

# Summary of Text to Speech System for Urdu Language

---

In this Paper discuss about the ways of synthesis and There are two important synthesis techniques

- Unit selection Base Synthesis Method
- HMM Base Model Synthesis Method

In **Unit selection Base Synthesis** suitable pre-recorded units are concatenated to obtain the speech of given Text and In the literature the development of Corpus for the Unit Selection Base Synthesis is very difficult so that in order to reduce the creation of corpus problem for Unit selection introduced the **HMM Base Model Synthesis**.

In **HMM model Base synthesis** is a statistical parametric bases speech Synthesis techniques. Given Text Convert it into the speech using parametric representation. The main advantages is that it stored the statistics rather than the waveform.

## Corpus Creation

---

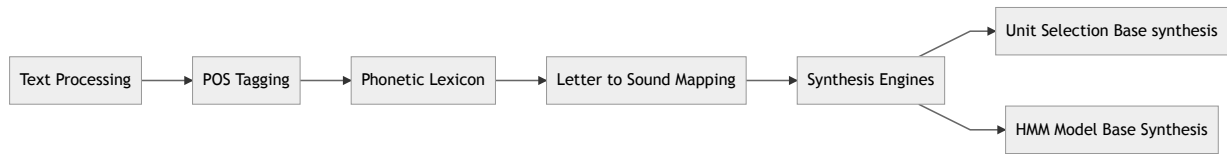
Corpus creation is the main difficulty level in TTS because every language has different system . For urdu They collected **80,081** sentence from different Resources and hired a professional Speaker to recording these sentences at sample rate **24khz**. At the last they achieved :

- 8081 Sentences
- 82,049 words
- 130,163 syllables

## Urdu TTS Development

---

For Urdu TTS development they used open source Festival . Given Text convert into the speech by executing several modules like First Step Text Analysis All non-standard words convert into the standard words and then convert graphemes into phonemes and then phonemes can be synthesis into Unit selection synthesis and HMM Base synthesis.



**Text Processing** Text processing modules takes text as input and change it into normalized text. Text processing module performs following steps:

- **Sentence segmentation**
- **Word boundary marking**
- **Classification of digits dates**
- **Symbols on the basis of context**
- **Text generation**

Sentences are segmented on the basis of punctuation mark (full stop, question mark), and line break. Each sentence is further tokenized at word level by using the space and punctuation marks as word boundary. Firstly, semantic tagger analyze each token and converts each token into text. Because, in the written form of a language, numbers, dates, symbols, and abbreviations can exist, that are pronounced differently when used in different context. Hence, a semantic tagger is responsible for identification of the token type in the context.

## Part of Speech Tags

---

Four basic POS tags are assigned to the out-of vocabulary words: NN (Noun), NMD(Nominal Modifier), PR(Pronoun), and VB(Verb). Trigram language model is used to select among the above mentioned tags for the out-of vocabulary words.

## Phonetic Lexicon

---

Pronunciation lexicon is an important part of natural language processing as it is used in assigning phonemic transcription to the words. The developed pronunciation lexicon for Urdu consist of threeparts;

- **One is the Urdu word**

- **Second part is its POS tag**
- **Phonemic transcription in CISAMPA**

Words have been extracted from different sources and the IPA symbols are used for transcription of these words. A utility has been developed to convert IPA transcription into CISAMPA.

## Evaluation

---

Evaluation used in this paper are evaluated by **humans** and **Mean opinion Score**.