The overall performance

The English whisper-large model is the best with the word error rate (WER) at 10.7%, word accuracy at 89.3% and LATTEScore at 80.85%. These average numbers, however, conceal detailed disparities in different conditions and levels of severity. As an example, speakers with Parkinson Disease retain 75-80% of meaning, whereas speakers with Multiple Sclerosis do not get more than 55-60%. The severity is also significant in that mild cases have 94.0% word accuracy, moderate have 68.3% and severe ones 72.7% word accuracy.

Detailed Performance of English Model

In the English models the dependence between size of model and performance is obvious, but the outcome differs widely depending on the type and extent of the impairment. The large model maintains meaning in 85-90% of mild cases, 75-80% of moderate cases, and 65-70% of severe cases- which represents a fall of about 20-25 points between mild and severe. The differences in the medical conditions are even more significant, the speakers with Parkinson Disease have 75-80% preservation of meaning, speakers with neurodevelopmental disorders 65-75, speakers with Cerebral Palsy 60-65, and speakers with Multiple Sclerosis only 55-60. This difference of 30 between the best and worst-performing groups is a serious concern of equity on which future model improvements will be based.

Elaborated Swahili Model Performance.

Whisper-small model has a 34.1% WER, 65.9 percent word accuracy, and a 39.4 percent LATTEScore. The whisper-large model, however, is even worse with a WER of 41.6, a word accuracy of 58.4, and a LATTEScore of 22.0. This negative correlation indicates that there were inherent problems in the model construction or Swahili non-standard speech training data. The level of performance is consistently low in all the severity levels: mild cases have only 35-45%

meaning preservation, moderate cases 25-35%, and severe cases 20-30%. Such a flat curve suggests that the models have problems with Swahili speech recognition, and not with the extent of impairment. Moreover, the difference between accuracy of words and preservation of their meaning shows a severe malfunction: the models know the words but cannot understand their word meanings and context.

Swahili Model Challenges

The Swahili models have substantial challenges in all severity levels and conditions and show poor performance irrespective of the type of impairment. The Swahili models do not show any pattern of clear, condition-dependent preservation of meaning like the English models, but they remain at 20-45% preservation of meaning and 58-66% word accuracy with all users. It seems the models are not recognizing the Swahili language structure and context.

Our multi-metric assessment system has played a critical role in exposing performance patterns in terms of conditions and level of severity. The reliability of the findings in each group supports our conclusions; in one of the examples, Multiple Sclerosis speakers do worse in all of the metrics (WER, word accuracy, and LATTEScore). The gap of severity is apparent in all English conditions, but it is most in the case of Cerebral Palsy and Multiple Sclerosis where severe speech features present the hardest circumstances in terms of recognition. Such uniform trends in measures categorize these groups as the top priority in the future improvement of the model.

Implementation and deployment recommendations.

In the case of implementation in the real world, the English large model is adequate to serve the mild and moderate speech impaired users in most scenarios. Nevertheless, its drastic deterioration of performance in cases of severe impairments and speakers with Cerebral Palsy or

Multiple Sclerosis suggests that these audiences need more work on the model and, probably, more communication assistance. The significant difference in performance between conditions is highly inequitable- a user with mild forms of Parkinson may have 85% of the meaning preserved and easily interact with the system, whereas a user with severe Multiple Sclerosis may have 55 percent and have a hard time using the system. This 30-point difference underlines the need to have condition specific solutions rather than a one-fit kind of solution.

Future Directions

In the future, we should focus on condition and severity aware improvements, especially those of speakers with Cerebral Palsy and Multiple Sclerosis at any severity levels. Meanwhile, it is also necessary to improve the fundamental problems with Swahili recognition, enhance it with more training data and create Swahili-specific acoustic models. The regularities which are witnessed in our metrics give a clear, equity-driven plan; the plan which appreciates not only accuracy of words but also preservation of true meaning to each user. The aim is to narrow our performance gap based on the condition by at least 30 points to less than 15 points and to attain at least 60 percent preservation of meaning of all groups in the next developmental cycle.