Data Collection Explanation

I think we only need the SA sentences and part of SX sentences in TIMIT. Also add some common voice commands for smart devices: e.g. open the door, set alarm in 30 seconds, read my recent message.

* **Common:** 2 sentences from TIMIT SA as original sentences in **Common** type. Add 3 more sentences worked as activation sentences to this category. So **Common** type is 5.
* **Compact:** This type almost inherited from TIMIT dataset, which I will choose 40 from SX category.
* **Fricative Command:** I will collect 20 common command used in voice assistant system; all of those commands should contain at least one fricative consonant.
* **None-Fricative Command:** Similar to fricative command, the non-fricative commands don’t contain fricative consonant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sentence Type | # Sentences | # Speakers | Total | #Sentences/Speaker |
| Common () | 5 | 100 | 500 | 5 |
| Compact TIMIT | 40 | 5 | 200 | 2 |
| Fricative Command | 20 | 10 | 200 | 2 |
| Non-fricative Command | 20 | 10 | 200 | 2 |
| Total | 85 |  | 1100 | 11 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sentence Type | # Sentences | # Speakers | # Repeat | Total | #Sentences/Speaker |
| Common () | 5 | 100 | 2 | 1000 | 10 |
| Compact TIMIT | 100 | 20 | 1 | 2000 | 20 |
| Fricative Command | 50 | 10 | 1 | 500 | 5 |
| Non-fricative Command | 50 | 10 | 1 | 500 | 5 (mark pure) |
| Total | 205 |  |  | 4000 | 40 |

The table above, #Sentences means how many sentences will have in the type. #Speakers means for same sentence it will be recorded by how many speakers. Total means total utterances we will collect in the type. #Sentences/Speaker means how many sentences in this type will be recorded by same speaker. Generally, #Sentences \* #Speakers = #Sentences/Speaker \*Total Speaker, the total speaker in our case is 100.

**Reference:**

TIMIT speech material

Sentence Type #Sentences #Speakers Total #Sentences/Speaker

------------- ---------- --------- ----- ------------------

Dialect (SA) 2 630 1260 2

Compact (SX) 450 7 3150 5

Diverse (SI) 1890 1 1890 3

------------- ---------- --------- ----- ----------------

Total 2342 6300 10

SA: The dialect sentences (the SA sentences) were meant

to expose the dialectal variants of the speakers and were read by all 630

speakers

SX: The phonetically compact sentences were designed to provide a good

coverage of pairs of phones, with extra occurrences of phonetic contexts

thought to be either difficult or of interest. Each speaker read 5

of these sentences (the SX sentences) and each text was spoken by 7 different

speakers.

SI: The phonetically diverse sentences (the SI sentences) were selected

from existing text sources - the Brown Corpus (Kuchera and Francis, 1967) and

the Playwrights Dialog (Hultzen, et al., 1964)