Profanity Filter to Beep Curse Word in All Regional Languages Video/Audio

Mukul Kumar Rajak (mkumarrajak@adobe.com)

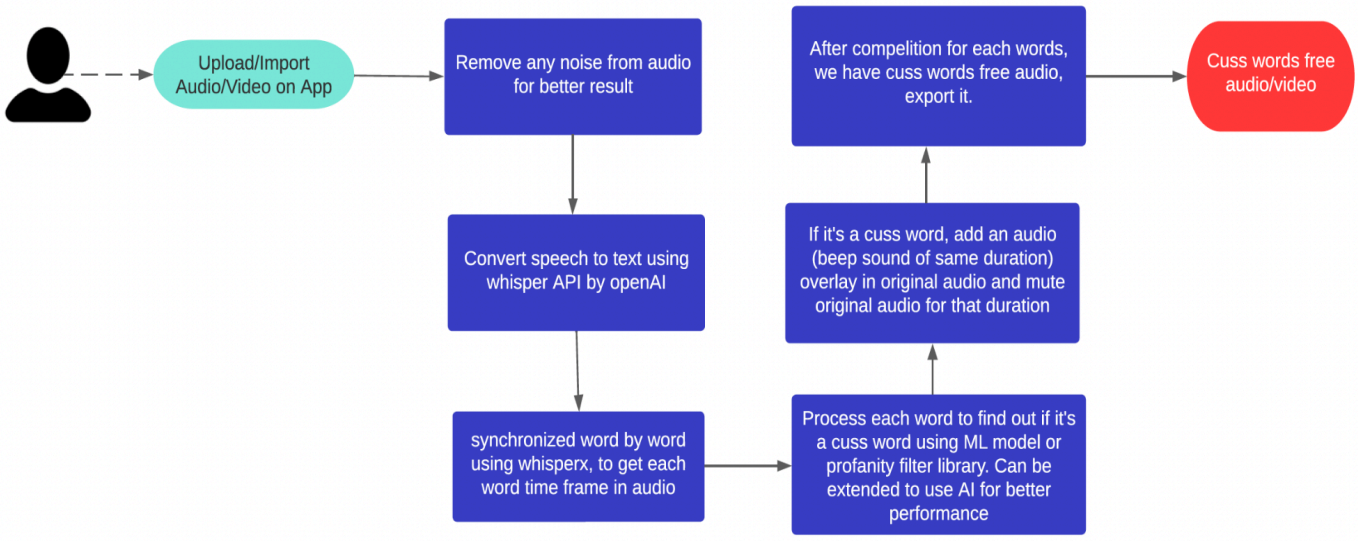
Achintya Dixit (acdixit)

PROBLEM STATEMENT

The problem statement is that Adobe podcast is an audio rendering software that produces high-quality audio output similar to that of a studio. However, the software currently lacks a feature that automatically removes cuss and abusive words from audio in one click. This feature is available in other non-Adobe products for English language audio (manual work required there), but not in Adobe podcast or other Adobe audio processing software such as Premier Pro and Premier Rush. The proposed solution is to develop a feature that can automatically remove offensive language from audio in one click, for english and other regional languages. The aim is to make Adobe podcast and other Adobe audio processing software more competitive in the market and to promote a more positive and respectful audio environment.

Though this problem seems simple, but there doesn't exist any such filter which automatically beeps cuss words in audio/video in one click.

SOLUTION



The possible solutions are:

1. Convert the audio to text using speech-to-text and then apply our machine learning model to detect the cuss/abusive words based on the threshold score that

how much that particular word is inappropriate we should either declare it to be an abusive word or normal word. And then return the timeframe where that word is and then we'll mute/beep the audio for that particular timeframe.

a. First we remove any noise from the audio
b. Then we convert audio to text using whisper API.
c. We are using Whisperx API which gives us word by word sychronised time duration of each word in audio in JSON format.
d. We then process that json file to check for cuss words for each word.
e. One by one we check if that word is cuss/abusive word or not (we are using `better_profanity` library).
f. If it's a cuss word, then we create a beep sound of same duration and we mute the original audio for that time frame and add beep sound by overlaying for that particular word timeframe.
g. At the end we have the audio which contains cuss/abusive word free audio. We can simply export it.

2. In Above solution, we can use chatGPT API while converting audio to text. It will directly remove the cuss/abusive word while converting, then we don't need to use any open source library and we will also get better performance.
3. By passing audio waveform and then doing similarity search on that waveform for finding cuss/abusive words.

BACKGROUND

Adobe is investing a lot in bringing its new product Adobe Premiere Rush, Adobe premier pro, Adobe Podcast.
Naturally there are many challenges like removing background noises from the audio. One such challenge that we are seeing is that people may want to remove abusive words from their audio. Currently we don't support any built in options in these apps to mute the abusive words.
A profanity filter is important for a number of reasons:

1. Maintaining a positive and respectful environment: The use of offensive language can create a negative and hostile environment for users, which can lead to arguments, bullying, and other forms of inappropriate behavior. A profanity filter can help to prevent the use of offensive language, thereby promoting a more

positive and respectful online community.
2. Protecting vulnerable populations: Children, in particular, may be more vulnerable to the negative effects of offensive language. A profanity filter can help to protect children and other vulnerable populations from exposure to harmful content.
3. Complying with legal and ethical standards: In some cases, the use of offensive language may be prohibited by law or ethical standards. A profanity filter can help organizations to comply with these regulations and avoid legal or ethical repercussions.
4. Protecting brand reputation: Companies and organizations may be held responsible for the content posted on their platforms. Offensive language can reflect poorly on a brand, which can lead to reputational damage and loss of trust among users.

Overall, a profanity filter can help to promote a more positive and respectful online environment, protect vulnerable populations, comply with legal and ethical standards, and protect brand reputation.

Please select all the Adobe products this idea could be useful for

Any platform where we edit, store audio content.

1. Adobe premiere pro
2. Adobe Podcast
3. Adobe premiere rush
4. Adobe express
5. After effects CC
6. Animate
7. Audiotion CC
8. Behance
9. Character Animator
10. Connect
11. Creative cloud
12. Spark video
13. stock

What is the problem you are trying to solve? Why is this problem difficult to solve? Demonstrate that it is a real problem by including use case scenarios. *

When user uploads the audio file, our job is to remove the cuss words, this feature is very important and we should have this feature in all the adobe app on which we edit or store any audio/video content. Eg: Adobe podcast and adobe premier pro app.

This problem is difficult solve as we don't have a proper API for converting speech to text and getting correct time frame for each word so that we can mute that particular duration of audio or add beep sound on it. We now have an api by whisperx which uses whisper by openAI to synchronise each word while converting speech to text and this synchronisation is currently available for {en, fr, de, es, it, ja, zh, nl, uk, pt} languages.

There exist some tool to remove cuss words, but that's only for english language.

However, it's important to note that these technologies are not perfect and can sometimes make mistakes, either by over-censoring non-offensive content or by failing to detect offensive language. As such, it's important to use these tools in conjunction with human moderation and oversight to ensure that the final output is appropriate and accurate.

The use of profanity and offensive language can be harmful and disrespectful, and can create a negative and hostile environment for listeners. It can also be particularly damaging in situations where the content is being consumed by young or vulnerable populations, such as children or individuals who are easily influenced by media. Additionally, in the case of other regional language like Hindi audio, offensive language can carry specific cultural and social connotations that may not be immediately apparent to non-Hindi speakers. Some Hindi profanities can be sexist, casteist, or derogatory towards certain religious or ethnic groups, which can lead to discrimination and prejudice. Therefore, removing offensive language from all the regional language audio can help to create a more inclusive and respectful media environment, promote positive and non-discriminatory language use, and protect vulnerable populations from harmful content.

What are the closest existing solutions? Why do they not solve the problem? Explain the technical differences and advantages of your invention over the existing solutions. *

Audacity: here user manually select the timeframe to remove the cuss words to make clean version audio.

Apple music: They already have clean version of the audio by user. Again it require manual work.

These required manual work. User have to manually provide clean version of the audio and to make clean audio, user have to listen the audio, and then needs to select the time frame which needs to be beeped. And that cuss word will depend on that user only, if he finds it abusive or not.

In 1-3 sentences, please highlight key aspects of your solution, including the advantages. Please clearly summarize what it is that you are attempting to patent. *

We are proposing an innovative solution to the problem described above –

- For different regional languages, we are trying to provide a profanity filter which will remove cuss words from the audio.
- Earlier it was possible to convert speech to text and then finding cuss words in that sentence, but it was difficult to find the time frame in which that particular word were spoken, so our option was to mute complete sentence, which was not a practical solution.
- So we are using https://github.com/m-bain/whisperX which gives us word by word time frame in audio, and we can find if that word is cuss word or not by using any ML or ai Model or any open source library which takes word as input and return true/false if it's cuss or not respectively, and then we mute/beep particular time frame in original audio easily. Currently whisperX provide synchronization for {en, fr, de, es, it, ja, zh, nl, uk, pt}. languages.

INNOVATION VALUE

- Innovative solution which fixes the issue – converting speech to text for profanity filter and then beeping particular time frame in original audio during which that word were spoken.
- With the proposed solution, the user can now remove profanity in one click automatically, for which user had to do the manual work of listening the whole audio and then manually beeping the audio.
- The idea is quite novel as this problem is quite common and there is no existing solution to this problem.

BUSINESS VALUE

- With each Adobe podcast and premiere pro upgrade, it is very important that we bring more smartness and intelligence to our tools so as to inhance our user experience. As our app is already remove noise from the audio, we must also give user an option to remove cuss words during cleaning of the audio, and hence making our app more intelligent. It will not only make our users happy but it will provide strong competitive edge.
- User uses our product to edit audio, videos and store videos, Even some movies are edited using our product. User had to manually beep cuss words while editing, which is time consuming. So this automation will improve their experience.
- Adding ML, ai based solution in Adobe's signature products like premiere pro and new product like podcast.

PATENT VALUE

- Very easy to check for infringement.
- It’s hard to find the work around. It’s not possible to find cuss word without converting audio to text and synchronising it word by word.
- Providing profanity for regional languages audio in just one click.

Is this AI/ML related? *

Yes

Describe the technical details of your solution. If possible, please include performance results and a comparison against the results of existing solutions. *

The possible solutions are:

1. Convert the audio to text using speech-to-text and then apply our machine learning model to detect the cuss/abusive words based on the threshold score that how much that particular word is inappropriate we should either declare it to be an abusive word or normal word. And then return the timeframe where that word is and then we'll mute/beep the audio for that particular timeframe.

- First we remove any noise from the audio
- Then we convert audio to text using whisper API.
- We are using Whisperx API which gives us word-by-word synchronised time duration of each word in audio in JSON format.
- We then process that JSON file to check for cuss words for each word.
- One by one we check if that word is cuss/abusive word or not (we are using the better_profanity library).
- If it's a cuss word, then we create a beep sound of the same duration and we

mute the original audio for that time frame and add beep sound by overlaying for that particular word timeframe.

- At the end we have the audio which contains cuss/abusive word free audio. We can simply export it.

2. In the Above solution, we can use chatGPT API while converting audio to text. It will directly remove the cuss/abusive word while converting speech to text, then we don't need to use any open source library and we will also get better performance.

3. By passing an audio waveform and then doing a similarity search on that waveform for finding cuss/abusive words.

All solutions can be used together. I've used the First solution and attached a Python colab notebook file which can be tested on sample audio provided in same folder.

Invention Status *

Prototyped

Do you have a demo? (Recommended) *

Yes

Provide a link to the demo *

https://adobe-my.sharepoint.com/:f:/p/mkumarrajak/Er-INzlF6ZRDiG2Oy3T908YBVwbNKW_rA0Of_ZhYXJUN2g?e=pg9z7Y

Please select the most appropriate Patent Review Committee (PRC) to evaluate your idea *

DM – NLP / Multimodal Knowledge / SemanticsDM