

## **CLE SEMINAR SERIES-III**

**Topic:** ASR Data Cleaning Guidelines

**Presenter:** Miss. Asima Hameed

**Presentation Date:** 6<sup>th</sup> May, 2014.

**Abstract:** ASR data cleaning guidelines are developed for cleaning of speech data. In these guidelines silence marking, noise, pronunciation and signal to noise ratio are evaluated in recorded files. The process aims to figure out the word without any outside disturbance so the word can be well acoustically analyzed. Praat is used for speech data analysis.

These guidelines are being used to clean the speech data recorded from different districts of Pakistan. The vocabulary is 139 district names of Pakistan that have been recorded over telephonic channel. Every speech file is evaluated on certain parameters regarding noise, pronunciation and signal to noise ratio. The recordings are conducted in natural atmosphere at 8000 Hz.