Speaking Rate Detection and Feedback

I think the accuracy of this speech recognition system is acceptable. It can detect approximately 70% of the words correctly. However, in some perspective, it's performance is quite bad.

- Speed Up
- Accent
- Liaison
- Some Names or Place, or a name or place were mentioned within a sentence
- Words has similar pronunciation

Besides words, it cannot identify the gap between sentence accurately. Sometimes, it will append two sentence together.

But, this calculating speaking rate highly depends on the recognizing rate. However, I found that the recognizing system cannot recognize some words when we speak in a fast speed. In such circumstance, though we have a high speaking rate, the result may calculate a lower rate.

As for the selected thresholds, actually, I think this thresholds needs further consideration. Or I think it may be more proper to adjust the thresholds according to the speaker. Because for students or lecturers whose native language is not English, 133 words/minute seems a quite tough task for them(At least for me). According to my test, only reporters in BBC or CNN can reach that speed. Besides the language problem, this threshold are hard to reach thus the error rate of this speech recognition system increase with the talking speed. For example, I found that this system may neglect some words we spoke. This inaccuracy may also leads to lower speech rate.

To further improve my detecting system, one way is to lower the sampling rate so that we can guarantee that user can at least finish a sentence(finish_transcript not null) after sampling.

I use three different feedbacks in this system, color+text hint and chart.

- Color + Text Hint

When the speech rate less than 133 words/minute, it will display a hint in red color tells that you should speed up your speech. Similarly, it will display a hint in red color tells you that you should slow down your speech when your speech rate exceeds 188 words/minute. Otherwise, it will give a hint in green shown good. In this way, user can get immediate feedback and hint and user doesn't need to pay lots of attention on the screen.

- Chart

This chart reveal the speech rate variation during the speech. The x-axis indicates the time slot will the y-axis shows the speech rate. I designed this chart for later reveal so that user can further improve his/her speech skills. To avoid distraction, I put the chart in the bottom.

Automated Agent

I think the automated agent can make more or less makes our live more convenient and efficient. For example, we can use a automated agent to classify each calls to different online agent, also, we can get some basic help from them anytime. In most case, they can perform appropriately, thus the tasks assigned to them is quite simple, like authentications, classifications. However, when dealing with complicated task like repairing. They cannot work well thus there could be various conditions, some of which cannot predict by programmers.

To implemented my own automated agent, besides using speech recognition, I use a tree structure with each node represents a specific status during the interaction. So that I can always keep the status, and avoid conditions like when I say hello again, the agent will response the services again even though I want to buy a product.

I think Google speech API cannot now fully substitute actual conversations. First, the pronunciation need further improvement. For example, it cannot change it tune of speak. It can say everything in a plain tune, but cannot ask questions or show surprise properly.

Video Link:

https://youtu.be/BVCWYXb5sHo