Benjamin Green
Department of Music
Faculty Advisor: Carolina Perez, B.M., M.F.A.
External Advisor: April Groulx, Au.D.

Digital Simulation of the Cochlear Implant's Signal Processing Algorithm Modeled for Speech and Investigation of Methods to Improve Musical Cognition and Recognition

Abstract

While the Cochlear Implant (CI) has been developed with a focus on drastically improving the recognition of speech in individuals with severe to profound hearing loss, current CI systems often fail to deliver sufficient auditory information for the user to fully appreciate the complexity of musical sounds (McDermott, 2004). Successful music perception demands more sophisticated CI hardware and software, which is an active area of research and development. The focus of this project was to create a simulation of the Digital Signal Processing (DSP) methods used in CIs modeled for speech recognition and then manipulate parameters from this simulation for improved transmission of musical sounds. The DSP simulation algorithm was created in Max/MSP, a visual programming language. The algorithm was manipulated at different stages of the audio signal chain in an attempt to convey the pitch and timbre of musical material as accurately as possible.

1. Introduction

1.1 Modern Cochlear Implant Systems

The CI, developed over two decades ago, started with the discovery that an electrical current could transfer sound to the brain via the stimulation of the auditory nerve. CIs work by placing a microphone attached to a digital signal processor (DSP) behind the user's pinna (the outer ear). The output signal of the DSP is transmitted to the surgically implanted components: a receiver/stimulator and an array of electrodes that stimulate auditory neurons (Figure 1.1a).

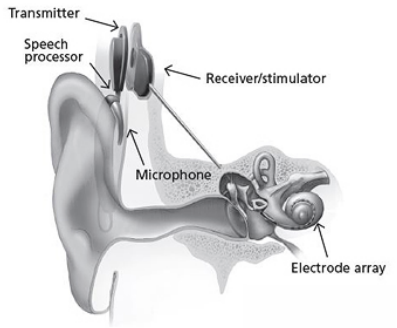

Figure 1.1a: Model of a CI (U.S. Department of Health and Human Services, 2016)

As the CI's microphone captures incoming acoustic signals, the DSP decodes the information by filtering it into different band-pass signals, much like a working inner ear but with fewer and wider bands. An amplitude envelope, also known as the temporal envelope, is calculated from each band-pass signal. The amplitude envelope derived from each band-pass signal is used to trigger the electrical pulses being transmitted by the electrodes. This process utilizes a modulating signal (the acoustic signal), a pulse carrier (a pulse train at a steady amplitude), and a pulse amplitude modulator (the amplitude envelope transformed into a pulse wave representing the extremes) (Figure 1.1b). These components are used in conjunction to determine the specific electrode to fire at a fixed pulse rate while at varying amplitudes. (Figure 1.1c)

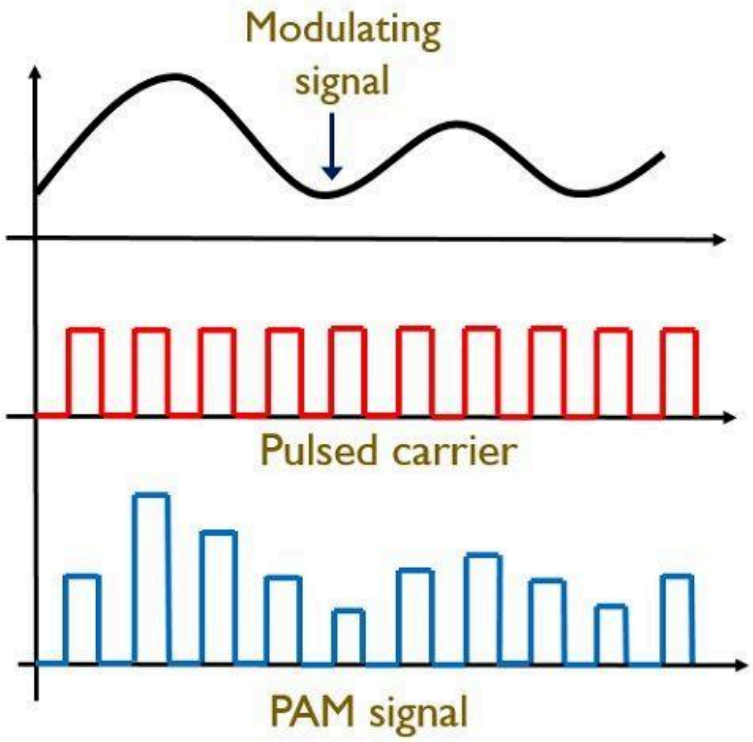

Figure 1.1b: Example of Pulse Amplitude Modulation **(**Y. R., 2021)

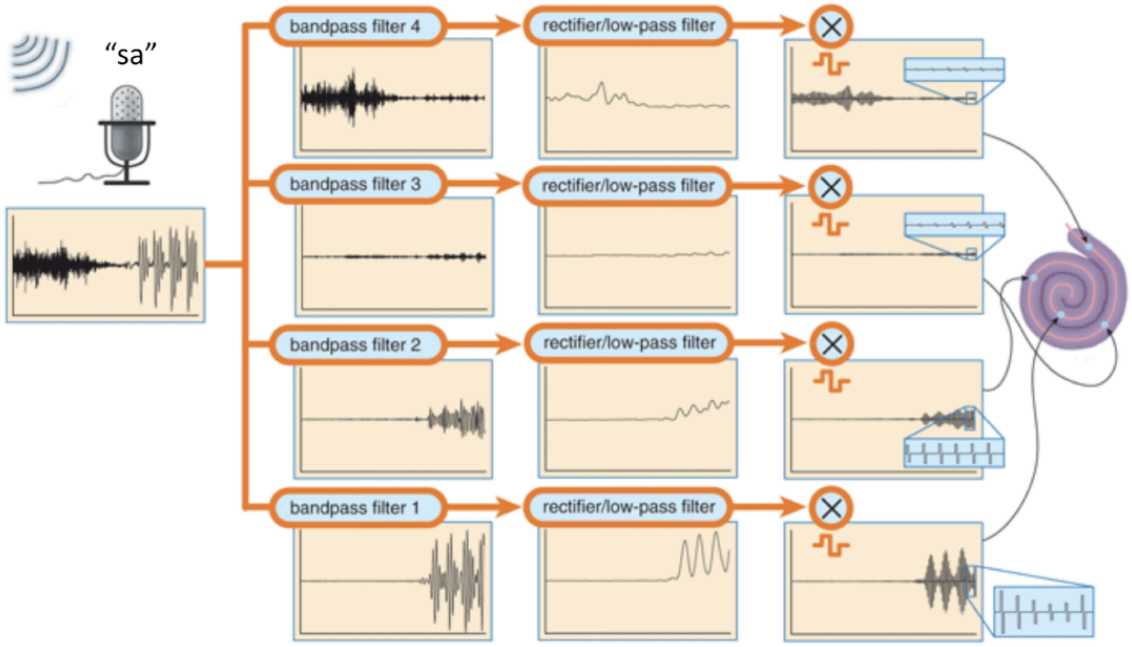

Figure 1.1c: Visual Representation of Working DSP **(**Skoe, 2021)

The CI started with as little as one electrode placed inside of the cochlea (the fluid-filled, spiral shaped cavity located in the inner ear). We now have the ability to implant up to 24 electrodes, allowing the frequency range of the device to increase drastically (Hainarosie et al., 2014). These advancements have motivated CI recipients to attempt to use their implants to regain their ability to perceive music (McDermott, 2004).

Many factors hinder the perception of music in CI users. The split of cognition between background ambience, melody, harmony, timbre, and rhythm create for an extremely difficult "one size fits all" solution. The limited number of electrodes in current CIs, combined with the coarse spectral analysis used for speech recognition, provide insufficient detail to accurately represent more complex acoustic sounds (Gfeller et al., 2019).

1.2 DSP for Speech Recognition

I created a DSP algorithm to understand how each step of the audio signal processing chain in current CIs affects the input signal. I wrote the DSP code in Max/MSP, a visual programming language used for audio processing. My algorithm was initially based on a block diagram of the Saphyr® SP sound processor, which is designed to help users understand speech.

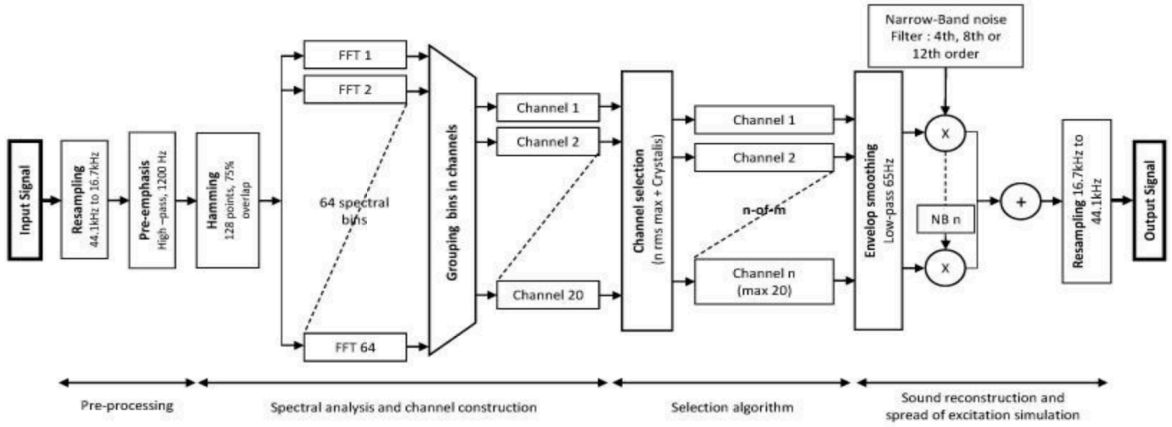

Figure 1.2a: Reference DSP (Cucis et al., 2021)

This system block diagram displays the audio signal flow, from left to right, in four stages (Figure 1.2a). The pre-processing stage reduces the possibility of errors later in the signal chain. The spectral analysis and channel construction stage transforms the signal from samples in time to the frequency and amplitude of the incoming stream of samples. The frequencies are partitioned into 64 spectral bins and then grouped into 20 non-overlapping channels. The selection algorithm determines which channels contain the highest average amplitude. Finally, the sound reconstruction and spread of excitation simulation represents frequency ranges from low to high.

In the pre-processing stage, the input signal is downsampled to reduce the amount of work required of the AD converter. A high-pass filter (pre-emphasis filter) is applied to reduce low-frequency information before the channels are constructed, which prevents unwanted low-frequency distortion.

The first step of the spectral analysis and channel construction stage is to prepare the audio samples to be converted into spectral information. A hamming window is applied to the signal to smooth out any discontinuities and minimize the amount of leakage between each non-overlapping band of frequencies. A Fast Fourier Transform (FFT) with the standard frame size of 128 points is then performed on the samples, yielding 64 usable individual frequency bands, each being equal within the range of 0 Hz to 16.7 kHz. The FFT yields 128 frequency bands within this range but half of them must be filtered out in accordance with the Nyquist

theorem. Bands 1-64 fall between 0 Hz and 8350 Hz, and the upper half of the frequency bands are filtered out in order to eliminate aliasing.

The first two and the last two frequency bands obtained from the FFT are discarded for reasons not mentioned in the reference material. The remaining 60 bands are assigned to 20 separate channels to mimic the average number of electrodes placed inside of the cochlea. Some channels contain more frequency bands than others, with more bands per channel in the higher frequencies due to the distribution method used. The distribution method of frequency bands begins by taking the lower cutoff frequency of the third band and adding it with the higher cutoff frequency of this same band. Then, the sum is divided by two to achieve the center frequency of the channel. As the frequency of both the lower cutoff frequency and the higher cutoff frequency increase, so does the center frequency. This results in a greater number of frequency bands per channel at higher frequencies.

In the selection algorithm, the average amplitude level is measured for each channel and an n-of-m rule is applied. The *n* channels with the highest average amplitude remain, while the channels that do reach the highest average amplitude, or below an average amplitude of 45 dB, are set to an amplitude value of zero.

2. Methodology

2.1 Audio Recordings

The first step in my process was to record speech and music test input signals in a semi-acoustically treated room, with a central loudspeaker at an azimuth of 0° directed at the diaphragms of the two microphones to replicate an audiometric testing environment (Spahr et al., 2012; Biever et al., 2022). The test input signals were a selection of AzBio sentences, which are sentences designed to test the performance of individuals with hearing loss. My goal was to include AzBio sentences used in audiometric testing, as well as musical examples to test my algorithm. I found out that there was no way to access the standardized AzBio sentence audio recordings without purchasing a CD and waiting on the extremely long shipping time. April Groulx, my external advisor and audiologist, informed me that it would be acceptable to record the AzBio sentences myself. I recorded my faculty advisor, Professor Carolina Perez, and myself speaking twenty AzBio sentences each at the UNCA recording studios.

I utilized AzBio sentences as speech stimuli to simulate real world audiometric testing. I also utilized pre-recorded music containing harmony, rhythm, and various timbres reproduced by the central loudspeaker. The reason for using music as a form of stimulus is to gather data obtained from the transformation of the complex signal to spectral information, and compare it to the results found from the spectral analysis of speech stimuli.

The microphones used to record the test signals were a pair of dbx RTM-A omnidirectional microphones. The microphones were placed at a similar distance to that of the human ears within a Styrofoam sculpture of a human head (Figure 2.1a). In order to simulate an audiometric listening test, the microphones were pointed directly at a central loudspeaker, five feet away and in a acoustically treated room. This environment is acoustically similar, though not identical, to the isolation booths of a nearby audiologist's office. My external advisor, audiologist April Groulx, recommended that each test each signal should be recorded at a level between 60-65 dBA, which is the sound pressure level of an average conversation (dBA is the loudness contour that is most closely related to human hearing). When recording the test signals, the loudspeaker output had to be calibrated using a sound pressure meter every time; 60-65 dBA is the range used with audiometric testing. The test input signals were recorded into the digital audio workstation Pro Tools and then transferred to Max/MSP for signal processing.

Figure 2.1a: Recording Studio at The University of North Carolina at Asheville

2.2 Simulation Algorithm

In Max/MSP, the next step was to transform the complex signal into the spectral-domain data utilizing common methods found in CI DSP algorithms. The spectral analysis performs a function on the input signal and transforms the complex signal into a group of frequency bands of deterministic sizes, each of which can be manipulated individually in intensity and resolution. The output signal of this transform yields a frequency analysis that can be viewed graphically in order to plot the spectral data. As each band of frequencies can be manipulated individually, it is possible to adjust them accordingly for a variety of audiometric and subjective reasons. This is extremely important and is used during CI mapping to adjust levels according to the user's preferences.

The pre-processing stage of my DSP begins with the input signal which immediately gets sent through a noise reduction block to reduce the high noise floor. This signal is then downsampled by a factor of 3 from 48 kS/s to a 16 kS/s signal to reduce the amount of information required of the processor. Due to the downsampling, I placed a high order low pass filter at 7.75 kHz to recreate the Nyquist frequency in order to prevent any aliasing. Pre-emphasis filtering replicates the reduced sensitivity of the human ear to low frequencies. I initially set the pre-emphasis filter at 1200 Hz as shown in the reference DSP. Later, I modified the cutoff frequency to 235 Hz to preserve vital bass frequencies in music.

In the spectral analysis and channel construction stage, the signal is then processed using an FFT with a frame size of 128, leaving us with 64 usable frequency bands to manipulate. Frequency bands 1-64 fall between 0 Hz and 8000 Hz and the latter half of the frequency bands are filtered out to eliminate aliasing. I discarded the first two and last two frequency bands following the reference DSP. The information retained in the remaining 60 bands, from 250 Hz to 7.75 kHz, covers the range of frequencies most sensitive to the human ear. These frequency bands are then grouped into separate channels and divided into narrow-band pass filtered signals. Subsequently, the envelope of this signal is smoothed in order to create a consistent amplitude envelope for the output signal and to prevent any aliasing.

In the algorithm created in Max/MSP, the sample-based signal is converted to individual bands from 0 Hz up to the set sample rate. Setting the FFT frame size, which should be a power of two, divides the signal up into a finite number of frames; larger frame sizes achieve increasingly narrow bandwidths for each individual frequency band. We calculate the frequency of each band by dividing the sample rate by the FFT frame size; this quotient gives the width of each frequency band in Hz. These frequency bands acquired from the FFT are then split into different channels and band-pass filtered before the output. After trying 4 different band-pass filtered channels, I increased the number of channels to 8 for increased spectral resolution. I tried

to achieve 12 channels, but due to hardware limitations I could not perform this task in a computationally efficient manner in Max/MSP.

The sound reconstruction and spread of excitation stage consists of the output of each band-pass filtered channel, which correspond to the frequency regions where each electrode would be placed. The band-pass filtered channels have a filter placed at 400 Hz, 700 Hz, 1100 Hz, 1600 Hz, 2100 Hz, 3500 Hz, 4500 Hz, and 7000 Hz. This allows for a small range of frequencies to be present in each channel's bandwidth. The channels are then summed together before being upsampled back to the original sampling frequency of 16 kS/s to 48 kS/s. This signal is then sent to the DA converter to be resynthesized via headphones or loudspeakers (Figure 2.2a).

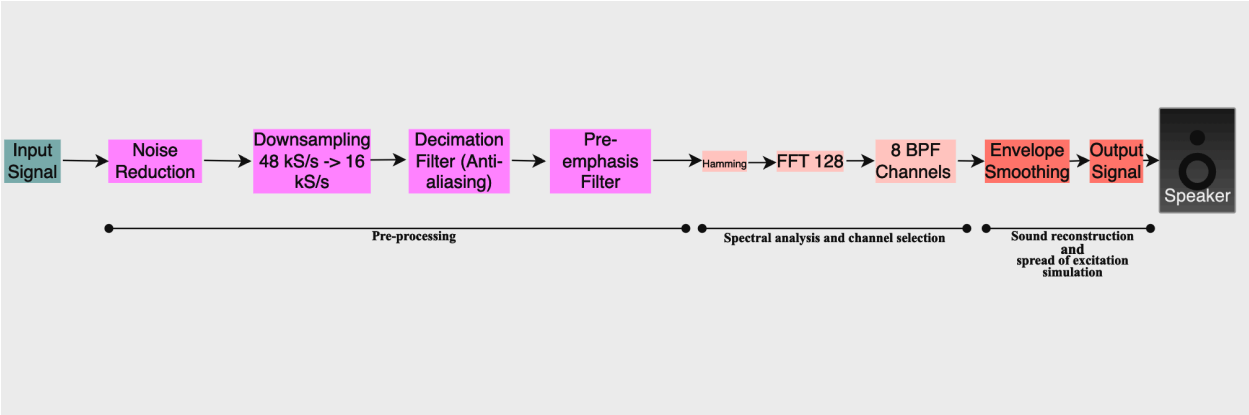

Figure 2.2a: Max/MSP Signal Flow

Vocoding can effectively simulate the CI because the noise vocoder degrades temporal fine structures, much like the CI, by using a group of band-pass filters to slice the signal into a group of independent narrow frequency bands and then modulate the amplitude of these frequency bands. The amplitude envelope of this signal then modulates the amplitude of various oscillators, which have a frequency corresponding to the center frequency of the band-pass filtered signal.

After programming the modulated oscillators in my own DSP algorithm, I found that modulating these oscillators would not be effective for this demonstration due to the computational efficiency of the algorithm. The program would often freeze and produce distortion and other artifacts at the output. The output of my DSP model therefore does not represent the relationship between implanted electrodes and the stimulation of auditory neurons. The resulting acoustic signal heard at the output of my DSP model simply allows listeners to hear the effect of different stages of the DSP on the input signal.

3. Results

3.1 Uses and Limitations of Simulation DSP

My CI DSP algorithm has been completed in its most basic form in the Max/MSP programming environment. At its input stage, the algorithm accepts audio recordings that simulate the sound picked up by a CI's microphone. The output of the algorithm is the digitally processed signal before it is converted to pulses fired by the electrodes implanted inside the cochlea. While my DSP algorithm demonstrates the ways in which the digital audio processing affects the signal picked up by the microphone, it does not simulate the human ear for an approximate representation of what a CI patient would hear.

Using my Max/MSP algorithm, I was able to create examples that demonstrate why it is difficult to perceive the combination of speech and music with acceptable quality. Pitch and timbre are not reproduced clearly in my DSP simulation, but rhythm is perceived accurately. With FFTs, smaller frame sizes capture smaller slices of the signal, which increases the timing resolution. Alternatively, larger frame sizes take a longer period of time to calculate, resulting in a decrease in timing resolution, but an increase in frequency resolution. Frequency resolution is poor in this simulation due to the bandwidth of 125 Hz per bin. This degradation of the signal dramatically affects the cognition of music. Increasing the frame size would result in an increase in pitch perception, but at the cost of important timing information required for the CI user to discern what occurs in their environment in real-time. For this simulation it was also necessary to keep the frame size low, as a larger frame size would require greater computational power to process the FFT. Further research, improved processing capabilities, and advocacy for this cause could lead to further improvements in music cognition and recognition.

3.2 Software and Programming Limitations

As I programmed the DSP algorithm in Max/MSP, altering the parameters of each component in the digital signal processing chain, I came across some software limitations. One of the limitations in Max/MSP is the FFT function lacking the ability to utilize one instance for all frequency band outputs in a channel based format. Another limitation is the large amount of processing power each FFT requires. There are other minor limitations of the software but these in particular created the biggest challenge. This in turn made the algorithm extremely inefficient and difficult to work with.

After creating the DSP algorithm that follows the fundamental functions of modern CIs, with the exception of the modulated oscillators at the center frequency of each frequency band, I ran into various limitations due to my own programming knowledge and the limitations of the Max/MSP software. With a better understanding of programming, this algorithm could be more computationally efficient and would be able to run on much less capable hardware. It is critical

that the DSP algorithm is as efficient as possible due to the hardware limitations of the physical devices used to deliver the stimuli to the electrodes in a functioning CI.

I will continue to make improvements as I gain knowledge of more complex programming environments, such as C and C++. These languages will extend the abilities of my current algorithm created in Max/MSP to reach a wider audience in order to get more audio engineers involved in the pursuit of increasing the cognition and recognition of music for CI recipients. This will in turn assist CI users with regaining their enjoyment of music.

4. Conclusion

My knowledge of the CI and of the current state of its DSP algorithms has progressed immensely. With the help of an audiologist and implementing acquired knowledge in the field of audio engineering, I was able to create a simulation of modern CI DSP algorithms. This project has allowed me to understand potential development strategies and formulate a set of hypotheses for further research and contributions to the fields of audiology and neuroscience.

The worlds of audiology and audio engineering form a key partnership, as the perception of sound is extremely important to both audio engineers and audiologists. The results of this research will contribute to the continuing study of CI signal processing by myself and others. As I personally suffer from severe hearing loss, my passion to help the progression of audible perception for CI users is unrelenting. My compassion for others suffering from hearing loss has led me to this great area of research. I will continue my quest to improve music cognition and recognition, rather than the recognizability of speech alone.

Works Cited

Biever, Allison, et al. “Evolution of the candidacy requirements and patient Perioperative Assessment Protocols for cochlear implantation.” *The Journal of the Acoustical Society of America*, vol. 152, no. 6, 1 Dec. 2022, pp. 3346–3359, https://doi.org/10.1121/10.0016446.

“Cochlear Implants.” *National Institute of Deafness and Other Communication Disorders*, U.S. Department of Health and Human Services, www.nidcd.nih.gov/health/cochlear-implants. Accessed 10 Mar. 2023.

Cucis, Pierre-Antoine, et al. “Word recognition and frequency selectivity in cochlear implant simulation: Effect of channel interaction.” *Journal of Clinical Medicine*, vol. 10, no. 4, 10 Feb. 2021, p. 679, https://doi.org/10.3390/jcm10040679.

Gfeller, Kate. “Adult Cochlear Implant Recipients’ Perspectives on Experiences with Music in Everyday Life: A Multifaceted and Dynamic Phenomenon.” *Frontiers in Neuroscience*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/31824240/. Accessed 13 Mar. 2023.

Hainarosie, M, et al. “The Evolution of Cochlear Implant Technology and Its Clinical Relevance.” *Journal of Medicine and Life*, U.S. National Library of Medicine, 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4391344/.

McDermott, HJ. “Music Perception with Cochlear Implants: A Review.” *Trends in Amplification*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/15497033/. Accessed 10 Mar. 2023.

Roshni, Y. “Difference between Pam, PWM and PPM (with Comparison Chart).” *Circuit Globe*, 18 Feb. 2021, circuitglobe.com/difference-between-pam-pwm-and-ppm.html.

Skoe, Erika. “Psychoacoustics Loudness.” Psychoacoustics, University of Connecticut, March 2021, Storrs.

Spahr, Anthony J., et al. “Development and validation of the AZBIO sentence lists.” *Ear & Hearing*, vol. 33, no. 1, Jan. 2012, pp. 112–117, https://doi.org/10.1097/aud.0b013e31822c2549.