A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis

Authors: Augustine H. Gray and John D. Markel

By Kaviraj, Komaljit, Vaibhav
Spectral Flatness

- Spectral Flatness is a measure of the noisiness/sinusoidality of a spectrum.
- For tonal signals it is close to 0 and for noisy signals it is close to 1.

SFM = 0

SFM = 0.51
Define log spectrum of time sequence

Normalized log spectrum of time sequence is given by

\[ V = V(\theta) = \log \left\{ \left| \mathbb{E}[\exp(j\theta)] \right|^2 / r_e(0) \right\}. \]

Where \( r_e(0) \) denotes energy of time sequence, given by

\[ r_e(0) = \sum_{n=-\infty}^{\infty} e_n^2 = \int_{-\pi}^{\pi} \left| \mathbb{E}[\exp(j\theta)] \right|^2 \frac{d\theta}{2\pi} \]

1. A perfectly flat or constant spectrum will yield a normalized spectrum of zero.

2. Nonzero values for \( V \) will represent deviations from a flat or constant spectrum.
Measuring deviations from flatness

• Let’s consider a couple of such measures:

\[ \eta(E) = \int_{-\pi}^{\pi} \frac{1}{2} V^2(\theta) \frac{d\theta}{2\pi} \]

\[ \mu(E) = \int_{-\pi}^{\pi} \left\{ \exp[V(\theta)] - 1 - V(\theta) \right\} \frac{d\theta}{2\pi} \]

\( \eta \)
Weights -ve and +ve excursions of the normalized log spectrum equally

\( \mu \)
Weighs the +ve excursions more heavily and the -ve excursions less heavily
μ OR η?

• As the peaks of speech log spectra (more precisely, the formants) play a more important role than do the valleys in the perception of speech, it would be preferable to use an integrand that is not symmetric, but more heavily weighs the positive excursions of $V$ than the negative excursions.

• Thus, $\mu$ has the desired properties.
As $V(\theta)$ represents the normalized log spectrum of the signal, the average of $e^V$ will be unity. Thus, our initial representation may be simplified and represented as:

$$\mu(E) = \int_{-\pi}^{\pi} \{ \exp[V(\theta)] - 1 - V(\theta) \} \frac{d\theta}{2\pi}.$$  

Thus, $-\mu(E)$ with a factor of 2 represents the zeroth quefrency of the cepstrum and $\exp[-\mu(E)]$ represents the ratio of the geometric to arithmetic means of the spectrum.
$\mathcal{Z}(E)$ – the Spectral Flatness Measure

$$\mathcal{Z}(E) = \exp \left[ -\mu(E) \right] = \exp \left[ \int_{-\pi}^{\pi} V(\theta) \frac{d\theta}{2\pi} \right]$$

With this normalization the spectral-flatness measure, $\mathcal{Z}(E)$ will lie between 0 and 1, and equal 1 for a perfectly flat spectrum.
Spectral-Flatness Transformations

\[ X(z) \rightarrow A_M(z) = 1 + \sum_{k=1}^{M} a_M k z^{-k} \rightarrow E(z) = X(z) A_M(z) \]

Residue calculus can be applied to show that

\[ \int_{-\pi}^{\pi} \log \{|A_M[\exp(j\theta)]|^2\} \frac{d\theta}{2\pi} = 0 \]

Hence we obtain the result

\[ \int_{-\pi}^{\pi} \log \{|E[\exp(j\theta)]|^2\} \frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} \log \{|X[\exp(j\theta)]|^2\} \frac{d\theta}{2\pi} \]
Log spectrum of $X(z)$

Using the previous result, and the definition of the measure of spectral flatness,

$$\mathcal{Z}(E) = \mathcal{Z}(X) \frac{r_x(0)}{r_e(0)} \quad \text{------} \quad (i)$$

To find Filter Coefficients: The results are described in terms of what are often called the $k$-parameters, $k_0$, $k_1$, …, $k_{M-1}$, whose basic property is that each is less than unity magnitude.

$$\alpha_0 = r_x(0) \quad \text{and} \quad \alpha_{m+1} = (1 - k_m^2) \alpha_m$$

for $m = 0, l, \ldots, M - 1$, which recursively leads to

$$r_e(0) = \alpha_M$$

Using these results, we may re-write Equation $(i)$ as

$$10 \log_{10} \mathcal{Z}(X) = 10 \log_{10} \mathcal{Z}(E) + 10 \log_{10} \mathcal{Z}(1/A_M)$$
Log spectrum of $X(z)$ continued...

- It has been shown that the energy of the time sequence associated with $\alpha_M^{1/2}/A_M(z)$ is equal to $r_x(0) = \alpha_0$, so that

$$\Xi(1/A_M) = \alpha_M/\alpha_0$$

$$10 \log_{10} \Xi(X) = 10 \log_{10} \Xi(E) + 10 \log_{10} \Xi(1/A_M)$$

- Thus, one can decompose the log spectrum of $X(z)$ in terms of both the log spectra of $E(z)$ and $1/A_M(z)$, and the spectral-flatness measure as was determined earlier.
# Spectral-Flatness of Two Driving Function Models

## Unvoiced Driving Function Model

- **Driving function:** Uncorrelated Gaussian noise.
- **Spectral Flatness measure** has an expected value of roughly \( \exp(-\Upsilon) \) or \(-2.5\) dB. \((\Upsilon \text{ is Euler's constant})\)

## Voiced Driving Function Model

- **Driving function:** Set of \( L + 1 \) equally spaced samples

\[
y_k = 0 \text{ for } k \neq l, l + P, l + 2P, \ldots, l + LP
\]

where, \( y_l \) is the first sample in the time window, \( y_{Z+P} \) is the last, and \( P \) represents a pitch period

- **Spectral Flatness** \( \in \left[ \frac{(L!)^2}{(2L)!}, 1 \right] \)
  - Eg. For two samples \( (L = 1) \) the measure lies between \(1/2\) and \(1\), or \(-3\) dB and \(0\) dB

- If all samples have the same size, then **Spectral-Flatness measure equals** \(1 / (L + 1)\)
“WILL THE REST FOLLOW SOON”

Spectogram of utterance “Will the rest follow soon.”
Spectral flatness measures at the input to the inverse filter $10\log_{10} \tilde{Z}(X)$

$M = 8, N = 128$

Rectangular Window, $F_s = 6.5$ kHz

Spectral flatness measures at the outputs $10\log_{10} \tilde{Z}(E)$

$M = 8, N = 128$

Hamming Window, $F_s = 6.5$ kHz

$M = 8, N = 128$

Hamming Window, $F_s = 13$ kHz

$M = 16, N = 256$

Hamming Window & Twice the Window Length, $F_s = 6.5$ kHz

$M = 8, N = 256$

Each data window has $N$ points. Thus, time windows have length $N/F_s = N \Delta t$
Conclusions from the figure

• Spectral-flatness measure of the inverse filter output varies far less than input’s.

• During unvoiced portions, the theoretical model predicted an average level of -2.5 dB which compares well with the experimental results.

• During voiced portions, the theoretical model predicted a wide range of values, \( [(L!)^2/(2L)!], 1 \) which compares well with the experimental results.

• 5 (d): the number of pitch periods per analysis window is twice \( \rightarrow \) the spectral-flatness measure is decreased.

• 5(a) and 5(b): Spectral-flatness measure is decreased during voiced portions by the use of a Hamming window.

• 5(b) and 5(c): Increasing the sampling rate reduces spectral-flatness measure of voiced sounds.
SPECTRAL FLATNESS AND ILL CONDITIONING

1. It takes more accuracy to analyze voiced sounds than unvoiced.

2. The use of a Hamming or Hanning window increases the amount of computational accuracy needed.

3. Increasing the sampling rate increases the amount of computational accuracy needed.

4. Proper pre-emphasis or pre-whitening can decrease the amount of computational accuracy needed.
Measure of Ill Conditioning

Ill conditioning may be quantified by $\rho_M$:

$$\rho_M = \left| R \right|^{1/M}/\alpha_0 = \left[ \prod_{m=0}^{M-1} (\alpha_m/\alpha_0) \right]^{1/M}$$

This expression for $\rho_M$ indicates that $\rho_M$ represents the geometric mean of the decreasing sequence $(\alpha_m/\alpha_0)$. It will decrease from 1, for $M = 1$, and approach the limiting value –

$$\lim_{M \to \infty} \rho_M = \rho_\infty = \alpha_\infty/\alpha_0 = \Xi(X)$$

The spectral-flatness measure is thus both a lower bound and a limiting value of $\rho_M$ and is ITSELF a measure of ill-conditioning.

Value 1 – Perfectly Conditioned Problem
Value 0 – Singular Problem
PREEMPHASIS OF THE SPEECH DATA

• Differencing the input speech signal accents the higher formants, thus allowing more accurate formant tracking results.

• One approach to preemphasis is to utilize a low order inverse filter and maximize the spectral flatness of its output.

• Proposed is a simple first order preemphasis filter to do the purpose.
First Order Preemphasis

Pre-emphasis Filter Form: \( 1 - \mu z^{-1} \).

Where
• \( \mu = \frac{r_s(1)}{r_s(0)} \),

• \( r_s(n) \rightarrow \) autocorrelation sequence for the input sequence data \( \{s_n\} \).

• If \( \{f_n\} \) is the time sequence of the preemphasis filter output then

• \( \Xi (F) = \Xi (S) \frac{r_s(0)}{r_f(0)} \), and a direct evaluation gives \( r_f(0) \) as:
  
  \[
  \Xi (F)_{\text{max}} = \Xi (S) / (1 - \mu^2)
  \]

  \[
  r_f(0) = (1 + \mu^2) r_s(0) - 2\mu r_s(1)
  \]
Spectral Flatness with Preemphasis
(a) Rectangular Window,
(b) Hamming Window.
Conclusions

• A spectral-flatness measure has been developed: numerical value from 0 to 1.

• Perfectly flat or constant spectrum has a flatness of 0 dB.

• The lower the spectral flatness the more ill-conditioned the problem.

• Pre-emphasis of the speech signal by means of a one-term linear predictor was shown to greatly enhance the spectral flatness of the signal.