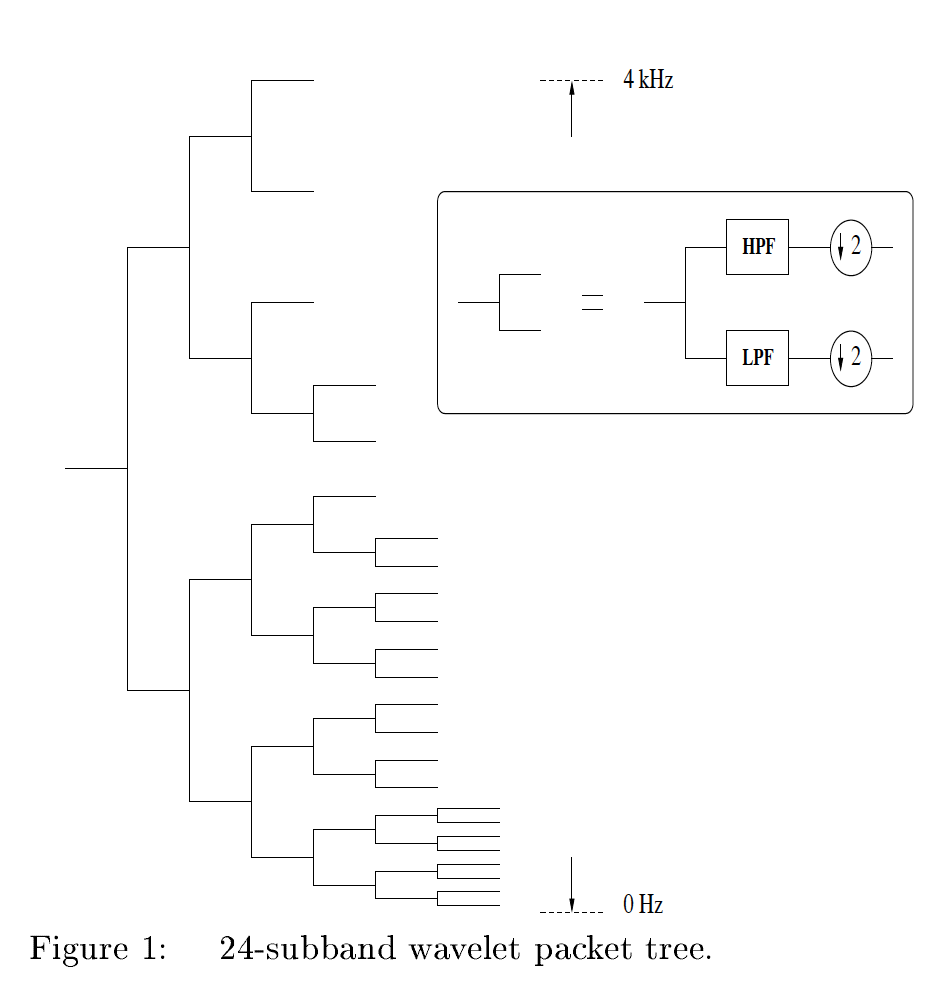
Good datasets: NIST Speaker Recognition Evaluations (SREs).(The goal of the NIST Speaker Recognition Evaluation (SRE) series is to contribute to the direction of research efforts and the calibration of technical capabilities of text independent speaker recognition.), TIMIT (6300 sentences spoken by 630 speakers)

DWT features

* Compare SBC (subband based cepstral), WPP, MFCC on TIMIT, both method is classified using GMM
* WPP is the best
* SBC: 24 subbad wavelet packet tree (provide, found by experiment) that approximates Mel-scale frequency division. Resulting subband emphasize 0~500Hz
* WPP: Daubechies’s 4tap filters, 3 level, shown in [6,9]
* 
* ~24ms frame, 10ms skip rate
* MFCC config can be found in [3]
* Used Daubechies’s orthogonal (key to decorrelation) filters
* Compute filterbank energies -> decorrelation of log energies with a DCT
* Wavelet packet tree -> smoother, that’s why it outperforms MFCC
* GMM rely on the fact that feature vectors are independent of each other (decorrelation is important)
* Very high accuracy (~95%)
* 32nd order Daubechies’ orthogonal filters to do wavelet packet transform

dwt packet irregular decomposition (can't find FB)

* energy based
* used ANN (GRNN)
* amplitude is normalized by mean and std
* Used irregular decomposition of WPT, compared with DWT, WPT, MFCC
* Better accuracy, same computational time/cost
* Proposed method can be further optimized (don’t have to use all features)
* Do transform, then calculate energy index
* GRNN architecture provided
* DWT, 7 level, 8 coeff
* WPT, tried 6 levels, 64 features
* Irregular decomposition: center freq given (8 level, tree can be redrawn)

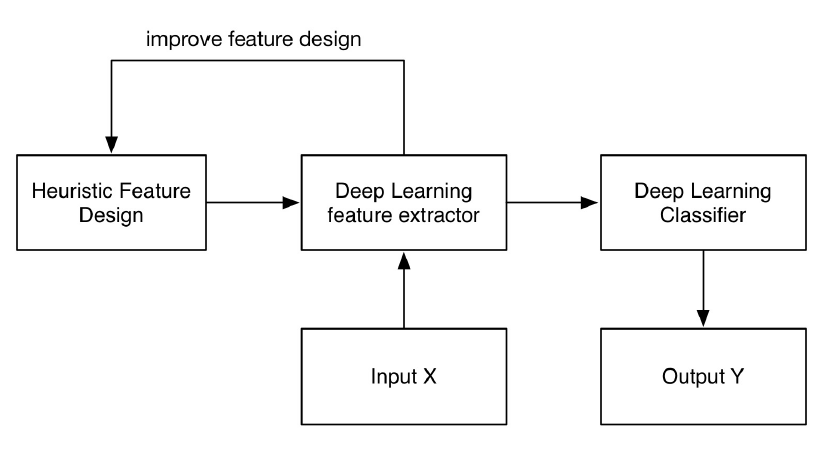
Using vowels features through a combined method of formants, wavelets, and neural network classifiers

* Formants and seven Shannon entropy
* 12 features
* Decompose the vowel signal at WP depth of level **two** with **Daubechies** type and calculate the Shannon entropy for eachsub-signal. The WP extracts additional features to the Shannonentropy and therefore enhances the recognition rate.

Invariant Scattering Convolution Networks

* Also called wavelet scattering
* Translation invariant
* ???

FILTER BANK LEARNING LAYER

* has a layer that train feature extractor
* 
* not what I think it is. They are learning the weights for each filter bank

LEARNING FILTER BANKS WITHIN A DEEP NEURAL NETWORK

* also learn the weights for each filter bank