Thesis Project Portfolio

Voice Restoration Device Using Machine Learning of Acoustic and Visual Output During Electrolarynx Use

(Technical Report)

A Review of Algorithmic Bias in the American Healthcare System

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Katherine M. Taylor

Spring, 2022

Department of Biomedical Engineering

Sociotechnical Synthesis

Over 3000 patients in the United States receive a total laryngectomy, or total removal of the vocal cords, annually (Kohlberg, Gal, & Lalwani, 2016). The electrolarynx, which emits vibrations that are transmitted to the throat and are shaped into words using the mouth, is the most commonly used voice restoration therapy post-laryngectomy. The electrolarynx is non-invasive, has no complications, and is most cost effective compared with other forms of voice restoration (Carr, Schmidbauer, Majaess, & Smith, 2000). Despite these positive characteristics, the use of any voice restoration therapy is correlated with decreased quality of life due to decreased communication abilities (Carr et al., 2000; Cox & Doyle, 2014; Gates et al., 1982). Thus, an ideal voice restoration therapy would result in more intelligible speech and an increased quality of life for laryngectomees.

The proposed ideal voice restoration therapy is a deep neural network (DNN) that translates lip movements and electrolarynx audio output into computer-generated speech. Approximately 100 videos were recorded of five different participants reading *The Rainbow Passage*, a short passage containing all English phonemes; these videos were used as training data for the DNN. In the development stage, the DNN was split into two pipelines, video and audio, that would be combined at the end of the DNN. The video pipeline utilized DeepLabCut, a software intended to capture the geometrical configuration of user-selected body parts using a convolutional neural network (CNN) as well as a recurrent neural network (RNN) with long short-term memory (LSTM) to predict phonemes from lip movements. The audio pipeline extracted Mel frequency cepstral coefficients (MFCCs) from each audio file and used a combination of a CNN and an LSTM to predict phonemes from the filtered audio data. Each pipeline was trained separately using the *Rainbow Passage* videos and was tested using videos of

the same participants speaking conversationally. Finally, the two pipelines were combined at the end using ensemble learning.

The overall results of this project were underwhelming due to many setbacks in both pipelines. While both pipelines were able to successfully perform feature extraction after several setbacks in that respect, a viable way to quickly label each frame of each of the 100 videos with the correct phoneme so that the DNN could be trained proved to be the breaking point. The project ended with a thorough documentation of all the steps taken throughout the year in order to provide a reference for a continuation of the project by a new group of students over the summer or as a capstone project next year.

While this algorithm intends to increase intelligibility of speech and quality of life for laryngectomees, the introduction of machine learning to the laryngology aspect of the healthcare field brings with it the possibility for bias. Even if the algorithm is successful as a whole, there is the possibility that the algorithm will increase quality of life at a different rate across different races, genders, and ages, for example. Bias in machine learning algorithms has become more prominent in recent times due to large cases such as the failure of the COMPAS sentencing algorithm; this bias can occur as a result of multiple factors including poor feature selection, bad training data, or a lack of transparency in the algorithm creation process (Martin, 2019; Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021; Yates, Gulati, & Weiss, 2013). Bias in the healthcare field can be particularly deadly as it can severely affect both the physical and mental health of each patient.

The use of the Responsible Innovation framework during the development process has the potential to help mitigate algorithmic bias before it starts. This framework seeks to reduce unintended consequences of technology development using a four-pronged process of

anticipation, reflectiveness, deliberation, and responsiveness (Stilgoe, Owen, & Macnaghten, 2013). Thinking of possible biases before development, reflecting on previous research, having conversations with a diverse set of stakeholders, and responding immediately to any bias that slips through the cracks are all excellent ways of avoiding bias specifically within healthcare settings.

There are multiple instances of healthcare-specific algorithms that failed due to bias as well as several algorithms that excelled due to mitigation techniques that the developers used. These algorithms were found during a comprehensive literature review and will be analyzed in the context of the Responsible Innovation framework. Specifically, the tenet of the framework that was most ignored in each of the failed algorithms as well as the tenet that was most important to the success of the thriving algorithms will be examined. Overall, this aspect of the thesis portfolio seeks to inform any future collaborators on this project of the attitude with which they approach this project, as well as any specific techniques they can implement to avoid algorithmic bias.

Table of Contents