??????????????????

??????????????????

???Sergio Bruccoleri



Would you mind repeating that?

This is a phrase we're hearing more often as people around the world wear face masks during the COVID-19 pandemic. To protect our health, we're making it a bit more difficult to communicate with our voices – a more-than-fair trade, but one that requires us to adapt, whether we are talking with a grocery store employee in the check-out lane or placing an order in a drive-through restaurant. Wearing masks means being more patient with each other. Wearing masks also means that AI-powered voice models need to be retrained.

The uptake of voice models depends on accuracy. We won't use voice models to make complex purchases or meet medical needs unless we are absolutely confident that they understand what we're saying. But how well do voice models understand us when we're wearing masks? This is a significant question as we encounter instances when we're wearing masks while we make voice commands to our phones and smart speakers.

How Centific Tested the Accuracy of Voice Models

At Centific, we decided to answer this question with a test using LoopTalk. For context, LoopTalk is our own voice-to-AI, text-to-AI transcription solution that enables the cloud-based creation of voice data, associated transcription, evaluation of AI engines, and the generation of bespoke clean, targeted and annotated, voice datasets. LoopTalk is powered by Centific's OneForma, which is our

AI-infused platform for data creation and curation tasks, leveraging the skills of more than 100,000 registered contributors who work remotely online from all around the globe. It is an entirely new approach to provide next-generation speech recording solutions. For our experiment, we:

- Relied on our own crowdsourced resources to form a test group. The group read scripts without wearing face masks. Then they read the same scripts while wearing face masks.
- Used LoopTalk to collect and compare the two sets of voice samples to discern whether the voice models experienced loss in accuracy when people read the scripts wearing mask.

Here is what we found:

- Voice models could handle simple sentences utterances that take around 5 seconds to be recorded. But when people wore masks and uttered more complex phrases, all voice models experienced a quality loss.
- Voice models experienced on average a 50-percent quality loss.
- The best-performing engine experienced a 25-percent quality loss.

So how can we mitigate against the impact of face masks on voice models? I believe the answer comes down to human intervention to help voice models adapt to changing times.

Voice Models Need Human Intervention

Voice models do not exist in a vacuum. People train the artificial intelligence that powers voice models based on many factors including cultural context. Today's voice models were trained using data sets created before the pandemic hit, and now voice models need to be retrained. But before the pandemic, people always needed to be in the loop to help voice models adapt to how human beings use language, including:

- Incorporating new idioms that crop of all the time.
- The use of voice models in an expanding number of contexts – such as in automobiles, which is a different environment than one's home.
- Training voice models in a more privacy-centric world.

Now it's time for people to train AI to handle how human beings talk while they wear masks. Doing so involves taking steps such as:

1) Engineer the application to mitigate the issue

A quick hack to mitigate problematic keywords and words in a voice-powered application is to use the data collected by the application itself to identify the words that get incorrectly transcribed; and to let the application make assumptions that correct the transcription in order to deliver the intended meaning to the user.

For example, a voice powered application in a fast food environment transcribing "May I get some orange shoes?" should take into account that what the user very likely meant is "orange juice" and repair the error from the model at an application level, or ask the final user for confirmation.

2) Increase the dataset

Engineering the application to make a course correction will deliver some quick results to make sure

customers still have a pleasant end user experience in the short-term. The long-term solution is all about increasing the dataset and to collect voice samples that are actually mimicking real-life scenario; which at this point in time will need to include muffled speech voices in a wide variety of environments (some more crowded than others!)

Our Centific LoopTalk application relies on our large freelancers pool to record speech samples following specific instructions to power voice-enabled applications with the customer model being part of the whole recording and improvement process.

Our clients have used our OneForma platform for voice recognition training in more than 50 languages, providing support to a wide range of products, from voice models to voice recognition in shows for accessibility. A large variety of accurate transcriptions means that end products can recognize a wide variety of voices, taking into account factors such as age, accent, and the voice pitch.

To learn more about how to adapt your voice-based experience, **contact us**.