

出门问问-声韵母切分标注

一、标注格式

1、文本格式：

1) 文本格式为*.txt，一行文字，一行拼音。文字行中文同“中文标注格式”，英文同“英文标注格式”。

2) 韵律标注：

中英混合文本，按**中文韵律**结构标注，包含**韵律词(#1)**、**韵律短语(#2)**、**语调短语(#3)**、**句末(#4)**四个层级的标注。(注：一个英文单词划分为一个韵律词，以“空格”表示)

例：

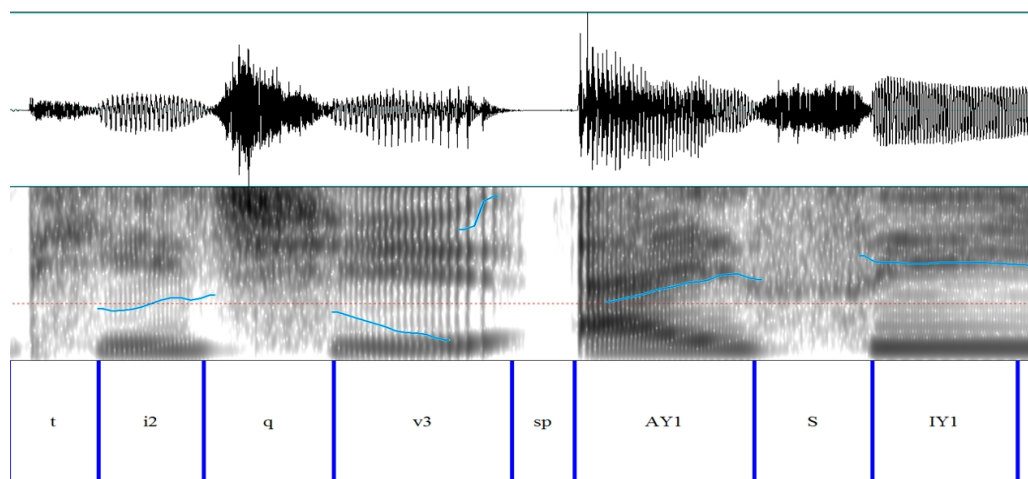
000001 阿凡达的#1 主题曲#2I See You#4。

a1 fan2 da2 de5 zhu3 ti2 qu3 / AY1 / S IY1 / Y UW1

2、声韵母（英文音素）边界切分格式：

中文（台湾普通话）切分到声韵母，英文切分到音素，标注格式为 interval 文件格式。

例：



二、切分规则

1、标注目的

- 1) 对中文汉字拼音（英文音标音素）进行声母韵母（英文音素）边界的调整。
- 2) 中文韵律词（英文单词或字母）之间有停顿（一般为30ms以上），加静音标注。
- 3) 如果标注的拼音或音素的名字有错（与实际发音不符），修改拼音或音素名，同时修改文本。

2、标注符号

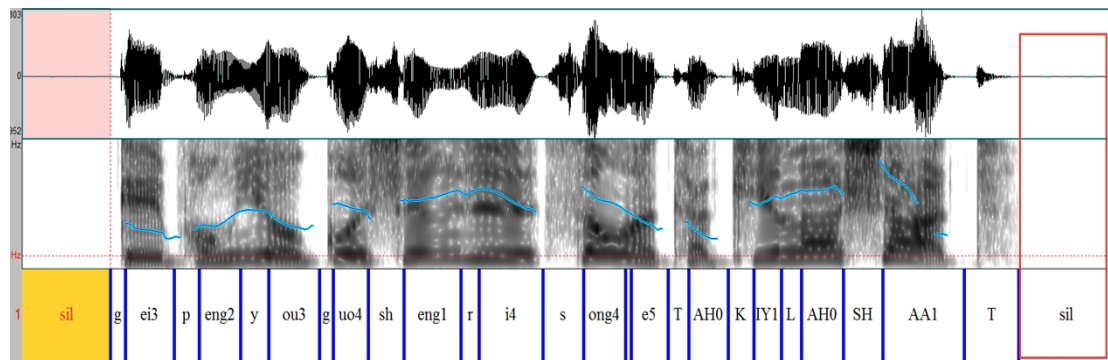
标注符号采用声韵母（英文音素）本身的符号，另外还包括有 ‘sil’ 和 ‘sp’ 的标注信息，具体解释如下：

- sil表示句子首尾静音段标记。
- sp表示句子中间大于30ms的停顿标记(韵律词或单词内部除外)。

3、标注一般规则

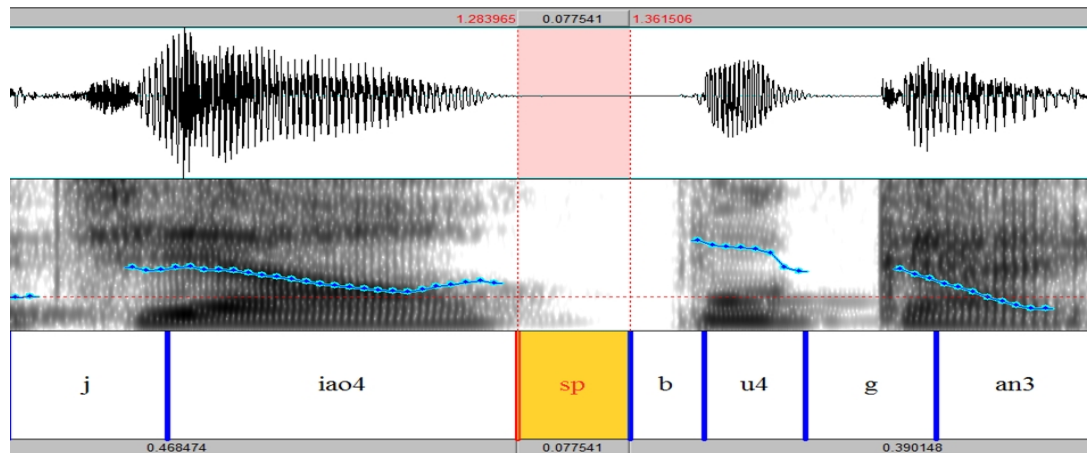
1) 句首、句尾用 “sil” 标注

例：



2) 句中静音段，插入 “sp” 标注

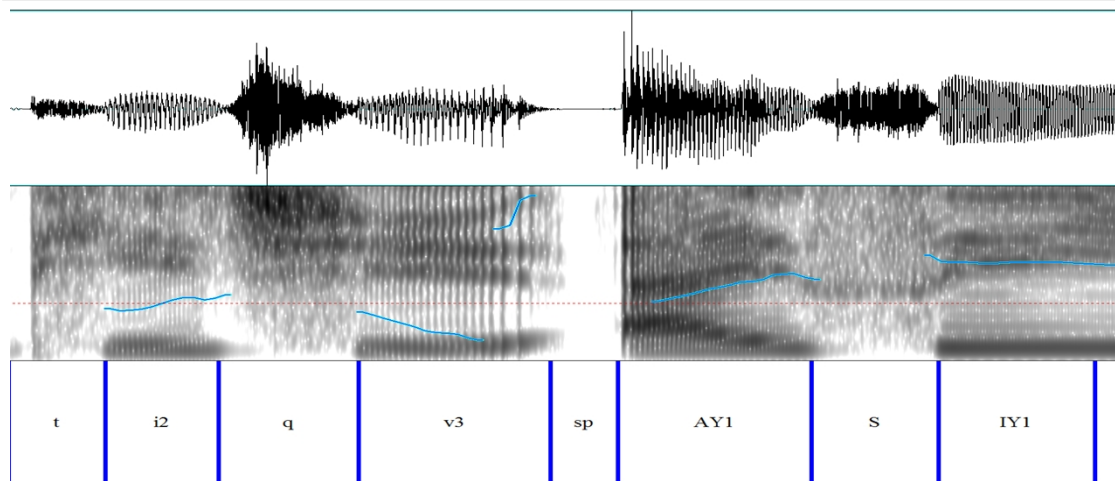
例：



3) 声韵母（或音素）本身以实际发音判断

标注采用音素本身的符号（汉语拼音“ü”由小写字母“v”代替），中文韵母+声调，英文元音+重音符号。

例：



4、问问切音特别规定

- 1) 由于非 native 发音人对英文字母 “W” 发音不标准问题，有时会根据发音人的读法更改标注音素，问问英文字母 “W” 统一标为 6 个音素 (D AH1 B . L Y UW0)；
- 2) 儿化音标注，如 “花儿” 的儿化读法一般标注为 “hual1”，问问标注为 “hua1 xr1” (台语中不涉及)；
- 3) 中文 “iu、ui、un” 一般采用 “iou、uei、uen” 标注边界切分文件，问问采用韵母原拼音符号；
- 4) 韵母 “i” 在声母不同的情况下，一般以 “i”、“ii”、“iii” 来区分在不同声母后的不同发音，问问全部采用 “i” 来标注；
- 5) 零声母的切分：一般标注中，以零声母 “y” 和 “w” 开头的音节，不进行声韵母切分，问问要求进行声韵母切分，并采用音素本身的符号标注。

例：

