


TIMIT Dataset:-

MIT

Texas Instruments

SRI (Stanford Research
Institute)

Corpus Speaker Distribution:-

6300 sentences

→ 630 speakers - 10 sentences each

8 dialect regions →

New England

Northern

North Midland

South Midland

Southern

New York City

Western

Army Boat (moved around)

~20% 438 men
~30% 192 women

Corpus Text Material :-

	Sentences	Speakers	Total	Sentences/ Speaker
Dialect (SA)	2	630	1260	2
Compact (Sx)	450	7	3150	5
Diverse (SI)	1890	1	1890	3

Dialect → Focuses on capturing regional variations in pronunciations

Compact → Densely pack specific phonetic combn's @ context to facilitate detailed phonetic & acoustic analysis

Diverse → Aimed at representing natural variation found in everyday speech

Core Test Set :-

24 speakers → 2 male, 1 female
from each dialect
region

5 compact
sentences
(SX)

3 Diverse
sentences
(SI)

= 192 sentences

Complete Test Set :-

168 speakers

1344 utterances

Files!

• Wav format → Uncompressed, lossless audio format

developed by IBM and
Microsoft

↓
LPCM (Linear Pulse-Code Modulation)

• Wav → RIFF Header

FMT chunk → describes audio format

Data chunk

↓
8 bit, 16 bit, 24 bit

Sample rate → 44.1 kHz
48 kHz

CD
DVD

Ⓛ Mono, Stereo,
Surround sound

With windows media player

→ installed K-lite Codec packs
for decoding & encoding of the
audio data
&
Video

To play the audio / .wav

• wav → Sphere-Headered Speech Waveform

• txt → Orthographic Transcription