# Literature Review

## Speech Processing

Automated speech recognition systems can be broken up into numerous categories based on the various criteria they meet. The criteria for systems usually consist of; speaker dependence, speech type, and recognition type,.

### Speaker dependence

whether the system is designed to be speaker dependent vs speaker independent;

### Speech type (continuous, discontinuous, isolated)

whether the system is designed to recognise singular words, sentences that are purposefully broken up by pauses, or sentences that are spoken naturally;

### Recognition type (word vs. phoneme)

A phoneme can be thought of as “the smallest contrastive linguistic unit which may bring about a change of meaning”[1] in a word.

whether the system performs recognition at the word or phoneme level

## Acoustical Pre-Processing

## Feature Extraction

The two most important speech characteristics are those contained in the spectral envelope (vocal tract characteristics) and those contained in the supra-segmental features (voice source characteristics) of speech. [2]

### Mel-frequency Cepstral Coefficients (MFCC)

## Classification Techniques

### Hidden Markov Models

### Artificial Neural Networks

### Dynamic Time Warping

## Robust speech recognition

## Auditory data

# References

[1] A. Cruttenden, "Gimson's pronunciation of English," 7th ed: Routledge, 2013, p. 41.

[2] S. Furui, "Recent advances in speaker recognition," in *Audio- and Video-based Biometric Person Authentication*. vol. 1206, J. Bigün, G. Chollet, and G. Borgefors, Eds., ed: Springer Berlin Heidelberg, 1997, pp. 235-252.