# 1 System Overview:

manual segmentation according to lyrics lines/sentences from script of aria

# 2 Within-Syllable distribution rules

# 3 Vocal/nonvocal segmentation

tested on dan-xipi-01, resulted in worse accuracy.

TODO: test on all with and without vocal

# 4 Experimental Setup

## 4.1 setting one: without score: each syllable gets equal duration

Using the duration-explicit decoding, but each syllable within a sentence gets equal duration

parameters:

-consonant_duration = 0.3,

- deviationINSec = 0.07 ( TODO: vowel gets longer deviationInSec. Try first reading syllable durs from textGrid and )

- resynthesis threshold = 70 db

## 4.2 setting two: each syllable gets duration from score

Lyrics are read from TextGrid annotaion layer 'lyrics-syllables-pinyin.' Lyrics sentences are automatically derived by considering each punctuation sign after a syllable as an indicator of an end of a sentence (ideally from from score, for now from annotation).

### 4.3 Results

| | deviationInSec | laosheng-erhuang_04 | laosheng-xipi_02 | dan-xipi_01 | dan-xipi_02 |
|---|---|---|---|---|---|
| without score | 0.1 | 26.37 | 25.00 | 45.30 | |
| | 0.5 | 28.26 | 21.74 | | |
| | | | | | |
| with score durarions | 0.1 | **32.01** | **45.05** | **63.77** | **18.39** |
| | 0.5 | 38.74 | 43.88 | 50.14 | 21.3 |
| | | | | | |

dan_xipi_01 has best score: seems it s due to female voice, not to tempo.

———-

Tempo (not done yet):

| | deviationInSec | slow | mid | fast |
|---|---|---|---|---|
| without score | 0.1 | 23.26 | 36.88 | 41.78 |
| | | | | |
| | | | | |
| with score | 0.1 | 36.22 | 46.24 | 51.44 |
| | 0.5 | 41.68 | 45.17 | 40.71 |
| | | | | |

Slow tempo has a lot of variation, not stable results. middle tempo is better

increasing the deviation brings improvement when tempo is slow, but affects results negatively for mid and fast tempo.