

STT & TTS

김수현 2022.02.12

목차

- STT
- TTS

STT란?

- Speech-to-Text
- 음성을 텍스트로 바꾸는 기술
- Apple Speech 프레임워크, Google Cloud, Naver CLOVA 등

	Apple Speech	Google Cloud	Naver CLOVA
난이도	중	상	상
속도	말이 길어질수록 버퍼링이 생김	Good	Good
정확도	70% 일상적인 말들은 인식을 아주 잘한다. 어려운 단어와 문장들은 음...	97% 완벽에 가까운 음성인식률로 감탄을 자아냄. 어려운 단어와 문장들도 웬만하면 다 인식을 잘 한다.	80% 일상적인 말들은 인식을 잘 한다. 어려운 단어와 문장들은 띄어쓰기가 엉망이 되는 경향이 있다.
가격	무료	0~60분 무료, 이후 15초당 \$0.006	15초당 4원
사용법	Speech STT 사용법	Google STT api 사용법	Naver CSR api 사용법

STT) Speech 프레임워크 (1)

- 마이크 & 음성인식 권한 요청

Privacy - Speech Recognition Usage Description	↕	String	Need the speech recognition capabilities for searching tags
Privacy - Microphone Usage Description	↕	String	Need the mic for audio to text

```
public func requestAuthorization() {
    SFSpeechRecognizer.requestAuthorization { status in
        DispatchQueue.main.async {
            self.authorizationStatus = status
        }
    }
}
```

STT) Speech 프레임워크 (2)

- AVAudioEngine : 오디오와 관련된 전반적인 일 담당

```
private let audioEngine = AVAudioEngine()
```

- SFSpeechRecognizer: 음성인식 (언어 설정 가능)

```
private let speechRecognizer = SFSpeechRecognizer(locale: Locale(identifier: "en-US"))!
```

- 인식된 음성은 받아올 String

```
@Published public var recognizedSpeech: String?
```

STT) Speech 프레임워크 (3) -음성(녹음)

- 음성을 듣고 인식하고 일을 하는 함수 만들어주기 ex. recordAndRecognizeSpeech()
- Start Recording Button을 탭하면, 앱은 **AVAudioSession**의 싱글톤 객체를 받아서, **녹음**을 위해 설정하고, active한 세션으로 만든다 (녹음 관련된 세팅)

```
// Configure the audio session for the app.
let audioSession = AVAudioSession.sharedInstance()
try audioSession.setCategory(.record, mode: .measurement, options: .duckOthers)
try audioSession.setActive(true, options: .notifyOthersOnDeactivation)
let inputNode = audioEngine.inputNode

// Configure the microphone input.
let recordingFormat = inputNode.outputFormat(forBus: 0)
inputNode.installTap(onBus: 0, bufferSize: 1024, format: recordingFormat) { (buffer: AVAudioPCMBuffer, when: AVAudioTime) in
    self.recognitionRequest?.append(buffer)
}

audioEngine.prepare()
try audioEngine.start()
```

STT) Speech 프레임워크 (4) -음성 인식

- SFSpeechAudioBufferRecognitionRequest : 실시간 오디오에서 **음성을 인식**하기 위해
- shouldReportPartialResults 프로퍼티를 true.
- 음성 인식 시스템이 결과가 인식되는 즉시 그 결과를 return 하도록 한다

```
// Create and configure the speech recognition request.  
recognitionRequest = SFSpeechAudioBufferRecognitionRequest()  
guard let recognitionRequest = recognitionRequest else { fatalError("Unable to create a SFSpeech  
recognitionRequest.shouldReportPartialResults = true
```


STT) Speech 프레임워크 (5) - 음성 인식 처리

- 음성 인식을 시작하기 위해 SFSpeechRecognizer의 `recognitionTask(with:resultHandler:)` 호출
- 이전에 `shouldReportPartialResults`가 true
 - 해당 메소드는 부분적으로 결과물들을 전달하기 위해 '주기적으로' 코드 블록 실행
- result 오브젝트의 `bestTranscription` 프로퍼티 안에 있는 텍스트를 사용해서 텍스트 뷰를 업데이트

```
recognitionTask = speechRecognizer.recognitionTask(with: recognitionRequest) { result, error
    var isFinal = false

    if let result = result {
        // Update the text view with the results.
        self.textView.text = result.bestTranscription.formattedString
        isFinal = result.isFinal
        print("Text \(result.bestTranscription.formattedString)")
    }

    if error != nil || isFinal {
        // Stop recognizing speech if there is a problem.
        self.audioEngine.stop()
        inputNode.removeTap(onBus: 0)

        self.recognitionRequest = nil
        self.recognitionTask = nil

        self.recordButton.isEnabled = true
        self.recordButton.setTitle("Start Recording", for: [])
    }
}
```


TTS란?

- Text-to-Speech
- 텍스트를 음성으로 바꾸는 기술
- Apple Speech 프레임워크, Google Cloud, Naver CLOVA 등

	Apple Speech	Google Cloud	Naver CLOVA
난이도	하	중	중
정확도	가장 AI 기계같은 말투	자연스러움	자연스러움
목소리	Siri 목소리	한글 목소리 남 여 각 두개 씩 지원	한글 목소리 수십가지 지원
가격	무료	0~100만 자 무료, 이후 문자당 \$0.000016	월 90,000원
사용법	Speech TTS	Google TTS api 사용법	CLOVA Voice 사용법

TTS 예제 코드

```
1 import SwiftUI
2 import AVFoundation
3
4 struct ContentView: View {
5
6     @State private var inputString = "Hello world! My name is Dev Techie"
7
8     var body: some View {
9         VStack {
10             TextField("Enter text", text: $inputString)
11                 .textFieldStyle(.roundedBorder)
12
13             Button("Text to speech") {
14                 let utterance = AVSpeechUtterance(string: inputString)
15                 utterance.voice = AVSpeechSynthesisVoice(language: "en-US")
16                 utterance.pitchMultiplier = 2.0
17                 utterance.rate = 0.3
18                 let synthesizer = AVSpeechSynthesizer()
19                 synthesizer.speak(utterance)
20             }
21             }.padding()
22
23     }
24 }
25
26 struct ContentView_Previews: PreviewProvider {
27     static var previews: some View {
```

참고 자료

- [STT & TTS] <https://hongssup.tistory.com/209>
- [STT] `SFSpeechAudioBufferRecognitionRequest`
- [TTS] <https://www.appcoda.com/text-to-speech-swiftui/>