

# Iowa Initiative for Artificial Intelligence

## Final Report

Project title:	Using Artificial Intelligence to Improve Voice Recognition in Surgery		
Principal Investigator:	Scott K. Sherman MD		
Prepared by (IIAI):	Avinash Mudireddy		
Other investigators:	John W. Cromwell, MD; Peter Nau, MD; James P. De Andrade, MD; Catherine G. Tran, MD; Mohammed O. Suraju, MD; Jeremy Chang, MD; Fiona A. Armstrong-Pavlik, BS.		
Date:	05/18/2024		
Were specific aims fulfilled:	Yes		
Readiness for extramural proposal?	Yes		
If yes ... Planned submission date	TBD		
Funding agency	TBD		
Grant mechanism	TBD		
If no ... Why not? What went wrong?			

### **Brief summary of accomplished results:**

The project successfully converted audio files from MP3 and M4A formats to WAV format. These WAV files were then used as inputs to the OpenAI Whisper model to transcribe the audio into text. The transcription process included a post-processing step to remove metadata and correct common transcription errors. Each file, ranging from 20 seconds to 8 minutes in length, was transcribed significantly faster than real-time. Optimal machine specifications for deploying the model were proposed to the project PI.

### **Research report:**

#### **Aims (provided by PI):**

##### **Background:**

Healthcare extensively uses voice dictation, where automated voice-to-text systems have largely replaced human transcription. Automated transcription results in significantly lower accuracy but substantial cost savings. Accurate transcription is critical for operative notes as they convey important medical information and serve as documents for billing and reimbursement. Dragon (Nuance Communications) is a commercial voice-to-text system used at the University of Iowa but comes with high costs and varying accuracy. WhisperAI, an open-source speech-to-text neural network developed by OpenAI, presents a promising low-cost alternative, potentially offering near-human performance and significant cost savings.

##### **Specific Aims:**

1. Minimum Computing Resources:
  - Determine the minimum computing resources needed for prompt transcription by WhisperAI.

- Hypothesis: Processing times of less than 1 second per 1 second of recording can be achieved on GPU-equipped desktop computers.
2. Accuracy in Operative Note Transcription:
    - Evaluate WhisperAI's accuracy in actual operative dictations.
    - Hypothesis: WhisperAI transcribes operative notes with lower error rates compared to commercial systems like Dragon.
  3. AI in Operating Room Speech Recognition:
    - Determine if sound quality in operating room video recordings is sufficient for WhisperAI to produce accurate transcripts.
    - Hypothesis: WhisperAI can identify key steps in operations from audio recordings.

#### **Data for Aims:**

The dataset included 250 sample dictation recordings and operating room videos ranging from 150MB to 4GB in size.

The project utilized both the University of Iowa's Argon HPC cluster and GPU-equipped desktop machines for transcription tasks.

#### **AI/ML Approach:**

OpenAI's Whisper model was implemented using Python, converting audio files from MP3 and M4A to WAV format.

A post-processing script corrected common errors in the transcriptions, significantly improving accuracy.

#### **Input Data:**

- Audio files in MP3 and M4A formats converted to WAV format.
- Operative notes and real-time operative recordings were used as input for transcription.

#### **Experimental methods, validation approach:**

Transcriptions were validated against original recordings to determine Word-Error-Rate (WER), Meaningful-Word-Error-Rate (MWER), and Punctuation-Error-Rate (PER).

Processing time and hardware requirements were benchmarked.

#### **Results:**

- WhisperAI showed significantly lower error rates compared to Dragon for both WER and PER.
- Transcription processing times were within the expected range, confirming the feasibility of using GPU-equipped desktops for real-time transcription.
- Performance Tests:
  - NVIDIA A100-SXM4-80GB GPU with AMD EPYC 7413 24-Core Processor: Processed 264 files in 8155.36 seconds.
  - NVIDIA GeForce RTX 2080 Ti-11GB GPU with Intel(R) Core(TM) i7-9800X 8-Core Processor: Processed 264 files in 8324.18 seconds.
  - NVIDIA A100 SXM4-40GB GPU with AMD EPYC 7742 64-Core Processor: Processed 264 files in 8162.73 seconds.

#### **DISCUSSION**

- WhisperAI demonstrated strong potential as a low-cost, high-accuracy alternative to commercial transcription programs.

- The project's success supports further development and potential clinical implementation, particularly in low-resource settings.
- Recommended optimal machine specifications to the project PI for efficient deployment of the WhisperAI model.

**Ideas/aims for future extramural project:**

Project is ready for next steps ...