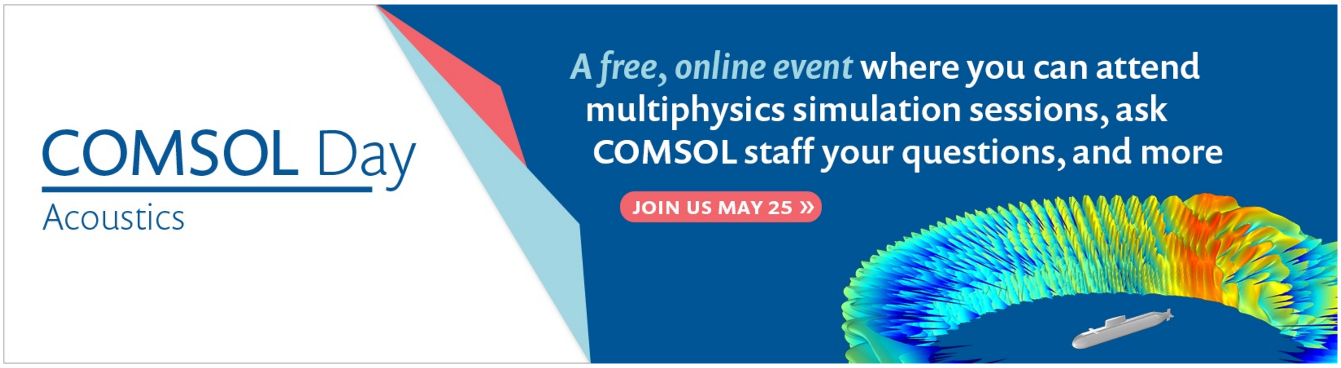
Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments

Adam Weisser, Kelly Miles, Michael J. Richardson, and Jörg M. Buchholz

Citation: The Journal of the Acoustical Society of America **149**, 2896 (2021); doi: 10.1121/10.0004774
View online: https://doi.org/10.1121/10.0004774
View Table of Contents: https://asa.scitation.org/toc/jas/149/4
Published by the Acoustical Society of America



JASA ARTICLE

Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments

Adam Weisser,[1,a)] Kelly Miles,[1,b)] Michael J. Richardson,[2,c)] and Jörg M. Buchholz[1,d)]

[1]*Department of Linguistics, Australian Hearing Hub, Macquarie University, 16 University Avenue, New South Wales 2109, Australia*

[2]*Department of Psychology and Perception in Action Research Centre, Faculty of Human Sciences, Macquarie University, Sydney, New South Wales 2109, Australia*

ABSTRACT:

Everyday environments impose acoustical conditions on speech communication that require interlocutors to adapt their behavior to be able to hear and to be heard. Past research has focused mainly on the adaptation of speech level, while few studies investigated how interlocutors adapt their conversational distance as a function of noise level. Similarly, no study tested the interaction between distance and speech level adaptation in noise. In the present study, participant pairs held natural conversations while binaurally listening to identical noise recordings of different realistic environments (range of 53–92 dB sound pressure level), using acoustically transparent headphones. Conversations were in standing or sitting (at a table) conditions. Interlocutor distances were tracked using wireless motion-capture equipment, which allowed subjects to move closer or farther from each other. The results show that talkers adapt their voices mainly according to the noise conditions and much less according to distance. Distance adaptation was highest in the standing condition. Consequently, mainly in the loudest environments, listeners were able to improve the signal-to-noise ratio (SNR) at the receiver location in the standing condition compared to the sitting condition, which became less negative. Analytical approximations are provided for the conversational distance as well as the receiver-related speech and SNR.

https://doi.org/10.1121/10.0004774

(Received 30 June 2020; revised 12 March 2021; accepted 1 April 2021; published online 29 April 2021)

[Editor: Jonas Braasch] Pages: 2896–2907

I. INTRODUCTION

Real-life human conversations require continuous adaptation between interlocutors with each other and the environmental conditions. A growing concern in hearing research has been to accurately estimate the associated conversational speech levels that are found in realistic noisy situations (Smeds *et al.*, 2015; Wu *et al.*, 2018). These values are used to inform the speech-in-noise performance of hearing-impaired individuals, as well as the design of suitable hearing devices, which can deliver good performance in everyday situations experienced by the device users (Naylor, 2016). Because of the variable conditions found in the real world, a talker uses different strategies to optimize the listener's speech reception: adapt their vocal effort (Cooke *et al.*, 2014; Lane and Tranel, 1971), turn their head to the listener (Brimijoin *et al.*, 2012; Grange and Culling, 2016), reduce the distance (Hadley *et al.*, 2019; Pearsons *et al.*, 1977), and facilitate lipreading, particularly in adverse listening conditions (Sumby and Pollack, 1954). Nevertheless, available literature that surveys real-world speech and noise levels provides only limited data on the effect of talker adaptation of their distance in conversation acoustics.

Only one published report provided field data on the distances between interlocutors—the "conversational distance"—in real-world conversations (Pearsons *et al.*, 1977). As part of that study, the conversational distance was measured as a function of background noise level in homes, hospitals, department stores, and onboard trains and airplanes. It was found that the conversational distance decreases with noise level. In the quietest environments (homes) the typical distance was about 0.9–1 m at 35–40 dBA, whereas in the loudest environments (trains and airplanes) the typical distance was 0.2–0.4 m at 75–80 dBA (see Fig. 22 from Pearsons *et al.*, 1977). However, it was not reported how these distances were measured in the different locations. Furthermore, it is not clear whether interlocutors in both trains and airplanes were sitting or standing when these estimates were made—a factor that can severely constrain talkers from effectively adapting their conversational distance.

A more recent study fixed the mouth-to-ear distances between two seated interlocutors to 0.5 and 1 m and simulated realistic noise environments by playing binaural recordings of acoustic environments through open headphones (Weisser and Buchholz, 2019). The test yielded realistic conversational speech sound pressure levels at known noise levels, from which the respective signal-to-noise ratios (SNRs) could be readily computed. By assuming that the

---

[a)] ORCID: 0000-0003-2261-8349.

[b)] ORCID: 0000-0002-4104-980X.

[c)] ORCID: 0000-0001-9159-2774.

[d)] Electronic mail: jorg.buchholz@mq.edu.au, ORCID: 0000-0001-6188-9761.

average surveyed distance data from Pearsons *et al.* (1977) (see their Fig. 22) holds true and interlocutors talk more softly the closer they are, it was possible to generate an interpolated model of the mean distance and SNRs at arbitrary noise levels between 53 and 80 dB SPL (sound pressure level; see Fig. 6) from Weisser and Buchholz, 2019). However, as this model relies on the not-fully-transparent distance data by Pearsons *et al.* (1977), the associated SNR prediction was subject to an unknown error. Additionally, the task in that study prevented the interlocutors from adapting their distance and body posture in real-time, which could have been an important factor in determining the optimal SNR for the specific noise stimuli that may vary in other ways beyond its average level. Instead of treating the SNR as a cause of distance adaptation, the fixed distance and noise level meant that SNR could only be adapted by raising the speech level. Therefore, there remains an uncertainty as to the realistic conversational speech levels and SNRs when interlocutors are free to optimize their relative positions.

In another recent study by Hadley *et al.* (2019), two interlocutors were free to adapt their conversational distance to improve the speech level by leaning forward with their torsos, while seated 1.5 m apart (ear-to-ear distance). Speech-shaped noise was generated by a ring of eight loudspeakers in an anechoic chamber, which varied every 15–25 s between five levels (54–78 dB, SPL). The conversational distance was measured using a head tracker and was found to decrease by an average of 1 cm for every 6 dB of noise level increase, which led to a negligible change in received speech level of 0.01 dB/dB. The talkers also increased their speech level by an average of 1.9 dB for every 6 dB of noise level increase and had a shorter mean length of utterance. Eye gaze was also tracked and, in line with previous studies (Bavelas *et al.*, 2002; Vertegaal *et al.*, 2001), found to vary subtly at high noise levels, as interlocutors were more focused on the talker's mouth while listening (compared to while talking). Head turns were tracked as well, but the change as a function of noise level was rather small, as listeners tended to primarily vary their yaw angle (azimuth) with noise level. While simultaneously introducing distance, head turns, and eye gaze as relevant dynamic variables in free conversations, this study strongly constrained the baseline of the interlocutors by seating them at a relatively large distance for these noise levels (Pearsons *et al.*, 1977), which may have led participants to compensate primarily by increasing their vocal effort, rather than by decreasing their conversational distance. Additionally, Hadley *et al.* (2019) applied a rather artificial noise stimulus that, on the one hand, offered excellent experimental control but, on the other hand, limited generalization of their results to the real world. This latter concern highlights the common compromise that has to be made between experimental control and ecological validity. In this regard, Pearsons *et al.* (1977) and Weisser and Buchholz (2019) focused on ecological validity by analyzing speech levels in different real or realistic acoustic environments while sacrificing some experimental control. In particular, they treated noise level as the main acoustic parameter, even though their environments varied in many other ways, including room reverberation as well as the type, location, and number of sound sources.

In addition to optimizing acoustic conditions during conversations, there is extensive literature that indicates that interlocutors strive to maintain a physically and psychologically comfortable personal space, which is also psychosocially, and culturally acceptable (Hall, 1990; Hayduk, 1983; Sommer, 2002). This means that conversational distances and dependent SNRs are also constrained by nonacoustic factors (see also discussion in Hadley *et al.*, 2019). For example, in two field studies that focused on differences between European cultural norms, the conversational distance of naive interlocutors was estimated in urban locations (Remland *et al.*, 1991, 1995). The mean range of head-to-head distance was 38–73 cm (or 25–58 cm torso-to-torso distance) across cultures and sexes, although talkers were not facing each other head-on but rather turned to each other at approximately right angles, or, at more obtuse angles (Remland *et al.*, 1991). The noise levels in these two studies were not considered, but it can be assumed that they were not negligible. This may have drawn the interlocutors closer than was reported in laboratory-based observations (e.g., Bell *et al.*, 1988; Sussman and Rosenfeld, 1982), similarly to other field studies (Shuter, 1977). Surprisingly, noise level and vocal effort were mentioned only in passing in the aforementioned literature (Hall, 1990, pp. 116, 118, 113–125, and 142–143) and also not included as central factors in later reviews (e.g., Harrigan, 2005; Hayduk, 1983; Matsumoto *et al.*, 2016; Sommer, 2002).

It can be concluded that while the speech and noise level values are sufficient to determine the SNR in a given acoustic situation, these values provide an incomplete picture without the associated conversational distances, as well as some of the personal context of the situation. The present study attempts to fill the gap in the acoustic literature by tracking the conversational distance data in free conversations between pairs of interlocutors in realistic noisy environments. The data should provide a rigorous supplement to the distance and SNR data in Pearsons *et al.* (1977), while controlling for several personal variables (sex, age, familiarity)—all of which are known to affect the conversational distance.

II. METHODS

A. Speech recording and processing

Each participant was equipped with a DPA d:fine^{TM} FIO66 omnidirectional headset (boom) microphone (DPA Microphones A/S, Alleroed, Denmark), which picked up their speech at a close distance to their mouths with a high SNR during overlapping speech. The microphone signals were sent by body-worn SK-D1 Sennheiser transmitters (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany) to two separate EM-D1 Sennheiser receivers (one microphone-transmitter set per participant) connected to the

RME UFX sound card (RME Audio AG, Haimhausen, Germany) of the test computer. The speech signals were recorded over the entire two-minute duration of the noise stimulus, but only the last 90 s were considered in the subsequent analysis to exclude any initial adaptation effects. The recordings were done at a sampling frequency of 44.1 kHz and 24 bit depth.

Since the sensitivity of the headset microphones depended on the specific positioning on the participants, they were individually calibrated following the procedure described in Beechey *et al.* (2018). In brief, each participant was asked to count aloud, while their voice was being recorded for 30 s with their headset microphone. Additionally, the voice was recorded with a calibrated $\frac{1}{4}$ in. omnidirectional G.R.A.S. type 46BL microphone (G.R.A.S. Sound & Vibration A/S, Holte, Denmark) placed at a reference position of 1 m in front of the participant. The participant-specific calibration gain was then derived by comparing the root-mean-square (rms) levels of the two speech recordings.

Due to the test room reverberation (Sec. II E), the speech recorded by the reference microphone contained significant reverberant energy. This was not the case for the very close headset microphone, which mainly picked up the participant's direct sound. Therefore, a "room gain correction" was applied to the calibration gain ensuring that only the direct sound was taken into consideration. This gain correction was derived by replacing the participant with a Genelec 8020C loudspeaker (Genelec Oy, Iisalmi, Finland) and then measuring the room impulse response (RIR) from the loudspeaker to the reference microphone (at a distance of 1 m). Afterward, the RIR was split into its direct and reverberant sound components using a frequency-dependent time window (Weisser *et al.*, 2019a). The effect of the room reverberation on the microphone calibration was then estimated by separately convolving the complete RIR and its direct-sound-only component with the 384 realistic sentences from the ECO-SiN corpus (Miles *et al.*, 2020). The difference of the resulting rms levels then provided a room gain correction of 7.2 dB, which was subtracted from the individual calibration gains. The corrected calibration gains were finally applied to the speech recorded with the headset microphones.

Even though the headset microphones maximized the SNR of the recorded speech, some acoustic crosstalk between headset microphones was still observed, in particular in the loudest test conditions where participants were rather close to each other. To minimize the crosstalk effect on the estimated speech levels, the same procedure as described in Beechey *et al.* (2018) was applied. In brief, the two recording channels were segmented using 30-ms long Hann windows with 50% overlap. The rms level of each segment was compared across channels and only the channel with the higher level was retained (i.e., the signal in the other channel was multiplied by zero). Using the "cleaned" (and calibrated) speech signals, their unweighted rms levels without pauses were calculated using the procedure described in IEC (2011, see Annex J). Finally, to provide sentence-equivalent speech levels that include natural pauses, a level correction (i.e., attenuation) of 1.87 dB was applied as further described in Weisser and Buchholz (2019). This resulted in the source-related speech levels reported below.

B. Noise stimuli

Participants listened to five scenes from the original Ambisonic Recordings of Typical Environments (ARTE) database (Weisser *et al.*, 2019b), where their levels have been reported: Library (53 dB SPL), Living Room (63.3 dB SPL), Café 2 (71.7 dB SPL), Train Station (77.1 dB SPL), and Food Court 2 (79.6 dB SPL).[1] Two additional scenes were not in the original database but were created in the same way: party without background music (85.0 dB SPL), and party with background music (92.0 dB SPL). These two scenes were necessary to simulate realistic conditions that strongly exacerbate SNR and may require people to get close together in order to converse. The acoustic environments were presented at their realistic levels, i.e., as they would be experienced by participants *in situ*, to elicit ecologically-valid Lombard speech (Beechey *et al.*, 2018; Weisser and Buchholz, 2019). Non-individualized binaural recordings were used for the test, which were low-pass filtered above 2000 Hz, to match the effect of the headphones on the speech that is slightly attenuated at mid-high frequencies by the open circumaural Sennheiser HD-800 headphones. The headphones were equalized and calibrated on a Brüel & Kjær type 4128 °C head and torso simulator (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark) and connected to the test computer *via* a RME Fireface UFX USB sound card and a wireless Sennheiser SR 300 IEM stereo transmitter, with two separate Sennheiser EK 300 IEM stereo receivers worn by the participants. Further details on the creation and playback of the noise stimuli are provided in Weisser and Buchholz (2019).

C. Motion capture

The position of participants was tracked using a Polhemus Latus motion tracking system (Polhemus LTD, Vermont, USA), which tracks the position and rotation (i.e., six degrees of freedom, 6 DoF) of wireless sensors or markers in reference to a set of motion tracking receptors that define the tracking area. The system uses an electromagnetic field to track the position of sensors, and because it does not rely upon line-of-sight, it is not affected by occlusion and consistent full 6DoF tracking was always achieved.

Two wireless motion sensors were used to track each participant's head location. The head was chosen, rather than the participant's torso or center-of-mass position, so that the head-to-head distance can be used to estimate the exact acoustic paths between mouth and ear. One sensor on each side was attached to a custom three-dimensional (3D) cap that was attached to the headphones to ensure that the grills were not blocked. Motion data were recorded at a sampling rate of 120 Hz in reference to two sensor receptors that were positioned on top of two 1.25 m high receptor stands, which were positioned 1.5 m apart along the orthogonal centerline between the participant's initial positions. This ensured that each motion sensor was always less than 1 m from a sensor receiver and that the positional measurement error was less than ±0.5 mm. These receptor locations also resulted in the accurate tracking of forward-back (anterior-posterior) and left-right (medial-lateral) movements of participants' heads when facing directly towards each other at the beginning of a trial, corresponding to the planar y- and x-dimensions of sensor motion, respectively. Last, motion sensor data were recorded on the same PC used to present the audio stimuli and were synchronized with the start and end of the audio stimulus presentations.

Prior to analysis, the recorded motion data were low-pass filtered using a 10 Hz, 4th-order Butterworth filter to remove measurement noise. Inter-participant distance was calculated as the planar distance between the average (x, y) location of the motion sensors attached to the left and right caps of the participants' headphone drivers during the last 5 s of the trial. This measure corresponded to the central position of each participant's head, halfway between the two ears. We chose to focus on conversational distance during the last 5 s of each condition to ensure that all participants had enough time to adapt their position to the communication scenario.

D. Subjects

Fifty-six subjects participated in this study. All were screened for normal hearing [pure tone thresholds better than 20 dB HL (hearing level) at 500, 1000, 2000, and 4000 Hz]. Participants were tested in pairs who were familiar and comfortable with each other—either friends, couples, or siblings (see Table I). Participants were recruited by word of mouth or flyers distributed across Macquarie University, so they were primarily students who were compensated for their participation. Due to technical problems during testing, the data from four pairs (eight participants) had to be removed from the analysis. Problems were mainly related to dropouts in the transmission of the wireless motion sensor data or the speech recordings. Two more pairs did not perform the experiment because they did not meet the audiometric requirements. Hence, only the data from 22 pairs (44 participants) were considered in the analysis. Out of these, 32 participants were female with an average age of 22.2 ± 4.2 years (± 1 standard deviation). Twelve participants were male with an average age of 24.4 ± 5.8 years. A two-sample t-test did not show any significant difference in age between the male and female participants ($p = 0.17$).

TABLE I. Number of talker-pairs grouped according to their relationship and sex.

	Couples	Siblings	Friends	Total
Male–male	0	1	1	2
Male–female	6	1	1	8
Female–female	0	0	12	12
Total	6	2	14	22

E. Procedure

The experiment took place in a room of dimensions $4.11 * 2.59 * 2.54$ m^3 and a reverberation time of $T_{30} = 0.7$ s. Participants wore open headphones, a boom microphone, and motion trackers – all wireless – that enabled them to move more or less freely in the room. Noise stimuli were played *via* the headphones at a realistic level (see Sec. II B). There were two conditions—one standing and one sitting. In the standing condition, participants held conversations while standing and always started from well-separated positions at around 2.5 m. In the seated condition, they were sitting on opposite sides of a table ($0.76 * 0.74$ m^2) and, in the very beginning, were allowed to move their chairs toward each other, to a comfortable distance. Both chairs and table were made from polypropelene resin that did not interfere with the motion capture signal. Once the noise stimulus began, they were allowed to move freely, until the noise stimulus finished. Subjects were instructed to speak naturally with one another: "…Try to relax and talk naturally… While your conversations will be recorded, we are not actually interested in what you are talking about—we are just interested in the way your voices sound. We have included a few conversation starters for you on this board in case you get stuck and can not think of anything to say. Really, we just need both of you to talk throughout the study. So if you notice that you are talking more than the other person, try asking a question…"

Two training conditions were played at the start of the experiment with the softest and loudest noise, to accustom the participants to the noise level range. The main experiment was conducted in two blocks, each containing the seven different noise stimuli in a random order. The two blocks referred to the seated and standing condition with their order counterbalanced. The complete experiment including setup took 1–1.5 h per pair.

F. Statistical analysis

All analyses were performed in R version 4.0.2 using the packages *lme4* (Pinheiro *et al.*, 2017) and *emmeans* (Lenth *et al.*, 2018). Linear mixed-effects (LME) models with a random intercept were used for all analyses to control for the individual effect of repeated measures over different levels. LME regression models with environment, sex, and condition as predictor variables were developed, in order to determine their effect interactions on speech level, conversational distance, adjusted speech level, and relationship. If there were no significant interaction, the main-effect model was presented. Tukey-adjusted p-values <0.05 were considered significant for all analyses.

III. RESULTS

In the following analysis, it is assumed that in the present communication task the applied acoustic environments can be characterized mainly by their sound pressure level, given their wide range of levels from 53 to 92 dB SPL. However, other acoustic properties may have confounded the participants' behavior, including the type, number, and location of the involved noise sources in the original environments.

A. Source-related speech level at 1 m distance

The unweighted (broadband) speech level at 1 m in front of the talkers is shown in Fig. 1 as a function of the unweighted noise level of the seven acoustic environments. The speech levels for the 32 female participants are shown in the upper panels and for the 12 male participants in the lower panels. The speech levels for the participant-pairs sitting at the table are shown in the left panels and when standing freely inside the test room in the right panels. The individual data are shown by the gray lines and their average is shown by the black lines with circles. The dashed lines are identical in all panels and refer to a second-order polynomial fit to the average data across all conditions and groups, with the noise level given by x and the speech level $y = a_2x^2 + a_1x + a_0$ and $a_2 = 0.007$, $a_1 = -0.635$, and $a_0 = 77.174$ dB SPL.

There were no significant interaction terms in the LME model. A main-effect model showed a significant effect of noise [$F(6, 549) = 236$, $p < 0.01$], sex [$F(1, 393) = 15$, $p < 0.01$], and condition [$F(1, 549) = 20$, $p < 0.01$]. *Post hoc* analysis indicated that speech level increases significantly with increasing noise level, excluding the change between the café and train station environments. In addition, the estimated marginal mean for speech level was 3.1 dB lower, on average, for females compared to males, and 1.1 dB lower, on average, for the sitting compared to the standing condition. This spread of speech levels is similar for the male and female participants in both the sitting and the standing conditions. Similarly to Weisser and Buchholz (2019), the variance has the tendency to decrease with increasing noise level. Overall, participants seem to adjust their speech level primarily to the noise level but not to the communication scenario (i.e., sitting versus standing). See Table S1 for detailed model main effects contrasts.[2]

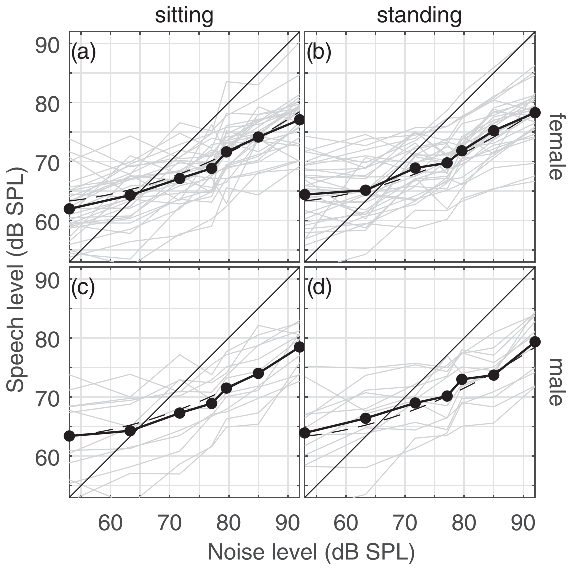


FIG. 1. Unweighted (broadband) speech level at 1 m distance in front of the talkers as a function of noise level of the different acoustic environments. The upper panels refer to the 32 female talkers and the lower panels to the 12 male talkers. The left panels refer to the condition where the participant-pairs were sitting at a table and the right panels where they were standing freely in the test room. The gray lines show individual data and the black lines with circles show average data. The dashed lines are identical in all panels and refer to a second-order polynomial fit to the average data across all conditions and groups.

B. Conversational distance

The distance as a function of noise level between the 22 participant pairs is shown in Fig. 2 separately for the sitting (left panel) and standing conditions (right panel). Above a certain noise level (about 72 dB SPL for the sitting condition and about 64 dB SPL for the standing condition), the talkers start getting closer to each other, as the noise level increases. This effect is less pronounced in the sitting condition, where movement is restricted by the chairs and the table between the participants. Moreover, particularly in the standing condition, participants exhibit different behavioral patterns with increasing noise level. Some participants stayed rather far from each other even in the loudest conditions, whereas others got very close with increasing noise level.

The only significant interaction was between noise and condition [$F(6, 550) = 60$, $p < 0.01$]. A *post hoc* analysis showed a significantly larger conversational distance in the standing condition for the three softest environments: Library (0.7 m), Living Room (0.7 m), and Café (0.3 m) environments. No significant differences in conversational distance were observed in the other four environments. See Table S2 for detailed model contrasts.[2]

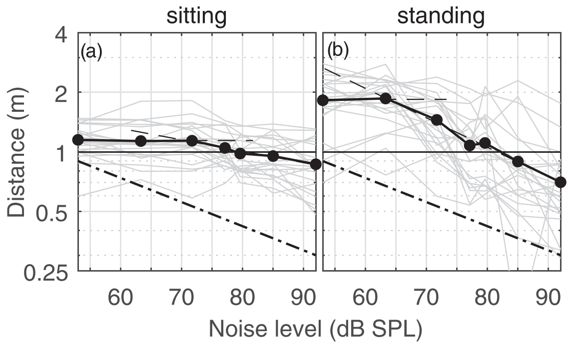


FIG. 2. The distance between the 22 talker pairs as a function of noise level of the different acoustic environments for the sitting (left panel) and standing (right panel) conditions. The dashed lines refer to broken stick line approximations to the distance data (see the text). The dashed-dotted lines refer to the approximations from Weisser and Buchholz (2019) that were based on data from Pearsons *et al*. (1977).

The change in distance with increasing noise level may be approximated by a broken stick function on a double-logarithmic scale (see dashed lines in Fig. 2). At soft noise levels, no change in distance is observed, which is then fitted with a horizontal line at a fixed distance of 1.14 m in the sitting condition and 1.84 m in the standing condition. For higher noise levels, a linear segment similar to Eq. (4) from Weisser and Buchholz (2019), is fitted, which is given by

$$\log D = -a \cdot L_n + b, \qquad (1)$$

with D the distance in meters, L_n the noise level in dB SPL, and a and b two parameters. For the sitting condition, these parameters were $a = 0.0058$ and $b = 0.465$, and for the standing condition $a = 0.0149$ and $b = 1.214$. Hence, applying Eq. (1), the talkers halve their distance with every increase in noise level of $\Delta_n = \log 2/a$ dB, which is $\Delta_n = 52$ dB in the sitting condition and $\Delta_n = 20.2$ dB in the standing condition. The noise level at which the average talker distance is equal to 1 m is 80.5 dB SPL in the sitting condition and 81.6 dB SPL in the standing condition. For comparison, the interpolations of the talker distance as a function of noise level from Eq. (4) of Weisser and Buchholz (2019), which were based on distance data reported by Pearsons *et al.* (1977), are shown by the dashed-dotted lines in Fig. 2. These curves indicate much shorter distance estimates based on that previous data set.

C. Speech level at the receiver location

The unweighted speech level at the receiver location is shown in Fig. 3 as a function of the noise level of the seven acoustic environments. This receiver-related (or distance-adjusted) speech level was derived from the source-related speech level shown in Fig. 1 by taking the distance of the talkers shown in Fig. 2 into account, following the transformation:

$$L_{rec} = L_{src} - 20 \log D, \qquad (2)$$

with L_{rec} the receiver-related speech level and L_{src} the source-related speech level given in Fig. 1. Hence, similarly to an omnidirectional sound source in free-field, it is assumed in Eq. (2) that with every halving of distance the receiver-related speech level increases by 6 dB. Combining this with the Lombard speech effect of Eq. (1), the change in conversational distance effectively raises the speech level at the receiver location by about 0.12 dB for every 1 dB increase in noise level in the sitting condition and 0.30 dB per 1 dB in the standing condition.

Similar to Fig. 1, the receiver-related speech level for the 32 female participants is shown in the upper panels and for the 12 male participants in the lower panels of Fig. 3. The speech levels for the participant pairs seated at the table are shown in the left panels and when standing freely inside the test room in the right panels. The individual data are shown by the gray lines and their average is shown by the black lines with circles. The dashed lines are identical in all panels and refer to a second-order polynomial fit to the average data across all conditions and groups with $y = a_2x^2 + a_1x + a_0$ and $a_2 = 0.011$, $a_1 = -1.030$, and $a_0 = 83.902$ dB SPL.

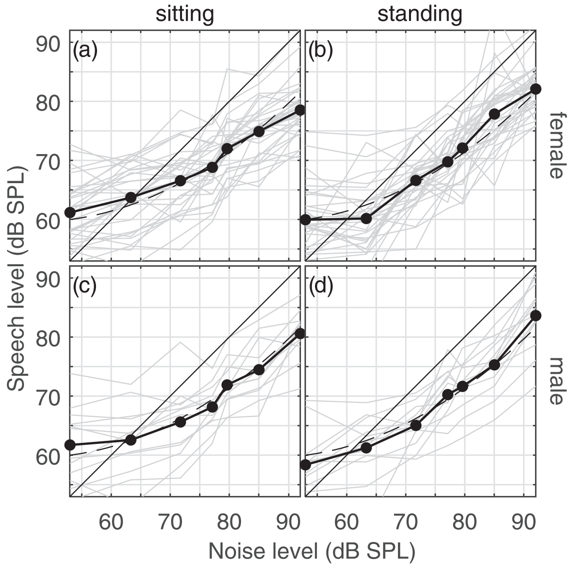


FIG. 3. Similar to Fig. 1, except that here the receiver-related speech levels were derived from Fig. 1 by taking into account the talker distance data from Fig. 2 (see the text). The dashed lines are identical in all panels and refer to a second-order polynomial fit to the average data across all conditions and groups.

The variance across participants is similar to the source-related speech levels in Fig. 1, with an average inter-participant standard deviation of 5.0 ± 0.8 dB. There was a main effect of sex [$F(1, 244) = 6$, $p < 0.05$] and a significant interaction between noise level and condition [$F(6, 548) = 6.5$, $p < 0.01$] for the source-related speech levels. *Post hoc* analysis revealed the estimated marginal mean for males was 2.3 dB higher, on average, than females, and significant differences between the sitting and standing condition for the living room environment (2.9 dB higher in the standing condition) and music party environments (3.4 dB higher in the sitting condition). See Table S1 for detailed model contrasts.[2]

In order to better compare the source-related speech levels from Fig. 1 with the receiver-related (or distance-adjusted) speech levels from Fig. 3, the group-average speech levels (i.e., the arithmetic means of the arithmetic group means) are replotted in Fig. 4 separately for the sitting (left panel) and standing (right panel) conditions. Additionally, the source-related (or distance-adjusted) speech level approximations from Weisser and Buchholz (2019) are shown by the dashed-dotted lines in Fig. 4, which were derived here by replacing the talker distance approximations (as a function of noise level) given in Eq. (4) of Weisser and Buchholz (2019) by the corresponding broken-stick approximations shown by the dashed lines in Fig. 2. Details are given in Sec. IV C.

D. SNR at the receiver location

The average SNR at the receiver location as a function of noise level is shown in Fig. 5 (solid lines with filled

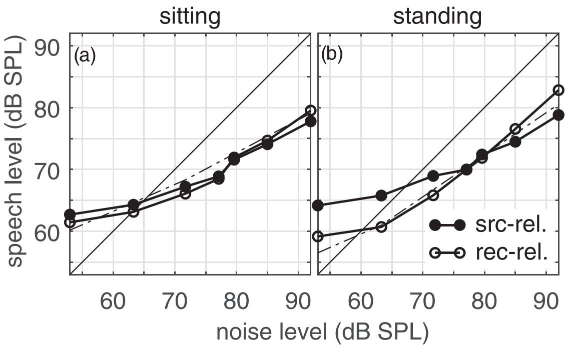


FIG. 4. Average source-related (filled circles) and receiver-related (open circles) speech levels from Figs. 1 and 3 for the sitting (left panel) and standing (right panel) conditions, averaged across male and female participants. Receiver-related speech level approximations from Eq. (8) of Weisser and Buchholz (2019) are shown by the dashed-dotted lines, which were here derived using the talker distance approximations from Fig. 2.

circles) for the sitting (left panel) and standing (right panel) conditions. The SNRs were derived by subtracting the noise levels shown on the abscissa from the receiver-related speech levels shown in Fig. 4. The SNR is well-approximated by a second order polynomial function as shown in Fig. 5 and further described in Sec. IV C. For comparison, the SNR approximations from Weisser and Buchholz (2019) are shown by the dashed-dotted lines. Similar to the speech levels in Fig. 4, these approximations were derived by replacing the conversational distance approximations (as a function of noise level) given in Weisser and Buchholz (2019) by the broken-stick approximations shown by the dashed lines in Fig. 2. Details are given in Sec. IV C.

E. Participant relationships

With respect to the present study, one inclusion criterion was that the participants were familiar and comfortable with each other to avoid extended silence between interlocutors or awkward conversations and behavior, in particular at the beginning of the experiment. Nevertheless, the recruited participants still differed in their level of intimacy with each other, as six pairs were couples and the rest were either friends (14 pairs) or siblings (2 pairs). To evaluate the effect of relationship status (as a measure of intimacy level) on conversational distance, the data from Fig. 2 were regrouped into couples and non-couples and plotted in Fig. 6. There was a significant interaction between relationship status and both noise level [$F(6,524) = 3.5$, $p < 0.01$] and condition [$F(1,524) = 97$, $p < 0.01$]. A *post hoc* analysis showed that couples in the standing condition were 0.28 m closer than non-couples, according to their estimated marginal means. In addition, there was, on average, a 0.29 m difference between non-couples in the sitting compared to standing condition. There were no differences between the pairs in the sitting condition. Couples also stood closer together as the noise levels increased. On average, conversational distance between couples was significantly reduced in the three loudest environments compared to the non-couples. Interestingly, the source-related speech levels for the couples were consistently lower during their conversations than for the non-couples, with an average decrease in 5.5 dB in the standing condition and 4.2 dB in the sitting condition. However, applying separate two-sample t-tests did not find any difference in source-related speech level between the two groups in any of the noise environments ($p > 0.05$). Thus, taking into account the closer distance of the couples removed any remaining differences between groups with basically identical distance-adjusted speech levels.

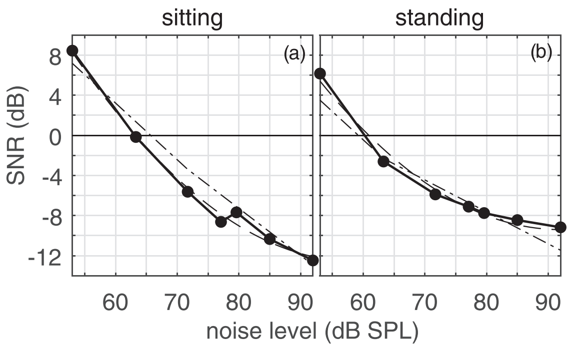


FIG. 5. Average SNR at the receiver location as a function of noise level of the different acoustic environments for the sitting (left panel) and standing (right panel) conditions. Second-order polynomial fits are shown by the dashed lines. The SNR approximations from Eq. (8) of Weisser and Buchholz (2019) are shown by the dashed-dotted lines, which were derived here using the conversational distance approximations from Fig. 2 (see the text).

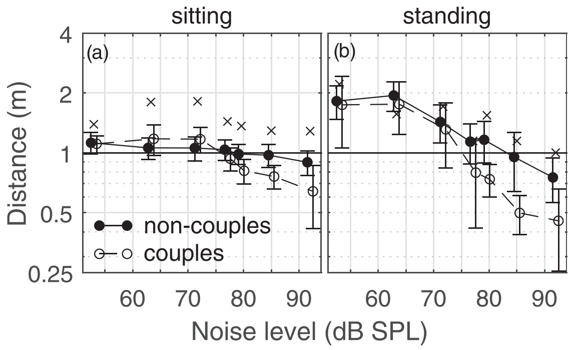


FIG. 6. Mean and confidence intervals of the conversational distance from Fig. 2, grouped into the five couples (open circles) and the 16 non-couples (filled circles). The data for the couple that was considered an outlier is indicated by the crosses.

IV. DISCUSSION

A. Conversational distance

The observed conversational distances in the present study are significantly larger than the ones reported in the only other study that simultaneously measured distances and speech levels between talkers (Pearsons *et al.*, 1977). As Fig. 2 shows, the distances may be up to 1 m larger in the standing condition and 0.6 m in the sitting condition, with few participant pairs showing similarly close distances to Pearsons *et al.* at the lowest noise level. The measured distances are also significantly larger than average distances found in interpersonal field studies that were done without controlling the noise or speech levels, even if measured outdoors or in quiet (Remland *et al.*, 1991, 1995).

The differences with Pearsons *et al.* (1977) may have several causes. First, the two studies employed fundamentally different methods to collect the data—the present study was highly controlled and set in a university laboratory, whereas Pearsons *et al.* collected their data *in situ*, as a field study. This may have influenced the participants' capability to relax and assume their 'normal' position, which would have been observable only in candid situations. Second, because of the first cause, interlocutors tended to stand in front of one another, whereas normally they may prefer to speak at an angle, but at closer distances than measured here (Remland *et al.*, 1995). Frontal talking tends to lead to direct eye contact, which is unconsciously avoided and therefore traded off with distance, so at closer distances, talkers have reduced eye contact than from farther distances (Argyle and Dean, 1965). Third, the present study applied a highly accurate motion capture system to track the distance of the talkers with a well-defined reference point halfway between the listener's ears, whereas neither the measurement procedure nor the reference points are reported in Pearson *et al.* Fourth, the Pearsons *et al.* data were not controlled for sitting and standing; although considering that their change (or slope) of the conversational distance with noise level is in between the change in the sitting and standing condition of the present study may suggest that some mixture of talker standing and sitting took place. Fifth, the space constraints may have been different in Pearsons *et al.*, where, for instance, participants may have sat directly next to each other in the airplanes or trains. Finally, differences in the participants' hearing status, age, or familiarity with each other may have also contributed to the differences in distance observed across the two studies, but not much is known about the participants in Pearsons *et al.*

Most studies on vocal effort focused on adaptation due to the Lombard effect, whereby talkers tend to raise their voice to compensate against the noisy environment (Lane and Tranel, 1971). The adaptive effect of distance on the acoustic conditions has usually been neglected, except for few studies that were either not fully disclosing their methods (Pearsons *et al.*, 1977), or physically limited by the seating geometry (Hadley *et al.*, 2019; Weisser and Buchholz, 2019). Nevertheless, distance has been widely explored as a variable in conversations, where it is shown to depend on a wide array of social and environmental parameters, such as the level of familiarity between the interlocutors, their gender, and their native culture (e.g., Hall, 1990; Sommer, 2002). Some of these parameters may have contributed to the large spread of the individual data that was observed in how participants adapted their conversational distance (Fig. 2) and speech level (Fig. 1). Despite the apparent spread of data, a very clear trend was observed of participant pairs decreasing their conversational distance once the ambient noise level rises above 60–65 dB SPL. The compensatory effect of this adaptation is very limited in the sitting condition (0.12 dB/dB of noise increase), which is in line with findings by Hadley *et al.* (2019). The effect is considerably more pronounced in the standing condition (0.3 dB/dB of noise increase) and presents itself as a viable adaptation strategy during conversation to optimize the SNR either by the talker, the listener, or both.

The slope of the interpersonal adaptation as a function of noise must be understood as a complex weighting of several parameters that are not strictly acoustical. Having a conversation of any kind is a social interaction, which may have implications for the interlocutors that are not explicitly reflected in the success or failure of speech communication. For example, close proximity may imply intimacy, which is not necessarily reciprocal in both interlocutors. This can be normalized depending on the situation (e.g., it may be acceptable in a loud party, despite a lack of intimacy). Another social aspect is eavesdropping-aversion, which some people may be oversensitive to. This would imply speech level adaptation that is not always mutually accepted or understood, and may not follow the mean patterns observed in the present study.

Even though the difference in the level of intimacy between groups may explain some of the variance seen in the individual conversational distances (Fig. 2), the small number of couples (six) does not allow for any strong conclusions. Moreover, the comparison may have been confounded by other factors, such as the sex of the participants. Whereby the couples were all mixed sex, the majority of the non-couples were female, i.e.,12 pairs were all female, two pairs were all male, and two pairs were mixed sex (Table I). However, applying separate two-sample t-tests to compare the distances between the eight mixed-sex pairs with the 12 all-female pairs did not reveal any significant differences ($p > 0.05$); reconfirming that intimacy may be the factor affecting conversational distance. Future studies need to assess larger groups with more equal sizes and include participants that are unfamiliar with each other to extend the range of intimacy level between participants.

B. Speech level and SNR

The present study applied the methods used by Weisser and Buchholz (2019), but some important differences exist that can serve to cross-validate the methods and results used in both. The main similarities were: common noise environments from the ARTE database that were used as binaural stimuli for the two interlocutors, the use of natural conversations between two participants, a sitting condition, and the use of the same open headphone model that is nearly acoustically transparent to external sound. The main differences of the present methods were: the use of motion capture and wireless audio equipment that relaxed the constraints on the relative positions of the talkers, inclusion of a standing condition, requirement of participant familiarity (within pairs), no speech-eliciting task, testing in a (normal) reverberant room and not in an anechoic chamber, and completely independent distance data that can be analyzed without reference to Pearsons *et al.* (1977).

Despite these differences, the measured speech levels reported in Weisser and Buchholz (2019) are very similar to the ones reported here. However, a direct comparison between the studies is difficult due to the different conversational distances at which the speech levels were measured. Whereas Weisser and Buchholz (2019) considered two fixed conversational distances of 0.5 m and 1 m, the participants in the present study were free to choose their own distances. The only comparable test condition is the Food Court (79.6 dB SPL) at which the talkers in the present study had an average distance of about 1 m, in both the sitting and standing conditions. The corresponding speech level, averaged across male and female talkers, in Weisser and Buchholz is 74.0 dB SPL, and very similar to the present study with 72.4 dB SPL in the sitting and 71.6 dB SPL in the standing condition. A more extensive comparison can be made by applying the distance-adjusted SNR equation from Weisser and Buchholz [see their Eq. (8)], but replacing their conversational distance approximation based on the Pearsons *et al.* (1977) data by the broken stick approximation described in Sec. III B (see Fig. 2 and Sec. IV C). The resulting SNR predictions are shown by the dashed-dotted lines in Fig. 5 and the corresponding speech level predictions are shown in Fig. 4. For the noise level range that is common to both studies, i.e., from 53.0 to 79.6 dB SPL, the predictions are in good agreement with the distance-adjusted (receiver-related) speech levels and SNRs measured in the sitting and standing condition.

This convergence of results is a strong indicator that the methods applied in the two studies were valid even though they may still be biased by similar constraints—in particular, the use of headphone reproduction versus the free-field conditions experienced in the real world. It is also telling that observations that were obtained *in situ* in five acoustic environment types, in an anechoic chamber (Weisser and Buchholz, 2019), and in a normal reverberant room (Sec. II E), yield highly similar results as far as the eventual speech levels and SNR between talkers is concerned.

The average source-related speech levels as a function of noise level at a reference distance of 1 m range is 62.7–77.8 dB SPL in the sitting condition and is 64.2–78.8 dB SPL in the standing condition, with some participants even reaching speech levels of more than 85 dB SPL in the loudest environment (Fig. 1). This speech level range is similar to the range reported by Pearsons *et al.* (1977) with 58–89 dBA for male and 55–82 dBA for female talkers, even though their reference distance is unclear. When compared to the vocal effort level categories reported in ANSI (1997), the present speech levels range from "normal" vocal effort (62.4 dB SPL) to "loud" (74.9 dB SPL) or even "shout" (82.3 dB SPL). Hence, talkers produced very high vocal effort levels to enable conversation in the loudest noise conditions, which they surprisingly did not compensate further by getting closer to each other. In real life, they may not be able (or willing) to sustain such high speech levels for a prolonged time, which may bring them eventually closer to each other, or might even make them lose interest in the conversation altogether. In this regard, the rather short noise durations of two minutes may have not been enough to elicit such behavior, and may provide another reason for why the distances observed in the present study were significantly larger than those observed by other studies (e.g., Pearsons *et al.*, 1977; Remland *et al.*, 1991, 1995).

The average slope of the source-related speech level as a function of noise level (Fig. 1) is about 0.46 dB/dB for noise levels above about 64 dB SPL. This slope increases to 0.68 dB/dB when the distance-adjusted speech level at the receiver location (Fig. 3) is considered. However, the slope decreases at lower noise levels due to the speech level converging towards its level in quiet at around 60–64 dB SPL. Similar values were found by Weisser and Buchholz (2019), who reported average slopes of around 0.43 to 0.46 dB/dB when the distance between talkers was fixed, and a slope of 0.64 dB/dB when the distance-adjusted speech level at the receiver location was considered. The slope for the distance-adjusted speech level is slightly higher than for the one reported by Pearsons *et al.* (1977) of 0.6 dB/dB and the one reported by Wu *et al.* (2018) of about 0.54 dB/dB for noise levels above 59.3 dB SPL. However, Weisser and Buchholz (2019) pointed out that due to the tendency of the slope to increase with increasing noise level, the concept of a single slope may be questionable and its value arguable. Nevertheless, there is a clear tendency for the slopes to become steeper when the effect of distance on the receiver-related speech levels is taken into account, highlighting again that the listeners adapt their distance to improve the SNR experienced by themselves and/or by the interlocutor.

Finally, it should be mentioned that the present analysis took only unweighted broadband levels into account both for the speech and noise signals, and hence for the SNR. However, the observed increase in speech (or vocal effort) level due to the Lombard effect is in general accompanied by a range of other acoustic as well as linguistic and communicative changes (e.g., Beechey *et al.*, 2018) that, among other aspects, significantly affect the long-term spectrum of speech (e.g., ANSI, 1997; Pearsons *et al.*, 1977; Weisser and Buchholz, 2019). Similarly, the applied noise stimuli did not only vary in their broadband level but also in their spectral, temporal, and spatial behavior (Weisser and Buchholz, 2019). As a consequence, the SNR in the different test conditions varied as a function of frequency (and time) and not only in its broadband value. Furthermore, only free-field levels were considered here, which do not take into account the acoustic shadow effect of the human ears, head, and torso, which can significantly change the SNR received by the ears (e.g., Weisser and Buchholz, 2019). This may be even further weighted by the most relevant frequency range for speech intelligibility of around 500–4000 Hz (ANSI, 1997). Even though all these important aspects apply equally to the present signals and SNRs, they have been extensively described in the literature and are therefore not discussed here any further.

C. Distance-adjusted modeling

Weisser and Buchholz (2019) derived approximations for realistic SNRs at the receiver location as a function of noise level (or acoustic environment) that took into account the effect of interlocutor distance. Thereby, they distinguished between two different cases, which depended on whether the interlocutors were fixed at a given distance, but still adapted their speech level to that distance, or they were free to move and adjust their distance as desired. Based on their own speech level data that were measured at two separate distances of 0.5 and 1 m, they approximated the distance-adjusted SNR averaged across male and female talkers by

$$\mathrm{SNR}_D = -16.54 \log D - 0.56 L_n + 37.91, \tag{3}$$

with D the interlocutor distance in meters, L_n the noise level in dB SPL, and log with base 10. They then applied data from Pearsons *et al.* (1977) to derive an approximation of the conversational distance as a function of noise level. Given that their distances were much shorter than the distances measured in the present study (Sec. IV A), their conversational distance approximation is replaced here by the one described in Sec. III B and given by Eq. (1). Considering also that the maximum observed distance is limited (Fig. 2), the conversational distance for the sitting and standing condition can be written as

$$\log D_{sit} = \min(-0.0058 L_n + 0.465 \log D_{sit,\max}), \tag{4}$$

$$\log D_{stand} = \min(-0.0149 L_n + 1.214 \log D_{stand,\max}), \tag{5}$$

with the maximum distance $D_{sit,\max} = 1.12$ m and $D_{stand,\max} = 1.87$ m. The corresponding SNR approximations for the case that the talkers can freely choose their distance are shown in Fig. 5 by the dashed-dotted lines. The corresponding speech levels can be derived by $L_s = \mathrm{SNR}_D + L_n$ and are shown in Fig. 4 (dashed-dotted lines). The approximations fit the measured data reasonably well, but deviate from the data at soft and loud levels and predict slightly too high values in the sitting condition.

The accuracy of the approximation can be improved by directly fitting a second-order polynomial function to the distance-adjusted SNR data, as given by

$$\mathrm{SNR}_D = k \cdot (L_n - L_0)^2 + \mathrm{SNR}_0, \tag{6}$$

with L_0 the noise level at the inflection point, SNR_0 the SNR at the inflection point, and k a 'compression' parameter. In the sitting condition $k = 0.001$, $L_0 = 99$ dB, and $\mathrm{SNR}_0 = -12.6$ dB. In the standing condition $k = 0.0097$, $L_0 = 92$ dB, and $\mathrm{SNR}_0 = -9.4$ dB. The corresponding approximations are shown in Fig. 5 by the dashed lines and are in excellent agreement with the measured SNRs.

Given that the distance-adjusted SNR approximations of Eq. (3) are in reasonably good agreement with the data measured in the present study, Eq. (3) remains a viable approximation for the distance-adjusted SNR (and speech level) for the case that the conversational distance is fixed at a given distance D.

D. Limitations

Coming up with a method that was relatively ecologically valid, but also provided the means to retain controlled acoustic and geometric data were not trivial. Therefore, there are a number of features in the eventual methods of the present study that may limit the generalizability of the results and could have exaggerated the individual spread in the data (Fig. 1) or added an overall bias.

First, the test took place in a normal reverberant room. This required some processing of the recorded speech signals to minimize the effect of room reverberation on the estimated speech levels (Sec. II E), but a certain error may have still remained. Furthermore, the reflected energy may have interacted with the vocal effort decision of the participants. However, this effect is known to be relatively small (Pelegrín-García *et al.*, 2011), and is not reflected in the prediction of the present study SNRs, compared to the previous study of Weisser and Buchholz (2019), which took place in an anechoic chamber.

Second, the participants may have experienced some kind of cognitive dissonance because of the discrepant information from the variable room acoustics presented on headphones along with the unvarying visual and acoustic information about the actual room. Proxemic studies suggest that the room dimensions (e.g., ceiling height, and indoor vs outdoor environments) affect the baseline conversational distance in which people feel comfortable (Cochran *et al.*, 1984; Cochran and Urbanczyk, 1982), so this discrepant information may have inflated the spread of individual data.

Third, several aspects of the conversation task that were unnatural. The initial standing condition of every conversation round had the two participants facing each other frontally. This resulted in participants mainly moving on a single axis to optimize their distance, whereas in reality movements within conversations may vary in more than one axis, and interlocutors often stand at an angle to one another (Remland *et al.*, 1991, 1995). This may lead to the additional psychological constraints of eye-contact avoidance, which is normally balanced with the conversational distance (Argyle and Dean, 1965). Another factor is that the participants knew that they are being observed by an unfamiliar experimenter, so they could have not been completely relaxed, as if they were truly alone, despite their familiarity.

Fourth, the amount of hardware—the headphones, multiple motion trackers, microphone transmitter, and headphones receiver—somewhat limited the participants' movement and may have made them appear unnatural to one another, at least in the beginning of the test. Despite the possible physical limitation associated with the hardware, the experimenters did not note any unnaturalness in the participants' conversations—perhaps because of their high degree of familiarity. Unfortunately, using this kind of natural stimuli can only be done using headphones if the two subjects are to be exposed to identically controlled noise levels, because simulating them in a 3D array can be controlled only to one (static) sweet spot in space, or one person, and otherwise requires a vastly increased technical effort (e.g., realizing multiple sweet spots that are continuously adjusted to the subject locations). Another limitation of the headphone reproduction was that the playback of the noise environments did not take the listeners' head movements into account. This would have interfered with the SNR improvements at the listeners' ears that may have been achieved in the real environments by rotating the head toward or away from their communication partner. This limitation may have influenced the observed conversation distances as well as speech levels.

Fifth, there was a potential confound between noise level and type of noise. The seven noise environments did not only differ in their level, but also in their temporal, spectral, and spatial properties, as well as by the presence of distracting talkers (Weisser *et al.*, 2019a). This may have influenced the measured speech levels and conversation distances as a function of noise level. However, their monotonic increase when plotted against noise level that is observed here and in (Weisser and Buchholz, 2019) suggests that this may have been a secondary effect. Moreover, in the applied noise environments, the overall level mainly increased because of an increasing number of sources and their increased individual level (e.g., people talking louder due to the Lombard effect), which seems to be a property of many realistic environments. Hence, it may be argued, that this confound may be a relevant aspect of real-world environments that should be included when considering realistic speech levels and conversation distances in noise. Nevertheless, future studies should investigate this potential confound and include noise conditions (such as a diffuse babble noise) where the level is the only parameter.

E. Outlook

Speech level has been an important parameter in the design of hearing instruments and fitting rationales for the hearing impaired, as speech audibility and more generally, communication, are considered the most critical aspects of hearing that have to be restored with amplification. With the lack of detailed data, it has been the practice to assume a range of conversational speech levels around 60–65 dB SPL (e.g., Keidser *et al.*, 2011), which mainly reflects speech levels in quiet. With the advent of sophisticated signal processing features in hearing instruments, the question of optimal SNR for the listener and the device has also become central to their design (Naylor, 2016; Smeds *et al.*, 2015).

The data in Weisser and Buchholz (2019) and the present study provide initial estimates for the range of acoustic adaptation that is possible to obtain by physically changing the conversational distance, but also the limitations of such a strategy in mitigating adverse SNR situations. It was found in both experiments that in the noisiest environments, the SNR can become negative, despite the distance adaptation. In the present study, the SNR measured was even worse given that the participants usually did not get as close to each other, as was prescribed in the close (0.5 m) condition of Weisser and Buchholz (2019). In a recent study, negative SNRs were estimated for normal-hearing speech tests done at 1 m distance in real restaurant and bar environments, when the noise level exceeded 75 dBA (see Fig. 4 of Brungart *et al.*, 2020). Negative SNRs were uncommon in the sample of hearing-impaired daily environments selected by Smeds *et al.* (2015) (see their Fig. 5), and were primarily measured for the loudest environments of 70 dB SPL or higher, with occasional negative SNRs estimated below. It is possible, however, that the prevalence of real-world negative SNRs is not represented well by any of these studies because of selection bias of the noise stimuli. In the case of the present study and in Brungart *et al.* (2020), the purpose was to increase the variance of the noise. The opposite may have been true in Smeds *et al.* (2015), as listeners may have stuck to familiar situations, which ensured optimal voice level and conversational distance, with fewer loud environments. Therefore, there is room for future studies that specifically sample hearing-impaired populations to obtain more real-world SNRs.

This study incorporated two technologies that were used in a novel way: wireless motion capture and audio equipment, itself containing multiple transmitter-receiver sets. Still, the richness and quality of the data obtained from these methods are unprecedented and we believe that they can be harnessed in other research niches that can benefit from bridging the gap between field- and laboratory-based measurements. Future studies will still have to investigate how far the applied outcomes generalize to behaviors observed in the real world.

V. CONCLUSIONS

The study simultaneously measured conversational speech levels and conversational distances in a broad range of realistic acoustic environments in young normal-hearing participants. Wireless technology was used to minimize physical constraints on the participants' movement while providing highly controlled results. The following conclusions can be drawn:

(1) Interlocutors adjust their speech levels primarily based on the noise level in the environment, in accord with the Lombard effect (Lombard, 1911), independently of the standing or sitting condition.
(2) Interlocutors reduced their conversational distance beyond a relatively constant distance when the noise level surpassed 64 dB SPL in the sitting condition and 72 dB SPL in the standing condition.
(3) Free-field SNRs at the receiver location were negative in noise above 60–63 dB SPL. Even though this is in broad agreement with existing field studies, negative

SNRs may reflect environments that are either avoided or rarely visited by hearing-impaired listeners.

(4) The distance adaptation of the talkers had a significant effect on the speech levels at the receiver location. This effect was rather modest in the sitting condition, with an average SNR increase of about 0.12 dB per 1 dB increase of noise level, but increased substantially in the standing condition, where there were no physical constraints to get closer, with an average increase of 0.3 dB per 1 dB increase in noise level.

ACKNOWLEDGMENTS

The research was supported by the Hearing Corporate Research Centre, an initiative of the Australian Government. Many thanks to Adrienne Crafford for assistance with the pilot testing, to Greg Stewart from the National Acoustic Laboratory for help with printing the 3D caps, and to Tim Beechey for his statistical advice. The behavioral measurements in this study complied with the ethics guidelines of the Macquarie University Research Ethics Committee, approval number 5201929708215, Project ID 2970.

[1] See additional details about these scenes in the ARTE database, https://zenodo.org/record/2655947.

[2] See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0004774 for the LME statistical model tables of the main and interaction effects.

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

Argyle, M., and Dean, J. (**1965**). "Eye-contact, distance and affiliation," Sociometry **28**, 289–304.

Bavelas, J. B., Coates, L., and Johnson, T. (**2002**). "Listener responses as a collaborative process: The role of gaze," J. Commun. **52**(3), 566–580.

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

Bell, P. A., Kline, L. M., and Barnard, W. A. (**1988**). "Friendship and freedom of movement as moderators of sex differences in interpersonal distancing," J. Social Psychol. **128**(3), 305–310.

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.

Brungart, D. S., Barrett, M. E., Cohen, J. I., Fodor, C., Yancey, C. M., and Gordon-Salant, S. (**2020**). "Objective assessment of speech intelligibility in crowded public spaces," Ear Hear. **41**, 68S–78S.

Cochran, C., Hale, W. D., and Hissam, C. P. (**1984**). "Personal space requirements in indoor versus outdoor locations," J. Psychol. **117**(1), 121–123.

Cochran, C., and Urbanczyk, S. (**1982**). "The effect of availability of vertical space on personal space," J. Psychol. **111**(1), 137–140.

Cooke, M., King, S., Garnier, M., and Aubanel, V. (**2014**). "The listening talker: A review of human and algorithmic context-induced modifications of speech," Comput. Speech Lang. **28**(2), 543–571.

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.

Hadley, L. R., Owen, B. W., and Whitmer, W. M. (**2019**). "Speech, movement, and gaze behaviours during dyadic conversation in noise," Sci. Rep. **9**, 10451.

Hall, E. T. (**1990**). *The Hidden Dimension* (Anchor Books Editions, New York).

Harrigan, J. (**2005**). *Proxemics, Kinesics, and Gaze* (Oxford University Press, Oxford, UK), pp. 137–198.

Hayduk, L. A. (**1983**). "Personal space: Where we now stand," Psychol. Bull. **94**(2), 293.

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

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, S. (**2011**). "The NAL-NL2 prescription procedure," Audiol. Res. **1**(1), 88.

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.

Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (**2018**). "emmeans: Estimated marginal means, aka least-squares means," R package documentation, https://cran.r-project.org/web/packages/emmeans/emmeans.pdf (accessed 15 October 2020).

Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of the elevation of the voice") (2006). Ann. des Maladiers de l'Oreille, du Larynx, du Nez, et du Pharynx **37**, 101–119, translated by P. H. Mason.

Matsumoto, D. E., Hwang, H. C., and Frank, M. G. (**2016**). *The Body: Postures, Gait, Proxemics, and Haptics* (American Psychological Association, Washington, DC), pp. 387–400.

Miles, K. M., Keidser, G., Freeston, K., Beechey, T., Best, V., and Buchholz, J. M. (**2020**). "Development of the everyday conversational sentences in noise test," J. Acoust. Soc. Am. **147**(3), 1562–1576.

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.

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

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., and Van Willigen, B. (**2017**). "Package 'nlme.' linear and nonlinear mixed effects models, version, 3(1)," Technical Report.

Remland, M. S., Jones, T. S., and Brinkman, H. (**1991**). "Proxemic and haptic behavior in three european countries," J. Nonverbal Behav. **15**(4), 215–232.

Remland, M. S., Jones, T. S., and Brinkman, H. (**1995**). "Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age," J. Social Psychol. **135**(3), 281–297.

Shuter, R. (**1977**). "A field study of nonverbal communication in germany, italy, and the united states," Commun. Monogr. **44**(4), 298–305.

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. Personality Social Psychol. **42**(1), 66.

Vertegaal, R., Slagter, R., Van der Veer, G., and Nijholt, A. (**2001**). "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, May 8–13, Seattle, WA, pp. 301–308.

Weisser, A., and Buchholz, J. (**2019**). "Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions," J. Acoust. Soc. Am. **145**(1), 349–360.

Weisser, A., Buchholz, J. M., and Keidser, G. (**2019a**). "Complex acoustic environments: Review, framework, and subjective model," Trends Hear. **23**, 1–20.

Weisser, A., Buchholz, J., Oreinos, C., Galloway, J., Beechey, T., Davila, J., and Keidser, G. (**2019b**). "The ambisonic recordings of typical environments (ARTE) database," Acta Acust. united Ac. **105**(4), 695–713.

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.