

How To Create Text-To-Speech Audio Files: A Practical Step-By-Step Course For Beginners



Introduction To Text-To-Speech

- Background & Brief History Of Text-To-Speech.
- Popular Text-To-Speech Engines.
- Text-To-Speech Terms.

A Brief History Of Text-To-Speech

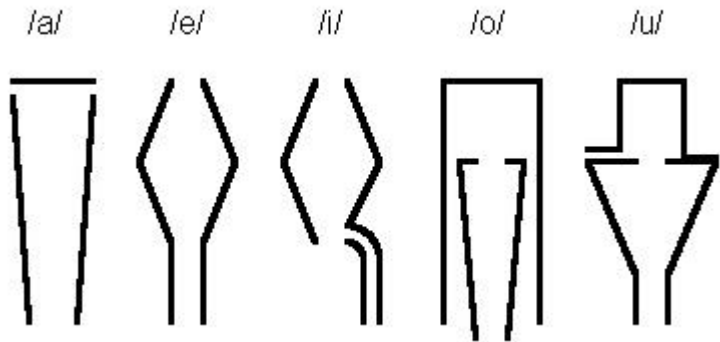


A Brief History Of Text-To-Speech

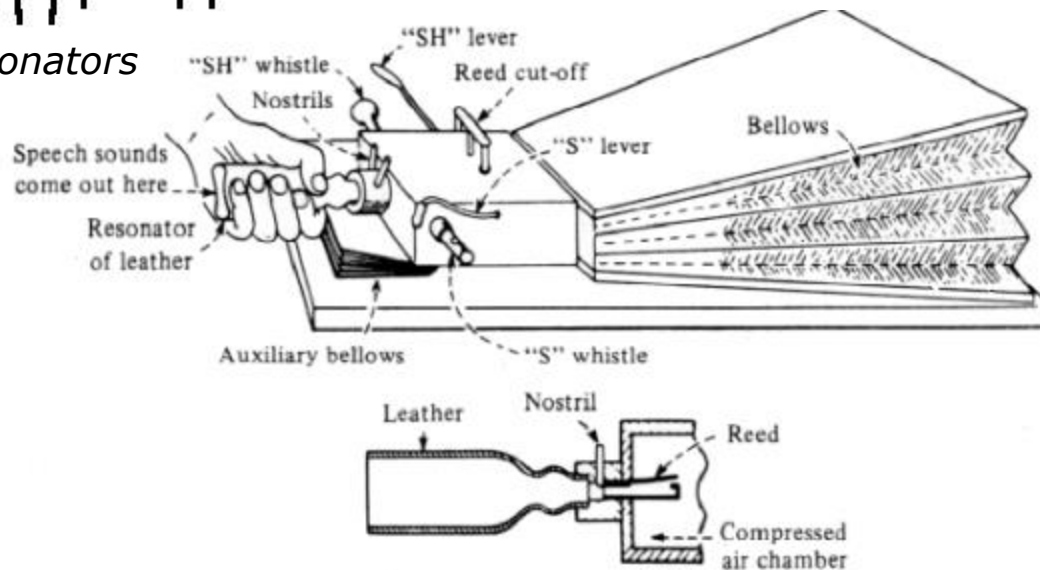


https://en.m.wikipedia.org/wiki/Brazen_head

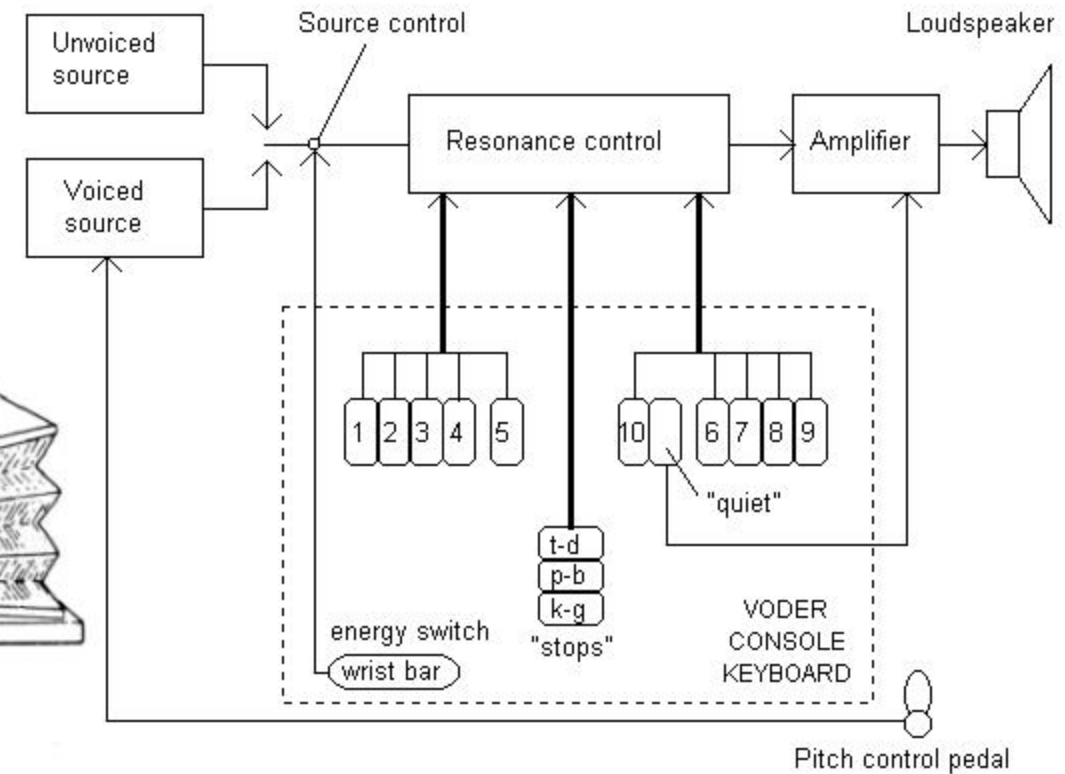
A Brief History Of Text-To-Speech



Kratzenstein's resonators



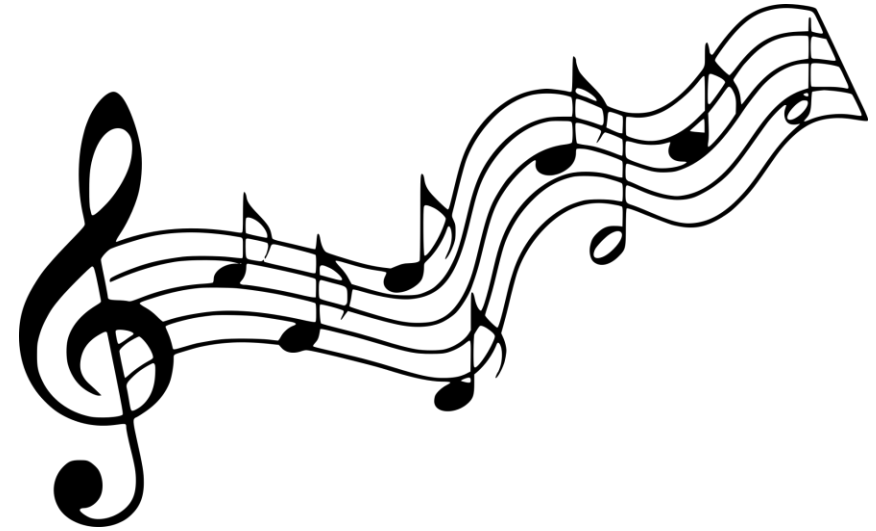
von Kempelen's speaking machine



VODER speech synthesizer

http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html

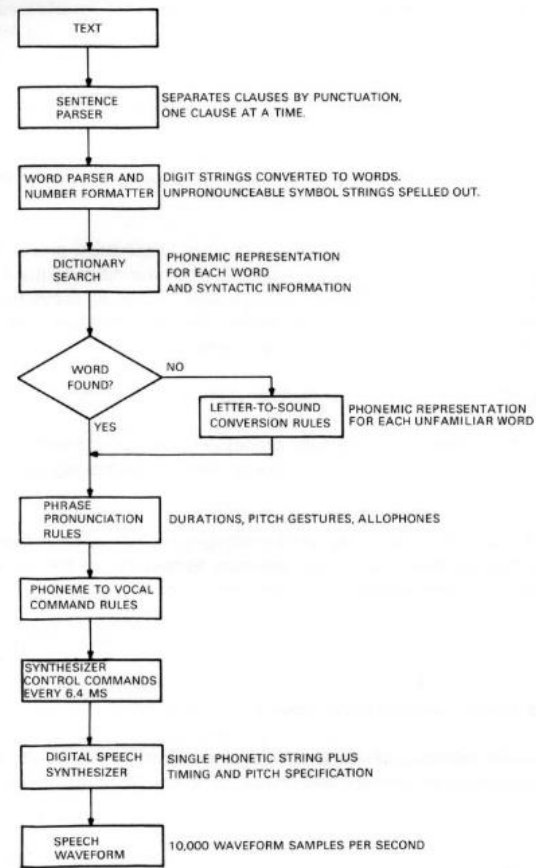
A Brief History Of Text-To-Speech



A Brief History Of Text-To-Speech

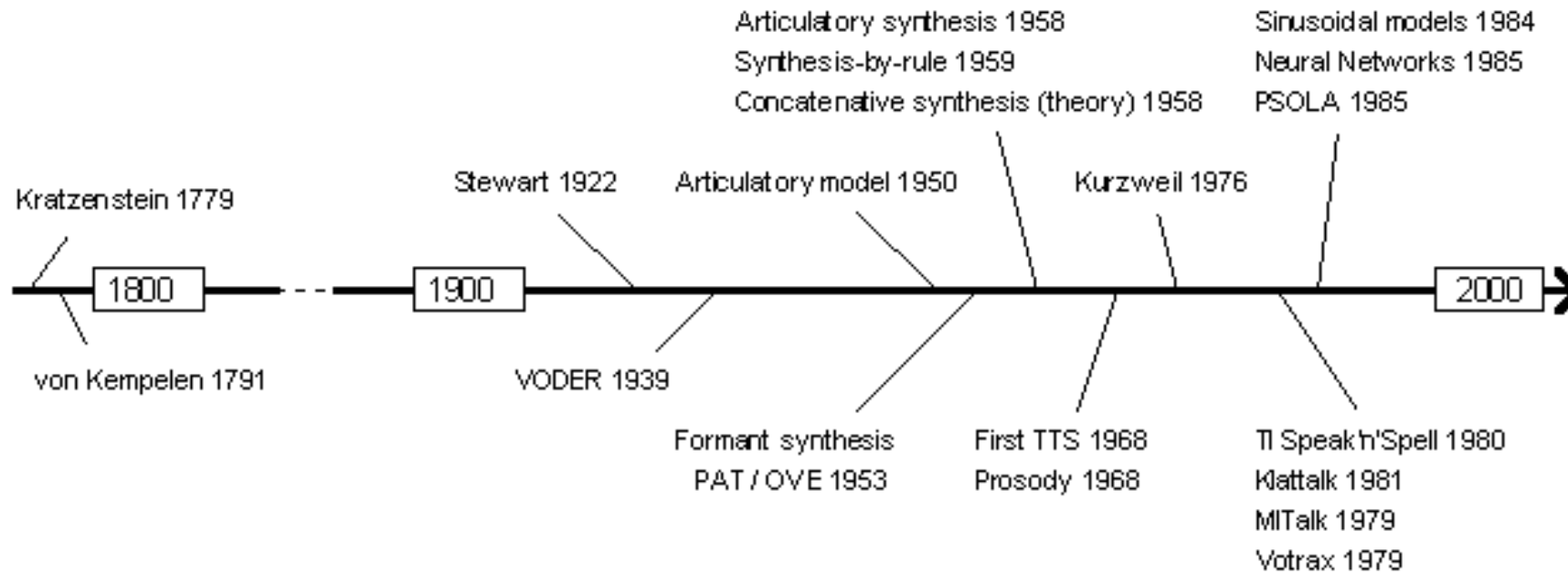


Figure 5-1 The DECtalk Voices

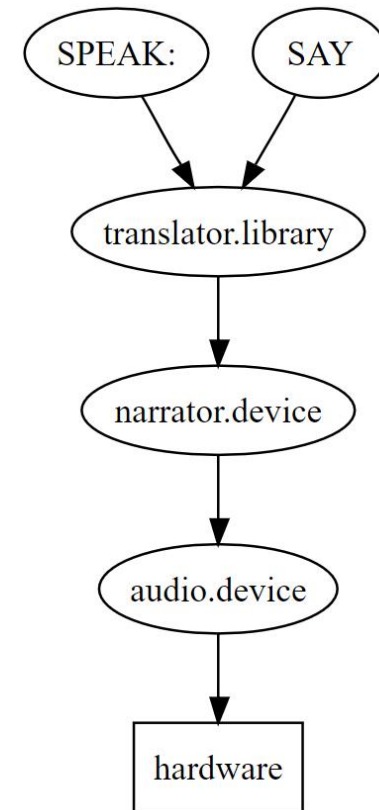


<https://archive.org/details/dectalk>

A Brief History Of Text-To-Speech



Milestones In Speech Synthesis



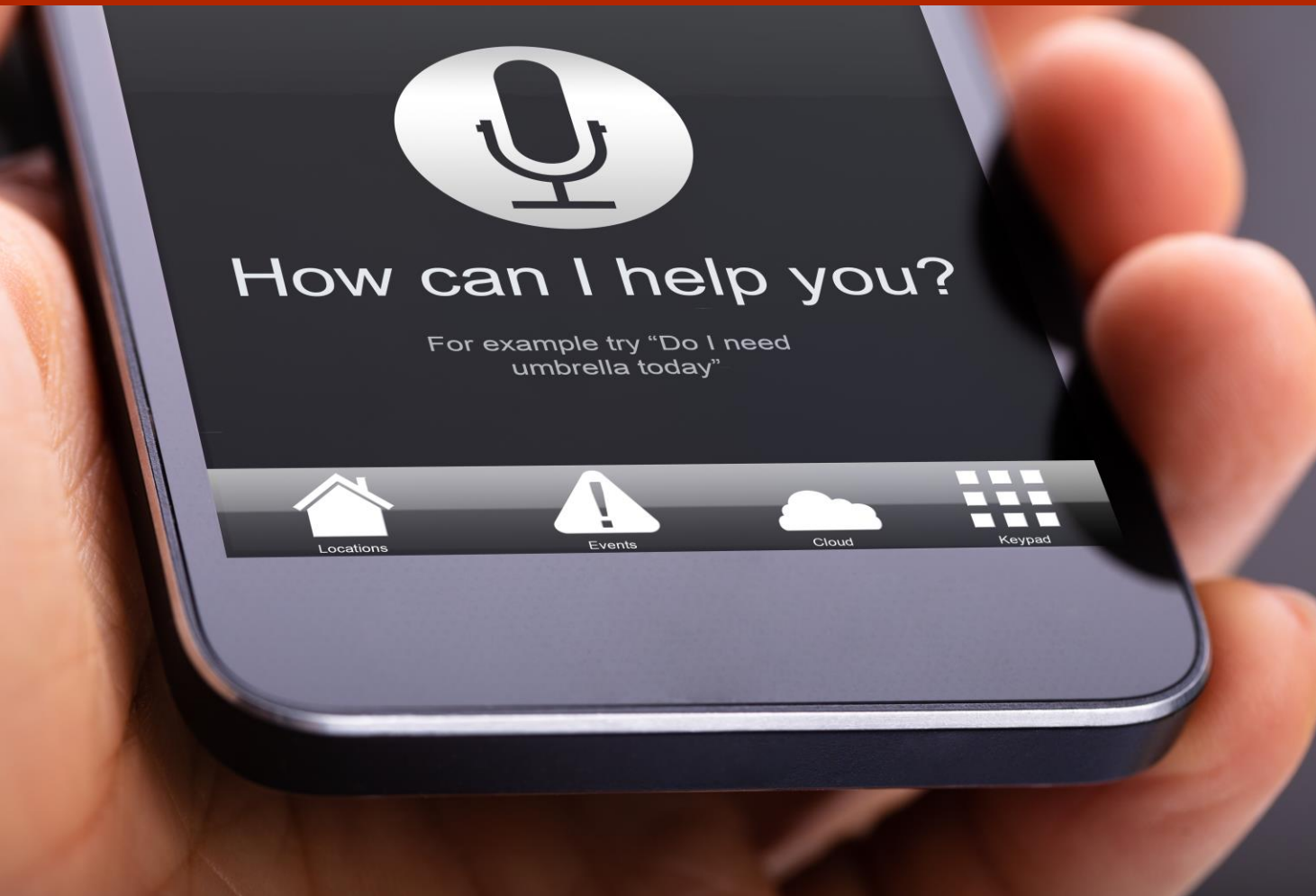
Amiga SoftVoice speech synthesis

<https://commons.wikimedia.org/wiki/User:Polluks>

A Brief History Of Text-To-Speech



A Brief History Of Text-To-Speech



<https://www.shutterstock.com>

A Brief History Of Text-To-Speech



A Brief History Of Text-To-Speech

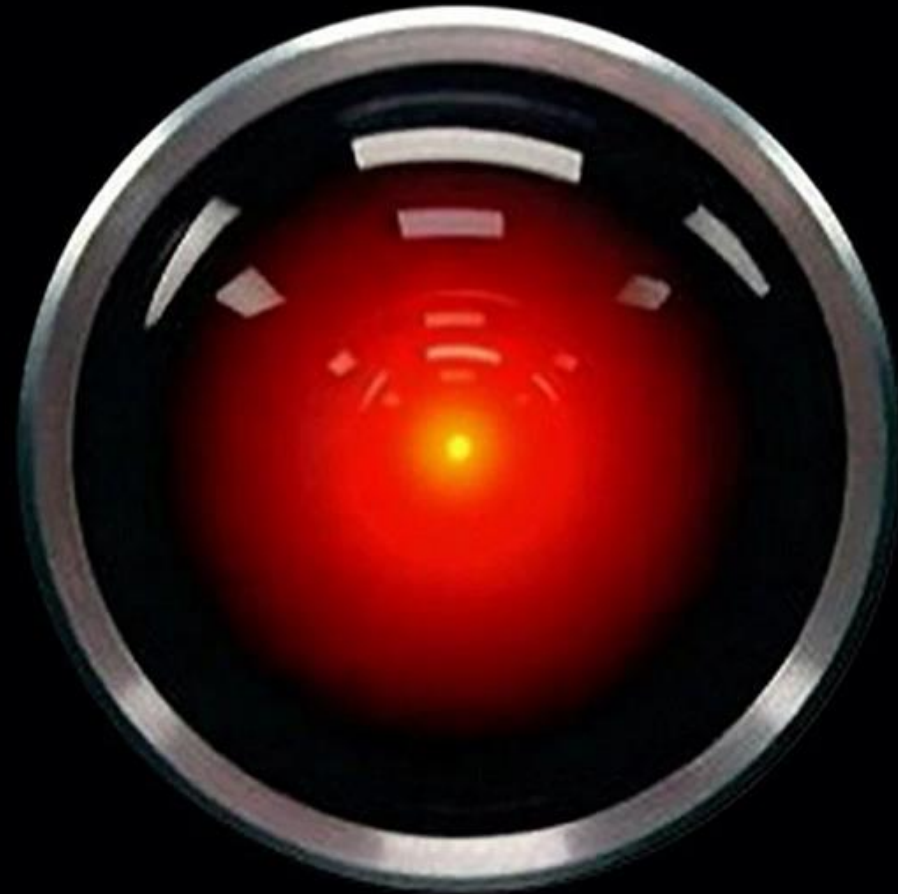


A Brief History Of Text-To-Speech



(The Office - Season 3, Episode 13 Travelling Salesman)

A Brief History Of Text-To-Speech



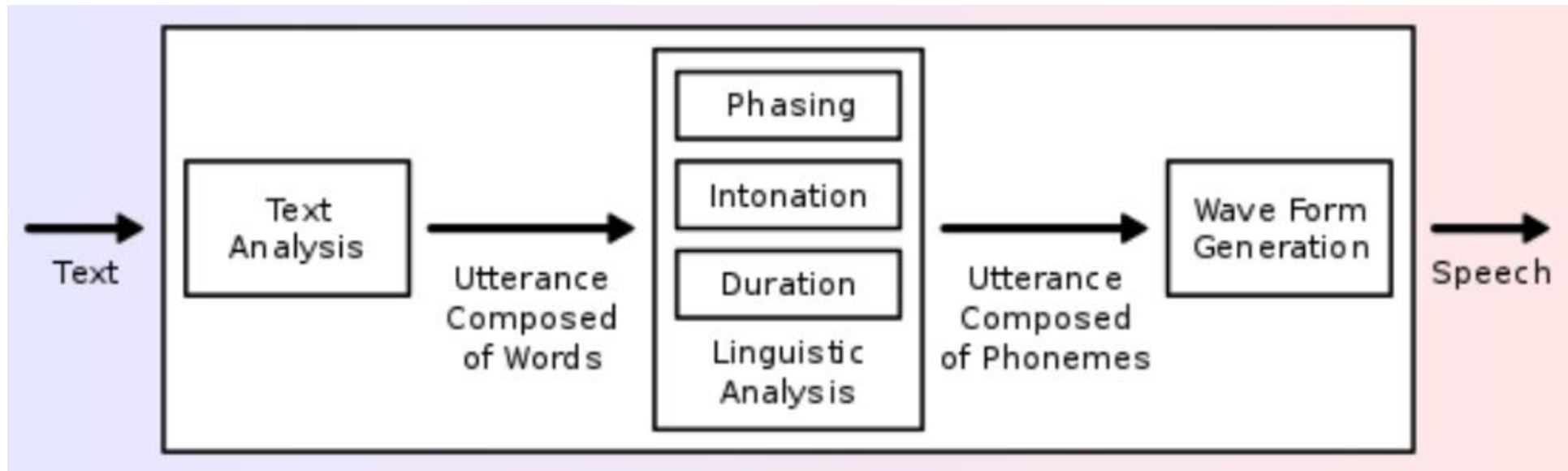
A Brief History Of Text-To-Speech



Text-To-Speech (TTS) Technologies

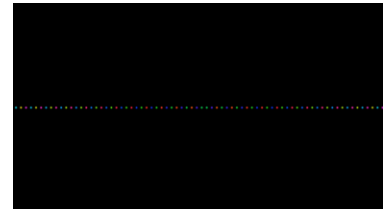
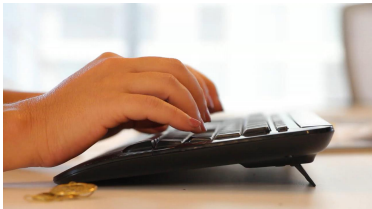


Text-To-Speech (TTS) Technologies



Overview Of A Typical Text-To-Speech System

https://en.m.wikipedia.org/wiki/Speech_synthesis



Text-To-Speech (TTS) Technologies

Qualities Of A Great Speech Synthesis System

- **NATURALNESS** - how closely the synthetic generated voice sounds like human speech.
- **INTELLIGIBILITY** - how easily the speech can be understood.



The ideal speech synthesizer aims to sound as natural and intelligible as possible!

Text-To-Speech (TTS) Technologies

Concatenative Speech Synthesis

- A very large database of short speech fragments (units) are recorded from a single speaker and recombined to form complete utterances.
- Stringing segments of recorded speech together.
- **PROS:** Produces natural-sounding synthesized speech.
- **CONS:** It's difficult to modify the voice (for example switching to a different speaker or altering the emphasis or emotion of their speech) without recording a whole new database.

Many Text-To-Speech Applications Use Concatenative Synthesis

Text-To-Speech (TTS) Technologies

Parametric Speech Synthesis

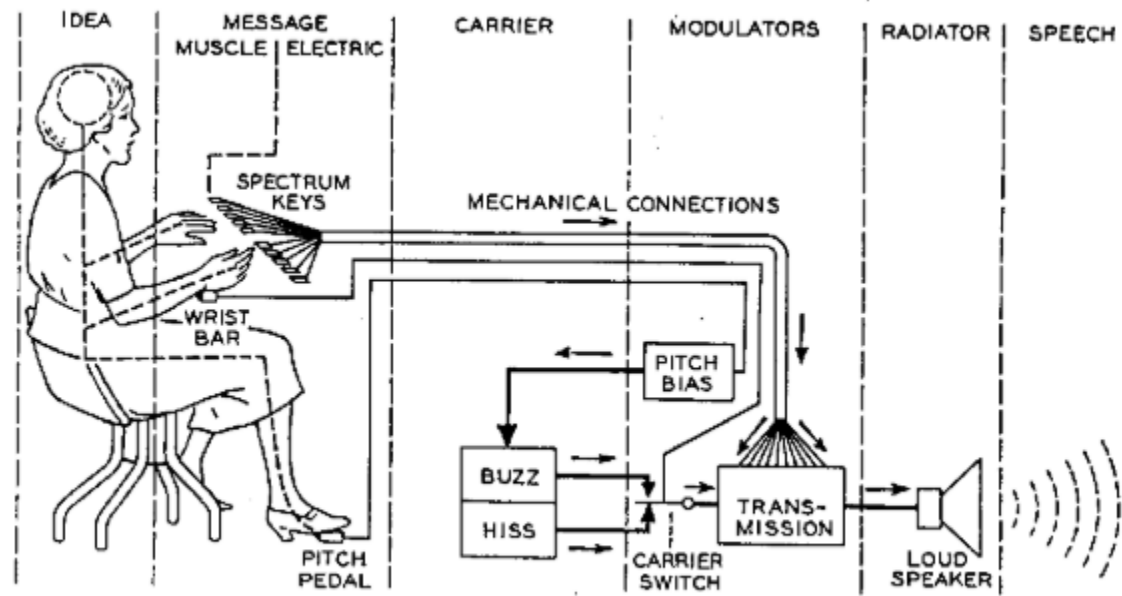
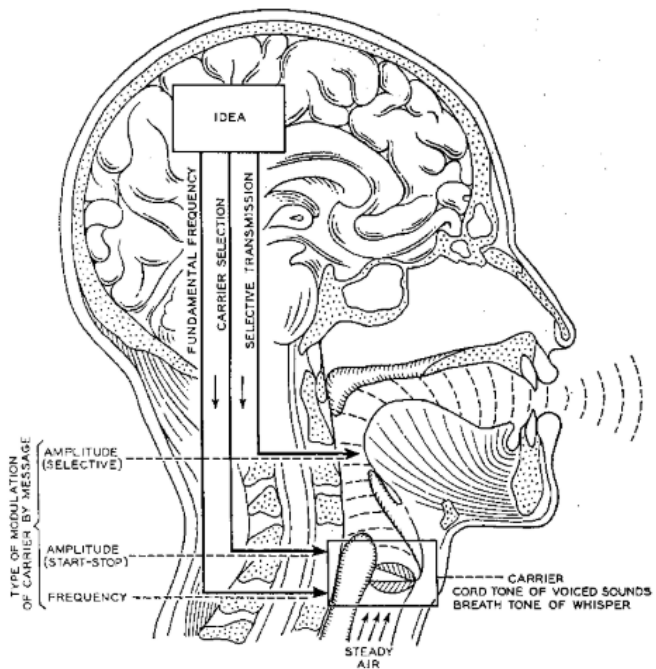
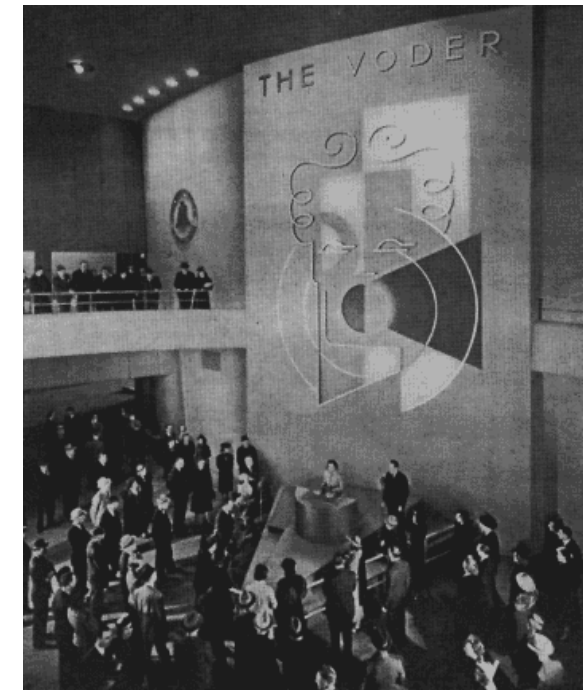


Fig. 8--Schematic circuit of the voder.



<https://ptolemy.berkeley.edu/eecs20/speech/voder.html>

Text-To-Speech (TTS) Technologies

WaveNet Speech Synthesis



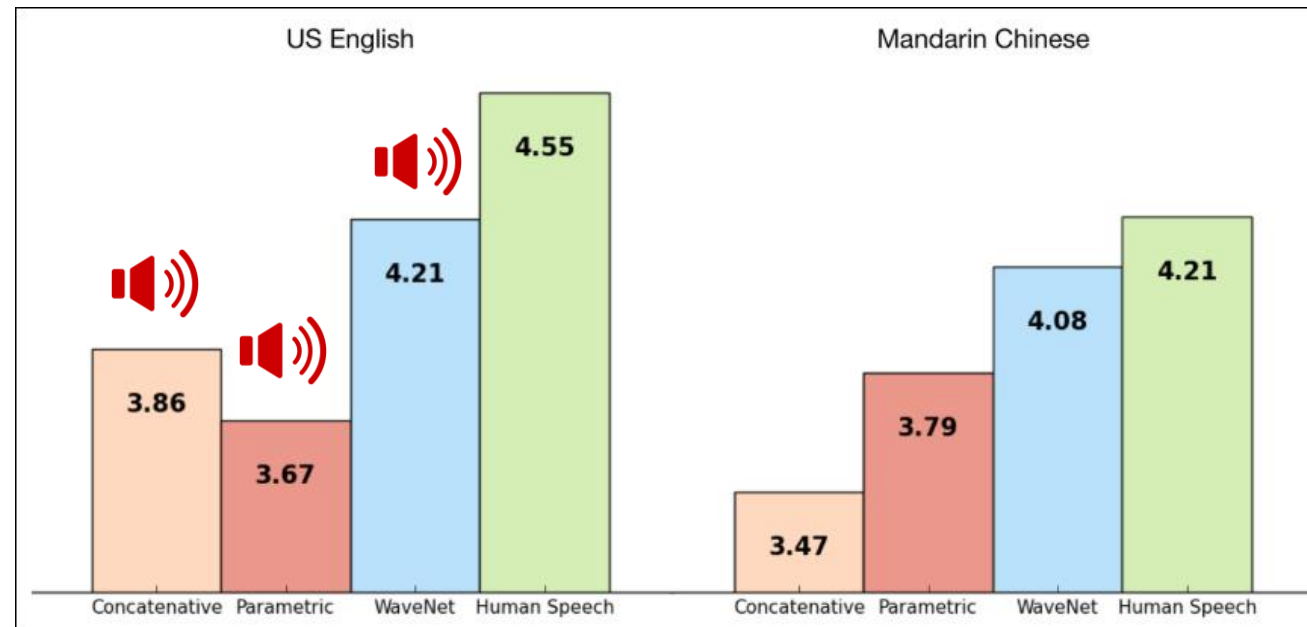
1 Second



Text-To-Speech (TTS) Technologies

WaveNet Speech Synthesis

- Same technology used to create speech for Google Assistant, Google Search, and Google Translate.
- Sounds more natural than other text-to-speech systems.
- Most people prefer WaveNet speech audio over other text-to-speech technologies.



<https://cloud.google.com/text-to-speech/docs/wavenet>

Text-To-Speech (TTS) Technologies



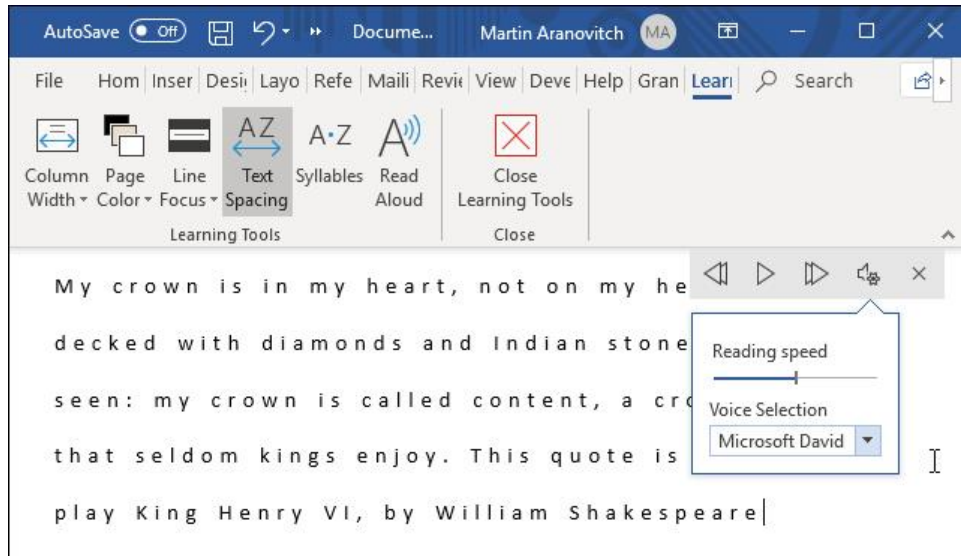
Text-To-Speech Engines



Text-To-Speech Engines

TTS engines provide users with access to text-to-speech functionality.

- Microsoft Speak (Word, Outlook, PowerPoint, Etc.)
- Amazon Polly
- Google Text-To-Speech



Text-To-Speech Terms

Here are some common terms we'll be using in this course:

- TTS (Text-To-Speech)
- Speech Synthesis (e.g. Concatenative, Parametric, WaveNet)
- Neural Networks
- Machine Learning AI (Artificial Intelligence) Voices
- SSML & Markup Tags
- Prosody (speech volume, pitch & speed)
- Phonemes & Phonetic Pronunciations



How To Create Text-To-Speech Audio Files: A Practical Step-By-Step Course For Beginners



End Of Lesson

How To Create Text-To-Speech Audio Files: A Practical Step-By-Step Course For Beginners

