

Hello dear developer,

Below please find a short description about the current project (designated, for the time being, as 'Flipper'). I will consider ONLY freelancers that would provide, in addition to their price offer, milestones (time and money). Many thanks in advance, good luck and have a best day.

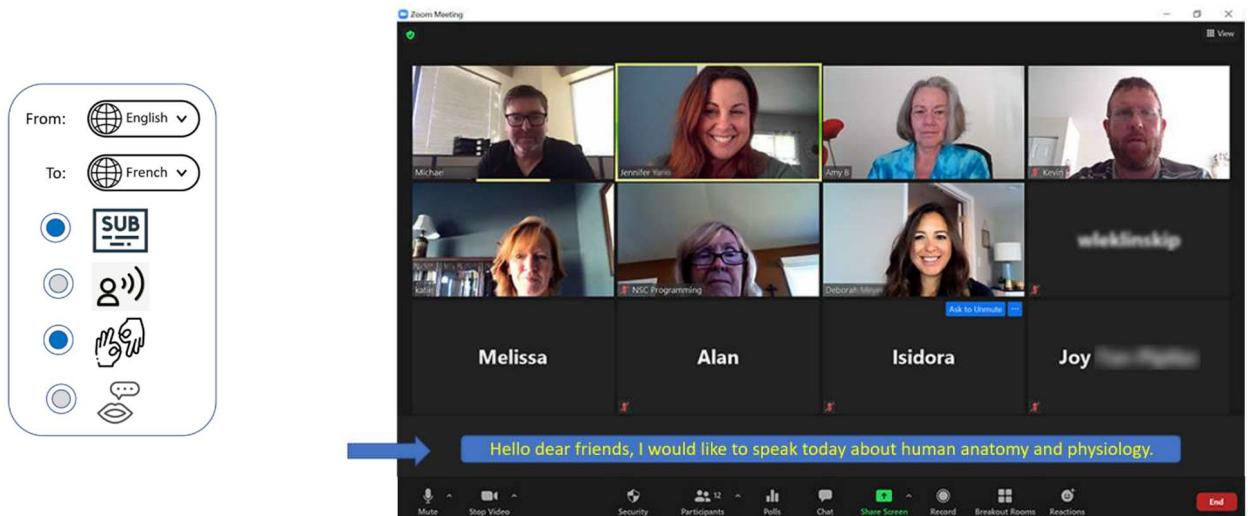
Flipper is about (a) translations from one language to another, and (b) should provide a utility for hearing impaired people. The needed translations in current project, can be done using common (hopefully free of charge) APIs. As far as I know, both "speech to text", "text to speech" and "translation from one language to another" have many free APIs that can be found in web. In addition, the project should utilize, an AI aspect as well (see below). Please note that all project stages, as described below, supposed to be complementary to each other. Namely, each stage should be based (and include) the stages before, to provide a complete tool.

Stage #1: To create a desktop tool that upon activated: (a) should "hear" a LIVE zoom meeting (while it occurs), (b) obtain the sound (=what they say in the zoom meeting, as a "source language"), (c) convert the obtained sound into text, (d) translate the text from the "source language" into a "target language", and (e) convert the translated text (from point d), into sound and play it (see mock designs).

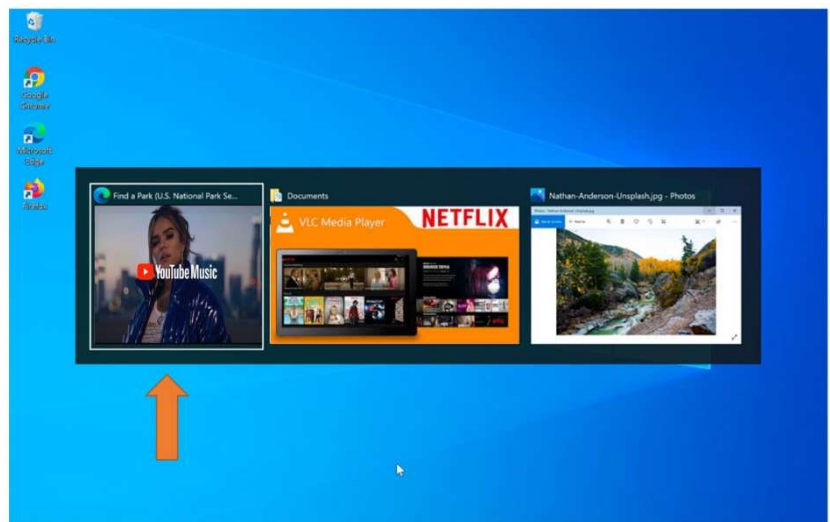
- User should be able to:
 - o (1) Turn on/off flipper
 - o (2) Select a "source language" (= the current spoken language on the zoom meeting)
 - o (3) Select a "target language" (= the language that the user wishes to translate the source language)
 - o (4) Turn on/off sub-titles
 - o (5) Turn on/off sound
- The 'Flipper' utility, should:
 - o (1) To "hear" the zoom sound and to convert it into a text (within the original (=source) language), by using a common / free "speech to text" API.
 - o (2) To show subtitles with the original language text.
 - o (3) To translate text from original language (=source language), into target language, by using a common / free "translation" API.
 - o (4) To show subtitles with the translated text (at the "target language").
 - o (5) To convert the translated text into sound and to play it (using common / free "text to speech" API). When activate this feature, the original zoom sounds, should be shut off and the user should hear only the speech that was done by the "text to speech" API.

BTW – It can be assumed that it will be some delay between the original spoken phrase, and the translation displayed on the screen. I hope this delay will be minimal and that the translation would be presented and heard very close in time, to the original sound. Please do the best that indeed, the time interval between the original sound and the target text/sound would be as minimal as it can be – namely, at the same moment the original phrase was said (and ended), translate it, show it on the screen, and play it. **Please note:** In case that "source language" and "target language" are the same (such case, indeed can be done), naturally, NO need for all translation processes (cause it's the same languages). Yet, all the rest, as describe below (such as subtitles, sign language, lip movement etc...), should work with source language.

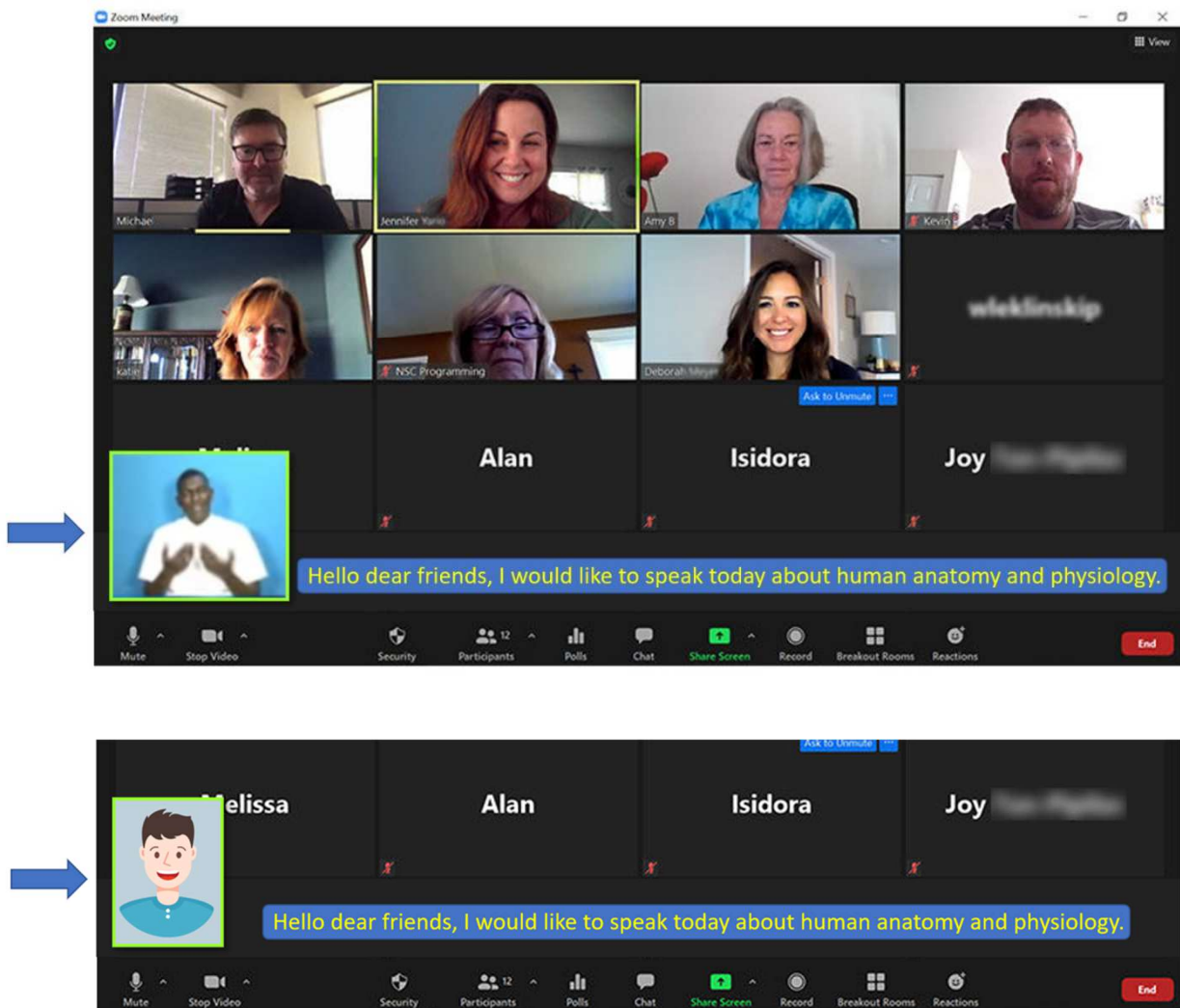
Below are some mock designs for both user panel and the desktop sub-titles:



Stage #2: To allow same features that developed on stage #1, on YouTube, VLC media player or any other web site that plays movies. Namely, to give the users the ability to select the "active window" by which the app would run. For example, to select the current browser's tab that plays YouTube or to select the current VLC window that is now playing some movie etc... see mock design:



Stage #3: To create, by using an AI tool, a "human figure" (=an avatar) AND a "face figure". The human figure should only be the upper (half) part of the body (because we need it only for sign language, in order to see his arms) and the "face figure" should allow the end-user to read the avatar lips. The human figure must be distinguished both by the movements of its mouth / lips and the movements of its arms/hands. Flipper should use this human figure, to show, **in sign language**, the new translated text (namely, in target language). Should be done in addition to stages #1 and #2. Below, please see mock design:



Few general issues:

- Regarding the mock design (as showed above): it is only a mock design, aimed to show in some figures my idea. The design is less important in this project, but indeed, it should be basically done to provide all needed features, as described above.
- I prefer (and wish) that the current POC would utilize a `free-of-charge` APIs for "speech to text", "text to speech" and "translation from one language to another". As mentioned above, there are plenty of such free APIs on the internet. However, in case of NO `free-of-charge` API was found, and/or a subscribing is needed, and/or pre-paid is needed... please let me know about that and I will take care of it. Thanks.

Q and A:

What specific languages should the tool (namely, Flipper) support in speech-to-text, text-to-speech, and translation? **Most important language is Hebrew for both speech-to-text and text-to-speech. Second important language is English (for both).**

Should Flipper recognize multiple speakers on a Zoom call and translate each separately, or focus only on the main speaker? **For the time being: let's exclusively focus on the main speaker.**

How quickly should the translation and subtitle display for live calls, and is any delay acceptable? ***Obviously, a short delay can be accepted. But please do the best it would be as short as it can be. It shouldn't be more than a few seconds.***

For the avatar, do you have a specific preference for its look or style, or should it just be a functional, clear figure? ***For the time being (first version as POC), it can be any clear figure, but it must be functional. Optional (not mandatory): to select between male avatar and female; to select between black or white avatars.***

Do you have any restrictions on using APIs that may have some minor costs? ***Not at all, but first, please do the best to search and find free options. I know and am aware that free options might be limited in quality, but, as much as I know, there are some Google APIs which are relatively good for such tasks. In case of needing some APIs for money, please let me know and I will subscribe, pay for it, and will provide credentials.***

Should there be options to adjust volume levels or display speed for translated text or audio output? ***For the first POC version, the answer is NO. First version should mainly be a POC. Thus, relatively minimal in its functions.***

Looking forward hearing from you. In case something is not clear and/or any question, do not hesitate to ask. Please provide milestones (time and money per each) and please provide a total cost and timetable for the entire project.

Many thanks in advance,

Looking forward