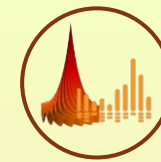# SPIRE Datasets

(Open-sourced datasets created at SPIRE Lab, IISc, Bangalore)

Vaani: 27000+ hours multi-modal spontaneous speech (~5% transcribed) using 258K images from 143K+ speakers in100+ Indian languages

RESPIN: 10000+ hours dialect-rich read speech from 18K+ speakers in 9 Indian languages in agriculture and finance

SPIRE SIES: 166+ hours of transcribed Indian English spontaneous speech from 1600+ speakers

SPICOR TTS_2.0: 57+ hours domain-rich Gujarati TTS Corpus from 2 speakers

SYSPIN: 900+ hours studio-quality, read speech, domain-rich TTS corpus from 18 speakers in 9 Indian languages

wSPIRE: 36+ hours of parallel Neutral and whispered transcribed speech from 92 speakers using 5 recording devices

SPICOR TTS_1.0: 97+ hours domain-rich Indian English TTS Corpus from 2 speakers

IndicTIMIT: 240+ hours of read Indian English recordings of 2342 TIMIT sentences from 80 speakers

SPIRE EMA: An electromagnetic articulography (EMA) corpus from 38 Indian speakers each speaking 460 English sentences

SPIRE VCV: Parallel Acoustic and EMA data from 10 speakers speaking VCV (5 vowels, 17 consonants) in three speaking rates