Evaluating Speech Intelligibility with Processed Sound

Sam J. Faulkner

Department of knowledge integration in partnership with the department of psychology

PSYCH 484: Independent Study

Supervisor: Dr. Katherine White

April 17, 2023

Abstract

This paper was created with the goal of researching the different impacts that background noise can have on listeners' ability to interpret speech. The brain is responsible for separating speech and noise, but this can be difficult if this organ is damaged or the noise is too overwhelming to separate out. I partnered with Augmented Hearing.io to see whether their noise reduction software can do some of this processing on behalf of the brain. This would reduce cognitive effort and help make conversations more accessible in noisy environments. To research this topic, I created a study that evaluated participants' ability to understand words that have often confused sounds in them. These words were presented with different types of voices, with different kinds of background noise, and both with and without processing from Augmented Hearing's algorithms.

Preliminary results indicate that intelligibility scores were not higher for the denoised speech compared to the noisy speech. This was not the expected result, however, there is still much to consider within the data. These preliminary findings are grounds for further studies and will hopefully lead to an improvement in future iterations of the speech processing software.

Acknowledgements

I would like to acknowledge the following people for their contributions to this project.

- Katherine White
- Greg Shaikevich
- The Lab for Infant Development and Language at the University of Waterloo
- Ashley Avarino
- Michael Kai Petersen
- Mette Carstensen
- Martin Bergmann
- The Augmented Hearing.io team
- Mathieu Feagan
- John Vanderkooy
- Dan Murray
- Joel Beilin
- Nicholas Richardson
- Charles Zhang
- Ewen MacDonald
- My friends and family for supporting me through this project and degree

Abbreviations

AH - Augmented Hearing.io

AI - Artificial intelligence

DB - Decibels

DSP - Digital signal processing

ESL - English as a second language

HA - Hearing aid

HL - Hearing loss

IPA - International phonetic alphabet

ITU - International Telecommunication Union

LIDL - Lab for Infant Language and Development (University of Waterloo)

LUFS - Loudness unit full scale

MOS - Mean opinion score

PESQ - Perceptual evaluation of speech quality

SIN - Speech-in-noise

SNR - Signal-to-noise ratio

STOI - Short time objective intelligibility

Glossary

Audiogram: An acoustic measurement based on how a listener responds to a series of pure tones in quiet. Each ear is tested at a variety of frequencies to determine which sections of inner hair cells in the cochlea are not responding. This is then plotted on a chart (audiogram) and used to inform the audiologist when programming the patient's hearing aids.

Full band: The sampling rate of a sound file that will not be affected by compression that shrinks the size of the file down, thereby giving a higher quality audio recording.

Hearable: A device that combines functions like listening to music with some adaptations to account for hearing loss or general hearing difficulties. For example, the Apple airpod is a hearable as it has transparency mode, which filters out background noise to focus on speech. Some devices can also do a basic ear bud calibration based on an audiogram that the user uploads.

Hearing aid (HA): Small programmable devices that sit on or in the ear to amplify sounds for folks with hearing loss.

Listening effort: The amount of energy a brain needs to expend to be able to make sense of speech in a conversation.

Listening fatigue: The tiredness that can come from sustaining increased concentration while engaging in conversation.

Loudness: Typically measured in LUFS, the loudness of a sound file is correlated with the peak amplitudes within the recording. The larger the peaks, the louder the signal.

Mean opinion score (MOS): A subjective measurement taken from participants' ratings of the quality of a sound transmission on a Likert quality scale. It can also have questions related to word transcription with specific test words determined based on their phonological relevance to communication in the test language.

Phoneme: The smallest form of distinct speech sounds. For example, in the word 'tie' there are two phonemes: t and i, whereas the word 'pander' would be broken up as p-ah-n-d-er. These sounds can vary across languages, but are often globally represented with the International Phonetics Alphabet (IPA). *I did not use the IPA characters for the sake of clarity in these examples.

Signal-to-noise ratio (SNR): The “signal” refers to the desired sound in a listening situation (for example, speech). The noise is anything that might be distorting the signal, for example, people speaking in the background, wind, or clanking sounds. The ratio compares the two – the lower the SNR, the harder it will be to interpret the signal from the noise.

Speech-in-noise test (SIN): An evaluation of how well a patient can separate speech from background noise when the volume of the noise is varied. There are multiple variations of this test which is explained in detail in section 2.3.4.

Speech intelligibility: The ability of a listener to discern meaning from what is being spoken.

Voicing: Some phonemes are produced with a vibration from the vocal cords, which is referred to as a “voiced” phoneme. An “unvoiced” phoneme would be produced without this vibration. For example, ‘d’ is a voiced consonant and ‘t’ is an unvoiced consonant.

Zoom fatigue: Tiredness experienced from having online meetings all day, often as a result of working remotely.

Contents

1. Introduction

Communication between people is a fundamental part of being human. The world in which we communicate is noisy and the degree to which this background noise influences our conversations depends on the volume and type of noise. It is natural to experience challenges understanding each other when the noise is overpowering, like at a concert or when an alarm is ringing. The real problem presents itself when the noise is not that loud, but the brain needs to expend more energy separating the speech from everything going on in the background (Moore et al., 2017). Because humans encounter this situation so frequently, the extra cognitive effort can take a serious toll on the ability to concentrate and participate in conversation (Plack, Barker, & Prendergast, 2014). There are numerous studies indicating that increased cognitive effort results in longer reaction times and higher self-reported feelings of fatigue (Moore et al., 2017; Murphy, 2021; Athey, 2016; Hicks & Tharpe, 2002). The consequences of this increased effort can escalate with age and as one's hearing naturally declines.

This is a phenomenon that occurs daily, but there are currently few solutions to improve communication in noisy environments. However, there are companies like Augmented Hearing.io (AH) that are working to improve the perception of speech in noise (SIN) using artificial intelligence (AI). My thesis will be looking at one approach to improving communication in noisy situations through a partnership with AH.

1.1. The Current Issue

There has been an increase in young people having trouble understanding SIN (WHO, 2023). This is in part due to the constant noise exposure and dangerously loud music that youth are experiencing in their everyday lives. However, with the rise in wireless earbud technology, it is becoming more fashionable to have devices in your ears (Mordor Intelligence, 2023). There is an opportunity to then make use of the existing earbud technology that consumers are already comfortable with to help people experiencing communication challenges. This is where AH comes in. Their noise removal software is being developed with the goal of having it available for live noise reduction and voice enhancement in earbuds. The goal is to reduce both the cognitive demand and increased fatigue that people experience while trying to sort through speech in noisy environments (Plack, Barker, & Prendergast, 2014).

1.2. Partnership with Augmented Hearing.io

Augmented Hearing.io is a sound processing company based in Lyngby, Denmark. They are a small start up consisting of four full time staff and around 2-3 temporary contractors and students. Their goal as a company is to create a software that ultimately achieves three things: noise reduction, general voice enhancement to improve speech intelligibility, and voice enhancement personalized based on the user's hearing profile and the listening context that they are in. My advisor from AH is Dr. Michael Kai Petersen, who is one of the co-founders and full time staff. I started working with this company as a summer internship, looking into non-human measurements of speech intelligibility. This analysis can be used to validate algorithm changes without having to pay for or take the time to survey participants. It is great for quick validations, but it still does not demonstrate how humans will perceive the audio. As a result, I decided to build upon this summer project as my final project for my bachelor's degree by running a human study.

1.3. Use Cases of AH's Technology

Noise reduction technology can be used in a wide variety of situations. Any situation that requires verbal communication is likely to be influenced by background noises. This is especially relevant in the COVID-19 pandemic era, as remote work has been on the rise. Just the reverberation of one's own voice in a room can result in sound distortion that affects comprehension (Adank, 2012). AH's software could be applied to online meeting platforms to improve the daily work listening experience. In addition to this, everyday life brings noise from traffic, people chatting, wind, music, and a plethora of other situations which add multiple layers of sources that the brain must try and sort through in order to understand the speech present. Just trying to have a conversation in a restaurant or crowded street can be a challenge that AH's software could alleviate.

This technology could also be applied in emergency communication scenarios to help mitigate the increased risk of hearing damage by reducing the noise (from sirens or fires for example) and enhancing communication by isolating the voice. In the case of emergency responders, since they are already typically afflicted by hearing damage due to the nature of their work (Flamme et al., 2019), having the extra processing could improve collaboration between service members while reducing the cognitive load on the individual user. This would reduce the risk of further hearing damage and miscommunications occurring in a highly stressful and time-sensitive environment.

Finally, beyond supplementing existing hearing aid (HA) technology, another application of this software could be for individuals with auditory processing disorder, attention deficit disorder, and other sensory sensitive learners. All of these disorders put extra demand on the brain already, making focusing on auditory stimuli a bigger challenge. The current solution is often to have FM or digital assistive technology systems where the affected listener has sound from, for example, a teacher's microphone, sent directly to their hearing devices (Rance et al., 2014; Johnston et al., 2009). This can be costly for schools and workplaces and still does not allow for quality peer to peer relationship development. Having AH's solution do that processing instead could allow for more successful information intake and easier socializing.

2. Background and Theory

2.1. Hearing Science Overview

The basic pathway of communication starts with the creation of sound from one person's vocal cords and alterations of the mouth and tongue (Ackermann, 2008). That sound then travels through the air, sometimes with other sounds getting mixed in from the outside environment. That sound is later received in the ear of another person where the vibrations of the sound are translated to nerve signals that travel to the speech processing centres in the brain (Mesgarani et al., 2014).

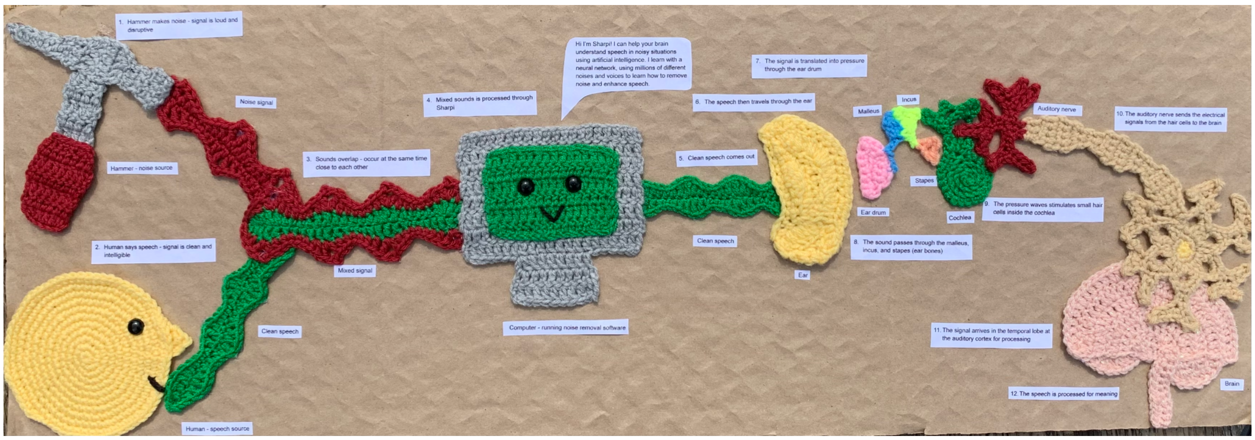

Figure 1: Crochet model of the hearing pathway

There are a number of opportunities for this process to break down along this pathway. The quality of the sound can be altered by lisps or quiet voices, the medium in which the sound travels can be too reverberant or dampening, the translation from sound wave to nerve signal can be impaired by a loss of certain frequency receptors, and the brain's ability to do the actual processing can fail between sorting the speech from the noise and the actual pulling of meaning from what was said (Adank, 2012). AH's software aims to act as one of the processing steps that the brain takes in separating speech from background noises. My project will be looking at how well AH's software can accomplish this processing.

2.2. AH's Neural Network

The solution being evaluated in this study is a software program that is based on a neural network model. This model is called Perceptnet, which is a type of gated recurrent neural network. This software works by learning from thousands of different sound files to understand what constitutes noise and what constitutes voice. This way, when the model is presented with a new speech and noise combination, it can use what it previously learned to separate the two and preserve the speech while removing the noise. This model differs from existing Digital Signal

Processing (DSP) chips currently on the market, as the neural network can provide information on exactly which parameters to focus on in a given situation. Since DSP is static, it will automatically break down in some noisy conditions rather than flexibly adjusting based on previous data, as AH's network can do (Kim, Kim, & Kim, 2019).

An important focus in this type of solution is the latency time. This is a challenge in digital audio processing because if the delay in sound processing from real time speech is longer than 50 ms, the brain can notice and it causes other issues such as confusion and difficulty in following conversation (Kim, Wang, & Maxa, 2020). One way to reduce the latency is by decreasing the number of parameters that are examined by the model. Each parameter, or characteristic of the sound file, such as the pitch at a given moment in time, contributes to the model's overall understanding of what is speech and what is noise. These parameters are responsible for the maintenance of the person's voice so the noise removal does not leave the speech sounding too robotic. Instead of comparing eight million different parameters, however, scaling it down to the most important one million will improve processing time drastically. This reduction becomes more possible as the model is further trained and begins to understand the most important characteristics to preserve. This network may evolve in the future as more noise removal algorithms are developed, but the current PerceptNet satisfies AH's preliminary targets.

2.3. Speech Intelligibility Versus Listening Fatigue

This project has gone back and forth between studying two different but related concepts: listening fatigue and speech intelligibility. It is important to mention this as there may be frequent references to listening fatigue in my literature review (Appendix A) and throughout this paper, despite the ultimate focus of the experimental study being on speech intelligibility. The concepts are closely tied together because the quality of the speech directly impacts the level of fatigue that one may experience when listening (Moore et al., 2017). I will present the two concepts in further detail below before ultimately explaining why speech intelligibility was chosen to be the focus of this project.

2.3.1. Listening Fatigue

At the start of the COVID-19 pandemic, many offices switched to remote work, which led to full days of Zoom or Teams meetings. At the same time, a new phenomenon was coined: "Zoom fatigue." This term refers to the added exhaustion that people began feeling after all those online meetings (Williams, 2021). There was uncertainty as to what caused this added exhaustion, since there was no longer a commute to and from work and people were able to work from the comfort of their own homes. One source of this increased fatigue appeared to be extra listening effort. For example, when users do not have their cameras on, they miss out on facial cues, which can increase listening effort (Williams, 2021). Additionally, the fatigue was coming from having to listen to unnatural sound (compressed and poorly transmitted voice, added background noise, or poor acoustics), which leads to increased cognitive effort required to adjust for these factors (Moore et al., 2017). Similar to eye strain, if the ears are straining to hear and put together the linguistic information coming across the computer, the brain will be exhausted.

Given the rise in remote work and its absolute necessity in many cases, having employees dealing with intense fatigue throughout a day of meetings is a serious problem (Williams, 2021). In online environments, the sound signal passes from a person's mouth to their microphone, through a computer to the recipient's computer, and out a speaker to their ears, where it is processed by the listener. Each of these steps may alter the sound in a way that makes it more challenging to process. Since Zoom fatigue does not happen during in-person meetings, there needs to be a way to replicate the in-person audio experience as closely as possible, so that the brain does not need to put in all that extra processing power. As the goal is for the AH software to return a signal that more closely approximates an in-person experience, this should lead to reduced listening effort, thereby reducing fatigue and improving the online meeting space.

2.3.2. Speech Intelligibility

Speech intelligibility refers to the ability to extract meaning from speech in a listening context. The present study is looking at whether speech intelligibility is enhanced as a result of AH's processing. The two most common automatic intelligibility measures are Perceptual Evaluation of Speech Quality (PESQ) and Short Time Objectivity Measure (STOI). However, these two methods have two issues that make them less optimal for project validation. For one, PESQ and STOI calculate speech intelligibility through automated means, relying on comparing an altered signal to the original "clean" signal and seeing how much the two differ (Taal et al., 2010). This means that regardless of whether the clean speech was actually the most intelligible of the two signals being compared, the new signal will always have a lower intelligibility score due to the original signal being altered. The second issue is that PESQ and STOI do not involve human listeners, so it is not possible to predict how the human speech perception system would truly interpret these signals.

Since AH's algorithm comes in three steps – noise reduction, user preferred voice enhancement, and context specific voice enhancement – the latter two steps cannot effectively be measured with PESQ and STOI. It is, however, an effective method of evaluation for the first step, noise reduction, as the goal is for the new signal, created after the background sounds are removed, to be as close to the original signal as possible. Even without the human perception aspect, having a quick comparison model still makes STOI and PESQ important tools, as it gives AH a way to quantify the quality of their noise reduction algorithm.

Given that these most common measures cannot effectively evaluate the enhancement aspects of AH's algorithm, there are some other measures that can be used. The clean signal can be put into a speech to text converter to get a baseline of how many errors are made, which can then be compared to the enhanced signal put through the same program. If the program interprets more words correctly from the enhanced signal, this demonstrates an increase in intelligibility. However, this solution has its own challenges. Depending on a speech to text algorithm to determine intelligibility does not actually represent how a human might interpret the same signal.

While creating a speech to text software to run intelligibility tests on an algorithm that is deeply known and understood by the AH team would address this issue, it is not a practical use of time when there are free and cheap existing softwares out there. It must also be noted that the speech to text algorithms are constantly evolving, so this may no longer be an issue in the next five years.

A human measure can also be used to evaluate speech intelligibility through the collection of a Mean Opinion Score (MOS) (ITU, 2016). This would address the second issue of PESQ and STOI not being entirely representative of the human hearing system. MOS is currently considered best practice, as it involves asking human participants to rate how intelligible a particular sound sample is (Streijl, Winkler, & Hands, 2016). This ultimately matters more than what a computer thinks, as it is the humans that AH is designing the product for. Even though this evaluation is considered best practice, it is not often done, as it is much more time consuming and expensive than getting a computer to evaluate the signal instead. Having to recruit human participants, design a study, run the study, and pay everyone afterwards before actually getting to analyze the data is not feasible on a smaller scale. This is why measures like PESQ and STOI were created and further considered for this project before ultimately choosing to run a human study. This guarantees that the audio is being interpreted through a human brain rather than a machine, thus providing information from actual future customers.

2.3.3. The Scope of This Project

For this project, I chose to focus on speech intelligibility in human listeners, as it was easier to assess quantitatively than listening fatigue. To truly evaluate the effect that increased listening effort can have on the brain, several physiological measurements need to be taken. A host of studies have used eye tracking technology in addition to dual-task paradigms that assess both the auditory and visual systems to see how the fatigue from one may influence the other (Dingemanse & Goedegebure, 2022; Zhang, Lehmann, & Deroche, 2021; Moore et al., 2017). This form of evaluation was not feasible given the time and resources of this study, however, gathering responses based on the MOS is. As I have researched different study methods for the literature review required to begin this project, I will be considering other possible future studies in section 6.3. However, the information from this intelligibility study alone will be enough to inform AH of their next steps in terms of improving their proprietary algorithm, which is the most important goal of the project.

2.3.4. Speech-In-Noise Tests

In order to evaluate speech intelligibility in human listeners while addressing the quality of the noise removal algorithm, I decided to explore speech in noise (SIN) tests. The basic concept underlying each test is that a speech signal is played at a defined volume and a noisy signal is played at a known signal-to-noise ratio (SNR) to the speech. The participant is then asked to repeat the speech. An easy listening experience for an average listener would be about

+10 SNR (Lee et al., 2022). A negative SNR means the noise is louder than the signal and that would make the signal much harder to distinguish. The exact number is typically measured in decibels (DB). There are a number of specific types of these SIN tests that will be detailed below, each one having certain strengths and weaknesses depending on the context they are being used in.

Different Versions of SIN Tests

- **QuickSIN:** This test, along with HINT (explained below), is the most commonly used method in evaluating SIN comprehension abilities. The test is a higher pitched "female" voice with a four speaker mix babble noise background. The SNR is tested between +25dB to 0dB using a variety of standard sentences. This test is typically used for adult evaluations. (Killion, Niquette, & Gudmundsen, 2004)
- **Hearing in Noise Test (HINT):** This test is conducted with a "male" (lower pitch) talker overlaid with speech shaped noise. It is unique in that it is typically evaluated with the presenting speakers in a variety of locations to also see what the impact of the distance from the sound source is. The test itself consists of ten sentences that are phonemically balanced. (Nilsson, Soli, & Sullivan, 1994)
- **Azbio:** This test is most often used to determine if a patient is a candidate for a cochlear implant. The stimulus consists of male and female voices in conversation style. There are fifteen sentence lists that the test can be drawn from which are more challenging than the HINT ones. (Spahr et al., 2012)
- **Words in Noise (WIN):** This is a test run across seven SNR variants using monosyllabic words. It is more sensitive compared to HINT because of the additional ratios it tests. (Wilson & Watts, 2005)

2.4. This Study's Test Method

The study I made is a type of SIN test. It most closely relates to the QuickSIN test, but is unique in that I am testing users' abilities to discriminate words based on a certain type of speech processing (done by AH's algorithm) rather than based on their own hearing abilities or sound processing that is occurring in a device that they have (such as a HA). I also chose to have a constant SNR of +10dB, which differs from testing along a scale of 0 - 25dB. More details on exactly how my test will be run is provided in section 3.

2.5. Research Question

My thesis is focused on validating the noise reduction and voice enhancement features of AH's algorithm. I specifically focused on how AH's algorithm impacts a user's ability to distinguish between similar English consonants. The results will inform how well the algorithm preserves the original speech when it removes different kinds of noise. I want to use this information to create recommendations for adjustments that the algorithm may need to improve speech intelligibility for future users. Knowing which words are more successfully understood is important in suggesting future improvements in the algorithm. If there is a consistent issue with the “k” sound, for example, this would warrant further consideration in the filters that are currently being applied to the test set files. Ultimately, I will be asking: Is the speech enhancement algorithm from Augmented Hearing.io capable of improving speech intelligibility scores in adult listeners, especially for hard to distinguish sounds? To do this, I will be evaluating the accuracy of responses that participants submit while listening to hard to distinguish words in different noisy conditions.

3. Study Methods

3.1. Recruitment

Participants were collected on the SONA recruitment platform. This is a site that is connected to students taking Psychology classes at the University of Waterloo. The study is posted on the website using a pre-approved description (from the ethics board of the University of Waterloo). Students interested in my study are then redirected from SONA to Gorilla, the platform that hosts the study. Gorilla was chosen for this type of study due to its audio integration and randomization capabilities. It is also a standard software used for studies at the Lab for Infant Language and Development (LIDL) which one of my advisors directs. Once the participants are on Gorilla, they must consent to beginning the study. This study is approved under the University of Waterloo’s ethics board, REB #44785.

3.2. Study Tasks

3.2.1. Tone Test

Participants begin the study by undergoing the AntiPhase Tone test as detailed by Milne et al. in *An online headphone screening test based on dichotic pitch* (2021). This is standard practice in the LIDL. This test was included to help participants adjust the volume on their computer before listening to the study audio and to give me a sense of whether the participant used earbuds as requested or if they used computer speakers. The AntiPhase Tone test plays three tones, each varying in loudness. The participant is prompted to pick the quietest tone of the set. This is repeated six times. A score of 4/6 is required to pass. Two of the tones are the same, but the third is shifted 180 degrees. When listening on a loudspeaker, the signal would be weakened due to the destructive nature of the phase shift, causing the listener to be unable to determine which tone is actually the quietest (Milne et al., 2021). Knowing whether a participant is using headphones is important as the earbuds help minimize background noise in the participant's environment as well as mitigate the effects of a room's acoustics.

3.2.2. Word Listening Task

After the tone test, study participants begin the main task of identifying words in noise. They are presented with a series of target words in isolation under different sound processing conditions. The selection of words is further discussed in section 4.1.1. Each participant is assigned to one of twelve conditions, based on the type of speaker, the type of noise, the placement of the target phoneme, and the type of noise removed using AH's software, see figures 2 and 3. On each trial, the participant is presented with a play button. After they click it, they must listen to a single audio file. The participant can only listen to it once as it may be possible to figure out a word with repetition over time, but that is not representative of a conversational context where words are exchanged quickly. After listening to the file, the participant is provided a text box to type in the word that they think they just heard. In the results, I will be accounting for spelling inconsistencies (e.g., where a target word was "do", but the participant typed "due") by correcting the final intelligibility scores as required. Each participant is presented with 108 audio files: 36 with no noise, 36 with noise, and 36 with noise removed. The 108 files are given in a random order, so the participant cannot predict whether the target phoneme is in the initial, medial, or final position, nor what kind of processing has occurred. However, each participant is only presented with one voice for consistency.

3.2.3. Demographics Survey

Finally, the study participants are presented with a questionnaire that has twelve questions related to their language background and listening habits. These questions are required in order to give context to who is participating in the survey. If there are only 18 year olds completing the study, their hearing profiles may be different than an 80 year old. Similarly, if English is not their first language, more phoneme confusions will be expected, as it can be harder to differentiate phonemes in a non-native language (Peng & Wang, 2019). Knowing this information will inform what kinds of participants to include in further studies, detailed in section 6.3, as well as provide some explanation for the results. The demographics questions are listed in section 4.2. After completing the survey, participants are redirected to a thank you page and are later awarded a 0.5 credit towards their SONA psychology course requirements for doing the study.

3.3. Conditions

Condition	**1**	**2**	**3**	**4**	**5**	**6**
Voice	Salli	Salli	Salli	Salli	Salli	Salli
Phoneme Placement	No noise: **Beg** Noise: **Mid** Denoised: **End**	No noise: **Mid** Noise: **End** Denoised: **Beg**	No noise: **End** Noise: **Beg** Denoised: **Mid**	No noise: **Beg** Noise: **Mid** Denoised: **End**	No noise: **Mid** Noise: **End** Denoised: **Beg**	No noise: **End** Noise: **Beg** Denoised: **Mid**
Noise	Babble	Babble	Babble	Wind	Wind	Wind
Processing	Babble denoised	Babble denoised	Babble denoised	Wind denoised	Wind denoised	Wind denoised

Figure 2: Conditions 1-6 in the study

Condition	**7**	**8**	**9**	**10**	**11**	**12**
Voice	Joey	Joey	Joey	Joey	Joey	Joey
Phoneme Placement	No noise: **Beg** Noise: **Mid** Denoised: **End**	No noise: **Mid** Noise: **End** Denoised: **Beg**	No noise: **End** Noise: **Beg** Denoised: **Mid**	No noise: **Beg** Noise: **Mid** Denoised: **End**	No noise: **Mid** Noise: **End** Denoised: **Beg**	No noise: **End** Noise: **Beg** Denoised: **Mid**
Noise	Babble	Babble	Babble	Wind	Wind	Wind
Processing	Babble denoised	Babble denoised	Babble denoised	Wind denoised	Wind denoised	Wind denoised

Figure 3: Conditions 7-12 in the study

Each participant receives a set of stimuli with no noise, a set of stimuli with one type of noise, and a set of stimuli with its noise removed counterpart. To avoid having participants listen to the same words under multiple processing types, participants received different sets of words (differing in target phoneme placement) for each condition. This, combined with the word randomization, mitigates the effects that stimuli presentation order may have on participant results.

4. Test Set Development

See appendix B for the test sets being used in the study.

4.1. Phoneme Selection

The test sets for my study were made with the information from Cutler et al. (2004)'s paper on bilingual speech intelligibility. In their study, they examined the perception of similar (confusable) speech sounds as a function of listener language background and noise condition (SNR level, etc.). They created confusion matrices that demonstrated average confusion of sounds across these conditions. I chose to focus on one specific matrix in the paper that summarized the average responses to ensure that the often confused phonemes were reflective of a larger population, rather than a small subset. My target words were determined from the top six most confused phoneme pairs from the aforementioned matrix. Some specific words were also taken from the International Telecommunication Union (ITU) standards on measuring speech intelligibility (2016). The ITU standards doubly confirmed the phonemes to be focused on in the context of sound quality, as these words were originally chosen to evaluate the efficacy of signal transfer across a telephone wire.

The target phonemes in my stimuli are consonants because, in the English language, they give the most semantic information to the listener (Varenina, 2018). This means that if a consonant is misinterpreted, it has a greater potential impact on meaning than a vowel. I identified pairs of words differing in a specific consonant phoneme, where the differing consonants are related by the fact that they require the same mouth shape but one is voiced and the other is voiceless. In other words, one of the consonants in each pair requires more vocal fold vibration than the other. This can make these specific phonemes easy to confuse for one another which is why I selected the specific target phoneme pairs in this study.

Across word pairs, the target phoneme occurs either at the beginning of the word, middle of the word, or end of the word. The placement of the phoneme matters, as the preceding phonemes can impact how the word is interpreted. The following pairs were used: T - D, S - Z, Ch - J, P - B, K - G, F – V (please note that these phonemes are written in English spelling, rather than the International Phonetic Alphabet (IPA) for ease of understanding by non-linguists). For example, one pair is tie - die. This is where the potentially confusable phoneme is at the beginning of a word. An example of a middle pair is etching - edging. An end placement example could be relief - relieve.

I generated three sets of 12 word pairs with each phoneme placement (beginning, middle, end) for counterbalancing purposes (assignment of word sets to processing conditions).

4.1.1. Word Selection

The words themselves were selected using a variety of different resources. The primary resource I used was the *Complete List of Minimal Pairs* chart from English Phonetics (n.d.), which consists of a phoneme matrix linking different phonemic word pairs. Each word was then put into the Corpus of Contemporary American English (COCA) database to ensure the words were of relatively equal frequency in the English language. If the frequencies were too unbalanced, then it is possible that a participant would hear one word over another due to familiarity with the word instead of mishearing the phonemic differences. It is important to note that frequency is based on spelling rather than phonetics, so the words being balanced are being directly compared by spelling. Therefore, the frequencies may not reflect spoken frequency, because of the presence of homophones (different words with the same spoken form). A full list of the stimuli used in the study is available in appendix B. Similarly, a list with all the stimuli and their corresponding COCA frequency score is listed in appendix C.

4.2. Demographic Questions

Each question was chosen to attempt to cover other reasons (outside of the software not improving intelligibility scores) that participants may have challenges with the study tasks.

1. How old are you?
   Drop down number list
2. Is English your first language?
   Yes/No
   a. If not, what was your first language?
      Open text box
3. How many years have you been speaking English?
   Drop down number list
4. How would you rate your English proficiency?
   Needs improvement, conversational, native/fluent
5. Have you ever been diagnosed with a hearing loss?
   Mild, Moderate, Severe, Profound
6. Do you use hearing aids?
   Yes/No
   a. If yes, how many hours (on average) do you wear them per week?
      0-10, 10-20, 30-40, 50+
7. On average, how many hours per week do you spend with earbuds in?
   0-2, 2-5, 5-10, 10+
   a. What volume do you typically listen at?
      Low, Medium, Loud, Extra loud

   b. If applicable, do you use active noise canceling features on your earbuds?
      Often, Sometimes, Never
8. What earbuds/headphones will you be using to complete the study?
   Open text box
9. How would you describe your ability to hear in a crowded restaurant?
   I can hear fine, I have some challenges but feels normal, I miss about half the conversation, I would not go out because I can't hear in those situations at all
10. Have you ever worked in construction or any other jobs that might have you dealing with loud noises often (ex. A first responder)?
    Yes/No
11. How long have you worked in that environment?
    0-6 months, 6 months-2 years, 2 years+
12. Do you have any known auditory processing disorder or other cognitive impairments that may impact your ability to process sound?
    Yes/No/Unsure

4.3. Voice Generation

The voices used to produce each word in the test set were generated through the AWS Polly neural text-to-speech program. Two voices were selected in order to compare how well the noise removal software could handle different frequency voices. On the AWS system, the two voices chosen for this study are called "Salli" and "Joey," both of whom have standard American accents. Salli is the more high pitched typical "female" range voice and Joey is the lower pitched typical "male" range voice. AWS Polly offers two forms of text-to-speech processing, one called "standard" and the other called "neural." The neural processing was chosen as that is the most lifelike version of artificially generated speech available, according to AH. The neural processing uses additional AI networks to further process the voices to sound more human-like. For this project, the more human-like the better because the voices are meant to be representative of speech a participant might encounter in their daily lives.

The test sets were generated using artificial intelligence instead of live human recordings because of the consistency in sound processing that is available from the text-to-speech options. If a human were to record the word list for this study, there are additional variables that would be more challenging to control. The distance from the microphone, the tone of voice and associated volume, and the quality of the recording microphone itself would all impact the final recording. Using a processor like AWS Polly means that all of this can be controlled through one standardized processing service.

4.3.1. Effects of Listening to AI-Generated Speech

In section 2.3.1, listening fatigue, I mentioned that unnatural sound from computer processing can lead to an increase in fatigue. My study is using computer generated voices instead of human voices and the listening tasks will be through a computer and ear buds rather than a live conversation. This setup is intentional in order to see how well the computer algorithms can replicate the quality of sound from these digital situations. While the voices are theoretically more robotic than a live person through Zoom, for example, speech created through the AWS Polly system is relatively natural sounding. Therefore, the effects of the artificial voice should be negligible on listening fatigue and speech intelligibility (Simantiraki, Cooke, & King, 2018).

4.4. Noise Overlay

Each stimulus voice file is recorded and then overlaid with one of two noise files before being further processed under a variety of conditions. The first noise file with multi-talker babble was given to me by the company, AH. Multi-talker babble is the most challenging noise to separate from speech because of the bits of speech that may be pulled out of the noise and mistaken as voice to focus on. The babble file was originally a stereo 32-bit float at 44.1 kHz. This file was then split from stereo to mono and one of the channels was deleted. The combined speech and noise files were saved at 48kHz since this is the closest option to full band, which AH's algorithm is optimized for. The higher the sampling rate, the more challenging it can be to look at different segments to separate out the speech from the noise since there are so many more bits to examine. However, since AH already accounted for that in their algorithm so, to ensure the highest quality possible, the study will be using fullband to benefit from that detail. All these steps were done to ensure that each of the sound files was uniformly processed. If one file is in stereo and the other is in mono, the stereo modality will change the way the brain perceives the sound, thereby influencing the listening results, especially when it will be overlaid with a mono track with speech.

The next type of noise overlaid is a type of noise called dynamic non-babble. This is a sound file like wind, for example, that has a variety of volumes and frequencies but has no speech so the noise itself is easier to distinguish from the voice. I chose to use wind for this type of noise as it is a common cause of communication challenges (Grenner et al., 2000). Note that I am not using noise files that are proven to increase stress levels, like ambient traffic sounds (Jafari, Kolb, & Mohajerani, 2018). This is so that the results of the study are not affected by the stress associations that may come from these sounds. The use of two different noise types (babble and wind) will allow for an examination of how the AH algorithm works under different conditions.

4.5. Controlling Signal-to-Noise Ratio

Each file is processed using the loudness normalization function on Audacity so that the standard broadcast guidelines for Advanced Television Systems Committee (ATSC) is achieved. The standard broadcast loudness level for Canada and the United States is -24 LUFS, or loudness units in full scale (Norcross, Lavoie, & Thibault, 2011). LUFS is a measurement that takes into account peak volume levels as well as how the human ear perceives these levels. The signal can then be normalized or reduced in overall volume relative to itself in order to satisfy the standards of the governing media regulators. This process needs to occur because of the lack of a consistent loudness scaling factor (Oetting et al., 2018). When the speech files are being overlaid with noise, they are then being altered to be at a specific ratio, SNR. If the speech file is at a higher LUFS than the noise file, the SNR will not be accurate, which is why normalization needs to occur.

For this experiment, an SNR of 10 was chosen as a baseline due to prior trials at AH. It also corresponds with a comfortable listening level, but with enough noise to lead to some difficulties (Lee et al., 2022). Because the noise removal portion of the software is more effective when the noise is much lower than the voice, making it easier to separate, knowing exactly where there is a breakdown that makes it unusable is helpful to create more targeted data sets for the neural network model to train on. The potential of varying the SNR will be discussed in section 6.3, future studies.

4.6. Predictions

Given the preceding methods, I will be making some predictions based on the literature review and prior tests that AH has run with their software. It is important to be able to compare the following expectations to the final results to troubleshoot why the results may differ and later inform the recommendations. Overall, I expect that the processing condition with the highest error rate will be the one with noise and no removal software applied. I expect both the noise removed and no noise conditions to be significantly more intelligible than the noise conditions. However, I do not have a prediction about which of the two will be more intelligible.

Next, we must look at the type of noise that is being suppressed. Babble noise has speech characteristics, which can sometimes trick AH's software into thinking that it should not be suppressed. Wind noise is more static and uniform, with no speech qualities to potentially confuse the program. As a result, it is expected that the wind noise removed conditions will be marginally more successful (i.e. have higher intelligibility scores) than the babble noise removed conditions.

After the noise, we must look at the two types of voices being used. Joey's voice is lower in pitch, which can sometimes sound like common background noises such as traffic and fan whirring. Salli's voice is a more unique frequency compared to the noise that the software is currently trained to recognize, so it is expected that Salli's voice will be of a higher quality when run through AH's software, and thus have a higher intelligibility score. It is important to note, however, that since AH's model is constantly growing and being trained, the difference between the voices may be less of a concern with the current study model (February 2023), rather than an early prototype when this project first started in the summer of 2022.

Finally, the placement of the target phoneme can also impact the final intelligibility results (Dufour & Grainger, 2019). Acoustic cues come from several different aspects of speech, from the voicing (or lack thereof) of a consonant to the length of a vowel. This is why the target phonemes need to be placed at three key locations in the stimuli set. For example, with the target phoneme in the final position, the listener gets the benefit of the preceding vowel, but loses some of the acoustic cues of the consonant (Woods, 2010). As there are so many different factors that can contribute to the effects of a placement in a given word, I will not be making general predictions for this section. The effects of placement in regards to the algorithm's processing may be more pronounced in future studies with longer sections of speech.

5. Results

5.1. Summary of Results

My study was run on SONA over a two week period. 152 people registered for this study. The average age was 18-20 years old. 62 participants were discarded for three main reasons: They did not complete the entire study, they did not complete the study wearing headphones as requested *and* failed the tone screening test, or they had a known hearing loss or auditory processing disorder. That left 90 participants with usable data points. Note that AH's noise reduction software can be used for people with hearing loss or processing disorders. However, these participants were discarded as there were too few to reach any meaningful conclusions.

In each condition, participants completed three sets of 36 differently processed sound files. For visualization purposes, figures 4-9 demonstrate the averages for each processing condition, as a function of voice and target phoneme position. These figures demonstrate that, overall, intelligibility scores were highest in the clean (no noise) speech and there is no advantage for the denoised conditions relative to the noisy conditions. Further examination of these patterns is available in section 6.1.

5.2. Intelligibility Scores

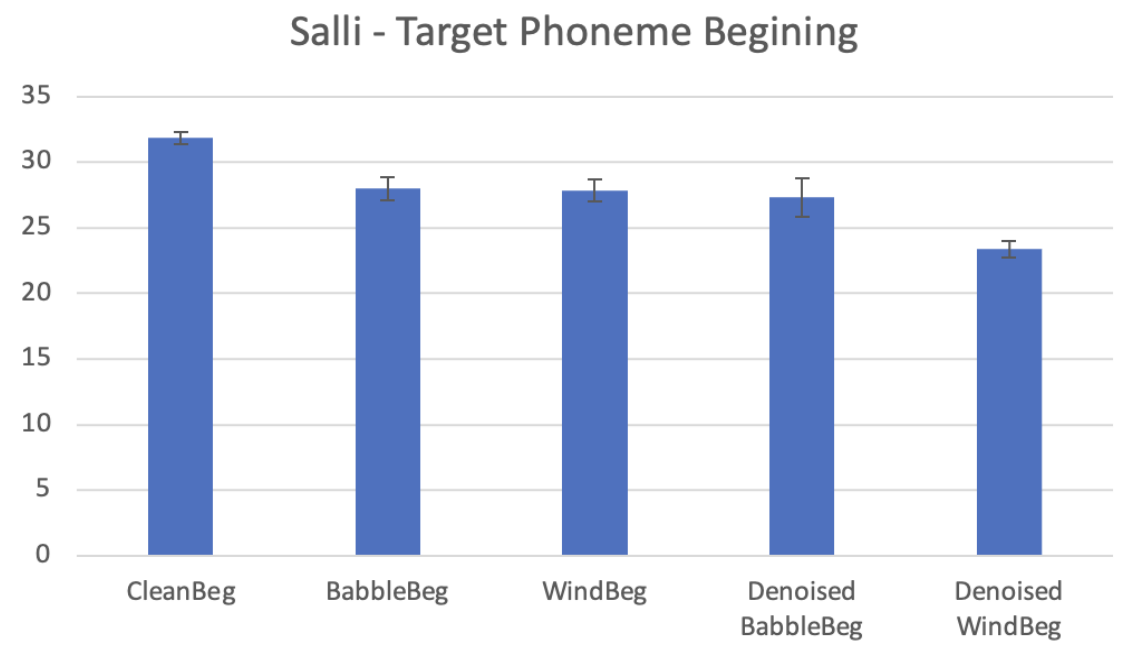

Figure 4: Number correct out of 36 for conditions 1-6 (Salli's voice) with target phoneme at the beginning

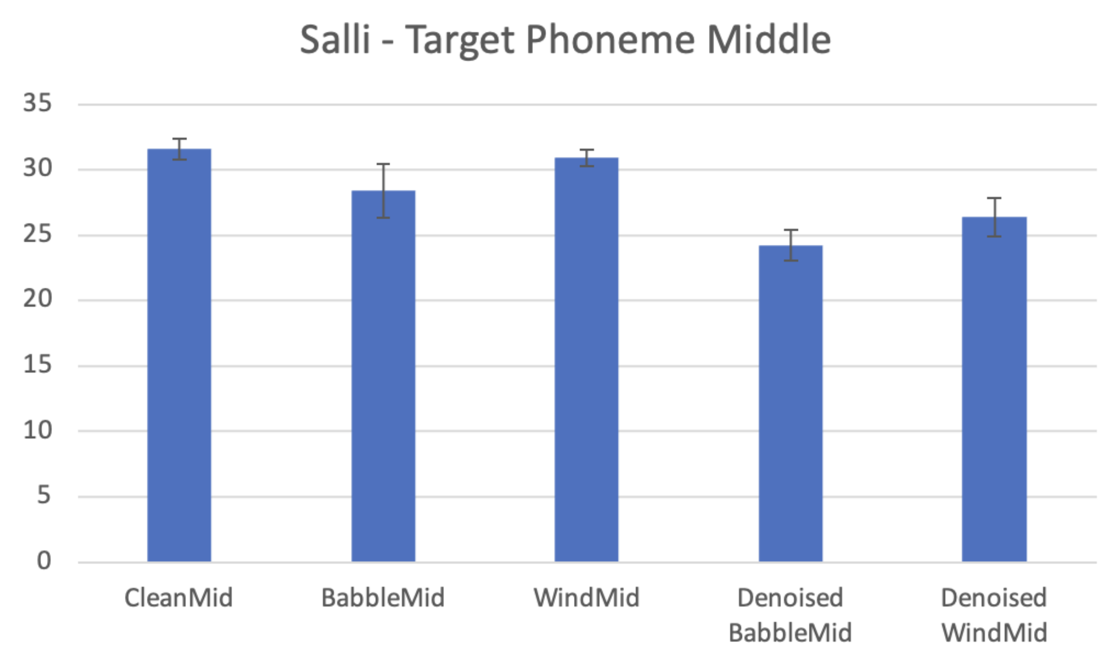

Figure 5: Number correct out of 36 for conditions 1-6 (Salli's voice) with target phoneme in the middle

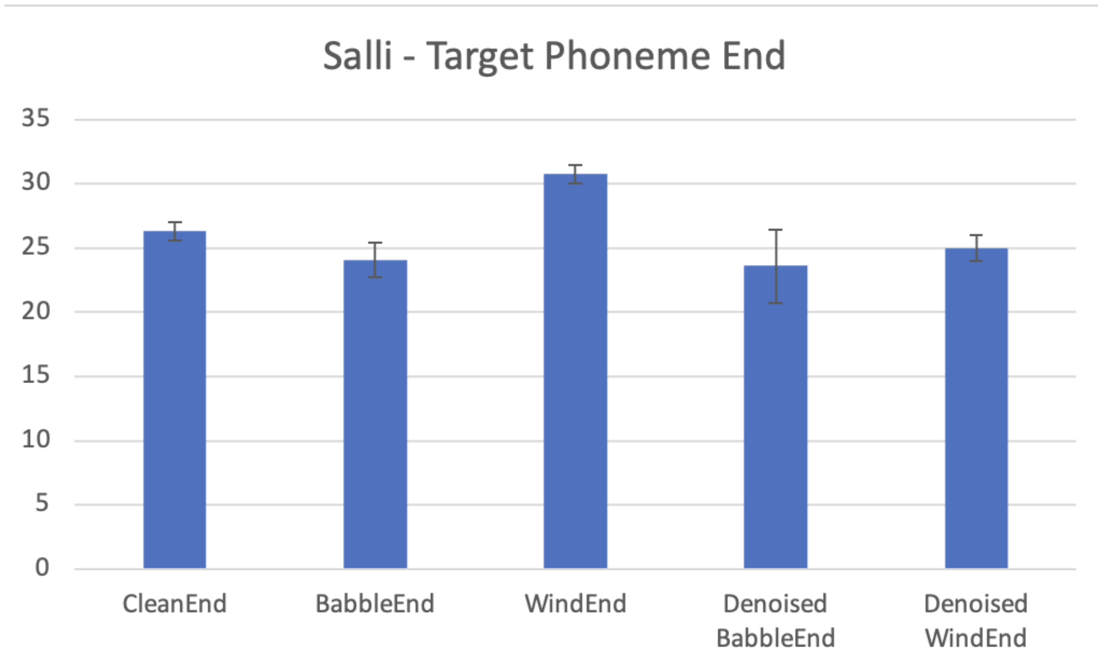

Figure 6: Number correct out of 36 for conditions 1-6 (Salli's voice) with target phoneme at the end

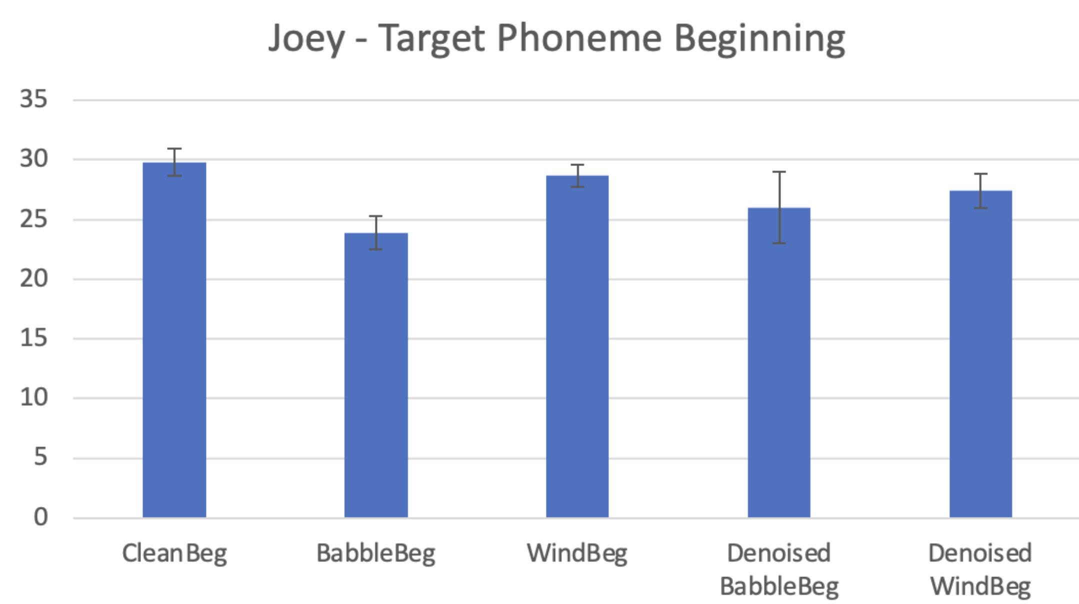

Figure 7: Number correct out of 36 for conditions 7-12 (Joey's voice) with target phoneme at beginning

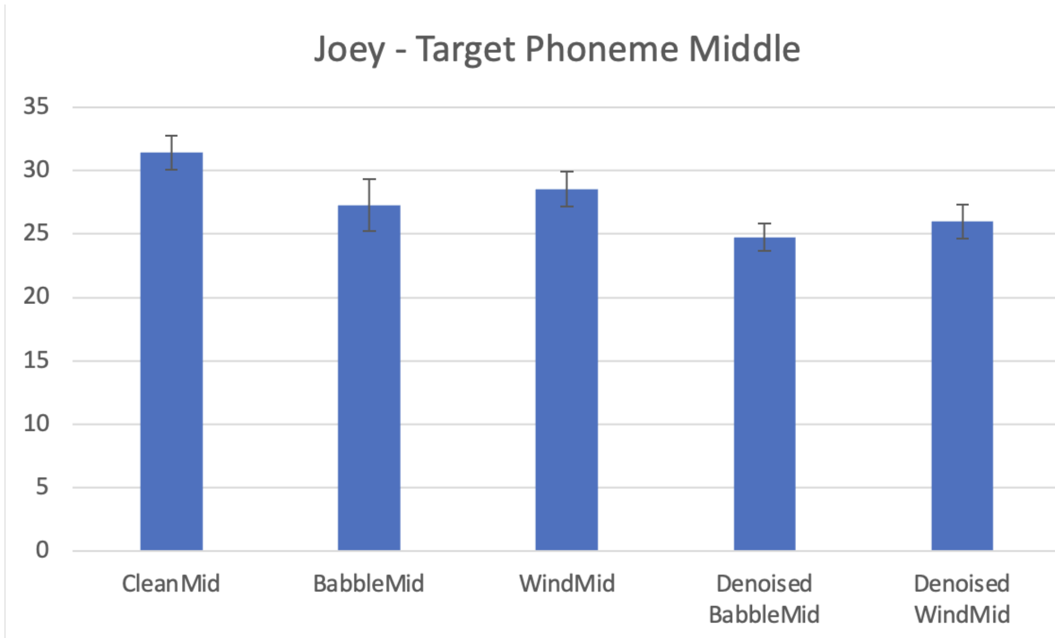

Figure 8: Number correct out of 36 for conditions 7-12 (Joey's voice) with target phoneme in the middle

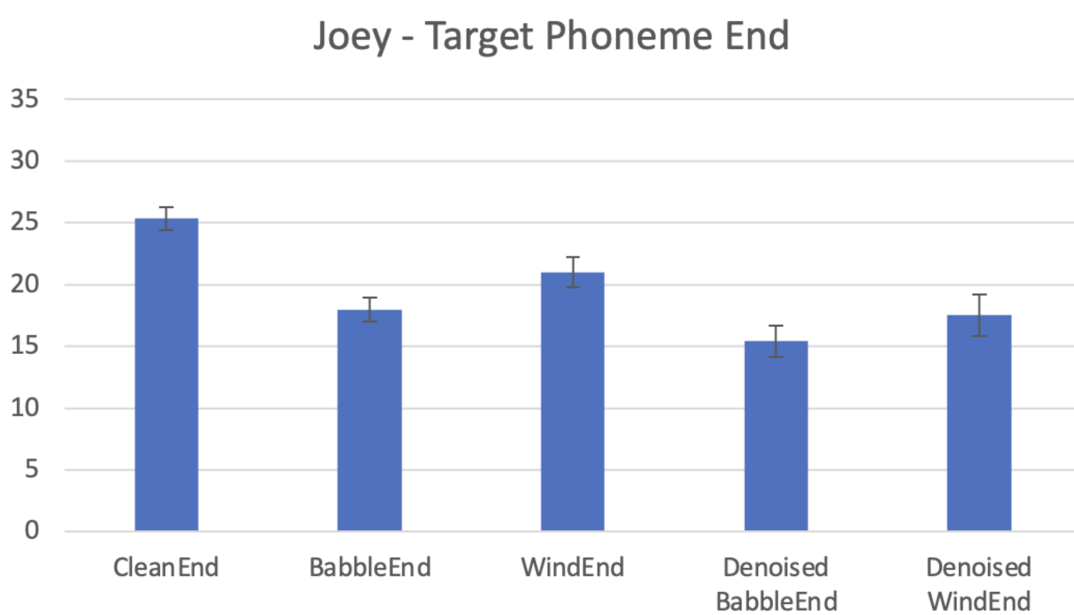

Figure 9: Number correct out of 36 for conditions 7-12 (Joey's voice) with target phoneme at the end

6. Discussion

6.1. Patterns of Results

Looking across all conditions and voice types in the above figures (4-9), the results indicate that the noisy and denoised conditions are not significantly different (which was confirmed informally by t-tests comparing each noise type and its corresponding denoised version). This means that the AH algorithm did not improve the intelligibility scores. This was not predicted, as theoretically, removing the noise should have improved intelligibility, potentially to the same level as the clean speech. This is not too consequential to the current software use, as a misprocessed sound resulting in a completely different interpretation of what the speaker was attempting to communicate which is going to be no worse than listening in a noisy situation anyways.

That being said, these unpredicted results could be explained by a few different factors. For one, the sample size contributing to each processing condition was highly variable, with one condition only having two participants after the others were deemed ineligible. It remains to be seen whether these patterns will hold with a larger data set. In addition to this, the algorithm itself is optimized for larger chunks of speech, as the solution is meant for use in conversation. This may result in less accuracy in removing noise from single word stimuli, as each audio file containing one word was individually processed through the software. Aside from these reasons, it is also possible that the algorithm is not at the stage of its development where it can achieve the same intelligibility scores as the clean speech. In order to advance AH's noise removal software, a series of recommendations for what training the model should focus on is included in section 6.2.

Outside of the noise removal algorithm, there are other observations that can be drawn from the study results. Looking across the two different voice types, Salli and Joey, it is also evident that the intelligibility scores were higher for the Salli voice as predicted. Between the babble and wind noise, the wind noise resulted in better intelligibility scores, again following previously made predictions. Finally, looking at the placement of the target phoneme, words with target phonemes at the end received the lowest scores (and this was confirmed statistically with t-tests comparing scores as a function of position, collapsing across processing).

6.2. Recommendations

Since one third of participants were lost due to a variety of contraindicating factors, further studies with more participants should be developed and run. Details about the kinds of studies and what to look for are covered in section 6.3. In addition to conducting further research, as mentioned in the previous section (6.1), this study does suggest that certain acoustic cues are not being maintained by the noise removal algorithm, resulting in lower intelligibility scores when compared to both clean and noisy speech files. As a result, the model should be developed to improve the preservation of these cues. This can happen through providing the model with more training data, especially with speech content containing the target phonemes covered in the study. Further training should also focus on eliminating the difference between the higher pitch and lower pitch voice. The model should be trained on more lower pitch voices with noise files in a similar frequency range. The more examples and variety of clean speech and noise that the model receives, the better it will be able to distinguish between voices in a similar pitch range to Joey. Ultimately, the greatest benefit for the software will be more training with diverse data sets, after which it may be beneficial to run similar studies to evaluate improvement.

6.3. Future Studies

There is a lot of opportunity to build on this study design for future research. Originally, the study was going to consist of three listening tasks rather than just one. Due to project time constraints, two of the tasks were set aside. However, the stimuli were still developed and can still be viewed in appendix D. The two other tasks were sentence-based. The sentences provide more cues to the brain than isolated words, which may impact the end result. Sentences also represent a more typical listening experience, which is why it is important to do further studies using these stimuli.

In addition to different kinds of tasks, the survey population should be varied in future iterations. As SONA was used as a recruitment service, the majority of participants are in the early years of their undergraduate degree. It is important to get more data with older participants who are less confident about their hearing ability but have yet to be diagnosed with a hearing loss. Since there were also not enough participants with a known hearing loss or auditory processing disorder to make any meaningful conclusions, these two populations should also be a focus for future recruitment. Finally, there is a good representation of English as a second language (ESL) participants in the data. However, performance was not considered as a function of language experience given time constraints. Moreover, it would be interesting to look at even more types of languages and different English abilities to see if the algorithm can make listening to English an easier experience for learners.

In the future, the survey design could be varied to optimize spreading the different condition types across participants. An alternate counterbalancing procedure in which the listeners are presented with 12 words in each processing condition rather than 36 in one and none in the other two would allow for more diverse participant data to further consider placement as a potential influence to AH's algorithm. Further consideration should take place about whether the stimuli should be presented in blocked sets based on the processing type. I chose to arrange stimuli in random order, but it may be more representative of real world situations by presenting one noise type at a time.

Further data could also be pulled from this study design in future research. Due to time constraints, there was not an opportunity to delve into whether there were certain minimal pairs that the algorithm handled better than others. If one minimal pair has lower intelligibility scores than another, the particular verbal formation of that pair may have certain acoustic cues that the algorithm is not preserving. This would also apply to the voiced versus unvoiced cues within each pair. As there is an entire international phonetics alphabet, there could be specific testing done to evaluate how the algorithm handles each phoneme. However, as I have created a stimuli set with the most semantically meaningful and often confused pairs in the English language, using the six minimal pairs from my study would likely yield beneficial results without needing to test the over 45 sounds in the IPA (International Phonetic Association, n.d.).

Lastly, there is more research that can be done with varied processing conditions. In the future, it will be important to vary the SNR level to see how well the algorithm can handle different levels of noise. The noise files could also be varied to test for specific population needs. For example, testing with construction noises in the background could be useful to see how well the solution could work in a construction industry. Along with the noises, the voices should be altered to reflect a wider population. Voices with other accents, higher or lower pitches, and different characteristics should be tested to ensure the algorithm can handle a wider variety of speech. This algorithm will also need to be tested in different languages as it begins to roll out. The focus of this particular study was on the maintenance of the quality of English consonants. English vowels, different phoneme confusion pairs, and whole other languages may react differently. This is especially true of tonal languages like Mandarin, since entirely different voice characteristics need to be preserved by AH's noise removal software (Ortega-Llebaria, Nemoga, & Presson, 2017).

6.4. Implications

As the noise removal software develops and more studies are run, this technology has large potential for improving speech intelligibility in a variety of environments. In addition to its usefulness in applications like Zoom, as discussed in section 1.3, there is a larger demographic that could benefit from this type of solution. Hearing people may experience a reduction in fatigue using this software, but people with untreated hearing loss (ie. not using hearing aids, sign language, or a mix of communication styles) often experience even higher levels of fatigue from noisy situations, even in high SNR levels and current solutions do not meet their needs (Bess & Hornsby, 2014).

Even with HAs, as they tend to amplify *all* noises around the user, it results in the noise being made louder along with the speech. This can make it more challenging to follow a conversation. Some HAs use beamforming technology, which means the microphones only pick up sounds from a certain direction, which can reduce some noise from behind the user (Green et al., 2022). The problem with this is that the user completely loses all awareness in the directions where the microphone is not pointing. This could mean, for example, that a HA user would not be able to hear their name called out if it was behind them and their microphones were set to only pick up forward sound. HAs also carry a lot of stigma due to their association with aging, as age-related hearing loss is one of the most common types (Uchida et al., 2018). Not only that, but HAs are prohibitively expensive, selling for upwards of $5000 a pair. These issues combined can lead to a reduced device uptake, resulting in future communication challenges, social isolation, and increased risk of dementia (WHO, 2023).

In addition to traditional hearing loss, hidden hearing loss is a new term that has been coined to address folks that have difficulty hearing not due to damage in the structure of the inner ear, as is common with noise exposure and age decline, but due to the loss of speech processing neurons in the brain (Monaghan et al, 2020). The result of this neuron loss is typically a normal audiogram, but significantly increased challenges understanding speech in noise (SIN) (Plack, Barker, & Prendergast, 2014). The hearing loss is then “hidden” to audiologists, as many only perform pure tone hearing tests (Portnuff & Bell, 2019). As a result, people may feel that they have a hearing problem, but do not get the support they need in an audiology clinic (Davidson et al., 2021). Not only that, but even if this issue was discovered through testing, traditional HAs are not typically useful, as it is a brain processing issue, not an ear sound reception issue. Given the relatively new discovery of hidden hearing loss, it is important to begin developing solutions to help the affected demographic.

Given all of these considerations – such as the difficulties that hearing aids bring, especially for milder forms of loss that do not require so much amplification, the brain processing difficulties young people are beginning to experience, and the increased noise in our real-world and online environments – there is a huge potential for AH’s software to make a difference.

References

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, *31*(6), 265–272. https://doi.org/10.1016/j.tins.2008.02.011

Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, *122*(1), 42–54. https://doi.org/10.1016/j.bandl.2012.04.014

Athey, H. (2016). *The effect of auditory fatigue on reaction time in normal hearing listeners at different signal to noise ratios*. JMU Scholarly Commons. https://commons.lib.jmu.edu/diss201019/108/?utm_source=commons.lib.jmu.edu/diss201019/108

Bess, F. H., & Hornsby, B. W. Y. (2014). Commentary: Listening can be exhausting – fatigue in children and adults with hearing loss. *Ear And Hearing*, *35*(6), 592–599. https://doi.org/10.1097/aud.0000000000000099

Complete List of Minimal pairs in English: More than 9000+. (n.d.). English Phonetics. https://englishphonetics.net/english-pronunciation-tools/minimal-pair/english-minimal-pair-list.html

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. https://doi.org/10.1121/1.1810292

Davidson, A., Marrone, N., Wong, B., & Musiek, F. (2021). Predicting Hearing Aid Satisfaction in Adults: A Systematic Review of Speech-in-noise Tests and Other Behavioral

Measures. *Ear &Amp; Hearing*, *42*(6), 1485–1498. https://doi.org/10.1097/aud.0000000000001051

Dingemanse, G., & Goedegebure, A. (2022). Listening Effort in Cochlear Implant Users: The Effect of Speech Intelligibility, Noise Reduction Processing, and Working Memory Capacity on the Pupil Dilation Response. *Journal of Speech, Language, and Hearing Research*, *65*(1), 392–404. https://doi.org/10.1044/2021_jslhr-21-00230

Dufour, S., & Grainger, J. (2019). Phoneme Order Encoding During Spoken Word Recognition: A Priming Investigation. *Cognitive Science*, *43*(10). https://doi.org/10.1111/cogs.12785

English-Corpora: COCA. (n.d.). https://www.english-corpora.org/coca/

Flamme, G. A., Goldfarb, D. G., Zeig-Owens, R., Hall, C. B., Vaeth, B. M., Schwartz, T., Yip, J., Vossbrinck, M., Stein, C. R., Friedman, L., Cone, J. E., & Prezant, D. J. (2019). Hearing Loss Among World Trade Center Firefighters and Emergency Medical Service Workers. *Journal of Occupational &Amp; Environmental Medicine*, *61*(12), 996–1003. https://doi.org/10.1097/jom.0000000000001717

Green, T. J., Hilkhuysen, G., Huckvale, M., Rosen, S., Brookes, M., Moore, A., Naylor, P. A., Lightburn, L., & Xue, W. (2022). Speech recognition with a hearing-aid processing scheme combining beamforming with mask-informed speech enhancement. *Trends in Hearing*, *26*, 233121652110686. https://doi.org/10.1177/23312165211068629

Grenner, J., Abrahamsson, U. B., Jernberg, B., & Lindblad, S. (2000). A comparison of wind noise in four hearing instruments. *Scandinavian Audiology*. https://doi.org/10.1080/010503900750042734

Hicks, C. B., & Tharpe, A. M. (2002). Listening Effort and Fatigue in School-Age Children With and Without Hearing Loss. *Journal of Speech, Language, and Hearing Research*, *45*(3), 573–584. https://doi.org/10.1044/1092-4388

International Telecommunications Union (ITU). (2016). *A Subjective test methodology for assessing speech intelligibility*. ITU. https://www.itu.int/rec/T-REC-P.807-201602-I/en

International Phonetic Association (IPA). (n.d.). *The International Phonetic Alphabet and the IPA Chart | International Phonetic Association*. https://www.internationalphoneticassociation.org/content/ipa-chart

Jafari, Z., Kolb, B. E., & Mohajerani, M. H. (2018). Chronic traffic noise stress accelerates brain impairment and cognitive decline in mice. *Experimental Neurology*, *308*, 1–12. https://doi.org/10.1016/j.expneurol.2018.06.011

Johnston, K., John, A., Kreisman, N. V., Hall, J., & Crandell, C. C. (2009). Multiple benefits of personal FM system use by children with auditory processing disorder (APD). *International Journal of Audiology*, *48*(6), 371–383. https://doi.org/10.1080/14992020802687516

Karawani, H., Jenkins, K., & Anderson, S. (2022). Neural Plasticity Induced by Hearing Aid Use. *Frontiers in Aging Neuroscience*, *14*. https://doi.org/10.3389/fnagi.2022.884917

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2395–2405. https://doi.org/10.1121/1.1784440

Kim, S. H., Kim, E. K., & Kim, J. N. (2019). High performance DSP platform for digital hearing aid SoC with flexible noise estimation. *IET Circuits, Devices & Systems*, *13*(5), 717–722. https://doi.org/10.1049/iet-cds.2018.5374

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_jslhr-19-00419

Lee, D., Lewis, J. D., Johnstone, P. M., & Plyler, P. N. (2021). Acceptable Noise Levels and Preferred Signal-to-noise Ratios for Speech and Music. *Ear And Hearing*, *Publish Ahead of Print*. https://doi.org/10.1097/aud.0000000000001157

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Monaghan, J. J., Garcia-Lazaro, J. A., McAlpine, D., & Schaette, R. (2020). Hidden Hearing Loss Impacts the Neural Representation of Speech in Background Noise. *Current Biology*, *30*(23), 4710-4721.e4. https://doi.org/10.1016/j.cub.2020.09.046

Moore, T. M., Key, A. P., Thelen, A., & Hornsby, B. W. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia*, *106*, 371–382. https://doi.org/10.1016/j.neuropsychologia.2017.10.025

Mordor Intelligence. (2023). Hearables Market - Growth, Trends, COVID-19 Impact, and Forecasts (2023-2028). In *Research and Markets* (No. 5239695). https://www.researchandmarkets.com/reports/5239695/hearables-market-growth-trends-covid-19?utm_source=BW&utm_medium=PressRelease&utm_code=p4mlzd&utm_campaign=1498648+-+Global+Hearables+Market+(2021+to+2026)+-+Growth%2c+Trends%2c+COVID-19+Impact%2c+and+Forecasts&utm_exec=jamu273prd#product--adaptive

Murphy, B. (2021, June 2). *Listening Fatigue in College Students*. https://ttu-ir.tdl.org/handle/2346/86970

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Norcross, S., Lavoie, M., & Thibault, L. (2011, March). *A Status Report on Loudness Control Technologies and Standardization for Broadcasting*. https://crtc.gc.ca/eng/publications/reports/rp110318.htm

Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., & Ewert, S. D. (2018). Restoring Perceived Loudness for Listeners With Hearing Loss. *Ear &Amp; Hearing*, *39*(4), 664–678. https://doi.org/10.1097/aud.0000000000000521

O'Neill, E. R., Parke, M. N., Kreft, H. A., & Oxenham, A. J. (2020). Development and Validation of Sentences Without Semantic Context to Complement the Basic English Lexicon Sentences. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3847–3854. https://doi.org/10.1044/2020_jslhr-20-00174

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese–English bilinguals perceive "Rose?" vs. "Rose." *Bilingualism: Language and Cognition*, *20*(2), 367–383. https://doi.org/10.1017/s1366728915000723

Peng, Z., & Wang, L. M. (2019). Listening Effort by Native and Nonnative Listeners Due to Noise, Reverberation, and Talker Foreign Accent During English Speech Perception. *Journal of Speech Language and Hearing Research*, *62*(4), 1068–1081. https://doi.org/10.1044/2018_jslhr-h-17-0423

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual Consequences of "Hidden" Hearing Loss. *Trends in Hearing*, *18*, 233121651455062. https://doi.org/10.1177/2331216514550621

Portnuff, C., & Bell, B. (2019). Effective Use of Speech-in-Noise Testing in the Clinic. *The Hearing Journal*, *72*(5), 40. https://doi.org/10.1097/01.hj.0000559502.51932.b1

Rance, G., Saunders, K., Carew, P. J., Johansson, M., & Tan, J. (2014). The Use of Listening Devices to Ameliorate Auditory Deficit in Children with Autism. *The Journal of Pediatrics*, *164*(2), 352–357. https://doi.org/10.1016/j.jpeds.2013.09.041

Simantiraki, O., Cooke, M., & King, S. (2018). Impact of Different Speech Types on Listening Effort. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.21437/interspeech.2018-1358

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., & Cook, S. (2012). Development and Validation of the AzBio Sentence Lists. *Ear &Amp; Hearing*, *33*(1), 112–117. https://doi.org/10.1097/aud.0b013e31822c2549

Streijl, R. C., Winkler, S., & Hands, D. S. (2016). Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*, *22*, 213–227. https://doi.org/10.1007/s00530-014-0446-1

Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2010). A short-time objective intelligibility measure for time-frequency weighted noisy speech. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. https://doi.org/10.1109/icassp.2010.5495701

Uchida, Y., Sugiura, S., Nishita, Y., Saji, N., Sone, M., & Ueda, H. (2019). Age-related hearing loss and cognitive decline — The potential mechanisms linking the two. *Auris Nasus Larynx*, *46*(1), 1–9. https://doi.org/10.1016/j.anl.2018.08.010

Varenina, L. (2018). Phonosemantics of English consonants and their combinations: students' perception. *XLinguae*, *11*(2), 508–517. https://doi.org/10.18355/xl.2018.11.02.41

Williams, N. (2021). Working through COVID-19: 'Zoom' gloom and 'Zoom' fatigue. *Occupational Medicine*, *71*(3), 164–164. https://doi.org/10.1093/occmed/kqab041

Wilson, R. H., McArdle, R., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech Language and Hearing Research*, *50*(4), 844–856. https://doi.org/10.1044/1092-4388(2007/059

Wilson, R. H., & Watts, K. L. (2012). The Words-in-Noise Test (WIN), List 3: A Practice List. *Journal of the American Academy of Audiology*, *23*(02), 092–096. https://doi.org/10.3766/jaaa.23.2.3

Woods, D. D., Yund, E. W., Herron, T. J., & Cruadhlaoich, M. a. U. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *Journal* *of the Acoustical Society of America*, *127*(3), 1609–1623. https://doi.org/10.1121/1.3293005

World Health Organization: WHO. (2023). Deafness and hearing loss. *Www.Who.Int*. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss#:~:text=The%20prevalence%20of%20hearing%20loss,affected%20by%20disabling%20hearing%20loss.

Zhang, Y., Lehmann, A., & Deroche, M. (2021). Disentangling listening effort and memory load beyond behavioural evidence: Pupillary response to listening effort during a concurrent memory task. *PLOS ONE*, *16*(3), e0233251. https://doi.org/10.1371/journal.pone.0233251

Appendix A: Annotated Bibliography

Measuring Intelligibility

- Factors Influencing Listening: Inside and Outside the Head (1984)
    - Mentions that the speaker must be aware of the properties of the room in which they are communicating to ensure appropriate loudness
- Mean Opinion Score (MOS) Revisited: Methods and Applications, Limitations and Alternatives (2016)
    - Compares subjective MOS with the measurement methods described in the ITU standards
    - Recommends MOS be used carefully within context, ie it will never be a perfect evaluation as one needs to make sure the right questions are being asked in the right environment
    - States the focus of quality design should be on the quality of the experience rather than the quality of the signal so the goal is to tie back to what the user is experiencing (hence the benefit of subjective MOS)
- International Telecommunication Union standards for subjective test methodology for assessing speech intelligibility (2016)
    - These standards detail exact procedures for running a subjective MOS evaluation
    - The test words from some of my test sets came from the pairs that are proposed in these standards
- A Short-Time Objective Intelligibility Measure for Time-Frequency Weighted Noisy Speech (2010)
    - Details how STOI was created as a form of measurement
    - STOI takes the original clean signal and compares it segment by segment to the new signal to see how similar they are
- Patterns of English Phoneme Confusions from Native and Non-native Listeners (2004)
    - Confusion matrix that displays which phonemes are often confused for one another in both native and non-native speakers
    - Used to inform the first iteration of my study to determine which phonemes to focus on
    - Also used to confirm focus on consonants rather than vowels as vowels in the English language are a lot more flexible than the distinction between consonants
- An Analysis of Perceptual Confusion Among Some English Consonants (1955)
    - Confusion matrix with oft confused phonemes
    - Earlier version of the Cutler (2004) study essentially but not looking at how non-native speakers compare

Speech-in-Noise Testing

- Effective Use of Speech-In-Noise Testing in the Clinic (2019)
    - Useful in determining the functional capacity of a patient's hearing system
    - States that SIN tests are used infrequently
- An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss (2007)
    - Study compared the results of people with normal (pure tone) hearing and people with a known hearing loss in a variety of SIN testing scenarios
    - Concluded that QuickSIN and WIN are more sensitive than BKB-SIN and HINT
- Development of a Quick Speech-in-Noise Test for Measuring Signal-to-Noise Ratio Loss in Normal-Hearing and Hearing-Impaired Listeners (2004)
    - Details for the QuickSIN test and how it was developed
    - The test takes about one minute to do and was based on a previous SIN test that was longer
- Development of the Hearing in Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise (1994)
    - Sentences used in test are phonemically balanced
    - Created to determine the speech reception thresholds for people with hearing loss so it was known where the challenges arose in conversation
- Development and Validation of the AzBio Sentence List (2012)
    - The goal of the study was to create new sentence lists that could be used in SIN testing
    - The testing is targeted at cochlear implant users as well as potential cochlear implant candidates
    - The study to validate the test put the sentences developed through a 5 channel cochlear implant processor before giving the results to normal hearing listeners to determine the success of the intelligibility
- The Words-in-Noise Test, List 3, a Practice List (2005)
    - Type of SIN test that evaluates monosyllabic words with multitalker babble in the background
    - The SNR is varied to get a more fine tuned idea of how well the participant can hear in different noisy situations

Psychoacoustics

- Phonetic Feature Encoding in Human Superior Temporal Gyrus (2014)
    - Phoneme charts based on where in the brain these specific phonemes activate
    - Looking specifically at how certain acoustic features from speech are encoded into the brain such as how consonants are stored vs vowels

- Functional Characterization of Human's Heschl's Gyrus in Response to Natural Speech (2021)
    - Looking at how the Heschl's gyrus is responsible for phonetic encoding specifically related to how this part of the brain is responsible for auditory processing
    - Used EEG to determine how the brain responds to certain stimuli
    - Determine the Heschl's gyrus is instrumental in converting key acoustic features to gain meaning out of speech
- Hierarchical Encoding of Attended Auditory Objects in Multi-Talker Speech Perception (2019)
    - Study looked at how the brain could focus on one speaker in a multi speaker situation
    - An invasive study was performed and determined that the primary auditory cortex was unaffected by attention but further processes occurred in the non-primary auditory cortex that could be predicted through a linear model
- Chronic Traffic Noise Stress Accelerates Brain Impairment and Cognitive Decline in Mice (2018)
    - Traffic noise exposure led to increased cognitive impairment and anxiety like symptoms in mice
    - Study aimed to further supplement other research that confirmed the need for public health measures to reduce the amount of traffic noise due to the detriment on human health
- Hearing Loss is Associated with Delayed Neural Responses to Continuous Speech (2022)
    - Study compared people with hearing loss to those without and discovered that those with hearing loss took longer to process speech
    - Seemed to be a gradual process over time as the hearing loss became more impactful on communication

Current Hearing Aid Design

- Neural Plasticity in Hearing Aid Use (2022)
    - Untreated hearing loss can result in cognitive decline
    - There are encouraging signs that the brain can start to adjust after two weeks with technological support
- Predicting Hearing Aid Satisfaction in Adults: A Systematic Review of Speech-in-noise Tests and Other Behavioral Measures (2021)
    - Initial surveys indicate that the severity of one's hearing loss does not correspond to their satisfaction with their hearing aids, as one may predict
    - Conclusion is that patients tested with SIN tests rated higher satisfaction than those who did pure tone only

- Restoring Perceived Loudness for Listeners With Hearing Loss (2018)
    - Emphasizes the importance of normalizing perceived loudness
    - This paper further looks into specific compression algorithms that can be used to address the additional challenges that hearing aid processing faces when mitigating loudness
- Subjective loudness ratings of vehicle noise with the hearing aid fitting methods NAL-NL2 and trueLOUDNESS (2019)
    - Study that occurred after the previous 2018 one (same head researcher Dirk Oetting)
    - Two hearing aid fitting methods were compared to see which one better normalized the loud noises coming from a vehicle racing track
    - trueLOUDNESS was determined to be more effective in restoring people with hearing loss's perception of loudness to those of folks without hearing loss

Hidden Hearing Loss

- Perceptual Consequences of Hidden Hearing Loss (2014)
    - Confirms that hidden hearing loss is not able to be measured through traditional audiometric testing
    - Noise exposure over lifetime can lead to increased challenges not only in hearing pure tones but in the additional sound processing like discriminating speech
- Primary Neural Degeneration in the Human Cochlea: Evidence for Hidden Hearing Loss in the Aging Ear (2019)
    - Evidence of auditory nerve connections degenerating over time, causing potential hearing challenges without being evident with a pure tone audiometry test
- Hidden Hearing Loss Impacts the Neural Representation of Speech in Background Noise (2020)
    - This paper is the best for detailing the actual pathways of the brain that are affected by hidden hearing loss
    - Found that the cause of this type of loss was from noise exposure
- Hidden Hearing Loss: Mixed Effects of Compensatory Plasticity (2020)
    - One study on hidden hearing loss found that while some neural pathways can be damaged, others might actually be better than average, seemingly to compensate
    - Confirms that this type of loss is not discernible on pure tone tests and instead can impact the ability to distinguish between multiple noise sources

Contextual Information From Sentences

- Development and Validation of Sentences Without Semantic Context to Complement the Basic English Lexicon Sentences (2020)
  - Includes a corpus of nonsense sentences to be used for the sentence tests which can be found here
  - Covers more variations of SIN tests
  - The experiment is a variation on a SIN test with new nonsense sentences that the researchers developed
- Context Effects in Sentence Comprehension (1974)
  - Distinguishes the effect that context can have in speech comprehension
  - Examples from this paper informed the direction of my study method through clarifying the importance that the placement of a word in a sentence can have on the comprehension of the phrase
- Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences (2006)
  - Explains that the structural set up of a sentence does not imply context and there are many ambiguities that arise from various expressions that intend to communicate meaning
- Building a large annotated corpus of learner English: The NUS corpus of learner English (2013)
  - This paper justifies the use of the NUCLE English word corpus that was used in the O'Neill et al. paper from 2020 that uses the corpus as a basis to creating nonsensical phrases
  - This justification further agrees with my basing the test sentences from my study off of the NUCLE corpus
- Phoneme-Order Encoding During Spoken Word Recognition: A Priming Investigation (2019)
  - Study looks at how the preceding/surrounding phonemes can impact what phoneme is perceived in the moment
  - This relates to the nonsensical test sentences in my study and justifies why I needed to add another version of sentences to the study (i.e. knowing where the target word was placed)

Listening Fatigue

- Listening fatigue in neurotypical college students (2021)
  - Perceived fatigue increased but cognitive task success unchanged after a full day of classes

    - While listening fatigue is not being looked at in my study, the prime demographic of who will be doing my study is within the same age range as the one in this study which can be relevant in terms of cognitive demand that students face after classes
- What is being measured when looking at listening fatigue? (2014)
    - Some researchers are using similar methods to look at listening fatigue but making different assumptions which leads to a variety of different definitions and understandings of the effects of listening fatigue
- The effect of auditory fatigue on reaction time in normal hearing listeners at different signal to noise ratios (2016)
    - Higher SNR = lower listening effort but didn't seem to correspond with fatigue
- Listening effort and fatigue: Are we talking about the same thing? (2013)
    - Displays the differences between effort and fatigue in the form of an infographic backed by research studies from the University of Manchester
    - This summary also proposes new words that more intuitively fit with what we would believe the definitions to be. For example: perceived listening effort as distinguished from physiological cost of listening
- Listening effort and fatigue in school-aged children with and without hearing loss (2002)
    - More effort but seems no change in subjective perceived fatigue
- Neural mechanisms of mental fatigue elicited by sustained auditory processing (2017)
    - Using measures such as EEG, ERP, etc.
    - Sustained auditory processing can illicit mental fatigue
    - Decreased brain activity over time was measured
- Disentangling listening effort and memory load beyond behavioural evidence: Pupillary response to listening effort during a concurrent memory task (2021)
    - While the methodology is too advanced for my thesis (uses eye tracking), the results and general information they pull to base the study on overlaps a lot
- Listening effort in cochlear implant users: The effect of speech intelligibility, noise reduction processing, and working memory capacity on pupil dilation response (2022)
    - Sound processing with cochlear implant users and pupillary response
    - Similar to above study but focusing more on speech intelligibility and the effects of a noise reduction algorithm
- Listening effort by native and non native speakers due to noise, reverberation, and talker foreign accent during English speech perception (2019)
    - Subjective measurement better captured impacts of alterations in speech signals on listening effort
    - Effects of listening effort based on noise, accent, and reverb

- Listening effort and perceived clarity in normal hearing children with the use of digital noise reduction (2014)
    - Compared two different noise reduction algorithms and gave tasks at various SNRs
    - Response time decreased with noise reduction algorithms (one more than the other)
- The relationship between speech cognition, behavioural listening effort, and subjective ratings (2018)
    - Participants rated their: (1) mental work, (2) desire to improve the situation, (3) tiredness, and (4) desire to give up
    - Looked at adults with known hearing loss and beamforming technology, but first half of study focuses on appropriate ways to evaluate listening effort
- Framework for understanding effortful listening (2016)
    - A panel of experts came together to create a standard approach to how "effortful" listening could be evaluated
    - The framework ties together both cognitive and physiological models in assessing listening effort
- Working through COVID-19" 'Zoom' gloom and 'Zoom' fatigue (2021)
    - Notes the cause of the fatigue being in part due to the missing facial cues that we don't get online
    - Fatigue is also caused from the unnatural soundscape and processing that all the conversations experience
- Auditory Cognition and Human Performance (2012): Book Review
    - Top researchers in the psychoacoustics field: Donald Broadbent, Colin Cherry, Anne Treisman, Reiner Plomp, Albert Bregman, Neville Moray
    - Mental workload theory: human have limited cognitive resources so the goal is to most efficiently balance the mental workload required for a task; this is affected if the brain needs to work harder for speech comprehension
    - Speech comprehension can already require more effort through difficulty with syntax, accents, unfamiliar words, etc.
    - Main point: Degraded speech signal increases cognitive effort

Functional Load

- Bridging phonological system and lexicon: Insights from a corpus study on functional load (2015)
    - Looked at tonal and non-tonal languages → including English and French
    - Determined that there's a consonantal bias across non-tonal
    - Has some references to types of qualities that go into what distinguishes a sound such as stress, tone, and segmentals

- Functional loads of pronunciation features in nonnative speakers' oral assessments (2014)
    - Looks at intelligibility and functional load as it interacts with different ways to form sound (ex. fricatives)
- High functional load inhibits phonological contrast loss: A corpus study (2013)
    - Study looked at minimal pairs in relation to functional load
    - Phoneme contrasts that have a high functional load are less likely to experience phonemic merge and thereby remain more distinct
- Functional load of fundamental frequency in the native language predicts learning and use of these cues in second-language speech segmentation (2016)
    - Looking at English, French, and Dutch and used eye tracking methods
    - Understanding how functional load of fundamental frequency intersects with second language learners and their understanding of the sound
- Bridging phonological system and lexicon: Insights from a corpus study on functional load (2015)
    - Consonantal bias across languages
    - Comparing nine different languages and their various phonemic representations

French Phonetics

- Phonetic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Quebecois French and Brazilian Portuguese (2018)
    - Nasal vowels have more contrastive structure but less variability than if it has less contrastive structure
    - Note: It was determined that French is outside the scope of this project, but it was important to gain some basic understanding as it will hopefully be used in further iterations of the study
- Assessing distinctiveness of phonological features in word recognition: Prelexical and lexical influences (2017)
    - Manner contrasts matters more than place or voicing in French
    - Place has higher functional load in nouns than voicing and manner
- Frenchville French: A case study in phonological attrition (2004)
    - There are different reasons as to why a shift in phonemic expression may occur that cannot solely be tied to functional load
    - Other reasons: cultural identity, acoustic salience, and other articulatory demands

Machine Learning

- Acoustic landmarks contain more information about the phone string than other frames for automatic speech recognition with deep neural network acoustic model (2018)
    - Rather than looking at each speech frame as identical, using landmarks in speech such as pauses and frequency jumps, one can develop a heuristic model to more efficiently detect speech
    - This is useful in distinguishing between speech and noise in terms of AH's algorithm
- It's About Time: Minimizing Hardware and Software Latencies in Speech Research with Real-Time Audio Feedback (2020)
    - 50 ms latency is when speech production begins to be impacted by the delay in speech processing
    - The processing latency must be calculated to include both hardware and software processing times as often only one of the two are considered and both can add up to be more substantial

Miscellaneous

- Nyquist sampling theorem: understanding the illusion of a spinning wheel captured with a video camera (2014)
    - Explaining the Nyquist theorem: essentially that you want double the sample rate of the highest expected frequency
    - This paper was used to inform what sampling rate should be used for the text to speech voice clips that will be used in my own study
- A status report on loudness control technologies and standardization for broadcasting (2011)
    - Canada guidelines for broadcast loudness mixing
    - This is how -24 LUFS was chosen as a value to normalize the sound files in my study
- FDA finalizes history rule enabling access to over-the-counter hearing aids for millions of Americans (2022)
    - Formal announcement when OTC hearing aids were allowed to be sold in the United States
- Hearing Loss Among World Trade Centre Firefighters and Emergency Medical Service Workers (2019)
    - The most interesting thing is they found increased hearing sensitivity which might be an indicator of hidden hearing loss
    - In general found an increased risk in developing hearing loss due to the conditions experienced on the job as an emergency worker

- An Online Headphone Screening Test Based on Dichotic Pitch (2021)
    - Test to determine whether participants are using speakers or headphones
    - Using the Huggins Pitch test alone has a 20% false positive rate
    - Useful in ensuring that participants are both using headphones (as requested) and adjusting the volume to a comfortable level before starting the test
- Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives (2008)
    - Gives a summary of how speech is formed and is then received into the brain
    - Highlights the importance of the cerebellum in speech production
- The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses (2012)
    - Lists variant situations in which someone might be trying to communicate (ex. background noise and difference in speech rate)
    - Believes that portions of the brain responsible for speech production are also responsible for speech comprehension

Appendix B: Test Sets for the Study

Beginning of Word Targets

Set1		Set2		Set3	
Tie	Die	Teen	Dean	To	Do
Zap	Sap	Zoo	Sue	Zip	Sip
Choke	Joke	Chin	Gin	Char	Jar
Peak	Beak	Pan	Ban	Pat	Bat
Game	Came	Guard	Card	Gap	Cap
Van	Fan	Vat	Fat	Veer	Fear

Middle of Word Targets

Set1		Set2		Set3	
Centre	Sender	Venting	Vending	Panter	Pander
Raising	Racing	Phases	Faces	Lazy	Lacy
Etching	Edging	Perching	Purging	Searching	Surging
Simple	Symbol	Staple	Stable	Repelling	Rebelling
Bugging	Bucking	Angle	Ankle	Plugging	Plucking
Waver	Wafer	Proving	Proofing	Divine	Define

End of Word Targets

Set1		Set2		Set3	
Mat	Mad	Cot	Cod	Pot	Pod
Eyes	Ice	Buzz	Bus	Laws	Loss
Batch	Badge	Lunch	Lunge	Rich	Ridge
Lap	Lab	Nip	Nib	Rope	Robe
Sag	Sack	Tag	Tack	League	Leak
Strive	Strife	Relieve	Relief	Calve	Calf

Appendix C: Stimuli Frequencies in COCA

Beginning of Word Target Phonemes*

Set1				Set2				Set3			
Die	106573	Tie	30026	Teen	15965	Dean	33918	To	25554050	Do	4501699
Sap	2883	Zap	1231	Zoo	10252	Sue	18603	Zip	6635	Sip	6428
Joke	**36066**	**Choke**	**4246**	Chin	15161	Gin	4185	Char	1274	Jar	9346
Peak	24736	Beak	2005	Ban	22905	Pan	26904	Pat	28493	Bat	14942
Game	311173	Came	402970	Guard	56589	Card	70441	Cap	26168	Gap	29157
Fan	48508	Van	51178	**Vat**	**1519**	**Fat**	**76374**	**Veer**	**1186**	**Fear**	**103493**

*Bolded words show that each set has one pair where there is a large inconsistency in the pair's frequency. Because there is one per set, they are equally distributed and should not have a large overall impact on the final intelligibility scores.

Middle of Word Target Phonemes

Set1				Set2				Set3			
Centre	11492	Sender	1335	Venting	1296	Vending	2436	Panter	30	Pander	829
Racing	18748	Raising	35440	Phases	6453	Faces	47544	Lacy	2014	Lazy	12589
Etching	1043	Edging	1635	Perching	224	Purging	851	Searching	24688	Surging	2216
Simple	120703	Symbol	17886	Stable	25655	Staple	4155	Repelling	359	Rebelling	614
Bucking	1025	Bugging	1475	Angle	24282	Ankle	19394	Plucking	1040	Plugging	1562
Waver	718	Wafer	973	Proofing	221	Proving	8491	Define	24870	Divine	18997

End of Word Target Phonemes*

Set1				Set2				Set3			
Mat	6279	Mad	39863	Cod	4334	Cot	2011	Pod	4314	Pot	26333
Ice	78320	Eyes	251742	Buzz	12322	Bus	49900	Loss	90774	Laws	78602
Batch	5926	Badge	6755	Lunch	51250	Lunge	1187	Ridge	15423	Rich	84507
Gap	**29157**	**Gab**	**617**	**Nip**	**1200**	**Nib**	**177**	Rope	14245	Robe	5807
Sag	1596	Sack	8432	Tack	3050	Tag	15788	League	72457	Leak	8753
Strive	6318	Strife	2661	Relief	38847	Relieve	5456	**Calf**	**4267**	**Calve**	**89**

Appendix D: Test Sentences for Future Studies

With Knowledge of Target Word Placement

Problem Phoneme at the Beginning of the Word

Set1
The next word I am going to say is die.
The next word I am going to say is tie.
The next word I am going to say is sap.
The next word I am going to say is zap.
The next word I am going to say is choke.
The next word I am going to say is joke.
The next word I am going to say is peak.
The next word I am going to say is beak.
The next word I am going to say is game.
The next word I am going to say is came.
The next word I am going to say is fan.
The next word I am going to say is van.

Set2
The next word I am going to say is teen.
The next word I am going to say is dean.
The next word I am going to say is zoo.
The next word I am going to say is sue.
The next word I am going to say is chin.
The next word I am going to say is gin.
The next word I am going to say is ban.
The next word I am going to say is pan.
The next word I am going to say is guard.
The next word I am going to say is card.
The next word I am going to say is vat.
The next word I am going to say is fat.

Set3
The next word I am going to say is to.
The next word I am going to say is do.
The next word I am going to say is zip.
The next word I am going to say is sip.
The next word I am going to say is char.
The next word I am going to say is jar.
The next word I am going to say is pat.
The next word I am going to say is bat.

The next word I am going to say is cap.
The next word I am going to say is gap.
The next word I am going to say is veer.
The next word I am going to say is fear.

Problem Phoneme in the Middle of the Word

Set1
The next word I am going to say is centre.
The next word I am going to say is sender.
The next word I am going to say is racing.
The next word I am going to say is raising.
The next word I am going to say is etching.
The next word I am going to say is edging.
The next word I am going to say is simple.
The next word I am going to say is symbol.
The next word I am going to say is bucking.
The next word I am going to say is bugging.
The next word I am going to say is waver.
The next word I am going to say is wafer.

Set2
The next word I am going to say is venting.
The next word I am going to say is vending.
The next word I am going to say is phases.
The next word I am going to say is faces.
The next word I am going to say is perching.
The next word I am going to say is purging.
The next word I am going to say is stable.
The next word I am going to say is staple.
The next word I am going to say is angle.
The next word I am going to say is ankle.
The next word I am going to say is proofing.
The next word I am going to say is proving.

Set3
The next word I am going to say is panter.
The next word I am going to say is pander.
The next word I am going to say is lacy.
The next word I am going to say is lazy.
The next word I am going to say is searching.
The next word I am going to say is surging.
The next word I am going to say is repelling.
The next word I am going to say is rebelling.

The next word I am going to say is plucking.
The next word I am going to say is plugging.
The next word I am going to say is define.
The next word I am going to say is divine.

Problem Phoneme at the End of the Word

Set1
The next word I am going to say is mat.
The next word I am going to say is mad.
The next word I am going to say is ice.
The next word I am going to say is eyes.
The next word I am going to say is batch.
The next word I am going to say is badge.
The next word I am going to say is lap.
The next word I am going to say is lab.
The next word I am going to say is sag.
The next word I am going to say is sack.
The next word I am going to say is strive.
The next word I am going to say is strife.

Set2
The next word I am going to say is cod.
The next word I am going to say is cot.
The next word I am going to say is buzz.
The next word I am going to say is bus.
The next word I am going to say is lunch.
The next word I am going to say is lunge.
The next word I am going to say is nip.
The next word I am going to say is nib.
The next word I am going to say is tack.
The next word I am going to say is tag.
The next word I am going to say is relief.
The next word I am going to say is relieve.

Set3
The next word I am going to say is pot.
The next word I am going to say is pod.
The next word I am going to say is loss.
The next word I am going to say is laws.
The next word I am going to say is ridge.
The next word I am going to say is rich.
The next word I am going to say is rope.
The next word I am going to say is robe.

The next word I am going to say is league.
The next word I am going to say is leak.
The next word I am going to say is calf.
The next word I am going to say is calve.

Without Context

These sentences were developed from the corpus created in the study *Development and Validation of Sentences Without Semantic Context to Complement the Basic English Lexicon Sentences* (O'Neill et al., 2020).

Problem Phoneme at the Beginning of the Word

Set1
His vegetables carried in a large tie.
My die drinks in the crowded school.
That doctor hates more sap.
Our places taste the zap.
The choke and fruit sing again.
A theatre planned foreign joke.
The course offers a thirsty beak.
The tired station cooks for her peak.
The driver earned from the game.
The foreign dancing came fresh.
The fan loves old cake.
The pretty van is angry.

Set2
The cold teen scored the questions.
The dean fell during their week.
They found the grape zoo brightly.
The birthday desserts sue the soup.
The chicken studies above the chin.
The artists gin the company.
The expensive ban gives the market.
A peaceful pan is exciting to visit.
The last guard broke everyone.
A large card was honest and upset.
The best vat is softly small.
The fat eats things through the shoes.

Set3
The hotel sells to the wood advice.
The people who write after do hurt.

The strangers zip on the blueberry water.
The flags sip the last team.
The funny pets char noodles.
An excited jar loves breaks.
The troubled grade is pat easily.
They wrote cake on the green bat.
The cap sells proud salt.
A helpful gap expected the plants
The babies veer over the river.
The great lady saves fear.

Problem Phoneme in the Middle of the Word

Set1
The centre kid was colourful and weak.
A child chased down the sender.
These racing months baked into a game.
My white mouse needs my raising.
Her etching singer lives clean.
She felt the grandparents edging across the oven.
Their fun truths simple first.
The nice symbol goat dried great.
The talented animals were bucking the grandmother.
A metal pepper was usually bugging sports.
The leaves waver tools together.
The sweet wafer wants glasses.

Set2
The grocery person is venting a farm.
The horrible computer was vending recently.
The stressful phases are the french bar.
The young faces are completely green.
A football lunch had a perching shirt.
The restaurant is purging the three holidays.
The stable street upset the tourist.
The teenagers took staple snacks.
A late angle bothered the house.
That terrible maid watches ankle.
A wife is proofing sweet apples.
The team is proving kind in the meal.

Set3
The large food is panter inspired today.

The three brushes pander newspaper.
The lacy sports sold monkeys.
Her support was lazy and weekly talking.
The searching mountain reads languages.
The smart bread is surging the pig.
Her adult is repelling early rain.
A snake is rebelling the unfair wedding.
Our day was plucking short questions.
A fresh sky is plugging high.
The blue juice needs to define my fight.
The divine dog patted soup.

Problem Phoneme at the End of the Word

Set1
The mat store was always relaxing.
The mad key was strange and in need.
The ice heard across the vegetables.
The fat eyes were difficult and funny.
The divorced pink batch was huge.
The ruined and sad badge is difficult.
They lap six games on the sun.
The family was assigned to the loud lab.
The daily confusing sag is not hungry.
The lesson saw the sack.
The classes strive dedicated garden.
The strife cut through the helpful news.

Set2
The simple cod is the youngest sister.
The television cot tasted tomorrow.
The buzz won our wooden workers.
Our bus buys more sky.
The lunch lives late on the fence.
The math lunge bakes suddenly.
The fruit nip five writers.
The nib played some dangerous drinks.
A far tack bought children.
The two gloves had their english tag.
The clear money was cooked with relief.
The last grandpas relieve the eggs.

Set3
The pod made the hot fly.

The tired pot was very tiny.
The army created big loss.
A bird roasts the full laws.
A ridge ate strong stories.
The rich stranger intrigued his instructor.
A rope came down the gifts.
The robe trusts the lady.
The four songs are a sick league.
A calm leak needs for flowers.
The school bag had the dry calf.
The calve starts from the football forest.