**Spectral Subtraction with Time-Frequency Filtering for Speech Enhancement**

***Abstract***

To develop a speech-to-text (STT) system using Kaldi speech recognition toolkit for continuous Kannada language/dialects. A continuous Kannada speech data is collected from 100 speakers/farmers of Karnataka state in field. The lexicon/dictionary and set of phenomes for Kannada language/dialects are created and transcribed the collected speech data using transcriber tool. The ASR models are developed at different phenome levels using Kaldi. In this work, an effort is made to develop a robust small vocabulary STT system for continuous Kannada language using Kaldi. The various acoustic modelling techniques are used to develop a robust ASR model and achieved a word error rate (WER) of 0.23%. The performance of the ASR model is analyzed by offline speech recognition.

***Keywords:*** Spectral Subtraction, time-frequency, PESQ, SNR, NCM.

1. **Introduction**

Speech enhancement is the fundamental application of speech processing. Spectral subtraction (SS) [1]

is by far the most popular method in speech enhancement, possibly because of its simplicity. A well-known shortcoming of the SS algorithms is the resulting residual noise consisting of musical tones. To overcome the musical noise problem, spectral smoothing has been suggested but it results in low resolution and variance [2]. In our previous work, an amalgamation of SS-VAD and linear predictive coding system to advance the SNR and enhanced audibility featured of encoded speech data was proposed [3]. It was observed that the resulting musical noise due to SS had an adverse effect on encoding performance.

1. **Methodology**

In this section, for completeness, we precisely describe the SS-VAD and implement the proposed SS-

TF for speech enhancement.

(1)

The Fourier transform of the above equation is

(2)

The Hanning window can be mathematically represented as

(3)

All these parameters were used for calculating the factor Z is given by

(4)

The spectral subtraction output can be written as follows:

(5)

The residual noise reduction is done by the mathematically modelling shown below.

(6)

(7)

The block diagram of proposed technique for background noise reduction is shown in Figure 1.

**Noise Estimation**

**Hamming window**

**Noisy Speech**

+

**FFT**

**Half wave**

**Rectification**

**Time-frequency**

**filtering**

**Attenuation**

**during non**

**speech activity**

**Enhanced speech**

**Overlap and Add**

**from SS method**

Figure 1: Block diagram of proposed speech enhancement technique.

Table 1: PESQ values for proposed and existing methods for the assessment of speech quality.

**Algorithm Types of Noise 0dB 5dB**

Airport 1.9085 2.1752

Exhibition 1.6571 1.9992

SS-VAD Restaurant 1.9950 2.0314

Station 1.6517 2.1396

Airport 1.9501 2.1758

Exhibition 1.6614 2.0123

Proposed (SS-TF) Restaurant 2.0900 2.0364

Station 1.6801 2.1294

1. **Conclusion**

We proposed the SS-TF filtering method for speech enhancement with promising results over the SS-VAD technique. Consistent speech quality and speech intelligibility performance for different noise types

and SNR levels was observed

**References**

[1]. Kuldip Paliwal,Kamil W\_ojcicki, and Belinda Schwerin. Single-channel speech enhancement using

spectral subtraction in the short-time modulation domain. Speech communication, 52(5): 450-475,

2010. DOI:10.1016/j.specom.2010.02.004

[2]. Radu Mihnea Udrea, Nicolae D Vizireanu, and Silviu Ciochina. An improved spectral subtraction

method for speech enhancement using a perceptual weighting filter. Digital Signal Processing,

18(4):581-587, 2008. DOI:10.1016/j.dsp.2007.08.002

[3]. Harald Gustafsson, Sven E Nordholm, and Ingvar Claesson. Spectral subtraction using reduced delay

convolution and adaptive averaging. IEEE transaction on speech and audio processing, 9(8):799-

807,2001. DOI:10.1109/89.996083