Module 3 - Topic 4

3.4.1 Introduction to Audio Models


Introduction
to Audio
Models

Audio is everywhere!

It is the primary way of communicating..

- Discussion
- Music
- Videos
- This very workshop!

What are Audio Models?

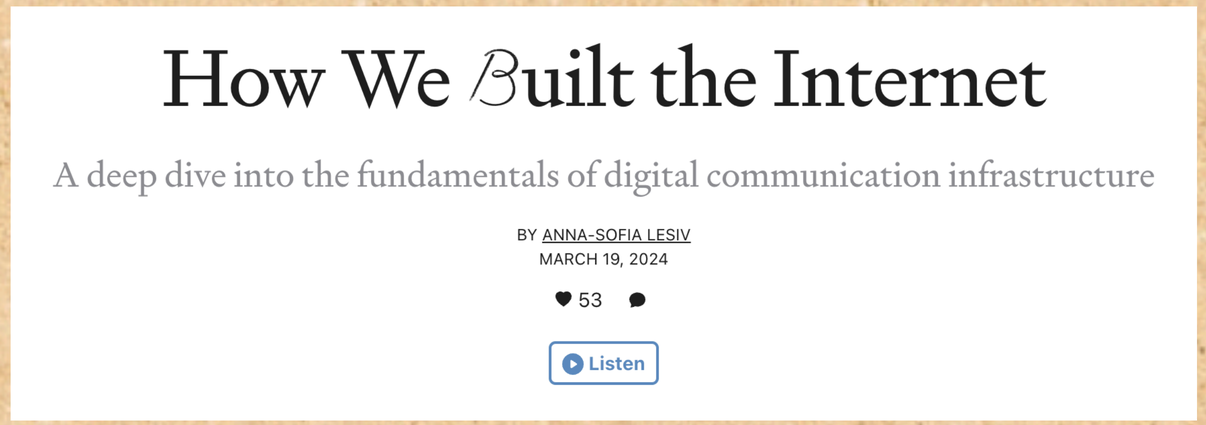
Audio Models are a subset of Generative AI that create audio content based on inputs. These systems can:

- Generate new, original audio content
- Create music, speech, and sound effects
- Transform text descriptions into audio
- Clone and modify voices

Things we do with Audio

Voice Assistants

Translation



Listen to Article

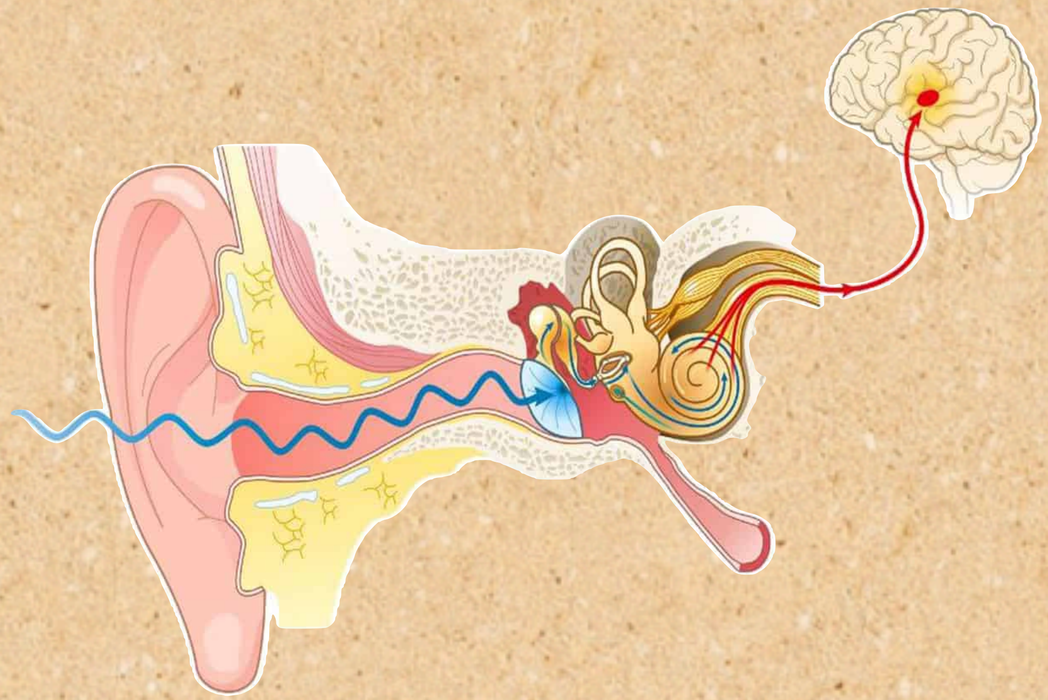
Meeting summariser

Types of Audio Models

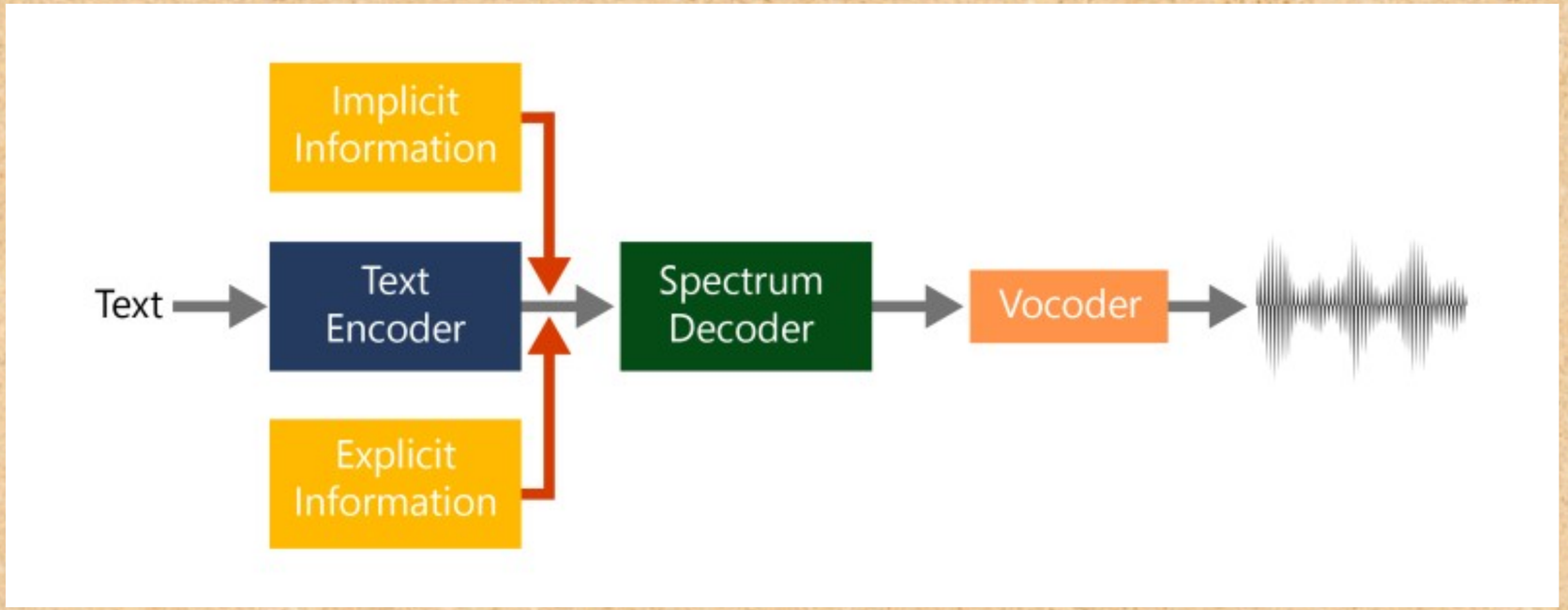
Different categories of audio models are:

- Speech to Text (STT)
- Text to Speech (TTS)
- Speech to Speech
- Music Generation
- Sound Effects Generation

How does AI understand Audio?



How Audio Models work?



Audio models break down the complex task of generating sound into 3 parts:

Text goes in, gets processed in stages, and natural sound comes out.

What do these components do?

Text Encoder: This component understands the text input and converts it into numerical values that represent both explicit content (like words, punctuation) and implicit meaning (like emotion, style, tone of voice).

Spectrum Decoder: This component takes the encoded information and creates a detailed plan of how the audio should sound, including patterns of frequencies and tones.

Vocoder: This component turns the audio patterns into actual sound waves that we can hear, ensuring the output sounds natural and clear.

In the next few lectures,

We will learn concepts of

- Speech to Text (STT)
- Text to Speech (TTS)
- Voice Cloning

Tools, Code and Apps!