

Basic Voicefox Model

- Send sample through a speech verifier
- If accepted, then pass the sample through STT
- Analyse the output against the actual reference text
- Check the statistical results against a set of rules
- If it passes all the rules then accept the sample

The Rules

- Once we have the raw text analyzed against the reference text, we check the Word Error Rate(the ratio of words that are wrong)
- If the WER is too large, then we fail it, this reduces the false reject rate, and if any of the words in the sample don't exist in our dictionary, we also fail it, this reduces the false accept rate
- This works because as is shown in the paper, the STT system isn't very good at correctly transcribing the synthesized samples, so by using this to our advantage, we can tell which are synthesized and which are real