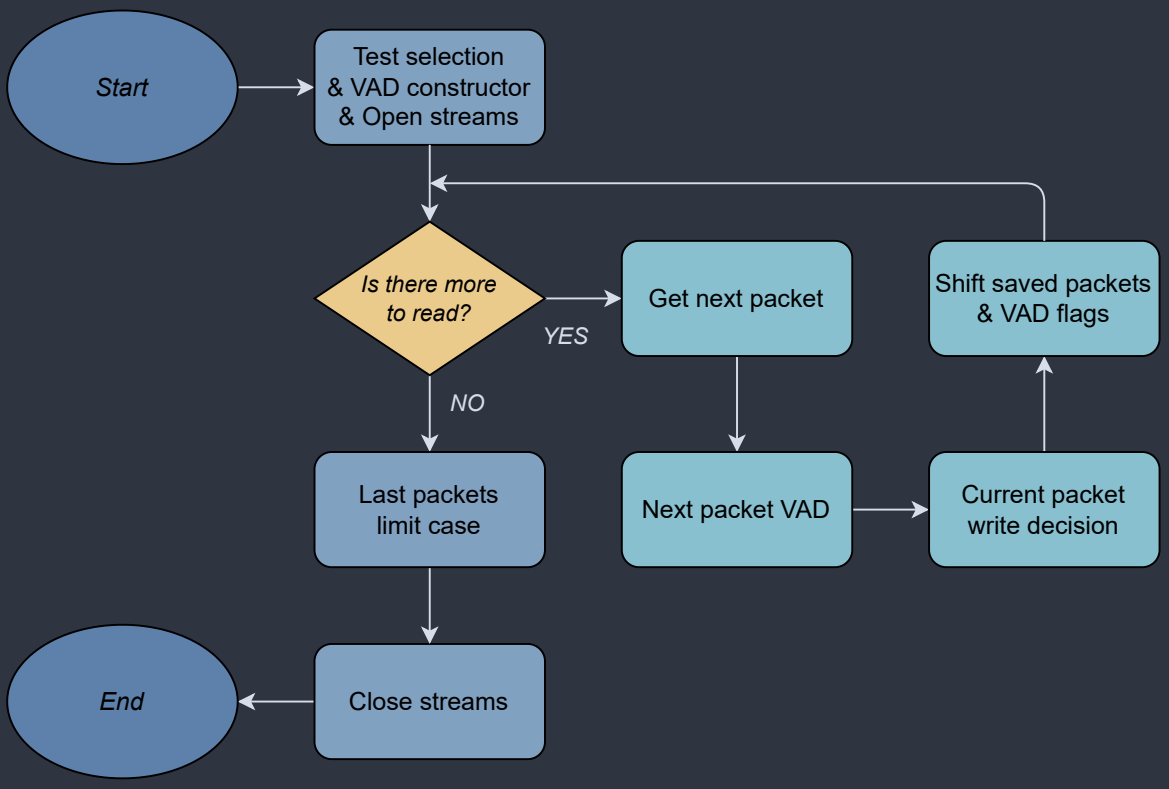
Andrea Valentinuzzi

Flowchart - Voice Activity Detection



The program reads a sample of audio file and detects if the audio is active or not, then smartly removes the background noise when not speaking.

LOGIC:
-# Read the audio file
-# Detect if there is voice or not
-# If there is voice, write the audio file in the output file
-# If there is no voice, write the silence in the output file
-# Repeat the process until the end of the file
-# Last iterations limit case
-# End

DECISIONS:
Every time the program reads a new packet, it is put in the vector packet[packet.size( ) -1] and the result of the voice detection for that packet is put in the vector has_voice[has_voice.size( ) - 1].
Then if any of the values of the vector has_voice is true, the current packet is written in the output file, since it means that there was voice detected nearby or in the current packet.
If instead all the values of the vector has_voice are false, the current packet is written in the output file as silence, since it means that there was no voice detected nearby or in the current packet.
After that the values of the packet and has_voice vectors are shifted to the left to make room for the new packet.

LIMITCASES:
The first iterations are not a limit case, since the writing is based on the size of the packet[0] vector, which is 0 for the first iterations.
The last iterations decide how to output the last packets, as voice or silence. They have to be done outside the loop since the input file stream returns false for ifs.good() at that point.

NOTE:
The best results are obtained when the program looks ahead and back one packet.
This allows for a more natural detection of voice, otherwise the words get cut off.
Looking ahead one packet means that the original audio will start playing 20ms (one packet) before the next voice section.
Looking back one packet means that the original audio will stop playing 20ms (one packet) after the last voice section.
This takes in account the fact that usually when speaking the voice takes some time to reach peak intensity (the first letter) and then fades out (the last letter).