ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

The Application of Time–Frequency Masking To Improve Intelligibility of Dysarthric Speech in Background Noise

Stephanie A. Borrie,[a] Sarah E. Yoho,[a,b] Eric W. Healy,[b] and Tyson S. Barrett[c]

[a] Department of Communicative Disorders and Deaf Education, Utah State University, Logan [b] Department of Speech and Hearing Science, The Ohio State University, Columbus [c] Department of Psychology, Utah State University, Logan

ARTICLE INFO

Article History:
Received September 24, 2022
Revision received December 13, 2022
Accepted January 10, 2023

Editor-in-Chief: Peggy B. Nelson
Editor: Christian E. Stilp

https://doi.org/10.1044/2023_JSLHR-22-00558

ABSTRACT

Purpose: Background noise reduces speech intelligibility. Time–frequency (T-F) masking is an established signal processing technique that improves intelligibility of neurotypical speech in background noise. Here, we investigated a novel application of T-F masking, assessing its potential to improve intelligibility of neurologically degraded speech in background noise.
Method: Listener participants (N = 422) completed an intelligibility task either in the laboratory or online, listening to and transcribing audio recordings of neurotypical (control) and neurologically degraded (dysarthria) speech under three different processing types: speech in quiet (quiet), speech mixed with cafeteria noise (noise), and speech mixed with cafeteria noise and then subsequently processed by an ideal quantized mask (IQM) to remove the noise.
Results: We observed significant reductions in intelligibility of dysarthric speech, even at highly favorable signal-to-noise ratios (+11 to +23 dB) that did not impact neurotypical speech. We also observed significant intelligibility improvements from speech in noise to IQM-processed speech for both control and dysarthric speech across a wide range of noise levels. Furthermore, the overall benefit of IQM processing for dysarthric speech was comparable with that of the control speech in background noise, as was the intelligibility data collected in the laboratory versus online.
Conclusions: This study demonstrates proof of concept, validating the application of T-F masks to a neurologically degraded speech signal. Given that intelligibility challenges greatly impact communication, and thus the lives of people with dysarthria and their communication partners, the development of clinical tools to enhance intelligibility in this clinical population is critical.

Given the substantial burden that background noise can have on the ability of listeners to understand spoken language (American National Standards Institute [ANSI], 1997; French & Steinberg, 1947), the development of effective signal processing techniques for hearing aids and other hearing technologies to mitigate these effects is of critical clinical import. Within the laboratory setting, time–frequency (T-F) masks, such as the Ideal Binary Mask (IBM), have been shown to be highly effective methods of reducing the negative impacts of background noise on speech intelligibility (Healy & Vasko, 2018; Srinivasan et al., 2006; Wang, 2005). In the most simplistic terms, these masks work by dividing a speech-noise mixture into a series of T-F units, discarding or attenuating portions of the signal that have a relatively unfavorable signal-to-noise ratio (SNR) and retaining those that are dominated by speech. Such masks have been shown to provide significant and clinically meaningful intelligibility benefits for both listeners with normal hearing and hearing impairment (e.g., Anzalone et al., 2006; Brungart et al., 2006; Healy & Vasko, 2018). Although the IBM is ideal, meaning it requires a priori knowledge of the

Correspondence to Stephanie A. Borrie: stephanie.borrie@usu.edu. Stephanie A. Borrie and Sarah E. Yoho share first author status. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

premixed speech and noise signals, algorithmic estimations of the IBM and other T-F masks have also demonstrated high efficacy in improving the intelligibility of speech in noisy conditions (e.g., Chen et al., 2016; Healy et al., 2013, 2015; Monaghan et al., 2017; Zhao et al., 2018). These types of noise reduction algorithms based on T-F masking show strong promise for eventual implementation into real-time signal processing devices such as hearing aids.

To date, T-F masks have been applied to neurotypical, intact speech that is highly or even perfectly intelligible in quiet conditions (e.g., Kjems et al., 2009; Li & Loizou 2008). However, speech is rarely perfectly intact and, as outlined below, can be degraded by neurological injury or disease. When either the listener *or* talker has a communication impairment, such as hearing loss or a speech disorder, the influence of background noise on intelligibility can be amplified (e.g., Baer et al., 1993; Yoho & Borrie, 2018). Although T-F masks were originally designed to address the burdens specifically associated with hearing impairment, the implementation of these masks for overcoming speech-in-noise difficulties with disordered speech, and specifically neurologically degraded speech signals, is an important and clinically relevant application.

Dysarthria is a motor speech disorder arising from neurological origins such as traumatic brain injury, Parkinson's disease, or stroke. The neurological speech disorder manifests in segmental and suprasegmental degradations that compromise the integrity of the acoustic signal, making it difficult for a listener to understand (Liss, 2007). These intelligibility challenges have been described as the most clinically and socially important aspects of dysarthria (Ansel & Kent, 1992) and have been causally linked with reduced participation in everyday life situations that involve communicating with others (Borrie et al., 2022). Reduced communicative participation can have determinantal consequences on individual well-being, including social isolation, loss of employment, and challenges in accessing services (e.g., Eadie et al., 2006; Walshe & Miller, 2011). As such, interventions and aids that preserve or enhance intelligibility of speakers with dysarthria are critical for not only improved communication outcomes but also quality of life.

While anecdotal reports from people with dysarthria and their communication partners indicate that communicating in noisy environments is exceedingly difficult and negatively impacts many aspects of life, research on listener perception of speakers with dysarthria has predominately focused on the neurologically degraded speech in quiet (i.e., no noise) conditions. Recently, however, a large-scale, systematic evaluation of the combined effects of environmental (i.e., background noise) and source (i.e., dysarthria) degradation empirically demonstrated that the addition of noise further decreases a listener's ability to understand a speaker with dysarthria (Yoho & Borrie, 2018), supporting the small collection of prior studies in this area (Adams et al., 2008; Dykstra et al., 2012; Lee et al., 2011). Additional support for the negative impacts of background noise on the intelligibility of dysarthric speech comes from studies that add noise to speech stimuli as a methodology decision to reduce ceiling effects for perception of mild forms of dysarthria (e.g., Fletcher et al., 2019; Tjaden et al., 2014). Taken together, there is satisfactory evidence that background noise, which is frequently present in real-world communication environments, markedly worsens the already challenging task of deciphering dysarthric speech.

Clearly, background noise has real and substantial impacts on understanding dysarthric speech, but there are currently no effective intervention strategies to overcome this nontrivial concern. Although formal data are lacking, anecdotal reports combined with the vast prevalence of hearing loss (approximately one third of the U.S. population aged 65 to 74 years and nearly half of those older than 75; National Institute on Deafness and Other Communication Disorders, 2022) suggest that the co-occurrence of dysarthria and hearing loss within communication partners, such as spouses, is high. Furthermore, the burden of combined hearing and speech disorders can be profound. Although, the impact of background noise on the intelligibility of dysarthric speech for even listeners with normal hearing is substantial. Therefore, the application of noise reduction via T-F masks to address this issue more broadly (i.e., for normal-hearing communication partners of speakers with dysarthria) may be beneficial.

There are a few different approaches to T-F masking, including the IBM (Hu & Wang, 2001; Wang, 2005) and the ideal ratio mask (IRM; Hummersone et al., 2014; Narayanan & Wang, 2013; Srinivasan et al., 2006; Wang et al., 2014). In the former, each T-F unit is assigned a value of either 0 or 1, depending on whether the unit is dominated by speech or noise, respectively, based on a designated SNR criterion. Units assigned 0 are discarded, whereas units assigned 1 are passed to the listener. In the IRM, each T-F unit is again attenuated based on speech or noise dominance, but instead of simply being discarded or retained, the assigned attenuation values fall along a continuum based on the degree of speech or noise dominance. Again, these techniques are considered "ideal" in that they involve a priori knowledge of the unmixed speech and noise signals and, therefore, allow complete control over the manner and extent to which T-F units are attenuated. For real-world, real-time applications in wearable devices, algorithmic estimation of these masks (typically achieved through machine learning) is required. There are advantages of each T-F mask type, including

the nature and computational resource of the algorithmic task required to estimate the masks, as well as the degree of speech intelligibility and subjective sound quality associated with each.

While both the IBM and the IRM have been shown to produce significant and substantial intelligibility gains relative to speech in noise prior to processing with the mask, the IRM is generally associated with superior sound quality. However, although the algorithmic task associated with the IBM is one of classifications (T-F units are simply classified into one of two bins based on a set criterion), the task associated with the IRM is one of regressions (the function relating attenuation to relative speech/noise dominance must be established) and approaches specific for classification exist (e.g., Hinton et al., 2015). Most recently, an approach termed the ideal quantized mask (IQM; Healy & Vasko, 2018) was developed to retain the best characteristics of both approaches. In this approach, each T-F unit is attenuated according to relative speech/noise dominance (such as the IRM), but this attenuation occurs as classification into a small number of steps (such as the IBM). It was recently shown that an IQM having only eight steps produced intelligibility and sound quality that exceeded that of the IBM and matched that of the IRM, despite the latter's infinite number of attenuation values (Healy & Vasko, 2018).

This Study

The primary purpose of this study was to assess the ability of T-F mask–based noise reduction (the IQM) to improve intelligibility of neurologically degraded speech in background noise. The secondary purpose was to compare the effects of background noise on the intelligibility of neurologically degraded versus neurotypical speech across a wide range of SNRs. Although the secondary aim involves basic information important for our understanding of communication involving neurologically degraded speech, the rationale for the primary aim is as follows: It is known that dysarthric speech is heavily impacted by background noise, and that noisy dysarthric speech produces particularly poor intelligibility (Yoho & Borrie, 2018). Furthermore, these same authors also found that the impact of dysarthria and that of noise are simply additive, meaning that the reductions in intelligibility resulting from noise are similar for both dysarthric and neurotypical speech, and that the low intelligibility of noisy dysarthric speech simply results from a lower baseline level of intelligibility in quiet. Accordingly, it may be hypothesized that effective noise reduction should have a similar effect on both types of speech (return it to noise-free levels of intelligibility). However, T-F masking–based noise reduction results in an acoustic speech signal that is sparser than that of the original speech, due to the attenuation of noisy T-F units. Neurotypical speech possesses sufficient redundancy and robustness to be understood perfectly when represented as a sparser-than-normal set of acoustic T-F units. What remains unknown is the extent to which dysarthric speech possesses this attribute and therefore the extent to which it can benefit from T-F mask–based noise reduction. In this study, the IQM was selected as a highly effective form of T-F masking, and an actual environmental recording containing multiple sound sources was selected as the background noise source. The magnitude of intelligibility benefit resulting from IQM processing for dysarthric and neurotypical (control) speech was assessed under conditions of *equal noise* (i.e., same SNR applied to both speech types) and under conditions of *equal performance* for speech in noise (i.e., different SNRs applied to dysarthric and control speech to equate performance in control conditions). Although the first condition allows comparison across speech types under equal acoustic conditions, the second condition allows comparisons involving similar performance baselines.

The execution of this study was in part made possible by a pandemic-motivated shift to online, large-scale data collection. Although the data for key conditions were collected in the laboratory using traditional methods, online data collection was used to comprehensively map the impact of background noise on intelligibility across a wide range of SNRs and the ability of IQM processing to restore intelligibility across these SNRs. This assessment was performed for both dysarthric and neurotypical speech. In addition to allowing a complete view of the impact of noise on both speech types, this technique allowed a complete view of the ability of T-F masking to restore intelligibility for neurotypical and neurologically degraded speech signals.

Last, reliability of laboratory versus online data collection was compared. Previous work on this data collection method comparison (e.g., Cooke et al., 2011; Lansford et al., 2016) and reliable data from prior studies examining intelligibility of dysarthric speech utilizing online data collection methods (e.g., Borrie et al., 2017; Yoho & Borrie, 2018) have demonstrated robust and congruent results. Novel here is the comparison of intelligibility performance across varying levels and combinations of speech degradation and types of speech signal processing.

Method

Participants

Participants were initially recruited from the student population of Utah State University (USU) and the

surrounding community of Logan, Utah. These 76 individuals formed the in-lab participant groups. As a result of the COVID-19 pandemic, data collection was shifted online, and additional participants were recruited via the crowd sourcing website Amazon Mechanical Turk (MTurk; http://www.mturk.com). These 346 individuals formed the online participant groups. In total, 422 adults, aged 18 to 71 years (M = 35.1, SD = 12.2), participated in the study. Table 1 shows the demographic and background characteristics of the participants, classified by data-collection method (in-lab vs. online). All participants were native speakers of American English living in the United States and reported no history of speech, language, or hearing impairment. All in-lab participants had pure-tone audiometric thresholds at or below 20 dB hearing level at octave frequencies from 250 to 8000 Hz (ANSI, 2004, 2010). To be included in the study, participants had to demonstrate task understanding and engagement, operationally defined as obtaining at least 80% of words correct for transcribing neurotypical (control) speech in quiet conditions. Participants received course credit or a monetary incentive. Data collection was approved by the USU Institutional Review Board.

Speech Stimuli

The stimuli consisted of 144 syntactically plausible but semantically anomalous phrases (e.g., *amend estate approach*; *had eaten junk and train*). Phrases were all six syllables in length and ranged from three to five words. These phrases, which restrict the listener's use of higher level cognitive-linguistic information to resolve the speech signal, were created specifically for examining speech perception in adverse listening conditions (Liss et al., 1998) and have been used extensively in the study of the perception of dysarthric speech (e.g., Borrie et al., 2012, 2017). The 144 phrases were divided into two 72-phrase sets and balanced for the total number of words, stress patterns of phrases, and speech segmentation error opportunities.

Speech Type

Two 72-year-old male native speakers of American English, one with dysarthria and one control, produced the stimuli for the study (one 72-phrase speech set each). The speaker with dysarthria presented with mild–moderate ataxic dysarthria secondary to cerebellar disease. His speech was characterized perceptually by excess and equal stress (scanning speech), prolonged phonemes and intervals, monotone, monoloudness, and imprecise articulation. The diagnosis was made by three independent speech-language pathologists (SLPs) with expertise in diagnosing motor speech disorders. The age and gender-matched control speaker presented with no neurological history or diagnosed speech production disorder, as confirmed by the same three SLPs above.

Processing Type

For each of the two speech types (dysarthria and control), there were three different processing conditions: speech in quiet (quiet), speech mixed with cafeteria noise (noise), and speech mixed with cafeteria noise and then processed by the IQM (IQM-processed). Prior to processing the speech, each separate speech utterance file was resampled to 16 kHz (16-bit precision) and trimmed to have 400 ms of silence at the beginning and end of each file. This speech represented the quiet condition.

Table 1. Demographic and background characteristics for in-lab versus online participants.

Variable	Overall N = 422	In-lab n = 76	Online n = 346	p value*
Mean age (*SD*)	35.1 (12.2)	20.3 (2.7)	38.3 (11.1)	< .001
Gender				< .001
Female	202 (47.9%)	55 (72.4%)	147 (42.5%)	
Genderqueer	2 (0.5%)	1 (1.3%)	1 (0.3%)	
Male	216 (51.2%)	19 (25%)	197 (56.9%)	
Nonbinary	2 (0.5%)	1 (1.3%)	1 (0.3%)	
Not Hispanic/Latinx	391 (92.7%)	72 (94.7%)	319 (92.2%)	.599
Race				.023
Asian/Pacific Islander	31 (7.3%)	1 (1.3%)	30 (8.7%)	
Black/African American	19 (4.5%)	0 (0%)	19 (5.5%)	
Caucasian/White	363 (86%)	74 (97.4%)	289 (83.5%)	
Native American	4 (0.9%)	1 (1.3%)	3 (0.9%)	
None of the above	5 (1.2%)	0 (0%)	5 (1.4%)	

*Chi-square for categorical variables and t test for continuous.

For the noise conditions, we utilized a noise file from an Auditec compact disc (http://www.auditec.com). This noise file (~10 min) was created using three overdubbed recordings from a busy hospital cafeteria and consists of various sound sources, including multiple voices, impact noises from dishes, and so forth. Thus, it represents a complex everyday sound and was selected for its ecological validity. To create the noise condition, each utterance file was mixed with a random segment of noise having equal duration and scaled to produce the desired SNR.

The creation of stimuli for the IQM condition followed the procedures of Healy and Vasko (2018). The speech utterance and the noise segment comprising each mixture was separated into T-F units using first a gammatone filterbank having 64 channels centered from 50 to 8000 Hz equally spaced on the equivalent rectangular bandwidth scale, then 20-ms Hanning windows with 10-ms shift. The IRM was then generated, defined as

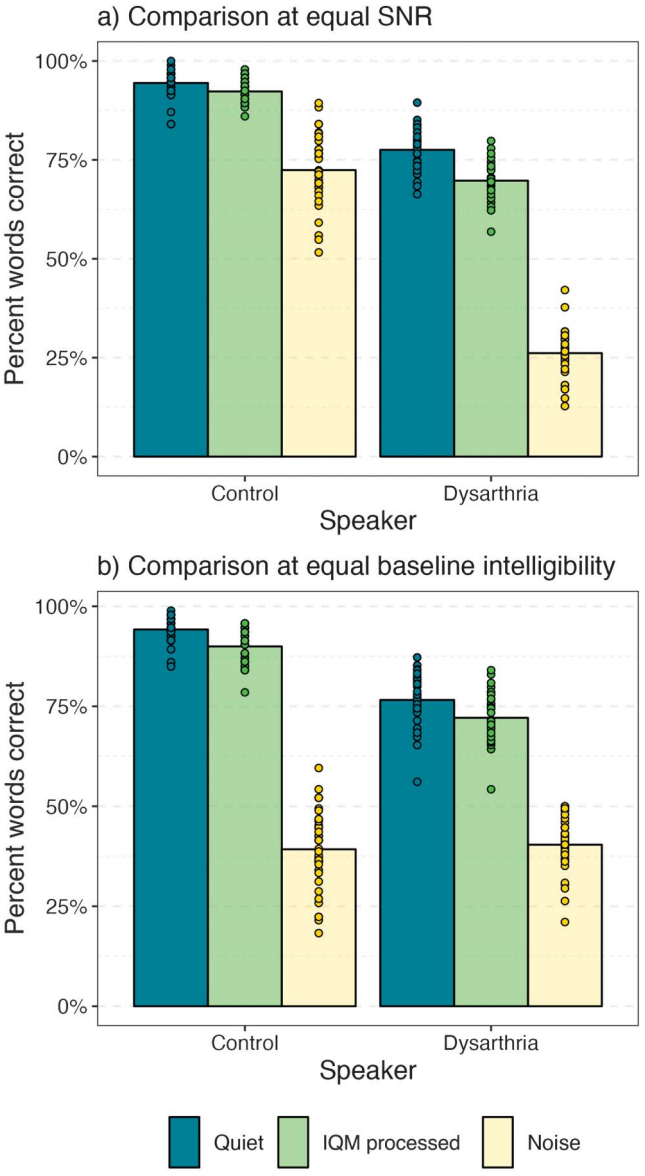
$$IRM(t, f) = \sqrt{\frac{S(t, f)}{S(t, f) + N(t, f)}} \quad (1)$$

where $S(t, f)$ and $N(t, f)$ denote speech and noise energy within a T-F unit at time t and frequency f, respectively. The IRM was transformed into the IQM having eight attenuation steps (IQM-8) by quantizing the IRM attenuation value for each T-F unit into the appropriate IQM attenuation step. The mapping of IRM to IQM attenuation, and the IQM attenuation values employed, may be found in Healy and Vasko (2018, see Figure 1). This mask, consisting of one of eight discrete attenuation values for each T-F unit, was then applied to the speech-plus-noise mixture to remove the background noise and result in IQM-processed speech. Following processing, each utterance file in each of the conditions was scaled to the same root-mean-square level. This processing was performed primarily in MATLAB.

Procedure

The preprocessed speech stimuli files were programmed in Gorilla (https://gorilla.sc/), an experiment builder for the collection of behavioral data. The basic function of the application was to present the participants with the speech stimuli and have them type out what they thought the person was saying. Prior to beginning all data collection, participants were asked to adjust the volume of a single utterance to a comfortable listening level over headphones. During testing, participants could only listen to each phrase once but could take as much time as necessary to type their responses. Based on pilot testing, SNRs resulting in equal performance (~40% correct) between the dysarthric and control speech-in-noise conditions were determined to be +5 dB and −3 dB, respectively. To address the question of intelligibility under conditions of equal performance, one group of in-lab participants (n = 39) received the speech stimuli with these differing SNRs between the speech types. A second group of in-lab participants (n = 38) heard the stimuli at the single fixed SNR of +1 dB for both dysarthric and control speakers. Each online participant was assigned a single SNR for both speech types, with SNRs ranging from −5 dB to 23 dB (−5, −3, −1, 1, 3, 5, 7, 9, 11, 17, 23 dB). These 11 SNRs

Figure 1. Comparisons of mean intelligibility performance (percent words correct) by speech and processing type for conditions of (a) equal signal-to-noise ratios (SNRs) and (b) equal performance for speech in noise across dysarthric and control speech. The points represent individual intelligibility scores for each individual participant. IQM = ideal quantized mask.



thus required 11 groups of participants. Each participant (both in-lab and online) heard all 144 phrases across the six conditions, blocked by type of speech (dysarthria or control) and processing type (quiet, noise, IQM-processed). An average of 32 participants (range: 29–39) were in each of the experimental groups. The presentation order of the blocks was randomized across all participants to eliminate potential order effects. On average, the speech perception task took the participants 30 min to complete.

Transcript Analysis

The participant transcripts were scored for correct words using autoscore, an open-source, computer-based tool for automated scoring of orthographic transcripts (http://autoscore.usu.edu/; Borrie et al., 2019). We applied the same scoring rules as previous studies on listener understanding of dysarthric speech (e.g., Borrie et al., 2022; Yoho & Borrie, 2018). Words were scored as correct if they matched the intended target exactly or differed only by tense or plurality. Homophones and obvious spelling errors were scored as correct using a preprogrammed default list of common misspellings. A percent words correct (PWC) score was tabulated for each participant, for each of the six conditions, by summing words correct from 24 speech phrases at each processing type and dividing by the total number of words. Thus, each participant had six intelligibility scores, one for each of the experimental conditions.

Statistical Analysis

To assess the magnitude of benefit from IQM processing on dysarthric speech in noise, we compared the benefit obtained from IQM processing for dysarthric and control speech under conditions of (a) *equal noise* (i.e., same SNR applied to dysarthric and control speech) and (b) *equal performance* for speech in noise (i.e., different SNRs applied to dysarthric and control speech). Prior to these statistical comparisons, the raw PWC scores were transformed using rationalized arcsine unit transformation to reduce potential issues of the distribution of proportion data (Studebaker, 1985). This transformation retains the other properties of the data while improving the distribution for comparisons. The comparisons were made using linear mixed-effects modeling, accounting for the repeated measures via random intercepts by individual. These models can be expressed as

$$PWC_{it} \sim N(\mu_{it}, \sigma) \tag{2}$$

$$\mu_{it} = \beta_0 + \beta_1 processing_{it} + \beta_2 speech_{it} + \beta_3 processing_{it} \times speech_{it} + \alpha_i \tag{3}$$

$$\alpha_i \sim N(\mu_i, \sigma_i) \tag{4}$$

where PWC_{it} is the transformed percent words correct for each individual and condition, the β_3 is the estimate of interest (the effect of processing type by speech type), and α_i is the random intercept by individual participant. Likelihood ratio tests were used to test for differences between conditions by speech type. Contrasts were assessed to estimate the magnitude of benefit from IQM processing within each speech type, with Bonferroni adjustment for the multiple comparisons. Reported pairwise contrasts (denoted "b" throughout) are in percentage units (e.g., $b = 20.0$ signifies a 20% point difference).

To assess the ability of IQM processing (i.e., noise removal) to restore intelligibility of dysarthric and control speech in noise to performance levels in quiet (i.e., no noise), as a function of the SNR, we used a psychometric ("logistic") curve. Specifically, we used the self-starting nonlinear least squares logistic model to estimate the effect of SNR on PWC by speaker type and processing level (both noise and IQM). The quiet condition was not modeled as that was not a function of SNR but was used as a reference for the other conditions. The model estimated three parameters: the asymptote (k), the inflection point (x), and the scale parameter (s), as shown by the following equation:

$$PWC = \frac{k}{1 + e^{\frac{x - SNR}{s}}} \tag{5}$$

The asymptote is the estimated highest level of PWC that can be achieved, the inflection point is the SNR value where PWC is higher than 50%, and the scale parameter is used for scaling the curve. The resulting estimates provide logistic curves that can be compared across conditions.

To assess reliability between in-lab and online data collection, we used independent-samples *t* tests to compare the laboratory and online results for each SNR, speech type, and stimulus pair wherein both laboratory and online results were available. This resulted in 12 comparisons. Notably, because all comparisons were planned a priori, we did not adjust for the multiple comparisons.

Assumptions of each statistical test were checked for problematic patterns. All analyses were performed in the R statistical environment (R Version 4.1.0; R Development Core Team, 2020). Data cleaning and visualization relied on the tidyverse packages (Wickham et al., 2019). Summary statistics were computed using the furniture and gtsummary packages (Barrett & Brignone, 2017; Sjoberg et al., 2021). The linear mixed-effects models relied on the lmer package (Bates et al., 2015) with likelihood ratio tests

using the built-in stats package. Statistical contrasts used the emmeans package (Lenth, 2022). The nonlinear least squares and *t* tests were performed using the built-in stats package. Analysis code and model output associated with this work are available at the study repository hosted at https://osf.io/z6dw5.

Results

Magnitude of Benefit

To quantify the magnitude of benefit from IQM-processing of dysarthric speech in noise, we assessed two configurations of conditions. First, we compared across conditions of equal SNR (+1 dB for the speech in noise and IQM-processed conditions for both speech types). Second, we compared across conditions of equal performance (~40% PWC with SNRs of +5 and −3 for the dysarthric and control, respectively, for both speech in noise and IQM-processed conditions).

The mean intelligibility scores for each condition are displayed in Figure 1. The upper panel displays the results for conditions of equal SNR, and the bottom panel displays the results for conditions of equal performance. For control and dysarthric speech in quiet, performance was 94% and 76%, respectively. Thus, these initial descriptive data confirm that dysarthria is associated with reduced intelligibility. In equal SNR conditions, intelligibility of control speech increased from 72% in the speech-in-noise condition to 92% in the IQM-processed condition, revealing a 20-percentage-point gain (and a 2-percentage-point difference between IQM-processed [92% correct] and speech in quiet [94% correct]). In contrast, in these equal SNR conditions, intelligibility of dysarthric speech increased from 26% in the speech-in-noise condition to 70% in the IQM-processed condition, revealing a 44-percentage-point gain (and a 6-percentage-point difference between IQM-processed [70% correct] and speech in quiet [76% correct]). In equal-performance conditions, intelligibility of control speech increased from 39% in the speech-in-noise condition to 90% in the IQM-processed condition, revealing a 51-percentage point gain (and a 4-percentage-point difference between IQM-processed [90% correct] and speech in quiet [94% correct]). In contrast, in these equal-performance conditions, intelligibility of dysarthric speech increased from 40% in the speech-in-noise condition to 72% in the IQM-processed condition, revealing a 32-percentage-point gain (and a 4-percentage-point difference between IQM-processed [72% correct] and speech in quiet [76% correct]).

For conditions of equal SNR, there was a significant interaction between processing type and speech type ($p < .001$). Contrasts indicated that the IQM-processed and speech-in-noise conditions were significantly different for both speech types ($b = 25.2$, $p < .001$ for control; $b = 42.3$, $p < .001$ for dysarthria), with large effect sizes, particularly for dysarthric speech. Thus, IQM processing, in this comparison, was more beneficial for dysarthric relative to control speech ($p < .001$). There were differences between the IQM-processed and speech-in-quiet conditions for both speech types ($b = 4.6$, $p = .007$ for control and $b = 8.4$, $p < .001$ for dysarthria), indicating that intelligibility was not fully restored for either neurotypical or neurologically degraded speech.

For conditions of equal performance in the speech-in-noise conditions, likelihood ratio tests indicated a significant interaction between processing type and speech type ($p < .001$). Contrasts indicated that the IQM-processed and speech-in-noise conditions were significantly different for both speech types ($b = 54.0$, $p < .001$ for control; $b = 30.6$, $p < .001$ for dysarthria), with large effect sizes, in this case, more so for the control speech ($p < .001$). Here, there were differences between the IQM-processed and quiet conditions for both speech types ($b = 7.6$, $p < .001$ for control and $b = 4.9$, $p < .001$ for dysarthria), once again indicating that intelligibility was not fully restored for either neurotypical or neurologically degraded speech.

Psychometric Functions

To comprehensively examine the impact of IQM processing on the restoration of intelligibility of dysarthric speech-in-noise, we psychometrically mapped intelligibility performance as a function of SNR. Estimates from the nonlinear least squares logistic models with their accompanying standard errors are shown in Table 2. All estimates were significantly different from zero. These curves are shown in Figure 2 with a reference line for each speech type to the speech in quiet performance levels (94% intelligibility for control speech and 76% intelligibility for dysarthric speech). Each point on the plot is the average for that specific SNR, speech type, and processing type. Several patterns (based on the model results and/or the figure) are immediately clear. First, as expected, intelligibility of the control speech was greater regardless of processing type—the speech-in-quiet condition as shown in the figure and the estimated asymptotes are higher for control relative to dysarthric speech. Second, the scale parameter indicates that the steepest slope was for the speech-in-noise condition for the control speech, followed by the IQM-processed condition for the control speech, then the speech-in-noise condition for the dysarthric speech and, last, the IQM-processed condition for the dysarthric speech. Third, relatively favorable SNRs that do not affect intelligibility of control speech do affect the intelligibility

Table 2. Estimates from the nonlinear least squares regression models.

Stimulus	Speaker	Asymptote		Inflection point		Scale	
		Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
IQM	Control	0.935	0.005	−18.6	6.180	4.40	1.950
IQM	Dysarthria	0.756	0.015	−18.3	4.320	7.77	2.590
Noise	Control	0.919	0.008	−2.36	0.105	2.61	0.117
Noise	Dysarthria	0.712	0.013	3.43	0.258	4.65	0.227

Note. *SE* = standard error; IQM = ideal quantized mask.

of dysarthric speech. For instance, for SNRs of +11, +17, and +23 dB, there is no distinguishable impact of noise on the intelligibility of the control speech; however, for the dysarthric speech, listeners perform below levels achieved in speech in quiet. Thus, taken together, the impact of noise is different across the two types of speech. Fourth, the IQM-processed condition, regardless of speech type, had similar inflection points. These inflection points, −18.6 and −18.3, are further negative than the speech-in-noise conditions, with the speech-in-noise condition for dysarthric speech having the highest value (3.4). This indicates that the IQM-processed speech had higher intelligibility at less favorable SNRs than the speech-in-noise conditions for both speech types. Importantly, it also indicates that the impact of IQM processing is not differently affected by the speech type, despite the evidence that the intelligibility of speech in noise is differently affected by speech type.

In-Lab Versus Online

Last, we compared intelligibility performance levels (i.e., PWC values) between the laboratory and online samples for each SNR, speaker, and stimulus pair, wherein both in-lab and online data were available. Figure 3 shows these comparisons, highlighting the consistent match between laboratory and online samples for the same experimental conditions. Independent samples *t* tests confirmed this finding, showing no significant differences across all 12 comparisons (all *ps* > .10).

Discussion

Here, we assessed the impact of noise on neurologically degraded as well as neurotypical speech and whether noise reduction via T-F masking could improve intelligibility

Figure 2. Logistic curves as a function of signal-to-noise ratio (SNR) estimated for each condition and speech type. The quiet condition is represented by the dashed line at the top of each of the curves. Black boxes represent the averages from the lab, whereas the other symbols represent averages from the online participants. *SE* = standard error; IQM = ideal quantized mask.

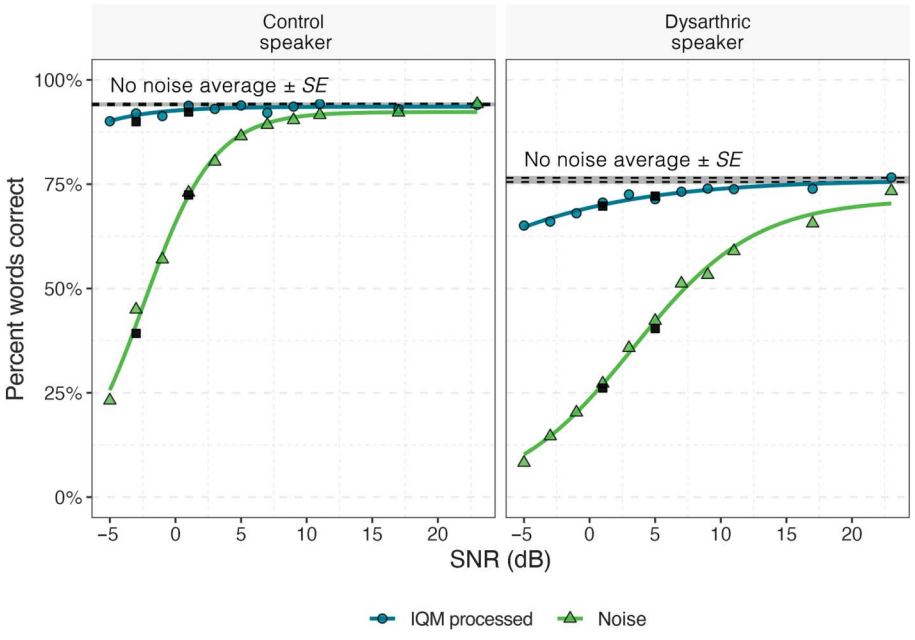
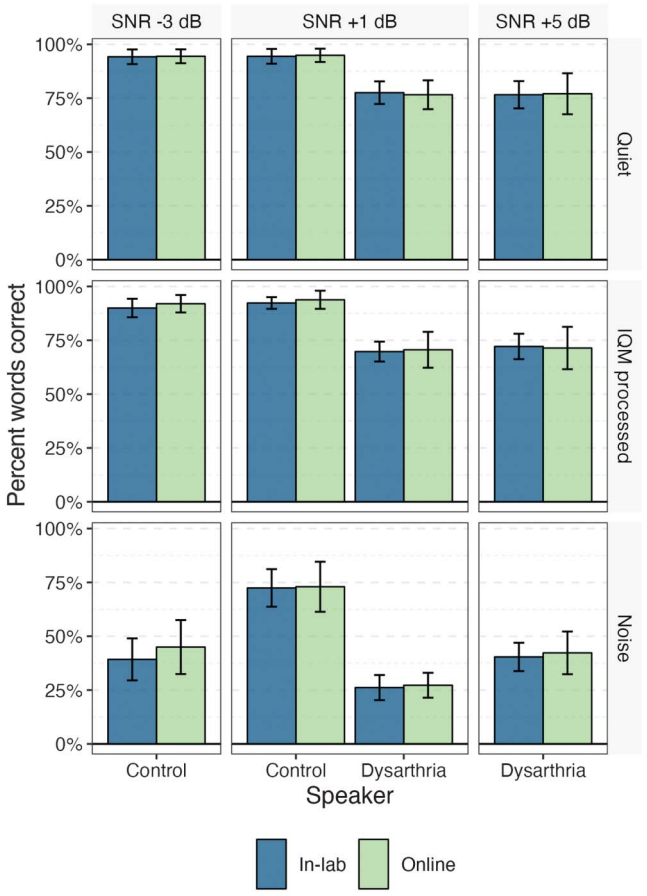


Figure 3. Comparisons of mean intelligibility performance between the lab and online samples across speech type, processing type, and signal-to-noise ratio (SNR) levels. Error bars are ± 1 *SE*. IQM = ideal quantized mask.



of these speech types in background noise. As described earlier, T-F masks work by attenuating portions of the noisy speech signal with relatively unfavorable SNRs. The IQM, an approach that applies one of several discreet attenuation steps to individual T-F units based on each unit's SNR, was selected given prior work demonstrating favorable intelligibility and sound quality relative to other masks (Healy & Vasko, 2018). Across all SNR conditions, we observed significant intelligibility improvements from speech in noise to IQM-processed speech, showing that IQM processing improved listener understanding of dysarthric speech in noise. Although a body of work had previously demonstrated the successful application of T-F masks, including the IQM, to listener understanding of neurotypical (i.e., control) speech in noise (e.g., Healy & Vasko, 2018; Kim et al., 2009; Sinex, 2013), to our knowledge, this is the first study to examine the application of T-F mask processing to impaired speech. Thus, this study demonstrates proof of concept for the application of T-F mask noise reduction to disordered, and specifically, dysarthric speech.

Overall, the data suggest that the benefit of IQM processing for dysarthric speech is comparable with that of neurotypical speech. In both speech types, intelligibility was not fully restored by IQM processing to levels of quiet at every SNR (see Figure 2); however, intelligibility was within 12 percentage points of quiet performance for even the least favorable SNR for both types of speech. At more favorable SNRs, IQM-processed intelligibility essentially matched scores in quiet for both speech types, indicating complete mitigation of the effects of noise. For conditions of equal performance (see Figure 1, bottom panel), intelligibility in the IQM-processed condition was within 4 percentage points of intelligibility in quiet for both control and dysarthric speech. For conditions of equal SNR (see Figure 1, top panel), intelligibility in the IQM-processed condition was within 2 and 6 percentage points of intelligibility in quiet for control and dysarthric speech, respectively. The current findings suggest that, as the noise levels become increasingly unfavorable, the nature of IQM removal process, in which additional small reductions in acoustic integrity result, impacts the ability of noise reduction to fully restore intelligibility of both neurotypical and neurologically degraded acoustic speech signals.

It was also observed that, for conditions of equal SNR, in which dysarthric speech was significantly less intelligible than control speech, the IQM produced greater intelligibility improvement for dysarthric speech relative to control speech. This finding is likely attributable to the lower baseline intelligibility in noise for the dysarthric speech and the resulting greater room to improve. This indicates that, within a particular noise environment, T-F mask noise reduction may be more beneficial for degraded speech. Within conditions of equal performance across the speech types in noise, the magnitude of benefit was greater for control speech. This finding is likely attributable to the lower overall intelligibility of the dysarthric speech in quiet (a lower ceiling), meaning that there was less opportunity for IQM processing to improve performance relative to control speech. Because of these differences between intelligibility in noise (baseline) and in quiet (ceiling) across the two speech types, perhaps the more reflective measure of the impact of IQM noise reduction is the comparison between IQM-processed and quiet speech.

Using the psychometric model-based predictions and online procedures, the ability of T-F masks to improve the intelligibility of dysarthric speech in noise is not only robust but also further supports the initial findings that, overall, the benefit from T-F masking is comparable across speech type. That is, while the current data indicate that background noise has a greater negative impact on understanding dysarthric speech, the ability to restore speech intelligibility to levels of speech in quiet is, in general, similar for dysarthric and control speech. Thus, the

presence of degraded segmental and suprasegmental acoustic information in neurologically degraded speech, which significantly drives down intelligibility relative to control speech, did not largely impact the effectiveness of IQM processing. Collectively, the study results allow us to conclude that IQM processing of speech in noise offers comparable and considerable intelligibility benefits for neurotypical and neurologically degraded speech. This finding is promising, highlighting the effectiveness of the application of T-F mask noise reduction to improve normal-hearing communication partners' ability to understand a speaker with dysarthria in background noise.

There has been limited investigation of the overall impact of background noise on dysarthric speech (Adams et al., 2008; Dykstra et al., 2012; Lee et al., 2011). In a recent study in this area, Yoho and Borrie (2018) found that, for relatively favorable SNR levels, increasing levels of speech-shaped noise decreased intelligibility at an equivalent rate for control and dysarthric speech. In this study, the psychometric function for control speech in noise was steeper than the function for dysarthric speech in noise, indicating that the intelligibility of neurotypical speech improved more rapidly as a function of SNR. This difference in findings across the two studies may be attributable to the noise types utilized in each (speech-shaped noise vs. cafeteria noise) and the range of SNRs evaluated. Continued investigation is needed to comprehensively characterize the impacts of different types of background noise on understanding of speakers with dysarthria, as well as any interactions between type of noise and dysarthria type or severity.

Importantly, we observed that that even minimal amounts of background noise, that is, levels that did not impact intelligibility of neurotypical speech, had a marked, negative impact on intelligibility of dysarthric speech. As illustrated in Figure 2, +11 dB, +17 dB, and +23 dB, reduced intelligibility of dysarthric speech by 16, 12, and 3 percentage points, respectively, whereas control speech was within 1-percentage point of performance in quiet for each of these SNRs. Thus, in environments with negligible background noise, levels that many listeners may not even consciously consider "noisy," the negative impact on listener understanding of dysarthric speech is substantial. This demonstration of the relative "fragility" of the dysarthric speech signal at very low levels of noise is supported by the earlier finding that removal of even small portions of the signal via IQM processing can have a negative impact in unfavorable SNRs (i.e., incomplete restoration). These findings provide empirical support for anecdotal reports that communicating in noisy environments is exceedingly difficult for people with dysarthria and their communication partners, even for listeners with normal hearing, and underscores the high importance of addressing the difficulties inherent in understanding disordered speech in background noise.

Finally, we also showed highly comparable intelligibility performance across two distinctly different data collection methods, in-lab versus online, and these comparable performances held across speech type, processing type, and noise levels. This was the case even though the two samples represented somewhat different populations, as indicated by demographic information. As noted earlier, others have found comparable intelligibility data for listeners transcribing dysarthric speech in-lab versus online methods (Lansford et al., 2016). Novel here is the comparison of intelligibility performance across varying levels and combinations of speech degradation and processing, with average intelligibility levels ranging from approximately 25% to 95%. That performance discrepancies did not emerge in more challenging listening conditions (e.g., lower SNRs) affords additional validation for the utility of online crowdsourcing as a feasible data collection method for studies examining listener understanding of degraded speech. Of note, as stated in the method, we excluded workers who showed evidence of poor task engagement, operationally defined as achieving less than 80% of words correct for transcribing control (i.e., neurotypical) speech in quiet. This resulted in excluding 131 potential participants from the analyses, leaving 435 participants to be included in the study. This highlights how intelligibility data collected through online crowdsourcing can present challenges; however, with some simple exclusion criteria based on unbiased markers of unsatisfactory task engagement, data are highly valid (see also Ziegler et al., 2021, for additional discussion and application of online crowdsourcing for valid and accurate intelligibility data in the assessment of dysarthria).

We note that the current comparison between laboratory and online data collection did not hinge critically on the scientific inquiry of this study. However, the benefits of online data collection are many. By crowdsourcing the experiment via MTurk, we were able to continue this work throughout a global pandemic that halted in-person data collection at the authors' universities for approximately 2 years. Furthermore, this collection allowed for a large, diverse sample of participants that more closely represents the general population (see Table 1) while still controlling for important variables such as country of residence, native language, and previous speech, language, and hearing history. With large numbers of participants available, the study could be extended from the original design shown in Figure 1 to include the collection of data for the complete psychometric functions displayed in Figure 2. Criticisms of online data collection methods have included lack of control over stimulus presentation levels and the testing environment. However, compelling comparable results not only in this study but also others, including those that examine listener processing beyond dysarthric

speech (e.g., Cooke et al., 2011; McAllister Byun et al., 2015; Slote & Strand, 2016), refute such criticism, collectively advancing online data collection methods as sufficiently rigorous for speech perception studies.

Limitations, Clinical Implications, and Future Directions

Although a subset of acoustic features (e.g., reduced vowel space and reduced speaking rate) have been identified as unifying the dysarthrias as a group of neurogenic speech disorders (see Weismer & Kim, 2010), it is important to acknowledge that there are different types of dysarthria, with different pathophysiology and constellations of deviant speech features. Here, we used stimuli from a speaker whose speech features represented the cardinal features of an ataxic dysarthria. Although this methodological design ensured suitable experimental control and allowed for the systematic evaluation of the IQM across many SNRs, it may limit the generalization of some of the results, key being the specific magnitude of benefit. However, despite acoustic differences across the dysarthrias, the broader findings of this work—that background noise substantially reduced intelligibility of a neurologically degraded speech signal and that a T-F masking–based noise reduction technique effectively improved intelligibility of the degraded signal—suggests effective application for other types of neurologically degraded, or otherwise impaired, speech.

Previously, we found that listeners with normal hearing struggle considerably to understand a speaker with dysarthria in the presence of background noise (Yoho & Borrie, 2018). Here, we extended those findings to a different type of noise (white noise vs. cafeteria noise) and a broader range of SNRs. Taken together, these data demonstrate clear and important implications for clinical practice—that people with dysarthria and their communication partners should be counseled on the importance of optimizing the communication environment, and that further study is needed to identify interventions to overcome the speech-in-noise challenge for this population. The current results indicate that T-F masks may be one promising solution to ameliorate these issues, even for listeners with normal hearing. For example, a communication partner of a person with dysarthria could potentially utilize a device similar to a hearing aid, one having effective noise reduction but without amplification, to reduce the deleterious effects of background noise on their partner's degraded speech. Future work in this area could include evaluation of the benefits of T-F processing across a range of dysarthria types and identify which acoustic cues are effectively transmitted via the T-F masks when the speech signal is degraded.

In addition, it is likely that individuals with impaired hearing could benefit greatly from noise-reduction strategies (e.g., T-F masks) that improve the reception of dysarthric speech. As stated in the Introduction section, hearing loss is highly prevalent. Prior studies have shown that understanding speakers with dysarthria is negatively impacted by hearing loss and advanced age (Lansford et al., 2018; McAuliffe et al., 2017). Given the common etiologies of dysarthria (i.e., stroke and degenerative disease), the speech disorder is also associated with advanced age. Although statistics on the co-occurrence of speech and hearing impairments in communication dyads (i.e., person with dysarthria and partner with hearing loss) are lacking, anecdotal evidence from clinicians who treat patients with dysarthria report the co-occurrence with alarming frequency. Given that the benefit of T-F mask noise reduction in understanding healthy speech in noise has been previously established in individuals with hearing loss (e.g., Anzalone et al., 2006, Healy et al., 2013) and, in this study, understanding dysarthric speech in noise for individuals with normal hearing, a future investigation into the benefit of T-F masks in understanding speakers with dysarthria in background noise for listeners with hearing loss is well justified.

Although this study examined ideal masking, that is, separation of the speech and noise signals based on a priori knowledge of each, the ultimate clinical goal is to develop algorithms capable of estimating such masks in real time. The current data demonstrate proof of concept that T-F masking has the capability to restore the intelligibility of dysarthric speech in noise to levels in quiet. Further work is needed to evaluate the ability of algorithmic estimation of T-F masks to improve the intelligibility of disordered speech, particularly when the speech is characterized by highly unpredictable acoustic degradations, such as the case of hyperkinetic dysarthria.

This challenge is nontrivial, given that the deep learning techniques typically used to estimate the T-F mask rely on a large number of training utterances to train the artificial neural network to identify clean speech in the noisy mixture. Such large data sets of neurologically degraded speech may not be readily available. However, the ability of modern neural networks to generalize to conditions different from those encountered during training is robust, as demonstrated in several recent works (e.g., Healy et al., 2020; Healy, Johnson, et al., 2021; Healy, Taherian, et al., 2021). This generalization has included the ability of a network trained using English-language speech materials to operate on Mandarin-language speech without hindrance (Healy, Tan, et al., 2021). Although the extrapolation from neurotypical speech to dysarthric speech is different from that across different languages, these demonstrations of vast generalization suggest that neural networks trained using neurotypical speech may prove effective for removing background noise from neurologically degraded speech, without the

need to train directly on neurologically degraded speech. Because these networks are also able to generalize to speech from talkers not in the training set (untrained talkers), training using the speech from any particular talker is not needed. Other questions remain, including the target that the network aims toward—is noise reduction more effective when the network is trained to output noise-free neurotypical speech or noise-free neurologically degraded speech? Although questions remain involving the best way to implement T-F mask or machine learning-based noise reduction for neurologically degraded speech, the current results offer promise that effective real-world noise-reduction systems can be developed to support the communication needs of people with dysarthria and their communication partners.

Conclusions

Here, we found that the presence of background noise significantly reduced intelligibility of a speaker with dysarthria, even at highly favorable SNRs that produce little to no reduction in the intelligibility of neurotypical speech. However, the application of T-F masks (the IQM) significantly increased intelligibility of the dysarthric speaker across a wide range of noise levels. In fact, at several SNRs, intelligibility of this speaker was restored to performance levels in quiet. Furthermore, the overall benefit of IQM processing for dysarthric speech in noise was comparable to that of the neurotypical control speech in noise. Given the substantial, negative impact of background noise on understanding speakers with dysarthria, future development of clinical tools to address these difficulties, which impact both people with dysarthria and their communication partners, could significantly improve communication outcomes and quality of life for these speaker-listener dyads. The current results are promising with regard to the application of T-F masks as a noise-reduction technique for neurologically degraded speech in background noise.

Author Contributions

Stephanie Borrie: Conceptualization (Lead), Project administration (Lead), Writing – original draft (Lead), Writing – review & editing (Lead), Methodology (Lead), Formal analysis (Supporting), Visualization (Supporting), Software (Supporting). **Sarah Yoho:** Conceptualization (Lead), Project administration (Supporting), Writing – original draft (Lead), Writing – review & editing (Lead), Methodology (Lead), Formal analysis (Supporting); Visualization (Supporting). **Eric Healy:** Conceptualization (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting), Methodology (Supporting), Formal analysis (Supporting), Visualization (Supporting). **Tyson Barrett:** Conceptualization (Supporting), Project administration (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting), Methodology (Supporting), Formal analysis (Lead), Software (Lead), Visualization (Lead).

Data Availability Statement

Anonymized listener data, analysis code, and model outputs associated with this work are available at the study repository hosted at https://osf.io/z6dw5.

Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders Grants (R21DC018641 to Sarah E. Yoho, R21DC018867 to Stephanie A. Borrie, and R01DC015521 to Eric W. Healy). Eric M. Johnson provided simple and elegant modification to IRM code provided by DeLiang Wang to perform IQM processing—we are grateful for their assistance. The authors also gratefully acknowledge research assistants in the Human Interaction Lab at Utah State University for assistance with data collection.

References

Adams, S. G., Dykstra, A., Jenkins, M., & Jog, M. (2008). Speech-to-noise levels and conversational intelligibility in hypophonia and Parkinson's disease. *Journal of Medical Speech-Language Pathology, 16*(4), 165–172.

American National Standards Institute. (1997). *Methods for calculation of the Speech Intelligibility Index (ANSI/ASA S3.5-1997)*. Acoustical Society of America.

American National Standards Institute. (2004). *Methods for manual pure-tone threshold audiometry* (ANSI S3.21-2004 (R2009)).

American National Standards Institute. (2010). *Specification for audiometers* (ANSI S3.6-2010).

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*(2), 296–308. https://doi.org/10.1044/jshr.3502.296

Anzalone, M. C., Calandruccio, L., Doherty, K. A., & Carney, L. H. (2006). Determination of the potential benefit of time-frequency gain manipulation. *Ear and Hearing, 27*(5), 480–492. https://doi.org/10.1097/01.aud.0000233891.86809.df

Baer, T., Moore, B. C., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development, 30,* 49–72.

Barrett, T. S., & Brignone, E. (2017). Furniture for quantitative scientists. *The R Journal, 9*(2), 142–148. https://doi.org/10.32614/RJ-2017-037

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Borrie, S. A., Baese-Berk, M., Van Engen, K., & Bent, T. (2017). A relationship between processing speech in noise and dysarthric speech. *The Journal of the Acoustical Society of America, 141*(6), 4660–4667. https://doi.org/10.1121/1.4986746

Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of Acoustical Society of America, 145*(1), 392–399. https://doi.org/10.1121/1.5087276

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27*(7–8), 1039–1055. https://doi.org/10.1080/01690965.2011.610596

Borrie, S. A., Wynn, C. J., Berisha, V., & Barrett, T. S. (2022). From speech acoustics to communicative participation in dysarthria: Toward a causal framework. *Journal of Speech, Language, and Hearing Research, 65*(2), 405–418. https://doi.org/10.1044/2021_jslhr-21-00306

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. L. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America, 120*(6), 4007–4018. https://doi.org/10.1121/1.2363929

Chen, J., Wang, Y., Yoho, S. E., Wang, D. L., & Healy, E. W. (2016). Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises. *The Journal of the Acoustical Society of America, 139*(5), 2604–2612. https://doi.org/10.1121/1.4948445

Cooke, M., Barker, J., Lecumberri, M. L. G., & Wasilewski, K. (2011). *Crowdsourcing for word recognition in noise*. In Proceedings of the Twelfth Annual Conference of the International Speech Communication Association, August 27–31, Florence, Italy (pp. 3049–3052).

Dykstra, A. D., Adams, S. G., & Jog, M. (2012). The effect of background noise on the speech intensity of individuals with hypophonia associated with Parkinson's disease. *Journal of Medical Speech-Language Pathology, 20*(3), 19–31.

Eadie, T. L., Yorkston, K. M., Klasner, E. R., Dudgeon, B. J., Deitz, J. C., Baylor, C. R., Miller, R. M., & Amtmann, D. (2006). Measuring communicative participation: A review of self-report instruments in speech-language pathology. *American Journal of Speech-Language Pathology, 15*(4), 307–320. https://doi.org/10.1044/1058-0360(2006/030)

Fletcher, A., McAuliffe, M., Kerr, S., & Sinex, D. (2019). Effects of vocabulary and implicit linguistic knowledge on speech recognition in adverse listening conditions. *American Journal of Audiology, 28*(3S), 742–755. https://doi.org/10.1044/2019_aja-heal18-18-0169

French, N. R., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America, 19*(1), 90–119. https://doi.org/10.1121/1.1916407

Healy, E. W., Johnson, E. M., Delfarah, M., Sevich, V. A., Krishnagiri, D. S., & Wang, D. L. (2021). Deep learning based speaker separation and dereverberation can generalize across different languages to improve intelligibility. *The Journal of the Acoustical Society of America, 150*(4), 2526–2538. https://doi.org/10.1121/10.0006565

Healy, E. W., Johnson, E. M., Delfarah, M., & Wang, D. L. (2020). A talker-independent deep learning algorithm to increase intelligibility for hearing-impaired listeners in reverberant competing talker conditions. *The Journal of the Acoustical Society of America, 147*(6), 4106–4118. https://doi.org/10.1121/10.0001441

Healy, E. W., Taherian, H., Johnson, E. M., & Wang, D. L. (2021). A causal and talker-independent speaker-separation/dereverberation deep learning algorithm: Cost associated with conversion to real-time capable operation. *The Journal of the Acoustical Society of America, 150*(5), 3976–3986. https://doi.org/10.1121/10.0007134

Healy, E. W., Tan, K., Johnson, E. M., & Wang, D. L. (2021). An effectively causal deep learning algorithm to increase intelligibility in untrained noises for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 149*(6), 3943–3953. https://doi.org/10.1121/10.0005089

Healy, E. W., & Vasko, J. L. (2018). An ideal quantized mask to increase intelligibility and quality of speech in noise. *The Journal of the Acoustical Society of America, 144*(3), 1392–1405. https://doi.org/10.1121/1.5053115

Healy, E. W., Yoho, S. E., Chen, J., Wang, Y., & Wang, D. (2015). An algorithm to increase speech intelligibility for hearing-impaired listeners in novel segments of the same noise type. *The Journal of the Acoustical Society of America, 138*(3), 1660–1669. https://doi.org/10.1121/1.4929493

Healy, E. W., Yoho, S. E., Wang, Y., & Wang, D. (2013). An algorithm to improve speech recognition in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 134*(4), 3029–3038. https://doi.org/10.1121/1.4820893

Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network*. arXiv. https://doi.org/10.48550/arXiv.1503.02531

Hu, G., & Wang, D. L. (2001). Speech segregation based on pitch tracking and amplitude modulation. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)* (pp. 79–82). https://doi.org/10.1109/aspaa.2001.969547

Hummersone, C., Stokes, T., & Brooks, T. (2014). On the ideal ratio mask as the goal of computational auditory scene analysis. In G. R. Naik & W. Wang (Eds.), *Blind source separation: Advances in theory, algorithms and applications* (pp. 349–368). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-55016-4_12

Kim, G., Lu, Y., Hu, Y., & Loizou, P. C. (2009). An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *The Journal of the Acoustical Society of America, 126*(3), 1486–1494. https://doi.org/10.1121/1.3184603

Kjems, U., Boldt, J. B., Pedersen, M. S., Lunner, T., & Wang, D. L. (2009). Role of mask pattern in intelligibility of ideal binary-masked noisy speech. *The Journal of the Acoustical Society of America, 126*(3), 1415–1426. https://doi.org/10.1121/1.3179673

Lansford, K. L., Borrie, S. A., & Bystricky, L. (2016). Use of crowdsourcing to assess the ecological validity of perceptual training paradigms in dysarthria. *American Journal of Speech-Language Pathology, 25*(2), 233–239. https://doi.org/10.1044/2015_ajslp-15-0059

Lansford, K. L., Luhrsen, S., Ingvalson, E., & Borrie, S. A. (2018). Effects of familiarization on intelligibility of dysarthric speech in older adults with and without hearing loss. *American Journal of Speech-Language Pathology, 27*(1), 91–98. https://doi.org/10.1044/2017_ajslp-17-0090

Lee, Y., Sim, H. S., & Sung, J. E. (2011). The intonation patterns of accentual phrase in Jeju dialect. *Phonetics and Speech*

Sciences, 6(4), 117–123. https://doi.org/10.13064/KSSS.2014.6.4.117

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.3*. https://CRAN.R-project.org/package=emmeans

Li, N., & Loizou, P. C. (2008). Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *The Journal of the Acoustical Society of America, 123*(3), 1673–1682. https://doi.org/10.1121/1.2832617

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

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457–2466. https://doi.org/10.1121/1.423753

McAllister Byun, T., Halpin, P. F., & Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders, 53,* 70–83. https://doi.org/10.1016/j.jcomdis.2014.11.003

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

Monaghan, J. J. M., Goehring, T., Yang, X., Bolner, F., Wang, S., Wright, M. C. M., & Bleeck, S. (2017). Auditory inspired machine learning techniques can improve speech intelligibility and quality for hearing impaired listeners. *The Journal of the Acoustical Society of America, 141*(3), 1985–1998. https://doi.org/10.1121/1.4977197

Narayanan, A., & Wang, D. L. (2013). Ideal ratio mask estimation using deep neural networks for robust speech recognition. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 7092–7096). https://doi.org/10.1109/icassp.2013.6639038

National Institute on Deafness and Other Communication Disorders. (2022, September 19). *Age-related hearing loss*. https://www.nidcd.nih.gov/health/age-related-hearing-loss

R Development Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sinex, D. G. (2013). Recognition of speech in noise after application of time-frequency masks: Dependence on frequency and threshold parameters. *The Journal of the Acoustical Society of America, 133*(4), 2390–2396. https://doi.org/10.1121/1.4792143

Sjoberg, D. D., Whiting, K., Curry, M., Lavery, J. A., & Larmarange, J. (2021). Reproducible summary tables with the gtsummary package. *The R Journal, 13*(1), 570–580. https://doi.org/10.32614/RJ-2021-053

Slote, J., & Strand, J. (2016). Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods, 48*(2), 553–566. https://doi.org/10.3758/s13428-015-0599-7

Srinivasan, S., Roman, N., & Wang, D. L. (2006). Binary and ratio time-frequency masks for robust speech recognition. *Speech Communication, 48*(11), 1486–1501. https://doi.org/10.1016/j.specom.2006.09.003

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28*(3), 455–462. https://doi.org/10.1044/jshr.2803.455

Tjaden, K., Sussman, J. E., & Wilding, G. E. (2014). Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

Walshe, M., & Miller, N. (2011). Living with acquired dysarthria: The speaker's perspective. *Disability and Rehabilitation, 33*(3), 195–203. https://doi.org/10.3109/09638288.2010.511685

Wang, D. L. (2005). On ideal binary mask as the computational goal of auditory scene analysis. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 181–197). Kluwer Academic. https://doi.org/10.1007/b99695

Wang, Y., Narayanan, A., & Wang, D. L. (2014). On training targets for supervised speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(12), 1849–1858. https://doi.org/10.1109/TASLP.2014.2352935

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhm, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *The Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Weismer, G., & Kim, Y. (2010). Classification and taxonomy of motor speech disorders: What are the issues? In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 229–241). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199235797.003.0013

Yoho, S. E., & Borrie, S. A. (2018). Combining degradations: The effect of background noise on intelligibility of disordered speech. *The Journal of Acoustical Society of America, 143*(1), 281–286. https://doi.org/10.1121/1.5021254

Zhao, Y., Wang, D. L., Johnson, E. M., & Healy, E. W. (2018). A deep learning based segregation algorithm to increase speech intelligibility for hearing-impaired listeners in reverberant-noisy conditions. *The Journal of the Acoustical Society of America, 144*(3), 1627–1637. https://doi.org/10.1121/1.5055562

Ziegler, W., Lehner, K., & KommPaS Study Group. (2021). Crowdsourcing as a tool in the clinical assessment of intelligibility in dysarthria: How to deal with excessive variation. *Journal of Communication Disorders, 93,* 106135. https://doi.org/10.1016/j.jcomdis.2021.106135