ملخص البحث للأسبوع الأول

بعد القيام بعملية بحث علمي لايجاد TtS models وجدنا الmodels التالية

# 1. Tacotron 2

• Advantages:

* Produces high-quality, natural-sounding speech.
* Capable of expressing various emotions and speaking styles.
* Tacotron 2's end-to-end training simplifies the model pipeline.

• Disadvantages:

* Requires a large dataset for training to achieve the best results.
* Training can be computationally intensive and time-consuming.

2. WaveNet

• Advantages:

* Generates audio samples directly, leading to highly natural and realistic sounding voice.
* Capable of capturing subtle nuances of the human voice.
* Has been used successfully in multiple languages, indicating potential for Arabic.

• Disadvantages:

* Very computationally intensive for both training and inference, requiring specialized hardware.
* Like Tacotron 2, it demands large amounts of high-quality training data.
* Real-time inference might require optimization or specialized hardware.

# 3. FastSpeech 2

# 

• Advantages:

* Offers faster speech generation compared to autoregressive models like Tacotron 2 and WaveNet, making it more suitable for real-time applications.
* Reduces the dependency on GPU-intensive computations for inference.
* Easier to train with less data than some other models.

• Disadvantages:

* Might not achieve the same level of naturalness and emotional variance as WaveNet or Tacotron 2.
* The quality of speech synthesis is highly dependent on the quality of the dataset.

# 4. Transformer TTS

• Advantages:

* + Utilizes the Transformer architecture, known for its efficiency and effectiveness in handling sequential data.
  + Can potentially reduce training time with parallel processing.
  + Capable of generating high-quality, natural-sounding speech.

• Disadvantages:

* + Like other neural models, it requires a substantial amount of training data for best performance.
  + May require significant computational resources for training.

# 5. ESPnet-TTS

• Advantages:

* + An open-source, all-in-one toolkit that simplifies the process of experimenting with different TTS models, including Transformer and Tacotron 2.
  + Supports multi-language TTS, including potential for Arabic with the right datasets.
  + Offers a flexible and extensible framework for research and development.

• Disadvantages:

* + Being a toolkit rather than a model, it requires more setup and configuration effort.
  + The performance is dependent on the underlying model and dataset used.

# نتائج البحث:

1. بعد التفكير والمناقشة قمنا باختيار ESPNET بسبب حاجتنا الى التخصيص (Customization) والدعم للغة العربية و ايضا لانها توفر بيئة عمل تسمح لنا من خلالها تجريب و استخدام عدة models لنقوم بالأختيار الأخير.
2. قمنا بالبحث عن Dataset مناسبة ووجدنا واحدة تحتوي على اللغة العربية وأيضاً اللهجة السورية الشامية.
3. نقوم بتثبيت تلك الأدوات للبدء بعملية بناء الموديل المرغوب به.

links:

1. Tactron 2:<https://pytorch.org/hub/nvidia_deeplearningexamples_tacotron2/>
2. Wavenet:<https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/>
3. Fastspeech2: <https://github.com/ming024/FastSpeech2>
4. Transformer TTS: <https://github.com/as-ideas/TransformerTTS>
5. ESPnet: <https://github.com/espnet/espnet>
6. Arabic graphemes and phonemes dataset: <https://huggingface.co/datasets/arabic_speech_corpus>