I tested four sentences, two with names of historical figures, and two with fictional names from video games. Then, I repeated the sentences twice, once with a clear voice, and a second time rushing my words a bit. For all the sentences, when I was speaking clearly, they were all transcribed correctly. However, when I spoke naturally (faster, not enunciating as clearly), the transcriptions became less accurate. The first sentence ("The first emperor of China was Qin Shi Huang") was correctly transcribed the second time, though the confidence score decreased from 0.71 to 0.65. The second sentence ("The kingdom in Legend of Zelda is called Hyrule") had "in" transcribed as "and", but the fictional names "Zelda" and "Hyrule" were correct, which surprised me. The third sentence ("My favorite Pokémon is Milotic") suffered the most in the second transcription, as it was transcribed as "My favorite Pokémon is my low tick" and had a confidence score of 0.55. The last sentence ("The Russian composer Tchaikovsky composed Swan Lake") was also transcribed incorrectly when I spoke faster, taking "Russian" as "rushing".

I think recognition falters in cases where the utterance is not pronounced clearly or enunciated particularly well because the machine simply assumes that whatever way someone is speaking is the default, and measures its transcriptions against that, rather than being able to notice that the person is speaking faster or with an accent and automatically filling in the blanks like humans do. This problem could be solved by putting greater focus on each predecessor word during transcription and then using context to better approximate what the next word should be. For example, with the sentence "My favorite Pokémon is Milotic", the previous group of words "My favorite Pokémon is" should give the machine a clue that whatever sound will follow the word "is" should be the name of a Pokémon. Then, if it still transcribes the utterance as "my low tick", there should be some mechanism in place that can cross reference the sound of "my low tick" with the names of Pokémon to identify what the person is trying to say, which is "Milotic".