Lecture 41: Audio Speech Options

Introduction:

- By default, the **Text to Speech (TTS)** model converts text into audio with standard settings. To **customize output** (such as speed and voice), we need to pass **options** using a SpeechPrompt.
- Key Concepts:
  - SpeechPrompt → Takes **text** and **options**
  - OpenAiAudioSpeechOptions → Used to configure customization.
  - Options can be built using the **builder pattern**.

Customizable Options:

- **Speed**
  - Controls playback speed of generated audio
  - Range: 0.25f (slowest) → 4.0f (fastest).
  - Example: 1.5f for faster playback.
- **Voice**
  - Multiple voices available.
  - Default: ALLOY.
  - Other options: ECHO, FABLE, ONYX, NOVA, SHIMMER
  - Example: Voice.NOVA.

Code Implementation:

```java
@RestController
public class AudioGenController {

    private OpenAiAudioTranscriptionModel audioModel;
    private OpenAiAudioSpeechModel audioSpeechModel;

    public AudioGenController(OpenAiAudioTranscriptionModel audioModel, OpenAiAudioSpeechModel audioSpeechModel) {
        this.audioModel = audioModel;
        this.audioSpeechModel = audioSpeechModel;
    }

    @PostMapping("api/tts")
    public byte[] tts(@RequestParam String text) {

        OpenAiAudioSpeechOptions options = OpenAiAudioSpeechOptions.builder()
                .speed(1.5f)
                .voice(OpenAiAudioApi.SpeechRequest.Voice.NOVA)
                .build();

        SpeechPrompt prompt = new SpeechPrompt(text, options);

        return audioSpeechModel.call(prompt)
                .getResult().getOutput();
    }
}
```

Key Points:

- **SpeechPrompt** allows passing both text and audio options.

- OpenAiAudioSpeechOptions is created using a builder.
- Customization improves the naturalness of speech output.
- Generated result is still returned as a **byte array** (can be saved as .mp3).

Summary:

- Default TTS can be enhanced with **speed** and **voice** options.
- Speed values range from 0.25f to 4.0f.
- Multiple voices are available (ALLOY, ECHO, NOVA, etc.).
- Use SpeechPrompt with OpenAiAudioSpeechOptions to apply customization.
- This completes the Audio Models section: **Speech to Text (STT)** and **Text to Speech (TTS)** with customization.