# Introduction

Speech is a complex signal containing comprehensive information related to intent, identity, emotion, and semantics and so on. Traditional speech systems such as speech recognition are effective on professionally recorded scripts: however, their performance is poor in the case of natural speech. This is primarily due to the difficulty in modeling and characterizing the ambiguity presented by emotions in natural speech. A corollary: It is these very emotions that allow for speech to be realistic. People extensively convey their intentions through emotions expressed in speech. The same textual phrase may express different semantics by incorporating different emotions. It is remarkable that humans are able to understand the intended message by perceiving the underlying emotions in addition to the phonetic information. It is apparent that there is a need to develop speech systems that can classify emotions along with the message. The essential goals of emotional recognition from speech are:

1. Recognize emotions present in speech.
2. Synthesize compelling emotions in speech.

Understanding speech emotions can be perceived as the classification of emotions while synthesis can be viewed as incorporating emotion specific domain knowledge during speech synthesis. The ability to effectively accomplish these two goals will facilitate novel human-computer interactions as well as, we believe, alleviate the contemporary conventions around human to human communications.

## Understanding Emotion

Due to the importance of emotion in our society, it has been studied in several scientific disciplines such as Biology, Psychology, Neuroscience, Anthropology, Communication and so on. As a result, distinctive perspectives on the concept of emotion have emerged, appropriate to the complexity and variety of the emotions. Most models are complementary rather than competitive. Classically, emotions can be viewed as a complex experience of consciousness, bodily sensation and behaviour. The emotions generally represent a synthesis of subjective experience, expressive behaviour and neurochemical activity. Correspondingly, there are more than 300 crisply identified emotions by researchers [1,2]. Generally, only a small subset of hese emotions are experienced on a daily basis. In this regard, most researchers are in agreement on the principles of the Palette theory. That is, any emotion is the composition of six primary emotions as any color is the combination of 3 primary colours [3]. Anger, disgust, fear, happiness, sadness and surprise are considered as the primal emotions by most researchers. These are also referred to as archetypical emotions.

When considering the production of emotions and its physiological production mechanisms, it has been illustrated that the nervous system is stimulated by the activity of high excitation emotions such as anger, happiness and fear. This chemical process causes a variety of physiological symptoms including an increased heart rate, higher blood pressure, changes in respiration pattern, the dryness of the mouth and so on. The resulting speech is correspondingly faster, louder and characterized with stronger high-frequency energy, higher average pitch and wider pitch range [4]. Conversely, in low excitation emotions such as sadness, the nervious systems is stimulated such that there is a decrease in heart rate, blood pressure which results in increased salivation, slow and low-pitched speech with little high-frequency energy. It is clear that acoustic features and the articulation of the speech signal contain highly correlated information regarding underlying emotions. It is important to note, however, that differentiation emotions is yet an arduous task as many emotions cause similar levels of stimulation the nervous system. From a speech signal perspective, there is significant physiological changes in speech patterns due to emotional experiences. In scientific literature, a speech signal is produced from the contribution of the vocal tract system excited by an excitation source signal [5]. In fact, the characterics of speech related to emotion can be attributed to:

1. The excitation source
2. The sequence of shapes of the vocal tract system
3. Supra-segmental characteristics (duration, pitch, energy)
4. Linguistic information.

## Understanding Speech

The speech production mechanism is a subtle process with tight couplings which result in utterances of sound which are typically perceived as a person speaking. Specifically, the air pressure created in the trahea gets released into the vocal tract and nasal cavity, in the form of air puffs. This pressure created beneath the vocal folds causes their vibration (glottal activity). This air flow is then chopped into a sequence of quasi-periodic pulses through the vibration of vocal folds. The resulting sequence of implse-like excitation is transformed intot different frequency components while passing through the pharynx, oral and nasal cavities. The vibration of coval folds serves as the major source of excitation the vocal tract. The oral cavity the behaves as a pseudo time varying resonater to enchance he intended frequency components. Certain articulator organs of the vocal and nasal tracts modulate the sequence of air puffs into speech that gets radiated and amplified through the lips. The vibration of the cocal folds produces the unvoiced sound units. The position of various articulators and the voicing process controls the different sound units that are produced. Additionally, the characteristics of the glottal activity and vocal tract shapes play a major role I the modulation of various emotions during product to speech. Figure 1 illustrates the human speech product system.

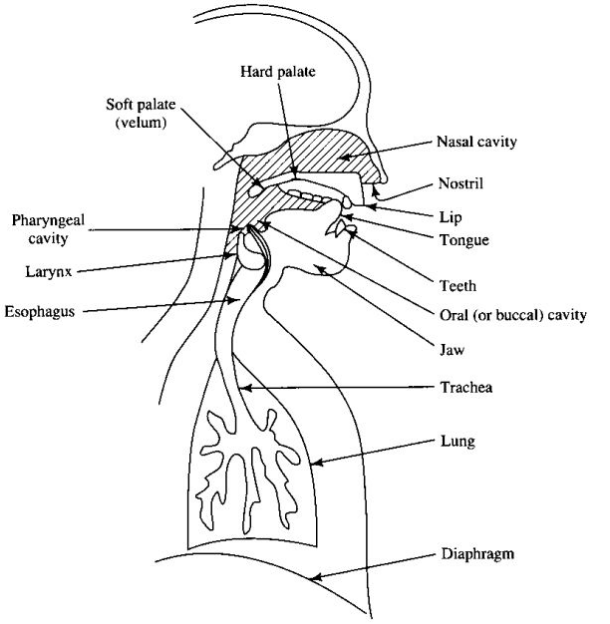


Figure 1 Human Vocal System

# State of Art