Cepstrum

The classical cepstrum is defined as the Inverse Discrete Fourier Transform (IDFT) of the log magnitude spectrum and is given by

$$x(n) = \int_{-\pi}^{\pi} \ln X(w) e^{j\omega n} d\omega$$

The cepstrum exists in a domain defined as *quefrency* which has units of time.

The mel-frequency cepstral coefficients used in a speech recognition front end are generated in a different manner. The cepstrum is obtained by doing a Discrete Cosine Transform (DCT), because the log magnitude spectrum is a real symmetric function.

$$C[k] = \sqrt{\frac{2}{N}} c_n \sum_{n=0}^{N-1} S_{avg}(n) \cos\left[\frac{\pi n(2k+1)}{2N}\right]$$

where N represents the total number of filterbank amplitudes, S_{avg} (n) represents the log filter bank amplitudes, $k = 0,1,\ldots,N$, and $c_n = 1/\sqrt{2}$ when $n = 0$, and $c_n = 1$ elsewhere. The resulting coefficients are an approximation to the classical cepstrum that represents an orthogonal and compact representation of the log magnitude spectrum.

Statistical Significance Tests

NIST uses several techniques to determine whether the difference in Word Error Rate (WER) difference between two systems is statistically significant.

MAPSSWE Test: MAPSSWE stands for MAtched Pairs Sentence-Segment Word Error test. This is a parametric test where the utterances in error are compared as a whole or in segments. The errors are assumed to be normally distributed and hence this test is essentially a t-test to estimate the mean difference of normal distribution with unknown variances.

Sign Test: This is a simple test that compares WER on a particular speaker in the test or a specified subset of the test set. It reports which system performs better on the subset of interest. Though this is very simple, its is very useful to analyze if there is a degradation of the performance in a particular subset (however small the magnitude of difference may be)

Wilcox Signed-Rank Test: This is similar to the Sign Test and is done on a particular subset of the test set. The difference is that it also takes into account the magnitude of the performance difference between the two systems.

McNemar Test: This is mainly applied to analyze the statistical significance of difference in performance two systems for a discrete test set. Test sets where we

evaluate an utterance as right or wrong only can be viewed using this test set. This test forms a nice metric to report statistical significance of sentence error rate.