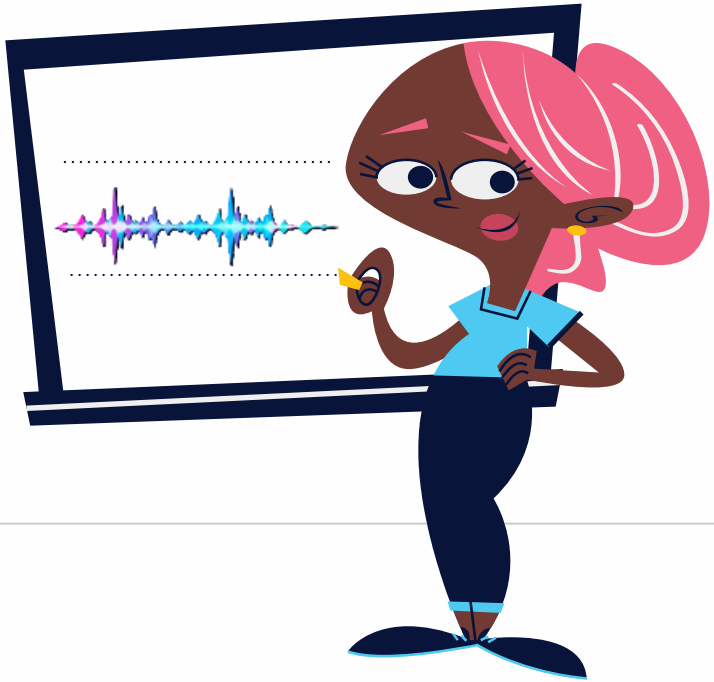
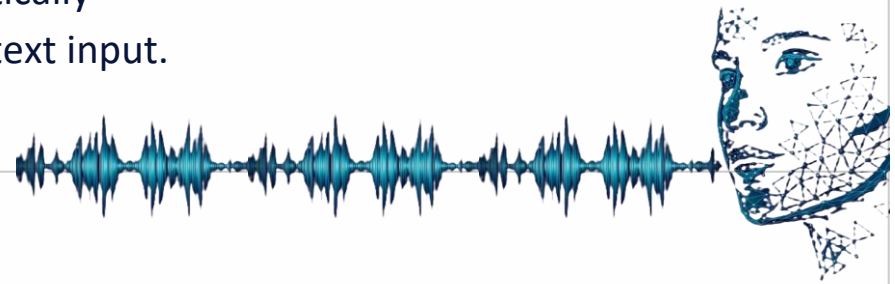
Speech Synthesis



What is Speech Synthesis ?

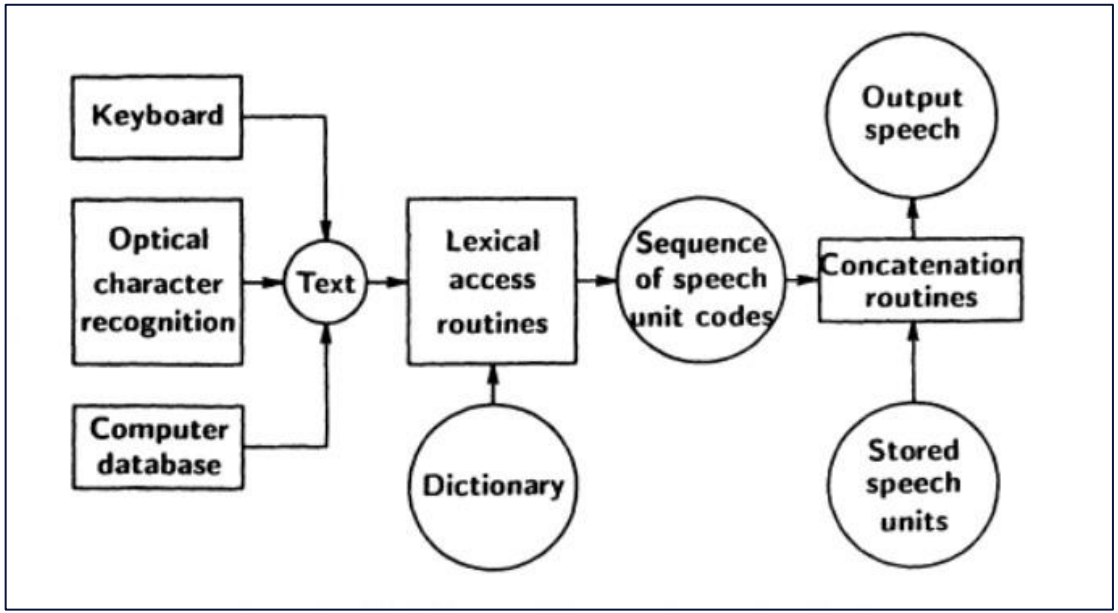
Speech synthesis is the process of automatically generating a speech signal from a written text input.



Comparison of Text Systems

Aspect	Limited Text Systems	Unrestricted Text Systems
Vocabulary Constraints	✓	✗
Input Flexibility	✗	✓
User Experience	Limited	Enhanced
Application Scope	Restricted	Evolving

Steps in Speech Synthesis



Synthesizer Methods

Today we'll be going over 3 different methods of Speech Synthesis :

a) Articulatory Synthesis

b) Format Synthesis

c) Linear Predictive Coding

Articulatory Synthesis

Articulatory synthesis is a method of speech synthesis that attempts to simulate the human vocal tract and its articulatory movements involved in speech production.

Model

Simulate

Reshape

Articulatory Synthesis

Advantages

- Enables control over parameters for natural speech.
- Develops understanding of speech production physiology.

Challenges

- Creating precise articulatory models is complex and needs high computation.
- Model parameter tuning needs expert knowledge.

Format Synthesis

Formant synthesis is a method of speech synthesis based on the manipulation of formants, which are resonant frequencies in the vocal tract.

Wave Generation

Filter & Combine

Adjust

Format Synthesis

Advantages

- Can produce intelligible speech.
- Control over individual formants.

Challenges

- lacks the naturalness and expressiveness .
- Fine-tuning the formant parameters to achieve natural-sounding speech is challenging.

Linear Predictive Coding (LPC)

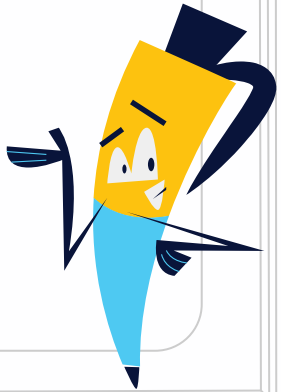
Linear Predictive Coding is a method commonly used in speech processing for representing the spectral envelope of a speech signal. It models the speech signal as the output of a linear filter driven by a sequence of input samples.

Analysis

Prediction

Estimation

Synthesis



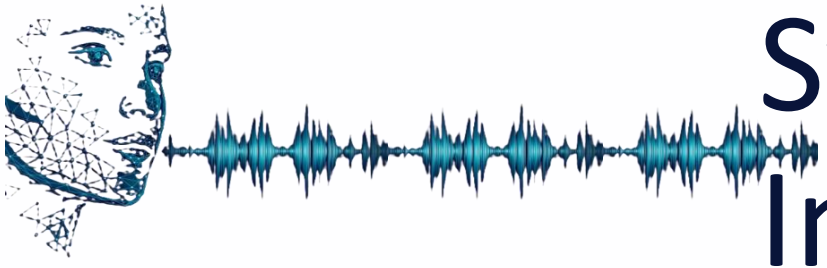
Linear Predictive Coding (LPC)

Advantages

- Provides a compact representation of speech signals, making it suitable for compression and transmission.
- Low computational complexity.

Challenges

- Performance can degrade in noisy environments or with rapidly changing speech characteristics.
- Fine-tuning LPC parameters for optimal speech synthesis may require expertise.



Synthesis of Intonation

Intonation

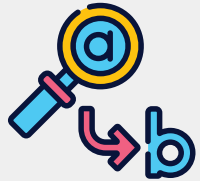
The synthesis of intonation refers to the process of generating the pitch contour or melody of speech. Intonation conveys linguistic and emotional information and plays a crucial role in speech perception and comprehension.

Methods:

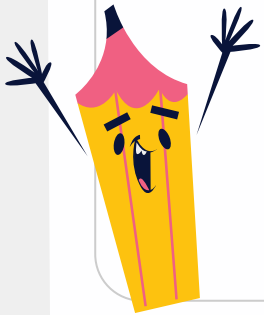
- Prosodic synthesis techniques, such as pitch contour modelling and prosody prediction.
- Statistical models, rule-based systems, and machine learning approaches can be employed to generate intonation patterns based on linguistic and contextual factors.
- Concatenative synthesis methods combine pre-recorded speech segments with varying intonation patterns to create natural-sounding speech.

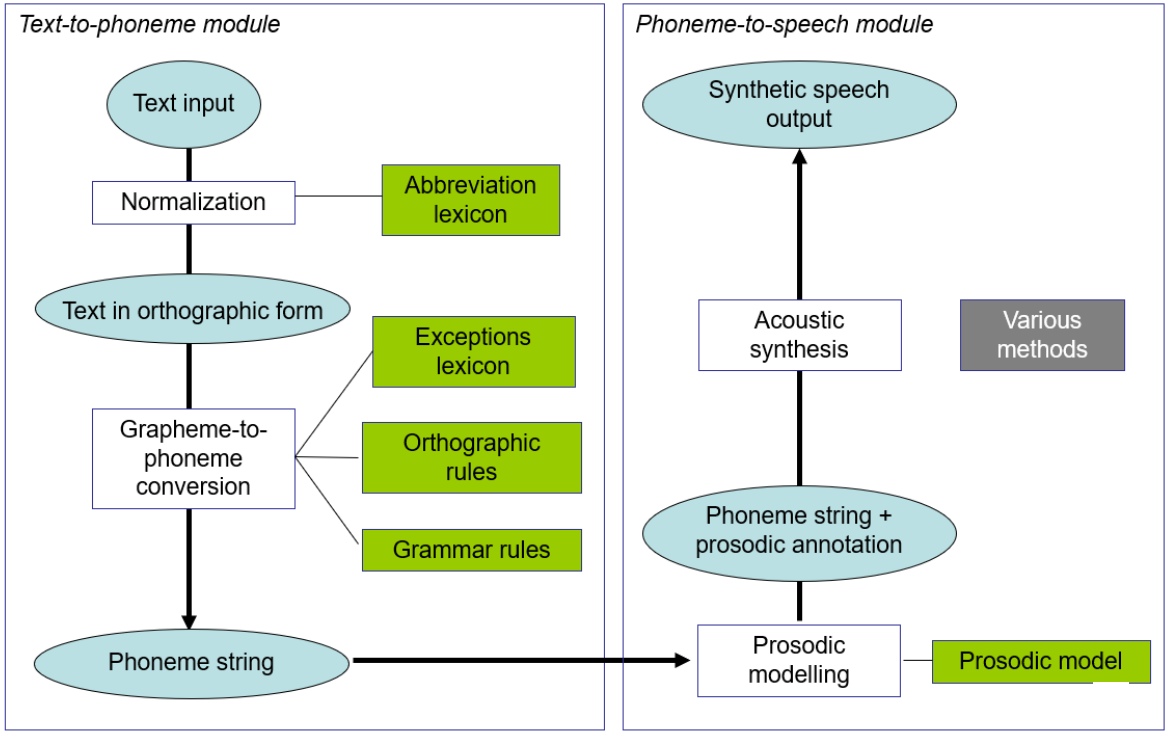
Considerations

Synthesizing natural intonation involves capturing nuances such as pitch variation, stress patterns, and rhythmic elements.



Contextual factors, such as sentence structure, discourse type, and speaker characteristics, influence intonation and must be considered in synthesis.





TTS Architecture

Limitations

Robotic Sound

Accent & Pronunciation

Emotional Nuance

Contextual Understanding