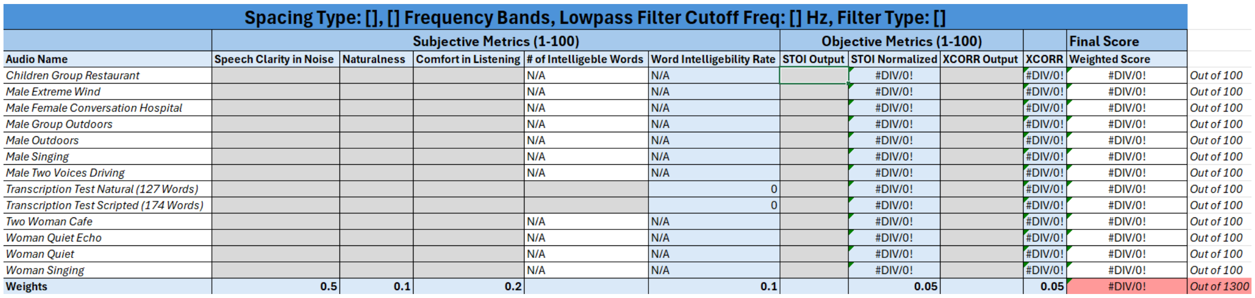
MTE252 Project - Phase 3 Report

Ethan Catz, Griffin Wilson, Kabir Raval, Oliver Chamoun

Evaluation Procedure Explained



Spacing Type: [], [] Frequency Bands, Lowpass Filter Cutoff Freq: [] Hz, Filter Type: []											
	Subjective Metrics (1-100)					Objective Metrics (1-100)				Final Score	
Audio Name	Speech Clarity in Noise	Naturalness	Comfort in Listening	# of Intelligeble Words	Word Intelligebility Rate	STOI Output	STOI Normalized	XCORR Output	XCORR	Weighted Score	
Children Group Restaurant				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Extreme Wind				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Female Conversation Hospital				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Group Outdoors				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Outdoors				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Singing				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Male Two Voices Driving				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Transcription Test Natural (127 Words)					0		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Transcription Test Scripted (174 Words)					0		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Two Woman Cafe				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Woman Quiet Echo				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Woman Quiet				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Woman Singing				N/A	N/A		#DIV/0!		#DIV/0!	#DIV/0!	*Out of 100*
Weights	**0.5**	**0.1**	**0.2**		**0.1**		**0.05**		**0.05**	#DIV/0!	*Out of 1300*

Figure 1: Evaluation Matrix

The evaluation scheme in the above figure assesses audio signals based on both subjective and objective metrics, with different weights assigned to each metric.

Subjective Metrics

Subjective metrics involve the perception of listeners and are rated on a scale from 1 to 100. In this scheme, the subjective metrics include:

- Speech Clarity in Noise, which measures how clearly the listener can understand the speaker despite background noise;
  - 0 – Pure noise/No perceivable human voice
  - 50 – Mildly understandable/Takes a fair amount of focus to understand
  - 100 – No extra effort needed to understand the speech/Perfectly understandable
- Naturalness, which assesses how natural and realistic the voice sounds;
  - 0 - Robotic/AI Generated/Crunchy
  - 50 – Phone Call with Dodgy Connection
  - 100 – Perfectly Natural Human Voice
- Comfort in Listening, which evaluates how comfortable the audio is to listen to, particularly over an extended period;
  - 0 – Very uncomfortable to listen/Would not listen to willing
  - 50 – Mildly comfortable/Not ideal but listenable
  - 100 – Perfectly comfortable to listen to for an extended period of time
- Number of Intelligible Words, which counts the words that are clearly understood by listeners; and Word Intelligibility Rate (WIR), which replaces the original Word Error Rate (WER) and measures how well words are understood by the listener.
  - The number of words that are intelligible divided by the number of total words as a percentage is the WIR.

The subjective scores are combined using weights:

- Speech Clarity in Noise has a weight of 0.5,
- Naturalness is weighted at 0.1,
- Comfort in Listening at 0.2,
- Word Intelligibility Rate at 0.1.

These weights indicate the relative importance of each metric in determining the final score, with Speech Clarity in Noise being the most significant at 50%.

Weightage Note: For audios that do not use the “Word Intelligibility Rate” metric, the 10% weightage for that metric is move to “Speech Clarity in Noise”, thus giving it a 60% weightage.

Sample formula for calculating final score seen in the right-most column:

```
=IF(E<>="N/A",B<>*(B$17+0.1)+C<>*C$17+D<>*D$17+H<>*H$17+J<>*J$17,B<>*B$17+C<>*C$17+D<>*D$17+F<>*F$17+H<>*H$17+J<>*J$17)
```

-> where <> is the row number

Objective Metrics

Objective metrics use algorithmic methods to quantify audio quality.

In the evaluation, STOI (Short-Time Objective Intelligibility) is used to measure speech intelligibility objectively, and its value is converted to a percentage to fit within the 1-100 scale. A value below 0.4 indicates unintelligibility.

XCORR (Cross-Correlation) is used to evaluate the similarity between the original and processed signals, providing an indication of how much information is retained after processing. The weights for the objective metrics are 0.05 each for STOI and XCORR.

Final Score

The Final Score is a weighted combination of both subjective and objective metrics, using the weights mentioned above. This weighted score is calculated to provide a comprehensive evaluation of the audio quality and intelligibility. The value is out of 100 and combines contributions from subjective listening tests and objective algorithmic assessments. In the context provided, the word intelligibility rate has replaced the original Word Error Rate metric, indicating a shift towards directly evaluating listener comprehension rather than focusing on errors.

Iterative Improvements on Cochlear Implant Signal Processing Design

Initial Issue and Problem Statement:

The initial problem we encountered was when we set the number of bandpass channels to a value greater than 13, which resulted in an output that did not contain lower frequencies. The output signal appeared to lack energy in lower frequency bands, resulting in a significantly compromised quality for the audio. The goal was to produce a signal that retained the intelligibility of the original input, even with different numbers of filter channels.

Below, we document the different iterations, changes made, and the corresponding results that allowed us to achieve improved output quality for both low and high values of N (number of channels).

Step 1: Diagnosis of the Issue with Logarithmic Frequency Spacing

- Problem Observed: When increasing the number of channels beyond 13, the output lacked the expected low frequencies. It showed only a few peaks without the desired full-band energy distribution.
- Analysis: This was likely due to how the center frequencies were being spaced logarithmically across the frequency band. Logarithmic spacing tends to favor higher frequencies with tighter spacing and could result in insufficient representation for low frequencies when the channel count is high.
- Initial Spacing Method: The original code used pure logarithmic spacing, which resulted in very tight filter bands for higher frequencies while relatively fewer channels covered the lower frequencies.

Step 2: Iterating on Frequency Spacing – Log. vs Linear Distribution

- Attempted Solution: To improve the representation of low frequencies, we introduced a hybrid frequency spacing method that combined linear and logarithmic spacing.

When the number of channels (N) was less than or equal to 30, we applied linear spacing between 100 Hz and 8 kHz. This ensured that each channel had equal representation, which is better suited for low channel counts where uniformity is crucial.

When N was greater than 30, we applied a hybrid method where the frequency bands were distributed using 60% linear spacing in the lower frequencies and 40% logarithmic spacing in the higher frequencies.

This provided better representation for both low and high frequencies, especially when a larger number of channels was used.

Outcome: The output was now well balanced across all frequency bands, which significantly improved the audio quality, especially for high channel counts. Lower frequencies were now represented much better.

Step 3: Improving Filter Design - Addressing Instabilities

- Problem Observed: We also observed that for high channel counts, some filters produced unstable outputs, which manifested as unexpected spikes or artifacts in the plots.
- Initial Filter Design: The initial design used 4th-order Butterworth filters for each frequency band.
- Solution: To mitigate instabilities, we:
    - Reduced the filter order from 4 to 2, which made the filters less sharp but increased stability. Higher-order filters can have steeper roll-offs but may be prone to numerical instabilities, especially when there are many closely spaced channels.
    - Implemented safeguards to ensure that filter cutoff frequencies (f_low and f_high) were within valid bounds (i.e., greater than 0 and below Nyquist frequency).

Outcome: Reducing the filter order stabilized the filtering process. This change reduced artifacts in the signal and made the filtering stage more reliable, especially when working with a high number of channels.

Step 4: Managing Overlapping Freq. Bands and Adjusting Filter Transition

- Problem Observed: The original filters did not transition smoothly between adjacent channels, causing some loss of audio continuity, especially at the boundaries between frequency bands.
- Implemented Solution:
    - We used the geometric mean of adjacent center frequencies to calculate the cutoff points for each band. This created smoother transitions between filters and ensured better overlap between adjacent frequency bands.
    - Manual Envelope Extraction: We also adjusted the envelope extraction step by increasing the window size of the FIR lowpass filter (moving average filter). This provided a smoother envelope for each channel, which improved the perceptual quality of the output.

Outcome: Using a geometric mean for the band transitions and increasing the window size for envelope extraction provided smoother, more natural outputs. The transition between different frequency bands was now less abrupt, and the perceived quality improved.

Step 5: Filter Type - IIR (Butterworth) Filters and Stability Improvements

Discussion of Filter Type:

- We used Butterworth filters due to their maximally flat frequency response in the passband, which avoids ripples and provides a natural-sounding output.
- IIR filters (Infinite Impulse Response) were chosen for their efficiency and simplicity. However, these filters require careful design, especially at high orders, to avoid instability.
- We considered FIR filters (Finite Impulse Response) but decided against them due to their need for a much larger number of coefficients to achieve similar characteristics, which could increase computational complexity.

Outcome: By keeping the Butterworth IIR filters with reduced order and using a geometric transition, we balanced performance and computational efficiency while achieving satisfactory output quality.

Step 6: Adaptive Strategy for Low and High Channel Counts

- Problem Observed: Even with improved filters, quality could be inconsistent when drastically switching between low and high numbers of channels.
- Implemented Adaptive Strategy:
    - We implemented an adaptive frequency allocation strategy:
    - Linear Spacing for Low Channel Counts: When N was low (e.g., below 30), we exclusively used linear spacing. This ensured equal distribution across the entire

frequency band, which is crucial when fewer bands are available to avoid any significant gaps.

  - Hybrid Spacing for High Channel Counts: When N was high (e.g., above 30), we mixed linear and logarithmic spacing to ensure better representation at both low and high frequencies.

Outcome: This adaptive strategy provided a consistent output across different channel configurations. The output quality improved significantly when using both fewer and more numerous channels, ensuring that each range of frequencies was represented as effectively as possible.

Final Testing and Verification

- Testing for Different Channel Numbers: We tested the implementation for different values of N, such as 16, 22, and 100. We observed:
- At 16 channels: The output contained both low and high frequencies, but the clarity was not as good as desired due to fewer channels. However, it was better than the original attempt with logarithmic-only spacing.
- At 22 channels: The quality was initially poor due to too tight of a transition and inadequate overlap. However, the improvements with linear spacing and smoother filter transitions provided more clarity in the output.
- At 100 channels: The hybrid approach provided excellent quality, with clear representation across all frequency bands.

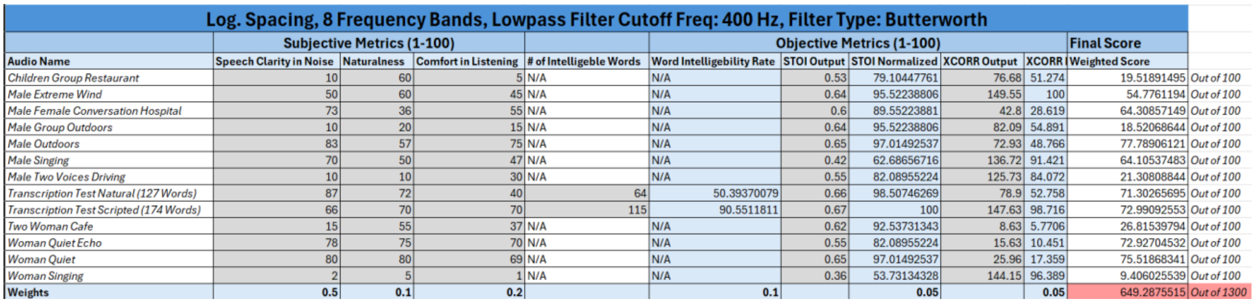
Summary of Key Changes and Iterations:

1) Logarithmic vs. Linear Frequency Spacing:
    a. Logarithmic spacing initially led to poor representation of low frequencies.
    b. Linear spacing or a hybrid approach ensured better frequency coverage for all channel configurations.
2) Filter Design Adjustments:
    a. Reduced filter order to stabilize the signal processing.
    b. Used geometric mean for band transition points to create smooth overlap between adjacent filters.
3) Adaptive Strategy for Spacing:
    a. Applied linear spacing for low channel counts (to distribute energy evenly).
    b. Applied hybrid spacing for higher channel counts (to maintain balance across the spectrum).
4) Lowpass Envelope Extraction:
    a. Used a manual FIR lowpass filter for extracting envelopes, with a larger window size to smooth the envelope signals.
    b. This helped in obtaining better-quality modulation, reducing harsh transitions.
5) Evaluation:
    a. Implemented STOI (Short-Time Objective Intelligibility) and cross-correlation to assess the quality of the processed audio output relative to the original input.

b. Adjusted the design based on these metrics to achieve a better balance of intelligibility and quality.

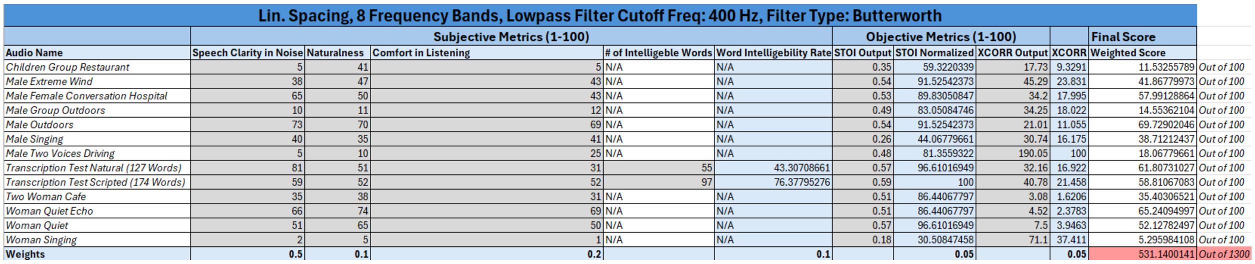
The iterative process documented here shows the importance of adapting filter parameters, frequency spacing, and processing strategies based on specific conditions (e.g., number of channels). By balancing complexity, stability, and quality through each iteration, we achieved an output that preserves both the fidelity and intelligibility of the original audio input, especially under varying numbers of channels.

Sample of Filled Evaluation Charts



Log. Spacing, 8 Frequency Bands, Lowpass Filter Cutoff Freq: 400 Hz, Filter Type: Butterworth											
	Subjective Metrics (1-100)				Objective Metrics (1-100)					Final Score	
Audio Name	Speech Clarity in Noise	Naturalness	Comfort in Listening	# of Intelligeble Words	Word Intelligebility Rate	STOI Output	STOI Normalized	XCORR Output	XCORR I	Weighted Score	
Children Group Restaurant	10	60	5	N/A	N/A	0.53	79.10447761	76.68	51.274	19.51891495	Out of 100
Male Extreme Wind	50	60	45	N/A	N/A	0.64	95.52238806	149.55	100	54.7761194	Out of 100
Male Female Conversation Hospital	73	36	55	N/A	N/A	0.6	89.55223881	42.8	28.619	64.30857149	Out of 100
Male Group Outdoors	10	20	15	N/A	N/A	0.64	95.52238806	82.09	54.891	18.52068644	Out of 100
Male Outdoors	83	57	75	N/A	N/A	0.65	97.01492537	72.93	48.766	77.78906121	Out of 100
Male Singing	70	50	47	N/A	N/A	0.42	62.68656716	136.72	91.421	64.10537483	Out of 100
Male Two Voices Driving	10	10	30	N/A	N/A	0.55	82.08955224	125.73	84.072	21.30808844	Out of 100
Transcription Test Natural (127 Words)	87	72	40	64	50.39370079	0.66	98.50746269	78.9	52.758	71.30265695	Out of 100
Transcription Test Scripted (174 Words)	66	70	70	115	90.5511811	0.67	100	147.63	98.716	72.99092553	Out of 100
Two Woman Cafe	15	55	37	N/A	N/A	0.62	92.53731343	8.63	5.7706	26.81539794	Out of 100
Woman Quiet Echo	78	75	70	N/A	N/A	0.55	82.08955224	15.63	10.451	72.92704532	Out of 100
Woman Quiet	80	80	69	N/A	N/A	0.65	97.01492537	25.96	17.359	75.51868341	Out of 100
Woman Singing	2	5	1	N/A	N/A	0.36	53.73134328	144.15	96.389	9.406025539	Out of 100
Weights	0.5	0.1	0.2		0.1		0.05		0.05	649.2875515	Out of 1300

Figure 2: Sample of Filled Evaluation Chart for Log Spacing



Lin. Spacing, 8 Frequency Bands, Lowpass Filter Cutoff Freq: 400 Hz, Filter Type: Butterworth											
	Subjective Metrics (1-100)					Objective Metrics (1-100)				Final Score	
Audio Name	Speech Clarity in Noise	Naturalness	Comfort in Listening	# of Intelligeble Words	Word Intelligebility Rate	STOI Output	STOI Normalized	XCORR Output	XCORR	Weighted Score	
Children Group Restaurant	5	41	5	N/A	N/A	0.35	59.3220339	17.73	9.3291	11.53255789	Out of 100
Male Extreme Wind	38	47	43	N/A	N/A	0.54	91.52542373	45.29	23.831	41.86779973	Out of 100
Male Female Conversation Hospital	65	50	43	N/A	N/A	0.53	89.83050847	34.2	17.995	57.99128864	Out of 100
Male Group Outdoors	10	11	12	N/A	N/A	0.49	83.05084746	34.25	18.022	14.55362104	Out of 100
Male Outdoors	73	70	69	N/A	N/A	0.54	91.52542373	21.01	11.055	69.72902046	Out of 100
Male Singing	40	35	41	N/A	N/A	0.26	44.06779661	30.74	16.175	38.71212437	Out of 100
Male Two Voices Driving	5	10	25	N/A	N/A	0.48	81.3559322	190.05	100	18.06779661	Out of 100
Transcription Test Natural (127 Words)	81	51	31	55	43.30708661	0.57	96.61016949	32.16	16.922	61.80731027	Out of 100
Transcription Test Scripted (174 Words)	59	52	52	97	76.37795276	0.59	100	40.78	21.458	58.81067083	Out of 100
Two Woman Cafe	35	38	31	N/A	N/A	0.51	86.44067797	3.08	1.6206	35.40306521	Out of 100
Woman Quiet Echo	66	74	69	N/A	N/A	0.51	86.44067797	4.52	2.3783	65.24094997	Out of 100
Woman Quiet	51	65	50	N/A	N/A	0.57	96.61016949	7.5	3.9463	52.12782497	Out of 100
Woman Singing	2	5	1	N/A	N/A	0.18	30.50847458	71.1	37.411	5.295984108	Out of 100
Weights	0.5	0.1	0.2		0.1		0.05		0.05	531.1400141	Out of 1300

Figure 3: Sample of Filled Evaluation for Lin. Spacing

Recommendation

We recommend using a Butterworth filter with a 400 Hz cutoff, logarithmic spacing for channel distribution, and 13 passbands for optimal performance. This setup provided clear, natural-sounding output with good intelligibility across samples. The Butterworth filter's flat frequency response avoids ripples, while logarithmic spacing ensures balanced frequency representation. The choice of 13 channels balances audio quality with hardware efficiency.

Through iterative testing, we found that simple logarithmic spacing with 8 to 13 channels produced the best results without excessive processing. However, more passbands are better and so using 13 is the best option in our case.

This configuration offers the best mix of clarity, naturalness, and efficiency, making it suitable for cochlear implant signal processing.

Appendix

Project GitHub Repo

The below GitHub repository includes all files related to this project, including all code and audio files. There are several versions of the code for phase 3. These versions are different iterations used along the development process described in the "Iterative Improvements on Cochlear Implant Signal Processing Design" section.

https://github.com/OliverChamoun/CochlearImplantProject.git

Audio Scripts for Longer Audios

Below are the scripts that were used to create the longer audios that were used in transcription testing. One was pre-scripted and the other was a natural off-the-cuff conversation. Both scripts had their word counts documented so that the word intelligibility rate could alter be calculated.

Scripted Conversation [174 Words]:

Speaker	Line
Person A	Hey, did you hear about the storm last night?
Person B	Oh, definitely. The wind was howling like crazy! I thought the windows were going to break.
Person A	Same here. I actually lost power for a couple of hours. What about you?
Person B	We got lucky—just a lot of flickering. But my dog was terrified. He kept hiding under the table.
Person A	Poor thing. My cat was surprisingly calm. Usually, she panics over the tiniest noise.
Person B	Maybe she just didn't care this time. Animals are weird like that.
Person A	They really are. Anyway, did you still want to go hiking this weekend, or are we calling it off?
Person B	I think we should still go! The weather's supposed to clear up by Saturday. Plus, I need the exercise.
Person A	True, me too. Alright, let's plan on it—same place, ten o'clock?
Person B	Works for me. Just don't forget your hiking boots this time!
Person A	Ugh, don't remind me! I won't make that mistake again; I promise.
Person B	Alright then, see you Saturday!
Person A	See you!

Natural Conversation [127 Words]:

Person A	Hey bro I heard you're on Duolingo.
Person B	Yea actually I hopped on Duolingo to learn some new languages.
Person A	What languages are you learning?
Person B	I actually started off learning Chinese but then I was interested in learning Spanish and even Russian, how about you?
Person A	Yeah, I hopped on as well and I was learning German at first but then I got interested in a little bit of Arabic and also some Spanish as well.
Person B	That's very interesting are you planning on using those languages abroad.
Person A	If I get accepted to my exchange application, then yeah definitely.
Person B	That's excellent! Where do you apply for exchange?
Person A	I applied to a few countries one of them was France and then Switzerland and also a school in Italy as well what about…

Sources

https://www.mathworks.com/help/audio/ug/measure-speech-intelligibility-and-perceived-audio-quality-with-stoi-and-visqol.html

https://www.mathworks.com/help/audio/ref/visqol.html