Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease

by Athanasios Tsanas et al.

Review by Meghna Ranjit

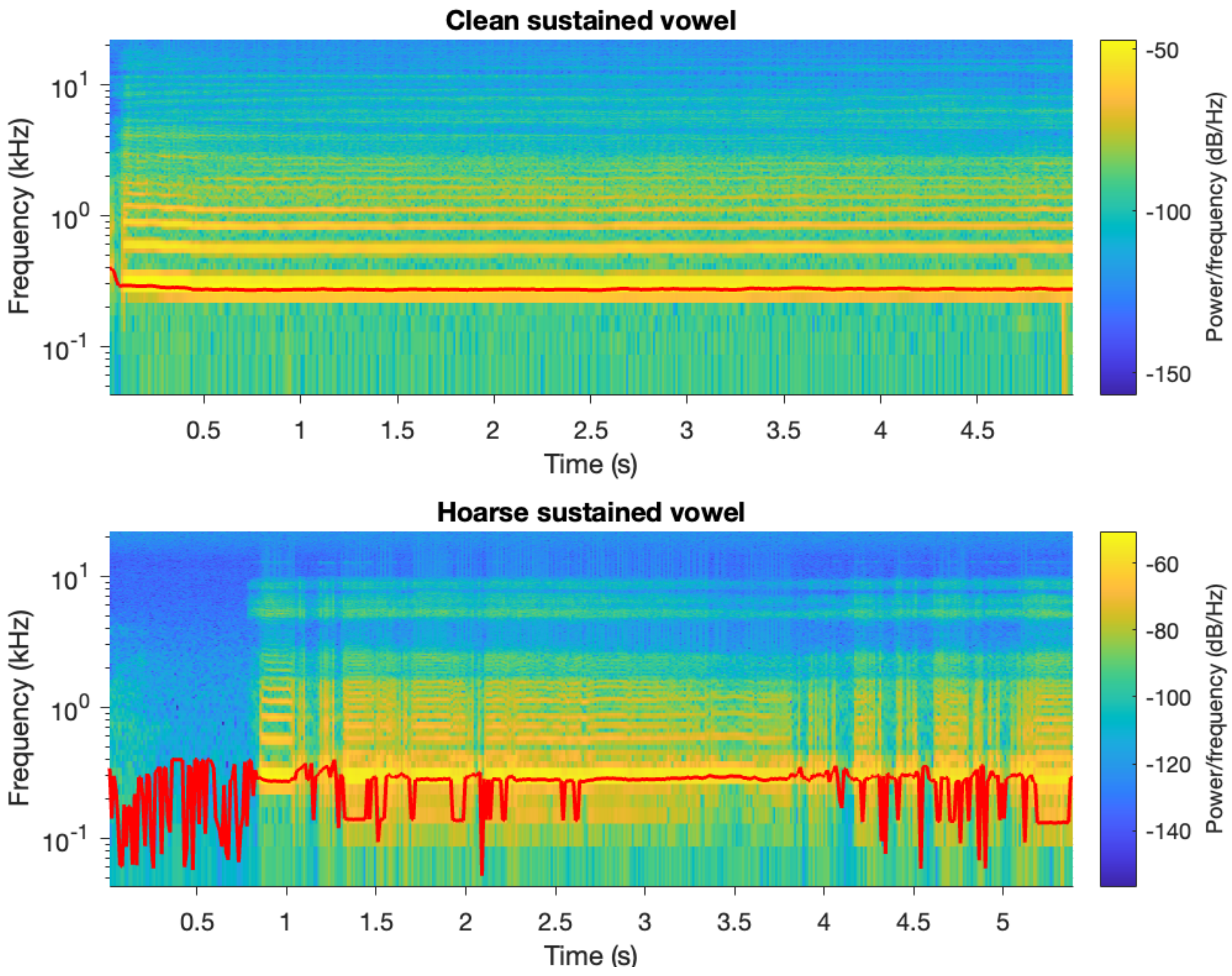
Degradations in speech due to PD

- Parkinson's disease results in symptoms including tremor, rigidity and loss of muscle control in general, as well as cognitive impairment.
- Impairment in the normal production of vocal sounds (*dysphonia*)
  - reduced loudness, breathiness, roughness, decreased energy in the higher parts of the harmonic spectrum, and exaggerated vocal tremor
- Problems with the normal articulation of speech (*dysarthria*)

- This paper gathers other relevant features for use to classify sustained vowels of Parkinson's disease patients from healthy controls

Identifying the fundamental frequency

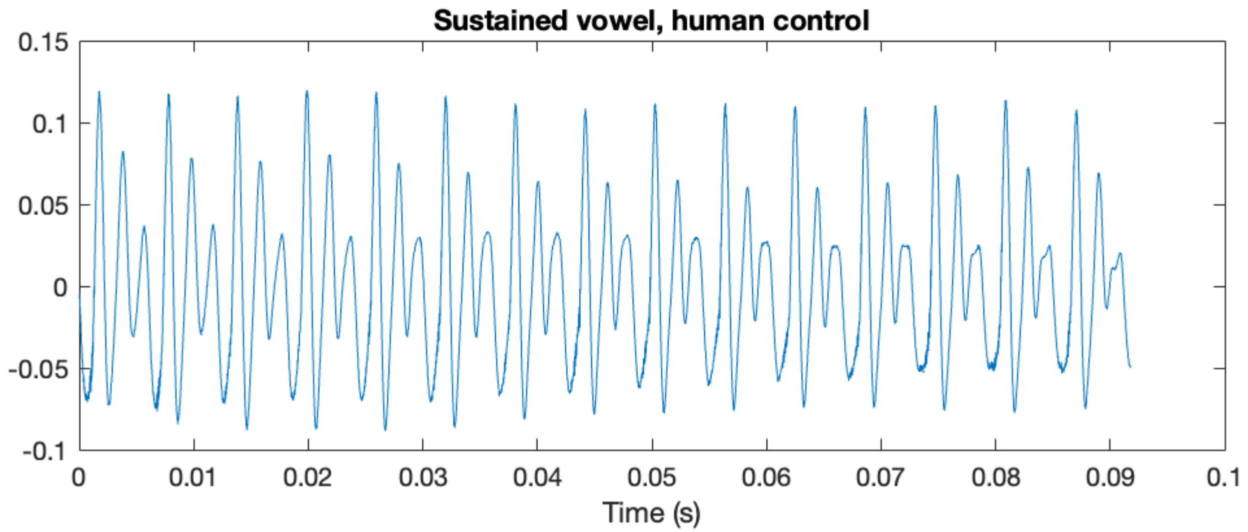
- Defined as the approximate frequency of the periodic waveform of voiced speech
- Critical to some subsequent dysphonic measures, which can be a problem
- Algorithms often miscompute F0 as twice or half actual values
- Paper computes F0 with 3 different algorithms (the Praat algorithm, Sun's algorithm, and the RAPT algorithm), combine all three of their outputs
- F0 statistics and divergence from healthy controls of the same age & gender used as features



Jitter

- Cycle-to-cycle fundamental frequency perturbations

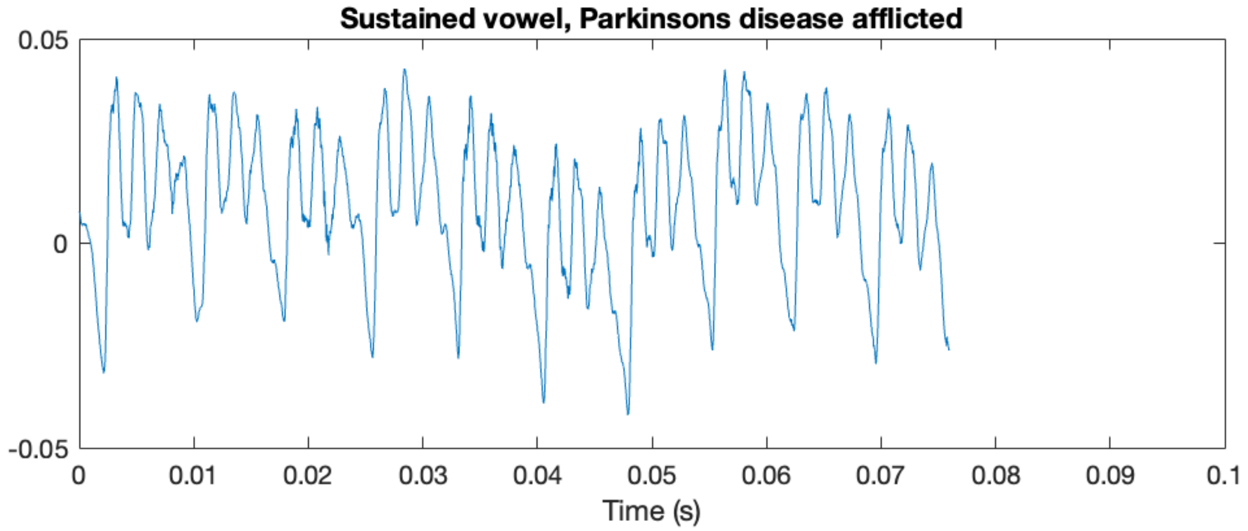
$$Jitter_{\mathrm{F_{0,abs}}} = \frac{1}{N}\sum_{i=1}^{N-1}\left|F_{0,\mathrm{i}} - F_{0,\mathrm{i+1}}\right|$$



Shimmer

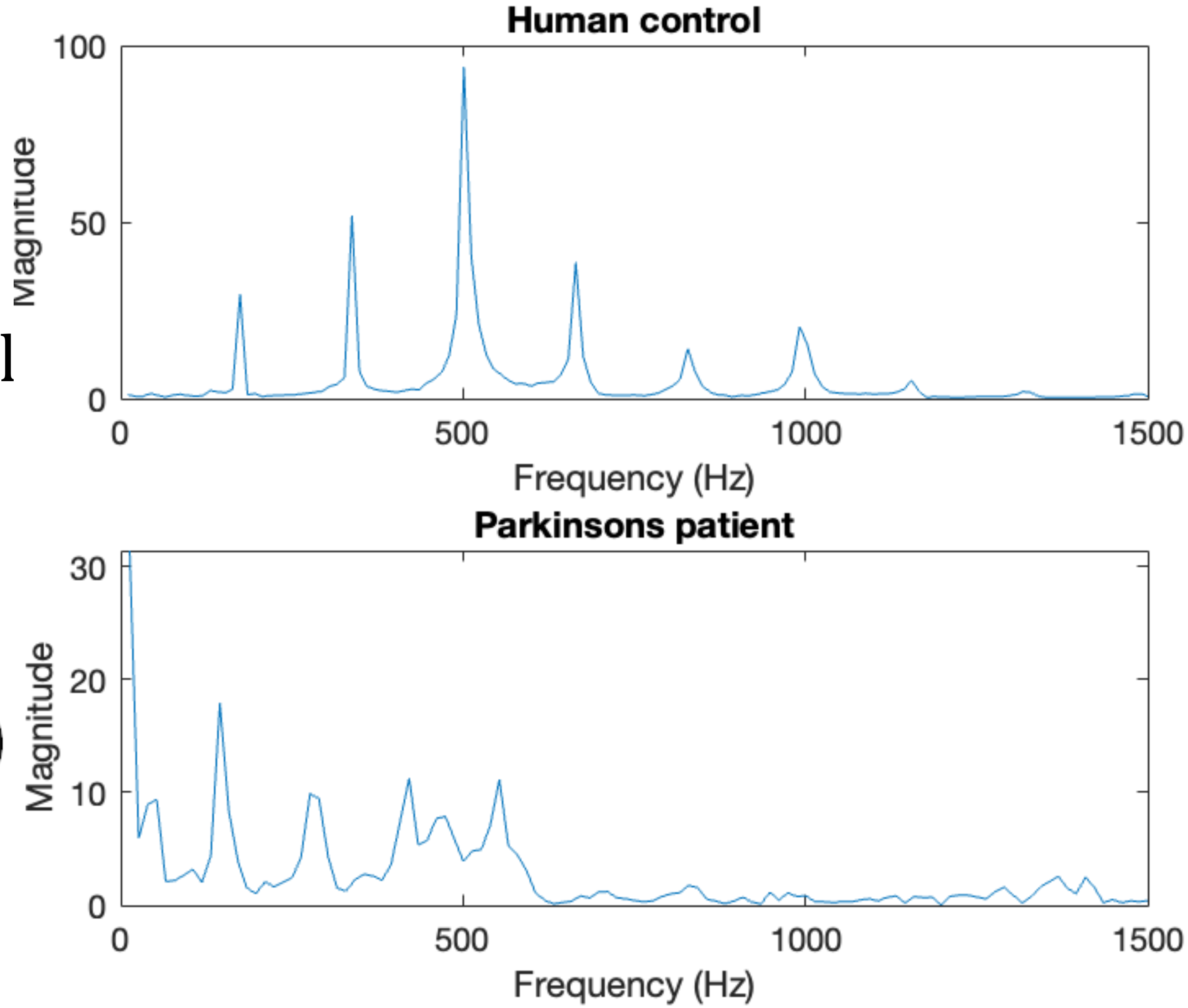
- Cycle-to-cycle amplitude perturbations

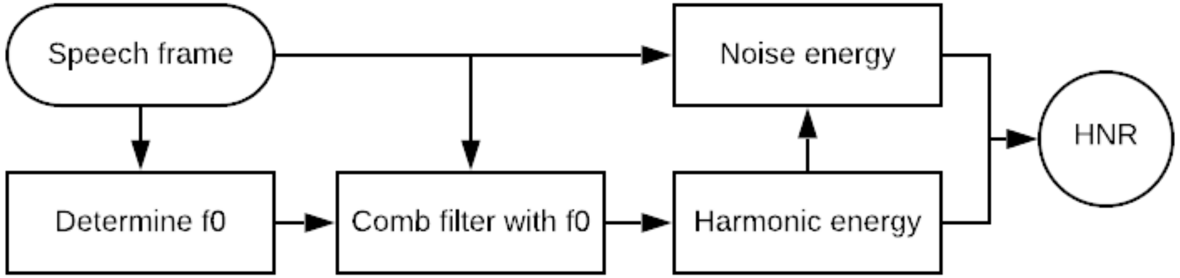
$$Shimmer_{\mathrm{dB}} = \frac{1}{N}\sum_{i=1}^{N-1} 20 \cdot \left|\log_{10}\frac{A_{0,\mathrm{i}}}{A_{0,\mathrm{i+1}}}\right|$$



Harmonics to Noise ratio

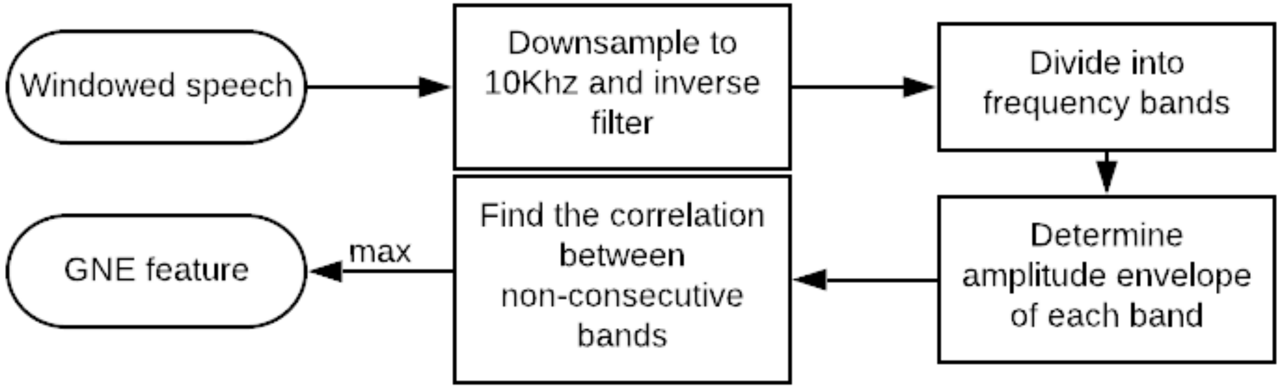
- The amount of noise in the speech signal.
- In PD patients, this results mainly due to incomplete vocal fold closure.

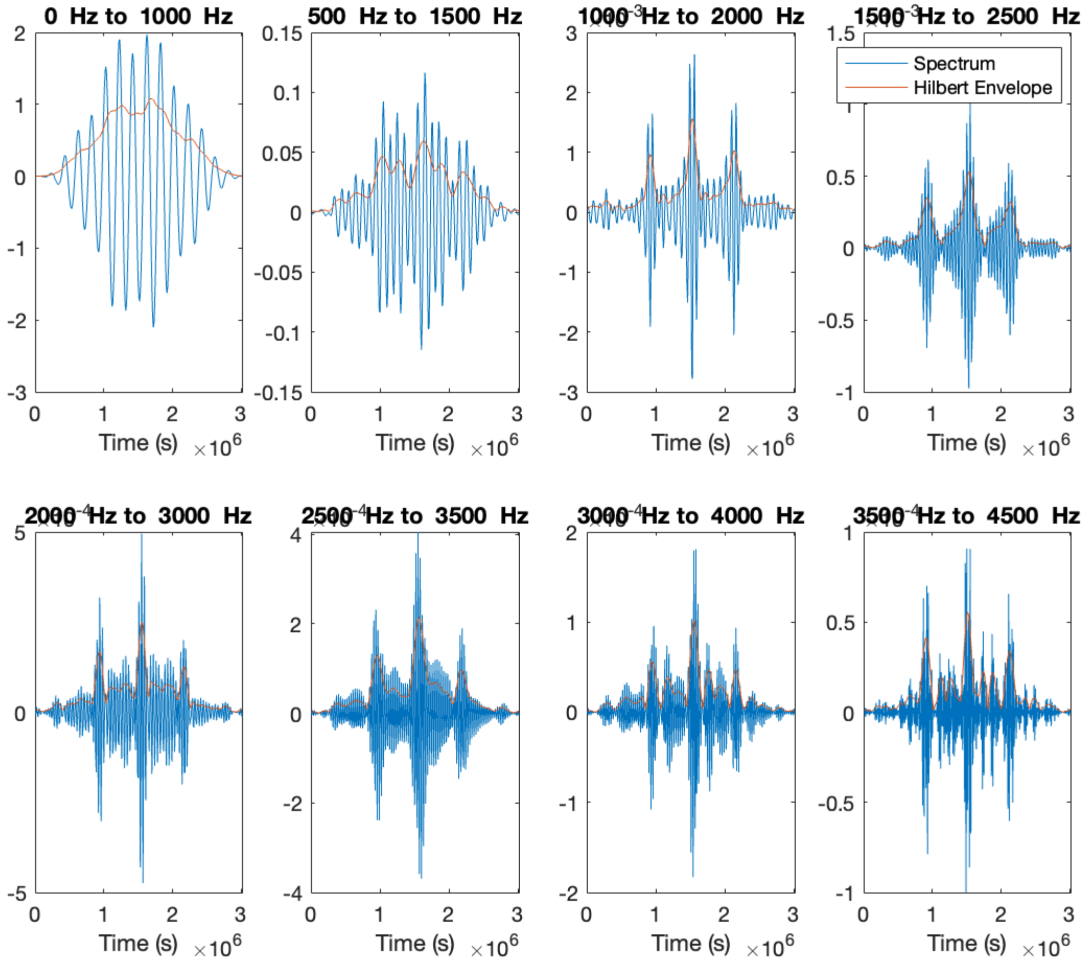




Glottal-to-Noise Excitation Ratio

- If the excitation is pure glottal closure, it would have similar energy patterns throughout the spectrum.
- Improper closure results in noise and dissimilar energy patterns





EMD excitation ratio

- EMD decomposes a multi-component signal into elementary signal components, known as *intrinsic mode functions*
- First few IMFs are high-freq components (often noise)

$$\mathrm{IMF_{SNR}} = \frac{\sum_{d=4}^{D} \mu_d}{\sum_{d=1}^{3} \mu_d}$$

- μ_d represents the mean values of SEO, TKEO, or Shannon's entropy

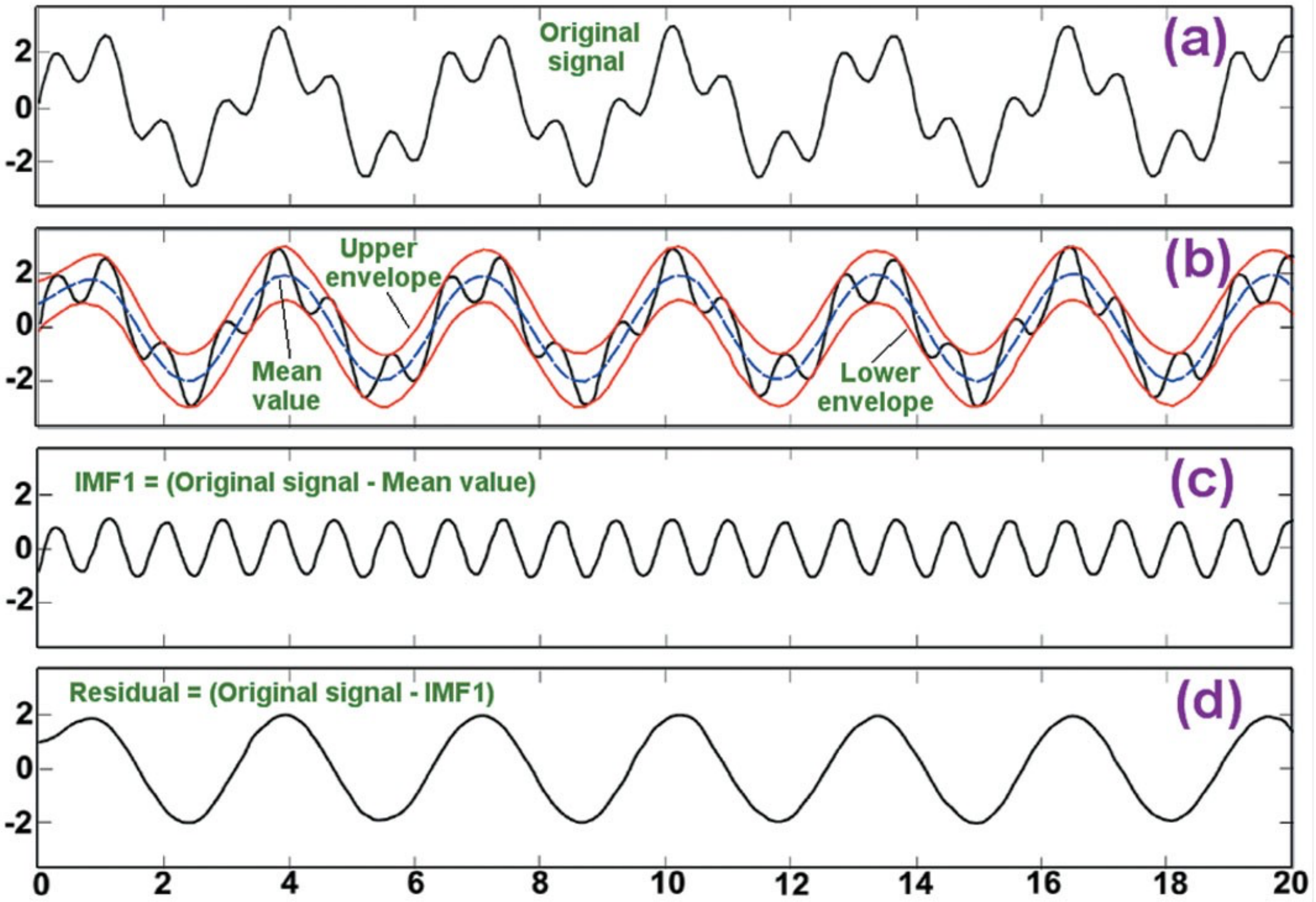


Image source: https://towardsdatascience.com/decomposing-signal-using-empirical-mode-decomposition-algorithm

Glottis Quotient

- Use DYPSA to calculate standard deviation of open glottis interval and closed glottis interval, as well as ratio between 95 percentile and 5 percentile estimates of ratio of open interval to cycle duration
- Close resemblance to jitter.

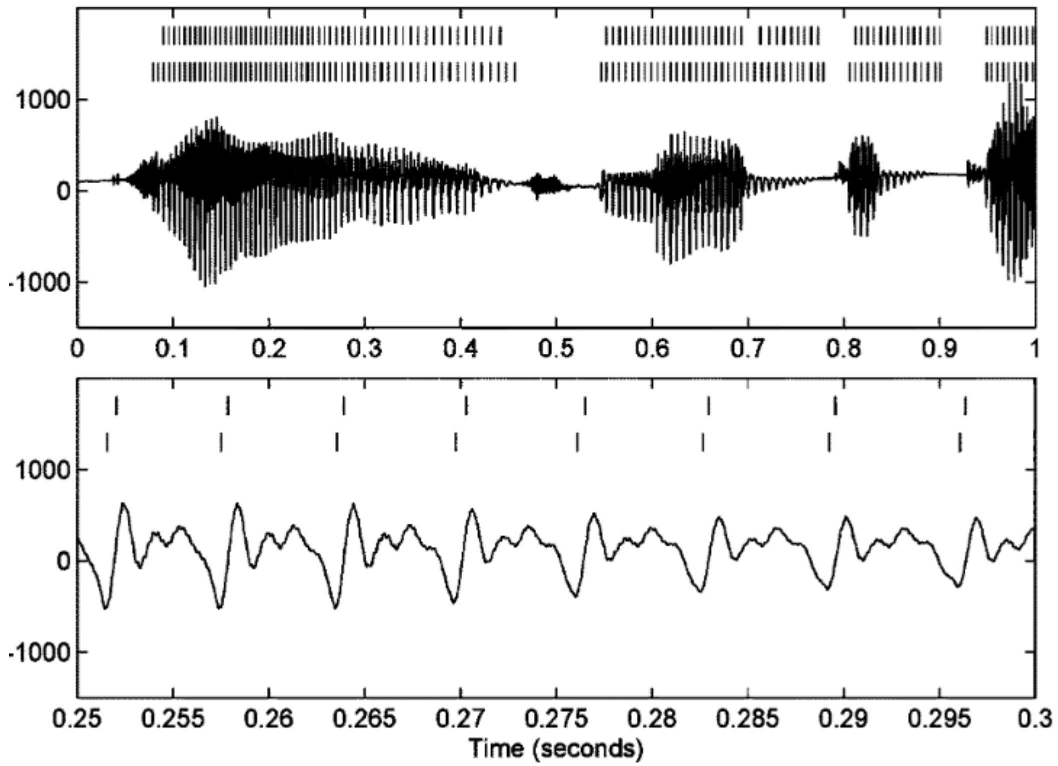
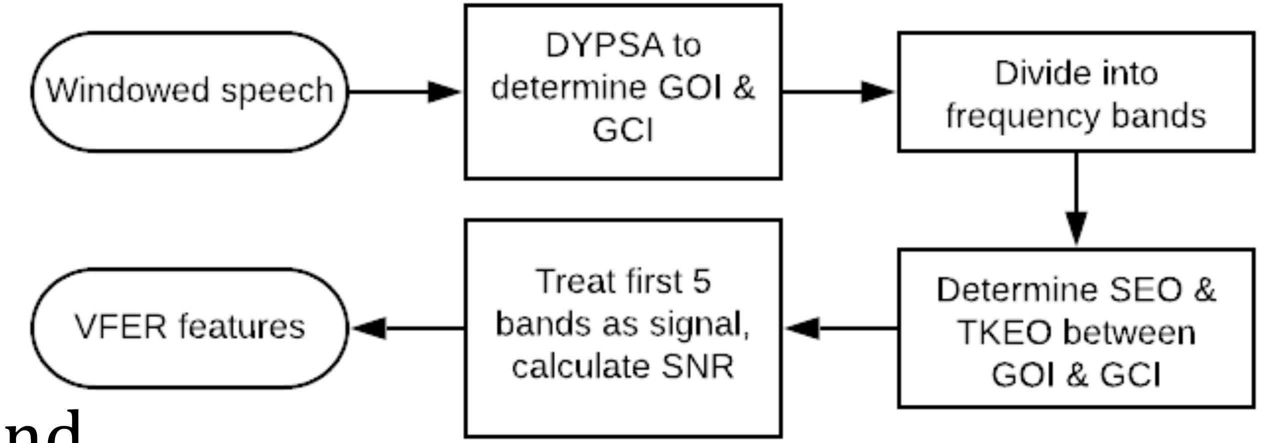


Image shows Glottal closing instances identified from speech signal through the DYPSA algorithm (top) and through an electroglottograph signal (bottom)

Image source: Naylor, Patrick A., et al. "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm." *IEEE Transactions on Audio, Speech, and Language Processing* 15.1 (2006): 34-43.

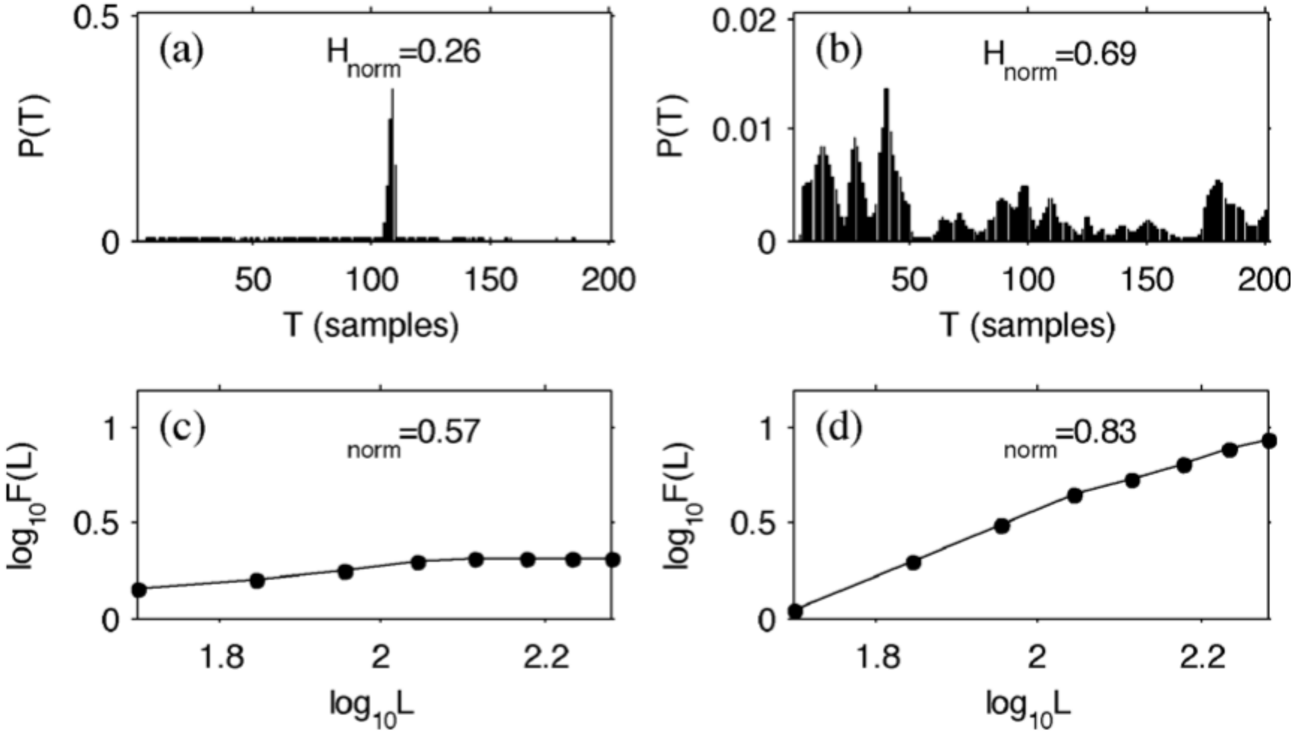
Vocal Fold Excitation Ratios

- Similar to Glottal-to-Noise excitation ratio features.
- Use DYPSA instead of inverse filtered speech
- No downsampling, in order to preserve high frequency detail
- SEO and TKEO used to characterize content of each band
- SNR + correlation mean and standard deviation are extracted as features



RPDE, DFA

- These features are based on *nonlinear dynamical systems* models.
- The *recurrence period density entropy* (RPDE) quantifies the extent to which can be considered as strictly periodic. The deviation from periodicity evaluated by the *entropy H* of the distribution of these recurrence periods P (T)
- The *detrended fluctuation analysis* (DFA) algorithm calculates the extent of amplitude fluctuation (F (L)) of the integrated speech signal over a range of time scales L
- The self- similarity of the speech signal is quantified by the slope α



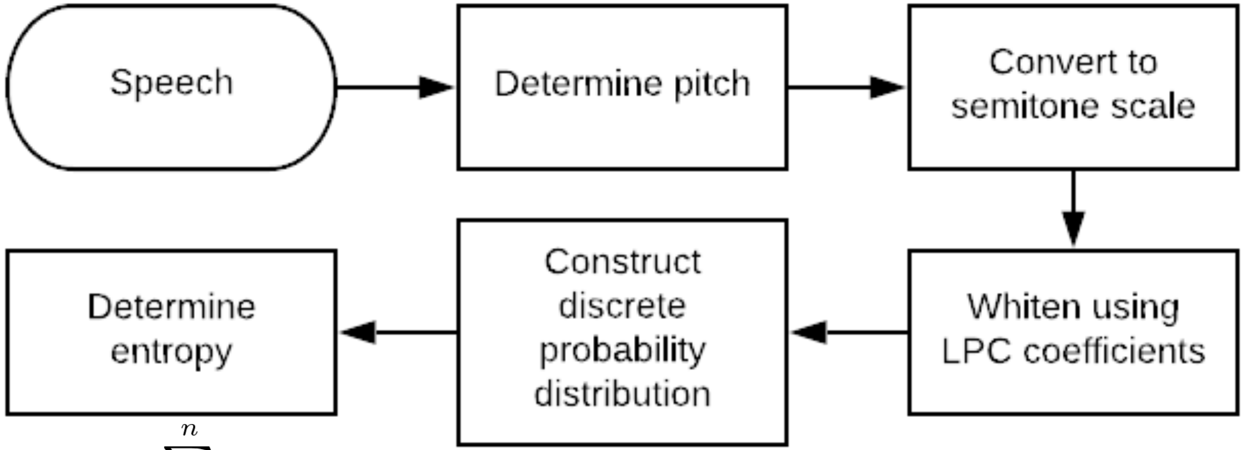
(a) and (b) Recurrence period density P(T) for recurrence times T.
(c) and (d) Log–log plot of scaling window sizes L against fluctuation amplitudes F(L).

Image source Tsanas, Athanasios, et al. "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity." *Journal of the royal society interface* 8.59 (2011): 842-855.

Pitch Period Entropy (PPE)

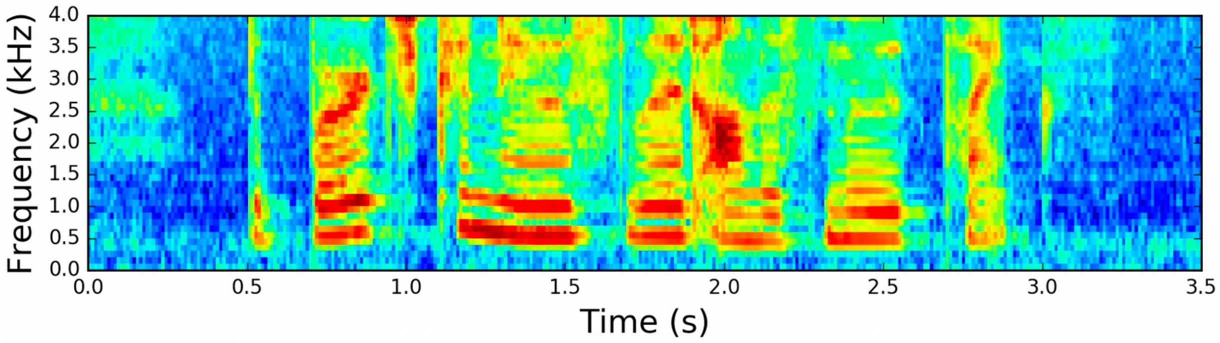
This feature is motivated by two observations

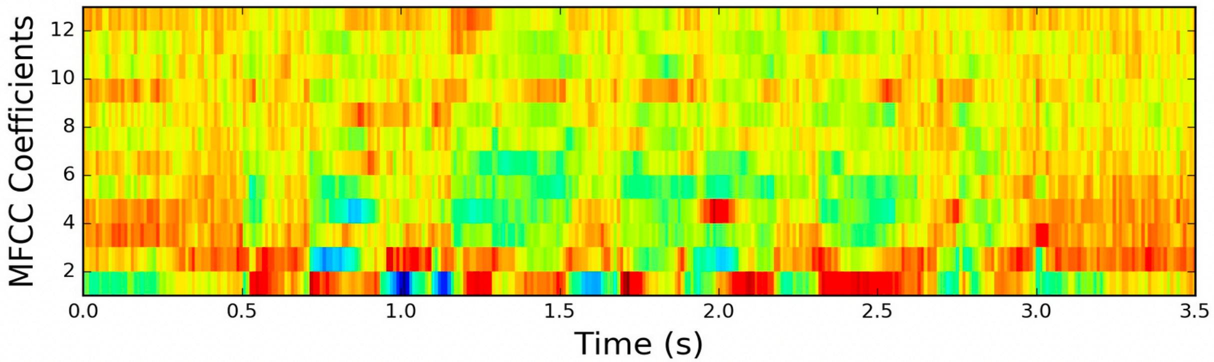
- a more relevant scale on which to assess abnormal variations in speech pitch is the *logarithmic* scale
- In order to better capture pitch period variation independent of natural variations, smooth variations should be removed prior to measuring the extent of such variations.



$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Mel Frequency Cepstral Coefficients (MFCCs)





- MFFCs aim to decompose the speech signal into vocal fold and vocal tract elements
- Good indicators of potential problems in the articulators.
- Detect misplacement of the articulators. PD is known to affect the articulation

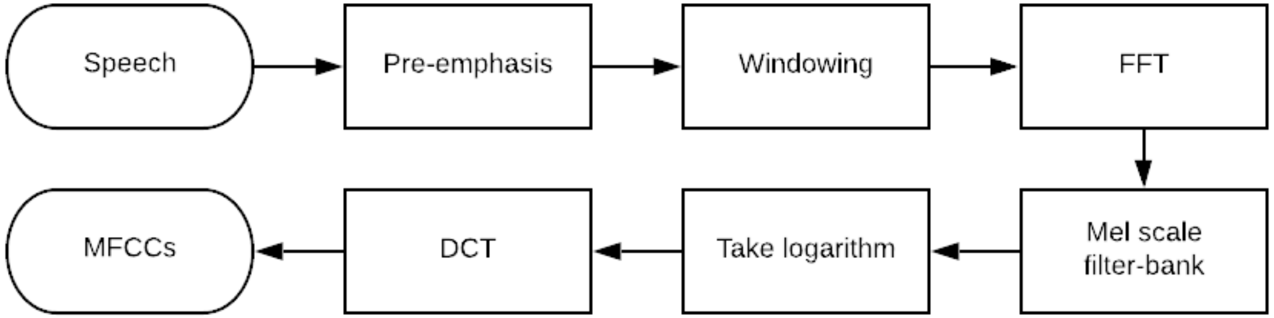
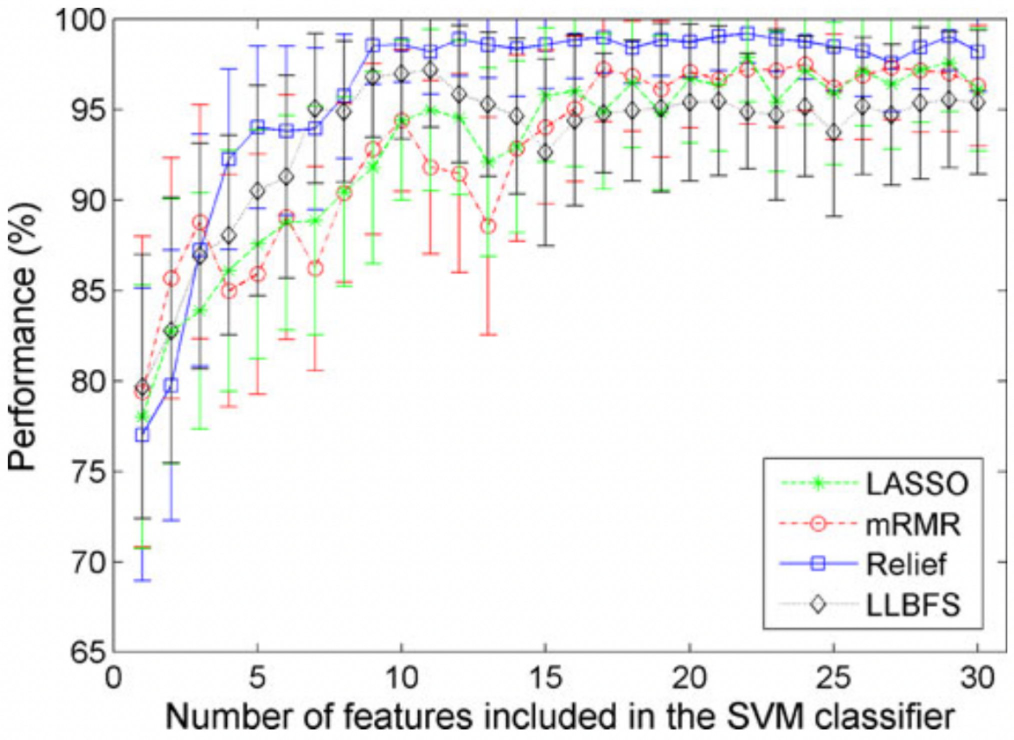


Image source: https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

Feature reduction

- *Feature selection* (FS) *algorithms* which offer principled approach to reduction of the number of features.
- Four efficient FS algorithms were compared
  1. Least absolute shrinkage and selection operator (LASSO): good performance in sparse (uncorrelated features) setting.
  2. Minimum redundancy maximum relevance (mRMR): greedy algorithm, neglects complementarity
  3. RELIEF: sensitive to feature interactions
  4. Local learning-based feature selection (LLBFS): extension to RELIEF, maximizes margin between classes



Results

- Classification accuracy using SVM with all the features: 97.7%
- Classification accuracy using SVM with ten features selected by RELIEF: 98.6%
- Prior state-of-the-art: Guo et al., Genetic Programming Expectation Maximization using RPDE, PPE and DFA, shimmer, jitter and HNR features: 93.1%

Dysphonia measure	Correlation coefficient
$VFER_{entropy}$	**-0.388**
$VFER_{NSR,TKEO}$	**-0.379**
11th MFCC coef	**0.369**
$VFER_{NSR,SEO}$	**-0.365**
4th delta MFCC	**-0.363**
$VFER_{mean}$	**-0.321**
RPDE	**0.292**
DFA	**0.287**
$Shimmer_{PQ11}$	**0.285**
HNR_{mean}	**-0.285**

References

[1] Tsanas, Athanasios, et al. "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease." *IEEE transactions on biomedical engineering* 59.5 (2012): 1264-1271.

[2] Tsanas, Athanasios, et al. "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity." *Journal of the royal society interface* 8.59 (2011): 842-855.

[3] Sáenz-Lechón, N., et al. "Screening voice disorders with the glottal to noise excitation ratio." (2009).

[4] Little, Max, et al. "Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease." *Nature Precedings* (2008): 1-1.

Appendix: energy estimation schemes

Teager-Kaiser Energy operator (TKEO):

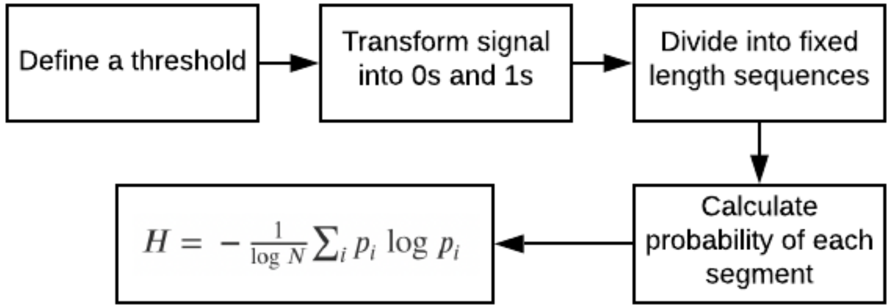
$$\Psi[\mathrm{x}(t)] = \left(\frac{d\mathrm{x}(t)}{dt}\right)^2 - \mathrm{x}(t)\frac{d^2\mathrm{x}(t)}{dt^2}.$$

$$\Psi[\mathrm{x}(\mathrm{n})] = x^2(n) - \mathrm{x}(n+1)\mathrm{x}(n-1)$$

Squared Energy Operator (SEO):

$$S[x(t)] \triangleq x^2(t).$$

Shannon entropy:



Questions

1. *The algorithm to find the fundamental frequency is not robust, and hence features that rely on finding F0 can be problematic.* Which of the features described clearly rely on F0 identification?
2. What can be the reason that reduction in the number of features through Feature Selection algorithms result in a better performance than with the whole feature set?
3. For the following features state which of these three speech degradations they characterize: a) incomplete vocal fold closure (which results in noise) b) irregular vibration of vocal folds c) change in articulation

   Glottis Quotient, EMD excitation ratio, Jitter, Vocal Fold Excitation ratios, Harmonics to Noise ratio, Pitch Period Entropy, Recurrent Period Density Entropy, MFCC, Glottal to Noise excitation ratio, Detrended Fluctuation Analysis
4. Describe in words the simplest procedure to determine the Pitch Period Entropy (PPE). What do larger values of PPE indicate?