

Project Report

Topic: **Multilingual Speech Recognition for Clinical Note-Taking**

Student Name: Precious Kolawole - Student Number: 101285187

Course Code: COMP 3106 A - Professor Matthew Holden

Background & Motivation

Imagine being a clinician, trying to balance listening to your patient while capturing every detail of the conversation. It's a delicate task, and the current methods of note-taking just don't cut it. Many clinicians either scribble notes by hand or type directly into electronic health records (EHRs). Both methods come with their own headaches. Handwritten notes need to be digitized later, which is time-consuming and prone to errors. On the other hand, typing into EHRs during patient interactions can feel awkward and impersonal. It's hard to maintain eye contact and fully engage with the patient when you're focused on a screen, leading to communication gaps and leaving patients feeling less satisfied.

The struggle doesn't stop there. Clinicians often find themselves spending more time on paperwork than on actual patient care. This administrative overload is exhausting and can negatively impact the quality of care provided. That's why we're proposing a speech recognition system to ease the burden of traditional clinical note-taking. But we want to take it a step further—what if we could support conversations in all languages? This would make our solution even more inclusive and accessible, breaking down language barriers in healthcare.

Prior Works

In the field of AI-assisted clinical note-taking and medical speech recognition, there have been several notable developments and implementations, even in the Canadian tech scene. Tali AI Assistant was developed in Toronto, and has been assisting physicians across the country through its Medical Dictation that accurately recognizes medical vocabulary and terms; Ambient Scribe that listens to physician-patient conversations, further generating clinical notes; and the French language support that is trained on Canadian French speakers. The Ottawa Hospital recently launched a trial of Microsoft's DAX Copilot that uses ambient, conversational, and generative AI to record and convert physician-patient conversations into medical notes; it integrates with the hospital's EPIC electronic health records system with the aim to reduce physician burnout and increase patient care time. Internationally, we have Nabla, a France-US company involved in medical transcription based on Whisper. Whisper, which is also the foundation of my model architecture, was used by Nabla to transcribe an estimated 7 million medical visits while implementing measures to mitigate hallucination issues.

Objective

We want to make life easier for clinicians and improve patient care by using automatic speech recognition and natural language processing in capturing patient-clinicians conversations in real-time

while turning it into an accurate structured note. Also, while there are a few approaches to this out there, most are limited to English - not very inclusive. We want to develop a system to handle multiple languages leveraging on speech to text and translation models such that it is accessible to clinicians and patients anywhere in the world

PS: Upon working on this, I had realized there are several works out there tackling the issue of clinical note taking being too administrative. Without the inclusion of the medical side as well, I saw a few models that have been built on the functionality of speech recognition and a few more to tackle multiple languages. However, in as much as there are all these solutions, this could be an opportunity to train myself on working with audio files or even using pretrained models, so with its little or no novelty, I could count it as a personal project.