AJSLP

Research Article

Effect of Noise on Speech Intelligibility and Perceived Listening Effort in Head and Neck Cancer

Tanya L. Eadie,[a] Holly Durr,[a] Cara Sauder,[a] Kathleen Nagle,[b]
Mara Kapsner-Smith,[a] and Kristie A. Spencer[a]

Purpose: This study (a) examined the effect of different levels of background noise on speech intelligibility and perceived listening effort in speakers with impaired and intact speech following treatment for head and neck cancer (HNC) and (b) determined the relative contribution of speech intelligibility, speaker group, and background noise to a measure of perceived listening effort.

Method: Ten speakers diagnosed with nasal, oral, or oropharyngeal HNC provided audio recordings of six sentences from the Sentence Intelligibility Test. All speakers were 100% intelligible in quiet: Five speakers with HNC exhibited mild speech imprecisions (speech impairment group), and five speakers with HNC demonstrated intact speech (HNC control group). Speech recordings were presented to 30 inexperienced listeners, who transcribed the sentences and rated perceived listening effort in quiet and two levels (+7 and +5 dB SNR) of background noise.

Results: Significant Group × Noise interactions were found for speech intelligibility and perceived listening effort. While no differences in speech intelligibility were found between the speaker groups in quiet, the results showed that, as the signal-to-noise ratio decreased, speakers with intact speech (HNC control) performed significantly better (greater intelligibility, less perceived listening effort) than those with speech imprecisions in the two noise conditions. Perceived listening effort was also shown to be associated with decreased speech intelligibility, imprecise speech, and increased background noise.

Conclusions: Speakers with HNC who are 100% intelligible in quiet but who exhibit some degree of imprecise speech are particularly vulnerable to the effects of increased background noise in comparison to those with intact speech. Results have implications for speech evaluations, counseling, and rehabilitation.

One of the most significant difficulties experienced by individuals diagnosed and treated for head and neck cancer (HNC) relates to deficits in verbal communication. When cancer involves the oral cavity, nasal cavity, and/or the pharynx, the tumor and/or the treatment may alter the structure, strength, range of motion, and/or precision of the articulators. Difficulties with speech may lead to social isolation and may affect relationships, the ability to return to work, and quality of life (Dwivedi et al., 2009). As a result, speech outcomes are important for measuring the impact and success of HNC treatment and for providing directions for follow-up care.

Traditional methods of assessing speech after HNC include both objective and subjective methods. Objective speech measures often include acoustic and physiological measures, whereas subjective measures may include perceptual judgments made by clinicians, as well as patient-reported outcomes. One additional method for assessing the impact of a speech disorder includes judgments made by unfamiliar communication partners. These measures typically include auditory–perceptual ratings of speech intelligibility or acceptability (Doyle & Eadie, 2005).

While measures of speech intelligibility may provide important information about the severity of a speech deficit secondary to HNC, the degree of speech signal disruption perceived by the listener may differ from the cognitive effort needed to complete this processing. In addition, real-life

---

[a]Department of Speech and Hearing Sciences, University of Washington, Seattle
[b]Department of Speech-Language Pathology, Seton Hall University, South Orange, NJ

Correspondence to Tanya L. Eadie: teadie@u.washington.edu

Portions of this paper were presented at the Conference on Motor Speech, February 20–23, 2020, Santa Barbara, CA.

Editor-in-Chief: Deanna Britton
Editor: Jessica E. Huber

Received May 26, 2020
Revision received August 13, 2020
Accepted September 22, 2020
https://doi.org/10.1044/2020_AJSLP-20-00149

Publisher Note: This article is part of the Special Issue: Select Papers From the 2020 Conference on Motor Speech.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

communication exchanges are often not performed in quiet environments. The effect of background noise is particularly important to understand because it not only impacts communicative exchanges in typical speakers, but there is preliminary evidence to show that it might differentially penalize those with other types of speech disorders (Chiu & Neel, 2020; Ishikawa et al., 2017; McAuliffe et al., 2009). This factor must be considered when interpreting the meaningfulness of any outcome measure in individuals with speech disorders.

Speech Intelligibility in HNC

The most common measure of speech performance in individuals treated for HNC includes ratings of speech intelligibility performed by listeners, such as unfamiliar communication partners (Dwivedi et al., 2009). Speech intelligibility has been defined as "the degree to which the speaker's intended message is recovered by the listener" (Kent et al., 1989, p. 484). Similar to approaches used in dysarthria, speech intelligibility after HNC is typically measured using transcription to determine a percentage of recognizable words within a speech sample (Constantinescu et al., 2017), perceived severity using rating scales (Borggreven et al., 2005), or speech recognition software (Windrich et al., 2008).

Studies examining the speech intelligibility of patients with HNC often include participants with specific tumor sites or they focus on outcomes related to specific treatments, such as surgical approaches or chemoradiation protocols (Dwivedi et al., 2009; Jacobi et al., 2010). While many studies have documented the effect of total laryngectomy (secondary to laryngeal cancer) on speech intelligibility and other related outcomes (e.g., Eadie et al., 2016; Holley et al., 1983), fewer studies have investigated the effects in other types of HNC or other treatments for HNC beyond total laryngectomy (Dwivedi et al., 2009). Across these studies, results indicate that there is wide variability in speech performance.

General trends indicate that intelligibility is typically higher preceding treatment of HNC, followed by a decline after treatment, which is then usually followed by long-term improvement after approximately 1 year (Jacobi et al., 2010). For example, patients with oral cancer treated using surgical resection of the tongue with or without a free flap inevitably show a reduction in speech intelligibility because of the important role the anterior portion of the tongue plays in articulatory precision (Hutcheson & Lewin, 2013). Single-word intelligibility was previously shown to decrease from a mean of 94% intelligible to 83% intelligible at 1 month posttreatment in a study of patients with oral cancer treated with surgery (Constantinescu et al., 2017). In another study, mean word intelligibility scores were 78% several months after surgery in patients who had undergone hemiglossectomies (Loewen et al., 2010). Similar observations have also been made in individuals treated using concomitant chemoradiation therapy, with deteriorations during treatment and progressive improvement after treatment (Jacobi et al., 2010; van der Molen et al., 2012). Importantly, typical values for voice and speech were not reached for individuals who had undergone chemoradiation protocols, even in the long term. All of these types of treatments must be considered when developing assessment protocols and identifying individuals who could benefit from rehabilitation.

It is clear that speech intelligibility may be affected in many patients with HNC beyond laryngeal cancer and total laryngectomy and that long-term performance may be widely variable. While measures of speech intelligibility are important for capturing the severity of a speech disorder, they may be limited. For example, consider a speaker who has undergone HNC treatment who is 100% intelligible but who may have some residual speech imprecisions (e.g., distortions of phonemes). While a communication partner may be able to understand all of the words in a conversation with this individual, the intelligibility score may not reflect the amount of work the listener expends (or perceives to expend) when deciphering the message. That burden is better captured by another measure: perceived listening effort.

Perceived Listening Effort in HNC

Listening effort is considered an umbrella term for two types of effort. *Processing effort* corresponds with the *objective* extent to which processing resources are consumed to accomplish a task and is measured instrumentally (Pichora-Fuller et al., 2016). On the other hand, *perceived listening effort* is the term used for "subjective estimates of how taxing a listening task is or was" (Lemke & Besser, 2016, p. 77S) and is fundamentally measured through questionnaires or perceptual rating scales. Perceived listening effort may be a stronger indicator of communicative success than speech intelligibility, in that it may involve consideration of the speech signal, the listener, and even interactions between them (Olmstead et al., 2020; Pichora-Fuller et al., 2016). Examining relationships between intelligibility and perceived listening effort and other factors that contribute to perceived listening effort is therefore important for understanding communicative success in individuals with speech disorders.

The majority of research on perceived listening effort has focused on how hearing loss affects communication (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). However, the usefulness of perceived listening effort as an outcome measure has also been highlighted in studies investigating outcomes in individuals with speech disorders as they communicate with typical hearing partners. In one study of dysarthric speakers (Whitehill & Wong, 2006), inexperienced listeners transcribed decontextualized sentences and judged perceived listening effort using rating scales. Overall, results showed that perceived listening effort and intelligibility were strongly and negatively correlated ($r_s = -.95$); speakers with higher intelligibility were judged less effortful to understand. One important and surprising finding from this study involved three speakers with high speech intelligibility (more than 85%): These speakers also demonstrated relatively high mean ratings of perceived listening effort. The results indicate that, while intelligibility and perceived listening effort are usually inversely related, they may be distinct concepts for individual speakers.

Other researchers have found similar results for speakers with other types of dysarthria (Beukelman et al., 2011), including those with Parkinson's disease (Chiu & Neel, 2020; N. Miller et al., 2007) and cerebral palsy (Landa et al., 2014), as well as individuals who have undergone total laryngectomy (Nagle & Eadie, 2012) and who use electrolaryngeal speech (Nagle & Eadie, 2018). Across studies, results have shown strong relationships between speech intelligibility and perceived listening effort. Yet, all authors reported wide variability of speech intelligibility scores, especially for speakers in the midrange of the perceived listening effort scales.

Results from all of these studies indicate that perceived listening effort is generally higher when communicating with someone with a speech impairment. However, even speakers with typical or high speech intelligibility may require increased perceived listening effort, suggesting that perceived listening effort may not be reflected in speech intelligibility measures alone. Beyond the severity of a speaker's impairment (i.e., precision of speech, voice quality, resonance, prosody), other factors related to the communication partner/listener (e.g., fatigue, working memory, hearing status, familiarity with the speaker), the stimuli (e.g., lexical density, predictability of the words), and the environment (e.g., competing noise) may affect perceived listening effort (Borrie et al., 2012; Chiu & Neel, 2020; Landa et al., 2014). Among these factors, the listening environment warrants further attention.

Considerations of Background Noise

While speech intelligibility and perceived listening effort are valuable when measured in controlled (quiet) research conditions, findings in these contexts may be limited due to the lack of generalizability to everyday settings that often include background noise. As a result, a growing number of studies have investigated the effect of background noise on speech outcome measures. For example, it is well known that the presence of noise adversely affects typical speech intelligibility (Sperry et al., 1997; Van Engen & Bradlow, 2007). However, for speakers with communication disorders, there is limited but growing evidence to suggest that such challenges are compounded in noise. McAuliffe et al. (2009) first explored the effect of background noise on speech intelligibility in three speakers with dysarthria and three control speakers in quiet and in background noise. Results revealed that intelligibility disproportionately declined with the addition of background noise, and this decline was only observed for the speakers with dysarthria in relatively low-noise conditions (+6 dB SNR).

Chiu and Forrest (2018) and Chiu and Neel (2020) also investigated speech intelligibility in those with Parkinson's disease in quiet and noisy conditions (multitalker babble set at a signal-to-noise ratio [SNR] of +6 dB). As expected, the authors found that words produced by speakers with Parkinson's disease were more difficult for listeners to recognize than healthy, older speakers. However, listening in the presence of noise was also significantly more penalizing for speakers with Parkinson's disease than healthy speakers. Relative to listening in quiet, there was a 43% decrease in speech intelligibility for speakers with Parkinson's disease in noise compared to a 21% decrease in healthy speakers in noise (Chiu & Forrest, 2018).

Similar effects have been reported in speakers who use different types of alaryngeal speech (Eadie et al., 2016; Holley et al., 1983; McColl et al., 1998). While methods have been variable across studies, the effect of noise has been particularly noticeable at relatively similar levels (e.g., +5 dB SNR in McColl et al., 1998; +6 dB SNR in Eadie et al., 2016). Ishikawa et al. (2017) also showed similar effects among those with laryngeal-based voice disorders. Speakers with dysphonia were significantly less intelligible than typical controls in two levels of background noise (+5 and +0 dB SNR), despite the fact that no differences were found between the speakers with dysphonia and healthy control speakers in quiet.

Purpose of the Study

Speakers with communication disorders may be more vulnerable to the effects of varying levels of background noise (Chiu & Forrest, 2018; Chiu & Neel, 2020; Eadie et al., 2016; Ishikawa et al., 2017). This is an important factor to consider because many speakers, including those treated for HNC, report particular difficulty communicating in noisy everyday environments (Baylor et al., 2011). Yet, to date, these effects have not been investigated in speakers with HNC, beyond total laryngectomy.

The main purpose of this study was to determine the effect of different levels of background noise (quiet, +7 dB SNR, +5 dB SNR) on speech intelligibility and perceived listening effort in speakers treated for nonlaryngeal HNC. These effects were investigated in a group of speakers treated for HNC who exhibited mild speech impairments and a group of speakers treated for HNC exhibiting intact speech who served as an HNC control group. This study also explored the unique contribution of noise in these two groups of speakers, as well as how interactions between these variables, might predict a measure of perceived listening effort, above and beyond speech intelligibility. This question was intended to provide additional information about the construct of perceived listening effort as a more global speech outcome measure, beyond speech intelligibility. Results have implications for counseling HNC patients and for developing strategies for communicating in adverse environments.

Method

This study included the following participants: (a) speakers who were diagnosed and treated for nonlaryngeal HNC who were (1) 100% intelligible in quiet but who had mild speech imprecisions or (2) 100% intelligible in quiet but who demonstrated intact speech (i.e., HNC control group) and (b) inexperienced listeners.

All participants were native speakers of English. The University of Washington Human Subjects Committee approved the procedures used in this study. Participants were paid for their participation. A subsample of participants in the HNC group was also enrolled in a broader investigation of communication outcomes.

Participants

Speakers

Speaker samples were selected from among a database of recordings that included adults who were previously diagnosed and treated for various types of HNC. They were recruited using a variety of methods such as local support groups, professional contacts, and those who underwent active treatment (surgery and/or [chemo]radiation) at the University of Washington Medical Center. Inclusion criteria for this study were adults (18+ years of age) with HNC diagnoses who had the potential to affect speech intelligibility, including those diagnosed with oral cancer (anterior two thirds of the tongue, floor of mouth, with or without extension to the mandible, the lips, or to the buccal mucosa), oropharyngeal cancer (posterior one third [base] of the tongue, soft palate, tonsils, or pharyngeal walls), or hypopharyngeal cancer (the piriform sinuses, lateral and posterior pharyngeal walls extending from the hyoid bone down to the posterior surfaces of the larynx). All participants used speech as their primary method of communication. There were no exclusions related to HNC treatment type, stage of the disease, or time posttreatment.

Participants were excluded if they had an HNC diagnoses that might primarily affect phonatory function (vs. articulation and resonance), such as diagnoses of laryngeal cancers; had previously altered anatomy of the upper aerodigestive tract, pre-existing speech/voice impairments unrelated to the tumor, or neurological disorders (e.g., Parkinson's disease, stroke); or were nonnative English speakers.

Listeners

Thirty inexperienced listeners were recruited from the community, ranging from 19 to 51 years of age (M = 26.3 years, SD = 7.08), and included 26 women and four men. While more women than men participated, no effect of gender or age has been found in previous studies investigating ratings of speech intelligibility (Pennington & Miller, 2007). We also restricted ages to those who were < 55 years of age because of some reported differences in perceived listening effort as a function of age (Larsby et al., 2005). Inexperienced listeners were those with no previous exposure or coursework examining the speech of those with HNC, or experience in performing perceptual ratings of speakers with HNC. Inexperienced listeners were selected because their judgments may reflect unfamiliar communication partners that individuals with HNC encounter in their everyday lives. All listeners were native speakers of English and passed a hearing screening at 25 dB SPL for the frequencies of 500, 1000, 2000, and 4000 Hz.

Data Collection

Speaker participants completed a variety of measures that included demographic information (e.g., age, ethnicity) and information related to HNC medical diagnosis and treatment. For those who had undergone treatment at the University of Washington Medical Center, medical information was extracted from the medical chart.

Recording and Selection of Speech Samples

In this study, speakers provided speech recordings that included sentences from the Sentence Intelligibility Test (SIT; Yorkston et al., 1996). Six sentences of increasing length from each speaker were included to assess speech intelligibility and perceived listening effort (consisting of five, seven, nine, 11, 13, and 15 words in length, resulting in 60 words per speaker), consistent with previous studies (Eadie et al., 2016). The six sentences included in this study were different for each speaker to control for learning effects across the listeners.

Speech samples were recorded in a room without background noise, using a headset microphone (AKG-C20, AKG Acoustics) with an offset mouth-to-microphone distance of 3 in. The headset was connected to a portable digital audio recorder (Zoom H6) and recorded at a sampling rate of 44.1 kHz with 16-bit quantization. All speech samples were then transferred to a computer using a sound card and acoustic software. Once converted to WAV files using acoustic software (Sony Soundforge, Sony Creative Software, Inc.), the samples were segmented into individual SIT sentences (Yorkston et al., 1996).

Selection of Speech Samples

The final set of 10 speakers with HNC selected for this study is presented in Table 1. They included (a) five speakers (I1–I5) with HNC who were rated as having 100% intelligibility in quiet, but who demonstrated imprecise phonemes, as evaluated by three experienced speech-language pathologists and (b) five speakers (C1–C5) with HNC who were also rated as having 100% speech intelligibility, but who exhibited no imprecise phonemes in quiet (i.e., intact speech), as judged by the same three listeners. All three speech-language pathologists were experienced in assessing and treating patients with HNC.

To select the 10 speakers with HNC for this study, the three experienced speech-language pathologists first listened to SIT samples from all speakers in the database and answered the following questions: (a) Is this speaker 100% intelligible in quiet (i.e., can you understand every word this speaker says?), and (b) does this speaker have typical articulatory precision and resonance (i.e., the two speech subsystems that would be most affected in this subgroup of HNC)? To be eligible for either speaker group, clinicians needed to answer "yes" to the first question, thereby excluding any speakers in the database with unintelligible speech.

Table 1. Demographics of head and neck cancer (HNC) speakers with imprecise speech (I) and HNC control speakers with intact speech (C).

Speaker	Age	Sex	Ethnicity	Elapsed time[a]	Location of tumor(s)	Treatment
I1	55	M	(NR)[b]	3 mo.	Soft palate	Surgery, radiation
I2	73	F	White	43 mo.	Oral cavity	Surgery, radiation, chemotherapy
I3	73	M	White	65 mo.	Hard/soft palate	Surgery
I4	74	F	(NR)[b]	31 mo.	Oral cavity	Surgery, radiation
I5	70	M	White	31 mo.	Tongue	Surgery, other
C1	57	M	White	37 mo.	Oro/hypopharynx	Surgery, radiation, chemotherapy
C2	51	F	White	75 mo.	Oral cavity	Surgery, radiation
C3	67	M	White	5 mo.	Oropharynx	Surgery, radiation, chemotherapy
C4	30	F	Asian	8 mo.	Hypopharynx	Radiation, chemotherapy
C5	67	M	White	6 mo.	L tonsil (oropharynx)	Surgery, radiation

Note. M = male; F = female; L = left.

[a] Elapsed time in months (mo.) between the date of HNC diagnosis and collection of speech recordings. [b] No response (NR) provided for this item in the demographic questionnaire.

To be eligible for inclusion in the HNC control group, clinicians needed to answer "yes" to the second question, demonstrating that speakers in this group had typical articulatory precision and resonance (i.e., intact speech). Speakers with any perceived articulatory or resonatory imprecisions (e.g., imprecise consonants, distorted vowels, prolonged phonemes, hypernasal phonemes [nonnasals]) were then considered eligible for the group of speakers with HNC with mild speech impairments. Twelve speakers who met the criteria were initially selected for consideration in the study; a final set of 10 was chosen in an effort to match speakers between the two groups for age and gender, with five in each group. There was 100% consensus on the selection of the samples for meeting inclusion/exclusion criteria in the two speaker groups.

Preparation of Stimuli

Sentences from each of the included speakers were first equated for peak amplitude (normalized) at 71.2–71.5 dB SPL, consistent with the method outlined by Ishikawa et al. (2017). Five hundred milliseconds of silence were then added before and after the speech signal following the intensity normalization procedure. This set of sentences was saved as the recordings in "quiet" for all 10 speakers included in this study.

A second and third set of stimuli was then mixed with a four-talker babble (one male and three females; Audiotec of St. Louis) from the Sentences in Noise Test. The intensity of the speech was held constant, and the noise was added to achieve the selected SNR. Multitalker babble was used as the noise because previous research has found that meaningful speech competitors had a significantly more adverse effect on word recognition performance compared to nonmeaningful competitors (e.g., white noise; Sperry et al., 1997). In addition, multitalker babble is representative of the most challenging adverse listening environment encountered in everyday speech communication situations (Gilbert et al., 2013).

Two levels of noise were added to all of the stimuli, resulting in three different versions of the same six sentences for each speaker: quiet (as described above), one set at an SNR of +7 dB, and one set at an SNR of +5 dB. These noise levels were chosen to reflect daily communication scenarios in which background noise is present, though not higher than the level of the speaker; these levels have also been used to differentiate speakers in previous studies (Ishikawa et al., 2017; McAuliffe et al., 2009; McColl et al., 1998). All speech samples were entered into a custom-made software program, which allowed the listeners to transcribe the sentences (eliciting a speech intelligibility measure) and submit a rating of perceived listening effort.

The speech samples were then arranged into two different listening sets (see Table 2). Each list consisted of five speakers and their six corresponding sentences (i.e., a total of 30 sentences per list). Three sublists were then created for each master list, such that different levels of noise were distributed among the sublists, and the order of the noise condition was counterbalanced. This design was used to ensure that no listener would hear a repeated stimulus and that each stimulus item would be presented across each noise condition. The stimuli were also screened for predictability using methods outlined by Beverly et al. (2010) and results for each sentence from the SIT (M. Cannito, personal communication, March 2013). Mean predictability scores were not found to differ across the sublists. This approach was performed in an effort to control the predictability of the sentences from the SIT across the listening conditions and speaker groups.

Procedure for Listener Ratings

The listening experiment was conducted in a quiet room in the University of Washington Vocal Function Laboratory. In advance of presenting stimulus items, listeners received instructions and familiarization on the protocol for transcribing speech samples and determining ratings of perceived listening effort. Listeners were seated at a desktop computer and listened to stimuli at a consistent sound level

Table 2. Stimulus sublists for listening tasks.

List 1					List 2				
Speaker	Stimuli[a]	Group 1a	Group 1b	Group 1c	Speaker	Stimuli[a]	Group 2a	Group 2b	Group 2c
I2	11	Quiet	SNR +7	SNR +5	I5	32	Quiet	SNR +7	SNR +5
C2	25	SNR +7	SNR +5	Quiet	C5	34	SNR +7	SNR +5	Quiet
I1	2	SNR +5	Quiet	SNR +7	I4	37	SNR +5	Quiet	SNR +7
C1	20	Quiet	SNR +7	SNR +5	C4	10	Quiet	SNR +7	SNR +5
C3	6	SNR +7	SNR +5	Quiet	I3	42	SNR +7	SNR +5	Quiet

Note. I = speakers with head and neck cancer with imprecise speech; SNR = signal-to-noise ratio; C = head and neck cancer control speakers with intact speech.

[a] Each number corresponds to a set of six Sentence Intelligibility Test sentences of increasing length.

through headphones (Samson RH600). To avoid exposing the listeners to the same sentence multiple times, they were randomly assigned to make judgments of speakers on one of six sublists (see Table 2; *n* = 5 listeners performed ratings for each sublist).

For each speaker, samples of increasing length were presented as a block, consistent with the SIT protocol (Yorkston et al., 1996). For each sentence, listeners completed two tasks: (a) transcribe the words they heard and (b) provide a rating of perceived listening effort. The transcription procedure was consistent with the protocol described in the Sentence Intelligibility Test (Yorkston et al., 1996). After the presentation of each sentence, listeners were instructed to type exactly what they heard into a software program. They were allowed one opportunity to replay the recording, totaling a maximum of two exposures. If uncertain, they were asked to make their best guess.

Ratings of perceived listening effort for each sentence took place immediately following each transcription. In advance of the task, listeners were provided the following instructions: “Perceived listening effort is the amount of work, attention or concentration it takes you to understand a speech sample. It also has been described as the listener’s contribution to a conversational exchange. When rating perceived listening effort, try to focus on your own effort and reactions to the speech sample. Try to hear every word and rate effort accordingly” (Nagle & Eadie, 2012, p. 238). Ratings were made on 100-mm visual analog scales, where 0 = *easy/no effort* and 100 = *extremely effortful.* All ratings took place in a single session that lasted 30–45 min. Listeners took breaks as needed to reduce the effects of fatigue and error.

Data Analysis

Descriptive Results

Listeners’ transcriptions were compared with the target stimuli to determine speech intelligibility for each sentence. For each sentence, intelligibility was averaged across five listeners, using a total word phonemic match model scoring approach (Hustad & Cahill, 2003). A score was allocated to each sentence based on the proportion of correctly transcribed words to the total number of words in the target sentence. The average of listeners’ percent correct words was used to determine the mean intelligibility scores for each speaker, accounting for variation in background noise (i.e., scores based on 5 listeners per set × 60 words per speaker = 300 words per speaker for each noise condition).

Ratings of perceived listening effort were similarly averaged for each individual sentence at each level of background noise; these scores were then used to generate an aggregated score for each speaker as a function of each level of background noise (i.e., scores based on 5 listeners per set × 6 sentences = 30 ratings per speaker for each noise condition).

Reliability

Measures of intrarater reliability for intelligibility were not included in this study due to learning effects with presentation of a repeated sentence. To assess interrater reliability of transcriptions for each set of five speakers evaluated by five listeners in each listening subgroup, intraclass correlation coefficients (ICCs) were calculated using ICC model (2, *k*; Shrout & Fleiss, 1979). The ICCs for speech intelligibility across the listening groups ranged from .79 to .97, with an overall mean of .91. The ICCs for perceived listening effort across the listening groups ranged from .87 to .96, with an overall mean of .94. These levels are consistent with prior research and are acceptable levels for data analysis (Sussman & Tjaden, 2012).

Analysis of Experimental Questions

To determine the effect of speaker group (i.e., HNC speech imprecision vs. HNC control) and noise conditions (i.e., quiet, +7 dB SNR, +5 dB SNR) on speech intelligibility and perceived listening effort, two 2 × 3 analyses of variance (ANOVAs) were performed. Noise (three levels) and speaker group (two levels) were set as fixed effects, and listener was set as random effects. Post hoc analyses were calculated where appropriate to determine main effects, as well as any significant interactions. Effect sizes (Cohen’s *d*; Cohen, 1988) were also calculated to determine the strength of any observed effects at each noise level.

Multiple regressions with sequential predictor entry were performed to examine the relative contributions of speech intelligibility, speaker group, and the interaction between speaker and group and noise, to perceived listening effort. After controlling for listener group in Block 1, speech intelligibility was entered into Block 2. Speaker group was entered into the model in Block 3, controlling for other variables. Finally, the interaction between speaker group and noise level at +7 dB SNR compared to quiet and +5 dB SNR compared to quiet were entered in Block 4 to determine how much speaker group by noise level contributed to perceived listening effort above and beyond the main effects of speech intelligibility and speaker group. Blocks of variables were reported with changes in r^2. Regression coefficients and effect sizes (sr^2) were used to demonstrate the unique contribution of each variable to perceived listening effort, all others being held constant. Standard errors were reported as measures of precision. The significance level for the regression models was set at $p < .05$. All analyses were performed using IBM SPSS Statistics, Version 26.

Results

Effect of Background Noise on Speech Intelligibility and Perceived Listening Effort

Descriptive Results

The mean speech intelligibility (%) and perceived listening effort (mm) averaged across speech samples used in the listening experiment are shown in Table 3. The average intelligibility scores of the group of speakers with intact speech (HNC control speakers) were 99.90% in quiet (*SD* = 0.92), 86.93% for +7 dB SNR (*SD* = 18.08), and 81.08% for +5 dB SNR (*SD* = 23.67). The average intelligibility scores for speakers with HNC with imprecise speech were 97.23% for quiet (*SD* = 7.01), 72.38% for +7 dB SNR (*SD* = 30.80), and 56.49% for +5 dB SNR (*SD* = 33.77).

The average perceived listening effort ratings of the group of HNC control speakers with intact speech were 1.15 mm in quiet (*SD* = 3.69), 40.93 mm for +7 dB SNR (*SD* = 31.19), and 34.05 mm for +5 dB SNR (*SD* = 25.19). The average perceived listening effort scores for the group of speakers with HNC with imprecise speech were 10.17 mm for quiet (*SD* = 13.97), 51.82 mm for +7 dB SNR (*SD* = 30.07), and 70.55 mm for +5 dB SNR (*SD* = 27.02).

Table 3. Mean speech intelligibility (%) and perceived listening effort ratings (mm) across all noise conditions for head and neck cancer (HNC) control speakers with intact speech (C) and speakers with HNC with imprecise speech (I).

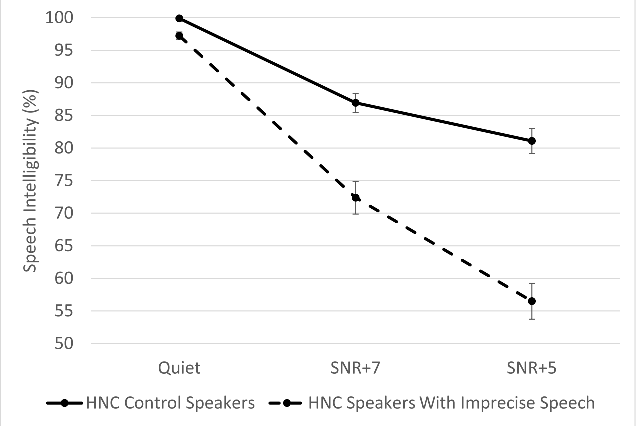
Speaker	Rating	Quiet	+7 dB SNR	+5 dB SNR
C1	Intelligibility	99.70 (1.66)	85.55 (18.70)	78.48 (21.63)
	Listening effort	0.87 (1.89)	23.17 (23.74)	33.17 (23.32)
C2	Intelligibility	100.00 (0)	86.09 (23.03)	82.82 (26.21)
	Listening effort	0.43 (1.38)	52.07 (29.65)	34.57 (27.73)
C3	Intelligibility	100.00 (0)	87.07 (16.61)	75.26 (27.43)
	Listening effort	1.07 (2.64)	54.23 (17.54)	46.90 (26.92)
C4	Intelligibility	100.00 (0)	92.95 (15.59)	91.74 (16.37)
	Listening effort	3.10 (7.16)	10.63 (17.83)	16.27 (19.50)
C5	Intelligibility	99.78 (1.22)	82.98 (14.94)	77.11 (22.90)
	Listening effort	0.30 (0.95)	64.53 (27.69)	39.37 (17.74)
I1	Intelligibility	95.88 (10.91)	39.71 (30.31)	30.87 (25.87)
	Listening effort	5.93 (8.08)	71.90 (21.55)	94.77 (5.73)
I2	Intelligibility	94.82 (13.34)	73.72 (27.59)	55.16 (32.53)
	Listening effort	21.07 (22.88)	48.20 (28.18)	56.33 (24.39)
I3	Intelligibility	99.27 (2.25)	70.34 (31.74)	55.08 (32.96)
	Listening effort	7.70 (9.08)	69.23 (29.97)	64.07 (29.52)
I4	Intelligibility	96.42 (7.85)	90.96 (10.67)	75.10 (27.65)
	Listening effort	4.73 (6.51)	41.93 (20.68)	68.67 (32.30)
I5	Intelligibility	99.78 (1.21)	87.18 (21.14)	66.25 (33.77)
	Listening effort	11.43 (18.97)	27.83 (24.90)	68.90 (18.56)

Note. The values are mean values across all sentences, averaged across five listeners per speaker per condition. The values in parentheses are standard deviations.

Effect of Background Noise on Speech Intelligibility

Mean intelligibility scores for the speaker groups are reported across all noise conditions in Figure 1. A 2 × 3 ANOVA was calculated to examine the effect of speaker group and background noise on speech intelligibility. Results revealed a significant Speaker Group × Noise interaction, $F(2, 894) = 17.92$, $p < .001$. To interpret the significant interaction effects, simple effects were tested using a Dunn–Sidak correction for multiple contrasts (p corrected = .009). Results revealed that there were no significant differences between the speaker groups in quiet (mean HNC control speakers = 99.90% vs. mean speakers with imprecise speech = 97.23%). However, as the SNR decreased, speakers with intact speech (HNC control speakers) performed significantly better than those with imprecise speech at both +7 dB SNR (mean HNC control speakers = 86.93% vs. mean speakers with imprecise speech = 72.38%, $p < .001$, Cohen's $d = 0.64$: modest effect) and +5 dB SNR (mean HNC control speakers = 81.08% vs. mean speakers with imprecise speech = 56.49%, $p < .001$, Cohen's $d = 1.09$:

Figure 1. Mean speech intelligibility (%) scores of speakers with head and neck cancer (HNC) with imprecise speech and HNC control speakers with intact speech across all noise conditions. The values are in percentages (%), where 0 % = *no words understood* and 100% = *all words understood*.



Note. Error bars are standard errors.

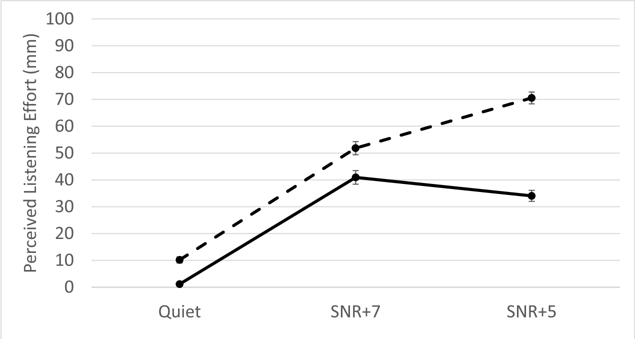
large effect). Presence of background noise significantly affected speakers with imprecise speech more than those with intact speech, with differences becoming significantly stronger from +7 to +5 dB SNR.

The effects within the two groups of speakers were also investigated to determine whether there was a significant change in speech intelligibility from one noise condition to the next. All changes from one noise condition to the next were significant for the group of speakers with imprecise speech (ps < .001). In contrast, while HNC control speakers showed a significant decrease from quiet to +7 dB SNR, there was no significant change in intelligibility from +7 to +5 dB SNR. The main effect of noise was also significant, $F(2, 894) = 135.04$, $p < .001$. Across the speaker groups, intelligibility scores were significantly lowest in the noisiest condition (+5 dB SNR: $M = 68.79\%$), followed by the +7 dB SNR condition ($M = 79.65\%$), followed by the quiet condition ($M = 98.56\%$; all contrasts were statistically significant, using Tukey's honestly significant difference post hoc tests). In addition, there was a significant main effect of speaker group, $F(1, 894) = 86.61$, $p < .001$, with HNC control speakers significantly more intelligible than speakers with HNC with imprecise speech, overall.

Effect of Background Noise on Perceived Listening Effort

The mean perceived listening effort scores for the speaker groups are reported across all noise conditions in Figure 2. A 2 × 3 ANOVA was calculated to examine the effect of speaker group and background noise on perceived listening effort. Consistent with the findings for speech intelligibility, a significant Speaker Group × Noise interaction was also found, $F(2, 894) = 30.73$, $p < .001$. To interpret the significant interaction effects, simple effects were tested using a Dunn–Sidak correction for multiple contrasts (p corrected = .009). Results revealed that there were significant differences between the speaker groups across all of the noise conditions, but that the effect was strongest in the +5 dB SNR condition. Specifically, significant differences were demonstrated between the speaker groups for perceived listening effort in quiet (mean HNC control speakers = 1.15 mm vs. mean speakers with imprecise speech = 10.17 mm, $p < .001$, Cohen's $d = .38$: small effect), at +7 dB SNR (mean HNC control speakers = 40.93 mm vs. mean speakers with imprecise speech = 51.82 mm, $p < .001$, Cohen's $d = .45$: small effect) and +5 dB SNR (mean HNC control speakers = 34.05 mm vs. mean speakers with imprecise speech = 70.55 mm, $p < .001$, Cohen's $d = 1.52$: large effect).

Figure 2. Mean perceived listening effort scores of speakers with head and neck cancer (HNC) with imprecise speech and HNC control speakers with intact speech across all noise conditions. The values are in mm, where 0 = *no effort* and 100 = *extremely effortful*.



Note. Error bars are standard errors.

The effects within the two groups of speakers were also investigated to determine whether there was a significant change in perceived listening effort from one noise condition to the next. Similar to the effects in speech intelligibility, all changes from one noise condition to the next were significant for speakers with HNC with imprecise speech (ps < .001). In addition, while HNC control speakers showed a significant decrease from quiet to +7 dB SNR, there was no significant change in perceived listening effort from +7 to +5 dB SNR. In addition to the interaction effects, the main effect of noise was significant, $F(2, 894) = 336.24$, $p < .001$. Across the speaker groups, perceived listening effort scores were significantly highest in the noisiest condition (i.e., +5 dB SNR, $M = 52.30$ mm), followed by the +7 dB SNR condition ($M = 46.37$ mm), and were least in the quiet condition ($M = 5.55$ mm); all contrasts were statistically significant using Tukey's honestly significant difference post hoc tests. In addition, there was a significant main effect of speaker group, $F(1, 894) = 138.31$, $p < .001$, with speakers with imprecise speech having higher perceived listening effort than HNC control speakers.

Relationship Between Speech Intelligibility and Perceived Listening Effort

Zero order correlation coefficients were first calculated to examine the association between speech intelligibility and perceived listening effort for ratings of all of the individual sentences, averaged across the five listeners in each listening group ($N = 900$ sentences). The results showed intelligibility was moderately to strongly negatively correlated with perceived listening effort ($r = -.70$, $p < .001$). As intelligibility decreased, perceived listening effort significantly increased. Correlations between perceived listening effort and the primary variables of interest in this study are reported in Table 4.

A multiple linear regression with sequential predictor entry was used to predict perceived listening effort. Assumptions of normality, linearity, and homoscedasticity were tenable for this data set. Listener group was dummy coded and entered into the model first to control for this variable and any potential clustering of data that would violate the assumption of independence. The predictor variable, speech intelligibility, was converted to standardized scores (z scores). Noise level was dummy coded and entered into the model, using quiet as the reference for comparison. Speaker group

Table 4. Zero-order correlations among primary variables of interest and perceived listening effort.

Variable	1.	2.	3.	4.
Predicted variable				
1. Perceived listening effort	—			
Primary predictor variable				
2. Intelligibility	−.70**	—		
3. Speaker group	.28**	−.26**	—	
4. Speaker Group × Noise at +7dB SNR	.23**	−.17**	.45**	—
5. Speaker Group × Noise at +5 dB SNR	.48**	−.43**	.45**	−.20**

**p < .001.

was effect coded (−1 = HNC control, 1 = imprecise speech). The interaction between speaker group and noise level was also added to the model.

Results of the multiple linear regression showed that listener group, entered in Block 1, accounted for a significant amount of the variance in perceived listening effort, 4.1%, $F(4, 894) = 7.56$, $p < .001$. Controlling for listener group, the main effect of intelligibility accounted for significant variance in perceived listening effort in Block 2, $R^2_{\text{change}} = .50$, $F_{\text{change}}(1, 893) = 962.11$, $p < .001$ ($R^2_{\text{total}} = .55$ and $R^2_{\text{adjusted}} = .54$). In Block 3, controlling for all other variables, the main effect of speaker group accounted for significant variance in perceived listening effort, $R^2_{\text{change}} = 0.01$, $F_{\text{change}}(1, 892) = 16.35$, $p < .001$ ($R^2_{\text{total}} = .55$ and $R^2_{\text{adjusted}} = .54$). In the final block, controlling for all other variables, the interaction between speaker group and noise accounted for an additional 8% of the variance in perceived listening effort (above and beyond the main effects), $R^2_{\text{change}} = 0.08$, $F_{\text{change}}(1, 890) = 96.60$, $p < .001$ ($R^2_{\text{total}} = .63$ and $R^2_{\text{adjusted}} = .62$).

In the final model, with all predictors entered, the average level of perceived listening effort for all speakers was 24.92 mm ($SE = 1.81$), holding all other variables constant, $t(890) = 13.80$, $p < .001$. Speech intelligibility was uniquely predictive of perceived listening effort ($b = -18.04$, $SE = 0.82$), $t(890) = -21.91$, $p < .001$, $sr^2 = 20.1\%$, after controlling for listener group. On average, for every 1 *SD* increase in speech intelligibility score, there was an 18.04-mm reduction in perceived listening effort, holding other variables constant. Speaker group was also unique predictive of perceived listening effort ($b = -5.98$, $SE = 1.00$, $t(890) = -6.01$, $p < .001$, $sr^2 = 1.5\%$). On average, perceived listening effort increased by 11.97 mm for speakers with imprecise articulation compared to those with intact speech when other variables were held constant. Finally, there was a significant interaction between speaker group and noise level. The effect of speaker group differed significantly at +7 dB SNR compared to quiet ($b = 29.21$, $SE = 2.55$), $t(890) = 11.47$, $p < .001$, $sr^2 = 5.4\%$, when other variables were held constant. The effect of speaker group also differed significantly at +5 dB SNR compared to quiet ($b = 35.34$, $SE = 2.73$), $t(890) = 12.93$, $p < .001$, $sr^2 = 7.0\%$, when other variables were held constant. Results of the multiple linear regression are presented in Table 5.

Discussion

Effect of Background Noise on Speech Intelligibility

Results from this study suggest that individuals with mild speech impairments are penalized to a significantly greater extent in background noise than HNC survivors with intact speech. As expected, all speakers with HNC experienced a decrease in speech intelligibility due to the degradation of the signal caused by background noise (i.e., speech intelligibility became increasingly worse across the three conditions: quiet, +7 dB SNR, +5 dB SNR). However, speakers with imprecise speech experienced greater deficits in the two levels of noise than HNC control speakers with intact speech, above and beyond what might be expected from noise alone.

Importantly, there were no significant differences in speech intelligibility between the speaker groups in quiet. Given that standard speech intelligibility measures are performed in quiet clinical environments, these results demonstrate how typical administration of these assessment tools and their derived scores might be insensitive for identifying speech difficulties and for making referrals for follow-up care. The effect of noise, therefore, warrants consideration when developing speech evaluation protocols and recommending treatment outcome measures.

Speakers included in this study may not be representative of others treated for nonlaryngeal-based HNC who may have more significant, long-lasting speech difficulties. Constantinescu et al. (2017) indicated that single-word intelligibility for oral cancer survivors ranged from 46.0% to 99.8% (average: 82.6%). Furia et al. (2001) reported that intelligibility scores of consonant–vowel–consonant syllables for speakers with glossectomies averaged between 20.0% and 42.2%. Thus, even in quiet conditions, individuals who have been treated for HNC often present with wide ranges of speech intelligibility. Results from this study should therefore be interpreted as an underestimate of the effect of noise on speech outcomes for many individuals treated for HNC.

As the noise levels increased in this study, the difference in speech intelligibility between the two speaker groups became most pronounced (i.e., strongest effect sizes at +5 dB SNR). These findings are consistent with previous studies, which have shown that speakers with other communication disorders may face more significant penalties in the presence of background noise than speakers without communication disorders (Chiu & Forrest, 2018; Holley et al., 1983; Ishikawa et al., 2017; McAuliffe et al., 2009). Results from this study also showed that, while the addition of noise impacted the speakers with HNC with imprecise speech from one noise level (+7 dB SNR) to the next (+5 dB SNR), HNC control speakers with intact speech were not similarly impacted. Speech intelligibility for these HNC control speakers was not significantly different between +7 dB SNR ($M = 87\%$) and +5 dB SNR ($M = 81\%$). Thus, it appears that listeners were able to tolerate this change in SNR for HNC control speakers, without it significantly affecting performance. To understand how these noise conditions

Table 5. Standardized/unstandardized coefficients for regression models predicting perceived listening effort.

Model		*B*	*SE*	β	*t*	Sig.	sr^2
1	(Constant)	31.75	2.70		11.78	< .001*	
	Listener Group 1c	0.83	3.81	0.01	0.22	.828	
	Listener Group 2a	11.64	3.81	0.13	3.05	.002*	.01
	Listener Group 2b	−2.43	3.81	−0.03	−0.64	.525	
	Listener Group 1a	12.85	3.81	0.14	3.37	.001*	.01
	Listener Group 2c	−4.73	3.81	−0.05	−1.24	.215	
2	(Constant)	32.03	1.87		17.12	< .001*	
	Listener Group 1c	−6.32	2.66	−0.07	−2.38	.018*	.00
	Listener Group 2a	14.33	2.65	0.16	5.41	< .001*	.02
	Listener Group 2b	−3.23	2.65	−0.04	−1.22	.223	
	Listener Group 1a	10.22	2.65	0.11	3.86	< .001*	.01
	Listener Group 2c	1.50	2.65	0.02	0.57	.572	
	Intelligibility (*z* score)	−24.08	0.78	−0.72	−31.02	< .001*	.50
3	(Constant)	32.67	1.86		17.54	< .001*	
	Listener Group 1c	−6.03	2.64	−0.07	−2.29	.022*	.00
	Listener Group 2a	12.91	2.65	0.14	4.87	< .001*	.01
	Listener Group 2b	−4.51	2.64	−0.05	−1.71	.089	
	Listener Group 1a	10.32	2.63	0.11	3.93	< .001*	.01
	Listener Group 2c	−0.06	2.66	0.00	−0.02	.983	
	Intelligibility (*z* score)	−23.12	0.81	−0.69	−28.70	< .001*	.42
	Speaker group (C −1; I +1)	3.27	0.81	0.10	4.04	< .001*	.01
4	(Constant)	24.92	1.81		13.80	< .001*	
	Listener Group 1c	−11.60	2.43	−0.13	−4.77	< .001*	.01
	Listener Group 2a	8.98	2.43	0.10	3.69	< .001*	.01
	Listener Group 2b	−7.70	2.42	−0.09	−3.18	.002*	.00
	Listener Group 1a	9.65	2.43	0.11	3.97	< .001*	.01
	Listener Group 2c	−4.74	2.45	−0.05	−1.93	.054	
	Intelligibility (*z* score)	−18.04	0.82	−0.54	−21.91	< .001*	.20
	Speaker group (C −1; I +1)	−5.98	1.00	−0.18	−6.01	< .001*	.02
	Speaker Group × Noise (+7 dB SNR vs. quiet)	29.21	2.55	0.32	11.47	< .001*	.05
	Speaker Group × Noise (+5 dB SNR vs. quiet)	35.34	2.73	0.39	12.93	< .001*	.07

Note. R^2 = .04*, Model 1; ΔR^2 = .50*, Model 2; ΔR^2 = .01*, Model 3; ΔR^2 = .08*, Model 4.
*p < .05.

affected the listeners and their rating strategies, ratings of perceived listening effort must also be considered.

Effect of Background Noise on Perceived Listening Effort

While the general findings for perceived listening effort were consistent with the expected hypotheses, a few notable differences arose. Although there were no differences in speech intelligibility between the two speaker groups in quiet, listeners showed a small but significant difference between the groups for perceived listening effort. Thus, even when speakers were fully intelligible, listeners perceived additional effort in navigating the changes to speech precision and/or resonance. This finding suggests that perceived listening effort, even in quiet, was a sensitive measure that was able to detect subtle differences between the groups.

Results also showed that increases in background noise corresponded with ratings of increased perceived listening effort. This effect was especially apparent in the group of speakers with imprecise speech, with the differences between the groups being most profound in the noisiest condition (+5 dB SNR; large effect). However, similar to the results for speech intelligibility, the HNC control speakers exhibited a different pattern. In the HNC control group, perceived listening effort significantly increased from quiet to the first noise condition (+7 dB SNR), but then it plateaued and did not show any differences between +7 and +5 dB SNR. These results are interesting because they also suggest that, at least for speakers with HNC with intact speech, listeners were not sensitive to increases in perceived listening effort in this range. Yet, for individuals with imprecise speech, the decrease in SNR from +7 to +5 dB SNR was especially penalizing, with listeners perceiving that they were working even harder to decipher the message between the two conditions.

Together, results from this study highlight how individuals treated for HNC who exhibit mild speech impairments, such as sound distortions, are particularly susceptible to even low levels of background noise. It appears that the differences from quiet to +7 dB SNR were enough to affect speech intelligibility and perceived listening effort in both groups of speakers. From +7 to +5 dB SNR, more prominent differences between the two HNC speaker groups strongly emerged. The increase in background noise did not affect listeners' ratings of intelligibility or their perceived effort for the HNC control speakers with intact speech; however, listeners were extremely sensitive to the added noise for speakers with HNC with imprecise speech. Depending on the speaker population and level of speech intelligibility, there may be a sensitive threshold where listeners are especially affected by a degraded signal, and in this range, they perceive that they are working harder to understand speakers. At this threshold, other strategies that listeners use to help them decode messages (e.g., "top-down" strategies) may not be as effective. This hypothesis aligns with previous research showing minimal increases in intelligibility for typical speakers (and normal hearing listeners) who decoded words in sentences at SNRs above approximately +6 dB SNR. Below this SNR, performance appears to precipitously decrease (G. A. Miller et al., 1951). This result has implications for clinical intervention, particularly understanding how to mitigate contextual factors that may significantly impact everyday communication exchanges.

Relationship Between Speech Intelligibility and Perceived Listening Effort

The second research question in this study examined how speech intelligibility, imprecise speech, and noise contributed to perceived listening effort as a global speech outcome measure, beyond speech intelligibility. Results showed that intelligibility was moderately to strongly negatively correlated with perceived listening effort ($r = -.70$, $p < .001$). As intelligibility decreased, perceived listening effort significantly increased, consistent with previous studies (Landa et al., 2014; Nagle & Eadie, 2012, 2018).

Results from the regression model showed that 63% of the variance in perceived listening effort ratings was predicted by intelligibility, speaker group, and noise as a function of speaker group. These findings are similar to the total perceived listening effort predicted by ratings of speech intelligibility and acceptability of typical speakers who used electrolaryngeal speech in the study by Nagle and Eadie (2018; total variance of perceived listening effort = 68%). Unsurprisingly, speech intelligibility accounted for a large percentage (50%) of the variance in perceived listening effort. Yet, controlling for this variable, speaker group accounted for 1% of additional variance, and when all other variables were held constant, noise significantly and uniquely predicted an additional 8% of the variance in perceived listening effort. Together, results show that perceived listening effort is a measure that is not only sensitive to speech intelligibility, but it is uniquely sensitive to the interaction between noise and speech precision.

Although 63% of the variance in perceived listening effort was predicted in the model, it is clear that other factors that go beyond the speech signal and the environment contribute to perceived listening effort. Evitts and Searl (2006) describe the cognitive–perceptual processes listeners use to analyze/decode the incoming speech signal, which may be independent of the acoustic/phonetic content of the signal itself. For example, perceived listening effort is also influenced by listener characteristics (e.g., hearing status, native language, familiarity with the speaker/stimuli) and state (e.g., fatigue, stress) (Landa et al., 2014; McGarrigle et al., 2014). Listeners may also experience differences in their effort threshold, which means that 100% of one individual's perceived listening effort may be different than another listener's threshold. How other factors contribute to a measure of perceived listening effort for speakers who have been treated for HNC should be a subject of future studies.

Limitations and Implications

Several limitations and implications of this study must be considered. First, this study only included five speakers in each of the two nonlaryngeal HNC speaker groups, which limits the generalizability of the results. Second, results need to be interpreted in the context of the speakers included in this study. Specifically, speakers with "speech impairments" were those who were 100% intelligible to experienced clinicians in quiet, but who demonstrated speech imprecisions. Third, the ecological validity of the design used in this study may have also been limited. Speech samples were recorded in a quiet environment, and the background noise was artificially added. This type of manipulation may not be entirely representative of the compensations that speakers make when they communicate with others in background noise. The Lombard effect is a phenomenon that accounts for a speaker's vocal response to background noise (Lombard, 1911), typically in the form of increased volume, duration, and pitch. Previous studies have shown that compensating for background noise with Lombard speech may enhance intelligibility (Garnier et al., 2006). However, other studies have shown that some speaker populations (e.g., tracheoesophageal speakers) may experience decreases in intelligibility in the same conditions due to maladaptive compensations made under such effortful conditions (McColl, 2006). Future studies should investigate how speakers treated for HNC might compensate by recording their samples in the real-time presence of background noise and then by measuring the listener outcomes associated with these potential compensations.

Finally, the levels of background noise selected for this study should also be considered. At +7 and +5 dB SNR, the signals were always presented at a higher sound intensity than the background noise. These conditions were designed to capture realistic environments in which background noise is present but not completely overpowering (e.g., comparable level to speaking with someone at a busy café or a social gathering). These levels were deliberately chosen to avoid a floor effect for all speakers. More challenging noise conditions might have elicited different interactions and need further study.

Conclusions

Few studies have examined the effect of background noise on speech intelligibility and perceived listening effort among those with nonlaryngeal HNC. This area of research is particularly important to consider in the face of changing demographics, in which the rate of HNC related to the human papillomavirus continues to rise among a younger population. This is an important consideration because these individuals have better prognoses and thereby return to work and may communicate in more complex (and noisy) settings on a regular basis.

Though this study included unfamiliar listeners, familiar communication partners might be able to understand more of the speakers' messages and use less perceived listening effort (Landa et al., 2014). Even so, individuals with HNC and their communication partners may benefit from further education and training. Best practices adapted from literature on hearing loss that promote better communication outcomes include selecting quiet environments, using principles of acoustic design to decrease noise, decreasing the distance between the speaker and listener(s), eliminating distractions, supporting visibility, using nonverbal cues, signaling topics, repairing communication breakdowns, and incorporating breaks to prevent fatigue (ASHA, 2015). Behavioral intervention strategies to enhance speech production have also been used to counteract the effects of background noise; these strategies include increasing vocal intensity, overarticulating sounds, reducing the rate of speech, and modifying intonation (Duffy, 2005).

Liss (2007) and others (Borrie et al., 2012; Lansford et al., 2019) have suggested interventions to improve a listener's ability to comprehend disordered speech signals. In this context, listener-targeted treatment is not meant to replace traditional behavioral intervention, but these programs serve as an adjunct by training listeners, such as family members. In cases where the speech signal is so severely impaired that direct speaker-based intervention would be of little benefit, interventions targeting the listener may provide an alternative method of improving communication interactions. How the speaker and the listener interact and adapt to these communicative exchanges has been a focus of recent study (Olmstead et al., 2020).

Results from this study suggest that perceived listening effort may be a more sensitive, global measure of speech than speech intelligibility alone. Measures of perceived listening effort warrant consideration when developing and standardizing assessment protocols. They may complement measures of speech intelligibility in quiet and noise, as well as other validated patient-reported outcomes that capture a person's speech difficulty in everyday contexts. Standard adoption of these types of measures will better serve our patients with HNC and their families as they continue to navigate everyday complex communication environments.

Acknowledgments

This study was funded by the National Cancer Institute (R01CA177635; PI: Eadie) and the University of Washington Speech & Hearing Sciences Vocal Function Laboratory.

References

American Speech-Language-Hearing Association. (2015). *Tips for improving your listening experience* [PDF]. https://www.asha.org/uploadedFiles/AIS-Improving-Listening-Experience.pdf

Baylor, C., Burns, M., Eadie, T., Britton, D., & Yorkston, K. (2011). A qualitative study of interference with communicative participation across communication disorders in adults. *American Journal of Speech-Language Pathology, 20*(4), 269–287. https://doi.org/10.1044/1058-0360(2011/10-0084)

Beukelman, D. R., Childes, J., Carrell, T., Funk, T., Ball, L. J., & Pattee, G. L. (2011). Perceived attention allocation of listeners who transcribe the speech of speakers with amyotrophic lateral sclerosis. *Speech Communication, 53*(6), 801–806. https://doi.org/10.1016/j.specom.2010.12.005

Beverly, D., Cannito, M. P., Chorna, L., Wolf, T., Suiter, D. M., & Bene, E. R. (2010). Influence of stimulus sentence characteristics on speech intelligibility scores in hypokinetic dysarthria. *Journal of Medical Speech-Language Pathology, 18*(4), 9–14.

Borggreven, P. A., Verdonck-de Leeuw, I., Langendijk, J. A., Doornaert, P., Koster, M. N., de Bree, R., & Leemans, C. R. (2005). Speech outcome after surgical treatment for oral and oropharyngeal cancer: A longitudinal assessment of patients reconstructed by a microvascular flap. *Head & Neck, 27*(9), 785–793. https://doi.org/10.1002/hed.20236

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55*(1), 290–305. https://doi.org/10.1044/1092-4388(2011/10-0349)

Chiu, Y.-F., & Forrest, K. (2018). The impact of lexical characteristics and noise on intelligibility on Parkinsonian speech. *Journal of Speech, Language, and Hearing Research, 61*(4), 837–846. https://doi.org/10.1044/2017_JSLHR-S-17-0205

Chiu, Y.-F., & Neel, A. (2020). Predicting intelligibility deficits in Parkinson's disease with perceptual speech ratings. *Journal of Speech, Language, and Hearing Research, 63*(2), 433–443. https://doi.org/10.1044/2019_JSLHR-19-00134

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Routledge Academic. https://doi.org/10.1002/bs.3830330104

Constantinescu, G., Rieger, J., Winget, M., Paulsen, C., & Seikaly, H. (2017). Patient perception of speech outcomes: The relationship between clinical measures and self-perception of speech function following surgical treatment for oral cancer. *American Journal of Speech-Language Pathology, 26*(2), 241–247. https://doi.org/10.1044/2016_AJSLP-15-0170

Doyle, P. C., & Eadie, T. L. (2005). The perceptual nature of alaryngeal voice and speech. In *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing* (pp. 113–140). Pro-Ed.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management* (2nd ed.). Elsevier Mosby.

Dwivedi, R. C., Kazi, R. A., Agrawal, N., Nutting, C. M., Clarke, P. M., Kerawala, C. J., Rhys-Evans, P. H., & Harrington, K. J. (2009). Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers. *Cancer Treatment Reviews, 35*(5), 417–424. https://doi.org/10.1016/j.ctrv.2009.04.013

Eadie, T. L., Otero, D. S., Bolt, S., Kapsner-Smith, M., & Sullivan, J. R. (2016). The effect of noise on relationships between speech intelligibility and self-reported communication measures in tracheoesophageal speakers. *American Journal of Speech-Language Pathology, 25*(3), 393–407. https://doi.org/10.1044/2016_AJSLP-15-0081

Evitts, P., & Searl, J. (2006). Reaction times of normal listeners to laryngeal, alaryngeal, and synthetic speech. *Journal of Speech, Language, and Hearing Research, 49*(6), 1380–1390. https://doi.org/10.1044/1092-4388(2006/099)

Furia, C. L., Kowalski, L. P., Latorre, M. R., Angelis, E. C., Martins, N. M., Barros, A. P., & Ribeiro, K. C. (2001). Speech intelligibility after glossectomy and speech rehabilitation.

Archives of Otolaryngology—Head & Neck Surgery, 127(7), 877–883.

Garnier, M., Dohen, M., Lœvenbruck, H., Welby, P., & Bailly, L. (2006, December). The Lombard effect: A physiological reflex or a controlled intelligibility enhancement? In H. C. Yehia, D. Demolin, & R. Laboissiere (Eds.), *Proceedings of ISSP '06, the 7th International Seminar on Speech Production* (pp. 255–262). http://hal.archives-ouvertes.fr/hal-00214307

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology, 24*(1), 26–36. https://doi.org/10.3766/jaaa.24.1.4

Holley, S. C., Lerman, J., & Randolph, K. (1983). A comparison of the intelligibility of esophageal, electrolaryngeal, and normal speech in quiet and in noise. *Journal of Communication Disorders, 16*(2), 143–155. https://doi.org/10.1016/0021-9924(83)90045-X

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12*(2), 198–208. https://doi.org/10.1044/1058-0360(2003/066)

Hutcheson, K. A., & Lewin, J. S. (2013). Functional assessment and rehabilitation: How to maximize outcomes. *Otolaryngology Clinics of North America, 46*(4), 657–670. https://doi.org/10.1016/j.otc.2013.04.006

Ishikawa, K., Boyce, S., Kelchner, L., Powell, M. G., Schieve, H., de Alarcon, A., & Khosla, S. (2017). The effect of background noise on intelligibility of dysphonic speech. *Journal of Speech, Language, and Hearing Research, 60*(7), 1919–1929. https://doi.org/10.1044/2017_JSLHR-S-16-0012

Jacobi, I., Molen, L., Huiskens, H., van Rossum, M., & Hilgers, F. (2010). Voice and speech outcomes of chemoradiation for advanced head and neck cancer: A systematic review. *European Archives of Oto-Rhino-Laryngology, 267*(10), 1495–1505. https://doi.org/10.1007/s00405-010-1316-x

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482

Landa, S., Pennington, L., Miller, N., Robson, S., Thompson, V., & Steen, N. (2014). Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding. *International Journal of Speech-Language Pathology, 16*(4), 408–416. https://doi.org/10.3109/17549507.2014.927922

Lansford, K. L., Borrie, S. A., & Barrett, T. S. (2019). Regularity matters: Unpredictable speech degradation inhibits adaptation to dysarthric speech. *Journal of Speech, Language, and Hearing Research, 62*(12), 4282–4290. https://doi.org/10.1044/2019_JSLHR-19-00055

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology, 44*(3), 131–143. https://doi.org/10.1080/14992020500057244

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and Hearing, 37*(Suppl. 1), 77S–84S. https://doi.org/10.1097/AUD.0000000000000304

Liss, J. M. (2007). The role of speech perception in motor speech disorders. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 187–219). Plural.

Loewen, I. J., Boliek, C. A., Harris, J., Seikaly, H., & Rieger, J. M. (2010). Oral sensation and function: A comparison of patients with innervated radial forearm free flap reconstruction to healthy matched controls. *Head & Neck, 32*(1), 85–95. https://doi.org/10.1002/hed.21155

Lombard, E. (1911). Le signe de l'elevation de la voix [The sign of the rise in the voice]. *Annals maladiers oreille, Larynx, Nez, Pharynx, 37,* 101–119.

McAuliffe, M. J., Schaefer, M., O'Beirne, G. A., & LaPointe, L. L. (2009, November). *Effect of noise upon the perception of speech intelligibility in dysarthria*. Poster session presented at the Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA, United States.

McColl, D. (2006). Intelligibility of tracheoesophageal speech in noise. *Journal of Voice, 20*(4), 605–615. https://doi.org/10.1016/j.jvoice.2005.09.005

McColl, D., Fucci, D., Petrosino, L., Martin, D. E., & McCaffrey, P. (1998). Listener ratings of the intelligibility of tracheoesophageal speech in noise. *Journal of Communication Disorders, 31*(4), 279–288. https://doi.org/10.1016/S0021-9924(98)00008-2

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology, 53*(7), 433–440. https://doi.org/10.3109/14992027.2014.890296

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329–335. https://doi.org/10.1037/h0062491

Miller, N., Allcock, L., Jones, D., Noble, E., Hildreth, A. J., & Burn, D. J. (2007). Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry, 78*(11), 1188–1190. https://doi.org/10.1136/jnnp.2006.110171

Nagle, K. F., & Eadie, T. L. (2012). Listener effort for highly intelligible tracheoesophageal speech. *Journal of Communication Disorders, 45*(3), 235–245. https://doi.org/10.1016/j.jcomdis.2012.01.001

Nagle, K. F., & Eadie, T. L. (2018). Perceived listener effort as an outcome measure for disordered speech. *Journal of Communication Disorders, 73,* 34–49. https://doi.org/10.1016/j.jcomdis.2018.03.003

Olmstead, A. J., Lee, J., & Viswanathan, N. (2020). The role of the speaker, the listener, and their joint contributions during communicative interactions: A tripartite view of intelligibility in individuals with dysarthria. *Journal of Speech, Language, and Hearing Research, 63*(4), 1106–1114. https://doi.org/10.1044/2020_JSLHR-19-00233

Pennington, L., & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clinical linguistics & Phonetics, 21*(5), 393–403. https://doi.org/10.1080/02699200701276675

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (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

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8,* 71–80.

Sussman, J. E., & Tjaden, K. (2012). Perceptual measures of speech from individuals with Parkinson's disease and multiple sclerosis:

Intelligibility and beyond. *Journal of Speech, Language, and Hearing Research, 55*(4), 1208–1219. https://doi.org/10.1044/1092-4388(2011/11-0048)

van der Molen, L., van Rossum, M. A., Jacobi, I., van Son, R. J. J. H., Smeele, L. E., Rasch, C. R. N., & Hilgers, F. J. M. (2012). Pre- and posttreatment voice and speech outcomes in patients with advanced head and neck cancer treated with chemoradiotherapy: Expert listeners' and patients' perception. *Journal of Voice, 26*(5), 664.e25–664.e33. https://doi.org/10.1016/j.jvoice.2011.08.016

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America, 121,* 519–526. https://doi.org/10.1121/1.2400666

Whitehill, T. L., & Wong, C. C. Y. (2006). Contributing factors to perceived listening effort for dysarthric speech. *Journal of Medical Speech-Language Pathology, 14*(4), 335–342.

Windrich, M., Maier, A., Kohler, R., Nöth, E., Nkenke, E., Eysholdt, U., & Schuster, M. (2008). Automatic quantification of speech intelligibility of adults with oral squamous cell carcinoma. *Folia Phoniatrica et Logopaedica, 60*(3), 151–156. https://doi.org/10.1159/000121004

Yorkston, K., Beukelman, D., & Tice, R. (1996). *Sentence Intelligibility Test*. Tice Technologies.