

The Effects of Age and Background Noise on Intelligibility and Perceived Listening Effort for
Tracheoesophageal Speakers

Carly Spies

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Tanya Eadie

Kristie Spencer

Kathleen Nagle

Program Authorized to Offer Degree:

Department of Speech and Hearing Sciences

©Copyright 2025

Carly Spies

University of Washington

Abstract

The Effects of Age and Background Noise on Intelligibility and Perceived Listening Effort for
Tracheoesophageal Speakers

Carly Spies

Chair of the Supervisory Committee:

Tanya Eadie

Department of Speech and Hearing Sciences

The current study evaluated the effect of background noise and/or a communication partner's age on speech intelligibility and perceived listening effort for individuals who use tracheoesophageal (TEP) speech. Nine adult male TEP speakers provided speech samples from an intelligibility test. One set of speech samples were recorded in quiet, and another set was mixed with multitalker babble. Twelve younger listeners and 12 older listeners with typical hearing transcribed sentences from the TEP speakers in quiet and noise and rated perceived listening effort using 100-mm visual analog scales. Results revealed that regardless of the age of listener, TEP speakers were significantly less intelligible in noise and that listeners required more listening effort to understand these speakers in noise. Results also revealed that when both hearing status and working memory were controlled, the age of the communication partner did not significantly impact speech intelligibility or perceived listening effort. These results add to the growing literature that people with communication disorders are penalized more than typical speakers when communicating in background noise, even if they show comparable metrics in quiet settings. These results need to be considered when developing strategies for assessing TEP

speakers and monitoring treatment success. While the results of this study did not show a statistically significant difference in intelligibility ratings and perceived listening effort based on the age of listener, it is also important to consider factors related to the communication partner/listener when providing therapy and counseling for TEP speakers and their families.

1. Introduction

The American Cancer Society (ACS) estimates that there will be approximately 13,020 new cases of laryngeal cancer in the USA in 2025, with the average age of diagnosis of laryngeal cancer being 66 years old (ACS, 2025). While the overall occurrence of laryngeal cancer is low in comparison to other cancers, its consequences may be particularly devastating due to the impact on an individual's voice and communication. The effects extend past the physical sequelae and may have a profound impact on an individual's emotional, psychological, and social wellbeing (Bornbaum & Doyle, 2019).

One surgical treatment option for individuals diagnosed with advanced laryngeal cancer continues to be total laryngectomy, either as a primary or salvage procedure (National Cancer Institute, 2025). The procedure, which includes total removal of the laryngeal framework, results in an altered airway and the loss of an individual's normal vocal mechanism. Following a total laryngectomy, there are three options for alaryngeal speech. These options are esophageal, electrolaryngeal, and surgical prosthetic (i.e., tracheoesophageal puncture; TEP) voice restoration. While esophageal and electrolaryngeal speech are still common, TEP speech is seen by many as the best option for voice rehabilitation following a total laryngectomy, at least in terms of speech outcomes (Kazi, Sayed, & Dwivedi, 2010).

The TEP technique is considered to be a safe and effective procedure for alaryngeal speech. TEP speech is obtained after a puncture is surgically created between the posterior wall of the trachea and the anterior wall of the esophagus, and may be done at the time of the primary surgery or as a secondary procedure. The puncture is followed by the insertion of a tracheoesophageal voice prosthesis, which allows for air to travel from the trachea to the esophagus, while preventing food and liquids from going in the other direction (Singer & Blom, 1980). One of the main advantages of TEP speech over other methods of alaryngeal speech is the rapid acquisition of speech, as well as the ability to produce smoother and more prolonged speech, due to the resumed access of air from the lungs available for speech production that occurs after the puncture (Knott, 2019). D'Alatri et al. (2012) found that unfamiliar listeners judge TEP speakers as more intelligible (i.e., more understandable) than speakers who use other alaryngeal speech methods. TEP speech has also been shown to be the most similar form of alaryngeal speech when compared to laryngeal speech in terms of acoustic measures such as fundamental frequency, speaking rate, and sound level/intensity (Doyle & Keith, 2005; Kazi et

al., 2010). However, even with these improvements in TEP speech, it is still identifiably different and less preferred by unfamiliar listeners when compared to typical laryngeal speech (van As, Koopmans-van Beinum, Pols, & Hilgers, 2003).

Although measures that capture the severity of a speech disorder, such as speech intelligibility and acoustic parameters of speech, are important when evaluating the performance of a TEP speaker, they are limited in scope and do not necessarily capture how well a person communicates in their everyday environment. To fully assess communication in everyday settings, we need to consider multiple factors, including aspects related to both the speaker as well as their communication partner, the measure that captures the efficiency of their communication, and factors related to the setting in which communication takes place. For example, one major limitation of most outcome measures, such as ratings performed by unfamiliar listeners, is that they are often assessed in quiet conditions. While TEP speakers tend to have higher intelligibility than alaryngeal speakers who use other options, there is some controversy about whether speakers who use this method actually report better outcomes in everyday contexts than other alaryngeal speakers (Eadie, Day, Sawin, Lamvik, & Doyle, 2013; Eadie, Otero, Bolt, Kapsner-Smith, & Sullivan, 2016; Evans, Carding, & Drinnan, 2009; Law, Ma, & Yiu, 2009). In fact, some studies have shown that speech intelligibility assessed in noise is a stronger predictor of voice-related quality of life than speech intelligibility in quiet (Eadie et al., 2016; Eadie et al., 2021). A possible reason for this increase in relationship strength is that an environment with background noise is more representative of daily communication settings than the strictly quiet setting of a lab or clinic.

In addition to considering the assessment context when evaluating a person's communication success, we also need to consider the type of outcome measure being evaluated. For example, several studies have shown that the perceived listening effort exerted by a communication partner, or "the attention and cognitive resources required to understand speech" (Eadie et al., 2021; Gosselin & Gangé, 2011), may be differentiated from speech intelligibility (Nagle & Eadie, 2012, 2018). That is, while a speaker may be 100% understandable to a communication partner, that partner may need to work at an increased level to understand all the words produced by that speaker.

Beyond the severity of a speaker's impairment (e.g., level of speech intelligibility) or the environment (i.e., speaking in quiet or in noise), factors related to the communication

partner/listener may also affect speech outcomes, such as intelligibility or perceived listening effort (Borrie et al., 2012; Chiu & Neel, 2020; Landa et al., 2014). These listener-related factors are important to consider for people who have had total laryngectomies because, on average, individuals are 66 years old when they receive their surgery. As a result, their spouses and peer social partners may also be of similar age. However, while differences between older and younger listeners' performances in speech perception have been investigated in other populations, such as individuals with dysarthria (Dagenais et al., 1999; Ingvalson et al., 2017; McAuliffe et al., 2013; McAuliffe, et al., 2017), very little is known about how older listeners perform as communication partners with individuals who have undergone total laryngectomies, including TEP speech (Clark, 1985; Law et al., 2009). Thus, the overall purpose of this study is to investigate speech intelligibility and perceived listening effort in TEP speakers who are evaluated by both younger and older adult listeners, in both quiet and noisy settings. Before these effects may be better understood, a review of outcomes after total laryngectomy will be summarized. This review is followed by what is known about background noise on these types of measures in speakers with and without speech disorders, including TEP speakers, as well as identifying factors related to the communication partner, such as the age of the listener, that could affect speech outcomes.

1.1 Speech intelligibility in TEP Speech

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). This is a common measure of alaryngeal speech in which clinicians or other listeners provide a metric related to intelligibility. Speech intelligibility can be measured using rating scales, such as 1 to 7 point equal-appearing interval scales or visual analog scales; alternatively, speech intelligibility can be measured via transcription to determine a percentage of recognizable words within a given sample. Both approaches have been used in dysarthria as well as speakers with various types of head and neck cancer (Constantinescu et al., 2017). According to van As et al. (2003), speech intelligibility for TEP speakers is primarily affected by “poor voice quality, altered speech rate, and the noise associated with improper timing of the tracheostoma”.

The overall intelligibility of TEP speech has been shown to be equal to or better than that of esophageal speech and better than that of electrolaryngeal speech (Hillman, Walsh, & Heaton,

2005; Tardy-Mitzell et al., 1985), further indicating why it is considered by many to be the best option for voice rehabilitation following a total laryngectomy. One study that included 16 TEP speakers showed that TEP speakers are highly intelligible to unfamiliar listeners (mean intelligibility = 92%; range = 73%-100%; Eadie et al., 2013). While the mean intelligibility is generally high for TEP speakers, there is wide variability within the group. In addition, while an intelligibility score may provide information on how much a communication partner understands, it does not provide information on how much effort or work the listener expends to understand what is being said. This may be better captured by another measure: perceived listening effort.

1.2 Perceived listening effort in TEP Speech

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 measured through questionnaires or perceptual rating scales. Olmstead et al. (2020) described how both the joint contributions of the speaker and the listener play a role in the speech intelligibility of speakers with dysarthria. They suggested that perceived listening effort exerted during a conversation may be a stronger indicator of communication success in individuals with speech disorders than speech intelligibility measures alone because they are inclusive of the communicative dyad (Olmstead et al., 2020).

Most research on perceived listening effort has centered on the impact of hearing loss on communication (McGarrigle et al., 2014; Pichora-Fuller et al., 2016). However, studies exploring outcomes in individuals with speech disorders interacting with typical hearing partners have also emphasized its value as an outcome measure (Nagle & Eadie, 2012, 2018; Whitehill & Wong, 2006). For example, Nagle and Eadie (2012) investigated perceived listening effort and speech acceptability in 14 highly intelligible TEP speakers. Inexperienced listeners rated the TEP speakers using a paired comparison method; each sample was presented in comparison with all other speech samples and listeners were asked which sample in a pair was “more acceptable” or required “less effort” using 100mm visual analog scales. Overall, group results showed a strong correlation between acceptability and listening effort, suggesting that these constructs are highly related (Nagle & Eadie, 2012). However, these results also highlighted the utility of the construct

of listening effort as a construct that can reliably differentiate speakers from one another, even when intelligibility remains constant.

Similar results have been found in studies involving speakers who used electrolaryngeal speech (Nagle & Eadie, 2018) and those with different types of head and neck cancer (Eadie et al., 2021). These studies, along with others who have included speakers with dysarthria (Beukelman et al., 2011; Chiu & Neel, 2020; N. Miller et al., 2007), indicate that 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 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 speaker, age of the speaker), the stimuli (e.g., lexical density predictability of words), and the environment (e.g., competing noise) may affect perceived listening effort (Borrie et al., 2012; Chiu & Neel, 2020; Landa et al., 2014). When considering these factors, the effect of noise in the communication environment warrants further attention.

1.3 Considerations of Background Noise in Speech

Measures of speech intelligibility and perceived listening effort are valuable when taken in controlled research conditions; however, these findings may be limited due to the lack of generalizability in functional contexts. When background noise is eliminated in research studies, it provides a controlled environment that helps ensure internal validity. However, the external validity of these studies, or their ability to be generalized into everyday contexts, is limited because background noise is present in most everyday communicative interactions. This has led to an increasing number of studies that have investigated the effect of background noise on speech outcome measures.

Previous research has shown that the presence of background noise adversely affects typical speech intelligibility (i.e., speakers without speech and language disabilities) (Sperry et al., 1997; Van Engen & Bradlow, 2007). For speakers with communication disorders, there is growing evidence to suggest that their speech-related challenges are compounded in noise. McAuliffe et al. (2009) first explored the effect of background noise on speech intelligibility in speakers with dysarthria. In this study, intelligibility was measured for dysarthric speakers (n=3)

in quiet and in the presence of background noise (+6dB signal-to-noise ratio or SNR). The performance of the dysarthric speakers was compared with three typical speakers under the same conditions. The results showed that the intelligibility of the dysarthric speakers disproportionately declined with the addition of background noise. In contrast, the typical speakers remained highly intelligible in both conditions.

Similar effects have been shown in speakers with different types of head and neck cancer (HNC; Eadie et al., 2021). Eadie et al. (2021) examined the effect of different levels of background noise on speech intelligibility and perceived listening effort in 5 speakers with mild speech imprecisions and 5 speakers with intact speech following treatment for HNC in quiet and two levels of noise. Their results showed that while both groups were 100% intelligible in quiet, HNC speakers with imprecise speech experienced greater deficits as background noise increased (i.e., quiet, +7 dB SNR, +5 dB SNR), beyond what would be expected from noise alone.

The effect of background noise on speech performance has also been investigated in individuals who have undergone total laryngectomies. For example, Holley et al. (1983) investigated the impact of background noise on speech intelligibility in one alaryngeal speaker who used both esophageal and electrolaryngeal speech, compared to one typical age-matched speaker. The listeners then evaluated the intelligibility of the speakers at the consonant-level, vowel-level, and sentence-level. Their results showed that as background noise increased (i.e., quiet, +3 dB SNR, -1 dB SNR), consonant-level intelligibility worsened for the alaryngeal speaker alone, demonstrating the susceptibility of the alaryngeal speaker's performance in noise.

The effect of background noise on speech outcomes also has been shown in several previous studies involving TEP speakers (Clark, 1985; Clark & Stemple, 1982; Eadie et al., 2016; McColl et al., 1998). Clark and Stemple (1982) compared speech intelligibility of four different speakers, including one TEP speaker, one individual who used esophageal speech, one individual who used an artificial larynx, and one typical laryngeal speaker. Twenty young listeners first made ratings of the speakers in quiet to determine which speaker they "preferred". The listeners then evaluated the intelligibility of the speakers in different levels of background noise (-5dB, -10dB, and 0 dB message-to-competition ratios) using a closed set task. The results showed that the typical, laryngeal speaker was always the most preferred and most intelligible. The TEP speaker was the most preferred mode of alaryngeal speech, with 55% of listeners rating it to be the "most pleasant" when compared to the two other alaryngeal speakers in quiet.

However, when background noise or competing messages were present, the TEP speaker was the least intelligible among the alaryngeal speakers, illustrating the differences in outcomes when listeners made ratings in background noise.

McColl et al. (1998) also investigated the effect of noise on TEP speech intelligibility. They recorded speech samples from one typical laryngeal speaker and one superior TEP speaker. Fifty listeners judged how well they understood the speakers using rating scales, with the samples being presented at nine different signal-to-noise ratio (SNR) levels. Results showed that both speakers were less intelligible in noise, but that the TEP speaker was significantly less intelligible than the laryngeal speaker across all conditions. The results also showed that the TEP speaker's intelligibility was more affected by noise than was that of the typical laryngeal speaker, even in low levels of background noise.

Likewise, Eadie et al. (2016) studied the effect of background noise on speech intelligibility in 24 TEP speakers with varied intelligibility levels that were at least one-year post laryngectomy. Each participant provided audio recordings of 5 sentences from the Sentence Intelligibility Test (SIT; Yorkston et al., 1996). Speech intelligibility was assessed in both quiet and in noise, at +6 dB SNR. Results from 66 listeners revealed that TEP speakers' intelligibility in quiet ranged from 74.13% to 100% ($M=93.27\%$). Similar to the results of the previously discussed studies (Clark & Stemple, 1982; McColl et al., 1998), speech intelligibility in noise was lower than speech intelligibility in quiet. However, the reduction in intelligibility from 93% to 69% in a +6 dB SNR condition also revealed the vulnerability of TEP speakers in noise, in that this decrease was more than what might be expected in typical speakers under similar conditions. One notable limitation of this study was that the only speech outcome that was assessed was speech intelligibility. Listeners did not evaluate how much work they needed to exert to understand the speakers. A second limitation of this study was that like the majority of previous studies, the raters who performed intelligibility ratings were mostly young; for example, the average age of the listeners in this study was 24 years, with a standard deviation of 5.4 years. Given that people who are diagnosed with laryngeal cancer are typically older than 60 years (average age of diagnosis is 66 years), it is important to determine how the age of the communication partner might also impact outcomes because older listeners may be more representative of social peers. In addition, older communication partners may have more

difficulty perceiving TEP speech in everyday settings, such as background noise, for a variety of reasons. These factors, including the age of the listener, will be discussed in the next section.

1.4 Age of Listener

Several studies have shown that older adults experience greater difficulty perceiving speech in noisy environments compared to younger adults (Pichora-Fuller, Schneider, & Daneman, 1995; Schneider, Speranza, & Pichora-Fuller, 1998). This challenge is evident in both older adults with normal hearing (Gordon-Salant & Fitzgibbons, 1995; Schneider, Daneman, Murphy, & See, 2000) and those with hearing loss (Lunner & Sundewall-Thoren, 2007; Souza, Arehart, Shen, Anderson, & Kates, 2015). However, speech perception in noise is not the only difficult listening situation that older adults face. When older adults communicate with an individual with a communication disorder, such as a person who uses TEP speech, they also may have challenges because of the speaker's reduced speech intelligibility, regardless of their hearing status.

It is difficult to discern the effect of age alone on the success of a communication partner who interacts with a person with a speech disorder because older adults are not typically included as participants in research studies. In the very few studies that have included older listeners, the effect of age has been inconsistent on outcomes such as speech intelligibility (Dagenais et al., 1999; Ingvalson et al., 2017; McAulliffe et al., 2017). For example, McAulliffe et al. (2017) examined the effect of loud and slow speech cues on younger and older listeners' comprehension and intelligibility of six speakers with dysarthria. The younger listeners were between the ages of 18-24, while the older listeners were between the ages of 66-80 years old. The findings of this study revealed that older listeners exhibited a significantly reduced ability to comprehend dysarthric speech than the younger listeners. The authors attributed this finding largely to the reduced hearing thresholds of the older adults who participated in their study, although they also posited that differences could be due to cognitive differences among their listener groups as well.

In contrast to these findings, others have reported no significant difference between older and younger listeners who evaluated other dysarthric speakers, particularly when hearing status was controlled (Dagenais et al., 1999; Ingvalson et al., 2017). For example, Ingvalson et al. (2017) included 15 younger and 15 older listeners who all passed similar hearing screening

thresholds in their study. They presented speech stimuli from four speakers with different types of dysarthria to their listeners and asked them to identify the words. They found that both groups of listeners were accurate in their speech perception (around 80% accurate), with no significant differences between the groups. In other words, there was no effect of listener age on speech perception. However, this study also found that in addition to hearing status, older adults may be using more cognitive support such as working memory, receptive vocabulary, and/or cognitive flexibility to perform their speech perception task. In contrast, only hearing status predicted speech perception performance in younger adult listeners.

To date, only two studies have investigated the effect of listener age on intelligibility in TEP speakers (Clark, 1985; Law et al., 2009). Law et al. (2009) examined the effect of listener age on speech intelligibility and acceptability in 49 speakers who used four different alaryngeal speech methods (7 esophageal speakers, 14 electrolaryngeal speakers, 13 who used a pneumatic device, and 15 TEP speakers). The study found that speech intelligibility and acceptability of alaryngeal speech, including TEP speech, significantly varied with the age of the listener. Specifically, younger listeners (aged 25 to 33 years) performed significantly better than older listeners (aged 60 to 65) for the accuracy of transcribed speech intelligibility, but older listeners rated the speech of alaryngeal speakers less severely (i.e., more acceptable) than younger listeners. These differences occurred despite the authors' control of hearing status in both listener groups, which contrasts with the findings of previous studies in dysarthria (Dagenais et al., 1999; Ingvalson et al., 2017).

Clark (1985) also studied the effect of the age of the listener on alaryngeal speakers' intelligibility. Their study included 11 younger listeners between the ages of 21 to 30 years, and 11 older listeners between the ages of 50 to 66 years, who all passed hearing screenings at levels normed for their age. The primary aim of this study was to evaluate the effect of listener age and background noise on speech intelligibility of different types of alaryngeal speech. Clark (1985) included one esophageal speech speaker, one TEP speaker, one speaker who used an artificial larynx, and one typical, laryngeal speaker. All speakers were recorded under the same conditions, which included three different levels of background noise (0dB, -5dB, and -10dB message-to-competition ratios (MCR)). This study used a closed set task to determine levels of speech intelligibility. The results revealed no significant differences between the two groups of listeners in the easiest listening condition for either the typical laryngeal speaker or the TEP

speaker. For example, in quiet, the typical laryngeal speaker was 100% intelligible to the younger listeners and 98% intelligible to the older listeners. Likewise, the TEP speaker was also 100% intelligible to the younger listeners and 92% intelligible to the older listeners, which was not statistically different between the groups. However, at -5 dB MCR, a significant difference was revealed. The typical laryngeal speaker was 97% intelligible to the younger listeners, and 81% intelligible to the older listeners. For the one TEP speaker, there was also a significant interaction. That is, older listeners showed significant difficulty understanding the TEP speaker, beyond what was shown for the typical speaker in noise. In -5dB MCR, the TEP speaker was 76% intelligible to the younger group and 55% intelligible for the older group, suggesting that not only was the TEP speaker difficult to understand in noise for all listeners, but that the effect of age of the listener might also have a differential effect in this amount of noise. This result is consistent with speech perception findings of older listeners in noisy environments in general but may also be preliminary evidence to suggest that older listeners find it particularly challenging to understand a TEP speaker in noise.

There are several limitations of the Clark (1985) study. First, only one speaker was used to represent each alaryngeal speech mode, so further research is needed to determine how representative the one TEP speaker represented TEP speakers in general. In addition, while speech intelligibility was assessed by listeners using a closed set task, it is unknown whether a more conservative type of intelligibility measure, such as transcription, might reveal similar results. Finally, while speech intelligibility of the one TEP speaker was similar between the older and younger listener groups in quiet, it is unknown whether an alternative speech outcome measure that might capture the cognitive resources involved in speech understanding, such as perceived listening effort, might also reveal differences in quiet. All these factors need to be considered when designing a study that investigates the effect of the age of the listener on TEP speech performance in quiet and in noise.

1.5 Experimental Questions

In summary, speech intelligibility and perceived listening effort have been used to quantify speech severity and listener burden in the setting of background noise in a limited number of patient populations, with very few studies including TEP speakers. In addition, this research has typically not included older listeners to represent the typical communication

partners that TEP speakers communicate with. By combining these variables, this study seeks to acknowledge that communication takes effort from both the listener and the speaker and that it rarely takes place in quiet settings. The literature also suggests that individuals with communication disorders may face additional challenges in the context of noise and that their speech may be more difficult to understand or may require more effort to decode than typical speakers in similar conditions. Therefore, the purpose of this study is to identify the effect of background noise and the age of the listener on speech intelligibility and perceived listening effort in TEP speakers. The results are important for effectively counseling TEP speakers and their families and for designing future rehabilitation approaches that are ecologically valid.

2. Methods

This study included the following participants: (a) speakers who had undergone total laryngectomy and used TEP speech as their primary mode of communication; (b) younger listeners; (c) older listeners who were similar in age to the TEP speakers. This study used an experimental design to investigate the effect of background noise (independent variable with two conditions: one quiet and one level of noise) and age of listener (independent variable with two levels: one younger and one older group) on two listener-rated measures: speech intelligibility (measured via transcription) and perceived listening effort (measured using a rating scale) in highly intelligible TEP speakers. The University of Washington Human Subjects Committee approved the procedures used in this study. Participants were paid for their participation.

2.1 Participants

2.1.1 TEP Speakers

Nine individuals (male) who had undergone total laryngectomy secondary to cancer were selected from a previously collected database for participation in this study (see Eadie et al., 2016). Only male speakers were included to control gender as a confounding variable, and because the majority of individuals who undergo total laryngectomy are male (ACS, 2025). In addition, only individuals who reported TEP speech as their primary method of communication were included. The mean age of the selected nine TEP speakers was 65 years (range: 52-73 years), which is representative of those undergoing total laryngectomy secondary to laryngeal cancer (ACS, 2025). Individuals had undergone total laryngectomy at least 1 year prior to

participation in the speech recordings, to allow time for adjustment and adaptation of their new speech method (Campbell, Marbella, & Layde, 2000). Only highly intelligible speakers (ranging from 94.2% to 99.6%) in quiet were selected for inclusion in this study to control the effect of noise on intelligibility, and to potentially allow for differences to be shown for measures of perceived listening effort (Nagle & Eadie, 2012). More information related to the nine selected speakers may be found in the study by Eadie et al. (2016), with key variables presented in Table 1 below.

Table 1. Participant Demographics of 9 Selected TEP Speakers Included in the Present Study

Speaker	Age	Sex (M = male)	% Intelligibility in Quiet (Based on listener ratings from Eadie et al., 2016)
D	70	M	98.52
II	73	M	97.90
J	66	M	97.72
JJ	56	M	99.57
P	72	M	94.23
SG	69	M	94.29
SN	52	M	98.48
SP	66	M	99.63
T	60	M	98.34

2.1.2 Listeners

Twenty-four inexperienced listeners (12 younger listeners and 12 older listeners) were recruited from the community. To be included in the younger listener group, participants needed to be between 18 and 30 years of age (Clark, 1985). To be included in the older listener group, participants needed to be between 52 and 73 years of age; these individuals were selected to be representative of typical peer communication partners for the TEP speakers. Inexperienced listeners were those who reported limited or no previous exposure or coursework in voice or speech disorders including alaryngeal speakers or experience in performing perceptual ratings of speakers using alaryngeal speech. Inexperienced listeners were selected because their ratings generalize to represent a wider cross-section of the population and are similar to unfamiliar communication partners that TEP speakers encounter in their everyday lives. To be included, all listeners were fluent in English and had hearing thresholds on a hearing screening lower than 40

dB HL at 500, 1000, 2000, and 4000 Hz (Ingvalson et al., 2017). These thresholds were selected in an effort to control the effect of hearing status on intelligibility and perceived listening effort. All listeners also completed the Digit Span subtest of the Wechsler Adult Intelligence scale (WAIS: Wechsler, 2008) in an attempt to grossly control working memory effects on the dependent measures of speech intelligibility and perceived listening effort.

Table 2 shows key demographic data for both listener groups. The 12 younger adult listeners were between the ages of 18–28 years old (mean age = 23 years); 9 females and 3 males were in the group, with all reporting being cisgender. Nine listeners in the younger group self-identified as White/Caucasian, two as Asian, and one did not report race. Two listeners self-identified as Hispanic, and 10 listeners did not identify as Hispanic or Latino. The average scaled scores of the WAIS Digit Span subtest across the group was 9.42 (range 6 – 12), with one listener performing slightly below the expected score for their age (see Table 2).

The 12 older listeners who participated in this study were between the ages of 52–73 years (i.e., the age range of the TEP speakers selected for this study) (mean age = 64 years), with 7 females and 5 males, and all identifying as cisgender. Eleven listeners identified as White/Caucasian, and one as Asian. None of the listeners reported identifying as Hispanic or Latino. The average scaled scores for the WAIS Digit Span across the group was 12.5 (range 8 – 16) which was significantly higher than the younger listening group. The 12 listeners in the older group demonstrated scores within normal limits for their age group.

Table 2. Listener Demographics of 24 Listeners Included in the Present Study

	Age (years)	Sex, F = female; M = male	WAIS Digit Span subtest (Scaled Score, normed for age)
Younger Listeners	28	F	8
	18	F	10
	19	F	10
	18	F	8
	21	F	12
	24	F	10
	20	F	10
	22	F	6*
	25	M	10
	22	F	10
	28	M	10
	28	M	9
Older Listeners	73	F	16
	65	F	8
	65	M	12
	64	F	15
	56	F	11
	65	F	12
	59	M	11
	57	F	13
	61	M	17
	72	M	9
	72	F	12
	61	M	14

Note: Scaled score mean=10; standard deviation=3; average= 7-13; *Indicates below average scaled score

2.2 Data Collection

2.2.1 Recording and selection of speech samples:

Speech samples included recordings from the Sentence Intelligibility Test (Yorkston et al, 1996). Six sentences of increasing length from each of the 9 TEP speakers were included (consisting of sentences with 5, 7, 9, 11, 13, and 15 words in length, resulting in 60 words per speaker). Speech samples were originally recorded in a sound-treated room using a headset

microphone (Shure PG-81, Shure, Niles, IL; or AKG-C20, AKG Acoustics, Vienna, Austria) with a 3-cm (offset) mouth-to-microphone distance. The microphone was connected to a preamplifier (M-Audio Fast Track Pro, Avid Technology, Burlington, MA), and the sample was acquired on a laptop computer using a specialized sound card and acoustic software (Sona-Speech II, Model 3650, KayPENTAX, Montvale, NJ). Additional samples were obtained in the same type of room, using the same microphone and mouth-to-microphone distance, but on a portable digital audiotape recorder (TASCAM DAPI, TASCAM, Montebello, CA). All speech samples were recorded at a sampling rate of 44.1 kHz with 16-bit quantization. Speech samples were transferred to a computer and converted into WAV files using acoustic software (Sony Soundforge, Sony Creative Software, Inc. Middleton, WI; Eadie et al., 2016).

Six sentences from each of the included speakers (from the SIT) were manually edited using Sony Soundforge acoustic software. The sentences 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 ms of silence was 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 9 TEP speakers included in this study.

A second set of stimuli was then mixed with a 4-talker babble (one male, and three females; Audiotec of St. Louis) from the Sentences in Noise (SIN) test. The intensity of the speech was held constant, and the noise was added to achieve the selected SNR. We used multi-talker babble as the noise because previous research found that meaningful speech competitors had a significantly more adverse effect on word recognition performance compared to non-meaningful competitors (e.g., white noise) (Sperry, Wiley, & Chial, 1997). In addition, multi-talker babble is representative of the most challenging adverse listening environment encountered in everyday speech communication situations (Gilbert, Tamati, & Pisoni, 2013).

One level of noise was added to all of the stimuli, resulting in two different versions of the same 6 sentences for each speaker: quiet (as described above) and +6dB SNR (representing an increased amount of background noise, but one that is 6dB lower than the signal). The noise level was chosen to reflect daily communication scenarios in which background noise is elevated, but did not previously demonstrate a ceiling or floor effect in younger listeners (Eadie et al., 2016; McColl et al., 1998). Of the speech samples mixed with noise, each audio file started with 500 ms of noise alone, followed by the speech signal mixed with the noise, and concluded

with 500 ms of noise alone (Van Engen & Bradlow, 2007). All speech samples were entered into a custom-made software program, which allowed the listeners to transcribe the sentences (speech intelligibility measure) and submit a rating of perceived listening effort.

2.2.2. Listener Procedures

The listening experiment for this study occurred in a quiet room at the University of Washington in the Vocal Function Laboratory, or in an alternative quiet setting of the participant's choice. In advance of presenting stimulus items, listeners received instructions and training on the protocol for transcribing speech samples and determining ratings of perceived listening effort.

Listeners were seated at a desktop or laptop computer and listened to instructions as well as experimental stimuli through headphones (Samson RH600). Each sentence set contained 6 sentences from the SIT. Each *listener* group (younger listeners and older listeners) was subsequently divided into four sub-groups, such that each speaker's intelligibility and perceived listening effort were based on averages across three inexperienced listeners in each listener group. This sample size has previously been shown to result in accurate intelligibility measures using transcription (Dahl et al., 2024). As a result, each listener listened to 4-5 speakers, with 2-3 of those in quiet and 2-3 speakers in noise (+6 dB) (see Table 3 below). This assignment ensured that no listener would hear a repeated stimulus (i.e., to avoid exposing a listener to the same sentence multiple times) and that each stimulus item was presented across each noise condition. Speakers were presented in a random order, but for each speaker, samples of increasing length (5, 7, 9, 11, 13, 15 word sentences) were presented as a block, consistent with the SIT protocol (Yorkston et al., 1996). For each sentence, listeners were asked to complete two tasks: (1) transcribe what they hear; and (2) provide a rating of perceived listening effort.

Table 3. Randomization of Listeners for the 9 Included Speakers

Listener Group	Quiet (~+20 dB SNR)	Noise (+6 SNR)
	Speaker (Sentence Set)	Speaker (Sentence Set)
Group 1	JJ(A), J(C), SG(F)	D(B), SN(D)
Group 2	SN(D), D(B)	SG(F), JJ(A), J(C)
Group 3	T(A), P(C)	II(B), SP(D)
Group 4	II(B), SP(D)	P(C), T(A)

Before the transcription task, both groups of listeners were provided adapted instructions from Eadie et al. (2016, p. 397) about the task and the speech samples: “You will be listening to adult speakers who have had total removal of their voice box due to cancer. These speakers are using a new method of speech called “tracheoesophageal speech”. We are interested in how well listeners can understand these speakers in both quiet and in background noise. You will hear samples presented in both quiet and in noise. We will play some sentences, and we would like you to type out the words that you hear. You may listen to the sentences up to two times. Some of these sentences will be difficult to understand. Do your best, and guess when you need to. You may listen to each sentence two times.” The transcription procedure described in these instructions is consistent with the protocol described in the *Sentence Intelligibility Test* (Yorkston et al., 1996).

After listening to the instructions, participants in both groups were presented with samples produced by two TEP speakers who were not otherwise included in the study. The first familiarization sample included one speaker reading a novel 12-word sentence in quiet, followed by the same sentence in noise. The listener was then provided with an additional sample, in both quiet and in noise, from another TEP speaker that immediately preceded the transcription and rating task. This protocol was used to prepare listeners to the general quality of TEP speech and allow them an opportunity to practice the transcription and rating task.

Ratings of perceived listening effort took place for each sentence immediately following each transcription. Listeners received information about listening effort prior to starting the task, using the adapted instructions from Nagle and Eadie (2012, p. 238): “After you have typed out the words, please rate your listening effort using the scale provided. Listening effort is the amount of work, attention, or concentration it takes you to understand a speech sample. To rate

the speech sample, drag the slider along the scale to indicate how effortful it was for you to understand the speaker”. A horizontally aligned visual analog scale was used, where 0= easy/no effort and 100= extremely effortful/ maximum effort (see Appendix A) All ratings took place in a single session that lasted approximately 45 minutes. Listeners were able to take breaks as needed to reduce the effects of fatigue and rater error.

3. Data Analysis

3.1 Descriptive Results

Listeners’ transcriptions from each group were compared with the target stimuli to determine speech intelligibility for each sentence for each speaker. For each sentence, intelligibility was averaged across 3 listeners who performed the transcriptions, using a total word phonemic match model scoring approach (Hustad & Cahill, 2003). A recent study showed that these values are accurate when there are at least 3 inexperienced listeners who transcribe a minimum of six sentences per speaker (Dahl et al., 2024). 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 criteria for scoring were based on the analysis described by Dahl et al. (2024). Words were marked correct if they were an exact match to the target, a homophone, a phonetically correct misspelling (e.g., “fone” for “phone”), a contraction error (e.g., “I’m for “I am” and vice versa), or an obvious spelling error (e.g., “specail” for “special”). Words were marked incorrect if they included an ambiguous spelling error (e.g., “ham” for “him”) or errors in plurality, verb tense, or possessive marking. If a listener transcribed all words in a sentence correctly but inserted words that were not present in the target sentence (e.g., “There is not enough oil” for “There is enough oil”), the denominator in the percent correct calculation was increased by 1 for each inserted word. 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 3 listeners per set x 60 words per speaker = 180 words per speaker for each quiet/noise condition) and each listener age group.

Ratings of perceived listening effort were similarly averaged for each individual sentence in each quiet/noise condition; these scores were then used to generate an aggregated score for each speaker as a function of each level of background noise in each listener group (i.e., scores

based on 3 listeners per set x 6 sentences = 18 ratings per speaker for each quiet/noise condition for each listener age group).

3.2 Reliability

To assess intrarater reliability for intelligibility and listening effort, 20% of sentences (i.e., 5-6 sentences) were repeated at the end of the listening session. Pearson's correlation coefficients were calculated to compare the consistency of the ratings within the same listener. The mean Pearson's correlation coefficient for speech intelligibility across the younger listening groups was 0.747 (SD = 0.388) and the mean Pearson's correlation coefficient for perceived listening effort was 0.737 (SD = 0.210). For the older listening group, the mean Pearson's correlation coefficient for speech intelligibility was 0.891 (SD = 0.131) and the mean Pearson's correlation coefficient for perceived listening effort was 0.830 (SD = 0.076).

To assess interrater reliability of transcriptions or perceived listening effort for each set of speakers evaluated by the listeners in each subgroup, intraclass correlation coefficients (ICCs) were calculated based on a mean rating, absolute agreement, two-way random effects model (2, k; Shrout & Fleiss, 1979). The ICCs for speech intelligibility across the younger listening groups ranged from 0.763 to 0.900, with an overall mean of 0.843. The ICCs for perceived listening effort across the younger listening groups ranged from .722 to .922, with an overall mean of 0.829. The ICCs for speech intelligibility across the older listening groups ranged from 0.889 to 0.913, with an overall mean of 0.903. The ICCs for perceived listening effort across the older listening groups ranged from 0.621 to 0.936, with an overall mean of 0.787. These levels are consistent with prior research and are acceptable levels for further data analysis (Sussman & Tjaden, 2012).

3.3 Analysis of Experimental Questions

To determine the effect of listener age (e.g., younger vs. older listener group) and noise condition (e.g., quiet and +6dB SNR) on speech intelligibility and perceived listening effort, two 2 X 2 mixed ANOVAs were performed. Noise (two within-subjects levels) and listener age (two between-subjects levels) were set as fixed effects, and speaker was set as random effects. Assumptions of sphericity and normality were met before interpreting any main or interaction effects.

4. Results

4.1 Descriptives: Speech Intelligibility and Perceived Listening Effort in Quiet and in Noise

The mean speech intelligibility (%) and perceived listening effort (mm) averaged across speech samples used in the listening experiment are shown in Table 4. The average intelligibility scores of the group of speakers transcribed by the younger listeners was 94.50% in quiet (SD = 3.62) and 69.39% (SD = 15.90) in noise. The average intelligibility scores of the group of speakers transcribed by the older listeners was 95.75% in quiet (SD = 2.01) and 61.04% in noise (SD = 24.19).

The average perceived listening effort scores of the speakers rated by the younger listeners was 14.95 mm in quiet (SD = 7.58) and 55.26 mm in noise (SD = 21.87). The average perceived listening effort scores of the speakers rated by the older listening group was 11.28 mm in quiet (SD = 5.02) and 62.71 mm (SD = 22.08) in noise. The means and standard deviations for each outcome measure for each of the 9 TEP speakers, as rated by the listener groups and across conditions, are presented in Table 4.

Table 4. Mean speech intelligibility (%) and perceived listening effort ratings (mm) for tracheoesophageal speakers in quiet and in noise when rated by younger and older listeners.

Speaker	Listener Group	Intelligibility		Listening Effort	
		Quiet	Noise	Quiet	Noise
JJ	Young	95.5 (6.10)	44.31 (23.68)	25 (9.81)	89.50 (6.94)
	Old	94.92 (6.05)	26.25 (17.30)	12.50 (11.75)	95.17 (6.40)
D	Young	95.94 (4.90)	61.51 (33.78)	21.67 (9.29)	53.96 (14.36)
	Old	97.86 (5.23)	37.79 (40.00)	5.59 (7.33)	52.28 (24.51)
J	Young	95.77 (6.64)	66.61 (24.96)	8.83 (7.56)	56.11 (25.23)
	Old	95.76 (5.79)	61.69 (23.99)	10.06 (13.03)	72.00 (22.52)
SN	Young	100.00 (0.00)	67.45 (34.08)	8.83 (5.15)	47.48 (6.91)
	Old	94.41 (7.15)	58.71 (23.20)	4.28 (3.51)	54.00 (20.86)
SG	Young	88.77 (14.40)	68.01 (22.85)	25.96 (16.43)	72.22 (10.54)
	Old	93.52 (8.33)	59.43 (26.36)	14.17 (13.92)	68.56 (20.760)
T	Young	91.85 (13.38)	54.14 (19.49)	14.67 (16.27)	78.72 (15.92)
	Old	93.95 (7.63)	40.47 (30.29)	19.72 (10.33)	90.56 (10.20)
II	Young	90.20 (12.58)	81.75 (12.69)	14.56 (20.16)	45.94 (21.29)
	Old	94.51 (7.58)	75.53 (16.67)	14.61 (7.83)	65.39 (19.68)
P	Young	94.43 (6.36)	95.30 (3.76)	9.39 (12.23)	22.72 (14.22)
	Old	99.01 (1.60)	91.41 (10.23)	13.78 (12.56)	30.00 (15.83)
SP	Young	97.78 (5.44)	85.46 (22.58)	5.67 (5.65)	30.61 (12.18)
	Old	97.93 (2.54)	98.09 (3.57)	6.78 (6.67)	36.44 (12.13)

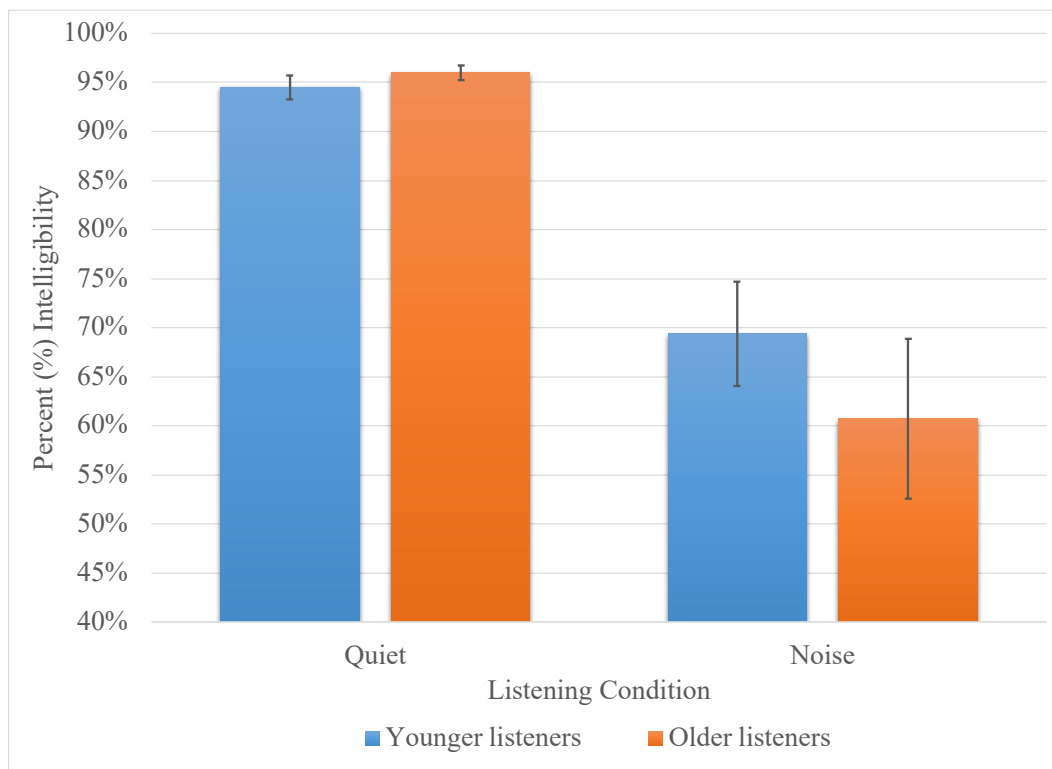
Note. The values are mean values across all sentences, averaged across three listeners per speaker per condition. The values in parentheses are standard deviations.

4.2 Effect of Background Noise and Age of Listener on Speech Intelligibility

Mean intelligibility scores for the two listener groups are reported across both noise conditions (quiet and +6dB SNR) in Figure 1. A 2 X 2 mixed ANOVA was calculated to examine the effect of listener age and noise condition on speech intelligibility. First, results revealed a significant main effect of noise, $F(1,16) = 39.816$, $p < .001$, $h^2p = .713$. Specifically, speech intelligibility was significantly higher for both listener groups when speakers were evaluated in quiet (mean intelligibility = 95.22%; SD = 2.92%) than in +6dB SNR background noise (mean intelligibility = 65.22%; SD = 20.32%), with a large effect size ($h^2p > 0.14$). However, there was no significant main effect for age of listener $F(1,16) = 0.501$, $p = .489$ (mean intelligibility = 81.94%; SD = 9.76% for the younger group; mean intelligibility = 78.34%; SD =

13.37% for the older group). Importantly, no significant interaction was found between age of listener and noise, $F(1,16) = 1.026, p = .326$. As can be seen in Figure 1, speech intelligibility significantly decreased with increased noise in both listener groups. While speech intelligibility decreased more in the older listener group in the presence of noise (mean decrease from 94.5% to 69.39% intelligible for the younger listener group; mean decrease from 95.97% to 60.71% for the older listener group), the difference was not statistically significant.

Figure 1. Mean speech intelligibility (%) scores of older and younger listeners across both noise conditions (quiet and +6dB SNR). The values are in percentages (%), where 0% = no words understood and 100% = all words understood.



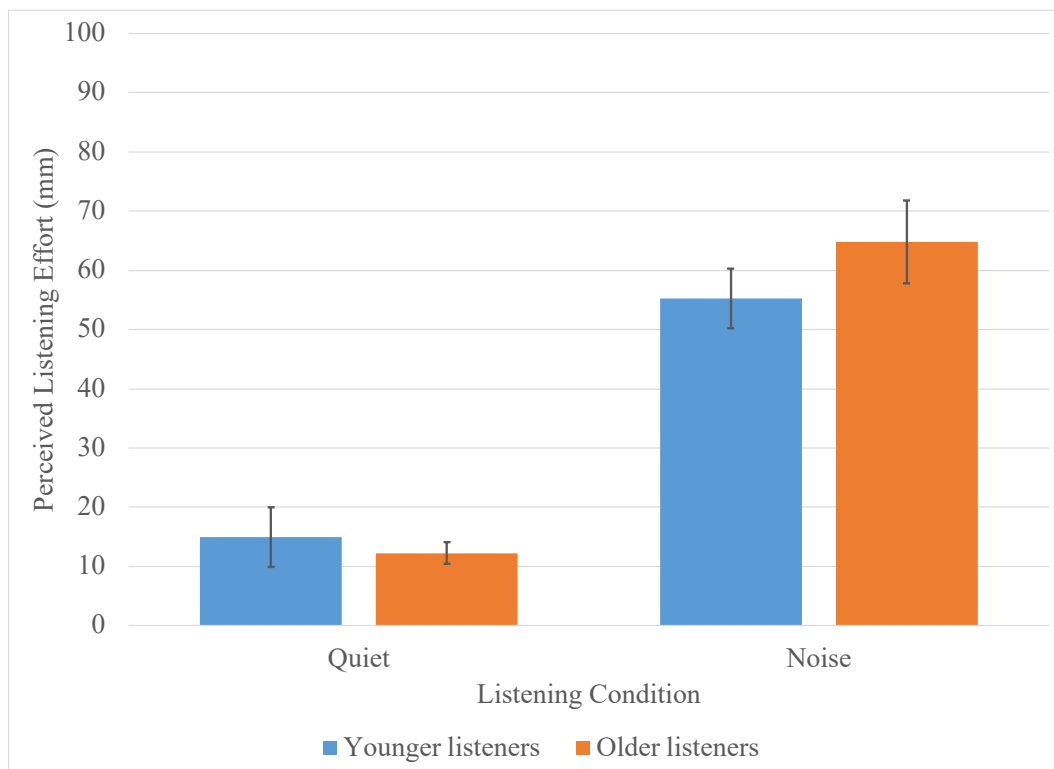
Note. Error bars are standard errors

4.3 Effect of Background Noise and Age of Listener on Perceived Listening Effort

The mean perceived listening effort (PLE) scores for the two listener groups are reported across both noise conditions (quiet and +6dB SNR) in Figure 2. A 2 X 2 mixed ANOVA was calculated to examine the effect of listener age and background noise on perceived listening

effort. Consistent with the findings for speech intelligibility, a significant main effect of noise was found, $F(1,16) = 109.099, p < 0.001$, with a large effect size ($h^2p = 0.872$). Specifically, perceived listening effort was significantly higher when listeners rated TEP speakers in background noise (average PLE = 58.99mm; SD = 21.66mm) than in quiet (average PLE = 13.11mm; SD = 6.52mm). However, like the speech intelligibility results, there was no significant main effect for age of listener, $F(1,16) = 0.91, p = .767$ (mean PLE = 35.11mm, SD = 14.72mm for the younger listener group; mean PLE = 38.50mm, SD = 13.21mm for the older listener group). Finally, no significant interaction was found between age of listener and noise, $F(1,16) = 1.604, p = .223$. As can be seen in Figure 2, perceived listening effort scores were significantly lower for both listener groups in the quiet condition than when speakers were rated in noise. While perceived listening effort increased more for the older listeners in noise (mean increase from 14.95mm to 55.26mm for the younger listeners; mean increase from 11.28mm to 62.71mm for the older listeners), the difference was not statistically significant.

Figure 2. Mean perceived listening effort scores of speakers using tracheoesophageal speech, given by older and younger of listeners in both noise conditions (quiet and +6dB SNR). The values are in mm, where 0= easy/no effort and 100= extremely effortful/ maximum effort.



Note. Error bars are standard errors.

5. Discussion

The purpose of this study was to identify the effect of background noise and the age of the listeners on two related, but potentially differentiated outcome measures - speech intelligibility and perceived listening effort - for 9 TEP speakers. Overall, the results showed that no matter the age of the listener, TEP speakers were significantly less intelligible in noise, and that listeners perceived increased listening effort to understand the speakers in noise. Importantly, in the group of listeners recruited in our study, age was not a significant factor, and it did not significantly interact with noise. That is, while older listeners tended to have more difficulty understanding TEP speakers with more perceived listening effort than younger listeners in noise, these differences were not significantly different. How the results from the present study compare to previous findings, and what this means for future research and clinical practice, is discussed next.

5.1.1 Effect of Background Noise and Age of Listener on Speech Intelligibility

Consistent with prior findings, the presence of background noise in this study significantly impacted intelligibility for TEP speakers, on average (Clark, 1985; Clark & Stemple, 1982; Eadie et al., 2016; McColl et al., 1998). Both younger and older listener groups showed a notable decline in intelligibility from the quiet condition to the background noise condition (+6 dB SNR) (95.23% in quiet to 65.05% in noise), supporting the conclusion that even relatively modest levels of background noise can impair a listener's ability to accurately understand TEP speech. These results confirm the impact of adverse noise conditions on intelligibility for this speaker population, regardless of the age of their communication partner.

Results from the present study differ from two previous studies, which found that older listeners rated alaryngeal speakers as significantly less intelligible than younger listeners (Clark, 1985; Law et al., 2009). First, Law et al. (2009) showed that a group of 3 older listeners (ages 60 to 65 years) were significantly less accurate than a group of 3 younger listeners (ages 25 to 33 years) for ratings of speech intelligibility in 49 alaryngeal speakers. The differences in the listener groups occurred despite the authors' control of hearing status. However, upon further inspection, it is not clear whether post hoc analyses were completed for alaryngeal speaker types in their study (Law et al., 2009). For example, although a significant main effect of listener age

was reported, it appears that it was driven primarily by ratings of speakers who used electrolaryngeal speech (e.g., average intelligibility = 77.3%, SD = 7.01% for the younger judges; average intelligibility = 62.0%, SD = 8.90% for the older judges) than for their group of 15 TEP speakers (average intelligibility = 61.5%, SD = 6.82% for the younger judges; average intelligibility = 59.1%, SD = 11.11% for the older judges). In other words, results from the current study might be consistent with those reported by Law et al. (2009), in that listener age did not appear to be a factor in their assessment of intelligibility for their group of TEP speakers. Another difference between the present study and Law et al. (2009) that could also account for the discrepancy in our results is that the speakers in that study spoke Cantonese, a tone language, where words that have the same sound but differ in pitch have different meanings, which may have further impacted intelligibility.

Results from the present study also are at odds with findings reported by Clark (1985), who also studied the effect of age of the listener on speech intelligibility. In contrast to the current study, Clark (1985) only used one representative speaker for each alaryngeal speech type, along with one laryngeal speaker. In that study, groups of older and younger listeners performed similarly for speech intelligibility when the one TEP speaker was presented in quiet; however, when evaluated in noise (e.g., at -5 dB MCR), the TEP speaker was significantly less intelligible to the older listeners when compared to the younger listeners. These results also occurred even though all listeners in their study passed hearing tests.

In the present study, there was a numerical difference in how much intelligibility declined when evaluating the TEP speakers in quiet to noise, with older listeners showing a greater average decrease (from 95.97% to 60.71%) compared to younger listeners (from 94.50% to 69.39%); however, this difference was not statistically significant. It is interesting to note that within the noise condition in our study, older listeners appeared to show more variability as a group (which can be seen in the error bars provided in Figure 1, and also in the larger SDs for each group for speech intelligibility). The variability in the noise condition as a whole, as well as within the older group specifically, may have played a role in obscuring any potential age-related effects at this noise level. Further study with more listeners in each group might help reveal whether this effect may have been due to reduced statistical power.

In this study, hearing status was grossly controlled in an effort to isolate age as a factor for two listener-rated outcomes. The results of our study were consistent with previous research

in dysarthria that shows when hearing status is controlled among older and younger listeners, there is no effect of the age of the listener on speech intelligibility or speech perception accuracy (Dagenais et al., 1999; Ingvalson et al., 2017). However, although Ingvalson et al. (2017) did not find differences between their older and younger listeners in the accuracy of their performance, they found that older adults may have been using more cognitive support, such as working memory and receptive vocabulary, to perform their speech perception task. As a result, we also sought to examine similar questions about a listener's age using a second listener-rated outcome measure: perceived listening effort.

5.1.2 Effect of Background Noise and Age of Listener on Perceived Listening Effort

As expected, results from this study showed that perceived listening effort significantly increased when TEP speech was presented in noise. The effect size for noise was large, with listeners rating their effort as significantly higher in the +6 dB SNR condition than in quiet. These findings suggest that even when speech remains somewhat intelligible (mean intelligibility in noise = 65.22%), the listening task becomes considerably more effortful, particularly for TEP speakers whose alaryngeal voice production may already challenge the listener's perceptual system.

Similar to the results for speech intelligibility, this study also showed no significant effect of listener age or an age by noise interaction for perceived listening effort. While older listeners reported slightly greater perceived listening effort in noise (62.71mm) compared to younger listeners (55.26mm), and despite a small decrease in perceived listening effort for older listeners in quiet (11.28mm vs 14.95mm), differences between the listener groups as a whole or within the noise condition alone were not statistically significant. This finding is consistent with the results for speech intelligibility and suggests that both listener groups were similarly impacted by the addition of background noise.

As already noted, in this study, we controlled hearing status to isolate the effects of listener age on the listener-rated outcomes. In their study with dysarthric speakers, Ingvalson et al. (2017) similarly controlled hearing status. They also did not find differences between their older and younger listeners in the accuracy of their speech perception task. However, they found that older adults may have been using more cognitive support to perform their speech perception task; their regression analysis showed that factors such as working memory and receptive

vocabulary significantly predicted speech perception scores, whereas only hearing status predicted the scores in younger listeners (Ingvalson et al., 2017). For this reason and contrary to the current results, we expected that the older listener group might have demonstrated higher perceived listening effort ratings than the younger listeners, even if speech intelligibility performance was the same.

One reason why we might not have found differences between our listener groups could relate to the characteristics of our recruited listeners. For example, the standard scale scores for the WAIS Digit Spans (Wechsler, 2008) were significantly higher in our older listener group (average = 12.5, SD = 2.68) than in our younger listener group (average = 9.42, SD = 1.51), despite the fact that only one listener in the younger listener group was slightly below the expected score for their age $t(17.31) = -3.475, p = 0.003$. The increased cognitive capacity in our older group may therefore have allowed our listeners to perform at a higher level than expected. Matching our two groups for this variable (or controlling it statistically) in the future might help reveal whether this factor could affect the results for either perceived listening effort or intelligibility as it relates to age.

5.1.3 Effect of noise on speech intelligibility and perceived listening effort: Summary

Together, the findings for speech intelligibility and perceived listening effort in this group of TEP speakers reveal important insights into how background noise impacts unfamiliar communication partners. The data show a clear pattern; multi-talker background noise (+6 dB SNR) significantly impacted the listeners' abilities to decipher the messages of the TEP speakers, resulting in decreased overall intelligibility and increased listening effort for both younger and older listener groups. This was shown in the substantial overall drops in intelligibility (95.22% in quiet to 65.22% in noise) and corresponding increases in perceived listening effort (13.11mm in quiet to 58.99mm in noise) across both listener groups. These results confirm that TEP speech becomes perceptually demanding in the presence of background noise, even when speakers are highly intelligible in quiet.

With regard to the effect of background noise, the results of this study closely align with McColl et al. (1998), who reported that a proficient TEP speaker was consistently less intelligible than a healthy control speaker in noise, and that the most pronounced differences emerged at +5 dB SNR. The current findings also are consistent with McAuliffe et al. (2009),

who showed that dysarthric speakers experienced notable drops in intelligibility from quiet to +6 dB SNR, while their control speakers were not significantly affected until background noise reached 0 dB SNR. These patterns point to a sensitive threshold, varying by speaker type, at which listeners begin to experience pronounced difficulty in understanding a person with a speech disorder.

Results from this study differ from those reported by Ishikawa et al. (2017), who found that the strongest differences in speech intelligibility between their speaker groups (dysphonia vs. control) took place at a noise condition of 0 dB SNR. The higher level of noise in that study may have been necessary to reveal group differences due to the milder nature of dysphonic deficits. In the present study, however, the unique characteristics of TEP speech make these speakers more vulnerable to even modest degradations in signal quality and the effects of competing background noise. While the underlying speech deficits in these studies differ, the consistent theme is that adverse listening conditions disproportionately penalize speakers with disordered speech.

Interestingly, while listener age was not a statistically significant factor in either intelligibility or perceived listening effort in this study, trends in the data might suggest that older listeners could potentially be more affected by noise. For example, older listeners showed a larger mean decrease in intelligibility (95.75% to 61.04%) and a larger increase in perceived listening effort (11.28mm to 62.71mm) compared to younger listeners (94.50% to 69.39% intelligible; 14.95mm to 55.26mm effort) when scores were compared from quiet to noise. In addition, it is interesting that this pattern was strongly consistent across both outcome measures in noise. Older listeners demonstrated poorer speech intelligibility and increased perceived listening effort for 8 out of 9 speakers for speech intelligibility, and 7 out of 9 speakers for perceived listening effort in noise (see Table 2). In other words, there did not appear to be any outliers who masked potential differences. Further research on whether increased levels or types of noise might differentially affect listeners of different ages may be warranted.

These findings emphasize the idea that a relatively small amount of background noise can have a disproportionately large impact on intelligibility and perceived listening effort for individuals who use TEP speech. This suggests a threshold at which the properties of alaryngeal speech, paired with environmental factors (i.e., multi-talker background noise), begin to exceed the perceptual capacity of the unfamiliar listener. Results from this study also showed that as a

group, speech intelligibility was related (negatively) to perceived listening effort, with similar findings across the two outcome measures. However, similar to previous studies, it appears that listeners used these constructs differently (Nagle & Eadie, 2012, 2018). For example, for both the younger and older listener groups, there were three speakers (II, P, and SP) who showed less than 20% decreases in intelligibility from quiet to noise, but except for speaker P (for the younger listener group), listeners rated their perceived listening effort for these same speakers as substantially increased ($>20\text{mm}$) from quiet to noise. The divergence between intelligibility and perceived listening effort for these particular speakers further emphasizes that PLE captures aspects of the communication experience not accounted for by intelligibility alone, particularly the cognitive-perceptual cost of understanding atypical speech in noise.

5.2 Limitations

Several limitations of this study need to be considered when interpreting the results. First, this study only included 9 TEP speakers, all of whom were highly intelligible in quiet. The small number of speakers in this study might limit generalizability of the results, as there could have been individual speakers that affected group results. Looking at the average performance in intelligibility and perceived listening effort, however, it does not appear that there were outliers (see Table 4). In addition, while there were only a limited number of speakers, the performance of each speaker was based upon the average of 3 listeners per age group and per noise condition (i.e., 3 listeners per set $\times$ 60 words per speaker = 180 words per speaker for each quiet/noise condition and each listener age group). Using the averages and this many ratings for each condition helps reduce the influence of potential outlier listener ratings on speech intelligibility and perceived listener effort ratings (Dahl et al., 2024; Shrivastav, Sapienza, & Nandur, 2005).

Second, the results need to be interpreted in the context of the speakers who were included in this study. Specifically, speakers included in this study were selected from a previously collected data base and were previously judged to be highly intelligible in quiet, (i.e., all were $>94\%$ in quiet) (Eadie et al., 2016). It is likely that TEP speakers with lower speech intelligibility in quiet would be more vulnerable to the effects of background noise. It is also worth noting that all speakers selected in this study were male to control for gender; therefore caution should also be exercised when generalizing these findings to speakers who are less intelligible in quiet as well as female TEP speakers. This is an important consideration as the

proportion of females who have undergone total laryngectomies has increased in the past few decades (ACS, 2025)

The ecological validity of the design used in this study may also have been limited due to the mismatch between the speaker and listener conditions. In this study, speech samples were recorded in a quiet environment and the background noise was artificially added afterwards (Ishikawa et al., 2017). This type of experimental manipulation may not be entirely representative of the compensations that speakers make when they communicate with others in everyday 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 of typical speakers have shown that compensating for background noise with Lombard speech may enhance intelligibility (Garnier, Løevenbruck, Welby, & Bailly, 2006). However, McColl (2006) found that the Lombard effect may actually negatively impact the intelligibility of TEP speakers in noise; as TEP speakers increase their effort to produce speech with higher intensity, it may become increasingly strained and less intelligible. In that study, three out of four TEP speakers were negatively impacted when producing speech in increased background noise. Future studies should therefore investigate how speakers compensate for the presence of background noise by obtaining real-time recordings (instead of adding the noise to the recordings).

The level of background noise selected for this study may also serve as a limitation. By only using +6 dB SNR when having listeners make ratings in background noise, the speakers' signals were always presented at a higher sound intensity than the background noise. While these conditions were selected to capture realistic environments in which background noise is present but not completely overpowering the speaker (e.g., comparable to speaking with someone in a restaurant or busy cafe), this is not completely representative of the environments TEP speakers may be in that are much louder (e.g., a concert or a large social gathering). The levels in this study were deliberately chosen to avoid a floor effect for all speakers, where the noise could become so difficult that all speakers would be equally impossible to transcribe. However, it is unclear whether the findings from +6 dB SNR would generalize to more demanding listening environments and needs future study. In addition, it is not known whether the use of an increasingly noisy environment might be more sensitive to revealing potential age-related differences.

Lastly, the demographics of the listener may also be considered a limitation. In this study, there were more female listeners than male listeners. While more women than men participated overall, no effect of gender or age has been found in previous studies investigating ratings of speech intelligibility (Eadie et al, 2021; Pennington & Miller, 2007) and did not appear to be a factor in this study either. However, this factor should still be taken into consideration when generalizing these results. Each of the listeners in the older listening group also had hearing thresholds lower than 40 dB HL at 500, 1000, 2000, and 4000 Hz and WAIS Digit Spans scores (Wechsler, 2008) that were within functional limits. In a study on cognitive predictors of perception of dysarthric speech in older adults, Lansford et al. (2025) found that older adults (ages 60-80) with better vocabulary knowledge, working memory, and cognitive flexibility had better speech intelligibility ratings. In this study, sentences were selected at random from the SIT (Yorkston et al., 1996), but Cannito et al. (2012) has found that some sentences are more predictable than others, which in turn allows both listening groups to use more top-down strategies to increase intelligibility scores (Beverly et al., 2010). As a result, caution should be exercised when generalizing these results to adults with hearing loss and/or cognitive changes that often happen with aging.

5.3 Clinical Implications

Very few studies to date have examined relationships among background noise, and its impact on speech intelligibility and listening effort when listener age is taken into consideration (Clark, 1985; Law et al., 2009). This area of research is particularly important because on average, individuals who have had total laryngectomies are 66 years old when they receive their surgery, and as a result, communication partners are often around similar ages. While the results of this study did not show a statistically significant difference in intelligibility ratings and perceived listening effort based on the age of listener, it is important to consider factors related to the communication partner/listener when making treatment decisions.

This study continues to highlight the differences between listener judgements performed in quiet conditions versus those collected in a more natural environment with background noise (e.g., a restaurant or busy cafe). As clinicians, we often perform speech evaluations and treatment in quiet. However, this data shows a significant gap between quiet and noisy conditions and overall communicative success. By continuing to assess and treat individuals in

quiet conditions, specifically with TEP speakers, we risk the possibility of failing to prepare our patients for the difficulties that may encounter outside the clinic, when background noise is present. We may also be failing to prepare familiar communication partners on the increased challenges and cognitive load that may be required in more naturalistic settings. As a result, there is a risk that many communication barriers for both speakers and listeners may not be factored into education, counseling and treatment for these patients.

When individuals are diagnosed with laryngeal cancer and undergo a total laryngectomy, there are many considerations that speech-language pathologists might consider incorporating into counseling:

- a) Communicating in quiet environments appears to be the most effective for enhancing speech intelligibility and decreasing the demands of perceived listening effort from unfamiliar communication partners. While communicating without background noise is not always feasible, it improves the chances of being understood and lowers listener burden significantly. This may play a role in psychosocial factors throughout the diagnosis and recovery process (e.g., being understood and successfully communicating one's message).
- b) When background noise is present, TEP speakers are penalized by background noise to a greater extent than those with typical laryngeal speech. Their speech signal becomes more degraded, the integrity of their message is reduced, and listener burden is increased. These listening conditions may place heavier burdens on both listeners and TEP speakers.

Familiar communication partners may be able to understand more of the speaker's message than the unfamiliar listeners used in this study. Even so, TEP speakers and their communication partners may benefit from education to facilitate effective communication across different listening conditions. Current best practices adapted from literature on hearing loss 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 with proper lighting, using non-verbal cues or gestures, signaling topics, repairing breakdowns by restating what has been understood so far, and incorporating breaks to prevent fatigue (ASHA, 2015). These considerations are important, particularly because the present study only included audio presentation of the speech samples and did not consider any additional benefit that might be gained from non-verbal cues, being familiar with the speaker, having the

context of the message, or other facilitators that might improve speech intelligibility and lessen listening burden.

5.4 Directions for Future Research

There continues to be room for further research of TEP speakers and overall communicative success in real-world environments, outside of the clinic. Future research with this population may lead to more direction on how to best help TEP speakers improve overall intelligibility and communication outcomes when background noise is present. Specifically, obtaining real-time recordings, where the speaker is speaking in the presence of background noise may provide further information on the Lombard effect and how this may impact TEP speakers. This may provide further information on treatment interventions to promote intelligibility, such as those found in “clear speech” (McColl, 2006). However, Searl (2007) cautioned that the implementation of these strategies may expand greater physical effort during speech production. Further, research indicates that TEP speakers who expend additional effort may increase their overall fatigue, leading to withdrawal and talking less (Searl & Knollhoff, 2018). Therefore, recording these samples in real-world environments may provide further insight on the impacts of the increased effort for the speaker, which is also critical when determining communicative success, and overall quality of life for TEP speakers.

References

- American Cancer Society. (2025). Key statistics for laryngeal and hypopharyngeal cancers. Retrieved from <https://www.cancer.org/cancer/types/laryngeal-and-hypopharyngeal-cancer/about/key-statistics.html>
- 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, 801–806.
- 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
- Bornbaum, C. & Doyle, P. (2019). Distress as a consequence of head and neck cancer. In Doyle (Eds.) *Clinical care and rehabilitation in head and neck cancer* (pp. 69-85). Springer Publications.
- 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.
- Campbell, B. H., Marbella, A., & Layde, P. M. (2000). Candidate's thesis: Quality of life and recurrence concern in survivors of head and neck cancer. *The Laryngoscope*, 110, 895–906.
- Cannito, M., Suiter, D., Beverly, D., Chorna, L., Wolf, T., Pfeiffer, R. (2012). Sentence intelligibility before and after treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice*, 26(2), 214-219.
- 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.
- Clark, J.G. (1985). Alaryngeal speech intelligibility and the older listener. *Journal of Speech and Hearing Disorders*, 50, 60-65.
- Clark, J.G., & Stemple, J. C. (1982). Assessment of three modes of alaryngeal speech with a synthetic sentence identification (SSI) task in varying message-to-competition ratios. *Journal of Speech and Hearing Disorders*, 25, 333-338.
- 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
- Dagenais, P. A., Watts, C., Turnage, L., & Kennedy, S. (1999). Intelligibility and acceptability of moderately dysarthric speech by three types of speakers. *Journal of Medical Speech-Language Pathology*, 7, 91–96.
- Dahl, K. L., Balz, M.A., Cádiz, M. D., & Stepp, C. E. (2024). How to efficiently measure the intelligibility of people with parkinson's disease. *American Journal of Speech-Language Pathology*, 1-15.
- D'Alatri, L., Bussu, F., Scarano, E., Paludetti, G., & Marchese, M. R. (2012). Objective and subjective assessment of tracheoesophageal prosthesis voice outcome. *Journal of Voice*, 26, 607–613. doi:10.1016/j.jvoice.2011.08.013
- Doyle, P. C., & Keith, R. L. (Eds.). (2005). Contemporary considerations in the treatment and rehabilitation of head and neck cancer: *Voice, speech, and swallowing*. Austin, TX: Pro-Ed.
- Eadie, T. L., Day, A. M. B., Sawin, D. E., Lamvik, K., & Doyle, P. C. (2013). Auditory-perceptual speech outcomes and quality of life after total laryngectomy. *Otolaryngology—Head & Neck Surgery*, 148, 82–88.
- Eadie, T.L., Durr, H., Sauder, C., Nagle, K., Kapsner-Smith, M., & Spencer, K. (2021) Effect of noise on speech intelligibility and perceived listening effort in head and neck cancer. *American Journal of Speech-Language Pathology*, 30, 1329-1342.
- 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.
- Evans, E., Carding, P., & Drinnan, M. (2009). The Voice Handicap Index with post-laryngectomy male voices. *International Journal of Language & Communication Disorders*, 44, 575–586.
- Garnier, M., Dohen, M., Løevenbruck, 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. doi:10.3766/jaaa.24.1.4
- Gordon-Salant, S., & Fitzgibbons, P. J. (1995). Recognition of multiply degraded speech by young and elderly listeners. *Journal of Speech and Hearing Research*, 38(5), 1150–1156.
- Gosselin, P. A., & Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, 54(3), 944-958.
- Hillman, R. E., Walsh, M. J., & Heaton, J. T. (2005). Laryngectomy speech rehabilitation: A review of outcomes. In P. C. Doyle & R. L. Keith (Eds.), *Contemporary considerations in the treatment and rehabilitation of head and neck cancer: Voice, speech, and swallowing* (pp. 75–90). Austin, TX: Pro-Ed.
- 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.
- 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, 198–208.
- Ingvalson, E. M., Lansford, K. L., Fedorova, V., & Fernandez, G. (2017). Receptive vocabulary, cognitive flexibility, and inhibitory control differentially predict older and younger adults' success perceiving speech by talkers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 60, 3632-3641.
- 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
- Kazi, R., Sayed, S. I., & Dwivedi, R. C. (2010). Post laryngectomy speech and voice rehabilitation: Past, present and future. *ANZ Journal of Surgery*, 80, 770–771.

- 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, 482–499.
- Knott, J. (2019). Clinical problem-solving in tracheoesophageal puncture voice restoration. In Doyle (Eds.) *Clinical care and rehabilitation in head and neck cancer* (pp. 189-208). Springer Publications
- 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.
- Lansford, K., Hirsch, M., Barrett, T., Borrie, S. (2025). Cognitive predictors of perception and adaption to dysarthric speech in older adults. *Journal of Speech, Language and Hearing Research*, 68(7S), 3507-3524. https://doi.org/10.1044/2024_JSLHR-24-00345
- Law, I. K.-Y., Ma, E. P.-M., & Yiu, E. M.-L. (2009). Speech intelligibility, acceptability, and communication-related quality of life in Chinese alaryngeal speakers. *Archives of Otolaryngology-Head & Neck Surgery*, 135, 704–711.
- 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>
- 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.
- Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*, 18(7), 604–617.
- McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., O'Beirne, G. A., & Anderson, T. (2017). Effect of dysarthria type, speaking condition, and listener age on speech intelligibility. *American Journal of Speech-Language Pathology*, 26(1), 113–123. https://doi.org/10.1044/2016_AJSLP-15-0182
- McAuliffe, M. J., Gibson, E. M., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *The Journal of the Acoustical Society of America*, 134(2), 1358–1368.
- 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.
- 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; quiz 288–289.
- 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.
- 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. (2018), Perceived listener effort as an outcome measure for disordered speech *Journal of Communication Disorders*, 73, 34-49.
- Nagle, K. F., & Eadie, T. L. (2012). Listener effort for highly intelligible tracheoesophageal speech. *Journal of Communication Disorders*, 45(3), 235-245.
- National Cancer Institute. (2024). Laryngeal cancer treatment (PDQ)- health professional version. Retrieved from: <https://www.cancer.gov/types/head-and-neck/hp/adult/laryngeal-treatment-pdq>
- 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>
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1), 593–608.
- Schneider, B. A., Daneman, M., Murphy, D. R., & See, S. K. (2000). Listening to discourse in distracting settings: *The effects of aging. Psychology and Aging*, 15(1), 110–125.
- Schneider, B., Speranza, F., & Pichora-Fuller, M. K. (1998). Agerelated changes in temporal resolution: Envelope and intensity effects. *Canadian Journal of Experimental Psychology*, 52(4), 181-194.
- Searl, J. (2007). Bilabial contact pressure and oral air pressure during tracheoesophageal speech. *Annals of Otology, Rhinology & Laryngology*, 116, 304–311.
- Searl, J., & Knollhoff, S. (2018). Sense of effort and fatigue associated with talking after total laryngectomy. *American Journal of Speech-Language Pathology*, 27(4), 1434-1444. https://doi.org/10.1044/2018_AJSLP-17-0218
- Shrivastav, R., Sapienza, C., & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *Journal of Speech, Language, and Hearing Research*, 48(2), 323-335. [https://doi.org/10.1044/1092-4388\(2005/022\)](https://doi.org/10.1044/1092-4388(2005/022))
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Singer, M. I., & Blom, E. D. (1980). An endoscopic technique for restoration of voice after laryngectomy. *Annals of Otology, Rhinology, & Laryngology*, 89, 529–533.
- Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing aid processed speech. *Auditory Cognitive Neuroscience*, 6, 526.
- 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.
- Tardy-Mitzell, S., Andrews, M. L., & Bowman, S. A. (1985). Acceptability and intelligibility of

- tracheoesophageal speech. *Archives of Otolaryngology*, 111, 213–215.
- van As, C. J., Koopmans-van Beinum, F. J., Pols, L. C. W., & Hilgers, F. J. M. (2003). Perceptual evaluation of tracheoesophageal speech by naive and experienced judges through the use of semantic differential scales. *Journal of Speech, Language, and Hearing Research*, 46, 947–959.
- 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>
- Wechsler, David (2008). *Wechsler Adult Intelligence Scale*. Pearson.
- Whitehill, T. L., & Wong, C. C. Y. (2006). Contributing factors to listener effort for dysarthric speech. *Journal of Medical Speech-Language Pathology*, 14(4), 335-342.
- Yorkston, K., & Beukelman, D. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. R. Berry (Ed.), *Clinical dysarthria* (pp. 155–163). San Diego, CA: College-Hill Press.
- Yorkston, K. M., Beukelman, D. R., Hakel, M., & Dorsey, M. (1996). *Speech intelligibility test for windows*. Lincoln, NE: Communication Disorders Software.

Appendix A: Instructions for Listening Session

Click play to hear the audio sample.

In the experiment, you will listen to the speech sample, and type out the words that you hear. You may listen to each sentence up to TWO times. Click on the "arrow" button to play the sample. Please listen to the entire sentence each time. If you are uncertain about a word, please make your best guess. If you are unable to guess, please use a dash (--) to represent the unknown word.

If there are several consecutive words you do not know, put a space between the dashes to represent each word (e.g., The ----- went down the road).

Type the words in the sentence here.

Please rate the speaker for LISTENING EFFORT. Listening effort is the amount of work, concentration, or attention it takes you to understand the speaker.

Easy / no listening effort  Extremely effortful / Maximum effort

