Developing an Intonation Labeled Dataset for Hindi

Authors : Esha Banerjee, Atul Kumar Ojha, Girish Nath Jha

Abstract : This study aims to develop an intonation labeled database for Hindi. Although no single standard for prosody labeling exists in Hindi, researchers in the past have employed perceptual and statistical methods in literature to draw inferences about the behavior of prosody patterns in Hindi. Based on such existing research and largely agreed upon intonational theories in Hindi, this study attempts to develop a manually annotated prosodic corpus of Hindi speech data, which can be used for training speech models for natural-sounding speech in the future. 100 sentences (500 words) each for declarative and interrogative types have been labeled using Praat.

Keywords : speech dataset, Hindi, intonation, labeled corpus
Conference Title : ICSLP 2021 : International Conference on Speech and Language Processing
Conference Location : San Francisco, United States
Conference Dates : November 01-02, 2021