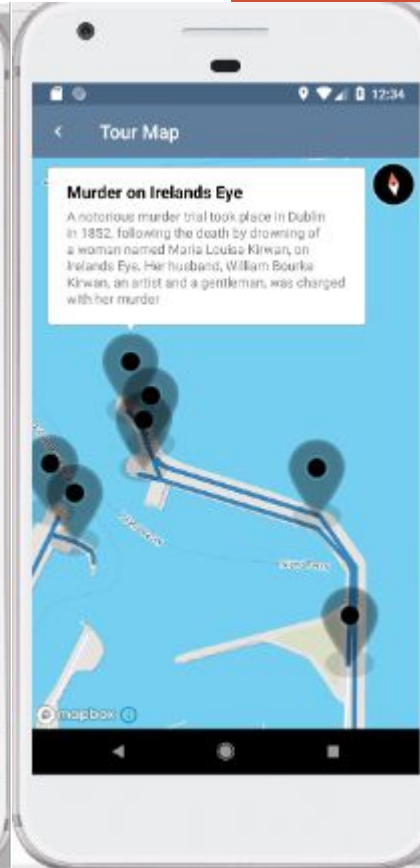
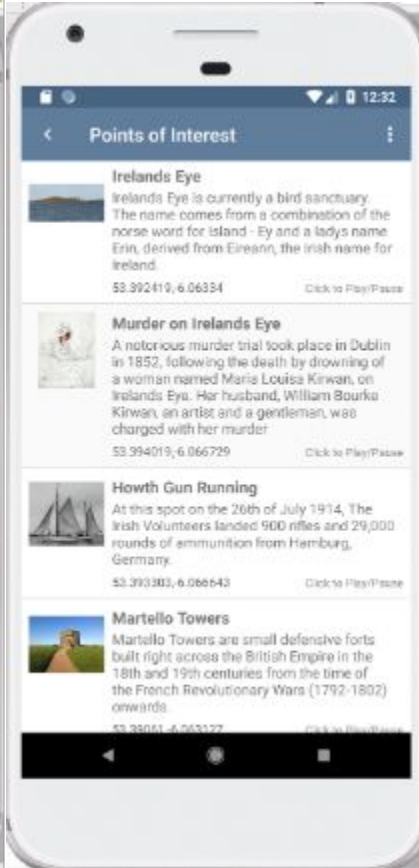




text-to-Speech and speech-to-text using Machine LEarning (SMILE)

Tara Hill National Park Teoranta
Flannan O Coileáin



Experiment Description

SMILE Motivation



Paradigm Shift from Static content Mode to Interactive content Mode

- **Recording Content with Human Actors**, including sourcing actors, recording facilities, sound engineers, etc. This has proven to be an extremely expensive and a very time consuming process.
- **Inability to Easily Modify Content after creation** when using the human actors. Text is provided to the actor and recording studio team, and THNP receives the finished product/experience as digital audio files. The digital audio content is static and very difficult, time consuming, and costly to change or update with new text and new digital audio content. For example, we recorded audio content with real actors stating a point of interest was on the left, when in fact, when the tourist arrived at the site, the point of interest was on their right hand side.
- **Providing Interactive Content to Tourist Users**, such as answering questions in real-time over Wi-Fi/4G/5G networks is impossible.



Experiment Setup and Project Results



Objective 2: DeepSpeech - Speech-to-Text (Appendix 3)



```
deepspeech --model deepspeech-0.9.3-models.pbmm --scorer deepspeech-0.9.3-models.scorer --audio dh_recordings/1.\wetransfer-3e6abe/1-Introduction.wav
```

```
(deepspeech-venv) jerryoc@ob2-speech-to-text-engine:~$ deepspeech --model deepspeech-0.9.3-models.pbmm --scorer deepspeech-0.9.3-models.scorer --audio 1-Introduction-16000hz.wav
```

```
Loading model from file deepspeech-0.9.3-models.pbmm
```

```
TensorFlow: v2.3.0-6-g23ad988
```

```
DeepSpeech: v0.9.3-0-gf2e9c85
```

```
Loaded model in 0.017s.
```

```
Loading scorer from files deepspeech-0.9.3-models.scorer
```

```
Loaded scorer in 0.000289s.
```

```
Running inference.
```

```
welcome to hot this is a town we connections to christianity by kings ghosts and fairies castles a murder wariious bands and productions such as you to and rivermen of course the nobel literate bennet a few it's a magical place soogans and getting here my name is mister stewart i will be a guide for the next few hours i've been in this area for a long long time so i will share some of the more interesting stories about this place with you before we begin the two or annette a few agents out of the way first remember this is the working harbor you need to be very aware of your surroundings and toting traffic other people and harbour edges into the sea this can be a very dangerous place second followed root on the map to each point of interest i have already decided the best route to take on your tour based on your carnoustie once we are close enough to each point of interest i will start telling you about it fieldrations or changed erection during the tour i will do my best to reorganize on your behalf now relax and enjoy decides that are front of you and beautiful hope
```

```
Inference took 63.313s for 82.000s audio file.
```

Speech Processing Tool Developed (Appendix 2)



Preprocessing Howth Audio Dataset for Text-to-Speech (TTS) training

```
jerryoc@ob2-speech-to-text-engine:~/audio-slience-splitter$ python3
```

```
process-raw-audio-totext-for-training.py /home/jerryoc/dp-DH-recordings-original-wavs/original
```

```
From: /home/jerryoc/dp-DH-recordings-original-wavs/original/11-Murder-on-Irelands-Eye.wav
```

```
To: /home/jerryoc/dp-DH-recordings-converted-wavs/converted/11-Murder-on-Irelands-Eye.wav
```

```
11-Murder-on-Irelands-Eye0|murder on Ireland's eye.
```

```
11-Murder-on-Irelands-Eye1|notorious murder trial took place in dublin in  
eighteen fifty two.
```

```
11-Murder-on-Irelands-Eye2|following the death by drowning of a woman named  
maria louisa carwin arline.
```

```
11-Murder-on-Irelands-Eye3|her husband william berg carwin and arisen a  
gentleman was charged with her murder
```

```
11-Murder-on-Irelands-Eye4|on the day of the murder september six eighteen fifty  
two william e maria were rowed out to the island early by four fishermen.
```

```
11-Murder-on-Irelands-Eye5|they arrived around ten.
```

Dataset Editing



The following shows a sample of data returned from the Speech Processing API developed, which is in a format suitable for processing by the Text-to-Speech ML libraries.

```
jerryoc@ob2-speech-to-text-engine:~/audio-slience-splitter$ python3 process-raw-audio-totext-for-training.py
/home/jerryoc/dp-DH-recordings-original-wavs/original
From: /home/jerryoc/dp-DH-recordings-original-wavs/original/11-Murder-on-Irelands-Eye.wav
To: /home/jerryoc/dp-DH-recordings-converted-wavs/converted/11-Murder-on-Irelands-Eye.wav
11-Murder-on-Irelands-Eye0|murder on Ireland's eye.
11-Murder-on-Irelands-Eye1|notorious murder trial took place in dublin in eighteen fifty two.
11-Murder-on-Irelands-Eye2|following the death by drowning of a woman named maria louisa carwin
arline.
```

A completed data sample utilised for training TTS is available after editing

```
11-Murder-on-Irelands-Eye0|Murder on Ireland's eye.
11-Murder-on-Irelands-Eye1|A Notorious murder trial took place in dublin in eighteen fifty two.
11-Murder-on-Irelands-Eye2|Following the death by drowning of a woman named Maria Louisa Kirwan on
Ireland's eye.
11-Murder-on-Irelands-Eye3|Her husband William burke Kirwan an artist and a gentleman was charged
with her murder.
11-Murder-on-Irelands-Eye4|On the day of the murder september sixth eighteen fifty two William and
Maria were rowed out to the Island early by four fishermen.
11-Murder-on-Irelands-Eye5|They arrived around ten a.m.
```

Text-to-Speech Model Training



Start training using the Glow-TTS speech DNN

```
export PYTHONPATH="${PYTHONPATH}:/home/scamallra/6.dp-dh-recordings/mozilla-tts"  
#Kick off the GlowTTS Model Training  
/usr/bin/python /home/scamallra/6.dp-dh-recordings/mozilla-tts/TTS/bin/train_glow_tts.py  
--config_path /home/scamallra/6.dp-dh-recordings/5.dp-models/dp_model-glow_tts.json
```

Instructions to train the THNP TTS multiband-melgan Vocoder model

```
export PYTHONPATH="${PYTHONPATH}:/home/scamallra/6.dp-dh-recordings/mozilla-tts"  
CUDA_VISIBLE_DEVICES="1" python /usr/bin/python3  
/home/ubuntu/6.dp-dh-recordings/mozilla-tts/TTS/bin/train_vocoder_gan.py --config_path  
/home/ubuntu/6.dp-dh-recordings/5.dp-models/irish-m2-vocoder-model-glow_tts.json
```


Text-to-Speech Example using Models produced

Text-to-Speech combining acoustic and vocoder models produced

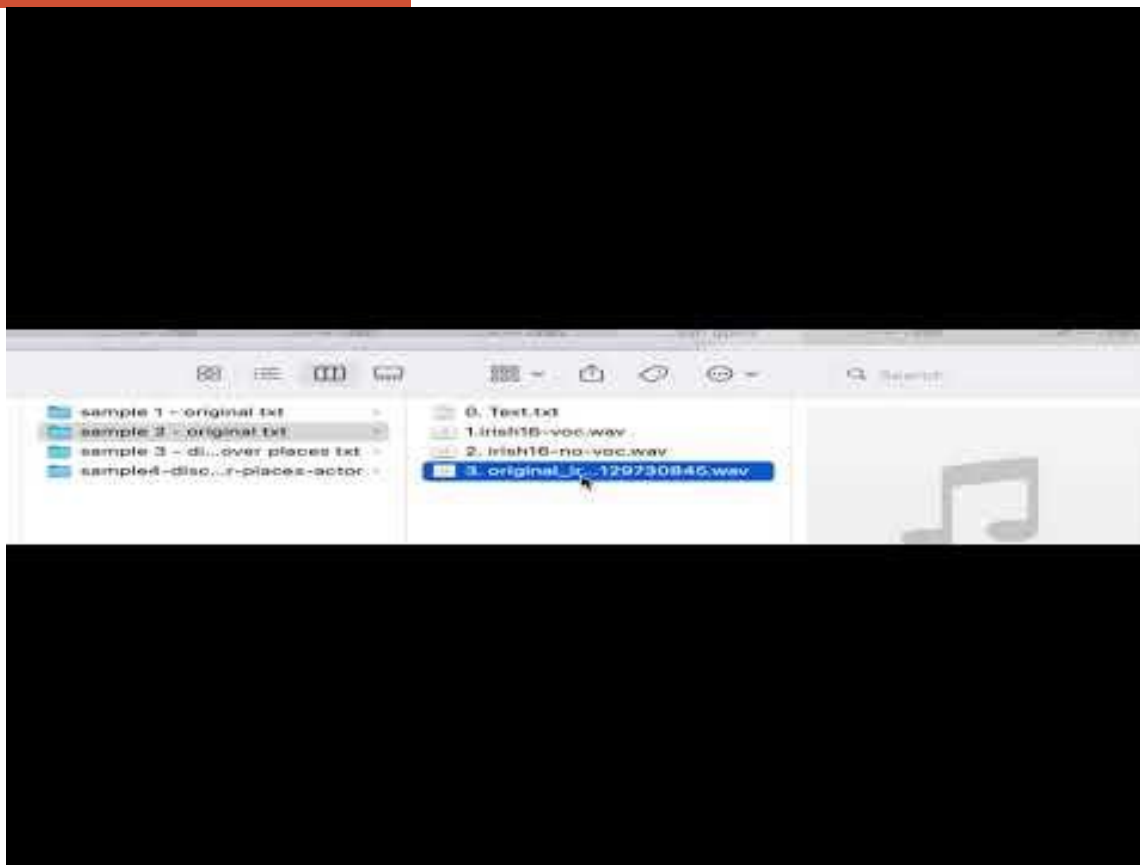
####To run your own TTS and Vocoder models trained, such as the THNP developed models, use the following format:

```
tts --text "Text for TTS" \  
  --model_path path/to/model.pth.tar \  
  --config_path path/to/config.json \  
  --out_path output/path/speech.wav \  
  --vocoder_path path/to/vocoder.pth.tar \  
  --vocoder_config_path path/to/vocoder_config.json
```

####For example, use model and vocoder trained based on THNP actor voice.

```
tts --text "William planned to spend the day painting, while maria intended to do some sunbathing and swimming." --model_path  
/home/ubuntu/6.dp-dh-recordings/6.output/glow-tts-rel_pos_transformer-December-05-2021_12+53PM-00000  
00/checkpoint_80000.pth.tar --config_path  
/home/ubuntu/6.dp-dh-recordings/6.output/glow-tts-rel_pos_transformer-December-05-2021_12+53PM-00000  
00/config.json --vocoder_path /home/ubuntu/dp_checkpoint_vocoder/checkpoint_1000000.pth.tar  
--vocoder_config_path /home/ubuntu/dp_checkpoint_vocoder/config.json
```

Objective 1 DEMO



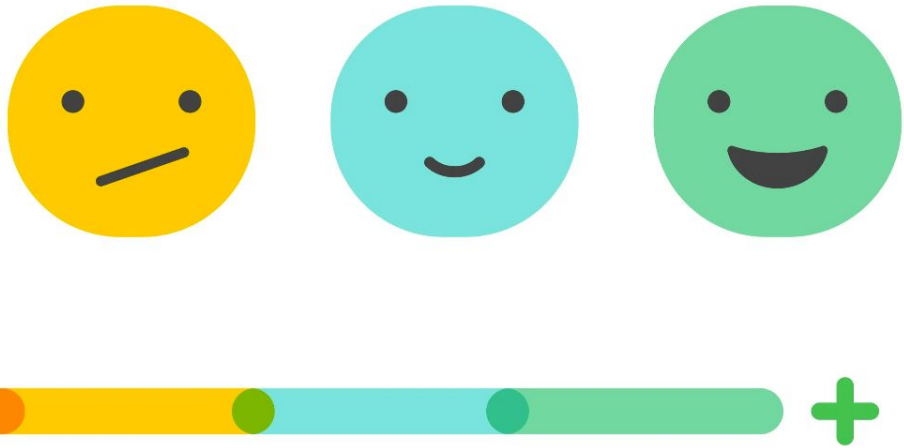
SMILE Mean Opinion Score (MOS) Results



Question	Acoustic and vocoder DNN models (mos)	Acoustic DNN model only (mos) (i.e., no vocoder)
The mean score of whether participants would listen to a given voice again (0–5)	2.5	2.5
mean quality (0–5)	2.125	3.125
mean clarity (0–5)	2.1875	3.375
mean comprehension grade (0–5)	3	4
Mean Opinion Score (MOS)	2.45	3.25



Business Impact



Feedback



Co-funded by the
European Union



Co-funded by the
Swiss Confederation

This project has received funding from the European Union's Horizon 2020 research and innovation programme, which is co-funded by the European Commission and the Swiss State Secretariat for Education, Research and Innovation, under grant agreement No 732638.

SMILE

WWW.FED4FIRE.EU