Outline: Voice Based Voice Search

Keyword: timestamp, word, voice search, ffmpeg, audio, video
Author: Soumya Pandey, Aman Rai, Prudhviraj Boddu, Veni S, Nilay Patel, Vignatha Manchala
Mentor: Dr. Tanmay Bhowmik
Language Used: Python (3)
Buyer Persona: Students, Employed people, who prefer voice search (age 15-35 years)

VOICE BASED VOICE SEARCH

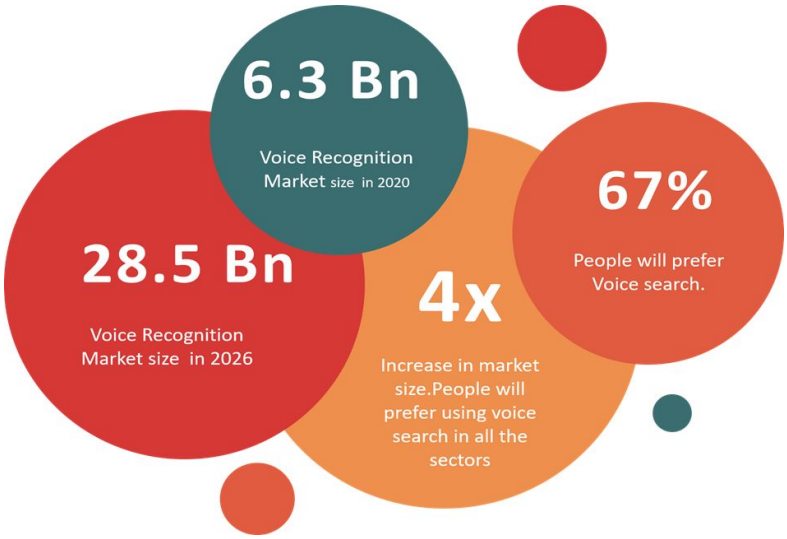
Introduction

.•*As voice search technology has leaped forward in recent years , the way many of us search online is changing .*

•*When we want to search for a particular word in a long durational video , our model takes voice input and searches the word in the video and gives required time stamps.*

•*We can select the timestamp and the video plays from the selected timestamp.*

Growth in Voice Searching



Methodology / Model Used

The project is designed as per the task that takes voice input from the user, also it allows the user to browse and select the video file of any format. The voice input could be in the form of a word, phrase or a sentence.

The input is then converted to text using speech recognition module. On the other hand, the selected video is converted into an audio file using ffmpeg and that audio is converted to text using the same speech recognition module.

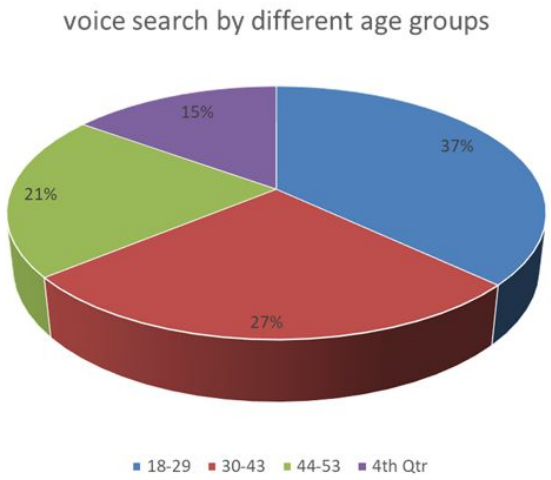
While converting the audio to text, the text is extracted in chunks of 3 seconds interval and text from each chunk is stored in a list. Also, while converting it uses more than one set of possible transcripts, the ones having slightly less probability than the actual result of conversion so that we acquire more accuracy with words.

Once the input and video both are converted into text format, string check is used to search the input in each chunk. This then generates timestamps from the video each time the input occurs.

Then the UI enables the user to select a particular timestamp from the list. The output video is then trimmed (from the original video selected) and plays from that particular time, hence, focussing on the part of the video in which the user is interested instead of watching the entire content.

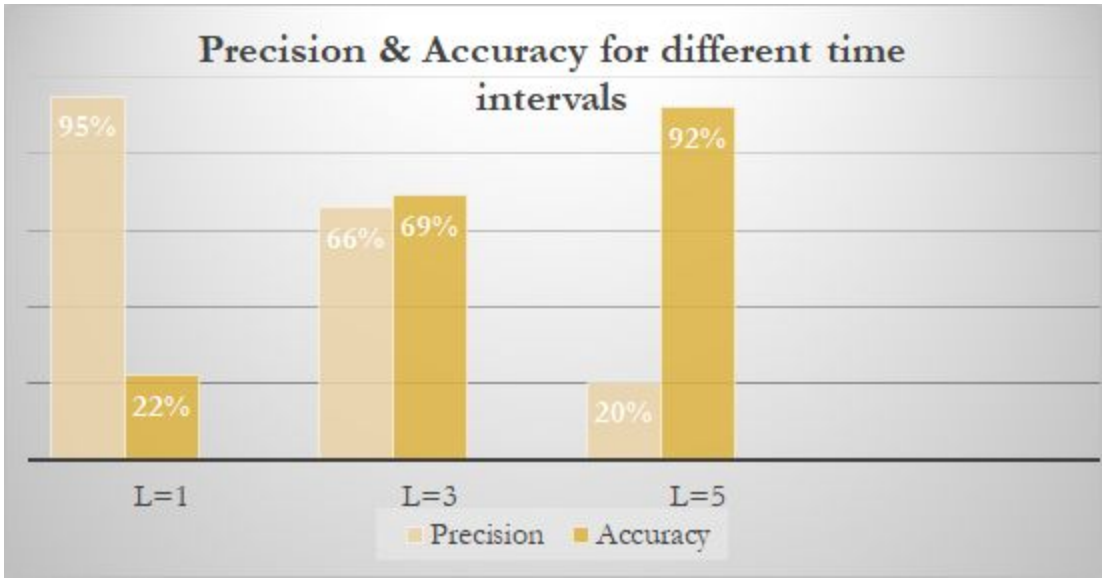
Annual Update

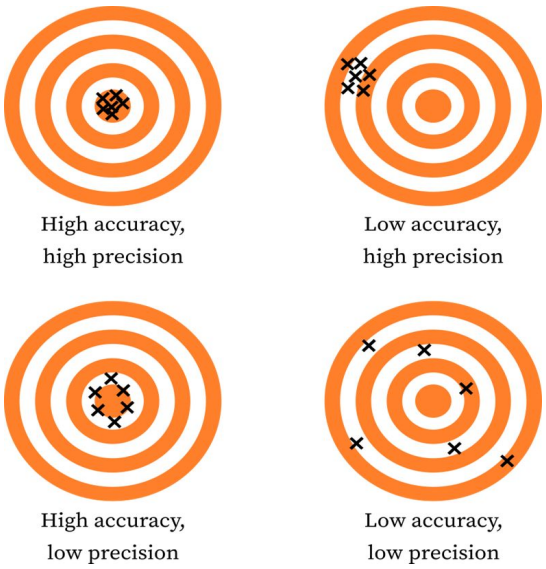
As you can see that the market of voice search is increasing rapidly and people from all age groups are preferring voice search over text search especially youngsters.



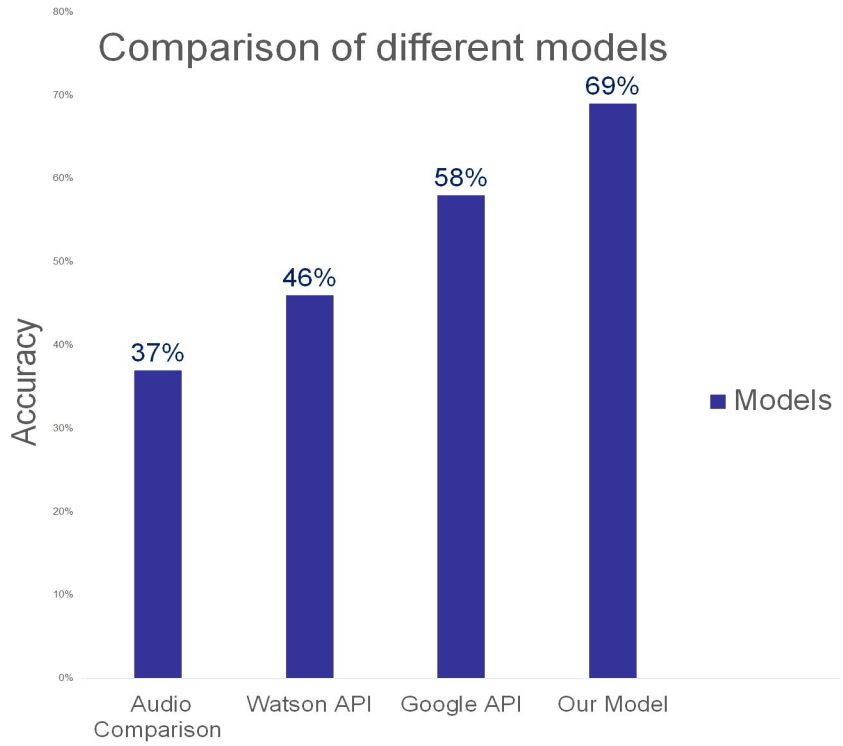
Results Archived

- *For every voice input we get timestamps of the word or sentence searched by the user and the video plays from the time selected by the user.*
- *When storing the text into an array , a time interval of 3 seconds is used to get more accuracy.*
- *If we use 5 seconds , the text in each index of the array will be more and we get less precision.*
- *And if we use less than 3 seconds , then the precision will be good but accuracy will be affected.*





Comparisons



Conclusion

- *The proposed model takes voice input for a video and gives the time stamps to the user. Then the video will play from the time selected by the user.*
- *The accuracy differs based on the video used by the user .The video may play with a bit error in the time.*
- *Voice based voice search can be implemented in different applications which contain video contents.*

Links

YouTube link: *https://youtu.be/w0bQFJeSBus*
Explanation Video :
https://drive.google.com/file/d/1X5Zp3Yl5l0iuE1Tzql9zWUDZ9JZRt3gF/view?usp=sharing

References

- Hurst-Hiller, O., & Farago, J. (2010). *U.S. Patent No. 7,672,931*. Washington, DC: U.S. Patent and Trademark Office.
- Ju, Y. C., & Wang, Y. Y. (2013). *U.S. Patent No. 8,589,157*. Washington, DC: U.S. Patent and Trademark Office.
- Mozer, T. F. (2010). *U.S. Patent No. 7,801,729*. Washington, DC: U.S. Patent and Trademark Office.
- Pettyjohn, N., Kulig, M., Jeffrey, N., & Saunders, E. (2015). *U.S. Patent No. 9,147,212*. Washington, DC: U.S. Patent and Trademark Office.
- Schalkwyk, J., Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., ... & Strope, B. (2010). "Your word is my command": Google search by voice: A case study. In Advances in speech recognition (pp. 61-90). Springer, Boston, MA.