Voice Recognition Flow

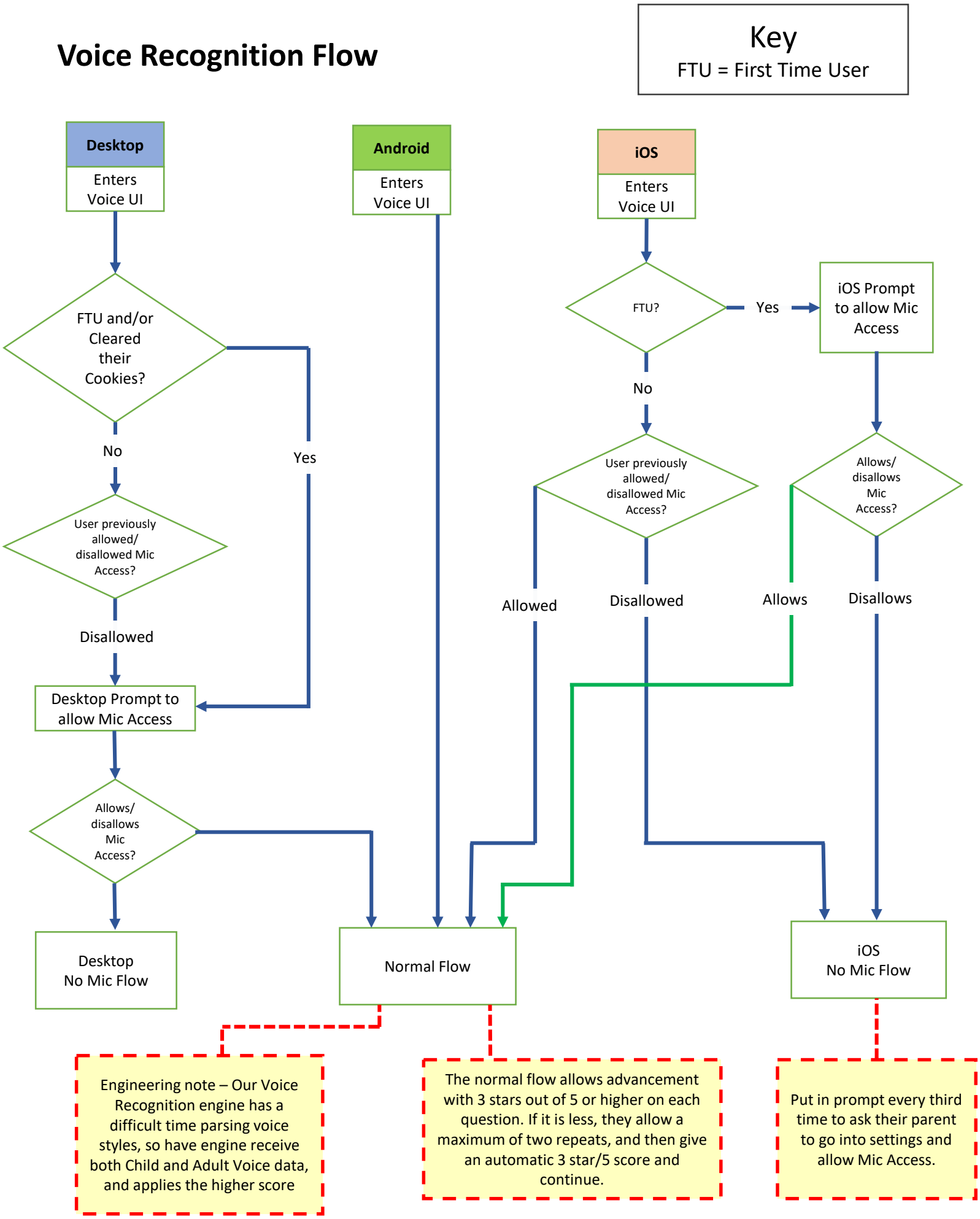
I worked on a Voice Recognition App that was meant to teach Chinese children English. The Chinese children would repeat a word, and the app would give them a score of 1 – 5 stars depending on how well they pronounced the word.

Challenges

We immediately ran into challenges:

- The Voice Recognition engine first needed to figure out whether an adult or a child was speaking. It would often guess wrong and give the child a zero.
- Voice Recognition engines in general are difficult. If a kid is outside, or any number of other things happen, it may give them an artificially low score.
- We wanted this to work across Desktop, Android and iOS apps. Each one of those have different mic allowance protocols.

Solutions

- After talking with many engineers, we solved the adult/child dilemma. An engineer had the Voice Recognition Engine take both scores, and only register the highest.
- We put in systems to sense repeated low scores. After a certain amount of time, we would just give them 3 stars and move on.
- For mic allowance protocols, we found three things:
  - Desktop – not a real problem. The browser can ask for Mic Allowance each time until the user clicks yes.
  - Android – not a problem at all. We can ask for Mic Allowance before the app install as a whole, so this won't be a problem.
  - iOS – this can be difficult. The user sees a Mic Allowance prompt the first time they open the UI, and if they say no, they can only allow it by going into settings, or reinstalling the app.
    - We user tested a prototype popup and found that 6/6 kids and 6/6 parents immediately allow access.
    - So we just accepted this, and built in a prompt reminder if the User says no.

Voice Recognition Flow