Refactoring Gentle as standalone component

(to use in autoEdit.io or elsewhere)

If interested in this project, I've started a conversation in the hyperaudio slack under #gentle_stt_refactor drop me an email if you need access pietro.passarelli@gmail.com

https://lowerquality.com/gentle/

Working on the refactoring this fork https://github.com/OpenNewsLabs/gentle

There are 3 ways of running Gentle it
docker, starting it as a local server, running it as a python command.
See README https://github.com/lowerquality/gentle

Running as local host server

```
curl -F "audio=@audio.mp3" -F "transcript=@words.txt"
"http://localhost:8765/transcriptions?async=false"
```

To connect with autoEdit, I've written a node module to talk to the API
https://github.com/OpenNewsLabs/gentle_stt_node see README.

at the moment autoEdit2 uses the local server.
Gentle is packaged as a mac app that starts a local server in version 9.0.1.
https://github.com/lowerquality/gentle/releases/tag/0.9.1

Downside of this approach
- the default local host port number could sometime change unpredictably
- the user is required to do extra setup to run the Gentle Mac app to use it

In order to integrate Gentle into autoEdit it is needed to make sure the component is well setup.
Gentle is written in python. That in itself is not an issue.

Running Gentle as pytnon command as described in README
https://github.com/lowerquality/gentle

```
python align.py audio.mp3 words.txt
```

Hower Gentle was born as a speech aligner, more then a transcription tool.
So the python command is expecting a text file to align those word with the audio.
Internally it generates a transcription from scratch.

Because of this inner working, I explained my use case to Rob (the dev who created Gentle) which is that of transcribing audio/video from scratch and he pointed me to version 0.9.1 https://github.com/lowerquality/gentle/releases/tag/0.9.1 where you can make an API call without providing a text file and it will return the transcription.

However the github code does not seem to support his out of the box, perhaps is not up to date with the 0.9.1 release (?)

So things to do with Gentle to make it a sensible standalone module that can be used in other context such as autoEdit 2. Without having to run a server to connect to it.

- Figure out where it calls ffmpeg, and see if the path to the ffmpeg binary can be passed as an argument.
- Seeing what is the most sensible way to run python from node
    - Option one is using node spawn to deal with child processes
    - Option two is using some other npm library
    - Option 3 is consider whether it might be worth it to re-write the python part in Node (?) - might be more time consuming but could make sense on the long run.

How I suggest to go about this

- Reading through the Gentle code is a great starting point, It does not have a lot of documentation so probably forking the project and writing some to make sense of the variosu components could be a good place to start https://github.com/lowerquality/gentle
- Identifying where ffmpeg is called
- Figuring out where the parameters that take the text input are, and what it takes to make those optional, and return transcription only skipping alignment when it is not provided.
    - Almost making it so that if you provide the text it would align
    - If you don't provide the text it would transcribe
    - If you provide the text, you could pass otpion for
        - "Conservative" alignment where any words not in the text are not added
        - "Complete" alignment where any words recognised not in the transcription are added to the results (and perhaps marked differently in the json)

Something else to consider is the Gentle is built on top of Kaldi.
Another option could be to go straight to the source in C++
https://github.com/kaldi-asr/kaldi

http://www.danielpovey.com/kaldi-lectures.html

Latest update from Rob (Gentle Dev) in regards to answering my question

> Hi Rob,
> Just wanted to circle back on this, It seems like the version 9.0.1 does not have corresponding up to date code in the github repo?
> Is that the case or just my impression?
>
> Coz if I run the terminal command I can only align audio to tex and I am required to pass a text file, and it is not possible to just transcribe something, which is possible through the server of version 9.0.1.
>
> Let me know if there's anything I can help to sort this out
>
> Best
>
> Pietro
> --------
>
> If you've downloaded the full language model, the example CURL command will do full transcription from the terminal, and it's trivial to adapt the code (pull requests welcome!) to support full transcription from the provided Python script.
> Gentle primarily is a forced aligner; the transcription is a "bonus."

Also worth checking out this fork of Gentle, see readme for how it differs from main branch
https://github.com/ronen/gentle

Something else worth considering is to speed up the transcription process similarly to Sam Lavine's approach of splitting into five minutes chunks getting then transcribe concurrently and then recombining it.
https://gist.github.com/pietrop/5008653567df73d813e525c6b89b23b6]
autoEdit is currently using a refactor of this in the transcriber module
https://github.com/OpenNewsLabs/autoEdit_2/tree/master/lib/interactive_transcription_generator/transcriber

https://github.com/alumae/kaldi-offline-transcriber

As Mark pointed out there is also some language model already available as open source

https://github.com/popuparchive/american-archive-kaldi

Refactor Notes 23Jan2017 - Pietro

on branch master

to try curl after `server.py` aligning

```
curl  -F 'audio=@examples/data/lucier.mp3' -F 'transcript=<examples/data/lucier.txt'
'http://localhost:8765/transcriptions?async=false'
```

to transcribe, in master branch, if you run does not do anything, need to test on `0.1.9`

```
curl  -F 'audio=@examples/data/lucier.mp3'  'http://localhost:8765/transcriptions?async=false'
```

python command line

```
python align.py examples/data/lucier.mp3 examples/data/lucier.txt
```

aligner work, but if you don't pass text file it says missing argument.
if you pass enty file eg `word.txt` like so `python align.py examples/data/lucier.mp3 word.txt`
then it returns `{}`.

Testing on `0.9.1` code.

works
```
curl  -F 'audio=@examples/data/lucier.mp3' -F 'transcript=<examples/data/lucier.txt'
'http://localhost:8765/transcriptions?async=false'
```

Works

```
python align.py examples/data/lucier.mp3 examples/data/lucier.txt
```

doesn’t work <———
```
curl  -F 'audio=@examples/data/lucier.mp3'  'http://localhost:8765/transcriptions?async=false'
```
However when testing os x version, see below

Testing with os x version

works Returns transcription
```
 curl  -F 'audio=@examples/data/lucier.mp3'  'http://localhost:8765/transcriptions?async=false'
```

CONCLUSION: This makes me think that, os x packaged version is not consistent with code of `0.9.1` version on github?

If in forced_aligner.py I add print words on line 24. After running from terminal

```
python align.py examples/data/lucier.mp3 examples/data/lucier.txt > test1.txt
```

I get something like this

```
[Word(duration=0.18 end=6.92 start=6.74 word=i), Word(duration=0.29 end=7.21 start=6.92 word=am), Word(duration=0.49 end=7.97 start=7.48 word=sitting), Word(duration=0.12 end=8.09 start=7.97 word=in), Word(duration=0.07 end=8.16 start=8.09 word=a), Word(duration=0.58 end=8.74 start=8.16 word=room), Word(duration=0.52 end=12.15 start=11.63 word=different), Word(duration=0.42 end=12.62 start=12.2 word=from), Word(duration=0.11 end=12.73 start=12.62 word=the), Word(duration=0.27 end=13.0 start=12.73 word=one), Word(duration=0.24 end=13.24 start=13 word=you), Word(duration=0.2 end=13.44 start=13.24 word=are), Word(duration=0.24 e
```

```

Which would sudgest that after initialising ` MultiThreadedTranscriber` and running it with `.transcribe(wavfile, progress_cb=progress_cb)` method it generates a transcription.

If I run this with no words inside of `lucier.txt`

```
python align.py examples/data/lucier.mp3 examples/data/lucier.txt > test1.txt
```

Then i get `{}`.

Also moved `lucier.wav` generated by system in example folder. That wav file is 8kb specs created by Gentle.
```