

# Day10 - 聲優太花錢？找 AI 幫你配音

昨天教大家從影像或聲音轉譯成文字，今天則是反過來，給一段文字後讓 AI 幫你配音。比起以前文字轉聲音的軟體，透過 AI 的聲音更為自然，角色跟速度也都能調整，讓我們看看如何操作吧。

為了彈性配置，application.yml 只存放 API KEY，其他參數都放在程式碼內。

```
@RestController
@RequiredArgsConstructor
class AiController {
    private final SpeechModel speechModel;
    @GetMapping("/tts")
    public ResponseEntity<byte[]> speech(String prompt) throws
        OpenAiAudioSpeechOptions speechOptions = OpenAiAudioS
            .withModel("tts-1")
            .withVoice(Voice.ALLOY)
            .withResponseFormat(AudioResponseFormat.MP3)
            .withSpeed(1.0f)
            .build();
        SpeechPrompt speechPrompt = new SpeechPrompt(prompt);
        SpeechResponse response = speechModel.call(speechPrompt);
        return ResponseEntity.ok()
            .header(HttpHeaders.CONTENT_DISPOSITION, "attachment")
            .body(response.getResult().getOutput());
    }
}
```

程式重點:

- Text To Speech 的模組是 `SpeechModel`，跟之前模組不一樣的是這個模組回傳的是 `byte[]`，要做一些處理才能下載檔案，所以回傳值也特別使用 `ResponseEntity<byte[]>`
- `.withModel("tts-1")` TTS 的模型，可以選 `tts-1` 以及 `tts-1-hd`
- `.withVoice(Voice.ALLOY)` 選擇要哪種聲音，有以下六種可選

```
@JsonProperty("alloy") ALLOY("alloy"),
@JsonProperty("echo") ECHO("echo"),
@JsonProperty("fable") FABLE("fable"),
@JsonProperty("onyx") ONYX("onyx"),
@JsonProperty("nova") NOVA("nova"),
@JsonProperty("shimmer") SHIMMER("shimmer");
```

- `.withResponseFormat( AudioResponseFormat . MP3 )` 輸出聲音的格式，雖然後面存檔會自訂附檔名，不過格式得匹配，不然程式會無法辨認

```
@JsonProperty("mp3") MP3("mp3"),
@JsonProperty("opus") OPUS("opus"),
@JsonProperty("aac") AAC("aac"),
@JsonProperty("flac") FLAC("flac");
```

- `.withSpeed(1.0f)` 講話的速度，雖然影音編輯軟體可以調整速度，不過透過 AI 來調整聲調才不會變太多，範圍介於 0.25 ~ 4
- 下載時要指定檔名可以透過 `HttpHeaders.CONTENT_DISPOSITION` 指定
- 因為結果直接是 `byte[]` 格式，直接放入 `body` 即可

```
return ResponseEntity.ok()
    .header(HttpHeaders.CONTENT_DISPOSITION, "att
    .body(response.getResult().getOutput());
```

經過測試，雖然偶有錯誤發生，不過聲音聽起來非常自然，可以跟手機版 ChatGPT 對話就知道效果，雖然有點 ABC 腔調，不過整體來說還蠻好聽的

到目前為止凱文大叔把基本 Model 的操作都介紹過，不過 AI 的變化非常快，搞不好幾個月後又有新的 Model，大家覺得前面介紹這些 Model 可以取代掉那些職業？

回顧一下今天學到甚麼：

- SpeechModel 的調用
- OpenAiAudioSpeechOptions 的參數設定

- 回傳 binary 格式要如何下載檔案