Problem Statement:

Comparison of Natural and Automated Speech using Duration Models

Teammates: Sourima Laha and Nayana Raj

One common grievance raised against automated speech generated by TTS engines is that they are devoid of important paralinguistic information such as prosodics or emotional content. Much of the existing work on generating natural sounding speech has focused on modification of the pitch contour of a sentence.[1] In this project, we compare natural speech from the SHRUTI Bangla Speech Corpus and synthesised speech generated by a TTS engine. The objective of this project is to construct duration models to collate the both using contextual effects in order to identify the vocal features that distinguish them.

---

[1] Muralishankar, R., et al. "Modification of Pitch using DCT in the Source Domain." *Speech Communication*, Feb. 2004, pp. 143–154.