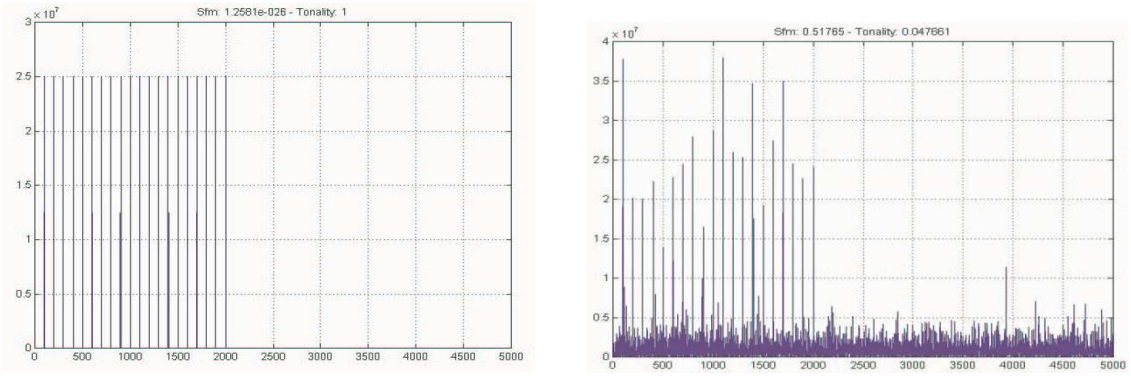
A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis

Authors: Augustine H. Gray and John D. Markel

A Report by Kaviraj Singh, Komaljit Meena, Vaibhav Vashisht

Definition: 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. Consider the two images below:

Figure 1(a) illustrates the spectrum of a tonal signal, i.e., a signal that contains sinusoids of various frequencies. Such a spectrum contains discrete peaks owing to which the *"flatness"* is 0. Figure 1(b) shows the spectrum of the same signal with noise added to it. The noise that is visible in the spectrum makes it more flat than in the previous case and the Spectral Flatness Measure is non-zero.



(a) Spectral Flatness = 0 (b) Spectral Flatness = 0.51

Figure 1: Comparison of the Flatness of two spectra.

I. To define a Spectral Flatness Measure for any given spectrum (here, a speech spectrum):

The normalized log spectrum of time sequence is given by

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

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

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

Given the normalized log spectrum of the speech signal, let us define two functions of $V(\Theta)$ and compare which one of the two is more suited to be a SFM (Spectral Flatness Measure).

A.
$$\eta(E) = \int_{-\pi}^{\pi} \tfrac{1}{2} V^2(\theta) \frac{d\theta}{2\pi}.$$
---------- Equation (1)

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

Plotting both A and B w.r.t. *V(Θ)* on a graph, we get the following shapes:

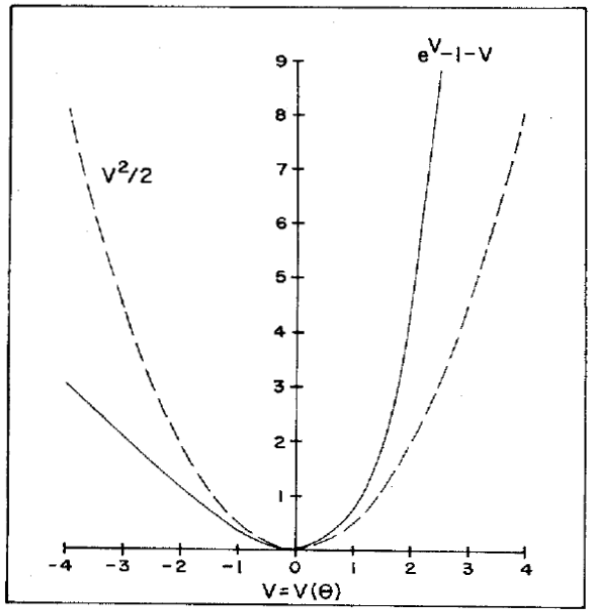


Figure 2: ***η*** *and* ***μ*** versus *V(Θ)*

Thus, the two measures defined have the following properties:

A: Weighs -ve and +ve excursions of the normalized log spectrum equally

B: Weighs the +ve excursions more heavily and the -ve excursions less heavily

Now, 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, ***μ* has the desired properties, and will be a better measure for our purposes**.

Developing the SFM: As *V(θ)* represents the normalized log spectrum of the signal, the average of e^V will be unity.

Thus, our initial representation in Equation (2) may be simplified as:

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

Here, - *μ (E)* with a factor of 2 represents the zeroth quefrency of the cepstrum and exp [- *μ (E)*] represents the ratio of the geometric to arithmetic means of the spectrum. Let this be called *Ƶ (E),* which is our Spectral Flatness Measure.

Thus, our Spectral Flatness Measure is:

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

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

II. Spectral Flatness Transformations

Consider an all-zero inverse filter $A_M(z)$

$$A_M(z) = 1 + \sum_{k=1}^{M} a_{Mk} z^{-k}$$

The output will be represented in the form of the input in the following manner:

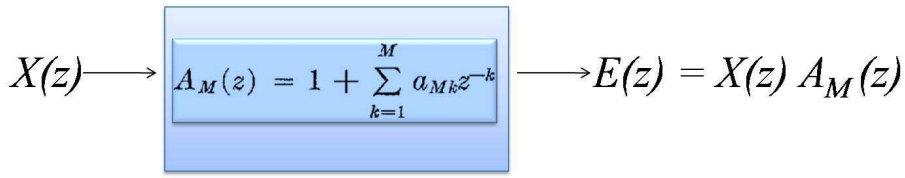


Figure 3: An All-Zero Model

Residue calculus can be applied to show that

$$\int_{-\pi}^{\pi} \log\left\{\left|A_M[\exp(j\theta)]\right|^2\right\}\frac{d\theta}{2\pi} = 0$$ ---------- Equation (3)

Using Equation (3), along with the expression obtained for $E(z)$, it can be shown that

$$\int_{-\pi}^{\pi} \log\left\{\left|E[\exp(j\theta)]\right|^2\right\}\frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} \log\left\{\left|X[\exp(j\theta)]\right|^2\right\}\frac{d\theta}{2\pi}$$

Using the previous result, and the definition of the measure of spectral flatness, we obtain the transformation:

$\mathcal{Z}(E) = \mathcal{Z}(X)\, r_x(0) / r_e(0)$ ------- **Equation (4)**

If the input to the filter AM(z) is fixed, the only portion of Equation (4) that can produce a change in the output spectral flatness is the term $r_e(0)$, the energy of that output. $\mathcal{Z}(E)$ will thus be a maximum when $r_e(0)$ is a minimum. Since minimizing $r_e(0)$ is one of the many criteria used to lead to the autocorrelation method of linear prediction [1], maximizing the spectral-flatness measure of the inverse filter output leads to precisely the same results.

Given that the aim is decomposition of log spectrum of $X(z)$ in terms of both the log spectra of $E(z)$ and $1/A_M(z)$, we can use the result obtained in Equation (4) to show that

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

III. Spectral-Flatness of Two Driving Function Models

A. Unvoiced Driving Function Model

One possible model in the case of unvoiced sounds for the driving function is uncorrelated Gaussian noise. The log spectrum of such a signal will have an expected value that is less than the logarithm of the expected value of the spectrum by an amount y, where *y* is Euler's constant 0.5772.Numerically evaluated flatness measure will have an expected value of roughly exp(- *y)* or -2.5 dB.

B. Voiced Driving Function Model

Driving function is the set of L + 1 equally spaced samples

$$y_k = 0 \text{ for } k \neq l, l+P, l+2P, \cdots, l+LP$$

where, y_l is the first sample in the time window, y_{Z+lP} is the last, and P represents a pitch period. The spectral flatness will lie between $(L!)^2/(2L)!$ and one. If there is only one such sample in the time window, the resulting spectral-flatness measure is 1, or **0** dB. If all samples have the same size, then Spectral-Flatness measure equals $1/(L+1)$

We draw the Spectogram of the utterance **"Will the rest follow soon."**

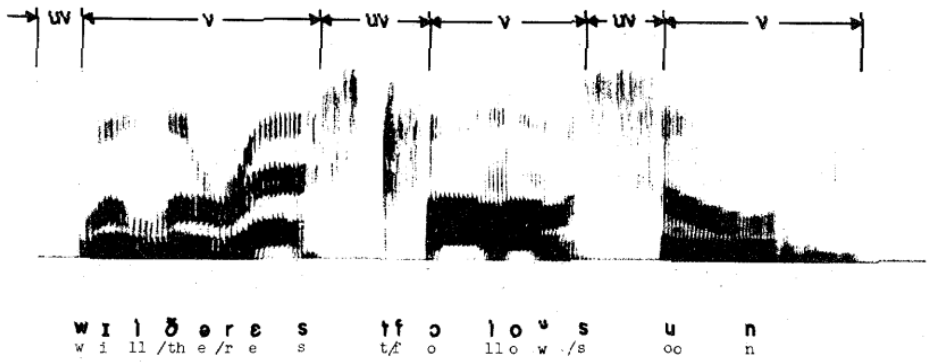


Figure 4: Spectogram of the utterance "Will the rest follow soon"

TABLE I

Example	Sampling Frequency F_s	Number of Samples N	Window Length	Type of Window
Fig. 5(a)	6.5 kHz	128	19.69 ms	rectangular
Fig. 5(b)	6.5 kHz	128	19.69 ms	Hamming
Fig. 5(c)	13.0 kHz	256	19.69 ms	Hamming
Fig. 5(d)	6.5 kHz	256	39.38 ms	Hamming
Fig. 6(a)	6.5 kHz	128	19.69 ms	rectangular
Fig. 6(b)	6.5 kHz	128	19.69 ms	Hamming

Table 1: Description of the analysis parameters in Figures 5 and Figure 6

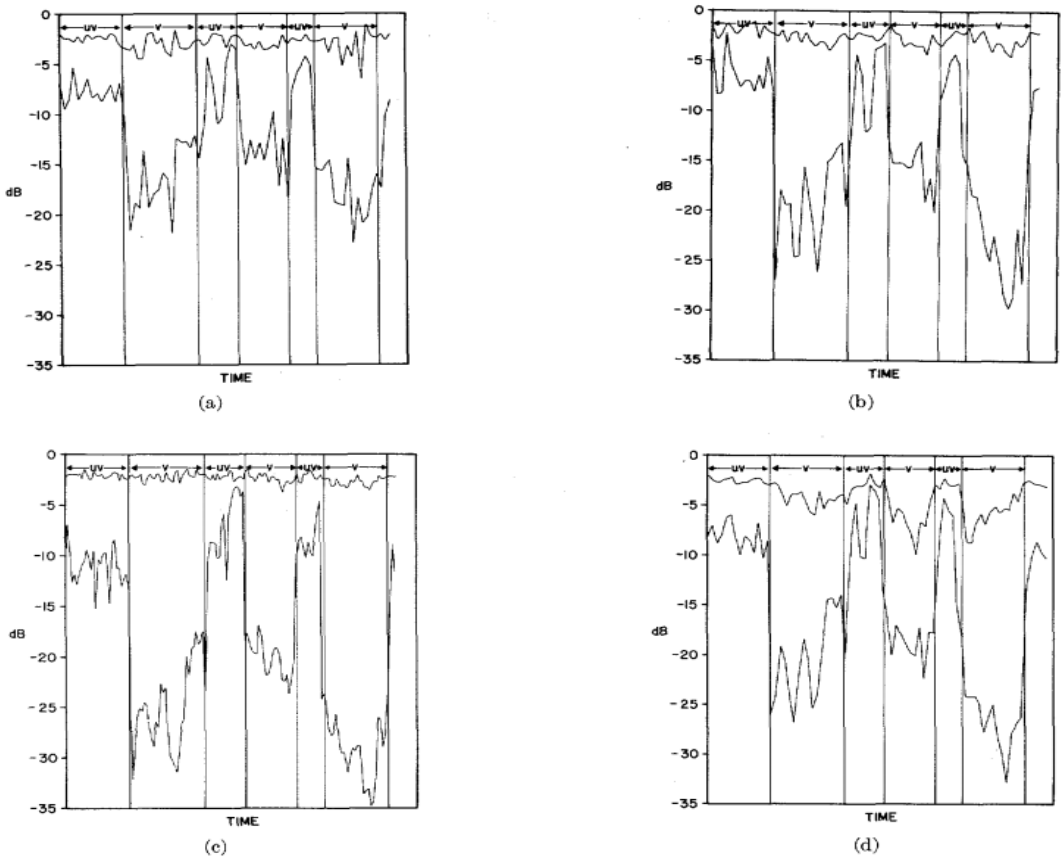


Figure 5: The lower curve represents the spectral-flatness measures at the input to the inverse filter, $\mathbf{10\log_{10} Ƶ(X)}$, and the upper curve represents the spectral-flatness measures at the outputs, $\mathbf{10\log_{10} Ƶ(E).}$ Each data window has N points thus, time windows have length $N/F_S = N\Delta t$

C. Conclusions from the figure

- The 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.
- In figure 5 (d) the number of pitch periods per analysis window is twice hence, the spectral-flatness measure is decreased.
- In figures 5(a) and 5(b) the spectral-flatness measure is decreased during voiced portions by the use of a Hamming window.
- In figures 5(b) and 5(c) increasing the sampling rate reduces spectral-flatness measure of voiced sounds.

IV. Spectral Flatness and Ill-Conditioning

To solve for the coefficients of the inverse filter, one must solve a set of *M* simultaneous algebraic equations. The matrix of coefficients of these equations is the *M* by autocorrelation matrix **R** whose elements rik for i = 1, 2, ..., M and k = 1, 2, … , M are given by the autocorrelation values of the inverse filter input sequence, rik = rx (I – k).

There are numerous "measures" of ill-conditioning of matrices. The most common of these are the N and M condition numbers of Turing and the P and H condition numbers of Todd.

In the given paper, an effort has been made to introduce a more elementary measure of ill-conditioning, which more closely corresponds with experimental results. This measure will be defined as a number which lies between zero and one, taking on the value of one for a perfectly conditioned problem, when R is the identity matrix, and zero for a singular problem, when R is a singular matrix.

One elementary approach is to utilize a normalized determinant of the matrix in question:

$$| \boldsymbol{R} | / r_x^M(0) = | \boldsymbol{R} | / \alpha_0^M = \prod_{m=0}^{M-1} (\alpha_m / \alpha_0)$$

A modification of the above will be our measure of ill-conditioning (Mth root of the above):

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

ρ_m represents the geometric mean of the decreasing sequence (α_m / α_0). It will decrease from 1, for M = 1, and approach the limiting value

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

Thus, the spectral-flatness measure is thus both a lower bound and a limiting value of ρ_m and as such can itself be considered a measure of ill-conditioning. From this, a number of conclusions may be drawn:

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.

V. Pre-emphasis of the Speech Data

The probability of numerically caused instabilities in the filter $1/A_M(z)$ is greatly reduced by pre-emphasis of the speech data. Pre-emphasis is mostly useful for ill-conditioned problem and is of importance to inverse filter analysis techniques.

One approach to pre-emphasis is to utilize a low order inverse filter and maximize the spectral flatness of its output. Proposed is a simple first order pre-emphasis filter to do the purpose.

A. First Order Pre-emphasis

The pre-emphasis filter is of the form $\mathbf{1 - \mu z^{-1}}$

Where $\boldsymbol{\mu = r_s(1) / r_s(0)}$, $\boldsymbol{r_s(n)}$ is autocorrelation sequence for the input sequence data $\mathbf{\{s_n\}}$.

If $\mathbf{\{f_n\}}$ is the time sequence of the pre-emphasis filter output then $\boldsymbol{\Xi(F) = \Xi(S)\, r_s(0)/ r_f(0)}$, and a direct evaluation gives $\boldsymbol{r_f(0)}$ as: $\boldsymbol{r_f(0) = (1 + \mu^2)\, r_s(0) - 2.\mu.r_s(l)}$ hence,

$$\boldsymbol{\Xi(F)_{max} = \Xi(S) / (1 - \mu^2)}$$

B. Experimental results

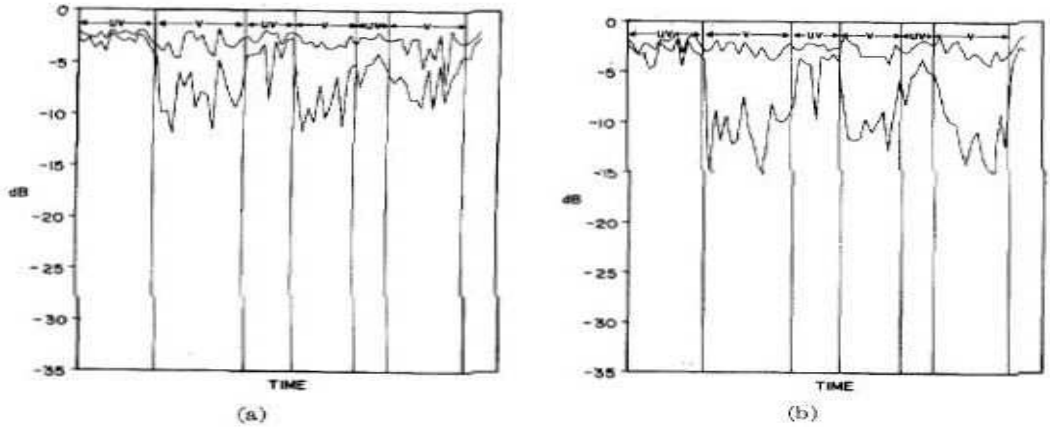


Figure 6: Spectral Flatness with Pre-emphasis: (a) Rectangular Window, (b) Hamming Window.

C. Conclusions

a. Comparing with Fig. 5(a), one can see that the input spectral flatness is greater, while the output spectral flatness is slightly less. This is to be expected, for the preemphasis filter increases the spectral flatness at the input of the inverse filter.
b. The combination prremphasis filter and inverse filter is eighth order in Fig. 6(a.), and thus cannot have an output spectral flatness which is as good as that of the optimum eighth-order inverse filter shown in Fig.5 (a)
c. The distance between the lower and upper curve, $\mathbf{\log_{10}(\alpha_M/\alpha_0)}$ indicates that the ill conditioning of the solution process is considerably reduced
d. A comparison of Fig. 6 (a) and (b) shows that the spectral flatness at the output of the inverse filters is essentially unchanged in its overall behaviour.

VI. Conclusions from the paper

- 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.

References

[1] J. D. Markel, "Digital inverse filtering - a new tool for formant trajectory estimation," *IEEE Trans Audio Electrocoust.*, vol., AU-20, pp. 129-137, June 1972.