5-2022

# Intelligibility of medically related sentences in quiet, speech-shaped noise, and hospital noise

Tessa Bent

Melissa Baese-Berk

Erica E. Ryherd

Sidney Perry

JASA ARTICLE

# Intelligibility of medically related sentences in quiet, speech-shaped noise, and hospital noise

Tessa Bent,[1,a)] Melissa Baese-Berk,[2] Erica Ryherd,[3] and Sydney Perry[1]
[1]*Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, Indiana 47408, USA*
[2]*Department of Linguistics, University of Oregon, Eugene, Oregon 97403-1290, USA*
[3]*Durham School of Architectural Engineering and Construction, University of Nebraska–Lincoln, Omaha, Nebraska 68182-0816, USA*

**ABSTRACT:**
Noise in healthcare settings, such as hospitals, often exceeds levels recommended by health organizations. Although researchers and medical professionals have raised concerns about the effect of these noise levels on spoken communication, objective measures of behavioral intelligibility in hospital noise are lacking. Further, no studies of intelligibility in hospital noise used medically relevant terminology, which may differentially impact intelligibility compared to standard terminology in speech perception research and is essential for ensuring ecological validity. Here, intelligibility was measured using online testing for 69 young adult listeners in three listening conditions (i.e., quiet, speech-shaped noise, and hospital noise: 23 listeners per condition) for four sentence types. Three sentence types included medical terminology with varied lexical frequency and familiarity characteristics. A final sentence set included non-medically related sentences. Results showed that intelligibility was negatively impacted by both noise types with no significant difference between the hospital and speech-shaped noise. Medically related sentences were not less intelligible overall, but word recognition accuracy was significantly positively correlated with both lexical frequency and familiarity. These results support the need for continued research on how noise levels in healthcare settings in concert with less familiar medical terminology impact communications and ultimately health outcomes.
 https://doi.org/10.1121/10.0011394

(Received 18 October 2021; revised 11 April 2022; accepted 29 April 2022; published online 26 May 2022)
[Editor: Deniz Baskent] Pages: 3496–3508

## I. INTRODUCTION

Noise in healthcare settings often exceeds recommended sound levels set by the World Health Organization (WHO) *Guidelines for Community Noise* (Berglund *et al.*, 1999). Hospitals, including adult and pediatric intensive care units, operating rooms, patient rooms, hallways, nursing stations, and chemotherapy clinics, have all been found to have noise levels that exceed the WHO guidelines (Gladd and Saunders, 2011; Darbyshire and Young, 2013; Tainter *et al.*, 2016; Busch-Vishniac, 2019). Additionally, the noise levels in hospitals appear to be getting worse, even with new construction (Busch-Vishniac *et al.*, 2005; Ryherd *et al.*, 2011). The noise in hospitals is a product of many sources, including building noise (e.g., heating, ventilating, and air conditioning), environmental sounds (e.g., service carts, ice machines, doors closing), equipment sounds (e.g., alarms, ventilators, phones ringing), and human sounds (e.g., conversations, coughing, activity noise) (MacKenzie and Galbrun, 2007; Ryherd *et al.*, 2011).

There are numerous potential deleterious impacts of these high hospital noise levels. One of the most researched areas is sleep disturbance. The results of these investigations generally show negative impacts of noise on both quantity and quality of sleep (Buxton *et al.*, 2012; Hsu *et al.*, 2012), although other analyses suggest that noise may not be the leading cause of lowered sleep quality (Basner and McGuire, 2018). A range of other possible negative physiological effects have been suggested (e.g., slowed wound healing, greater need for pain management, cardiovascular changes, extended hospital stays, and increases in rehospitalization) (Ryherd *et al.*, 2011; Hsu *et al.*, 2012). There also is evidence that noise in healthcare settings is a significant issue for staff, including increases in stress, tension headaches, annoyance/irritation, and concentration difficulties (Morrison *et al.*, 2003; Ryherd *et al.*, 2008; Ryherd *et al.*, 2012).

One impact of hospital noise that has received relatively less attention is the potential detriment to effective oral communication, even though successful communication between healthcare providers and patients is paramount to high quality healthcare delivery. Some research has employed surveys regarding the subjective views of patient-provider communication broadly and suggests that "communication" was the top concern for patients (O'Hara *et al.*, 2018); three types of communication issues were identified: staff to patient (most frequently cited by patients), staff to staff, and patient to staff. In a recent systematic review of patient-provider communication (Shukla *et al.*, 2019), most studies reported communication problems for older patients who were hard

[a)]Electronic mail: tbent@indiana.edu

 

of hearing, including communication difficulties because of noise levels (Mulley and Ng, 1995). One study analyzing interactions between nurses and patients during admission interviews noted that miscommunication occurred in 50% of the conversations due to acoustic reasons, primarily background noise in combination with patient hearing loss (VanCott, 1993). Background noise for individuals with hearing loss has also been associated with decreased performance on cognitive tests, potentially leading to higher rates of medical misdiagnosis (Dupuis *et al.*, 2016; Jorgensen *et al.*, 2016), higher rates of hospital readmission (Chang *et al.*, 2018), and the potential for misunderstanding of discharge and medication instructions (Mormer *et al.*, 2017).

Although frequently cited as a potential problem and as "an area well deserving of much greater attention" [Busch-Vishniac (2019), p. 7800], there is a dearth of objective speech intelligibility data in hospital environments. The noise in hospitals has the potential to be highly detrimental to speech intelligibility not only due to the noise levels, as discussed above, but also resulting from the noise types present in the environment. The noise in hospitals is likely to contain a mixture of relatively steady state noise, fluctuating noise, and interfering speech sounds (MacKenzie and Galbrun 2007; Okcu *et al.*, 2012). These sound sources all can be detrimental to speech intelligibility, and their impact may differ across populations. For example, listeners who are hard of hearing are more impacted by noise (Souza and Turner, 1994). Fluctuations in background noise amplitude can benefit listeners with normal hearing because listeners can extract information from the sections with more favorable signal-to-noise ratios (SNRs) and put that information together to identify words and sentences (Miller and Licklider, 1950; Cooke, 2006). These fluctuations, however, are less beneficial for listeners who are hard of hearing (Festen and Plomp, 1990) and for older listeners with normal hearing compared to younger listeners with normal hearing (Dubno *et al.*, 2002). There may also be detriments of informational masking within hospitals when the speech of others in the environment competes with the target speaker. Speech intelligibility can decrease when the interfering sounds include meaningful speech (Sperry *et al.*, 1997; Summers and Molis, 2004), as may occur when other patients, healthcare providers, or visitors are speaking near the patient's location.

Only two published articles have evaluated speech intelligibility in hospital noise specifically. Ryherd *et al.* (2013) assessed the noise levels in five different hospitals, including a range of unit types and locations. They employed the speech intelligibility index (SII), which is a calculation of an acoustical prediction shown to correlate with speech intelligibility in multiple adverse conditions (ANSI, 1997). Their analysis demonstrated that none of the units had noise levels that would allow for "good" intelligibility, but rather levels leading to predicted "marginal" to "poor" intelligibility with normal voice levels. The SII measurements were also correlated with staff perception of communication challenges.

The only study investigating speech intelligibility in hospital noise with an objective behavioral measure was Pope *et al.* (2013). In their study, hospitalized patients were presented with sentences in quiet, mixed with white noise, or mixed with hospital noise (with or without voices) using SNRs of +1.5, –3.5, and −8.5 dB. The sentences included contexts where the final word was highly predictable (e.g., "For dessert, I'd like some apple pie") and contexts where the final word was low in predictability (e.g., "Mom talked about the pie"). Participants were less accurate at identifying the final keywords when the sentences were presented in hospital noise with voices than without voices and were better in all noise conditions for the high vs low context sentences. Recall patterns mirrored the word identification patterns. The authors note that the results suggest that there may not be issues with patients understanding routine, high context speech (e.g., "It's time to take your meds"), but they raise serious concerns about patient understanding and recall for low context information (e.g., discharge instructions that are unfamiliar and thus lower in supportive context). One other study tested speech intelligibility in dental noise at an SNR of +5 dB (Mendel *et al.*, 2008) with listeners with and without hearing loss using sentences that were not medically related. The results showed significant but very small reductions in performance with dental noise and for listeners with hearing loss. However, performance in all conditions was near ceiling (i.e., above 90% correct).

Beyond the soundscape in healthcare settings, the linguistic characteristics of the messages to be conveyed must be considered. Information about medical diagnoses, screening procedures, treatment plans, prognosis, and etiology of illness may include many terms that are wholly unfamiliar to patients or very infrequently encountered. There is a large body of speech perception research that highlights how linguistic factors can cause serious challenges for successful communication. In addition to the impact of sentence context, as shown in Pope *et al.* (2013) and many prior studies (e.g., Miller and Isard, 1963; Duffy and Giolas, 1974; Kalikow *et al.*, 1977), lexical characteristics substantially impact a listener's ability to understand spoken language in noise. For example, higher frequency words are more accurately identified compared to lower frequency words (Howes, 1957; Brysbaert *et al.*, 2018). The level of subjective familiarity for lexical items also influences how quickly and accurately words are perceived (Epstein *et al.*, 1968; Connine *et al.*, 1990) with more familiar words showing higher accuracy and faster identification than less familiar words. How familiarity and frequency interact during word recognition has not been established. Intelligibility studies that include words varying in frequency typically control for familiarity by only including words with high familiarity (e.g., Bradlow and Pisoni, 1999) or by varying word familiarity without manipulating word frequency (e.g., Epstein *et al.*, 1968; Sakamoto *et al.*, 2004). However, there is evidence that both word familiarity and frequency are important during lexical access, with some evidence for divergent effects in lexical decision and naming tasks (Colombo *et al.*, 2006; Connine *et al.*, 1990). Including both word frequency and familiarity will also be important for future studies with other populations, such as older listeners, since word frequency and familiarity effects do not show the same patterns across the adult lifespan (Newman and German, 2005).

These classic findings from the speech perception literature should be considered for evaluating the likely success of patient-clinician communication in healthcare settings, especially because interactions in these settings are likely to include less frequent and less familiar terminology than in other everyday interactions. The concerns about patient-clinician communication are heightened by the low rates of health literacy in the United States, including estimates that there are $90 \times 10^6$ Americans with low health literacy, which has been called a crisis and one of the causes of health disparities (Carmona, 2006). Health literacy has been defined as "the capacity to obtain, process and understand basic health information and services needed to make appropriate health care decisions" [Ratzan *et al.* (2000), p. 1]. Low health literacy may therefore lead to miscommunication. The use of unfamiliar medical jargon (Hadlow and Pitts, 1991) and semi-technical terms (Smith and Davis, 2018) by healthcare professionals may lead to misunderstandings and miscommunication.

To avoid these miscommunications, healthcare providers are advised to avoid some terminology or explain potentially unfamiliar terms (Chapple *et al.*, 1997). However, there is evidence that doctors do not explain medical terms when they are first introduced to patients (Koch-Weser *et al.*, 2009), nor do they consistently use "everyday language" with patients (Bourhis *et al.*, 1989; Denton *et al.*, 2020). Furthermore, various healthcare providers, including physicians, nurses, and pharmacists, overestimate the level of comprehension by patients (Byrne and Edeani, 1984; Yoshida and Yoshida, 2014), and patients may also overestimate their own understanding of medical terminology (Chapman *et al.*, 2003; Neill *et al.*, 2020). In one discourse analysis of admissions interviews, 25% of conversations between nurses and patients had miscommunications due to unfamiliar lexical items (VanCott, 1993). Finally, healthcare professionals and patients may have different understandings of the same medical terms or use different terms for the same concepts (Lerner *et al.*, 2000; Zeng *et al.*, 2001). However, healthcare providers need to tailor their communication to the level of knowledge of their patients, with some analyses showing much higher levels of knowledge for medical terminology for patients with chronic conditions who access health information online (Fage-Butler and Jensen, 2016).

Although medical terminology has the potential for leading to miscommunications particularly in noisy places like hospitals, it is also possible that the presence of hospital noise itself could benefit the perception of medically related terminology. There is strong evidence that listeners use social information about the speaker when interpreting speech. For example, the presentation of a visual image of a speaker can influence how accented they sound or how intelligible they are for listeners (Babel and Russell, 2015; McGowan, 2015). Less is known about how the environmental context impacts speech perception (Hay, 2018), but there is some evidence that individuals store knowledge about the physical environment in which linguistic encounters take place and use that information to guide their speech production and perception (Hay *et al.*, 2017). Therefore, the presence of hospital noise may set the expectation for words that are medically related and would raise their activation levels, conferring a perceptual benefit to listeners for these congruent noise/sentence pairings compared to incongruent pairings (e.g., hospital noise and an unrelated sentence, such as "Mom talked about the pie").

The current study was designed to build on the sparse findings regarding objective measures of speech intelligibility in hospital noise while simultaneously integrating speech materials that address another possible concern with effective communication across patients and healthcare providers: medical terminology composed of less familiar and less frequent words. The intelligibility of sentences with and without medically related words in three listening conditions (quiet, speech-shaped noise, and hospital noise) was evaluated using an online testing methodology. The medically related sentences contained keywords of three types: high frequency/high familiarity, low frequency/high familiarity, and low frequency/low familiarity. The primary research questions and associated hypotheses were as follows:

(1) How does intelligibility differ when sentences are presented in hospital noise compared to speech-shaped noise?
Hypothesis: Intelligibility will be significantly worse in hospital noise compared to speech-shaped noise.
(2) Does sentence intelligibility differ for sentences with medically related words compared to sentences that are traditionally employed in speech perception research (i.e., not including medically related words)?
Hypothesis: Intelligibility will be similar across medical and non-medical sentences for sentences where the keywords have similar lexical characteristics.
(3) How do word familiarity and word frequency characteristics with medically related sentences impact speech intelligibility? How do these lexical characteristics interact with noise type?
Hypothesis: Lower frequency and lower familiarity words will be more difficult to identify in noise. There will be an interaction between lexical characteristics and noise such that worst performance will be seen with low frequency, low familiarity words in hospital noise.

## II. METHOD

### A. Participants

Listeners included 69 monolingual American English-speaking adults. All participants indicated no current speech, language, or hearing impairments. Most participants had studied another language, but none of the included participants indicated fluency in a language other than English. Additional demographic information about the participants can be found in Table I.

TABLE I. Background information about the participants in the three listening conditions.

| | Listening condition | | |
|---|---|---|---|
| | Hospital noise | Quiet | Speech-shaped noise |
| Age (years) | 25.0 (19–35) | 26.2 (19–34) | 26.3 (19–34) |
| Gender identity | 10 female<br>12 male<br>1 gender fluid | 9 female<br>13 male<br>1 nonbinary | 11 female<br>12 male |
| Race | 16 white; 4 Black or African American; 3 biracial | 15 white; 2 Black or African American; 1 Asian American; 5 bi- or multi-racial | 18 white; 1 Black or African American; 2 biracial; 1 prefer not to say |
| Ethnicity | 4 Hispanic or Latinx<br>19 not Hispanic or Latinx | 1 Hispanic or Latinx<br>21 not Hispanic or Latinx<br>1 prefer not to say | 5 Hispanic or Latinx<br>18 not Hispanic or Latinx |
| Interaction with medical professionals | 10 minimal<br>11 low moderate<br>1 moderate<br>1 frequent | 8 minimal<br>10 low moderate<br>2 moderate<br>3 frequent | 8 minimal<br>10 low moderate<br>3 moderate<br>2 frequent |
| Highest level of education | 0 some high school<br>6 high school diploma<br>8 some college<br>2 associate degree<br>6 bachelor's degree<br>1 master's degree | 0 some high school<br>1 high school diploma<br>9 some college<br>2 associate degree<br>8 bachelor's degree<br>3 master's degree | 1 some high school<br>6 high school diploma<br>5 some college<br>0 associate degree<br>8 bachelor's degree<br>3 master's degree |
| Medically related course work | 6 no courses<br>7 one course<br>6 two courses<br>4 three or more courses | 3 no courses<br>6 one course<br>7 two courses<br>7 three or more courses | 6 no courses<br>5 one course<br>8 two courses<br>4 three or more courses |

An additional 14 participants were tested, but their data were not included due to low effort responses (i.e., intelligibility scores more than three standard deviations below the mean for their condition; $n = 5$), daily interactions with medical professionals ($n = 6$), bilingual language background ($n = 2$), or a reported noise level above 7 in their environment from a scale of 1 = very quiet to 10 = very loud ($n = 1$).

### B. Stimuli

The sentence stimuli were taken from a recently developed corpus of medically related sentences (Perry *et al.*, 2021). For additional detail about the development of the corpus, see Perry *et al.* (2021). Recordings of the full sentence set along with the familiarity, frequency, and predictability data described below can be accessed through our Open Science Framework project repository (Bent *et al.*, 2022). The corpus includes 160 sentences with 40 sentences for each of four types. Each sentence included three keywords and was between four and nine words in length (average = 6.9). Three of the sentence types are medically related sentences with different keyword frequency and familiarity profiles. These medically related sentences were divided into high familiarity/high frequency, high familiarity/low frequency, and low familiarity/low frequency (Table II). Word frequency categorization was determined by the data in the SUBTLEX-US database, a 51-million-word corpus generated from American subtitles (Brysbaert and New, 2009; van Heuven *et al.*, 2014). Using the Zipf scale, which ranges from 1 to 7, keywords in the lower frequency sentences had average Zipf scores between 1.7 and 3.99, and keywords in the higher frequency sentences had average scores from 4.3 to 5.3. Familiarity scores for the keywords in the medically related sentences were gathered from 41 monolingual American English listeners, who rated the words on a scale of 1 to 7, where 1 = "You have never seen or heard this word before," and 7 = "You recognize the word and are confident you know the meaning of the word" (Perry *et al.*, 2021). Sentences in the low familiarity category had average ratings of 3.6–5.5, and those with high familiarity words had average ratings of 6.7–7.0. The final sentence type includes sentences adapted from the Hearing In Noise Test (HINT), a standard sentence set used to evaluate intelligibility in noise (Nilsson *et al.*, 1994). Frequency scores for the keywords in the HINT sentences were also taken from the SUBTLEX-US corpus, and familiarity scores were taken from Nusbaum *et al.* (1984). The keywords in the HINT sentences are non-medically related words with high frequency and high familiarity, as is the case with nearly all sentence stimuli commonly used in research and clinical applications. Future research could develop non-medically related sentences with lower frequency and familiarity characteristics to compare to the medically related sentences employed here.

The predictability of the keywords for all sentences (standard and medically related) was determined through a Cloze sentence test procedure. In this task, 48 monolingual

https://doi.org/10.1121/10.0011394



TABLE II. Average and ranges by sentence for familiarity, frequency, and predictability across the four sentence types. Sentence type was determined by the average rating for the three keywords in the sentence. Familiarity scores are on a scale from 1 to 7 where 1 = "You have never seen or heard this word before," and 7 = "You recognize the word and are confident you know the meaning of the word." Frequency scores are Zipf scores, which range from 1 (less frequent) to 7 (most frequent). Predictability scores indicate percent of participants who correctly guessed the keyword in a Cloze testing procedure.

| Sentence type | Familiarity | Frequency | Predictability |
|---|---|---|---|
| Medical high familiarity/high frequency | 6.98 (6.90–7.00) | 4.79 (4.31–5.31) | 7% (0%–19%) |
| Medical high familiarity/low frequency | 6.90 (6.69–7.00) | 3.37 (2.74–3.99) | 2% (0%–24%) |
| Medical low familiarity/low frequency | 4.79 (3.55–5.52) | 2.36 (1.65–3.35) | 2% (0%–15%) |
| Standardized non-medical | 6.98 (6.98–7.00) | 4.88 (3.82–5.74) | 17% (0%–63%) |

American English listeners were presented orthographically with a version of the sentence with one word missing. Their task was to complete the sentence with the first word that came to mind. Predictability scores were then determined by the percent of participants who accurately guessed the target keyword.

The 160 sentences were recorded by four monolingual American English speakers, including two cisgender male and two cisgender female speakers. All speakers grew up in the Midwest, including in Indiana, Illinois, and Ohio, and identified their dialect as either Midland (n = 2) or North ($n = 2$). Three of the speakers identified as white and one as biracial (Black and white). All identified as not Hispanic or Latinx. Speakers were between the ages of 18 and 29 (average = 22.5 years). The average fundamental frequencies for the two female speakers were 201 and 213 Hz; for the two male speakers, they were 119 and 139 Hz. Recordings were made in a sound-attenuated booth using a Marantz (Kanagawa, Japan) PDM670 digital recorder and a Shure (Niles, IL) Dynamic WH20XLR headset microphone with a sampling rate of 22 050 Hz. Sentences were equated for root mean square (rms) amplitude. Throughout the recording sessions, a researcher (S.P.) monitored the speakers' pronunciations and asked them to repeat sentences with incorrect pronunciations by supplying the correct pronunciation. The sentences were read in the same order by all participants. The speakers were told to speak conversationally like they were speaking to a patient. They were not told that the sentences would later be mixed with noise.

For the experiment, the sentences were presented in quiet, speech-shaped noise, or hospital noise. The speech-shaped noise was created by taking the long-term average spectrum of a set of sentences and using it to filter a white noise. The hospital noise was synthesized in previous work (Messingher, 2013). It included noise sources typical of healthcare facilities, such as conversation, medical alarms, footfall, and ventilation, and was calibrated to match spectral content, fluctuations over time, and other acoustic characteristics typical of noise measured in real-world hospital settings. The hospital noise was compressed using Audacity to remove extreme peaks in the signal prior to mixing with the sentences. Long-term average spectral analysis of the two noise conditions was conducted in SigView version 5.3.2, normalized to 60 dBA to allow relative comparisons as shown in Fig. 1. Despite differing source content and fluctuations, the average spectral content of the two signals was not substantially different across the majority of spectral bands. More specifically, 68% of the one-third octave bands plotted in Fig. 1 had a difference of 3 dB or less between the two signals. Larger differences of up to 7.8 dB were observed at certain bands and in the higher frequencies (4000 Hz and above).

For the two noise-added conditions, each sentence was mixed with a unique, randomly selected portion of one of the noise files, either speech-shaped noise or hospital noise, at a SNR of −1 dB that was 1 s longer than the sentence. The SNR was selected based on pilot testing that suggested that performance in the noise conditions would be at neither ceiling nor floor. Furthermore, this SNR is in a range of what would be expected in hospital settings based on prior measurements and typical loudness for conversational speech (Pope *et al.*, 2013).

### C. Procedure

Participants were recruited through Prolific (https://www.prolific.co/). If they met the study criteria, the study would appear as one in which they were eligible to participate. The inclusion criteria on Prolific required participants to be monolingual English speakers, be American citizens, be currently living in the United States, be between the ages of 18 and 35 years, and have no hearing difficulties. Participants were also excluded if they had participated in one of the norming tasks for the development of the corpus (e.g., word familiarity rating task). After opting to start the study, participants were directed to a Qualtrics survey. This survey included the study information sheet, on which they could indicate their consent to participate by clicking a button. They then completed a headphone screening to ensure they were wearing headphones rather than listening over loudspeakers (Woods *et al.*, 2017). In this screening, participants are presented with a series of three pure tones and are asked to select the quietest tone. One of the tones is 180° out of phase across the stereo channels and therefore should result in phase cancellation. The task is designed to be easy if the participants are wearing headphones but difficult to perform using a speaker. They were given three opportunities to complete the screening. If they failed all three attempts, they could not continue with the study. Seven participants failed the headphone screening. If they passed the headphone screening, they then continued to a demographic

https://doi.org/10.1121/10.0011394

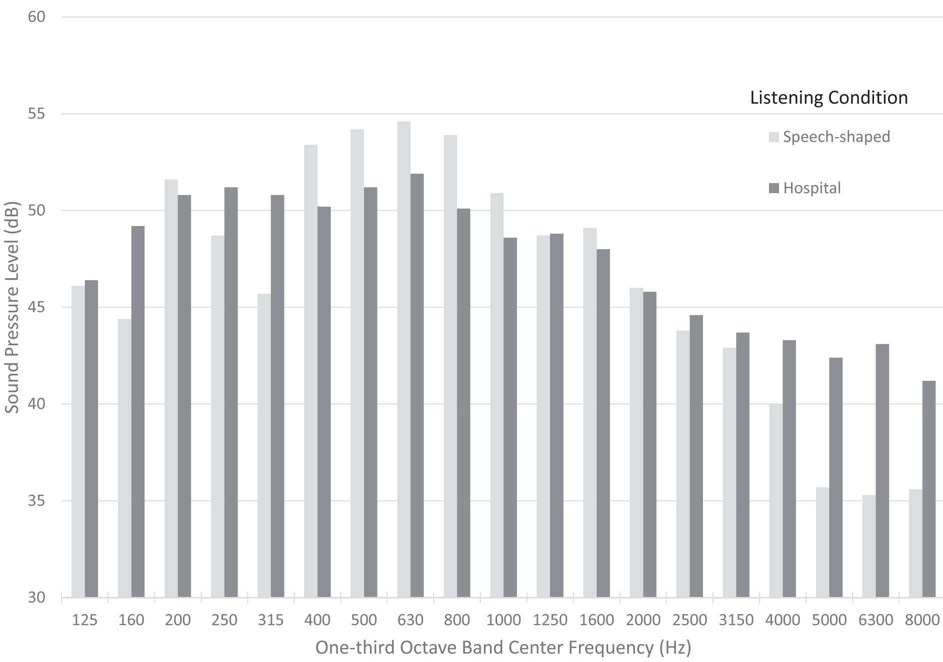


FIG. 1. One-third octave band sound pressure levels of the speech-shaped and hospital noise conditions, normalized to 60 dBA.

and language background questionnaire. There were also questions about their interactions with medical professionals and their current environment. After completing the survey, they were redirected to Pavlovia (https://pavlovia.org/), the online platform for PsychoPy (Peirce *et al.*, 2019), for the intelligibility task. Participants used their own computers and headphones to complete all tasks.

Participants were assigned to one of three listening conditions: quiet, speech-shaped noise, or hospital noise, with 23 participants in each condition. For all listening conditions, participants heard all 160 sentences. Forty of the sentences were produced by each of the four talkers with assignment between talker and sentence counter-balanced across listeners within a listening condition. Sentences were presented in random order for each listener.

Before completing the experimental trials, participants were presented with four practice trials. The practice trials were presented in the same listening condition as the experimental trials and were produced by a different speaker than those in the experimental trials. After each sentence was presented, participants typed in what they heard (i.e., the entire sentence). They could hear each sentence only once, were not given any feedback, and could take as long as needed to enter their responses. They were given two breaks of 10 s minimum provided after every 54 trials. In addition, there was a trial counter in the top left of the screen, so they could keep track of their progress throughout the experiment.

## III. ANALYSIS

Responses were scored for keyword accuracy. Prior to scoring, two researchers (author S.P. and one other) completed a spellcheck on the participants' responses. Obvious typos (e.g., "diffrent" changed to "different"), homophones (e.g., "pair" for "pear," "pane" for "pain"), and words in which there was an extra space or a missing space (e.g., "pace maker" for "pacemaker," "redspots" for the phrase "red spots") were counted as correct. Because the words in the low familiarity sentence set were, by design, much less familiar to listeners, the spellings were accepted if they could be pronounced as the target word (e.g., "perpherated" for "perforated," "disfunction" for "dysfunction," "silia" for "cilia"). After each of the researchers completed the spellcheck, the two separate spellchecked versions were scored for accuracy. The two sets of scores were then compared. The two researchers discussed discrepancies and came to a final decision with the assistance of a third rater (author T.B.). Each keyword was then given a 0 (incorrect) or 1 (correct). Words with added or deleted morphemes were counted as incorrect.

These word recognition accuracy scores, with each keyword entered separately into the model, were then analyzed using generalized linear mixed effects models with a logit link function to account for the binomial outcome measures (i.e., correct or incorrect). Fixed effects for the analysis included listening condition (quiet, speech-shaped noise, or hospital noise), sentence type (medical or standard), frequency of target word, and familiarity of target word.[1] The listening condition variable was Helmert coded to compare the quiet listening condition to the average of the two noise conditions and then to compare each noise condition to the other. The final model also included interactions between both listening condition comparisons and frequency and both listening condition comparisons and familiarity. The model also included random intercepts for items, for speakers, and for participants in addition to random slopes for participants by sentence type.

## IV. RESULTS

As can be seen in Fig. 2, listeners were overall much more accurate in the quiet condition than in either of the noise conditions. The results of the mixed effects model mirror this observation. The comparison between the quiet

https://doi.org/10.1121/10.0011394

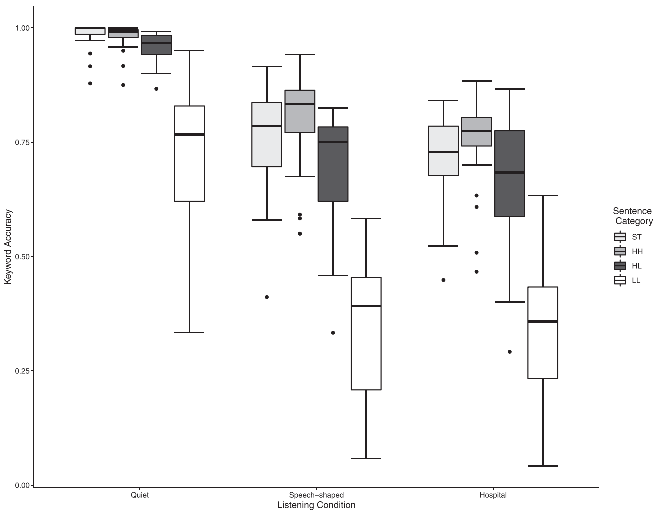


FIG. 2. Performance on the intelligibility task across listening conditions (23 listeners in each of the three conditions) and sentence categories. The lines in the center of the boxes indicate the median value (i.e., 50th percentile); the boxes indicate the interquartile range (i.e., 25th to 75th percentile); whiskers indicate the largest and smallest values within 1.5 times the interquartile range; dots indicate values greater than 1.5 times the interquartile range. Medical sentences with low familiarity and low frequency words (LL) are in white, medical sentences with high familiarity and low frequency (HL) words are in dark gray, medical sentences with high familiarity and high frequency (HH) words are in medium gray, and standard, non-medical sentences (ST) are in light gray.

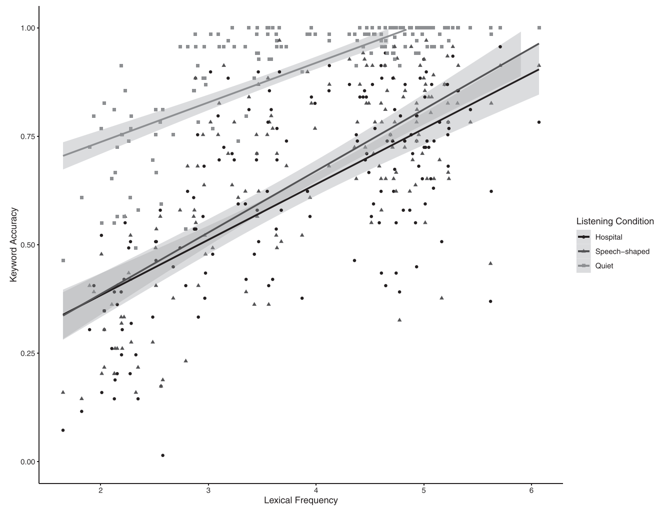


FIG. 3. Keyword accuracy by lexical frequency (Zipf transformed) of target words. The best fitting linear regression line is plotted, with a shaded section for 95% confidence intervals. Average accuracy scores across items and subjects for words presented in hospital noise are represented in black circles and a black regression line; words presented in speech-shaped noise are represented in dark gray triangles and a dark gray regression line; words presented in quiet are represented in light gray squares and a light gray regression line.

listening condition and the two noise conditions significantly contributed to the model fit ($\beta = 0.759$, $z = 2.622$, $p = 0.009$). However, the two noise conditions were quite similar to one another ($\beta = 0.075$, $z = 0.262$, $p = 0.794$).

There were also clear differences among sentence types, many of which were captured in the fine-grained frequency and familiarity results described below. However, it did not appear that there were major differences between medical and non-medical sentences. Indeed, the standard (ST in Fig. 2) sentences and medically related high frequency/high familiarity sentences demonstrated very similar performance. Instead of differences among sentence types (i.e., standard vs medical sentences), it appeared that the primary differences were driven by lexical frequency and familiarity. The model results support this conclusion. Sentence type (i.e., medical vs standard) did not significantly impact model fit ($\beta = 0.186$, $z = 1.263$, $p = 0.207$).

Both lexical familiarity and frequency improved model fit ($\beta = 0.455$, $z = 13.358$, $p < 0.001$ and $\beta = 0.110$, $z = 2.680$, $p = 0.007$, respectively). Figures 3 and 4 show accuracy as a function of lexical frequency (Fig. 3) and lexical familiarity (Fig. 4) across the three listening conditions. Listeners correctly identified more frequent words compared to less frequent words and identified more familiar words compared to less familiar words.

The two interactions between word familiarity and listening condition were not significant ($\beta = 0.033$, $z = 0.670$, $p = 0.503$ for the interaction between quiet vs noise and word familiarity and $\beta = -0.050$, $z = -1.214$, $p = 0.225$ for the interaction between hospital noise and speech-shaped noise and word familiarity). This finding reflects the observation in Fig. 4 that the slopes of the regression lines are approximately similar for all three listening conditions. However, the two interactions between word frequency and listening condition were significant. The interaction between the comparison of the quiet condition and the two noise conditions with frequency was significant ($\beta = 0.506$, $z = 7.736$, $p < 0.001$), suggesting that the noise conditions were more impacted by frequency than the quiet condition.

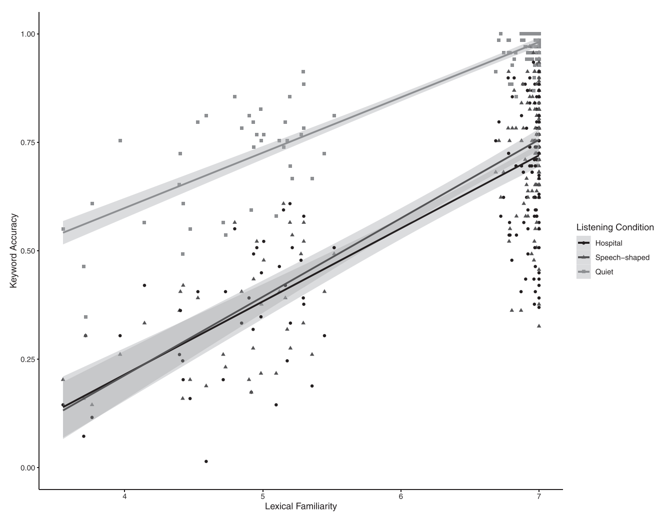


FIG. 4. Keyword accuracy by lexical familiarity of target words. The best fitting linear regression line is plotted as well, with a shaded section for 95% confidence intervals. Average accuracy scores for words across items and subjects presented in hospital noise are represented in black circles and a black regression line; words presented in speech-shaped noise are represented in dark gray triangles and a dark gray regression line; words presented in quiet are represented in light gray squares and a light gray regression line.

Similarly, the interaction between the comparison of the hospital noise vs speech-shaped noise and frequency was also significant ($\beta = 0.109$, $z = 2.767$, $p = 0.006$), which reflects the steeper slope of the regression line in Fig. 3 for the speech-shaped noise as compared to the hospital noise.

## V. DISCUSSION

Our understanding of the impact of hospital noise on speech intelligibility has been limited by the types of stimuli employed in prior studies, which did not include speech stimuli with medically related terminology (Pope *et al.*, 2013), or by the methods used, which did not include objective measures of how much listeners understand from speech stimuli (Shukla *et al.*, 2019). This study, therefore, provides novel data regarding the intelligibility of sentences with medical terminology in hospital noise.

Sentence intelligibility did not differ significantly overall between the speech-shaped noise and the hospital noise conditions, contrary to our hypothesis, although both noise conditions led to significantly lower speech intelligibility than the quiet condition. The reason that these two maskers resulted in similar performance is deserving of further investigation as these results may not hold for hospital noise maskers with different characteristics.

There are some aspects of the hospital noise that were expected to be more detrimental to speech intelligibility than the speech-shaped noise. Specifically, the hospital noise included voices, which were not included in the speech-shaped noise. However, the voices in the hospital noise used in this study were not very distinct. Therefore, hospital noise maskers in which the voices are more intelligible may result in greater informational masking (Summers and Molis, 2004) and thus more challenges in identifying the target words (Pope *et al.*, 2013).

The amplitude fluctuations in the hospital noise and the variety of realistic sound sources (e.g., speech, medical alarms, occupant-generated sounds) also have the potential for greater distraction and less adaptation to the noise. However, we compressed the hospital noise to remove extreme peaks to keep the SNRs throughout each sentence more consistent. This compression may have made the hospital noise and the speech-shaped noise conditions more similar than found in real-world healthcare settings. Conditions testing hospital noises that contain large amplitude variations with attention capturing sounds (e.g., loud alarms) may lead to different results than found in this study. Although these amplitude variations could have detrimental impacts on speech intelligibility, amplitude dips can also benefit listeners' abilities to understand the speech by providing access to parts of the speech signal that can be used to piece together messages (Miller and Licklider, 1950; Cooke, 2006), although these benefits are smaller for some listener populations (Dubno *et al.*, 2002; Festen and Plomp, 1990). Future work should investigate how the potential benefits and detriments from the amplitude fluctuation in hospital noise ultimately impact speech intelligibility across listeners differing in age and hearing ability.

The analysis of the long-term average spectra for the two maskers also showed similarities, including characteristics typical of effective maskers, such as energy at all frequencies with significant energy in the speech frequency range (Bradley, 2003). Therefore, the extent of energetic masking for the two noise types may have been very similar. The long-term average spectra were not specifically controlled for in this study because we wanted to test two different types of noise that inherently had some spectral differences; however, the minimal long-term average differences actually observed may in part explain the lack of significant results. It would be interesting to further explore how larger, controlled differences in spectral content impact results. In our experiment, we also only employed one SNR. It would be useful to investigate how noise levels interact with hospital noise type to impact word recognition performance.

There may have been a congruency benefit for the medically related sentences when presented in the hospital noise that could have offset the aspects of the hospital noise that tend to be more challenging for word identification. Additional research is needed to further explore how noise type by lexical item congruency may impact speech perception. Although usage-based accounts, such as exemplar models (Foulkes and Docherty, 2006), would predict benefits for these congruent situations, there is currently little evidence about how the physical environment impacts speech perception and production (Hay, 2018).

One of the goals of the current study was to investigate sentence intelligibility with sentences that included medical terminology compared to sentences that are more traditionally used in speech perception research. For ecological validity, it is important to assess how hospital noise will impact word recognition for the types of words encountered in medical settings. Overall, there were not significant differences between the standard speech perception sentences and sentences with medical terminology. In particular, the medical sentences with high frequency and high familiarity words and the standardized non-medical sentences showed very similar performance across listening conditions (i.e., quiet, speech-shaped noise, hospital noise). Therefore, the mere presence of medically related words does not cause listeners difficulties with understanding speech. However, there were substantial decrements for speech intelligibility for words that were lower in familiarity and frequency. The decrements for decreasing levels of word frequency were steeper in the noise-added conditions than in quiet.

These results further highlight the calls for healthcare providers to use "everyday language" with patients. The avoidance of medical jargon can help decrease miscommunications generally, but it also will reduce the possibility that patients will not recognize the words, particularly in adverse listening conditions. Furthermore, the sentences, even in the low familiarity/low frequency condition, were designed to have words that were low in familiarity rather than completely unknown words. On the familiarity scale, which ranges from 1 to 7, we excluded words with average familiarity rates below 2.5. The ratings 1, 2, and 3 specifically refer to words with the following characteristics: (1) "You have never seen or heard this word before"; (2) "You think that you might have seen or heard this word before"; and (3) "You are pretty sure that you have seen or heard the word before, but you are not positive." It is likely that medical terminology encountered in healthcare settings would include words that would fall into these unknown and lesser-known categories. Thus, the impact of hospital noise on health-related communication may in fact be overestimated in our study.

Although patients now have access to greater resources for gaining knowledge of medical terminology through the internet (Fage-Butler and Jensen, 2016), it is still likely that many medical encounters will include terms that are unfamiliar to patients or for which patients have only vague meanings. For example, during consultations for initial diagnoses or in the prescription of new medications, patients may be presented with completely unfamiliar terms. Not only should healthcare providers ensure that patients understand the meanings of these words, but they should also be cognizant that the listening environment itself could lead to word recognition difficulties for patients.

In addition to the characteristics of words and sentences, listening situations with higher cognitive load can cause significant detriments to speech intelligibility and recall (Hunter and Pisoni, 2018). Considering that many patients seeking treatment in hospitals are likely to be in high stress situations and therefore under high cognitive load, their ability to understand and remember information presented by healthcare providers may be negatively affected. Indeed, a study by Dunn *et al.* (1993) of cancer consultations found that 1–3 weeks after the consultation, patients only remembered about 25% of the information presented and only approximately 40% of points deemed most important by the doctor. It will be essential in future research to include measures of recall in addition to measures of initial word recognition accuracy (e.g., Pope *et al.*, 2013).

This study was also limited by the methodology employed, specifically only requiring participants to type in what they heard (i.e., word recognition) without a measure of comprehension. Therefore, we only measured whether a listener could correctly interpret the acoustic signal and immediately write down the words. Word recognition is an essential first step for accurately perceiving and remembering medical information, but the full understanding of the words is obviously essential. Future studies should consider different methodologies that incorporate comprehension measures.

The lexical items included in either word recognition or language comprehension tasks could also be categorized in other ways beyond their frequency and familiarity characteristics as done here. For example, Fage-Butler and Jensen (2016) provide several categories of medical terms. Their category of dictionary-defined medical terms maps best to the words used in this study, but other categories would also be important to investigate, such as medical initialisms and medication brand names. They also point out the use of colloquial technical terms, such as "endo" for endocrinologist, which do not appear in medical dictionaries but may be used by patients and should be considered within condition specific contexts. Other authors divide medical terminology in different ways, such as the seven categories of medical terminology of Koch-Weser *et al.* (2009), which included some overlap with Fage-Butler and Jensen (2016) (e.g., drug names) but also has categories for names of medical specialties, symptoms, and diseases and disease processes. Investigating listeners' perception and understanding of these different semantic categories of medical terminology is another important avenue for future work.

The participants in the current study were all monolingual English-speaking young adults from the United States with self-reported normal hearing who had relatively high levels of education. Changing any of these listener characteristics could lead to decrements in performance and potentially different impacts of both the noise type and lexical characteristics. It will be essential to test participants from different age groups. Since the incidence of health problems increases with age, testing older adults will be an important next step. The incidence of hearing loss also substantially increases with age, with only approximately 3% of individuals in the United States having a hearing loss of some type in their 20s but close to 90% for individuals 80 years of age and older (Lin *et al.*, 2011). The increase in hearing loss, which makes understanding speech in noise more challenging, suggests that many older adults will have more difficulty with the initial steps toward comprehending and remembering orally presented health information. Without success in this initial step of spoken communication, the chance that patients will fully understand their diagnoses and be able to comply with discharge instructions substantially decreases.

Not only were the participants in our study younger than many hospitalized patients, but our participants had on average higher levels of education. Specifically, ∼80% of the listeners in this study had some education beyond high school compared to 62% of the United States population (U.S. Census Bureau, 2020). The impact of word frequency and familiarity may have different effects on participant samples with different education levels. For example, the familiarity ratings of the words were gathered from participants with similar educational profiles to those in the intelligibility tests. As familiarity ratings are subjective, it is certainly possible that some words included here as having moderate to high levels of familiarity may be less familiar to other populations. Future studies should gather more information from the participants about their experiences with medical terminology beyond what was asked here. For example, questions on the background questionnaire could be added about their exposure to medical terminology from a wider range of sources (e.g., media and friends or family with health conditions). An objective vocabulary assessment could be incorporated into the protocol as well to test their health literacy and knowledge of medical terminology. Relatedly, questions could be incorporated about the listeners' exposure to hospital noise, as listeners who have spent more time in hospitals specifically may show long-term adaptation, which confers an advantage for the perception of speech in hospital noise.

The age of the listeners would also influence how the lexical characteristics impact word recognition. Vocabulary knowledge increases with age (Verhaeghen, 2003), but the ability to extract meaning from context for unfamiliar words decreases with age (McGinnis and Zelinski, 2000). Thus, there may be trade-offs for older adults when trying to understand medical information in noisy environments with some disadvantages (e.g., higher prevalence of hearing loss) but some advantages (e.g., greater vocabulary knowledge). Furthermore, middle-aged and older adults are better at estimating their understanding of words, an important consideration since patients may overestimate their knowledge of medical terminology, potentially leading to miscommunications with healthcare providers (Chapman *et al.*, 2003; Neill *et al.*, 2020).

One final population that will be essential to incorporate in future studies is non-native speakers of English. There should be consideration both for how unfamiliar accents of healthcare providers and patients may make speech communication more challenging in noise (Munro, 1998; Adank *et al.*, 2009) and for how a patient's knowledge of the language may impact their understanding of medical terminology (Dahm, 2012).

For future studies that incorporate listener populations different than those tested here, assessments of hearing and language abilities may need to be incorporated into the protocol. Some of these assessments are quite amenable to online testing. For example, testing older adults may require more knowledge about their hearing abilities. A questionnaire about hearing abilities (e.g., the Speech, Spatial, and Qualities of Hearing Scale) or a hearing screening (e.g., the digits-in-noise hearing screening) could be employed in online testing protocols since these assessments can be administered via phone or computer (Folmer *et al.*, 2017; Moulin *et al.*, 2019). However, these tests do not provide the same detail as an audiogram. If a full audiogram is desired, then in-person testing may be required. Similarly, for non-monolingual listeners, more detail about their language learning history and proficiency levels should be incorporated into the experimental protocol. Although there are limitations for online testing compared to in-person protocols and online testing may not be appropriate for all populations, there is strong evidence of replication when data collection is conducted online vs in person (e.g., Crump *et al.*, 2013), including studies showing replication of intelligibility differences across listening conditions (e.g., Cooke and Garcia Lecumberri, 2021; Slote and Strand 2016).

The materials in this study were presented in an audio-only modality with stimuli that were recorded in a nearly ideal recording environment (i.e., sound attenuating booth). The inclusion of only audio information may have underestimated participants' abilities to understand the speech, since it is well established that the provision of visual information from speakers improves intelligibility in adverse listening conditions (Sumby and Pollack, 1954). However, the COVID pandemic has highlighted another potential communication barrier: the need for healthcare providers in hospitals and long-term care facilities to don personal protective equipment, including face masks. Even before COVID, face masks were commonplace in hospitals and long-term care facilities to help prevent the spread of infection. Thus, even in real-world hospital settings, patients may not receive the benefits of visual information from the speaker. Furthermore, recent studies show that face masks significantly deteriorate directional output of speech (Pörschmann *et al.*, 2020), which could make it more difficult to separate the target speech from competing sounds in the hospital environment, and are especially detrimental for conveying higher frequency information, which is essential for many consonant sounds (Corey *et al.*, 2020). Many recent studies have shown detrimental impacts of face masks on speech intelligibility, recall, and listening effort especially in noisy conditions or for listeners with hearing loss (Homans and Vroegop, 2021; Rahne *et al.*, 2021; Smiljanic *et al.*, 2021; Toscano and Toscano, 2021; Truong *et al.*, 2021; Yi *et al.*, 2021), although the impact of speaking style can ameliorate some of the negative impacts (Cohn *et al.*, 2021; Smiljanic *et al.*, 2021). It will be essential to expand on these recent studies by investigating the impacts of face masks on speech intelligibility and recall under hospital noise conditions with materials relevant for healthcare settings.

## VI. CONCLUSION

The data presented here demonstrate that the noisy conditions found in many hospitals coupled with the use of less familiar medical terminology have the potential to lead to miscommunications between healthcare providers and patients. The method used in our study only required word recognition by participants; therefore, the measurement of language comprehension and recall will be essential future directions to understand how hospital noise and lexical characteristics may influence the transmission of essential medical information, such as discharge instructions and diagnoses. Finally, the assessment of listeners from different populations (e.g., older listeners, non-native speakers) and the use of varied stimulus conditions (e.g., audio-visual, masked speech, hospital noise with different levels or characteristics) are important future directions.

## ACKNOWLEDGMENTS

We would like to thank Mary Schiff for her assistance with data scoring. We also acknowledge support for this research from a Collaborative Research Award from the Indiana University Institute for Advanced Study and from an Opportunity Award from the James S. McDonnell Foundation (https://doi.org/10.37717/2021-3028).

[1]All words in the medically related sentences have familiarity ratings, as we collected those ratings ourselves (Perry *et al.*, 2021). However, some words in the standardized non-medical sentence condition are missing familiarity information because familiarity ratings for these words came from the Hoosier Mental Lexicon database, and not all words were included in that database; however, we do not believe this is problematic for our analyses as mixed models are robust to missing data. Further, it is clear that familiarity and frequency are correlated measures. For example, it is not possible to have low-familiarity but high-frequency words. However, despite the two measures being correlated in our current data set, they do not contribute to a multicollinearity issue in our analyses. Calculation of variation inflation factor (VIF) to detect for multicollinearity suggests that all VIF values are below 2.5. The consensus among statisticians is that values >10 are indicative of collinearity problems, although even if one chooses a more conservative threshold (e.g., 5), our factors still do not pose a collinearity challenge.


Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (**2009**). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," J. Exp. Psych. Hum. Percept. Perform. **35**, 520–529.

ANSI (**1997**). ANSI S3.5-1997, "Methods for the calculation of the speech intelligibility index" (American National Standards Institute, New York).

Babel, M., and Russell, J. (**2015**). "Expectations and speech intelligibility," J. Acoust. Soc. Am. **137**(5), 2823–2833.

Basner, M., and McGuire, S. (**2018**). "WHO environmental noise guidelines for the European region: A systematic review on environmental noise and effects on sleep," Int. J. Environ. Res. Public Health **15**, 519.

Bent, T., Baese-Berk, M., Ryherd, E., and Perry, S. (**2022**). "Open Science Framework project repository," https://osf.io/rvz2f/ (Last viewed April 11, 2022).

Berglund, B., Lindvall, T., Schwela, D. H., and Goh, K. T. (**1999**). "Guidelines for community noise," Technical Report (World Health Organization, Geneva).

Bourhis, R. Y., Roth, S., and MacQueen, G. (**1989**). "Communication in the hospital setting: A survey of medical and everyday language use amongst patients, nurses and doctors," Soc. Sci. Med. **28**, 339–346.

Bradley, J. S. (**2003**). "The acoustical design of conventional open plan offices," Can. Acoust. **31**, 23–31.

Bradlow, A. R., and Pisoni, D. B. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**(4), 2074–2085.

Brysbaert, M., Mandera, P., and Keuleers, E. (**2018**). "The word frequency effect in word processing: An updated review," Curr. Dir. Psychol. Sci. **27**, 45–50.

Brysbaert, M., and New, B. (**2009**). "Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," Behav. Res. Methods **41**, 977–990.

Busch-Vishniac, I. (**2019**). *Next Steps in Hospital Noise Research* (Universitätsbibliothek der RWTH Aachen, Aachen, Germany).

Busch-Vishniac, I. J., West, J. E., Barnhill, C., Hunter, T., Orellana, D., and Chivukula, R. (**2005**). "Noise levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645.

Buxton, O. M., Ellenbogen, J. M., Wang, W., Carballeira, A., O'Connor, S., Cooper, D., Gordhandas, A. J., McKinney, S. M., and Solet, J. M. (**2012**). "Sleep disruption due to hospital noises: A prospective evaluation," Ann. Intern. Med. **157**, 170–179.

Byrne, T. J., and Edeani, D. (**1984**). "Knowledge of medical terminology among hospital patients," Nurs. Res. **33**, 178–181.

Carmona, R. H. (**2006**). "Health literacy: A national priority," J. Gen. Intern. Med. **21**, 803.

Chang, J. E., Weinstein, B., Chodosh, J., and Blustein, J. (**2018**). "Hospital readmission risk for patients with self-reported hearing loss and communication trouble," J. Am. Geriatr. Soc. **66**, 2227–2228.

Chapman, K., Abraham, C., Jenkins, V., and Fallowfield, L. (**2003**). "Lay understanding of terms used in cancer consultations," Psychooncology **12**, 557–566.

Chapple, A., Campion, P., and May, C. (**1997**). "Clinical terminology: Anxiety and confusion amongst families undergoing genetic counseling," Patient Educ. Couns. **32**, 81–91.

Cohn, M., Pycha, A., and Zellou, G. (**2021**). "Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech," Cognition **210**, 104570.

Colombo, L., Pasini, M., and Balota, D. A. (**2006**). "Dissociating the influence of familiarity and meaningfulness from word frequency in naming and lexical decision performance," Mem. Cognit. **34**(6), 1312–1324.

Connine, C. M., Mullennix, J., Shernoff, E., and Yelen, J. (**1990**). "Word familiarity and frequency in visual and auditory word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **16**, 1084–1096.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooke, M., and Garcia Lecumberri, M. L. (**2021**). "How reliable are online speech intelligibility studies with known listener cohorts?" J. Acoust. Soc. Am. **150**(2), 1390–1401.

Corey, R. M., Jones, U., and Singer, A. C. (**2020**). "Acoustic effects of medical, cloth, and transparent face masks on speech signals," J. Acoust. Soc. Am. **148**, 2371–2375.

Crump, M. J., McDonnell, J. V., and Gureckis, T. M. (**2013**). "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research," PLoS One **8**(3), e57410.

Dahm, M. (**2012**). "Coming to terms with medical terms–Exploring insights from native and non-native English speakers in patient-physician communication," Hermes **25**(49), 79–98.

Darbyshire, J. L., and Young, J. D. (**2013**). "An investigation of sound levels on intensive care units with reference to the WHO guidelines," Crit. Care **17**, R187.

Denton, C. P., Laird, B., Moros, L., and Flores, J. L. L. (**2020**). "Challenges in physician-patient communication for optimal management of systemic sclerosis-associated interstitial lung disease: A discourse analysis," Clin. Rheumatol. **39**, 2989–2998.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.

Duffy, J. R., and Giolas, T. G. (**1974**). "Sentence intelligibility as a function of key word selection," J. Speech Hear. Res. **17**, 631–637.

Dunn, S. M., Butow, P. N., Tattersall, M. H., Jones, Q. J., Sheldon, J. S., Taylor, J. J., and Sumich, M. D. (**1993**). "General information tapes inhibit recall of the cancer consultation," J. Clin. Oncol. **11**, 2279–2285.

Dupuis, K., Marchuk, V., and Pichora-Fuller, M. K. (**2016**). "Noise affects performance on the Montreal Cognitive Assessment," Can. J. Aging **35**, 298–307.

Epstein, A., Giolas, T. G., and Owens, E. (**1968**). "Familiarity and intelligibility of monosyllabic word lists," J. Speech Hear. Res. **11**, 435–438.

Fage-Butler, A. M., and Jensen, M. N. (**2016**). "Medical terminology in online patient–patient communication: Evidence of high health literacy?," Health Expect. **19**, 643–653.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Folmer, R. L., Vachhani, J., McMillan, G. P., Watson, C., Kidd, G. R., and Feeney, M. P. (**2017**). "Validation of a computer-administered version of the digits-in-noise test for hearing screening in the United States," J. Am. Acad. Audiol. **28**(2), 161–169.

Foulkes, P., and Docherty, G. (**2006**). "The social life of phonetics and phonology," J. Phon. **34**(4), 409–438.

Gladd, D. K., and Saunders, G. H. (**2011**). "Ambient noise levels in the chemotherapy clinic," Noise Health **13**, 444–451.

Hadlow, J., and Pitts, M. (**1991**). "The understanding of common health terms by doctors, nurses and patients," Soc. Sci. Med. **32**, 193–196.

Hay, J. (**2018**). "Sociophonetics: The role of words, the role of context, and the role of words in context," Top. Cogn. Sci. **10**(4), 696–706.

Hay, J., Podlubny, R., Drager, K., and McAuliffe, M. (**2017**). "Car-talk: Location-specific speech production and perception," J. Phon. **65**, 94–109.

Homans, N. C., and Vroegop, J. L. (**2021**). "The impact of face masks on the communication of adults with hearing loss during COVID-19 in a clinical setting," Int. J. Audiol. **61**, 365–370.

Howes, D. (**1957**). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–305.

Hsu, T., Ryherd, E., Waye, K. P., and Ackerman, J. (**2012**). "Noise pollution in hospitals: Impact on patients," J. Clin. Outcomes Manag. **19**, 301–309.

Hunter, C. R., and Pisoni, D. B. (**2018**). "Extrinsic cognitive load impairs spoken word recognition in high- and low-predictability sentences," Ear Hear. **39**, 378–389.

Jorgensen, L. E., Palmer, C. V., Pratt, S., Erickson, K. I., and Moncrieff, D. (**2016**). "The effect of decreased audibility on MMSE performance: A measure commonly used for diagnosing dementia," J. Am. Acad. Audiol. **27**, 311–323.

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**, 1337–1351.

Koch-Weser, S., Dejong, W., and Rudd, R. E. (**2009**). "Medical word use in clinical encounters," Health Expect. **12**, 371–382.

Lerner, E. B., Jehle, D. V., Janicke, D. M., and Moscati, R. M. (**2000**). "Medical communication: Do our patients understand?," Am. J. Emerg. Med. **18**, 764–766.

Lin, F. R., Niparko, J. K., and Ferrucci, L. (**2011**). "Hearing loss prevalence in the United States," Arch. Intern. Med. **171**, 1851–1852.

MacKenzie, D., and Galbrun, L. (**2007**). "Noise levels and noise sources in acute care hospital wards," Build. Services Eng. Res. Technol. **28**, 117–131.

McGinnis, D., and Zelinski, E. M. (**2000**). "Understanding unfamiliar words: The influence of processing resources, vocabulary knowledge, and age," Psychol. Aging **15**, 335–350.

McGowan, K. B. (**2015**). "Social expectation improves speech perception in noise," Lang. Speech **58**(4), 502–521.

Mendel, L. L., Gardino, J. A., and Atcherson, S. R. (**2008**). "Speech understanding using surgical masks: A problem in health care?" J. Am. Acad. Audiol. **19**, 686–695.

Messingher, G. (**2013**). "Relating hospital noise to staff outcomes in real and simulated settings," master's thesis, Georgia Institute of Technology, Atlanta, GA.

Miller, G. A., and Isard, S. (**1963**). "Some perceptual consequences of linguistic rules," J. Verbal Learning Verbal Behav. **2**, 217–228.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Mormer, E., Cipkala-Gaffin, J., Bubb, K., and Neal, K. (**2017**). "Hearing and health outcomes: Recognizing and addressing hearing loss in hospitalized older adults," Semin. Hear. **38**, 153–159.

Morrison, W. E., Haas, E. C., Shaffner, D. H., Garrett, E. S., and Fackler, J. C. (**2003**). "Noise, stress, and annoyance in a pediatric intensive care unit," Crit. Care Med. **31**, 113–119.

Moulin, A., Vergne, J., Gallego, S., and Micheyl, C. (**2019**). "A new speech, spatial, and qualities of hearing scale short-form: Factor, cluster, and comparative analyses," Ear Hear. **40**(4), 938–950.

Mulley, G. P., and Ng, K. Y. (**1995**). "Problems encountered by hearing-impaired people in hospitals," Lancet **345**, 1640.

Munro, M. J. (**1998**). "The effects of noise on the intelligibility of foreign-accented speech," Stud. Second Lang. Acquis. **20**, 139–154.

Neill, B. C., Golda, N., Seger, E. W., Wick, J., Whitsitt, J., Huber, A., Chu, T., Potts, G. A., Chow, P., Moore, S., Fakhoury, J. W., Rajpara, A., and Hocker, T. L. H. (**2020**). "Determining patient understanding of commonly used dermatology terms: A multicenter cross-sectional survey," J. Am. Acad. Dermatol. **83**, 933–935.

Newman, R. S., and German, D. J. (**2005**). "Life span effects of lexical factors on oral naming," Lang. Speech **48**(2), 123–156.

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**, 1085–1099.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). "Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words," in *Research on Speech Perception Progress Report No. 10* (Indiana University, Bloomington, IN), pp. 357–376.

O'Hara, J. K., Reynolds, C., Moore, S., Armitage, G., Sheard, L., Marsh, C., Watt, I., Wright, J., and Lawton, R. (**2018**). "What can patients tell us about the quality and safety of hospital care? Findings from a UK multi-centre survey study," BMJ Qual. Saf. **27**, 673–682.

Okcu, S., Ryherd, E., Pelton, H., and Zimring, C. (**2012**). "Characterizing impulsiveness of hospital sound environments," Noise Control Eng. J. **60**, 246–257.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (**2019**). "PsychoPy2: Experiments in behavior made easy," Behav. Res. **51**, 195–203.

Perry, S., Bent, T., Ryherd, E., and Baese-Berk, M. (**2021**). "A novel corpus developed to evaluate the impact of hospital noise on speech intelligibility," in *Proceedings of INTER-NOISE 2021—2021 International Congress and Exposition of Noise Control Engineering*, August 1–5, Washington, DC, pp. 2157–2163.

Pope, D. S., Gallun, F. J., and Kampel, S. (**2013**). "Effect of hospital noise on patients' ability to hear, understand, and recall speech," Res. Nurs. Health **36**, 228–241.

Pörschmann, C., Lübeck, T., and Arend, J. M. (**2020**). "Impact of face masks on voice radiation," J. Acoust. Soc. Am. **148**, 3663–3670.

Rahne, T., Fröhlich, L., Plontke, S., and Wagner, L. (**2021**). "Influence of surgical and N95 face masks on speech perception and listening effort in noise," PLoS ONE **16**, e0253874.

Ratzan, S., Parker, R., Selden, C., and Zorn, M. (**2000**). *National Library of Medicine Current Bibliographies in Medicine: Health Literacy* (National Institutes of Health, Bethesda, MD).

Ryherd, E., Okcu, S., Hsu, T., and Mahapatra, A. (**2011**). "Hospital noise and occupant response," ASHRAE Trans. **117**, 248–255.

Ryherd, E. E., Moeller, M., Jr., and Hsu, T. (**2013**). "Speech intelligibility in hospitals," J. Acoust. Soc. Am. **134**, 586–595.

Ryherd, E. E., Okcu, S., Ackerman, J., Zimring, C., and Waye, K. P. (**2012**). "Noise pollution in hospitals: Impacts on staff," J. Clin. Outcomes Manag. **19**, 491–500.

Ryherd, E. E., Waye, K. P., and Ljungkvist, L. (**2008**). "Characterizing noise and perceived work environment in a neurological intensive care unit," J. Acoust. Soc. Am. **123**, 747–756.

Sakamoto, S., Iwaoka, N., Suzuki, Y., Amano, S., and Kondo, T. (**2004**). "Complementary relationship between familiarity and SNR in word intelligibility test," Acoust. Sci. Technol. **25**(4), 290–292.

Shukla, A., Nieman, C. L., Price, C., Harper, M., Lin, F. R., and Reed, N. S. (**2019**). "Impact of hearing loss on patient-provider communication among hospitalized patients: A systematic review," Am. J. Med. Qual. **34**, 284–292.

Slote, J., and Strand, J. F. (**2016**). "Conducting spoken word recognition research online: Validation and a new timing method," Behav. Res. **48**(2), 553–566.

Smiljanic, R., Keerstock, S., Meemann, K., and Ransom, S. M. (**2021**). "Face masks and speaking style affect audio-visual word recognition and memory of native and non-native speech," J. Acoust. Soc. Am. **149**, 4013–4023.

Smith, A., and Davis, A. (**2018**). "Disambiguating the use of common terms across related medical fields: The problem of intervention," Lexicography **4**, 63–80.

Souza, P. E., and Turner, C. W. (**1994**). "Masking of speech in young and elderly listeners with hearing loss," J. Speech. Lang. Hear. Res. **37**, 655–661.

Sperry, J. L., Wiley, T. L., and Chial, M. R. (**1997**). "Word recognition performance in various background competitors," J. Am. Acad. Audiol. **8**, 71–80.

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

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers," J. Speech Lang. Hear. Res. **47**, 245–256.

Tainter, C. R., Levine, A. R., Quraishi, S. A., Butterly, A. D., Stahl, D. L., Eikermann, M., Kaafarani, H. M., and Lee, J. (**2016**). "Noise levels in surgical ICUs are consistently above recommended standards," Crit. Care Med. **44**, 147–152.

Toscano, J. C., and Toscano, C. M. (**2021**). "Effects of face masks on speech recognition in multi-talker babble noise," PLoS ONE **16**, e0246842.

Truong, T. L., Beck, S. D., and Weber, A. (**2021**). "The impact of face masks on the recall of spoken sentences," J. Acoust. Soc. Am. **149**, 142–144.

U.S. Census Bureau (**2020**). "Educational attainment in the United States: 2020," https://www.census.gov/data/tables/2020/demo/educational-attainment/cps-detailed-tables.html (Last viewed April 11, 2022).

VanCott, M. L. (**1993**). "Communicative competence during nursing admission interviews of elderly patients in acute care settings," Qual. Health Res. **3**, 184–208.

van Heuven, W. J., Mandera, P., Keuleers, E., and Brysbaert, M. (**2014**). "SUBTLEX-UK: A new and improved word frequency database for British English," Q. J. Exp. Psychol. **67**, 1176–1190.
Verhaeghen, P. (**2003**). "Aging and vocabulary score: A meta-analysis," Psychol. Aging **18**, 332–339.
Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (**2017**). "Headphone screening to facilitate web-based auditory experiments," Atten. Percept. Psychophys. **79**, 2064–2072.
Yi, H., Pingsterhaus, A., and Song, W. (**2021**). "Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic," Front. Psychol. **12**, 682677.
Yoshida, Y., and Yoshida, Y. (**2014**). "Patient's recognition level of medical terms as estimated by pharmacists," Environ. Health Prev. Med. **19**, 414–421.
Zeng, Q., Kogan, S., Ash, N., and Greenes, R. A. (**2001**). "Patient and clinician vocabulary: How different are they?," Stud. Health Technol. Inform. **84**, 399–403.