

Speech is a non-stationary signal and contains modulations in both spectral and temporal domains. Based on the type of modulations studied, most speech processing algorithms can be classified into short-time analysis algorithms, narrow-band analysis algorithms, or joint spectro-temporal analysis algorithms. While traditional methods of speech analysis study the modulation along either time (Short-time analysis algorithms) or frequency (Narrowband analysis) at a time. A new class of algorithms that work simultaneously along both temporal as well as spectral dimensions, called the spectro-temporal analysis algorithms, have become prominent over the past decade.

Joint spectro-temporal analysis (also referred to as 2-D speech analysis) has shown promise in applications such as formant estimation, pitch estimation, speech recognition, etc.

Over the past decade, 2-D speech analysis has been independently motivated from several directions. Broadly these motivations for 2-D speech models can be grouped into speech-production motivated, source-separation/machine-learning motivated and neurophysiology motivated.

In this thesis, we develop 2-D speech model based on the speech production motivation. The overall organization of the thesis is as follows: We first develop the context of 2-D speech processing in Chapter one, we then proceed to develop a 2-D multicomponent AM-FM model for narrowband spectrogram patch of voiced speech and experiment with the perceptual significance of number of components needed to represent a spectrogram patch in Chapter two. In Chapter three we develop a demodulation algorithm called the inphase and the quadrature phase demodulation (IQ), compared to the state-of-the art sinusoidal demodulation, the AM obtained using this method is more robust to carrier estimation errors. The demodulation algorithm was verified on call voiced sentences taken from the TIMIT database. In chapter four we develop a demodulation algorithm based on Riesz transform, a natural extension of the Hilbert transform to higher dimensions, unlike the sinusoidal and the IQ demodulation techniques, Riesz-transform-based demodulation does not require explicit carrier estimation and is also robust to pitch discontinuous in patches. The algorithm was validated on all voiced sentences from the TIMIT database. Both IQ and Riesz-transform-based methods were found to give more accurate estimates of the 2-D AM (relates to vocal tract) and 2-D carrier (relates to source) compared with the sinusoidal modulation. In Chapter five we show application of the demodulated AM and carrier to pitch estimation and for creation of hybrid sounds. The hybrid sounds created were found to have better perceptual quality compared with their counterparts created using the linear prediction analysis. In Chapter six we summarize the work and present with possible directions of future research.