



# HAQ.NEWS Contest Report

Date: 07/09/2024  
By: Patrick Armstrong  
[patrickarmstrong.com](http://patrickarmstrong.com)

## Contents

Contest Summary .....	3
Objective.....	3
Hypothesis .....	3
Assessment Components .....	3
Tools and Equipment.....	3
1. Audio Capture.....	3
2. Editing and Analysis .....	3
3. Equipment: .....	4
Procedure .....	4
1. Audio Capture.....	4
2. Editing .....	4
3. Listening and Analysis .....	5
4. Visual Inspection .....	6
Findings .....	6
AI Voice Characteristics.....	7
Human Voice Characteristics.....	7
It is important to note .....	7
Results.....	7
Recommendations to improve AI speech simulation .....	7
Conclusion .....	9

## Contest Summary:

**HaQ.NEWS Podcast**, run by Jared Folkins hosted a contest, challenging participants to identify an episode featuring his daughter Gracie's real voice among AI-generated versions of her voice. This task leveraged the rapid advancements in AI speech generation and tested participants' ability to differentiate between human and AI-produced audio.

## Objective:

Identify the episode of Jared Folkins' podcast that features his daughter's authentic voice among AI-generated versions. Provide a detailed explanation and empirical evidence to support the findings.

## Hypothesis:

In the AI audio technology arms race, the leader must deliver the highest quality possible. Therefore, an AI version will likely produce exceptionally high-quality speech. Conversely, the average person typically lacks high-quality recording processing equipment, an acoustically treated vocal booth, and the expertise to process and edit vocal audio.

## Assessment Components:

In a sound-isolated environment, I critically listened to each episode of the HaQ.NEWS podcast. Taking note of sound clarity, audio levels, speech patterns, and speaker presentation. I then recorded samples of each podcast episode to process, compare and visually inspect the podcast episodes.

## Tools and Equipment:

### 1. Audio Capture:

#### Tools:

- Audacity – Audio recorder and editor.

#### Purpose:

- Capture podcast audio from the web browser.

### 2. Editing and Analysis:

#### Tool:

- Ableton Live (Digital Audio Workstation software)
- Ableton Live's stock audio effects:
  - Multiband Equalizer and Spectrum Analyzer

#### Purpose:

- Edit, organize, process and analyze audio clips.

### 3. Equipment:

Sennheiser HD 650 Headphones:

Reference-quality headphones are designed to reproduce sound as accurately and transparently as possible. They aim to provide a flat frequency response, meaning they reproduce audio without adding coloration or emphasizing certain frequencies over others. Also providing clear, detailed, and accurate sound reproduction across the entire frequency spectrum (bass, midrange, treble) and handle a wide range of volumes without distortion, allowing for precise monitoring of audio dynamics.

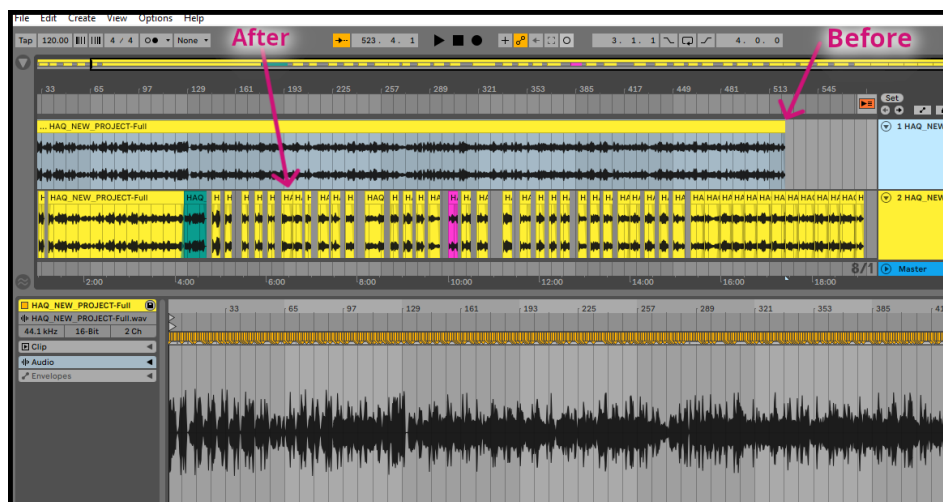
### Procedure:

#### 1. Audio Capture:

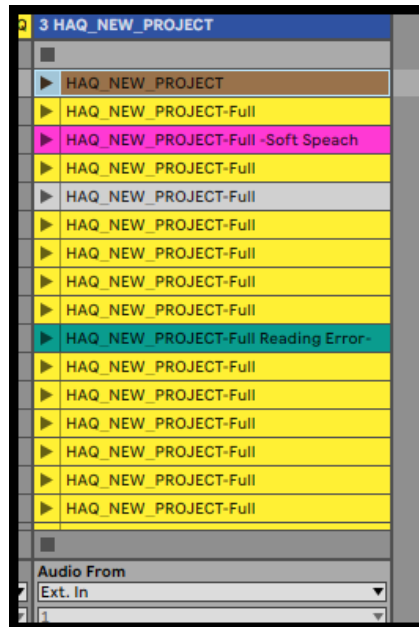
- Captured 30-second audio segments from the beginning of each episode using the program Audacity.
- Compiled the segments into a consolidated and continuous audio file.

#### 2. Editing:

- Loaded the audio file into the audio editing program Ableton Live.
- Re-segmented the extended audio clip encompassing multiple podcast episodes into smaller segments to isolate each instance where the speaker vocalizes the name and date of each podcast episode.
- Imported each clip into Ableton Live's Session view to facilitated rapid alternation and comparison of audio clips.



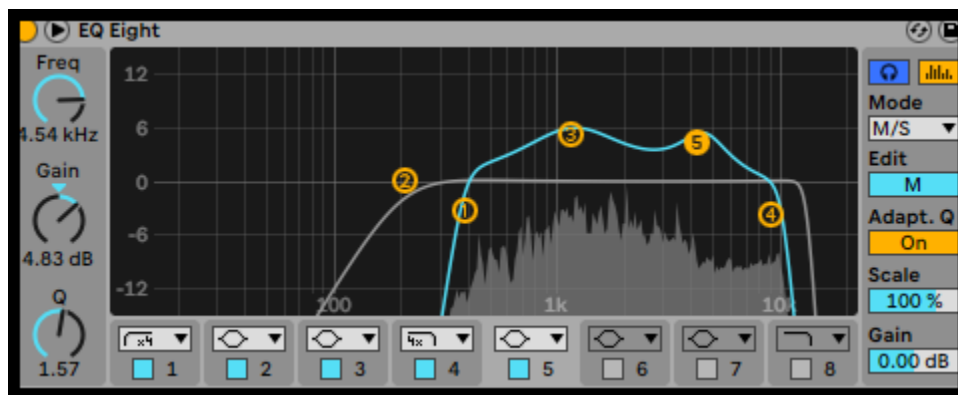
Large audio file segmented into small audio clips with the speaker vocalizes the name and date of each podcast episode.



Ableton's Session view.

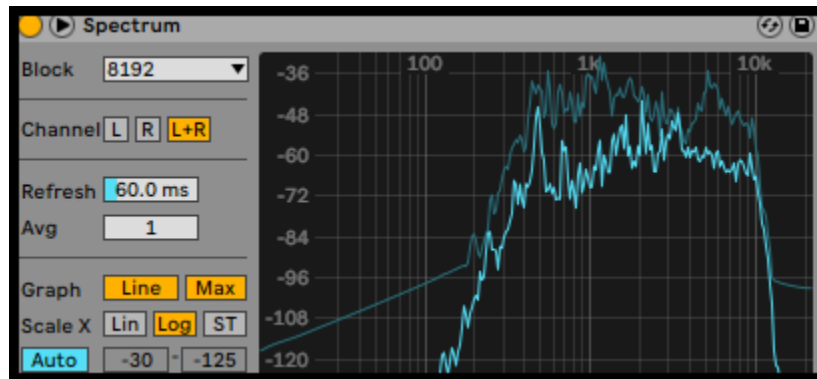
### 3. Listening and Analysis:

- Listened to the audio clips through reference-quality headphones.
- Applied audio filters to isolate the vocal frequency range, thereby eliminating extraneous sounds and removing sound that falls outside of the mono range.

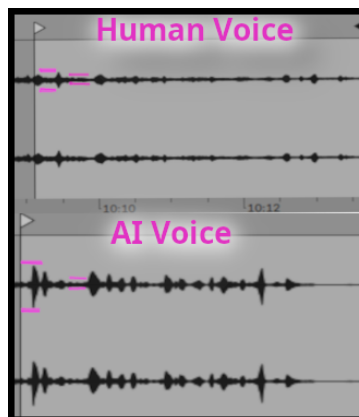


View of the filtered vocal frequencies.

#### 4. Visual Inspection:



Monitored the played back of the audio clips through a spectrum analyzer to search for inconsistencies in the audio files.



Compared the dynamic range of the audio clips side by side.

#### Findings:

After review, I identified an audio segment characterized by reduced amplitude and softer intonation in articulation, indicating the potential presence of human speech. This segment visually appeared more compressed compared to the other audio clips, which all had a very similar characteristic.

### **AI Voice Characteristics:**

- High-quality, polished sound.
- Vocal performance that is clear, enunciated, and properly pronounced, comparable to that of a professional speaker.
- Little or no noise floor in the recording (This is an artifact of digital recording and production. Many audio software companies offer tools to recreate this characteristic).

### **Human Voice Characteristics:**

- Softer, whispery quality.
- Compressed sound with less dynamic range.
- Visual inspection showed lower peaks and less pronounced valleys between spoken words.
- This audio clip appeared to feature the speaker utilizing thoracic breathing, whereas the other clips seemed to demonstrate diaphragmatic breathing.

### **It is important to note:**

The AI introduced deliberate imperfections into the speech and performance to deceive the listener.

- That the AI versions added inconsistencies, pauses and stutters to mimic human speech.
- The AI versions included breathing sounds that would be typically edited out in professional recordings.

### **Results:**

The analysis confirmed the hypothesis: The clip with softer speech, compressed audio, and less dynamic range was identified as featuring Gracie's real voice. The evidence provided sufficient basis for me to finalize my decision and submit my entry that episode 3/14/2024 featured Gracie's authentic voice.

### **Contest Results:**

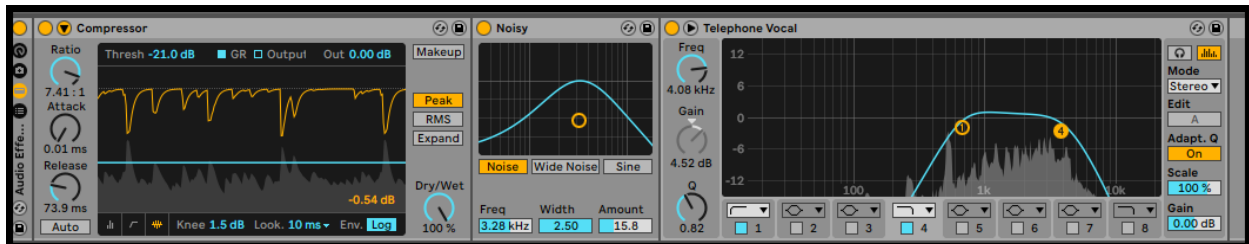
A few weeks later, I received notice from Jared Folkins, confirming my success in the contest.

### **Recommendations to improve AI speech simulation:**

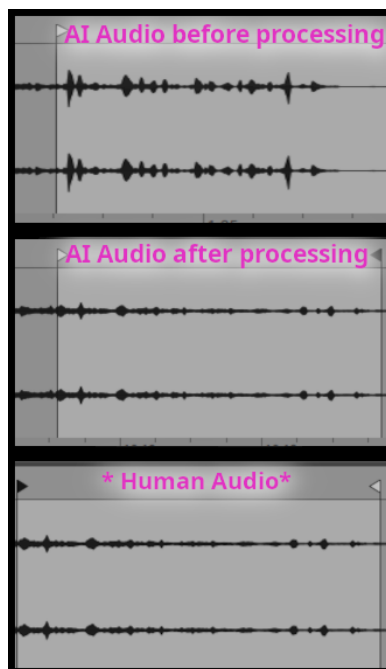
(However, by the time you read this, changes may already have been implemented.)

- Add a quality control option.
- Focus on refining dynamic range and compression to mimic lower-quality human recordings.
- Use an audio compressor with a strong ratio, a short attack and a low threshold.
  - o This will lower the vocal range of the AI's voice.

- Filtering the audio cutting the low frequencies at 680Hz and the high frequencies at 4000Hz.
  - o This will simulate telephone like audio quality.
- Add a Vintage vinyl/tape simulator.
  - o This process will introduce subtle noise and saturation to the AI recording, thereby creating a lo-fi effect.
- Continue adding natural inconsistencies and breathing sounds for authenticity.



I utilized Ableton's stock audio effects to enhance the audio and demonstrate how to obtain a more authentic sounding recording.



This is a visual comparison of the AI audio file before and after processing, compared to the human voice. In the processed AI file, note the reduction in spike amplitude compared to the human voice.



## Conclusion:

The HaQ.News podcast contest provided a compelling platform to assess the current capabilities of AI speech generation technology. By analyzing audio samples and employing auditory and visual inspection techniques, I successfully identified an instance featuring Gracie's authentic voice amidst AI-generated versions. The findings underscored the advancements in AI's ability to produce high-quality speech, often indistinguishable from human recordings while also revealing subtle imperfections deliberately introduced to emulate natural human speech patterns.

The identified segment exhibited distinctive characteristics such as softer articulation, compressed audio, and reduced dynamic range, indicative of human speech traits. These observations align with the hypothesis that in the competitive landscape of AI audio technologies, efforts are directed towards achieving optimal quality, often surpassing the typical recording conditions available to the average person.

Moreover, the contest highlighted areas for AI refinement. Recommendations included refining dynamic range and employing audio processing techniques such as compression and frequency filtering to achieve more authentic results.

In conclusion, the contest not only validated the hypothesis but also emphasized the evolving complexity of AI-generated speech and the ongoing pursuit of authenticity in audio replication. This experience underscores the need for continuous improvement in AI technologies to better replicate human speech nuances convincingly.

- Patrick Armstrong



Thank you to [Jared Folkins](#)  
Check out his Podcast @ <https://haq.news/>