Video-Translator

Learn in your own language.

The Concept

The best medium to share any information with others is a picture or a video. With growing technology every day to make the human community closer with each other, however this has also lead to formation of clusters, clusters that share same tastes, interests and most importantly; languages.

Though, today the internet is flooded with various video resources, a user's topic of interest is 'A' however he/she can only watch certain set 'B' (B is a subset of A) mostly because of the language barriers. Our project is an initiation to make the dream of a world not bounded by languages.

Our main focus is to try to solve for the domain of educational videos or MOOCs or any documentaries or news is available in only one language without any subtitles, talking about subtitles, they are more of a distraction and sometimes are out of sync or incorrect. Using Video-Translator, we tried to solve the above problem. It will transcribe the video, translate it and then merge it with the original video and will try to do it with as much synchronization as possible. This will enable the users to understand things in their own language and be comfortable with it.

Target Audience or Market:

We believe that knowledge should not be confined by languages, so the target audience are the people who are not comfortable or don't know the languages used in the educational videos or news or documentaries which is mostly in English. This includes many South American countries as the majority of the population their speaks Spanish and/or Portuguese or countries like India where many people either don't understand English or are not comfortable with it or can't grasp it properly.

Personas

- Documentaries like those shown on Animal Planet, Discovery or National Geographic.
- News recordings (not debates).
- Many courses available on Coursera.
- Recipe Videos
- Product Reviews
- Speeches (does not include the competitive election marches)
- Talk Shows
- Lifestyle channels

Dubbing a video nowadays costs around $1000 to $1500 for an hour. (Discovery and National Geographic Channel data). Using our services will just cost them around $60 to $100.

Feedback

We took feedback from 3 people from the product labs in our university. They are as follows-

- The idea is really good but you need to find out the stakeholders as early as possible.
- The idea is very nice and innovative, however you must ensure that the design created is with a broad future prospect so that tomorrow you can stand firm in any competition and earn profit.

- I had seen the output example and even understood the German in it, it's good enough but not 100% accurate, talking about the idea, surely quite innovative thinking; assembling pieces to build a tank.

How it works:
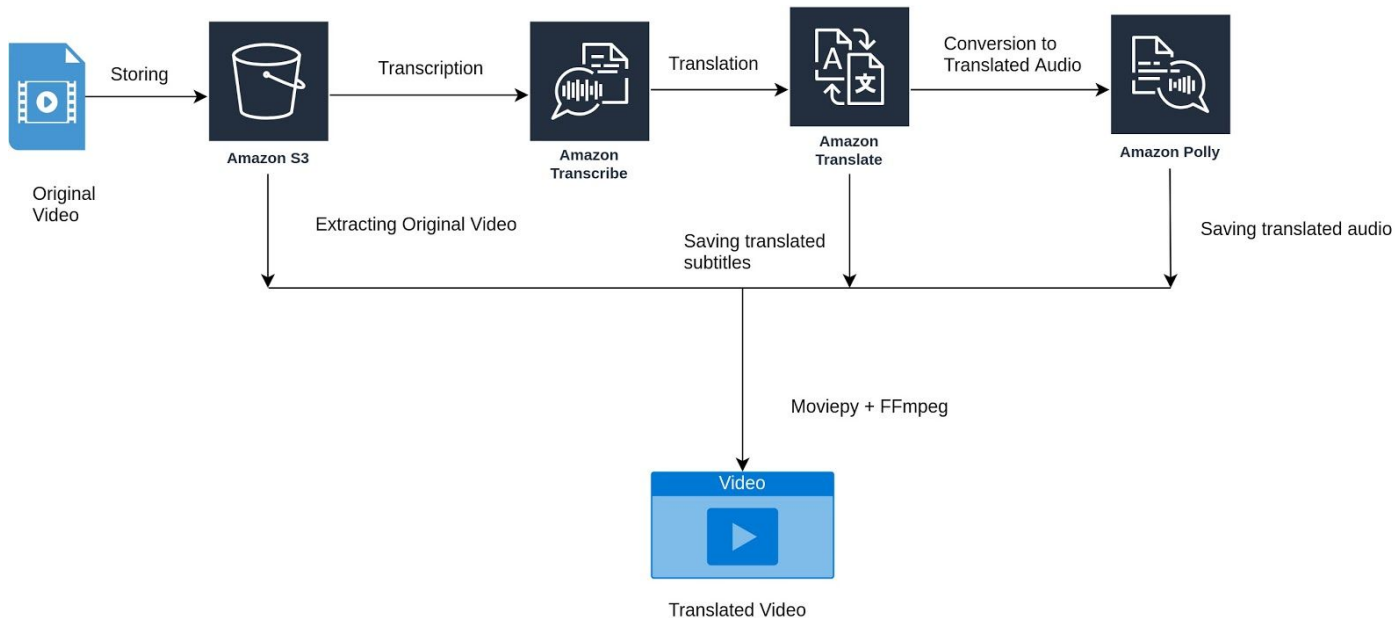
- The files are uploaded to S3 bucket.
- The audio file is extracted and transcribed by Amazon Transcribe
- Then translation of the saved data takes place using Amazon Translate.
- Now text is converted to speech using Amazon Polly.
- The original video and the new audio is merged using ffmpeg.
- SRT files are arranged and merged using moviepy.
  .

The project is highly modular and easy to edit and/or update. The implementation and integration of various APIs was a challenge and we overcame it using the well defined documentation of AWS. We also had to handle the output of various APIs and errors related to it.

Due to lack of features of directly converting speech from one language to another, we had to create a complex pipeline and used AWS AI Services to its fullest.
Also emotions can be captured to a very high extent using custom speech and acoustic models. It can capture more human like voices and also feelings like loudness while getting angry or decrease in pitch and loudness when shy.

The only thing we aren't able to solve is to find the mapping of idioms and certain dialogues to their actual meaning and not the literal one.

Core Technologies
- S3
- Amazon Transcribe
- Amazon Translate
- MoviePy

- Amazon Polly
- FFmpeg

The Business Plan:

We are mostly aiming for video sharing platforms like YouTube, Vimeo or MOOCs or any self uploaded videos. The content uploader can ask for the videos to be dubbed in a certain language. We will take a nominal fee per user and we will dub the video for them (More explanation in the Business Model Section). The content uploader will keep the translated video with themselves.

Competition:

Since the essential of a good Business Model is making the product stand out against the competition. The comparisons made are on the basis of the data available online.

Since we didn't find any other service provider providing a similar service as a platform or API, i.e. to dub a video by directly taking video input and providing a dubbed video as an output, so we are considering our biggest competitor as Real Time Translator. They average cost of any real time translator is around $200 and they are not real time as such. Highly advanced one which can differentiate between multiple people cost above $300 (refer Business Model section). Also there is a danger of losing or misplacing them and they might breakdown too. Also by our calculation, for an average course on Coursera, a user has to pay approximately Rs. 10 extra per course only. That is a very extreme case where we assumed a course has just 1000 users and we are taking Rs. 1500 as a profit margin). So basically from the above calculation, you can take around 1400 coursera courses from the amount you save on buying a Real Time Translator. An average length of coursera course is around 5 hours. So the user can watch 7000 hours worth of video.

Business Model

We have come up with rough plan for the cost estimation of deploying from service end and the per user estimation.

The Business Model works considering video platforms (like YouTube, Coursera, Plural, Twitch, Vimeo etc) as primary stakeholders .

Here we are describing two use cases :

- YouTube (Free Platform)
- Coursera (Paid Platform)

YouTube :

- A YouTuber (here considered those with videos getting like more than half million views) will make a request to "Video-Translator" to get a dubbed version of the actual video. The YouTuber will pay us a Nominal cost (a rough estimate of Rs 500-700/hr, this is upper-bound) plus the margin that we desire. Further we will, for a normal %age share (0.01%-1% depending on the type of YouTuber) for the amount received from YouTube for the translated versions of video.

```
Rough Cost = Video_length(in hrs) * 600 + Profit_margi
n + %age share for the number of views
```

Coursera:

- A course available on coursera costs around $39-$79. On an average course has 5 weeks with each week having an average of 50 min content. Now the Coursera make a request to 'Video-Translator' to get the content translated into different languages
- So

```
Appx Cost paid by coursera per language =
hours_per_week*No_of_weeks*cost_per*hour  + Profit Margin
Consider and example:

Appx cost paid by coursera to Video-Translator per language per course =
⅚hrs/week * 5weeks * Rs600/hr + Rs 1500 = Rs 4000
```

- Now considering the side of Coursera, it take Rs 6000 as its the deployment cost + other profit margin for the translated course content. Assuming that Coursera had already surveyed about the language preference and has got it translated into a language with a demand of 1000+ users so the cost for the translated course per user will be (Actual Cost)[$39-$79] + (Translation Charges)[Rs 10].