Noise and a Speaker's Impaired Voice Quality Disrupt Spoken Language Processing in School-Aged Children: Evidence from Performance and Response Time Measures

Isabel S. Schiller[1], Dominique Morsomme[1], Malte Kob[2], and Angélique Remacle[1,3]

[1]Faculty of Psychology, Speech and Language Therapy, and Education, University of Liège, Belgium

[2]Erich Thienhaus Institute, Detmold University of Music, Germany

[3]Fund for Scientific Research – F.R.S. – FNRS, Brussels, Belgium

Disclosure: The authors have no conflict of interest to declare.

Funding: Published with the support of the Belgian University Foundation. Isabel S. Schiller was supported by a PhD grant from the University of Liège (no. RD/DIR-vdu/2016.7166), and Angélique Remacle was supported by the National Fund for Scientific Research – F.R.S. – FNRS, Brussels, Belgium

Corresponding Author:

Isabel S. Schiller

Rue de l'Aunaie, 30 (B38)

4000 Liège, Belgium

Phone: +32 4 366 32 64

E-mail: isabel.schiller@uliege.be

Abstract

Purpose: Our aim was to investigate isolated and combined effects of speech-shaped noise (SSN) and a speaker’s impaired voice quality on spoken language processing in first-grade children.

Method: In individual examinations, 53 typically developing children aged 5 to 6 years performed a speech perception task (phoneme discrimination) and a listening comprehension task (sentence-picture matching). Speech stimuli were randomly presented in a 2x2 factorial design with the factors *noise* (no added noise vs. SSN at 0 dB signal-to-noise ratio) and *voice quality* (normal voice vs. impaired voice). Outcome measures were task performance and response time (RT).

Results: SSN and impaired voice quality significantly lowered children’s performance and increased RTs in the speech perception task, particularly when combined. Regarding listening comprehension, a significant interaction between noise and voice quality indicated that children’s performance was hindered by SSN when the speaker’s voice was impaired but not when it was normal. RTs in this task were unaffected by noise or voice quality.

Conclusions: Results suggest that speech signal degradations caused by a speaker’s impaired voice and background noise generate more processing errors and increased listening effort in young school-aged children. This finding is vital for classroom listening and highlights the importance of ensuring teachers’ vocal health and adequate room acoustics.

Keywords: spoken language processing, speech in noise, voice quality

Noise and a Speaker's Impaired Voice Quality Disrupt Spoken Language Processing in School-Aged Children: Evidence from Performance and Response Time Measures

Because of the trajectory of spoken language acquisition, children are highly vulnerable to adverse listening conditions (Elliott, 1979). Phonological awareness continuously improves during the first years of school (Anthony & Francis, 2005), which may partly explain why younger pupils in particular have difficulties understanding acoustically degraded speech (Astolfi, Bottalico, & Barbato, 2012; Johnson, 2000). Generally, children benefit from high-quality speech signals and quiet surroundings for effective listening, but such conditions are rare. In classrooms, for example, noise levels frequently exceed official guidelines (Silva, Oliveira, & Silva, 2016), and the prevalence of voice disorders in teachers is between 20% and 50% (Martins, Pereira, Hidalgo, & Tavares, 2014). Investigating school-aged children's ability to perceive and comprehend speech that is degraded by noise and impaired voice quality is therefore critical.

The complex system that allows us to understand and retain speech is known as spoken language processing (SLP) (Medwetsky, 2011). We can broadly divide SLP into low-level speech perception and high-level listening comprehension. During speech perception, acoustic information is mapped onto linguistic representations (e.g., phonemes, syllables, or words) (Holt & Lotto, 2010). This auditory-perceptual mapping is a prerequisite for listening comprehension. Following Klatte, Lachmann, and Meis's (2010) use of the term, we define listening comprehension as the process whereby listeners integrate semantic, syntactic, and pragmatic information to construct the meaning of verbal messages.

As a whole system, SLP is closely related to working memory. Among other theories (reviewed in Wingfield, 2016), this link has been described in the Ease of Language Understanding model (Rönnberg et al., 2013), which provides a cross-modal explanation of how language is understood under different conditions. According to this model, impoverished speech signals may result in a mismatch between the perceptual input and a listener's phonological-lexical representations. To resolve this mismatch, the listener must deliberately allocate cognitive resources (i.e., explicit processing), which slows down processing because long-term memory must be consulted.

The effect of noise on school-aged children's SLP has repeatedly been demonstrated in listening tasks. For example, Jamieson, Kranjc, Yu, and Hodgetts (2004) tested 5- to 8-year-old children's ability to discriminate among phonologically similar words at classroom-typical signal-to-noise ratios (SNR), using a word-picture matching task presented in classroom noise. Decreasing SNRs significantly lowered task performance, particularly in younger children. Several further studies have shown noise-induced declines in speech perception (which focuses on low-level speech intelligibility) (Bradley & Sato, 2008; Crandell & Smaldino, 1996; Klatte et al., 2010; Prodi, Visentin, Borella, Mammarella, & Di Domenico, 2019), listening comprehension (which focuses on understanding longer utterances) (Klatte et al., 2010; Nirme, Haake, Lyberg-Åhlander, Brännström, & Sahlén, 2019; Sullivan, Osman, & Schafer, 2015), and working memory (Osman & Sullivan, 2014; Sullivan et al., 2015). However, most of these studies examined children around the age of 8 to 10 years old. We believe it is important to investigate the effects of noise on pupils in the early school years (i.e., children aged 5 to 7 years old), because the first grades are critical for language development (Anthony & Francis, 2005). Children's performance during this period may predict future academic performance, such as reading skills (Rabiner, Godwin, & Dodge, 2016).

The effects of noise are influenced not only by SNRs but also by the source of noise (Astolfi et al., 2012; Klatte et al., 2010; Peng, Zhang, & Yan, 2016; Prodi & Visentin, 2015; Prodi, Visentin, Borella, et al., 2019). This may be explained by energetic and informational masking as well as spectro-temporal aspects. Energetic masking refers to physical interference by noise (i.e., poor intelligibility due to shared acoustic characteristics of the noise signal and the speech signal) (Mattys, Brooks, & Cooke, 2009), while informational masking refers to "…everything that reduces intelligibility once energetic masking has been accounted for" (Cooke, Garcia Lecumberri, & Barker, 2008, pp. 414–415). Under conditions of high energetic masking, small dips (or glimpses) in the noise signal may improve listeners' speech-in-noise processing (Cooke, 2006; Klatte et al., 2010). There is, for example, some indication that competing speech is more detrimental to children's listening comprehension, while a steady-state noise has a stronger impact on speech perception (Klatte et al., 2010).

In addition to noise, children's SLP may be hampered when listening to a dysphonic speaker (i.e., a speaker with an impaired voice). Dysphonia is commonly used as a synonym for hoarseness, and refers to a coarse or rough voice quality (Schwartz et al., 2009). While noise degrades transmission (Mattys, Davis, Bradlow, & Scott, 2012), impaired voice modulates the speech signal directly during speech production; thus, at the source. Brännström, Kastberg, et al. (2018) suggested that the effect of impaired voice may be less problematic than the effect of noise. Morsomme, Minell, and Verduyckt (2011) studied the effect of voice quality on phonological discrimination and passage comprehension in 8-year-old children. When listening to a voice that was moderately to severely impaired, children performed significantly worse than when listening to a normal voice. This aligns with past findings that revealed negative effects of impaired voice on children's acceptability judgments (Brännström, Kastberg, et al., 2018), passage comprehension (Chui & Ma, 2018; Rogerson & Dodd, 2005), and word recall (Morton & Watson, 2001).

Research suggests that the effects of voice quality may be mediated by source/degree of dysphonia and task demands. For example, more pronounced effects have been found when the impaired voice was mimicked (Chui & Ma, 2018; Morsomme et al., 2011; Rogerson & Dodd, 2005) rather than provoked by means of vocal loading tasks (Lyberg-Åhlander, Holm, et al., 2015; Brännström, Kastberg, et al., 2018). In previous work, we pointed out that this probably relates to differences concerning dysphonia severity and perceptual voice characteristics (e.g., hyperfunction or breathiness) (Schiller, Remacle, & Morsomme, 2019a). Regarding task demands, the impact of impaired voice appears to be more detrimental when the listening task creates a considerable processing load (Lyberg-Åhlander, Haake, Brännström, Schötz, & Sahlén, 2015; Lyberg-Åhlander, Holm, et al., 2015). Processing load may increase not only due to linguistic factors but also due to acoustic interference (Rönnberg et al., 2013); thus, listening to dysphonic speech in noisy conditions should be particularly challenging.

The combined effect of noise and impaired voice on children's SLP has rarely been investigated. Two studies (Brännström, von Lochow, Lyberg-Åhlander, & Sahlén, 2018; Von Lochow, Lyberg-Åhlander, Sahlén, Kastberg, & Brännström, 2018) assessed listening comprehension at different SNRs (i.e., no added noise, speech noise at +10 dB SNR, and speech noise at +5 dB SNR) and voice qualities (normal voice and mildly to moderately impaired voice) in children between the ages of 7 and 12 years. Neither study revealed a significant interaction between noise and voice quality or a main effect of voice quality on children's performance. Only noise triggered a decline in performance. Considering that separate effects of each factor have previously been observed, these results are counterintuitive. On the other hand, in line with a review by Lyberg-Åhlander, Brännström, and Sahlén (2015), both studies provided indications of a complex interplay between listening conditions, task demands and children's executive functioning, which might have complicated the detection of significant effects. Clearly, this topic needs further investigation.

To better understand the listening effort required to listen to acoustically degraded speech, performance measures can be enriched with response time (RT) measures (McCreery & Stelmachowicz, 2013; McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017; Visentin & Prodi, 2018). Listening effort refers to the effort associated with "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016, p. 10S). Simply put, degraded listening conditions contribute to increased listening effort, but only when the listener intends to listen. According to the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016), listeners produce more errors and require longer processing times when their processing capacity is close to depletion. A recent study confirmed that collecting RTs in single-task paradigms (i.e., listening tasks that consist of one task only) is a useful technique for indirectly measuring listening effort in children from the age of 6 and older (McGarrigle, Gustafson, Hornsby, & Bess, 2019).

Indeed, McCreery and Stelmachowicz (2013) found that speech-shaped noise (SSN) and poor signal quality, induced by limiting the bandwidth, prolonged school-aged children's RTs in a speech perception task. Likewise, in the study by Prodi, Visentin, Borella, et al. (2019), children responded significantly slower in a speech perception task and a listening comprehension task when speech was presented in classroom noise. In another study by Prodi, Visentin, Peretti, Griguolo, and Bartolucci (2019), SSN increased 5- to 7-year-old children's response latencies in a word-picture matching task. Two other studies found no detrimental effects of noise on children's RTs (McGarrigle et al., 2017; Nakeva von Mentzer, Sundström, Enqvist, & Hällgren, 2018). Regarding voice quality, Sahlén et al. (2017) found that listening to an impaired voice increased RTs in girls but not boys in a listening comprehension task. The combined effect of noise and impaired voice on RTs has never been studied.

The goal of this study was to investigate isolated and combined effects of noise and a speaker's impaired voice quality on speech perception and listening comprehension in first-grade children (5 to 6 years old). Speech perception primarily refers to the process of auditory-perceptual mapping. Listening comprehension focuses on the processing of meaning (i.e., content level of speech). Specifically, we sought to determine to what extent noise and impaired voice influenced children's performance and RTs in a phonological discrimination task and a sentence-picture matching task. Four hypotheses were tested:

- H1: Noise or impaired voice quality reduces children's performance and increases RTs in speech perception.
- H2: A combination of noise and impaired voice quality results in even poorer performance and longer RTs in speech perception than each factor alone.
- H3: Noise or impaired voice quality reduces children's performance and increases RTs in listening comprehension.
- H4: A combination of noise and impaired voice quality results in even poorer performance and longer RTs in listening comprehension than each factor alone.

Methods

Participants

Figure 1 depicts the participant recruitment and selection procedure. Out of 94 first-grade children who participated in the experiment, 53 children (28 girls) between 5 and 6 years old (*M* = 6;4) were eligible for inclusion in the statistical analysis. Participants were recruited from five randomly selected primary schools within the French-speaking community of Belgium. During information sessions, the children were given consent forms and questionnaires for their parents. The questionnaires concerned the child's age, mother tongue, auditory development, and speech-language development.

We applied the following inclusion criteria: (a) between 5 and 6 years of age; (b) French native speaker; (c) normal auditory development; (d) normal speech-language development; (e) hearing thresholds ≤ 25 dB HL bilaterally at octave frequencies between 500 and 4000 Hz (audiometric screening); (f) score ≥ 25th percentile (i.e., normal and above-normal performance) in a receptive lexical subtest (i.e., LexR subtest of the Épreuve du Langage Orale (ELO) [Oral Language Assessment]; Khomsi, 2001); and (g) score ≥ 25th percentile (i.e., normal and above-normal performance) in an auditory selective attention test (i.e., AA subtest of the Bilan NEuroPSychologique de L'Enfant 2 (NEPSY-II) [Developmental NEuroPSYchological Assessment]; Korkman, Kirk, & Kemp, 2007).

Children's compliance with inclusion criteria (a) to (d) was determined based on parental report (questionnaire), while compliance with criteria (e) to (g) was assessed on the day of the experiment during individual examinations in a quiet room at school. These examinations consisted of the pure-tone audiometric screening (using a MADSEN Itera II audiometer with TDH-39 earphones), the receptive lexical test (ELO material), and the selective attention test (NEPSY-II material).

- Figure 1 -

Ethics Statement

All participating children gave their oral informed consent. Written informed consent was obtained from their parents. The ethics committee of the Faculty of Psychology, Speech and Language Therapy, and Education (University of Liège, Belgium) approved the study (file no. 1617-54).

Speech Perception Task

Speech perception was assessed by means of a phonological discrimination task. For this purpose, we created a digitized version of the Épreuve Lilloise de Discrimination Phonologique

(Macchi et al., 2012). List 1 of this test is designed for French-speaking children aged between 5 and 6;6 years and contains 36 spoken pseudo-word pairs (i.e., words that follow phonotactic rules but have no meaning, which controls for semantic priming effects). Speech items demonstrate either structural oppositions (e.g., kaʃifugR/ – /kafiʃugR/) or phonemic oppositions (e.g., /zil/ – /zij/) and their length ranges between one and three syllables. Children's task is to decide whether the two pseudo-words in each pair are identical or different.

In our version of the task, children discriminated between the pseudo-words by touching a screen (see Appendix A for a picture of the experimental set-up). The task was presented on a laptop with an integrated touch screen (Dell Latitude 5480). We used the experimental software OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Children were instructed to discriminate between pseudo-words by touching the correct response symbol on the screen (i.e., symbols denoting the options "same" versus "different"). Speech stimuli were played via earphones (AKG K 271 MK II) in a randomized order. Performance was measured in terms of a binary outcome variable (1 = correct response, 0 = incorrect response). RTs were automatically collected in OpenSesame and comprised the time from stimulus offset to the moment the touch response was registered. This means that, irrespective of the listening condition, RTs were measured in quiet surroundings. The permitted response time was unlimited. Response symbols remained visible on the screen until the response was registered.

Listening Comprehension Task

Listening comprehension was assessed with a sentence-picture matching task from the ELO material (Khomsi, 2001). Again, a digitized version of the task was created for this study. Designed for children aged 5 to 10 years old, the ELO sentence-picture matching task contains 32 sentences (21 of which are recommended for the use with 6-year-olds), which vary in length and syntactic complexity. Each sentence is presented orally with a set of four pictures (one target picture and three morphosyntactic or semantic distractors). Children’s task is to match each sentence to the corresponding picture.

For the purpose of this study, the 21 ELO sentences were presented via earphones and pictures were presented on a computer screen (see Appendix B for a picture of the experimental set-up). Presentation material and software were the same as for the speech perception task. Children were instructed to listen to each sentence and select the matching picture on the screen. Sentences were presented in a randomized order and performance and RT measures were collected in the same way as for the speech perception task.

Listening Conditions and Stimuli Preparation

Speech stimuli (i.e., pseudo-word pairs and sentences) were prepared according to four listening conditions: (C1) normal voice in quiet, (C2) impaired voice in quiet, (C3) normal voice in noise, and (C4) impaired voice in noise. For speech-in-quiet conditions (C1 and C2), we achieved highly favorable SNRs ranging between +31 dB and +33 dB (a certain amount of noise is introduced automatically during the recording procedure). For speech-in-noise conditions (C3 and C4), we applied a 0 dB SNR to simulate typical classroom conditions (Crandell & Smaldino, 1996; Howard, Munro, & Plack, 2010).

We recorded the speech stimuli in accordance with the recommendations of Barsties and De Bodt (2015). The speaker was a 51-year-old female speech-language therapist. During a single recording session, she recorded all stimuli in her normal voice and an imitated dysphonic voice. These speech files are available in the NOrmophonic and DYsphonic Speech samples database (Schiller, Remacle, & Morsomme, 2019b). A previous study validated both voice qualities using perceptual and acoustic evaluations (Schiller et al., 2019a). For the perceptual evaluation, five speech-language therapists listened to part of the speech samples and rated them on the parameters of the GRBAS scale (i.e., overall Grade, Roughness, Breathiness, Asthenia, and Strain) (Hirano, 1981), as well as their authenticity and consistency. They perceived the normal voice as non-pathological (i.e., all GRBAS parameters rated 0), authentic, and consistent. The imitated dysphonic voice was perceived as rough and asthenic, and moderately to severely dysphonic (median GRBAS scores: Grade = 3, Roughness = 3, Breathiness = 2, Asthenia = 3, Strain = 1), with acceptable authenticity and consistency. Interrater reliability was moderate (K = 0.52). For the acoustic evaluation, we calculated the Acoustic Voice Quality Index (Maryn, Corthals, Van Cauwenberge, Roy, & De Bodt, 2010), which is based on a sustained vowel /a/ concatenated with connected speech, as an objective measure of dysphonia. Its score ranges from 0 (normal voice) to 10 (severe dysphonia). Moreover, we extracted jitter, shimmer, and harmonics-to-noise ratios (HNR) from a sustained vowel /a/. The results were in line with the perceptual evaluations. The normal voice yielded a non-pathological Acoustic Voice Quality Index score of 2.53. Perturbation measures were also low. The imitated dysphonic voice yielded an Acoustic Voice Quality Index score of 6.89, indicating a moderate to severe voice pathology, and high perturbation measures (jitter = 2.8%, shimmer = 9.2%, HNR = 10.8). In sum, our voice evaluation suggested that (a) the speaker's normal voice was non-pathological, and (b) she succeeded in imitating a moderate to severe dysphonia.

Before the speech-in-noise conditions were created, all auditory stimuli were equalized to a mean level (root-mean-square) of 65 dB, using Praat, version 6.0.29 (Boersma & Weenink, 2017). Speech stimuli were then merged with SSN to create an SNR of 0 dB. We used the STIPA signal (DIN EN IEC 60268-16; Deutsches Institut für Normung e.V. [German Institute for Standardization], 2019), an amplitude-modulated SSN covering several octave bands in the frequency range of speech (125 Hz to 8 kHz). Houtgast et al. (2002) developed this signal as a test signal for the Speech Transmission Index. The quasi-stationary characteristics of the STIPA signal preclude the risk of erratic noise events masking certain phonemes more than others. At the same time, the signal approximates the spectral and temporal characteristics of speech, which is favorable because competing speech is a common type of interference in classroom listening. Speech-in noise conditions (C3 and C4) were prepared such that the noise and speech signal always started and ended simultaneously. No noise was played between the items.

The long-term average spectra of the two voice qualities and the noise signal are presented in Figure 2. Two important aspects should be mentioned: first, the normal voice shows more spectral components than the impaired voice in frequency regions up to about 2000 Hz, which are critical for speech intelligibility (Ardoint & Lorenzi, 2010; Ishikawa, Nudelman, Park, & Ketring, 2020). Compared to SSN, the normal voice is more intense up to frequencies of about 1000 Hz (covering the fundamental frequency and the range of the first formant), which may contribute to vowel disambiguation. Second, the impaired voice generally shows more spectral components in higher-frequency regions, with a peak between 3300 Hz and 4100 Hz. This suggests a higher proportion of noise components (i.e., components potentially degrading speech intelligibility), which aligns with the low HNR (i.e., 10.8 vs. 25 in the normal voice).

- Figure 2 -

Procedure

Prior to the experiment, we ran a pilot study to evaluate the appropriateness and clarity of our material and experimental procedure. Five 5- and 6-year-old children were tested in quiet rooms in their homes. The pilot test confirmed that the study design was suitable, the instructions were comprehensible, and the 0 dB SNR was appropriate. Several children were not familiar with the touch screen, so we incorporated a short practice phase in the procedure for the main experiment.

The main experiment was conducted in separate rooms at each of the participating schools. Noise levels were measured with a PCE-353 sound level meter (PCE Holding GmbH, Germany) and ranged between 35 and 43 dB(A). A potential effect of ambient noise on the results cannot be fully ruled out, as the earphones used to present the auditory stimuli were not noise-attenuating. Children were assessed individually. Each assessment lasted about 20 minutes. In a fully crossed design, all children performed both listening tasks: speech perception and listening comprehension. Stimuli were presented randomly according to the four listening conditions. For example, a child might listen to one item in a normal voice in quiet and the next item in an impaired voice in noise. We used OpenSesame software (Mathôt et al., 2012) to randomize sequence allocation based on participant number. The examiners were three second-year speech-language therapy students who were supervised by the first author [I.S.] to ensure standardized test administration.

During the experiment, we first seated the children in front of the laptop and taught them how to use the touch screen. Based on a sample speech signal, children were encouraged to set a comfortable intensity level. The experimenter then asked, “Is this level comfortable for you or is it too loud or too quiet?” and allowed time for further adjustments if necessary. Afterward, the experimenter launched the experiment, which started with the listening comprehension task followed by the speech perception task. Our rationale for this predefined order was that the task instructions for the listening comprehension task were less abstract, which helped children to become familiar with the response method. Each task began with a few practice trials (listening comprehension: n = 3; speech perception: n = 4). The practice trials used different material from the tasks and were later discarded from the statistical analyses. The children were instructed to listen carefully to each item and then to respond as accurately as possible by selecting the corresponding symbol (speech perception task) or picture (listening comprehension task). They received no instructions about how quickly they should respond and were unaware that RTs were collected. Considering the children’s young age, we did not want to create any pressure regarding response speed. When a child touched the screen, it went black. The examiner launched the next item after verifying that the child was still attending to the task. Between the two tasks, the children were allowed a short break of about 1 or 2 minutes.

Statistical Analysis

Data were analyzed using R software, version 3.3.2 (R Core Team, 2019). Response variables were task performance and RT. Performance was assessed in terms of children's probability of correct responses. RT (in ms) comprised the time from stimulus offset to screen touch. Only RTs from correct trials fed into the statistical models, following the lead of earlier studies (Balota, Aschenbrenner, & Yap, 2013; McCreery & Stelmachowicz, 2013; Zhang, Barry, Moore, & Amitay, 2012). The rationale was that RTs from incorrect trials are difficult to interpret as errors may have different causes. RTs of less than 200 ms ($n = 30$) were considered unrealistically short (potentially representing fast guesses) and removed (Balota et al., 2013; Whelan, 2008; Zhang et al., 2012). We also removed RTs that were not immediately registered ($n = 21$). These RTs were removed based on the experimental record (i.e., the experimenter noted when a child touched the screen twice, which occurred if the first touch response was too soft). Overall, performance data include 3,021 trials and RT data came from 2,005 of these trials (i.e., 66%). The relationship between these response variables was investigated with Spearman's correlations.

Statistical modeling involved generalized linear mixed-effect models (GLMMs) using the glmer function of the lme4 package, version 1.1-15 (Bates, Maechler, Bolker, & Walker, 2015). Unlike classical ANOVAs, GLMMs allow individual predictions rather than averaging data over items or participants (Baayen & Milin, 2010). With respect to the binary outcome variable task performance, we chose GLMMs because they have been claimed to generate more reliable results for categorical variables than ANOVAs (Jaeger, 2008). Regarding RTs, our data were positively skewed, which is a typical result (Whelan, 2008). They also contained missing values. We opted for GLMMs as they do not require prior data transformation to yield normal distributions (Lo & Andrews, 2015) and are powerful in dealing with missing data (Quené & Van Den Bergh, 2004).

To assess task performance, we fitted the GLMMs with a binomial distribution and a logit link function. Similar to Visentin and Prodi (2018), we modeled RTs with a gamma distribution and log link function. For each of the two tasks, we fitted one GLMM for task performance and one for RT. *Noise* (no added noise vs. SSN at 0 dB SNR), *voice quality* (normal voice vs. impaired voice), and the *noise* x *voice quality* interaction were treated as fixed factors. The models controlled for random effects of *participant*, *item* and *trial* by means of random intercepts. *School* and *gender* were two further factors we initially considered but then dropped because they did not significantly improve the models.

Models were established by increasing their complexity in a stepwise process. Each new model was compared to the previous simpler model (e.g., *noise* x *voice quality* vs. *noise* + *voice quality*) by means of the Akaike Information Criterion (Akaike, 1974) using R's anova function. When listening comprehension performance was modeled, the interaction term improved the model fit and was therefore kept as a fixed factor. The other three final models that predicted performance and RTs for speech perception and RTs for listening comprehension included *noise* and *voice quality* as separate fixed effects. We assumed an $\alpha = .05$ significance level. For significant effects, we performed pairwise comparisons using the lsmeans package (Lenth & Lenth, 2018), adjusting for multiple comparisons by means of Tukey's HSD test.

Results

In the following sections, we present the effects of noise and voice quality on performance and RT measures according to task. Regarding RTs, we generally found that children took significantly more time when responding incorrectly than when responding correctly ($\chi^2(1) = 117$, $p < .001$). For speech perception, mean RTs were 1895 ms (*SE* = 75) for incorrect trials and 1730 ms (*SE* = 65) for correct trials; for listening comprehension, the means were 4153 ms (*SE* = 281) and 3513 ms (*SE* = 232), respectively. The RT results discussed below concern only data from correct trials.

Effects of Noise and Impaired Voice on Speech Perception

Performance and RT measures for each condition of the speech perception task are presented in Figure 3 and Figure 4, respectively. Figure 3 shows that performance was best in the control condition (C1: *M* = .89, *SE* = .02, range = 0.33–1), decreased in the impaired voice condition (C2: *M* = .83, *SE* = .04, range = 0.11–1) and the SSN condition (C3: *M* = .72, *SE* = .05, range = 0.22–1), and dropped to close to chance level when the two factors were combined (C4: *M* = .60, *SE* = .06, range = 0.22–0.89). Likewise, Figure 4 shows that RTs were shortest in the control condition (C1: *M* = 1630 ms, *SE* = 98, range = 986–3708 ms), increased in the impaired voice (C2: *M* = 1737 ms, *SE* = 105, range = 1014–3775 ms) and SSN conditions (C3: *M* = 1792 ms, *SE* = 108, range =1095–3911 ms), and were longest when the two factors were combined (C4: *M* = 1910 ms, *SE* = 116, range = 985–5569 ms).

- Figure 3 and Figure 4 -

Table 1 presents the GLMM results for the speech perception task. Both *noise* and *voice quality* significantly affected children's performance and RTs irrespective of gender. Compared to the control condition (C1), post hoc Tukey's HSD pairwise comparisons showed that either impaired voice (C2) or SSN (C3) significantly reduced children's speech perception performance (C1–C2: $z = –4.5$, $p < .001$; C1–C3: $z = –9.16$, $p < .001$) and lengthened their RTs (C1–C2: $z = 3.52$, $p = .002$; C1–C3: $z = 5.14$, $p < .001$). Moreover, the combination of noise and impaired voice (C4) was significantly more disruptive than either factor alone, both in terms of performance (C2–C4: $z = –9.16$, $p <.001$; C3–C4: $z = –4.5$, $p < .001$) and in terms of RTs (C2–

C4: z = 3.52, p = .002 and C3–C4: z = 5.14, p < .001). Most of the remaining comparisons between conditions were also significant (Performance: C1–C4: z = –9.48, p < .001; C2–C3: z = –3.57, p = .002; RT: C1–C4: z = 6.1, p < .001 and C2–C3: z = 1.19, p = .632). Speech perception performance did not correlate with RT (r_s = –.08, p = .244). The absence of a correlation between the task performance and RT variables indicated that there was no speed-accuracy tradeoff, which would have occurred if fast responders made more errors than slow ones (Ratcliff, Gomez, & McKoon, 2004).

- Table 1 -

Effects of Noise and Impaired Voice on Listening Comprehension

Figure 5 presents performance measures and Figure 6 shows RT measures for each condition of the listening comprehension task. As illustrated in Figure 5, children's performance under the normal voice in quiet condition (C1) was equal to their performance with a normal voice in noise (C3: M = .60, SD = .06, range = 0–1). When listening to the impaired voice, however, children performed better in quiet (C2) than in noise (C2: M = .66, SD = .05, range = 0.2–1; C4: M = .50, SD = .06, range = 0–1). Figure 6 shows that RTs were relatively equal across all conditions (C1: M = 3415 ms, SE = 316, range = 1284–9032 ms; C2: M = 3408 ms, SE = 314, range = 1084–8347 ms; C3: M = 3509 ms, SE = 323, range = 863–24264 ms; C4: M = 3501 ms, SE = 324, range = 1196–23186 ms).

- Figure 5 and Figure 6 –

Table 2 presents the GLMM results for the listening comprehension task. Again, results were unaffected by children's gender. There was a significant interaction between *noise* and *voice quality* on children's task performance, indicating that SSN only impeded performance when the speaker's voice was impaired. Post hoc comparisons confirmed that the performance difference between the two impaired-voice conditions was significant (C2–C4: z = –3.38, p <

.01), while there was no performance difference between the two normal voice conditions (C1–C3: $z = 0.17$, $p = 1$), and none of the other pairwise comparisons was significant. Neither *noise* nor *voice quality* significantly affected RTs. Finally, performance and RTs were not correlated ($r_s = .024$, $p = .73$), again suggesting that there was no speed-accuracy tradeoff.

- Table 2 -

Discussion

Effects of Noise and Impaired Voice on Speech Perception

In this study, we explored the effects of noise and a speaker's impaired voice on first-grade children's speech perception and listening comprehension. The results of the speech perception task showed that each factor generated a decrease in performance and an increase in RT. This was in line with H1 (i.e., noise or impaired voice quality reduces children's performance and increases RTs in speech perception).

Regarding the effect of noise on speech perception performance, the results were generally in compliance with the findings of Jamieson et al. (2004) and Klatte et al. (2010), who assessed speech-in-noise perception in 5- to 8-year-olds. Their noise sources were classroom noise (Jamieson et al, 2004; Klatte et al., 2010) and speech noise (Klatte et al., 2010). A comparison with age-matched children from these studies supported the hypothesis that noise effects vary with noise source, task complexity, and SNR; in our study, speech-shaped noise at 0 dB SNR lowered phoneme discrimination performance by ~20% compared to the control condition. Klatte et al. (2010) found a similar effect size for classroom noise (~22%) but a lower effect size for speech noise (~6%) in a word-picture matching task presented at comparable SNRs. In Jamieson et al.'s (2004) study, classroom noise did not affect word-picture matching until an SNR of –6 dB. To better predict the effects of different noise sources on children's speech perception, more studies should be conducted, in which several types of noise are contrasted (e.g., Peng et al., 2016). Nevertheless, inter-study comparisons are hampered due to methodological differences.

Our results showed a significant increase in RTs of ~170 ms in noise at 0 dB SNR compared to quiet. This supports earlier findings by McCreery and Stelmachowicz (2013), Prodi, Visentin, Borella, et al. (2019), and Prodi, Visentin, Peretti, et al. (2019), who administered speech perception tasks to children aged 6 to 12 years, 11 to 13 years, and 5 to 7 years, respectively. For example, McCreery and Stelmachowicz (2013) measured an RT increase of ~90 ms in noise when SNRs dropped from +9 dB to +3 dB SNR. For Prodi, Visentin, Borella, et al. (2019), classroom noise (but not traffic noise) presented at ~0 dB SNR resulted in an RT increase of ~130 ms compared no additional noise. However, McGarrigle et al. (2017) found no effects of noise on children's RTs. In Nakeva von Mentzer et al.'s (2017) study, children actually responded faster in noisy than in quiet conditions. Possible reasons for these unexpected findings might be floor/ceiling effects (McGarrigle et al., 2017) and an unbalanced test order (Nakeva von Mentzer et al., 2017). We controlled for these factors by using an existing task with available reference data and by ensuring a randomized sequence. Our results indicate that noise may slow down children's SLP even when auditory-perceptual mapping is successful (recall that we only analyzed RTs from correct trials). Concurring with the cognitive mechanisms described in the Ease of Language Understanding model (Rönnberg et al., 2013) and the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016), we interpreted this RT increase as an indication of listening effort resulting from excessive processing costs.

Our study provides the first evidence of the negative effect of impaired voice on phonological discrimination in 5- to 6-year-old children. Listening to an impaired voice lowered performance by ~11% and increased RTs by ~100 ms. The disruptive effect of impaired voice concurs with the findings of Morsomme et al. (2011), although their listeners were older (8 years) and the results involved only performance measures. We assume that the negative effect of impaired voice is due to imprecise phoneme realizations, an example being the devoicing of voiced phonemes (Schoentgen, 2006). In line with this assumption, a recent study showed that dysphonia reduces vowel intelligibility (Ishikawa et al., 2020). As opposed to when listening to a normal voice, children seem to have required more processing time to discriminate such non-prototypical phoneme candidates (e.g., when discriminating the pseudo-words /tɔ̃kl/ and /tɔ̃gl/).

In line with H2 (i.e., a combination of noise and impaired voice quality results in even poorer performance and longer RTs in speech perception than each factor alone), the combination of noise and impaired voice had more detrimental effects on children's performance and RTs than each factor in isolation. When listening to an impaired voice in noise, children's performance decreased by ~33% and RTs increased by ~270 ms compared to the control condition. In the absence of any contextual cues, the speech perception task required children to rely solely on auditory-perceptual mapping. This was no longer possible as intelligibility became too low to restore missing phonemes. Importantly, the effect of noise did not simply outweigh the effect of impaired voice but added to it. In the present study, we applied an imitated, moderately to severely dysphonic voice. It would be interesting to investigate whether the results would change if the degree of dysphonia was lower.

Effects of Noise and Impaired Voice on Listening Comprehension

Contrary to H3 (i.e., noise or impaired voice quality reduces children's performance and increases RTs in listening comprehension) and previous studies (Chui & Ma, 2018; Klatte et al., 2010; Morsomme et al., 2011; Osman & Sullivan, 2015; Prodi, Visentin, Borella, et al., 2019; Rogerson & Dodd, 2005; Sullivan et al., 2015), we found that noise and impaired voice quality did not have separate effects on children's performance or RTs in the listening comprehension task. One reason might be that this task offered syntactic and semantic contextual cues the children could use to compensate for reduced intelligibility. Considering that comprehension performance collapsed when the two factors were combined, the benefit of contextual cues seems to have diminished as listening conditions became too adverse. In addition, the strong variance in performance and RT data suggests that the lack of main effects of either noise or impaired voice could also relate to item heterogeneity (i.e., variations in sentence length and syntactic complexity). Although our GLMMs controlled for the effect of item, the fact that working memory demands varied between the sentences is not ideal. Consider, for example, that children's speech-in-noise listening performance has been shown to correlate with their working memory loading (Sullivan et al., 2015). In line with this, impaired voice appears to be most disruptive at an intermediate degree of task difficulty, while the effects diminish as the task becomes either too simple or too complex (Lyberg-Åhlander, Brännström, et al., 2015). Analyzing performance and RT data for each individual sentence might therefore have revealed more detailed information regarding this factor, but it was beyond the scope of the present study.

Our results partially confirmed H4 (i.e., a combination of noise and impaired voice quality results in even poorer performance and longer RTs in listening comprehension than each factor alone). The central result was the significant interaction effect between noise and voice quality on children's performance (but not RTs). When the speaker's voice was normal, performance was unaffected by noise. However, when the speaker's voice was impaired, noise decreased performance by ~23%. Analyses of the long-term average spectra (Figure 2) indicated that the spectral properties of the speech signals might have contributed to this finding. For example, the normal voice was characterized by more spectral components in frequency regions up to about 2000 Hz (regions that are important for speech intelligibility). As shown by Schiller et al. (2019a), the normal voice was also more favorable in terms of HNR (i.e., 25 dB vs. 10.8 dB). These factors suggest that the impaired voice was more susceptible to energetic masking by noise than the normal voice. Although our results did not entirely concur with H4, they demonstrate that a combination of noise at a typical classroom level (Howard et al., 2010) and a speaker's impaired voice may severely affect children's listening comprehension. We speculate that this effect is twofold: (a) speech intelligibility declines with the increasing spectral overlap of speech and noise signals; and (b) listening becomes more effortful as more cognitive capacity is taken up by the processing of the speaker's atypical voice quality or the inhibition of irrelevant noise.

In contrast to this study, the two previous studies that investigated the combined effects of noise and impaired voice on children's listening comprehension found neither an additive effect nor a significant interaction (Brännström, von Lochow et al., 2018; Von Lochow et al., 2018). Let us consider some possible reasons: first, we applied a 0 dB SNR, which likely resulted in a higher ratio of masked speech segments than the more favorable SNRs applied by von Lochow et al. (2018) (i.e., +5 dB) and Brännström, von Lochow et al. (2018) (i.e., +10 dB). Second, we used speech-shaped noise while the other two studies used actual speech noise (i.e., noise coming from one or more speakers, inducing different proportions of energetic and informational masking; Mattys et al., 2009). Third, we used an imitated impaired voice with a moderate to severe degree of dysphonia, whereas the other two studies used provoked impaired voices with a mild to moderate degree of dysphonia. Although previous studies have suggested that even mild voice impairments may affect performance (Chui & Ma, 2018; Rogerson & Dodd, 2005), it is still possible that our impaired voice was more disturbing. Finally, Von Lochow et al. (2018) and Brännström, von Lochow et al. (2018) tested children with a mean age of 8 years and 10 years, respectively, who might have possessed more advanced SLP skills to cope with adverse listening conditions than our 5- and 6-year-old participants. This concurs with the assumption that children become less affected by masking and more proficient at using contextual cues in noisy situations as they get older (Elliott, 1979).

Overall Considerations

In this study, both noise and impaired voice were found to hamper children's processing of spoken language. But how can we distinguish between their effects on the speech signal and on SLP? Regarding effects on the speech signal, this is relatively straightforward: impaired voice modulates the speech signal during production. Acoustically, it is characterized by correlates such as increased noise components or F0 and amplitude irregularities (Schoentgen, 2006). Noise interferes with the speech signal during its transmission by creating overlapping acoustic information (Cooke et al., 2008; Mattys et al., 2009). Regarding effects on children's SLP, the differentiation is less clear-cut. As our results indicated, both factors may reduce intelligibility – impaired voice by distorting speech (e.g., devoicing of voiced phonemes), and noise by masking it – and increase listening effort. An important difference concerns the quantification of exposure; noise interference can be quantified by means of SNR. To measure the degree of dysphonia, researchers rely on subjective ratings or acoustic analyses. We therefore question the claim that noise may be more disturbing than impaired voice (Brännström, Kastberg, et al., 2018). Although the findings from the speech perception task would support this claim, we argue that drawing such a comparison is problematic since noise and impaired voice do not share a common metric. In future, it may be interesting to explore whether SNR and HNR can be related in a way that allows the comparison of interfering noise and "phonation noise" (i.e., noise caused by dysphonia).

Limitations

There are some limitations on this study that should be considered. First, adhering to the common practice in speech-in-noise perception studies (Crandell et al., 1996; Klatte et al., 2010;

McGarrigle et al., 2017; Peng et al., 2016), the speech recordings were made in quiet conditions. While this approach ensures a high recording quality, it does not account for the fact that speakers adapt their voice use in noisy situations – the Lombard effect (Lombard, 1911). Such vocal adjustments may improve speech intelligibility (Garnier & Henrich, 2014), and it is therefore possible that our speech-in-noise conditions posed a greater listening challenge than if Lombard speech had been used (e.g., Brännström, von Lochow et al., 2018; von Lochow et al., 2018).

Second, we prepared the auditory stimuli such that speech and noise started and ended simultaneously in each speech-in-noise condition. The rationale was to keep the length of the items stable across the four different listening conditions, randomized across participants. We concede that this method has the risk that noise onsets may potentially affect children's performance. Introducing a lead time (i.e., launching noise prior to the speech signal) could avoid this problem and might therefore be the preferred method. For example, Visentin and Prodi (2018) and Brännström, von Lochow et al. (2018) launched noise signals 1000 ms before the start of the speech signal.

Third, in line with some previous studies (Prodi, Visentin, Peretti, et al., 2019; Visentin & Prodi, 2018), we defined RT as the time between the offset of the auditory stimulus and the point when the child touched the screen. However, RTs to speech stimuli may vary with a listener's motivation (Lyberg-Åhlander, Brännström, et al., 2015), and motivation is likely affected by item length and complexity. To better account for this aspect, it would have been interesting to also measure RTs from the onset of the auditory stimulus and relate them to the RTs reported here.

Conclusion

This study shows that listening to speech in noise and/or to a speaker's impaired voice may disrupt children's ability to process spoken language. Speech-shaped noise and impaired voice impeded 5- and 6-year-old children's performance and lengthened their RTs in a speech perception task, particularly when combined. It seems that, even when no processing errors are made, adverse listening conditions still slow down children's phoneme perception. The results of the listening comprehension task revealed that children's speech-in-noise performance declined significantly when the speaker's voice was impaired but not when it was normal. Taken together, our findings suggest that a combination of noise and impaired voice may be especially detrimental for SLP in school-aged children, which has crucial implications for the educational context. Children would probably need to explicitly employ processing capacity to understand a dysphonic teacher in a noisy classroom. This may be particularly difficult for children with language or hearing impairments, or non-native speakers. Another important discovery was that noise and impaired voice affected SLP at quite an early stage. Disruptions during speech perception are likely to carry over to higher-order SLP, potentially affecting auditory working memory, syntactic parsing, and semantic processing. Future experiments in more realistic settings and with different noise sources are needed to confirm the ecological validity of our findings.

Acknowledgments

The authors would like to thank the participating schools for their assistance during the experiment. We also acknowledge the help of Florence Bastings, Louise Lemmens, and Aurélie Gillot in collecting the data.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Ardoint, M., & Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing Research*, *260*(1–2), 89–95. https://doi.org/10.1016/j.heares.2009.12.002

Astolfi, A., Bottalico, P., & Barbato, G. (2012). Subjective and objective speech intelligibility investigations in primary school classrooms. *Journal of the Acoustical Society of America, 131*, 247–257. https://doi.org/10.1121/1.3662060

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1563–1571. https://doi.org/10.1037/a0032186

Barsties, B., & De Bodt, M. (2015). Assessment of voice quality: Current state-of-the-art. *Auris Nasus Larynx*, *42*(3), 183–188. https://doi.org/10.1016/j.anl.2014.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Boersma P., & Weenink D. (2017). Praat (version 6.0.29) [Computer software]. Retrieved from http://www.fon.hum.uva.nl/praat/

Bradley, J. S., & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *Journal of the Acoustical Society of America*, *123*(4), 2078–2086. https://doi.org/10.1121/1.2839285

Brännström, K. J., Kastberg, T., von Lochow, H., Haake, M., Sahlén, B., & Lyberg-Åhlander, V. (2018). The influence of voice quality on sentence processing and recall performance in school-age children with normal hearing. *Speech, Language and Hearing*, *21*(1), 1–9. https://doi.org/10.1080/2050571X.2017.1309787

Brännström, K. J., von Lochow, H., Lyberg-Åhlander, V., & Sahlén, B. (2018). Immediate passage comprehension and encoding of information into long-term memory in children with normal hearing: The effect of voice quality and multitalker babble in noise. *American Journal of Audiology*, *27*, 231–237. https://doi.org/10.1044/2018_AJA-17-0061

Chui, J. C. H., & Ma, E. P. M. (2018). The impact of dysphonic voices on children's comprehension of spoken language. *Journal of Voice*, *33*(5), 801.e7–801.e16. https://doi.org/10.1016/j.jvoice.2018.03.004

Cooke, M. P. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, *119,* 1562–1573. https://doi.org/10.1121/1.2166600

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*(1), 414–427. https://doi.org/10.1121/1.2804952

Crandell, C. C., & Smaldino, J. J. (1996). Speech perception in noise by children for whom English is a second language. *American Journal of Audiology*, *5*(3), 47–51. https://doi.org/10.1044/1059-0889.0503.47

Deutsches Institut für Normung e.V. (2019). *Sound system equipment – Part 16: Objective rating* *of speech intelligibility by speech transmission index* (IEC 100/3202/CDV:2019); German and English version prEN IEC 60268-16:2019.

Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America*, *66*(3), 651–653. https://doi.org/10.1121/1.383691

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech and Language, 28*(2), 580–597. http://dx.doi.org/10.1016/j.csl.2013.07.005

Hirano, M. (1981). Psycho-acoustic evaluation of voice. In G. Arnold, F. Winckel, & B. Wyke (Eds.), *Disorders of human communication* (pp. 81–84). New York: Springer.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception and Psychophysics*, *72*(5), 1218–1227. https://doi.org/10.3758/APP.72.5.1218

Houtgast, T., Steeneken, H. J. M., Ahnert, W., Braida, L. D., Drullman, R., Festen, J. M., … van Wijngaarden, S. J. (2002). *Past, present and future of the speech transmission index*. Soesterberg, The Netherlands: TNO Human Factors.

Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology*, *49*(12), 928–932. https://doi.org/10.3109/14992027.2010.520036

Ishikawa, K., Nudelman, C., Park, S., & Ketring, C. (2020). Perception and acoustic studies of vowel intelligibility in dysphonic speech. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2019.12.022

Jaeger, T. F. (2008). Memory and language categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jamieson, D. G., Kranjc, G., Yu, K., & Hodgetts, W. E. (2004). Speech intelligibility of young school-aged children in the presence of real-life classroom noise. *Journal of the American Academy of Audiology*, *517*(15), 508–517. https://doi.org/10.3766/jaaa.15.7.5

Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research*, *43*(1), 144–157. https://doi.org/10.1044/jslhr.4301.144

Khomsi, A. (2001). *ELO: Évaluation du langage orale* [Measurement instrument]. Toronto, ON: ECPA Pearson.

Klatte, M., Lachmann, T., & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise and Health*, *12*(49), 270–282. https://doi.org/10.4103/1463-1741.70506

Korkman, M., Kirk, U., & Kemp, S. L. (2007). *NEPSY-II*, 2nd ed. [Measurement instrument]. San Antonio, TX: Harcourt Assessment.

Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. *The American Statistician*, *34*(4), 216–221.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Lombard, E. (1911). Le signe de l'elevation de la voix. [The sign of the raised voice.] *Annales des Maladies de l'Oreille et du Larynx, 37*, 101–119.

Lyberg-Åhlander, V., Brännström, K. J., & Sahlén, B. S. (2015). On the interaction of speakers' voice quality, ambient noise and task complexity with children's listening comprehension and cognition. *Frontiers in Psychology*, *6*, 871. https://doi.org/10.3389/fpsyg.2015.00871

Lyberg-Åhlander, V., Haake, M., Brännström, J., Schötz, S., & Sahlén, B. (2015). Does the speaker's voice quality influence children's performance on a language comprehension test? *International Journal of Speech-Language Pathology*, *17*(1), 63–73. https://doi.org/10.3109/17549507.2014.898098

Lyberg-Åhlander, V., Holm, L., Kastberg, T., Haake, M., Brännström, K. J., & Sahlén, B. (2015). Are children with stronger cognitive capacity more or less disturbed by classroom noise and dysphonic teachers? *International Journal of Speech-Language Pathology*, *17*(6), 577–588. https://doi.org/10.3109/17549507.2015.1024172

Macchi, L., Descours, C., Girard, É., Guitton, É., Morel, C., Timmermans, N., & Boidein, F. (2012). *ELDP: Épreuve Lilloise de Discrimination Phonologique* [Measurement instrument]. Retrieved from http://orthophonie.univ-lille2.fr/stocks/stock-contents/epreuve-lilloise-de-discrimination-phonologique.html

Martins, R. H. G., Pereira, E. R. B. N., Hidalgo, C. B., & Tavares, E. L. M. (2014). Voice disorders in teachers. A review. *Journal of Voice*, *28*(6), 716–724. https://doi.org/10.1016/j.jvoice.2014.02.008

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice*, *24*(5), 540–555. https://doi.org/10.1016/j.jvoice.2008.12.014

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McCreery, R. W., & Stelmachowicz, P. G. (2013). The effects of limited bandwidth and noise on verbal processing time and word recall in normal-hearing children. *Ear and Hearing, 34*(5), 585–591. https://doi.org/10.1097/AUD.0b013e31828576e2

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95–112. https://doi.org/10.1016/j.jecp.2017.04.006

McGarrigle, R., Gustafson, S. J., Hornsby, B. W., & Bess, F. H. (2019). Behavioral measures of listening effort in school-age children: Examining the effects of signal-to-noise ratio, hearing loss, and amplification. *Ear and Hearing*, *40*(2), 381–392. https://doi.org/10.1097/AUD.0000000000000623

Medwetsky, L. (2011). Spoken Language Processing Model: Bridging auditory and language processing to guide assessment and intervention. *Language, Speech, and Hearing Services in Schools*, *42*, 286–296. https://doi.org/10.1044/0161-1461(2011/10-0036)

Morsomme, D., Minell, L., & Verduyckt, I. (2011). Impact of teachers' voice quality on children's language processing skills. *VOCOLOGIE: Stem en Stemstoornissen, July*, 9–15.

Morton, V., & Watson, D. R. (2001). The impact of impaired vocal quality on children's ability to process spoken language. *Logopedics Phoniatrics Vocology*, *26*, 17–25. https://doi.org/10.1080/14015430118232

Nakeva von Mentzer, C., Sundström, M., Enqvist, K., & Hällgren, M. (2018). Assessing speech perception in Swedish school-aged children: Preliminary data on the Listen-Say test. *Logopedics Phoniatrics Vocology*, *43*(3), 106–119.

https://doi.org/10.1080/14015439.2017.1380076

Nirme, J., Haake, M., Lyberg-Åhlander, V., Brännström, J., & Sahlén, B. (2019). A virtual speaker in noisy classroom conditions: Supporting or disrupting children's listening comprehension? *Logopedics Phoniatrics Vocology*, *44*(2), 79–86. https://doi.org/10.1080/14015439.2018.1455894

Osman, H., & Sullivan, J. R. (2014). Children's auditory working memory performance in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, 57(4), 1503–1511. https://doi.org/10.1044/2014_JSLHR-H-13-0286

Peng, J., Zhang, H., & Yan, N. (2016). Effect of different types of noises on Chinese speech intelligibility of children in elementary school classrooms. *Acta Acustica united with Acustica, 102*(5), 938–944. https://doi.org/10.3813/AAA.919008

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... & Naylor, G. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37,* 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Prodi, N., & Visentin, C. (2015). Listening efficiency during lessons under various types of noise. *Journal of the Acoustical Society of America*, *138*(4), 2438–2448. https://doi.org/10.1121/1.4932053

Prodi, N., Visentin, C., Borella, E., Mammarella, I. C., & Di Domenico, A. (2019). Noise, age and gender effects on speech intelligibility and sentence comprehension for 11- to 13-year-old children in real classrooms. *Frontiers in Psychology*, *10*, 2166. https://doi.org/10.3389/fpsyg.2019.02166

Prodi, N., Visentin, C., Peretti, A., Griguolo, J., & Bartolucci, G. B. (2019). Investigating listening effort in classrooms for 5- to 7-year-old children. *Language, Speech, and*

Hearing Services in Schools, *50*(2), 196–210. https://doi.org/10.1044/2018_LSHSS-18-0039

Quené, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1–2), 103–121. https://doi.org/10.1016/j.specom.2004.02.004

R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. Retrieved from https://www.R-project.org/

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, 45(2), 250–267. https://doi.org/10.17105/SPR45-2.250-267

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Rogerson, J., & Dodd, B. (2005). Is there an effect of dysphonic teachers' voices on children's processing of spoken language? *Journal of Voice*, *19*(1), 47–60. https://doi.org/10.1016/j.jvoice.2004.02.007

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. https://doi.org/10.3389/fnsys.2013.00031

Sahlén, B., Haake, M., von Lochow, H., Holm, L., Kastberg, T., Brännström, K. J., & Lyberg-Åhlander, V. (2017). Is children's listening effort in background noise influenced by the speaker's voice quality? *Logopedics, Phoniatrics, Vocology*, *43*(2), 47–55.

https://doi.org/10.1080/14015439.2017.1324914

Schiller, I. S., Remacle, A., & Morsomme, D. (2019a). Imitating dysphonic voice: A suitable technique to create speech stimuli for spoken language processing tasks? *Logopedics Phonatrics Vocology*. https://doi.org/10.1080/14015439.2019.1659410

Schiller, I. S., Remacle, A., & Morsomme, D. (2019b). NODYS: NOrmophonic and DYsphonic Speech samples. *Mendeley Data, v1*. https://doi.org/10.17632/g2fmkw8t85.1

Schoentgen, J. (2006). Vocal cues of disordered voices: An overview. *Acta Acustica united with Acustica*, *92*(5), 667–680.

Schwartz, S. R., Cohen, S. M., Dailey, S. H., Rosenfeld, R. M., Deutsch, E. S., Gillespie, M. B., ... & McMurray, J. S. (2009). Clinical practice guideline: Hoarseness (dysphonia). *Otolaryngology – Head and Neck Surgery*, *141*(1_suppl), 1–31. https://doi.org/10.1016/j.otohns.2009.06.744

Silva, L. T., Oliveira, I. S., & Silva, J. F. (2016). The impact of urban noise on primary schools. Perceptive evaluation and objective assessment. *Applied Acoustics*, *106*, 2–9. https://doi.org/10.1016/j.apacoust.2015.12.013

Sullivan, J. R., Osman, H., & Schafer, E. C. (2015). The effect of noise on the relationship between auditory working memory and comprehension in school-age children. *Journal of Speech, Language, and Hearing Research*, *58*(3), 1043–1051. https://doi.org/10.1044/2015_JSLHR-H-14-0204

Visentin, C., & Prodi, N. (2018). A matrixed speech-in-noise test to discriminate favorable listening conditions by means of intelligibility and response time results. *Journal of Speech, Language, and Hearing Research*, *61*(6), 1497–1516. https://doi.org/10.1044/2018_JSLHR-H-17-0418

von Lochow, H., Lyberg-Åhlander, V., Sahlén, B., Kastberg, T., & Brännström, K. J. (2018). The effect of voice quality and competing speakers in a passage comprehension task: Perceived effort in relation to cognitive functioning and performance in children with normal hearing. *Logopedics Phoniatrics Vocology*, *43*(1), 32–41. https://doi.org/10.1080/14015439.2017.1307446

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing*, *37*, 35S–43S. https://doi.org/10.1097/AUD.0000000000000310

Zhang, Y. X., Barry, J. G., Moore, D. R., & Amitay, S. (2012). A new test of attention in listening (TAIL) predicts auditory performance. *PLoS ONE*, *7*(12), e53502. https://doi.org/10.1371/journal.pone.005350

Figures

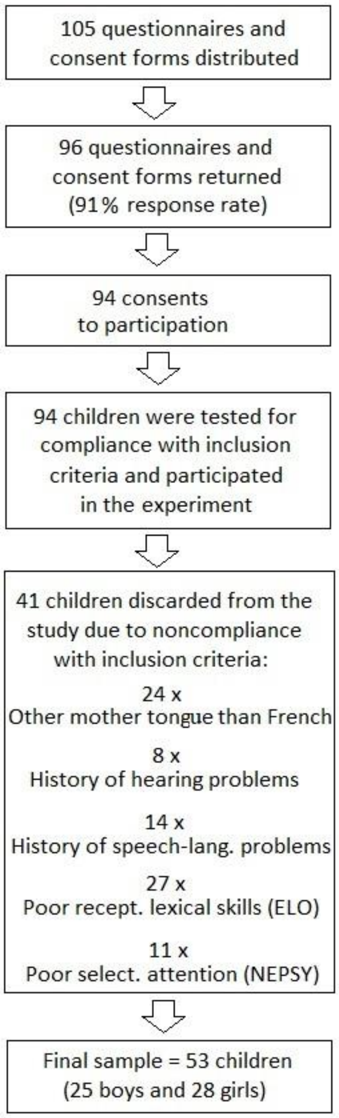

Figure 1. Flowchart presenting the recruitment of participants and selection of the final sample.

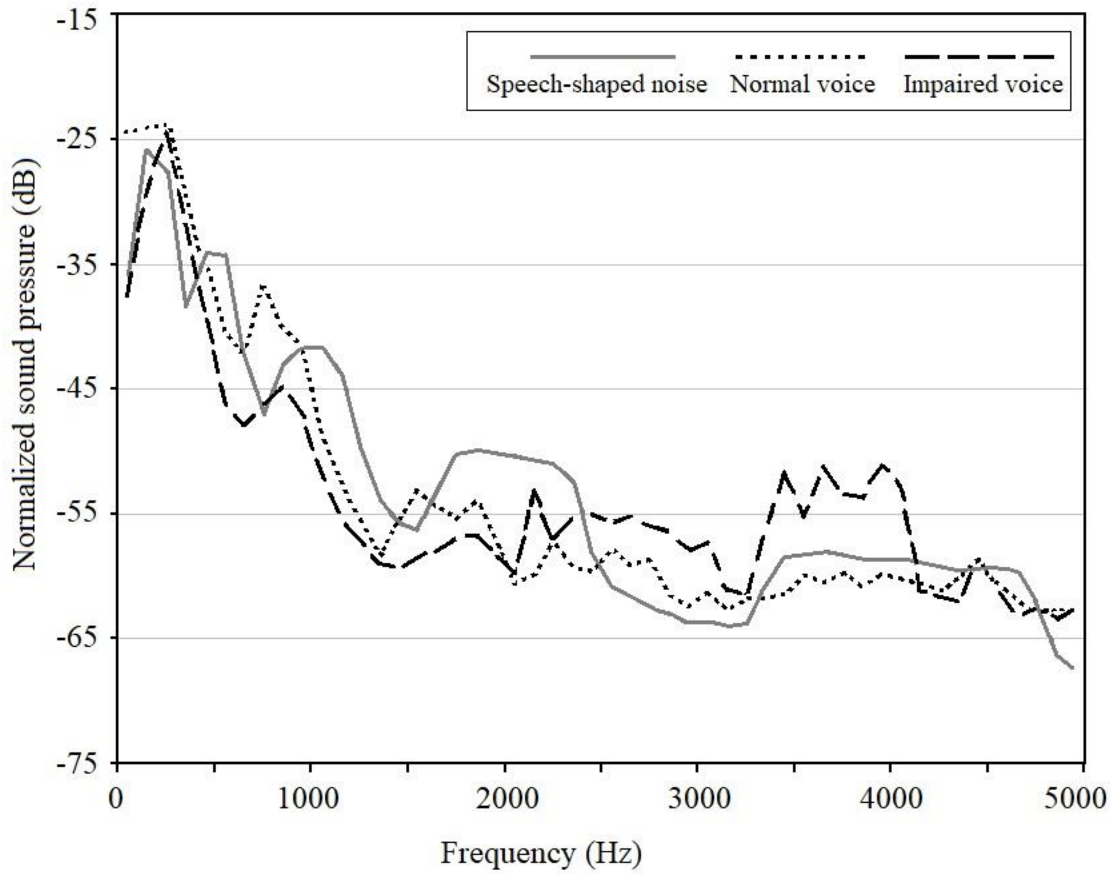

Figure 2. Long-term average spectra of the normal voice, impaired voice, and speech-shaped noise. Signals were normalized to a mean level (root-mean-square) of 65 dB.

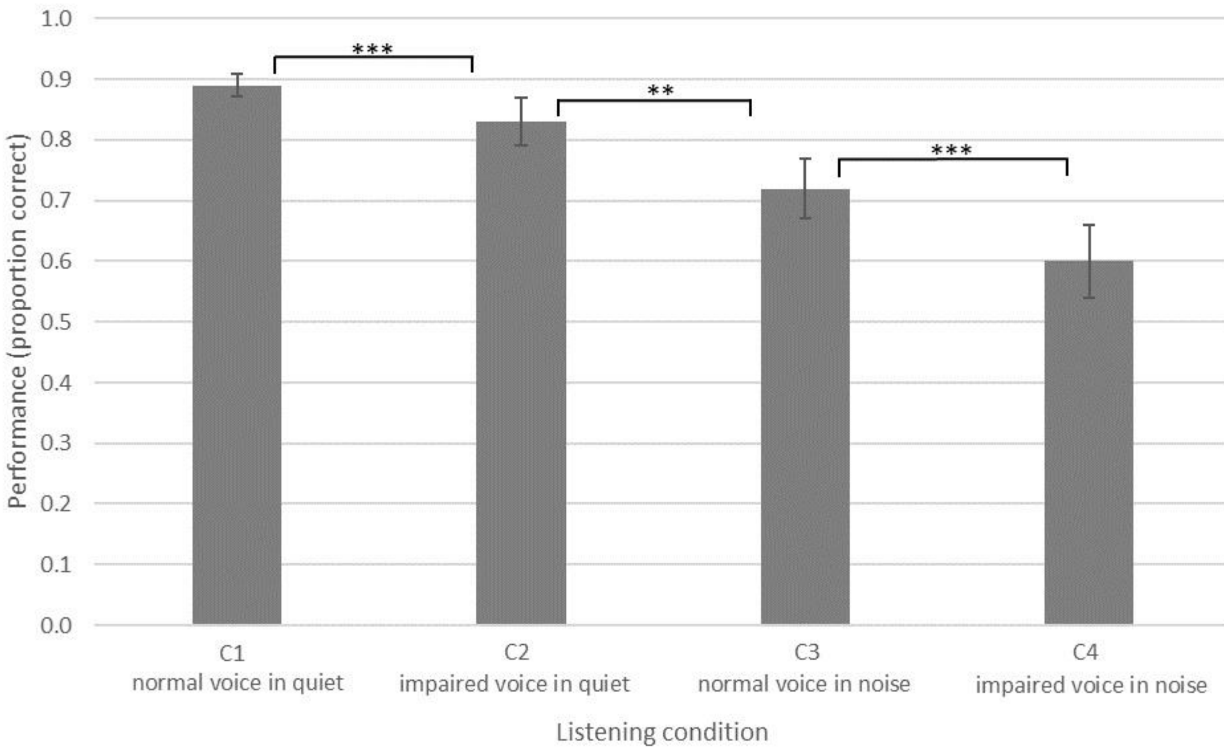

Figure 3. Mean speech perception performance as a function of listening condition. Performance measured as probability of correct responses. Error bars represent standard errors (*SE*). ***$p <$ .001.

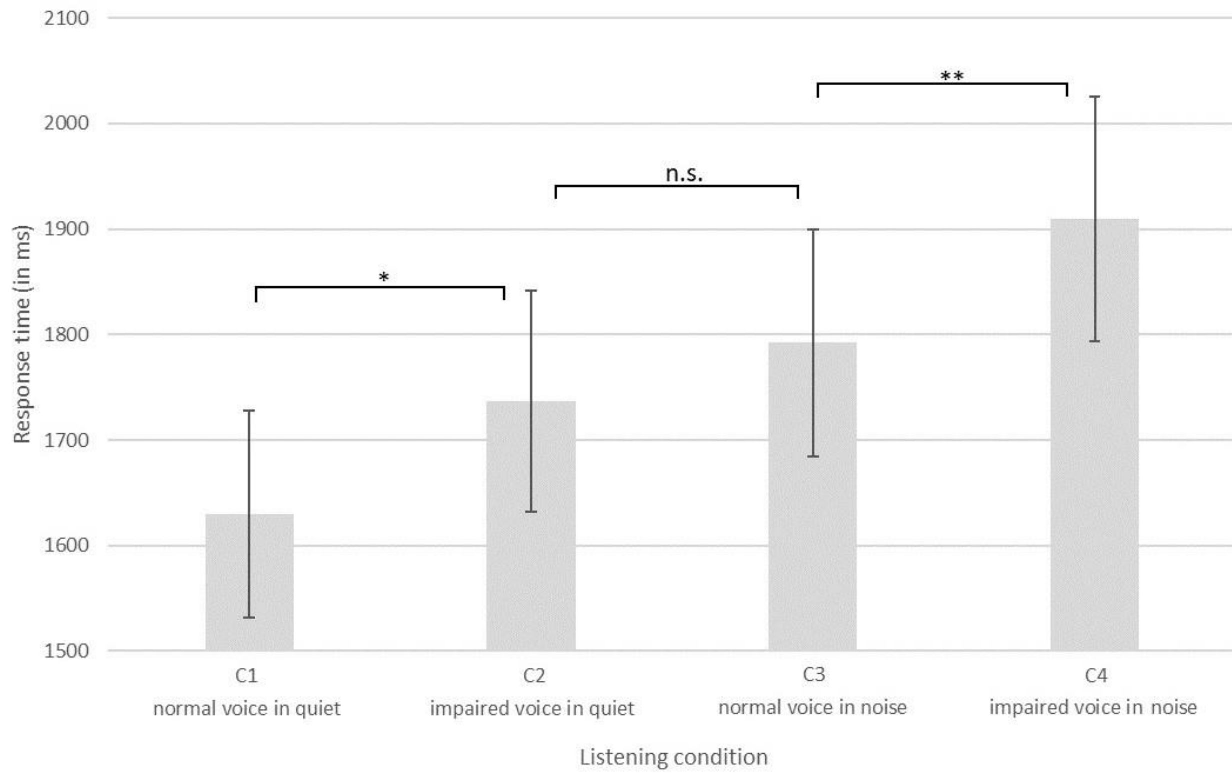

Figure 4. Mean response time in speech perception task as a function of listening condition. Error bars represent standard errors (*SE*). *$p < .05$, **$p < .01$, and ***$p < .001$.

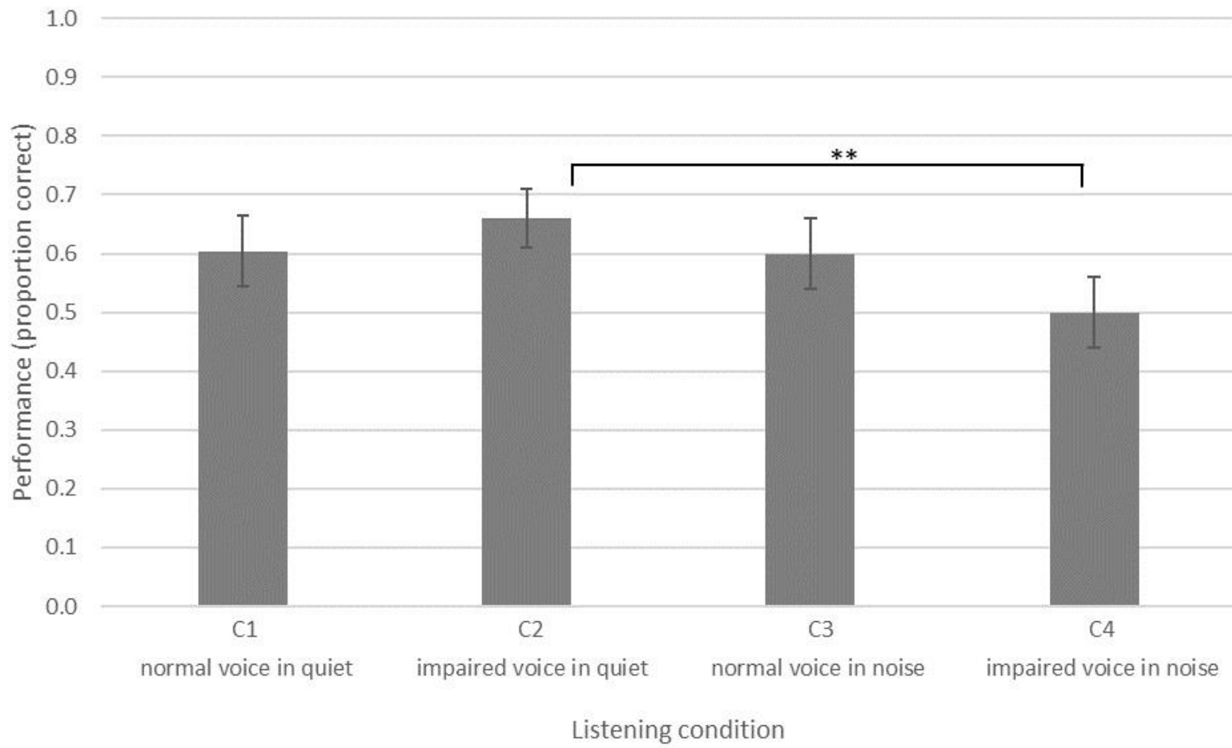

Figure 5. Mean listening comprehension performance as a function of listening condition. Performance measured as probability of correct responses. Error bars represent standard errors (*SE*). $^{**}p < .01$.

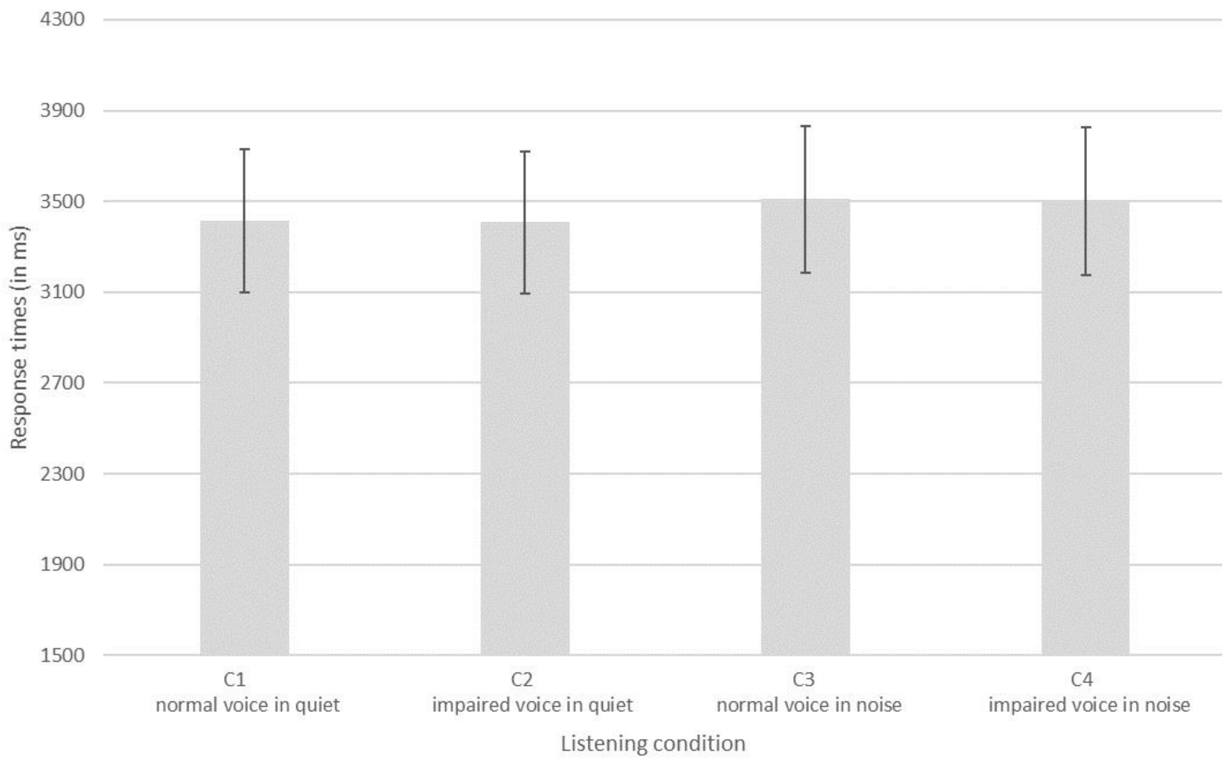

Figure 6. Mean response time in listening comprehension as a function of listening condition. Error bars represent standard errors (*SE*).

Tables

Table 1

GLMM Results for the Speech Perception Task in Terms of Performance and Response Time

Fixed factor	**Performance**				**Response time**			
	β	**95% CI**	***z***	***p***	**β**	**95% CI**	***t***	***p***
Noise (SSN vs. no added noise)	–1.16	[–1.40, –0.91]	–9.16	< .001	0.1	[0.06, 0.13]	5.14	< .001
Voice quality (impaired vs. normal)	–0.55	[–0.78, –0.31]	–4.5	< .001	0.06	[0.03, 0.1]	3.52	< .001

Note. Performance measured as probability of correct responses. Response times for correct trials measured in milliseconds.

β = fixed effect coefficient, CI = confidence interval.

Table 2

GLMM Results for the Listening Comprehension Task in Terms of Performance and Response Time

Fixed factor	**Performance**				**Response time**			
	β	**95% CI**	***z***	***p***	**β**	**95% CI**	***t***	***p***
Noise (SSN vs. no added noise)	–.03	[–0.4, –0.34]	–0.17	.863	0.03	[–0.05, 0.1]	0.72	.47
Voice quality (impaired vs. normal)	–.23	[–0.14, –0.60]	–1.21	.226	0.0	[–0.07, 0.07]	–0.05	.957
Noise x voice quality	–.60	[–1.13, 0.09]	–2.28	.023	–	–	–	–

Note. Performance measured as probability of correct responses. Response times for correct trials measured in milliseconds.

β = fixed effect coefficient, CI = confidence interval.