes did not track the head rotation, so whatever events were taking place in the scenes turned with the head.

C. Noise stimuli

In order to elicit Lombard speech at a realistic vocal effort that corresponds to the noise level of the environment, subjects listened to binaural versions of 13 scenes from the Ambisonic Recordings of Typical Environments (ARTE) database (Weisser, 2018). Eleven scenes were recordings of real places in urban environments, and played back at their original SPL. Two additional scenes were of diffuse speech-weighted noise at 60 and 70 dB SPL. The 13 scenes were ordered according to their broadband SPL shown in Fig. 1 and comprised of (1) a library, (2) an open plan office, (3) soft speech-weighted diffuse noise, (4) people gathering quietly inside a small church, (5) a living room with a television in the front playing advertisements and kitchen noise in the back, (6) people inside the same church as before but making more noise, (7) loud speech-shaped diffuse noise, (8) a busy indoor café, (9) a dinner party with people around a table and music in the background, (10) a busy street recorded on

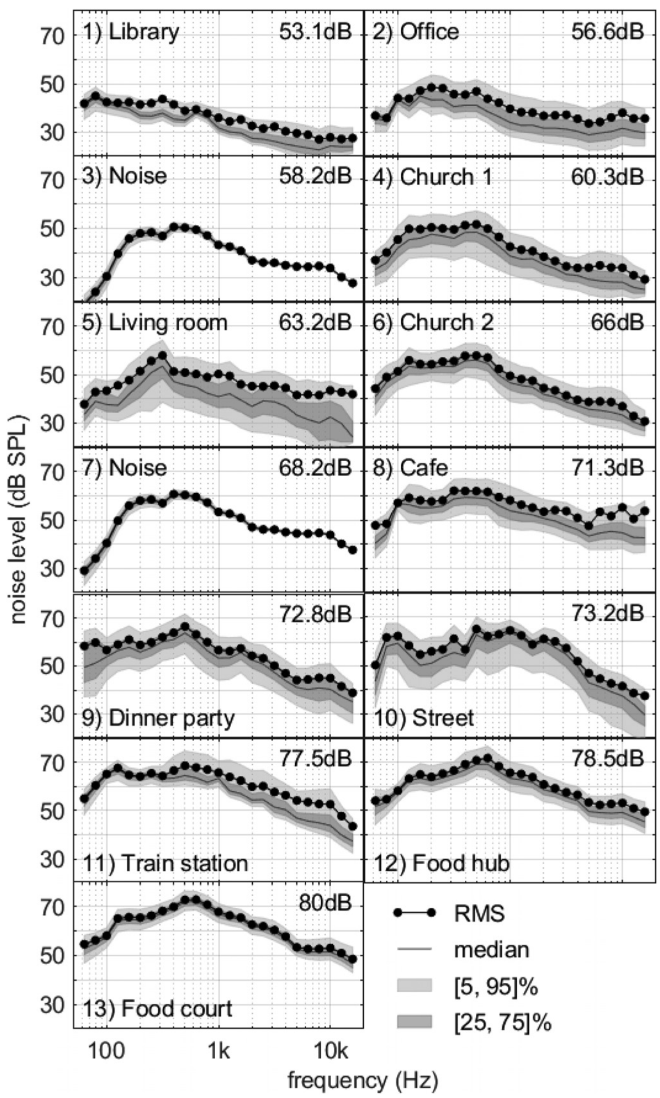

FIG. 1. SPLs in third-octave bands for all 13 acoustic environments, ordered according to their broadband level. Solid lines with circles indicate the long-term RMS level. The different shaded areas indicate the quantiles applied to a short-term level analysis within the third-octave bands using a 125 long Hann window.

a nearby balcony, (11) a large train station including loud announcements, (12) a busy university food hub, and (13) a large food court of a shopping mall during lunch. Further details of these scenes can be found in Weisser (2018).

Figure 1 shows the SPLs in third-octave bands for all 13 acoustic scenes, averaged across ears, and ordered according to their broadband level shown in the figure panels. The long-term RMS level is indicated by the solid lines with circles, and the different percentiles of a short-term level analysis are indicated by the gray shaded areas. The short-term analysis was performed on the power envelope within each frequency channel after temporal smoothing with a 125 ms long Hann window. Besides the expected upward shift with increasing broadband levels, all spectra show a low-pass characteristic with details varying across scenes, although not in any obvious systematic fashion. The different environments contain different amounts of level fluctuations, which is indicated by the spread of the short-term levels (i.e., the width of the shaded areas), with the diffuse noises (scenes 3 and 7) exhibiting the lowest and the living room exhibiting the strongest fluctuations (scene 5).

The binaural recordings were played back individually to subjects seated in the anechoic chamber of the Hearing Hub, Macquarie University, using Sennheiser HD 800 open, circum-aural, diffuse-field equalized headphones (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany). The binaural recordings were realized by simulating the playback of the multi-channel sound files via a 41-channel three-dimensional (3D) loudspeaker array to the ears of a Brüel and Kjær type 4128C Head and Torso Simulator (HATS; Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark), a method that is further described in Weisser (2018). The playback via headphones had the main advantage over loudspeaker-based sound reproduction methods in that (i) the two test subjects heard exactly the same noise signals and (ii) it minimized the cross talk of the noise signals to the subjects' headset microphones, which maximized the SNR of the recorded speech signals and strongly simplified the subsequent analysis.

The binaural sound files were processed by two subject-specific minimum phase filters. The first filter realized an individual headphone equalization and inverted the third-octave smoothed magnitude response of the transfer function measured between the headphones and microphone probes placed in the listener's ear canals, just in front of the ear drums. These microphone probes were custom made, but resembled the probe tube microphones used for standard real-ear hearing aid measurements (similar to the Etymotic Research ER7, Elk Grove Village, IL). The second filter simulated the passive attenuation that is introduced by wearing the headphones when listening to external sounds, such as the interlocutor's voice during the conversation task. The filters were derived by playing diffuse noise to each listener via a 3D loudspeaker array, which was then recorded at the listener's ear drums with the same probe microphones as used before. These recordings were performed once while the subjects were wearing the headphones and another time while the headphones were taken off. The filters were then designed such that they approximated the spectral difference between these two recordings, analyzed in third-octave bands. Applying these filters to the binaural playback of the noise signals during the conversation task ensured that the effective SNR was approximately the same as it would have been experienced in the real world. However, the overall spectra of the speech and noise signals was slightly low-pass filtered by the passive attenuation of the headphones. This is illustrated in Fig. 2, which shows the passive attenuation of the headphones averaged across all subjects (solid line), including ± 1 standard deviations (shaded area), as well as an example filter response (dashed line) designed here to match the average attenuation of the headphones. It can be noticed that the attenuation starts to become significant for frequencies above about 1.2 kHz, which is well described by the filter response. However, the attenuation then increases again above 4 kHz, which is due to the limited SNR of the probe microphones. To avoid this limitation, the filters were designed such that the attenuation was kept constant above 4 kHz. However, it is expected that the actual attenuation increases further at high frequencies, which is not considered in the filter design.

D. Speech recording and processing

During the conversation task, each subject wore a DPA d:fine™ FIO66 omnidirectional headset (boom) microphone (DPA Microphones A/S, Alleroed, Denmark), which was positioned a few centimeters away from the talker's mouth. The microphones were individually calibrated on the subjects at the very beginning of the experiment and not moved until the experiment was completed. During the calibration process the subjects were reading aloud from a written passage for 30 s. Their voice was thereby recorded by their headset microphone as well as by a calibrated 1/4 in. Type 46BL G.R.A.S. low-noise omnidirectional microphone (GRAS Sound & Vibration A/S, Holte, Denmark) placed in front of them, first at 0.5 m then at 1 m. Afterward, two separate calibration filters were derived, which mapped the speech signal recorded by the headset microphone to the location of the measurement microphone. The filters were created by approximating the spectrum of the speech recorded with the calibrated measurement microphone divided by the spectrum of the speech recorded with the

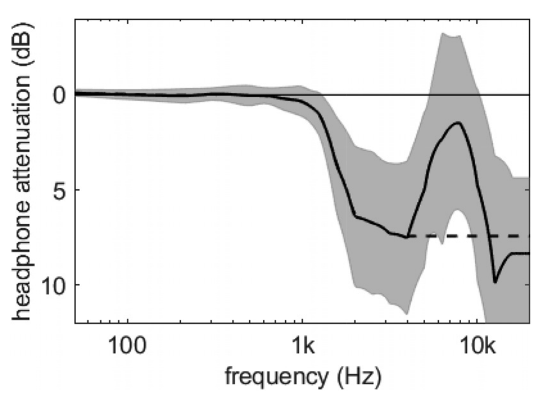

FIG. 2. Passive diffuse-field attenuation of Sennheiser HD800 headphones (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany) measured *in situ* across all test subjects. The solid line refers to the average attenuation, the shaded area refers to ± 1 standard deviations, and the dashed line refers to the magnitude response of a filter designed to match the average response.

headset microphone, both analyzed in third-octave bands. The resulting mapping function mainly realized a broadband gain that differed across subjects due to the small differences in the individual headset microphone placement relative to the mouth. Speech levels were then calculated from the headset microphone after convolution with the calibration filter for the given distance. Since conversations were recorded, the individual speech recordings contained extensive speech pauses. To avoid the uncontrolled duration of the speech pauses from affecting the derived speech levels, the method described in IEC (2011, Annex J) was applied, which removes speech pauses from the level calculations. However, since in standard speech tests sentences are commonly applied with their root-mean-square (RMS) level calculated across the entire sentence (including pauses), the derived speech levels were here transformed into sentence-equivalent levels. This was done by first calculating the average RMS level over 1280 BKB (Bamford-Kowal-Bench)-like sentences (Bench *et al.*, 1979) from a speech corpus available at the Australian Hearing Hub and widely used for both clinical and research purposes, and then recalculating their RMS level following the above method with pauses removed. Removing the pauses increased the average sentence level by 1.87 dB. This value was then subtracted from the present speech levels derived for the conversation recordings.

III. RESULTS

A. Measured data

The boom microphone recordings of all the conversations were analyzed with respect to absolute noise and speech level in the different scenes. Figure 3 shows the measured broadband speech levels as a function of the corresponding noise level. Individual data are shown by the gray lines and data averaged across subjects are shown by the black lines. The speech levels for the near-talker distance (0.5 m) were normalized to a distance of 1 m by subtracting 6 dB from the measured speech levels. Linear functions were fitted to the average data as indicated by the dashed lines, with an average RMS error of 0.81 dB. Their estimated slopes α and intercepts β are given in the individual figure panels. The estimated slopes are very consistent across conditions with speech levels increasing by about 0.43 dB for each 1 dB increase in noise level. Comparing average data, male speech levels are about 1.7 dB higher than female speech levels, which is slightly more pronounced in the far-talker condition with 2.1 dB versus 1.2 dB. When referenced to a distance of 1 m, average speech levels decrease for the female talkers by 0.6 dB from the far to the near-talker distance (i.e., when the distance is halved) and decrease by 1.5 dB for the male talkers. Hence, people slightly reduce their vocal effort (i.e., the source) levels when moving closer to the conversation partner under constant noise conditions (Pelegrín-García *et al.*, 2011). This was confirmed using a mixed effect linear model of the distance-normalized levels in which distance and gender (no interaction) are modeled as fixed effects, and scene and subject are modeled as random effects with by scene and by subject slopes and intercepts. The modeling was done using the *lme4* package in *R* (Bates *et al.*, 2015). At 95% confidence interval level, only distance had a significant effect on level (Tukey's test for pairwise comparison z-ratio $= -3.18$, $p = 0.0015$, with 1.07 dB between the far and near condition means) but not gender (z-ratio $= 1.761$, $p = 0.0782$), and no interaction between gender and distance. As Fig. 3 illustrates, the individual spread of levels of the female talkers is larger than the male talkers throughout all scenes. From Fig. 3 it can be also deduced that the individual speech levels vary substantially across subjects, which is pronounced in the quietest environments. Whereas the mean inter-subject standard deviation across the four conditions is 3.8, 3.3, and 3.0 dB for the three quietest scenes (e.g., the library, office, and soft diffuse noise scenes), it is about 2.5 dB on average for the other (louder) environments. Additionally, some subjects seem to adjust their vocal effort to the given acoustic scene more than others, with individual slopes varying between 0.3 and 0.65. Moreover, some subjects keep their vocal effort level constant across the softer scenes and adjust only to the louder scenes, while others adjust their vocal effort level across all scenes.

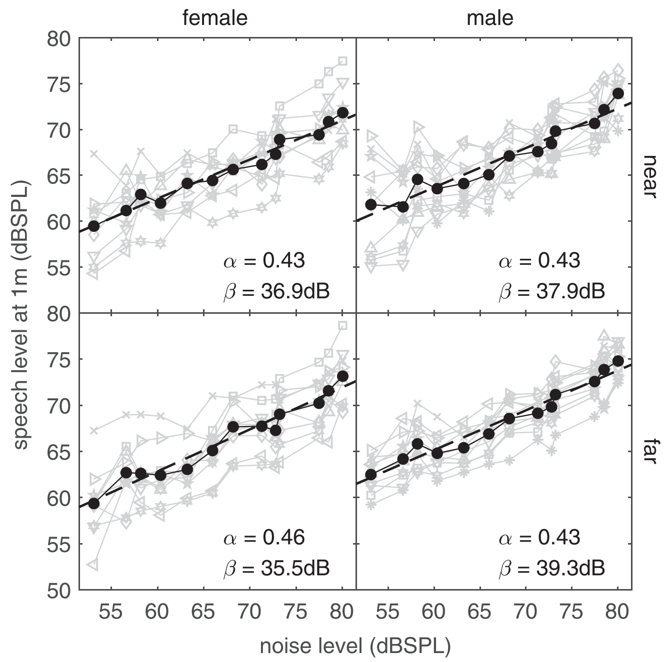

FIG. 3. Broadband SPLs of the recorded speech as a function of the corresponding noise level normalized to a talker distance of 1 m. The left column refers to the female talkers and the right column refers to the male talkers. The upper row refers to the near-talker distance (i.e., 0.5 m) and the lower row refers to the far-talker distance (i.e., 1 m). The gray lines refer to individual levels and the black lines refer to the average levels across subjects. Dashed lines indicate linear fits to the average data, with the values for the fitted slopes, α, and intercepts, β, given in each panel.

Figure 4 shows the average speech levels in third-octave bands in all 13 acoustic scenes, ordered according to their noise levels shown in Fig. 1. For improved readability, only the speech levels for the far-talker condition are shown here, but the levels for the near-talker condition look very similar when normalized to a distance of 1 m. Speech levels averaged over all female talkers are indicated by filled circles and for the male talkers by open circles. The male and female spectra differ mainly at low frequencies, where the lower fundamental frequency of the male talkers shifts the

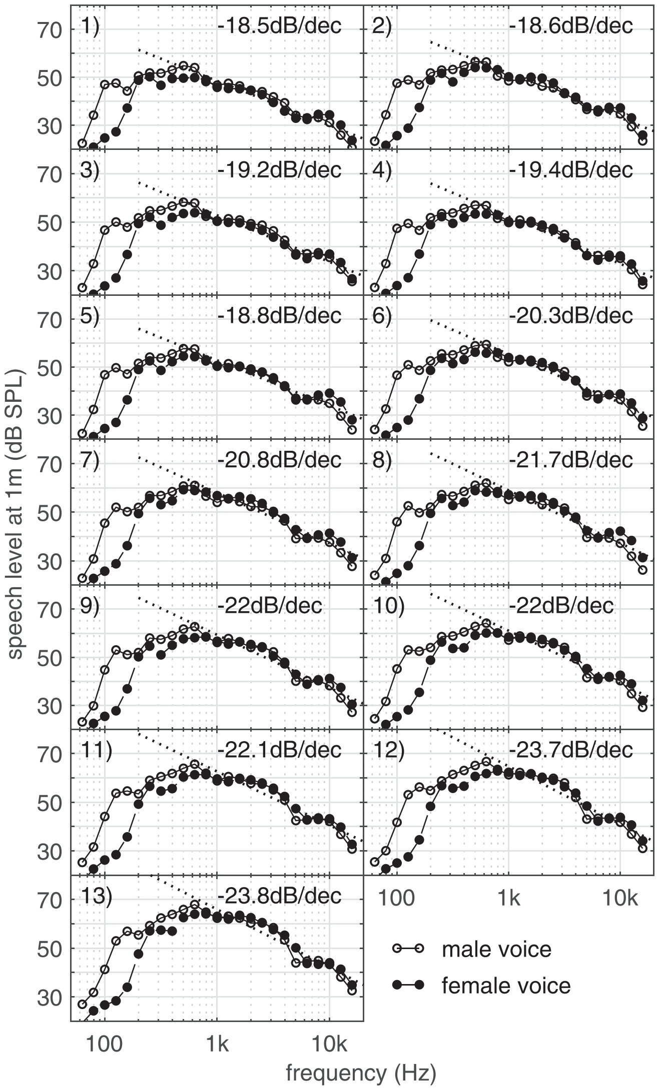

FIG. 4. Long-term speech levels in third-octave bands for all 13 scenes for the far-talker distance (i.e., 1 m) averaged across subjects. The filled circles refer to the female talkers and the open circles refer to the male talkers. Dotted lines indicate linear fits to the data averaged across gender and calculated over a frequency range from 800 to 16 000 Hz. The values for the fitted slopes (dB per decade) are given in each panel.

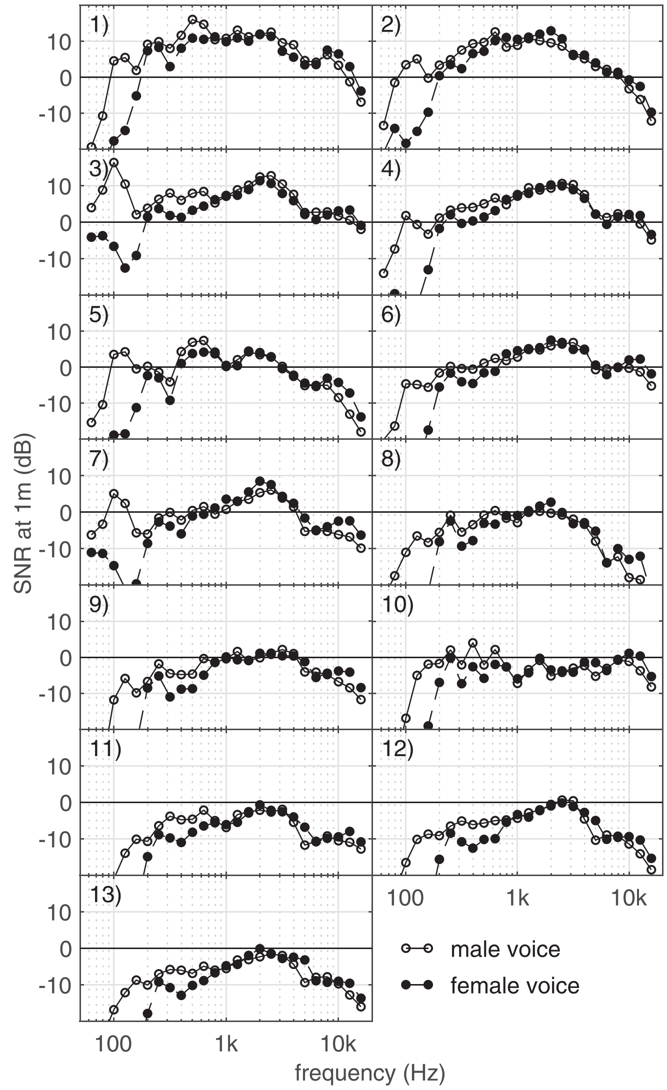

FIG. 5. Long-term SNRs in third-octave bands for all 13 environments for the far-talker distance (i.e., 1 m) averaged across subjects. The filled circles refer to the female talkers and the open circles to the male talkers.

low-frequency cutoff of speech from about 200 Hz to 100 Hz. To highlight the increasing frequency roll-off (or slope) of the speech spectra with increasing noise (and speech) levels, the speech spectra from 800 to 16 000 Hz, averaged across gender, were fitted by linear functions on a double-logarithmic scale. The fitted functions are shown in Fig. 4 by the dashed lines with the slope given in each figure panel. The slopes increase from −18.5 dB/decade in the softest scene (i.e., the library) to −23.8 dB/decade in the loudest scene (i.e., food court 2).

Figure 5 shows the average SNRs in third-octave bands for all 13 acoustic scenes, which were derived directly by subtracting the noise levels shown in Fig. 1 from the speech levels shown in Fig. 4. Accordingly, only the SNRs for the far-talker condition are shown. The SNRs for the near-talker condition are very similar when normalized to a distance of 1 m, but when considering their original (non-normalized) speech levels at 0.5 m distance, the SNRs across all frequencies improve (i.e., the curves in Fig. 5 shift upward) on average by 5.4 dB for the female talkers and 4.5 dB for the male talkers. It can be seen that the overall SNR decreases with increasing noise level, which directly reflects the (broadband) speech level behavior shown in Fig. 3, according to which every 1 dB increase in (broadband) noise level is only partially compensated by a 0.43 dB increase in (broadband) speech level. Considering the spectral shape of the SNR, it can be seen that for most environments the highest (i.e., best) SNR can be observed in a frequency range of about 1–4 kHz.

B. Distance-adjusted SNR modeling

Figure 6 shows the broadband SNR for all 13 acoustic scenes averaged across subjects, with the female data in the left panels and the male data in the right panels. The upper panels show the SNRs measured at the listener location for the near-talker (open circles) and far-talker (solid circles) conditions, i.e., at a distance of 0.5 and 1 m, respectively.

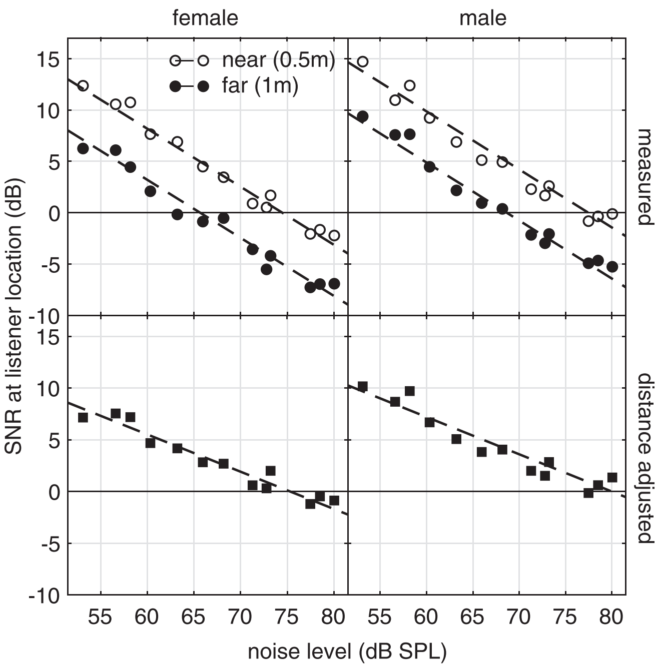

FIG. 6. Broadband SNRs as a function of the level of the corresponding acoustic scene, averaged across subjects. The left column refers to the female talkers and the right column refers to the male talkers. The upper row refers to the measured SNRs separately shown for the far (1 m) and near (0.5 m) talker distances, as indicated by the filled and open circles, respectively. The lower row refers to the distance-adjusted SNR as further described in the main text. Dashed lines indicate linear fits to the measured SNR data.

The dashed lines indicate linear fits to the SNR data, and are given by

$$\mathrm{SNR} = -0.565L + 40.4 + \Delta. \tag{1}$$

With L the noise level in dB SPL as given in Fig. 1, and Δ a gender and distance dependent intercept adjustment with $\Delta = \Delta_G + \Delta_D$. For male subjects $\Delta_G = 0.84$ dB and for female subjects $\Delta_G = -0.84$ dB. For the near-talker condition $\Delta_D = 2.49$ dB and for the far-talker condition $\Delta_D = -2.49$ dB. Equation (1) approximates the measured SNR data well (Fig. 6) with an average RMS error of 0.85 dB.

The distance-adjusted SNR is shown in lower panels of Fig. 6, and takes into account that in the real world, conversation partners move closer to each other when the noise level increases to improve the SNR at the partners' ears while limiting their vocal effort level. These SNRs were derived by first estimating the distance that conversational partners would have chosen in a given acoustic scene, and then using this distance to derive the corresponding speech level from the data shown in the upper panel of Fig. 6. This was done by linearly interpolating the data on a double-logarithmic scale between the measured (and fitted) SNRs of the near- and far-talker conditions. Based on data from Pearsons *et al.* (1977, Fig. 22), the conversation distance D can be predicted from the A-weighted noise level L_A of the scene by

$$\log(D) = \alpha_D L_A + \beta_D, \tag{2}$$

with a base-10 logarithm of the distance D in meters. Applying the two reference points provided by Pearsons *et al.* (1977), i.e., $D_1 = 0.5$ m at $L_{A,1} = 70$ dBA and $D_2 = 1.0$ m at $L_{A,2} = 43$ dBA, the slope and intercept are found to be $\alpha_D = -0.011$ and $\beta_D = 0.479$, respectively.

Since the unweighted noise level L (in dB SPL) was applied throughout this study, the A-weighted noise level L_A is approximated in the following by (see Sec. IV A):

$$L_A = 1.11L - 11.08. \tag{3}$$

Combining Eqs. (2) and (3) results in the final conversational distance approximation, given by

$$\log(D) = -0.0122L + 0.6. \tag{4}$$

The corresponding distance-adjusted SNR_D can be derived by applying a linear interpolation on a double logarithmic scale between the SNR measured at the two distances $D_{\mathrm{near}} = 0.5$ and $D_{\mathrm{far}} = 1$, which takes into account that interlocutors talk (slightly) more softly when they are moving closer to each other, i.e.,

$$\mathrm{SNR}_D = \alpha_{\mathrm{SNR}} \log(D) + \beta_{\mathrm{SNR}}, \tag{5}$$

with

$$\alpha_{\mathrm{SNR}} = \frac{\mathrm{SNR}_{\mathrm{far}} - \mathrm{SNR}_{\mathrm{near}}}{\log(D_{\mathrm{far}}) - \log(D_{\mathrm{near}})} \approx 3.32(\mathrm{SNR}_{\mathrm{far}} - \mathrm{SNR}_{\mathrm{near}}), \tag{6}$$

$$\beta_{\mathrm{SNR}} = \mathrm{SNR}_{\mathrm{near}} - \alpha_{\mathrm{SNR}} \log(D_{\mathrm{near}}) \approx \mathrm{SNR}_{\mathrm{near}} + 0.3\alpha_{\mathrm{SNR}}. \tag{7}$$

Applying Eq. (5) separately to the measured male and female SNR data results in the distance-adjusted SNR_D shown by the squared markers in the lower panels of Fig. 6. In the case that the SNR approximation by Eq. (1) is taken into account as well as the $\log(D)$ approximation provided by Eq. (4), Eq. (5) can be simplified to

$$\mathrm{SNR}_D = -0.361L + 28 + \Delta_G, \tag{8}$$

with $\Delta_G = 0.84$ dB for male talkers, $\Delta_G = -0.84$ dB for female talkers, and $\Delta = 0$ for the distance-adjusted SNR_D averaged over male and female talkers. Equation (8) approximates the distance-adjusted SNR data shown in Fig. 6 well, with an average RMS error of 0.85 dB.

Equation (8) describes the case that the talkers can freely adjust their distance, which is not always true. For instance, when sitting around a table, the talker's movements are constrained and their distance may be rather invariable and independent of the noise level (Brimijoin *et al.*, 2017). In such a case, if the distance D is known, then the distance-adjusted SNR_D can be derived from Eq. (5) with

$$\mathrm{SNR}_D = -16.54\log(D) - 0.56L + 37.91 + \Delta_G, \tag{9}$$

where the parameters $D_{\mathrm{far}} = 1$ m, $D_{\mathrm{near}} = 0.5$ m, $\Delta_D = 2.49$ dB, and $\mathrm{SNR}_{\mathrm{far}} - \mathrm{SNR}_{\mathrm{near}} = -2\Delta_D$ were applied in Eq. (5) to obtain $\alpha_{\mathrm{SNR}} \approx -16.54$. Moreover, by setting $\mathrm{SNR}_{\mathrm{near}} = \mathrm{SNR}$ and $\Delta = \Delta_D$, Eq. (1) was applied to Eq. (6) to obtain $\beta_{\mathrm{SNR}} \approx -0.56L + 37.93 + \Delta_G$.

IV. DISCUSSION

The present study measured conversational Lombard speech levels and SNRs using binaural reproduction of realistic acoustic scenes at two different realistic interlocutor distances. Together with the high level of control provided by the binaural playback, as well as the speech calibration, recording, and analysis method, the study provided speech levels and SNRs with a higher degree of precision than in previous field studies, while still maintaining considerable ecological validity in the conversation. The results are discussed below.

A. Reference levels and SNRs

Throughout this study, the considered speech and noise broadband levels, as well as SNRs, were spectrally unweighted, calculated from the pressure signals picked up by an omnidirectional microphone, and averaged across the entire duration of the scene-specific recorded conversation. The unweighted acoustic measure, dB SPL, provides a straightforward, standard, and unbiased representation of the considered signals, is widely used across literature, and allows direct comparison with other speech and noise signals. However, *A*-weighted levels are probably as often used across literature, focus more on the frequency region that is relevant for speech perception, and are less impacted by environmental low-frequency noise in the playback or test environment such as emitted by ventilation systems. To allow comparison of the present data with the aforementioned literature, Fig. 7 summarizes the effect of applying *A*-weighting on the measured speech and noise broadband levels, as well as the corresponding SNR. It can be seen that the *A*-weighting reduces the noise level by about 2–6 dB (crosses), an effect that, at least for the acoustic environments considered in this study, decreases with increasing noise levels. This behavior can be reasonably well approximated by a linear function (dashed line) with a slope $\alpha = 0.11$, an intercept $\beta = -11.08$ dB, and a RMS error $E = 0.64$ dB. The *A*-weighting affects the speech levels in a similar way, although to a lesser extent, in particular for the female speech (circles). For the male speech (triangles), $\alpha = 0.066$, $\beta = -5.78$ dB, and $E = 0.12$ dB, and for the female speech, $\alpha = 0.068$, $\beta = -7$ dB, and $E = 0.14$ dB. Due to the slightly different effect that the *A*-weighting has on the speech levels than on the noise levels, the effect on the resulting SNR is significant as shown in the left panel of Fig. 7. On average, the *A*-weighting increases the SNR for the female speech by 2.3 dB and for the male speech by 1.2 dB.

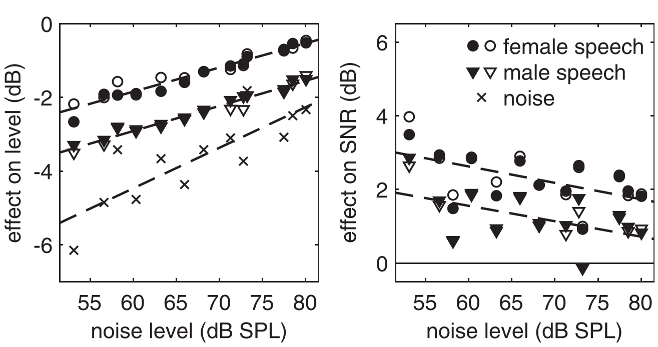

FIG. 7. Effect of applying *A*-weighting on the speech and noise levels (left), as well as on the SNR (right) as a function of (unweighted) noise level. Circles refer to female speech, triangles refer to male speech, and crosses refer to the noise. Filled symbols refer to the near and filled symbols refer to the far-talker conditions. Dashed lines present linear fits to the data.

These linear fits are not entirely consistent with data from Pearsons *et al.* (1977, Fig. 19), which approximated the drop in speech level after *A*-weighting using a quadratic equation. In the quietest environment, at 55 dB SPL the difference is about the same, ranging between −2 and −6 dB. However, at 75–85 dB SPL no correction had to be made (0 dB) to convert the two, whereas in the present data this is true for the male, but not for the female data. That said, the spread in the data of Pearsons *et al.* is much larger than in the present study, perhaps because they applied a very different speech elicitation task in which subjects were asked to recite different passages in the anechoic chamber, at casual, normal, raised, loud, and shout levels—a task that is not conversational.

Even though unweighted and *A*-weighted (broadband) levels measured with an omnidirectional microphone are widely used—and probably best for controlling listening tests using loudspeakers—the pressure at the listener's ears is more representative when psychoacoustic data, in particular speech intelligibility outcomes, are evaluated. Therefore, the effect of placing a listener inside the acoustic scene on the speech and noise levels picked up by the listener's ears is illustrated in Fig. 8, using a Brüel and Kjær HATS, type 4128C. The effect of a listener's head on the noise levels averaged across all 13 acoustic scenes and across the left and right ears is shown by the filled circles, and the corresponding standard deviation across scenes is shown by the gray shaded area. The curve more or less represents the diffuse-field head-related transfer function (HRTF) of the HATS, and mainly highlights the pressure boost provided by the ear canal resonance at around 2–3 kHz. The effect of the listener's head on the speech signals is described here by the free-field HRTF for frontal sound incidence and shown by the solid line with triangles. The effect on the resulting SNR is

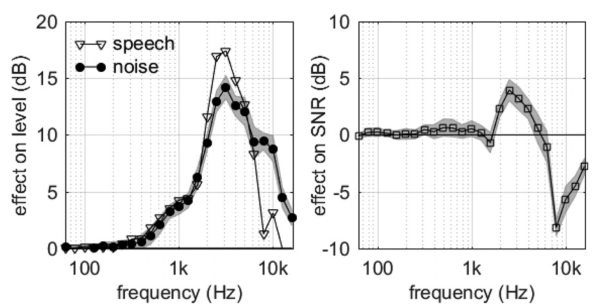

FIG. 8. Effect of the listener's head inside the different acoustic scenes on the speech and noise levels (left), as well as the SNR (right) in third-octave bands and measured at the ear drums of a HATS, referenced to the corresponding pressure picked up by an omnidirectional microphone. This effect averaged across all scenes, as well as over the left and right ear, is indicated by the filled circles and the corresponding standard deviation (STD) by the gray shaded areas. The effect of the listener's head on the speech signals is described by the head-related transfer function (HRTF) of the HATS for frontal sound incidence and shown by the open triangles.

shown in the right panel, which mainly received an improvement of up to 4.5 dB at around 1.5–5 kHz, as well as a significant drop above 5 kHz. Since the frequency range between 1 and 4 kHz is particularly important for speech intelligibility (Accredited Standards Committee S3, Bioacoustics, 1997), the presence of the head may (slightly) improve speech intelligibility. However, neither of the above acoustic measures directly reflects the speech intelligibility that would have been observed by the subjects during their conversations in the different (simulated) acoustic scenes.

B. Comparison to existing studies

The acoustic parameters and respective analysis described by the present study focused on some of the main changes in level associated with the Lombard effect and its implication on SNRs in realistic scenes. One of the main results is the degree of speech level adaptation (increase) for each 1 dB increment in noise level. When inspected independently of distance between the interlocutors, then a consistent slope of 0.43–0.46 dB/dB was obtained at 0.5 and 1 m. This slope is very similar to the average slope of about 0.5 dB/dB reported in Lane *et al.* (1970) for conditions in which subjects communicated in noise. Interestingly, they reported significantly shallower slopes when subjects performed a non-interactive task such as reading a paragraph aloud in noise, which highlights the importance of the interactive puzzle-task that was applied in the present study. Pearsons *et al.* (1977) reported a steeper slope of 0.6 dB/dB in different everyday scenes with levels of 50–70 dB SPL, but almost no change above and below this range of noise levels. Similarly, Wu *et al.* (2018) also derived a piecewise linear fit with a slope of 0.34 dB/dB below 59.3 dB SPL noise and 0.54 dB/dB above it. Considering the behavior of the present data shown in Fig. 3 more closely, the slope of the speech level function increases almost monotonically with increasing noise level, and may be very well described by a similar piecewise linear function as proposed by Wu *et al.* (2018). However, due to the rather steady increase in slope, it is unclear where to place any inflection point(s), which lead here to the fitting of a single linear function with an "average" slope.

In the speech intelligibility index (SII) standard (Accredited Standards Committee S3, Bioacoustics, 1997), four vocal effort levels are given: normal (62.3 dB SPL; 59.2 dBA), raised (68.4 dB SPL; 66.4 dBA), loud (74.8 dB SPL; 73.9 dBA), and shout (82.3 dB SPL; 82.2 dBA). A similar speech level range was reported by Pearsons *et al.* (1977) with male talkers speaking on average at 58–89 dBA and female talkers at 55–82 dBA. Male talkers spoke on average at 58–89 dBA and female talkers at 55–82 dBA (Pearsons *et al.*, 1977, Table I). The speech levels shown in Fig. 3 range from about 62 to 75 dB SPL for male subjects and about 59 to 73 dB SPL for female subjects. This is a very similar range as covered by the normal, raised, and loud speech levels reported in the SII standard as well as in Pearsons *et al.* (1977). The observation that subjects in the present study did not use "shouted" speech suggests that, within conversations, subjects either do not like to shout at each other, or the loudest scene that was applied here did not require the subjects to shout at each other to make themselves understood.

Long-term speech spectra shown in Fig. 4 illustrate a gradual increase in long-term spectral tilt as the scenes become louder. The slope of both male and female talkers decreases by approximately 5 dB from −19 dB/dec to −24 dB/dec, between the softest and loudest scenes. Even though not explicitly stated, these slopes are very similar to the slopes that can be derived from the third-octave levels provided by the SII standard (Accredited Standards Committee S3, Bioacoustics, 1997, Table III), which range from −18 to −24 dB/dec for normal to loud vocal effort levels, as well as the slopes shown in Pearsons *et al.* (1977, Figs. 16 and 17). The similarity of the speech levels and spectra across studies is rather surprising, given the very different methods that were applied. Whereas the present study evaluated conversational speech that was increased in level by presenting realistic acoustic scenes with increasing overall level, the other two studies elicited different speech levels by simply asking the subjects to adjust their vocal effort level while reading speech aloud in a quiet anechoic chamber. Even though the latter task may be considered rather unnatural and unrealistic, it seemed to result in very similar speech levels as in the present study.

Regardless of the derived speech levels, in realistic situations the SNR must not be taken as independent from distance, since the listener can often adaptively correct the distance to improve the effective SNR that they receive, as is illustrated in Fig. 6. In the case that the distance is fixed at 1 m, negative SNRs are observed for noise levels above 69 dB SPL for male and 66 dB SPL for female talkers [top panels of Fig. 6 and Eq. (1)], which is in the typical source level range of standard speech intelligibility tests. However, an SNR of around 0 dB is still higher than commonly used in standard speech tests, which mostly apply SNRs that are well below 0 dB (Sec. I). This mismatch in SNRs becomes even worse when allowing for distance adaptation by the listener, which at these noise levels would result in SNRs of 5.3 dB and 3.7 dB for male and female speech, respectively. In that case, negative SNRs are not encountered until noise levels of 80 dB SPL and 75 dB SPL for male and female talkers, respectively [Eq. (8) and the bottom of Fig. 6], as the average listener listening to a female talker moved 15 cm closer (from 60 to 45 cm) and by the same amount if listening to male speech [from 55 cm to 39 cm; see Eq. (4)]. However, it should be noted that these numbers are gross approximations based on data from Pearsons *et al.* (1977), which were collected with undisclosed methods and contained possible confounds (see Sec. I). Nevertheless, the distance-adjusted SNR approximations shown in the bottom panels of Fig. 6 are similar to SNR data from Wu *et al.* [2018, Fig. 3(B)], which found SNRs of around 0 dB at a noise level of 74 dB SPL on average. Very similar trends are also seen in the field studies of Smeds *et al.* (2015, Fig. 5), but without a numerical fit to the data. These trends suggest that distance was likely a latent variable in producing these largely positive SNRs observed in the two field studies, and may have also contributed to their large SNR variations of more than 20 dB at any given noise level. In this regard, multiple uncontrolled social, cultural, psychological, and situational factors and norms will have influenced the conversation distance within the field studies (see Sec. I) and contributed to the large spread in SNRs.

C. Effect on speech tests

One of the main incentives of this study in measuring realistic SNRs was to evaluate the ecological validity of the speech and noise level ranges employed in clinical speech tests that are commonly used to evaluate speech in noise reception of hearing-impaired listeners and the benefit provided by hearing devices. As was reviewed in the Introduction, these tests assume that background noise and speech levels vary independently, so that the same SNRs can be produced at different combinations of the two levels. As already shown by previous studies (Pearsons *et al.*, 1977; Smeds *et al.*, 2015; Wu *et al.*, 2018) and as the systematic analysis of the current study demonstrates, noise and speech levels cannot be considered independent in many realistic conditions since interlocutors raise their voice level and decrease the distance to their communication partner so that a favorable SNR is maintained despite difficult acoustic conditions. This means that for a rather generally applicable speech test to be of higher ecological validity, the SNR should follow the operational ranges that are suggested by Fig. 6. This may be done in different ways, and depends on whether an adaptive test or a test with fixed SNR should be applied.

Adaptive SRT tests (e.g., Nielsen and Dau, 2009; Nilsson *et al.*, 1994) fix either the noise or speech levels, and vary the other one adaptively until the listener-specific 50% intelligibility is found in terms of SNR, thereby avoiding ceiling and floor effects. However, even though the individual speech levels shown in Fig. 3 vary considerably around the mean, the SNR within an SRT measurement should not be varied more than the observed standard deviation of ± 2.5 dB, if it is to represent realistic speech (and noise) levels. Alternatively, the speech and noise levels could be adjusted simultaneously to adapt the SNR, following the function shown in Fig. 6 and approximated by Eq. (8) or (9).

Similarly, for a fixed-SNR speech test, the noise and speech stimuli would have to change non-adaptively in predefined steps, so that they can reflect representative conversation scenes that take into account the background noise, Lombard speech, and talker-receiver distance. This entails that different regions of SNRs would be covered by different combinations of realistic scenes and effortful speech that cannot be varied adaptively without bounds. The selection of the SNR range of interest may be done according to real-world surveyed noise levels experienced by hearing-aid users (e.g., Wagener *et al.*, 2008, Table II), which are the true independent variable here.

The above constraints that can apply to speech tests pertain to conversations in which the listener is also an active participant, rather than a passive listener to other talkers. In passive listening, the acoustic conditions and social constraints could lead to more challenging SNRs than in active conversation. In many such cases, the passive listener is unable to adapt their distance sufficiently to hear the conversation in full, because of physical constraints, as objects standing in between or standing too close to the other interlocutors. Passive listening may also be more significantly affected by the head orientation of the other talkers, as their speech would suffer from high-frequency attenuation when they look away from the passive listener due to the directional properties of human speech (Monson *et al.*, 2012). Additional complications may arise due to the various social factors that are at play in conversations, as the proxemics literature suggests. For example, a cultural or gender difference between talkers (Sussman and Rosenfeld, 1982) can be asymmetrical across the interlocutors, creating interpersonal tension and instability of distance between them. These constraints may be exacerbated if one of the interlocutors has a hearing impairment and wishes to get to the unaware normal-hearing talker closer than they are comfortable with. Also, when a hearing-impaired listener actively participates in a conversation, then the communication partner may adjust their communicative behavior to effectively increase the SNR, which would not be the case in passive listening.

D. Limitations

The applied laboratory-based methods involved some practical compromises that may have limited the overall degree of realism that was accomplished. For instance, even though the headphones that were applied to present the realistic acoustic scenes to the subjects were specifically chosen for their high acoustic transparency, they still provided an attenuation to the external speech signals for frequencies above about 1.2 kHz (Fig. 2). However, any potential effect on the produced speech levels was minimized by applying the same attenuation to the noise stimuli, which ensured that natural SNRs were more or less maintained. Hence, even though it cannot be excluded that the passive attenuation of the headphones had any effect on the produced speech levels, this effect would have been rather small. Another issue with the binaural headphone reproduction is that they did not incorporate head-tracking, so in parts when the subjects turned their heads, the scenes may have sounded less realistic.

Moreover, the subjects were tested in an anechoic chamber and therefore could only hear the direct sound of the interlocutor's voice with no sound reflections from the walls. Therefore, the speech they heard had lower energy than would have been experienced in real life, it was clearer and may have not matched the subjective expectations based on the acoustics of the noisy scenes presented over headphones. However, due to the close distance between the two subjects, the direct-to-reverberation energy ratio would have been rather high (and positive) for all the considered scenes, suggesting that providing realistic reverberation would have had a rather small effect on the produced speech levels. Similarly, the missing reverberation could have affected the own-voice perception-related adaptive communication behavior of the subjects. However, it is known that reverberation has a very limited effect on the adaptive behavior of speech (Pelegrín-García *et al.*, 2011), and therefore, may have had only a negligible effect on the final speech levels and SNRs.

Finally, a significant part of the analysis is based on the only available data on realistic talker-listener distances in noise (Pearsons *et al.*, 1977), which provides very limited details on the applied methods and analyses, and may be confounded by the sampling of the acoustic scenes. Hence, there is a need for highly controlled studies that further characterize the effect of realistic noise on the distance, as well as speech levels that two (or more) subjects adapt to when communicating with each other.

V. CONCLUSION

The present study measured the speech levels and SNRs of natural conversations between two people at two fixed distances when they were subjected to 13 realistic acoustic scenes of different levels, which were presented via open headphones. Based on data by Pearsons *et al*. (1977), the derived data allowed estimating the realistic SNRs and speech levels for male and female talkers as a function of noise level, assuming that people would adjust (i.e., reduce) their distance as levels increase. The realistic, distance-adjusted SNRs were found to be positive for noise levels of up to about 75 dB SPL, confirming previous results from field studies but by applying far more rigorous and controlled experimental methods. These findings support more robust assumptions about realistic SNRs that are encountered in typical situations, which are important for the design of realistic speech intelligibly tests, as well as fitting rules for hearing aids.

ACKNOWLEDGMENTS

Many thanks to Katrina Freeston for her help with administering the test and to Greg Stewart for his dedicated technical support. Also, thanks to Kelly Miles for assistance with the statistical analysis. This study was funded by the Oticon Foundation and Macquarie University via an iMQRES Ph.D. scholarship.

Accredited Standards Committee S3, Bioacoustics (**1997**). ANSI s3.5-1997, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**(1), 1–48.

Beechey, T., Buchholz, J. M., and Keidser, G. (**2018**). "Measuring communication difficulty through effortful speech production during conversation," Speech Commun. **100**, 18–29.

Bench, J., Kowal, Å., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**(3), 108–112.

Bode, D., and Carhart, R. (**1973**). "Measurement of articulation functions using adaptive test procedures," IEEE Trans Audio Electroacoust. **21**(3), 196–201.

Brimijoin, W. O., Hadley, L. V., Naylor, G. M., and Whitmer, W. M. (**2017**). "Parametric measurements of natural conversation behaviour reveal effects of background noise level on speech, movement, and gaze," in *6th International Symposium on Auditory and Audiological Research (ISAAR)*.

Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (**2012**). "Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task," Hear. Res. **283**(1–2), 162–168.

Carhart, R., and Tillman, T. W. (**1970**). "Interaction of competing speech signals with hearing losses," Arch. Otolaryngol. **91**(3), 273–279.

Ching, T. Y., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**(1), 9–21.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987a**). "Development of the connected speech test (CST)," Ear Hear. **8**(5), 119S–126S.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987b**). "Intelligibility of average talkers in typical listening environments," J. Acoust. Soc. Am. **81**(5), 1598–1608.

Culling, J. F. (**2016**). "Speech intelligibility in virtual restaurants," J. Acoust. Soc. Am. **140**(4), 2418–2426.

Grange, J. A., and Culling, J. F. (**2016**). "The benefit of head orientation to speech intelligibility in noise," J. Acoust. Soc. Am. **139**(2), 703–712.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**(2), 79–87.

Hagerman, B. (**1984**). "Clinical measurements of speech reception threshold in noise," Scand. Audiol. **13**(1), 57–63.

Hanks, W. D., and Johnson, G. D. (**1998**). "Hint list equivalency using older listeners," J. Speech, Lang. Hear. Res. **41**(6), 1335–1340.

IEC (**2011**). IEC 60268-16, edition 4.0 (2011-06), "*Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*," (International Electrotechnical Commission, Geneva, Switzerland).

Junqua, J.-C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. **93**(1), 510–524.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**(5), 1337–1351.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech, Lang. Hear. Res. **14**(4), 677–709.

Lane, H., Tranel, B., and Sisson, C. (**1970**). "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**(2B), 618–624.

Levitt, H., and Rabiner, L. R. (**1967**). "Use of a sequential strategy in intelligibility testing," J. Acoust. Soc. Am. **42**(3), 609–612.

Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of the elevation of the voice"), Ann. Maladiers Oreille. **37**, 101–119.

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., and Ng, E. H. N. (**2016**). "Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions," Ear Hear. **37**, 145S–154S.

MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**(1), 29–43.

Monson, B. B., Hunter, E. J., and Story, B. H. (**2012**). "Horizontal directivity of low- and high-frequency energy in speech and singing," J. Acoust. Soc. Am. **132**(1), 433–441.

Naylor, G. (**2016**). "Theoretical issues of validity in the measurement of aided speech reception threshold in noise for comparing nonlinear hearing aid systems," J. Am. Acad. Audiol. **27**(7), 504–514.

Nielsen, J. B., and Dau, T. (**2009**). "Development of a Danish speech intelligibility test," Int. J. Audiol. **48**(10), 729–741.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). *Speech Levels in Various Noise Environments* (US Environmental Agency, Office of Health and Ecological Effects, Office of Research and Development, Washington, DC).

Pelegrín-García, D., Smits, B., Brunskog, J., and Jeong, C. (**2011**). "Vocal effort with changing talker-to-listener distance in different acoustic environments," J. Acoust. Soc. Am. **129**(4), 1981–1990.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech, Lang. Hear. Res. **29**(2), 146–154.

Ricketts, T. A., and Hornsby, B. W. (**2005**). "Sound quality measures for speech in noise through a commercial hearing aid implementing," J. Am. Acad. Audiol. **16**(5), 270–277.

Smeds, K., Wolters, F., and Rung, M. (**2015**). "Estimation of signal-to-noise ratios in realistic sound scenarios," J. Am. Acad. Audiol. **26**(2), 183–196.

Sommer, R. (**2002**). "Personal space in a digital age," in *Handbook of Environmental Psychology*, edited by R. B. Bechtel and A. Churchman (John Wiley & Sons., Inc., New York), pp. 647–660.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**(2), 212–215.

Sussman, N. M., and Rosenfeld, H. M. (**1982**). "Influence of culture, language, and sex on conversational distance. J. Pers. Soc. Psychol. **42**(1), 66.

Wagener, K. C., Hansen, M., and Ludvigsen, C. (**2008**). "Recording and classification of the acoustic environment of hearing aid users," J. Am. Acad. Audiol. **19**(4), 348–370.

Warren, R. (**1968**). "Vocal compensation for change of distance," in *Proceedings of the 6th International Congress of Acoustics*, International Commission for Acoustics Tokyo, pp. 61–64.

Weisser, A. (**2018**). "Complex acoustic environments: Concepts, methods, and auditory perception," Ph.D. thesis, Macquarie University.

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., and Oleson, J. (**2018**). "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss," Ear Hear. **39**(2), 293–304.

Zahorik, P., and Kelly, J. W. (**2007**). "Accurate vocal compensation for sound intensity loss with increasing distance in natural environments," J. Acoust. Soc. Am. **122**(5), EL143–EL150.