DSPSpeech-20

A speech dataset of 20 isolated words each with 680 recordings from 34 individuals

Status

- Vocabulary: **20 words** in the following table
- Speaker: **34 individuals**
- Amount: **13,599 recordings** (20 of each word per speaker)
- Size of a single file: **66 KB**
- Total size: **891.5 MB**

Vocabulary

Each audio file is a recording of an isolated word. The audio files are named as `speech_data/{speakerID}/{speakerID}-{label}-{index}.wav`. For example, `speech_data/16307130194/16307130194-12-16.wav` is the `16`th word `Speech` recorded by speaker `16307130194`.

00	01	02	03	04	05	06	07	08	09
数字	语音	语言	识别	中国	总工	北京	背景	上海	商行

10	11	12	13	14	15	16	17	18	19
复旦	饭店	Speech	Speaker	Signal	Process	Print	Open	Close	Project

Parameters

Format	AudioFormat	NumChannels	SampleRate	BitRate	BitDepth	length
`.wav`	PCM	1	16 kHz	256 kbps	16 bits	~2 s

These are the important parameters regarding the audio files. If you'd like to get more information about the audio files, you can look closely at the header of the files. While doing so, keep in mind that it's **little-endian** format!

Source

The recordings are contributed by 34 students of *Digital Signal Processing Spring 2019* and recorded on website https://github.com/czhongyu/audio-collector