

# Taiwanese Learner

Taiwanese, also known as Taiwanese Hokkien, Taigi, 臺語, 臺灣閩南語, is a variety of the Hokkien language spoken natively by about 70%+ of the population of Taiwan. However, due to the 國語運動 implemented before, less and less people use this language nowadays. Although there are some native language courses in schools, they are far from enough, most young people don't use Taiwanese in their life, and most people don't know how to write down Taiwanese, they even think Taiwanese don't have a writing system.

In fact, Taiwanese has a well-structured writing system. It can be both written in Hà̍n-jī (漢字) and Tâi-lô phing-im (臺羅拼音). "Tâi-lô phing-im" is an example of 臺羅拼音, a word seperated by '-' or ' ' represent a 漢字, the symbol above a word represent it's tone. Today, TA A-huâ (阿華) wants a tool to help him pronounce Taiwanese.

The tool will take a sentence of 臺羅拼音 as input which has a maximum length of 50, then generate a sentence and a single .wav file that contains the pronounation of each words seperately and then the whole sentence.

As for input, because the tones' symbols of them are not presented in ASCII (e.g.: â, ò), we will use numbers to represent tones instead. Taiwanese has 8 tones, so we will use number 1~8. For example, Tâi-lô phing-im will be represented as Tai5-lo5 phing1-im1. Input's tones always represent as numbers.

After taking the input, the tool will use a web API to download the pronunciation from Internet. (Tip: you can use Unix commands like `curl` or `wget` )

This is a website that help you pronounce Taiwaness:

<https://suisiann.ithuan.tw/講>

This is the web API to download the pronounation:

<https://hapsing.ithuan.tw/bangtsam?taibun=<input>>

For example, if you want the pronounation of Guá-sī tâi-uân-lâng (我是臺灣人), you can use:

<https://hapsing.ithuan.tw/bangtsam?taibun=Gua2-si7+tai5-uan5-lang5>

and get 1 .wav file (Tip: replace space with '+' in the url, spaces can cause some problems)

Apart from the whole sentence, you also need to grab the pronounation of each words from the sentence. Beware, Taiwanese has something called Liân-thòk piàn-tiāu (連讀變調), that means when pronouncing a sentence, except for **the last word in the sentence**, all other words need to change their tones while keeping their tones' numbers the same. But when you grab a word from a sentence and pronouce it alone, you also need to change it's tone's number, so it can retain it's tone. To make it simpler, TA A-huâ provided simplified rules of 連讀變調 in **tone.pdf**. For example, "Gua2-si7 tai5-uan5-lang5" will become "Gua1. si3. tai7. uan7. lang5.". You need to use the API for these 5 words separately ("Gua1.", "si3.", "tai7.", "uan7.", "lang5.") and get 5 .wav files.

After that, you need to combine those 6 (5 seperated words and 1 full sentence) .wav files into 1 .wav file called **out.wav**, you can check .wav file format in Wiki.

So if you input Gua2-si7 tai5-uan5-lang5. :

```
$ ./hw0505
input: Gua2-si7 tai5-uan5-lang5.
output: Gua1. si3. tai7. uan7. lang5. Gua2-si7 tai5-uan5-lang5.
out.wav has been generated.
```

**out.wav** file will sound like

Gua1. si3. tai7. uan7. lang5. Gua2-si7 tai5-uan5-lang5.

In **out.wav**, the first half part should sound the same as the last half part. TA A-huâ also provided **example.wav** as the output file of it.

Here is another example:

```
$ ./hw0505
input: Gueh8-niu5.
output: Gue3. niu5. Gueh8-niu5.
out.wav has been generated.
```

The input will always be in correct 臺羅拼音 format, and will always consist of 1 sentence. That means it will only contain one dot, lower-case letters, one upper-case letter, non-continuous spaces, non-continuous dashes, non-continuous numbers. You also don't need to handle any error cases, like fail to create files, can't connect to Internet etc. You are allowed to create temporary files in the same directory.

TA A-huâ also provided **example.pdf**, which includes 臺羅拼音 sentences and their corresponding Tâi-gí hàn-jī (臺語漢字) and Huâ-gí huan-ik (華語翻譯).

If you are interested in learning Taiwanese, you can take 閩南語概論 or 初級台語 courses in NTNU