

Signal Processing

Speech Signal Processing

Speech Information Processing

Role of language

- **Semantics:** The meanings of words, and relations among them.
- **Syntax:** The order of words, role of function words.
- **Phonology:** Individual phonemic segments, features, stressed and unstressed vowels.
- For example,
 - What is the phonemic inventory of English?
 - How does it function? The concept of contrast (e.g., **pat** vs. **bat**).
 - Why do we believe that it is psychologically real?
 - Why does the same phoneme give rise to different, acoustic realizations in different utterances? (e.g., In fluent speech, "Joe ate his soup" loses the /h/ of "his", and the /t/ of "ate" doesn't look like a /t/ in "Tom".)
 - What are the principles that lead to modifications of segments in different environments?
 - How are phonemes usually described in terms of features translated into phonetic representations? (e.g., /z/ is + voiced, /s/ is - voiced; same relation for many pairs, like f-v; patterning of sounds is beautifully captured by feature concept.)

New terms that you will know..

Voiced, Unvoiced, pitch, intensity, timbre, formants, speech production, vocal tract, vocal cord, phonemes, manner and place of articulation, co-articulation, linear prediction, homo-morphic filtering, spectrograms, speech coding, speech enhancement, speech recognition, deep learning for speech, acoustic modelling, time/pitch scale modification, speech synthesis, Human auditory system, speech perception, speech quality measures (MOS, PESQ).....

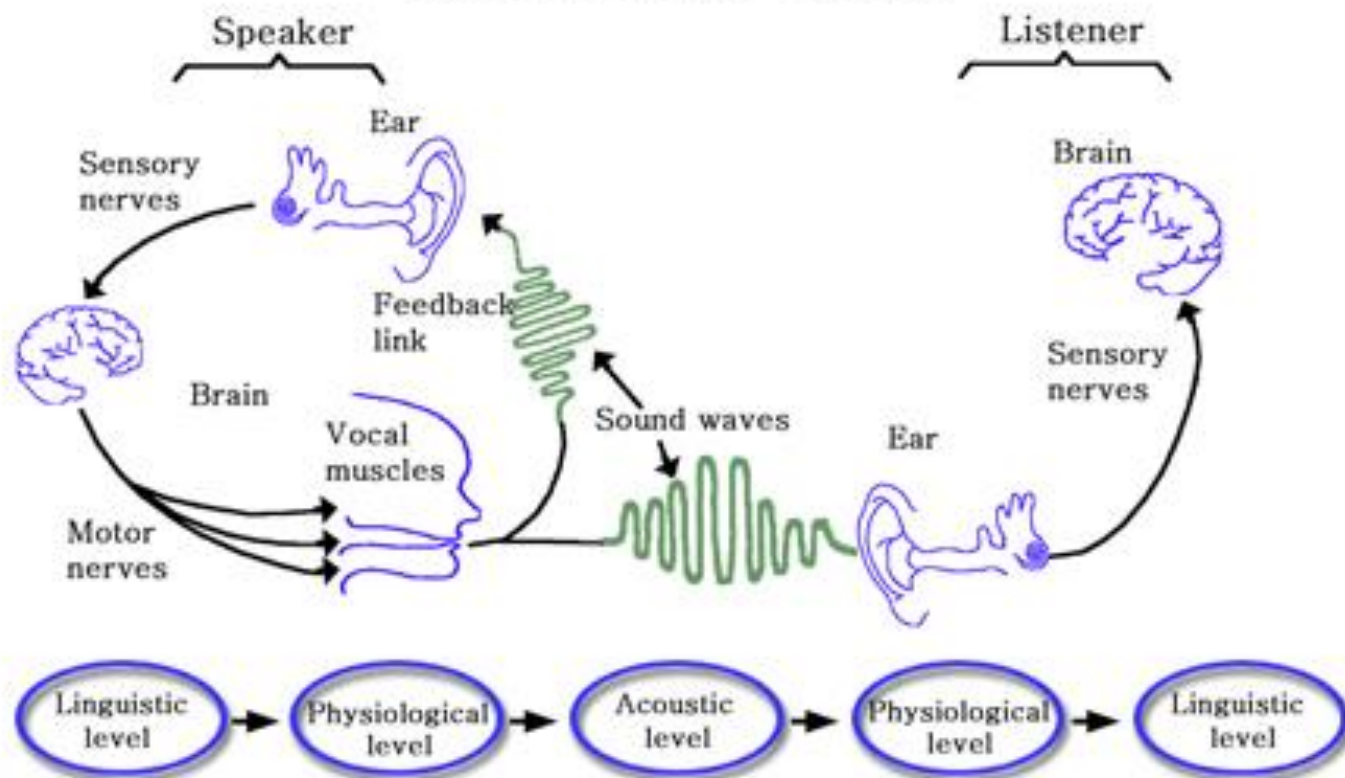
Speech Information Processing

Speech Information Understanding and Modeling

What is information in speech and
how it is encoded?

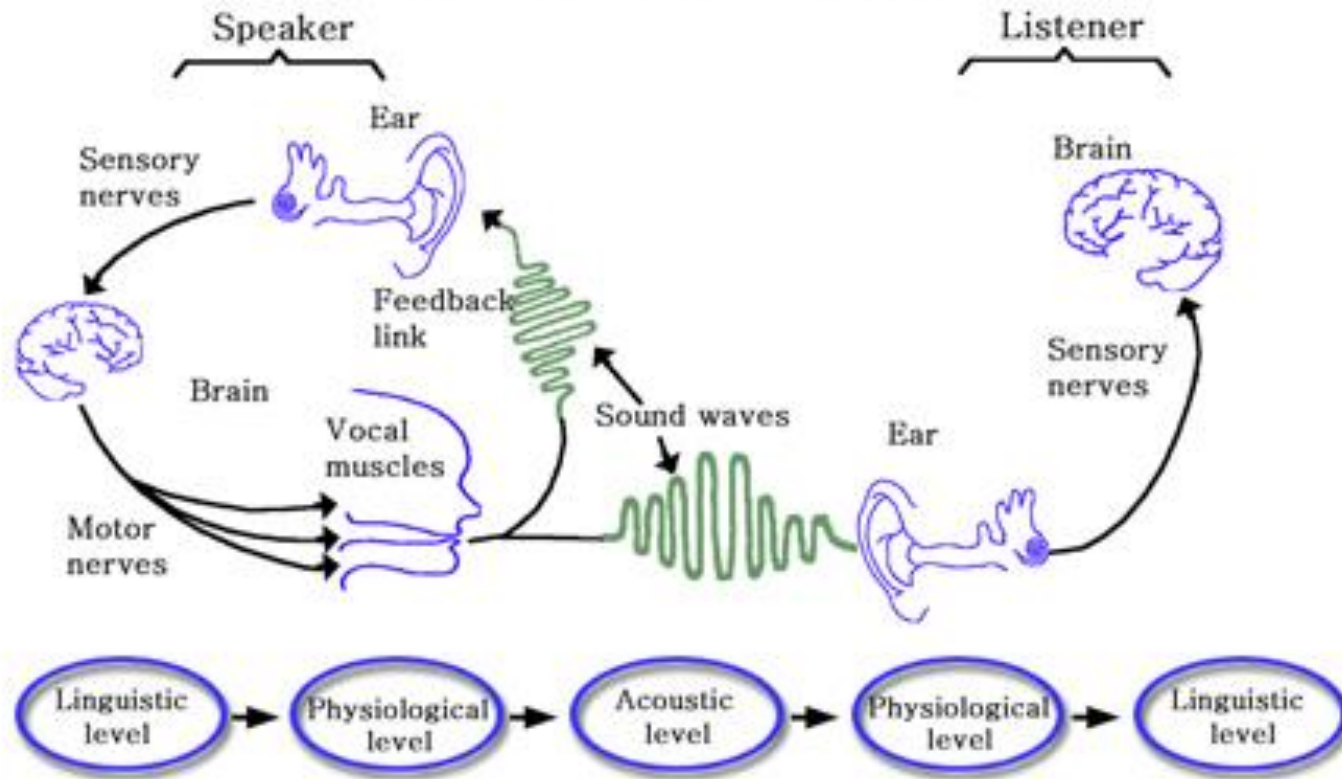
Lets give it a try

The Speech Chain



- Speech Research –
 - Speech Science
 - - Linguistics
 - - Physiology of Speech Production
 - - Acoustics
 - - Auditory Nervous System
 - - Psychophysics of Auditory System
 - - Cognitive Psychology
 - - Computer-based Algorithms
 - Speech Technology
 - Speech Recognition
 - Speaker Recognition
 - Speech Synthesis
 - Speech coding

The Speech Chain



Speech Coding

Speech Synthesis

Speech Recognition

Speech Understanding

Speaker Recognition

Language Recognition

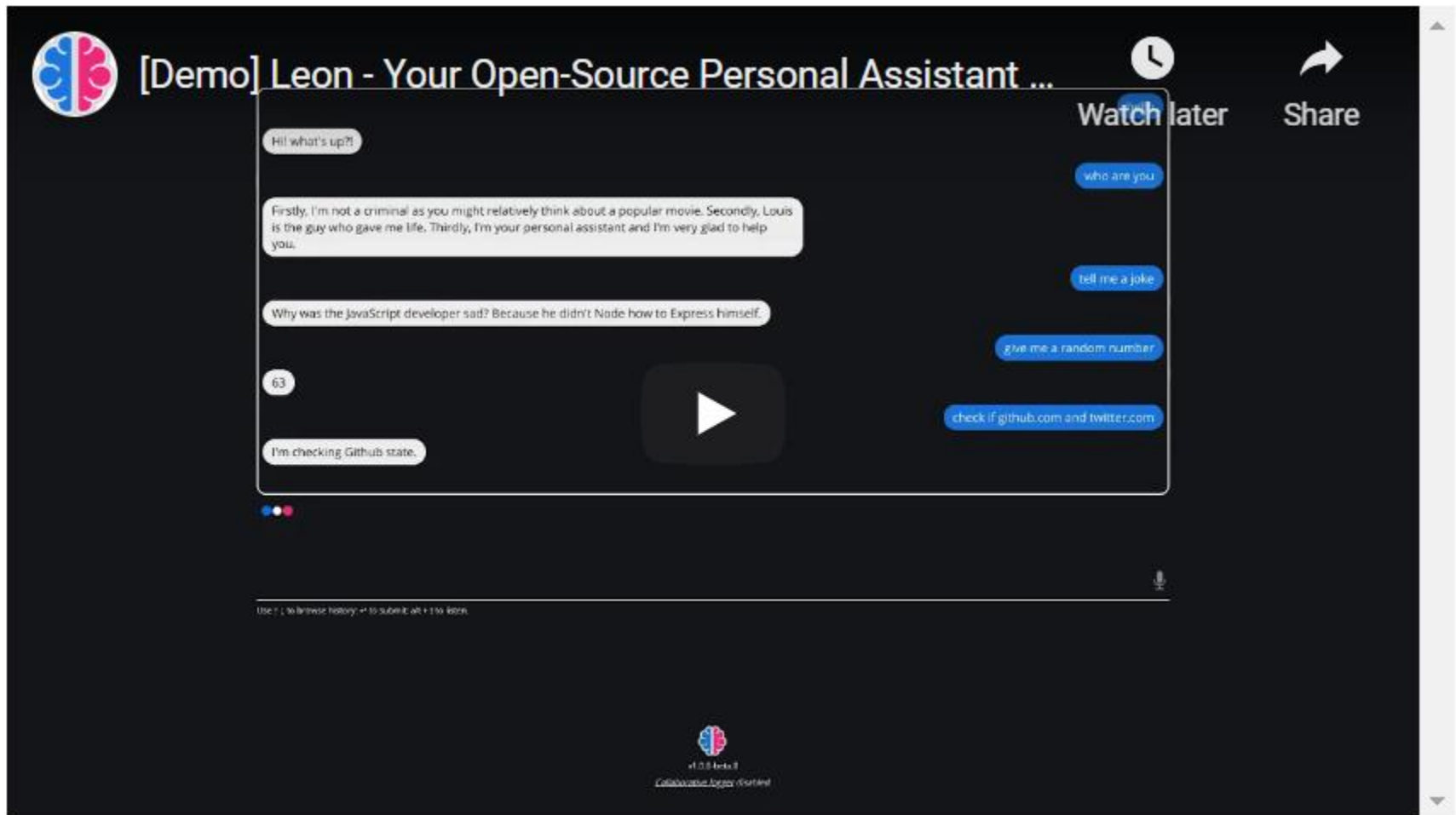
- Challenges to machine speech processing:
 - Definition of information content
 - Multiple levels of information
 - Subjectivity of the listener
 - Robustness to
 - Interfering signals
 - Partial information
 - Algorithmic complexity

Speech Recognition



Source: <https://mycroft.ai/blog/deepspeech-update/?cn-reloaded=1>

Speech Recognition



The screenshot displays the Leon AI demo interface. At the top left is a brain logo with blue and pink halves. The title "[Demo] Leon - Your Open-Source Personal Assistant ..." is followed by a clock icon, a "Watch later" button, and a "Share" button with a right-pointing arrow. The main chat area has a dark background. On the left, user messages are in white rounded rectangles: "Hi! what's up?!", "Firstly, I'm not a criminal as you might relatively think about a popular movie. Secondly, Louis is the guy who gave me life. Thirdly, I'm your personal assistant and I'm very glad to help you.", "Why was the JavaScript developer sad? Because he didn't Node how to Express himself.", "63", and "I'm checking Github state.". On the right, assistant responses are in blue rounded rectangles: "who are you?", "tell me a joke", "give me a random number", and "check if github.com and twitter.com". Below the chat is a video player with a large white play button. At the bottom left, there are three colored dots (blue, green, red) and a small text line: "Use F1 to browse history. ** to submit at 1 to 1000". At the bottom center is the brain logo, the text "v1.0.0 beta.0", and the text "Collaborative forges (outland)".

Source: <https://github.com/leon-ai/leon>

Speech Recognition

wav2letter

Facebook AI Research Automatic Speech Recognition Toolkit

Source: <https://www.techleer.com/articles/455-wav2letter-a-facebook-ai-researchfair-automatic-speech-recognition-toolkit/>

Speech Recognition

annyang! SpeechRecognition that just works

annyang is a tiny javascript library that lets your visitors control your site with voice commands. annyang supports multiple languages, has no dependencies, weighs just 2kb and is free to use.



Source: <https://github.com/TalAter/annyang>

Speech Recognition

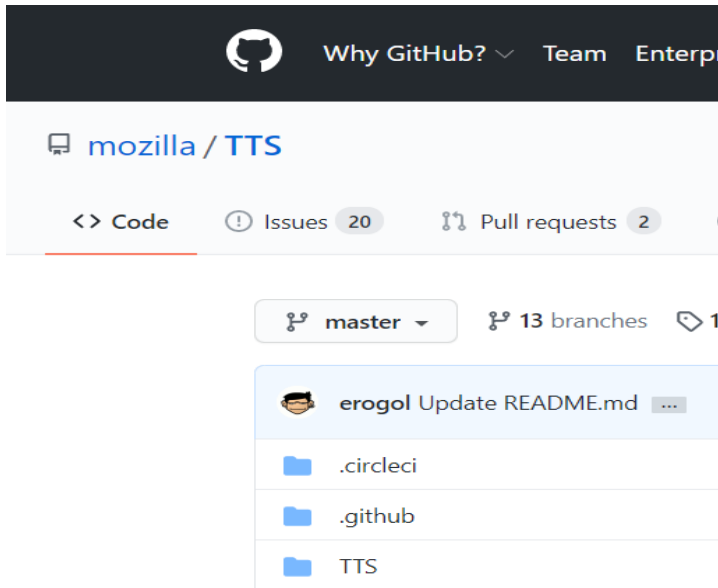
SpeechRecognition 3.8.1

```
pip install SpeechRecognition
```



Source: <https://pypi.org/project/SpeechRecognition/>

Text-to-Speech Synthesis



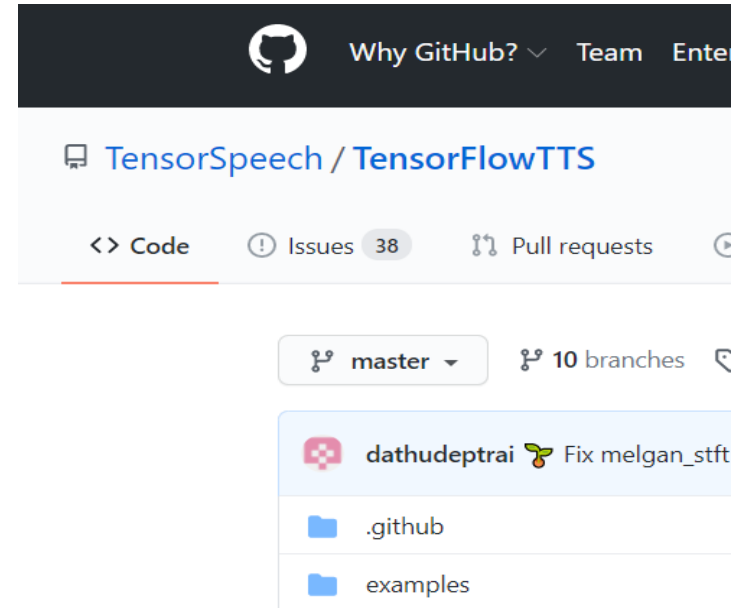
mozilla / TTS

<> Code Issues 20 Pull requests 2

master 13 branches

erogol Update README.md

- .circleci
- .github
- TTS



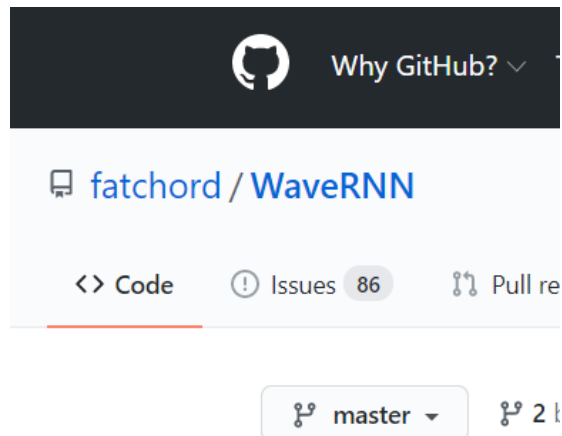
TensorSpeech / TensorFlowTTS

<> Code Issues 38 Pull requests

master 10 branches

dathudeptrai Fix melgan_stft

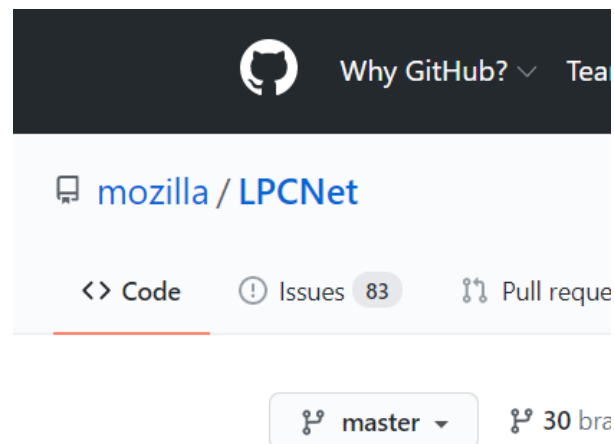
- .github
- examples



fatchord / WaveRNN

<> Code Issues 86 Pull requests

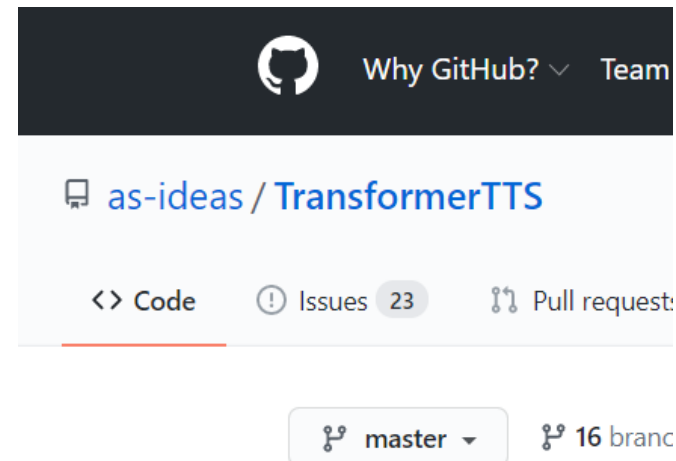
master 2 branches



mozilla / LPCNet

<> Code Issues 83 Pull requests

master 30 branches

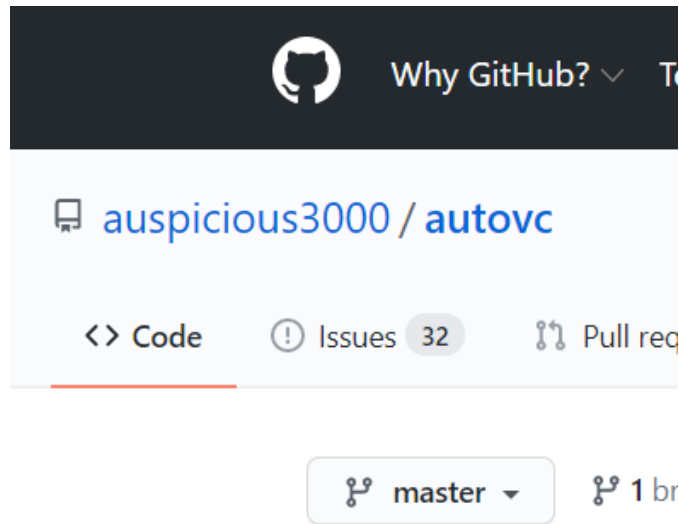


as-ideas / TransformerTTS

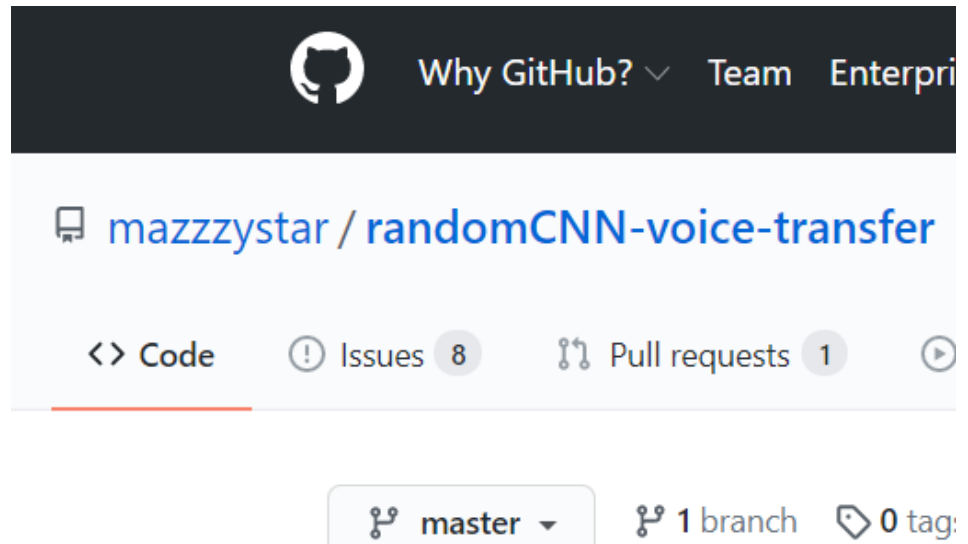
<> Code Issues 23 Pull requests

master 16 branches

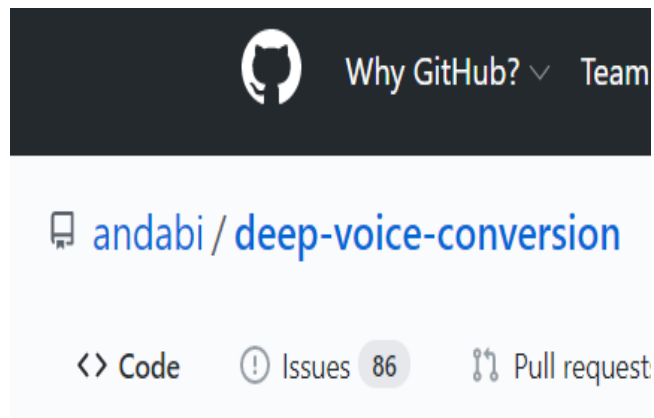
Voice conversion/ Style Transfer



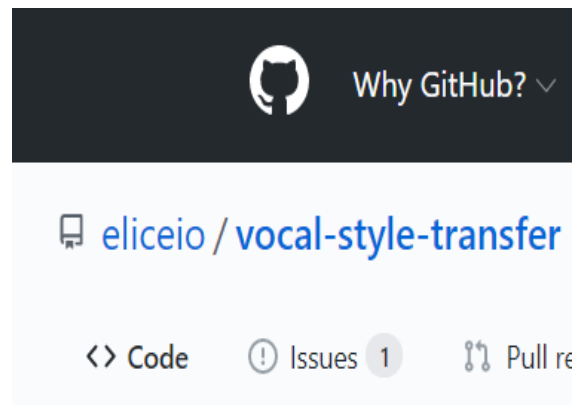
GitHub repository card for **auspicious3000 / autovc**. The card shows the repository name, a 'Code' button, and a badge for 32 issues. Below the repository name, it shows the 'master' branch and 1 branch.



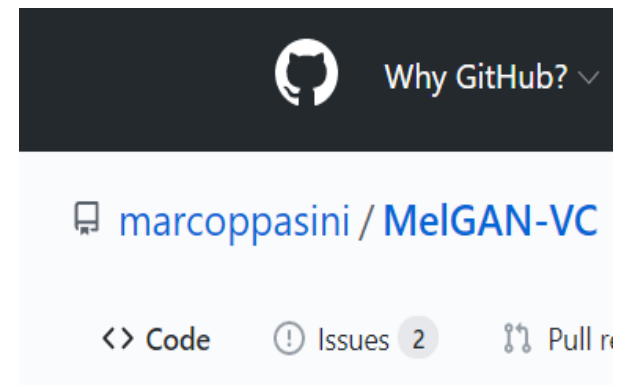
GitHub repository card for **mazzystar / randomCNN-voice-transfer**. The card shows the repository name, a 'Code' button, and badges for 8 issues and 1 pull request. Below the repository name, it shows the 'master' branch, 1 branch, and 0 tags.



GitHub repository card for **andabi / deep-voice-conversion**. The card shows the repository name, a 'Code' button, and a badge for 86 issues. Below the repository name, it shows the 'master' branch and 1 branch.

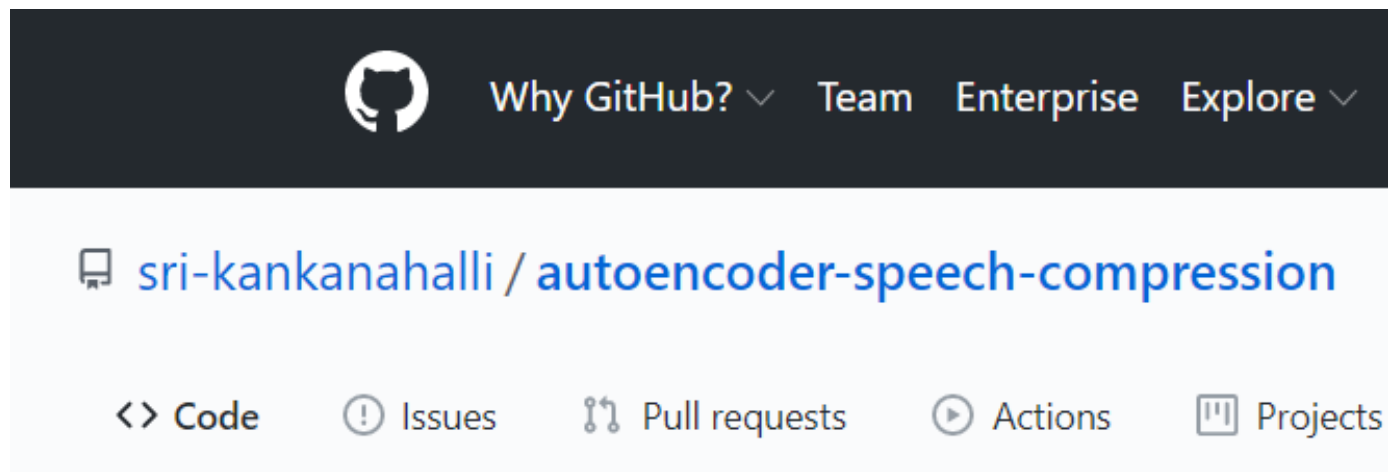
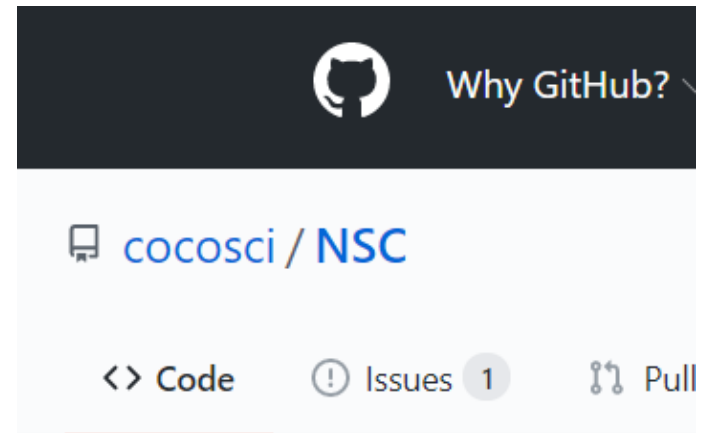
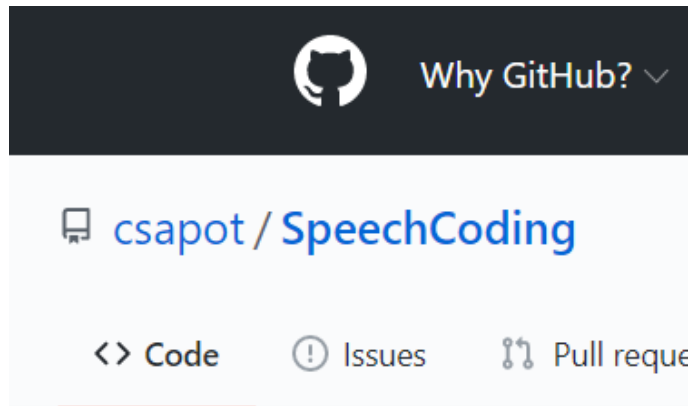


GitHub repository card for **eliceio / vocal-style-transfer**. The card shows the repository name, a 'Code' button, and a badge for 1 issue. Below the repository name, it shows the 'master' branch and 1 branch.

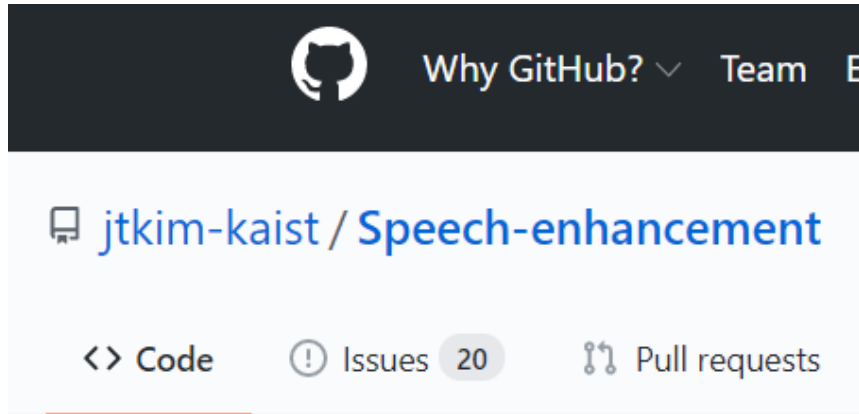


GitHub repository card for **marcoppasini / MelGAN-VC**. The card shows the repository name, a 'Code' button, and a badge for 2 issues. Below the repository name, it shows the 'master' branch and 1 branch.

Speech Coding



Speech Enhancement



And so on...

Voice activity detection

Speaker recognition

Speaker diarization

Emotion recognition

Conversational systems

Voice cloning

Pronunciation evaluation

Language identification

.....