Databases for speech synthesis

- Group activity: design the script to be recorded

Group activity: design the script to be recorded

- Step 1
  - find a **source** of text
  - things to consider include: copyright, domain, size, readability, NSWs, ...
- Step 2
  - **clean** the text
  - things to consider include: vocabulary, readability, normalisation,…
- Step 3
  - design a simple '**richness**' measure
  - write this as a function that computes a score for one sentence

Group activity: design the script to be recorded

- But first...
  - Why are we building a synthetic voice?

- Think about
  - What will it be used for?
  - What sort of things does it need to say?
  - Who will listen to it?
  - *plus various technical requirements (computation, memory, platform, latency,...)*

the "use case"

Example use cases

Summary of today’s class: what we learned about data

- Define the use case first

- A recipe for creating a dataset
  - Identify a source of data
  - Obtain (a lot) more data than we require
  - Curate the data
    - define “good data”
    - filter out “bad” data (simplest: discard it ; alternatively: fix/repair)
    - select a dataset of the desired size

- We only considered text data, but the recipe works for speech too (later in the course)

A look forward to neural approaches

- is data selection still relevant?

Data requirements for neural approaches

- In simple terms
  - many neural models need a lot more speech data than we could record in the studio
  - *not true for all models (e.g., the one we are using in the assignment)*

- So, for data-hungry models we have no choice but to
  - Find pre-existing speech
  - Automatically curate a dataset from that
  - **IMPORTANT: do not do this for the assignment!**
    - You must *only* use purposely-recorded speech (your own + corpora approved by us)

Data requirements for neural approaches

- Is data selection still relevant?
- Yes, in at least two ways:

1. Large-scale curation of 'found' speech data still involves selecting "good" data
   - may lack reliable transcription, other labels, and meta-data

2. Purposely-recorded speech is still useful (in fact, *essential* for most commercial products)
   - generally much higher quality
   - we can control everything: speaker identity, content, speaking style, ...