Title
I/F NOISE"" IN MUSIC and SPEECH

Author
Voss, Richard P.

Publication Date
1975-06-01

Submitted to Science

LBL-3959
Preprint c.1

"1/F NOISE" IN MUSIC AND SPEECH

Richard F. Voss and John Clarke

June 1975

Prepared for the U. S. Energy Research and
Development Administration under Contract W-7405-ENG-48

Submitted to Science Magazine LBL-3959

UNIVERSITY OF CALIFORNIA

Lawrence Berkeley Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

"1/F NOISE" IN MUSIC AND SPEECH

Richard F. Voss and John Clarke

June 1975

"1/f Noise" in Music and

Speech

Loudness fluctuations in music and speech, and

pitch fluctuations in music,exhibit a 1/f spectrum.

Richard F. Voss

John Clarke

Department of Physics, University of California, and

Inorganic Materials Research Division, Lawrence Berkeley

Laboratory, Berkeley, California 94720

The power spectrum of many physical quantities is "1/f-like", varying as $1/f^{\gamma}$, where $0.5\lesssim\gamma\lesssim1.5$, over many decades. Thus vacuum tubes[1], carbon resistors[2], semiconducting devices[3], continuous[4,5] or discontinuous[6] metal films , ionic solutions[7], films at the superconducting transition[8], Josephson junctions[9], nerve membranes[10], sunspot activity[11], and the flood levels of the river Nile[11] all exhibit what has come to be known as "1/f noise".[12] In this paper, we report measurements of the power spectra of various fluctuating quantities associated with music and speech. We have found that loudness fluctuations in music and speech, and pitch (melody) fluctuations in music, exhibit 1/f spectra. We discuss the implications of these measurements.

Although most frequently used in the analysis of random signals or "noise", the power spectrum is an extremely useful characterization of the average behavior of any quantity varying in time. The power spectrum, $S_V(f)$, of a fluctuating quantity, $V(t)$, is a measure of the average "power", $\langle V^2\rangle$, in a unit bandwidth about the frequency f. $S_V(f)$ may be measured by passing $V(t)$ through a tuned filter of frequency f and bandwidth Δf. $S_V(f)$ is then the average of the squared output of the filter divided by Δf. A second characterization of the average behavior of $V(t)$ is the autocorrelation function, $\langle V(t)V(t+\tau)\rangle$. $\langle V(t)V(t+\tau)\rangle$ is a measure of how the fluctuating quantities at times t and $t+\tau$ are related. $S_V(f)$ and $\langle V(t)V(t+\tau)\rangle$ are not independent, but are related by the Wiener - Khintchine relations:[13]

$$\langle V(t)V(t+\tau)\rangle = \int_0^\infty S_V(f)\cos(2\pi f\tau)\,df, \quad (1)$$

and

$$S_V(f) = 4\int_0^\infty \langle V(t)V(t+\tau)\rangle \cos(2\pi f\tau)\,d\tau. \quad (2)$$

Many fluctuating quantities, $V(t)$, may be characterized by a single correlation time, τ_c. In such a case, $V(t)$ is correlated with $V(t+\tau)$ for $|\tau|<\tau_c$, and is independent of $V(t+\tau)$ for $|\tau|>\tau_c$. From Eq.(2), it is then possible to show that $S_V(f)$ is "white" (independent of frequency) in the frequency range corresponding to time scales over which $V(t)$ is independent ($f \ll 1/2\pi\tau_c$); and is a rapidly decreasing function of frequency, usually $1/f^2$, in the frequency range over which $V(t)$ is correlated ($f \gg 1/2\pi\tau_c$). A quantity with a 1/f power spectrum cannot, therefore, be characterized by a single correlation time. In fact, the 1/f power spectrum implies some correlation in $V(t)$ over all time scales corresponding to the frequency range for which $S_V(f)$ is 1/f-like.[14] In general, a negative slope for $S_V(f)$ implies some degree of correlation in $V(t)$ over time scales of roughly $1/2\pi f$. A steep slope implies a higher degree of correlation than a shallow slope.

In our measurements on music and speech, the fluctuating quantity of interest was converted to a voltage whose power spectrum was measured by an interfaced PDP-11 computer using a Fast Fourier Transform algorithm. The most familiar fluctuating quantity associated with music is the audio signal, V(t), such as the voltage used to drive a speaker system. Fig. 1(a) shows a linear-linear plot of the power spectrum, $S_V(f)$, of the audio signal from J. S. Bach's 1st Brandenburg Concerto (Angel SB-3787) averaged over the entire concerto. The spectrum consists of a series of sharp peaks in the frequency range 100Hz to 2kHz corresponding to the individual notes in the concerto and, of course, is not 1/f-like. Although this spectrum contains much useful information, our primary interest is in more slowly varying quantities.

One such quantity is the loudness of the music. The audio signal, V(t), was amplified and passed through a bandpass filter in the range 100Hz to 10kHz. The filter output was squared and the audio frequencies filtered off to give a slowly varying signal, $V^2(t)$, proportional to the instanteous loudness of the music. ($V^2(t)$ is thus similar to the reading given by recording level meters.) The power spectrum of the loudness fluctuations of the 1st Brandenburg concerto, $S_{V^2}(f)$, averaged over the entire concerto is shown in Fig. 1(b). On this linear-linear plot, the loudness fluctuations appear as a peak close to zero frequency.

Figure 2 is a log-log plot of the same spectra as in Fig. 1. In Fig. 2(a), the power spectrum of the audio signal, $S_V(f)$, is distributed over the audio range. In Fig. 2(b), however, the loudness fluctuation spectrum, $S_{V^2}(f)$, shows the 1/f behavior below 1 Hz. The peaks between

1 Hz and 10 Hz are due to the rhythmic structure of the music.

Fig. 3(a) shows the power spectrum of loudness fluctuations for a recording of Scott Joplin piano rags (Nonesuch H-71248) averaged over the entire recording. Although this music has a more pronounced rhythm than the Brandenburg Concerto, and, consequently, has more structure in the spectrum between 1 Hz and 10 Hz, the spectrum below 1 Hz is still 1/f-like.

In order to measure $S_{V^2}(f)$ down to even lower frequencies an audio signal of greater duration than a single record is needed, for example, that from a radio station. The audio signal from an AM radio was filtered and squared. $S_{V^2}(f)$ was averaged over approximately 12 hours, and thus included many musical selections as well as announcements and commercials. Figures 3(b), 3(c), and 3(d) show the loudness fluctuation spectra for three radio stations characterized by different motifs. Figure 3(b) shows $S_{V^2}(f)$ for a classical station. The spectrum exhibits a smooth 1/f dependence. Fig. 3(c) shows $S_{V^2}(f)$ for a rock station. The spectrum is 1/f like above 2×10^{-3}Hz, and flattens for lower frequencies, indicating that the correlation of the loudness fluctuations does not extend over time scales longer than a single selection, roughly 100 sec. Fig. 3(d) shows $S_{V^2}(f)$ for a news and talk station, and is representative of $S_{V^2}(f)$ for speech. Once again the spectrum is 1/f-like. In Fig. 3(b) and Fig. 3(d), $S_{V^2}(f)$ remains

1/f-like down to the lowest frequency measured, $5x10^{-4}$Hz, implying correlations over time scales of at least 5 min. In the case of classical music this time is less than the average length of each composition.

Another slowly varying quantity in speech and music is the instantaneous pitch. A convenient means of measuring the pitch is by the rate, Z, of zero crossings of the audio signal, V(t). Thus an audio signal of low pitch will have few zero crossings per second and a small Z, while a high pitched signal will have a high Z. For the case of music, Z(t) roughly follows the melody. Figure 4 shows the power spectra of the rate of zero crossings, $S_Z(f)$, for three radio stations averaged over approximately 12 hours. Figure 4(a) shows $S_Z(f)$ for a classical station. The power spectrum is closely 1/f. Figure 4(b) shows $S_Z(f)$ for a jazz and blues station. Here the spectrum is 1/f-like down to frequencies corresponding to the average selection length, and is flat at lower frequencies. Although not shown here, $S_Z(f)$ for a rock station is similar to Fig. 4(b): 1/f for frequencies greater than the frequency corresponding to the average selection length. Figure 4(c), however, which shows $S_Z(f)$ for a news and talk station, exhibits a quite different spectrum. The spectrum is that of a quantity characterized by two correlation times: the average length of an individual speech sound, roughly 0.1 sec., and the average length of time for which a given announcer talks, about 100 sec. For most musical selections the pitch or melody has correlations that extend over a large range of time scales, and has a 1/f power spectrum. For normal English speech, on the other hand, the pitches of the individual speech sounds are unrelated. As a result, the power spectrum is "white" for frequencies less than about 3 Hz, and falls as $1/f^2$ for $f \gtrsim 3$Hz. In fact, in both Fig. 4(a) and Fig. 4(b), one observes shoulders at about 3 Hz corresponding to the announcements averaged in with the music.

It is remarkable that quantities associated with music and speech exhibit the same 1/f power spectrum and have the same sort of correlations extending over all time scales as so many physical variables. The most successful theories of 1/f noise in physical systems have focused on the mechanism of the noise generation, rather than on the exact shape of the spectrum. For example, 1/f noise in continuous metal films[5], superconducting films biased at T_c[8], and Josephson junctions[9] has recently been shown to be due to temperature fluctuations with a 1/f spectrum. A different physical mechanism dominates in semiconductors where the most accepted theory is that surface traps with various trapping times modulate the conductivity[15]. In neither case is the 1/f spectrum fully explained. Although we make no attempt to associate a physical mechanism with the 1/f spectra in speech and music, we speculate that various measures of "intelligent" behavior should show a 1/f-like power spectrum. Such behavior is expected to show some correlation over all time scales yet not depend too strongly on its past. The power spectrum, therefore, should lie somewhere between the "white"spectrum of an uncorrelated quantity and the $1/f^2$ spectrum of a quantity which depends strongly on its past (such as the position of a particle undergoing Brownian motion). Human communication is one example where correlations extend over various time scales. In music much of the communication is directly by the melody which exhibits a 1/f spectrum. In speech, on the other hand, the communication is not directly related to the pitch of the individual sounds but rather to the meanings attached to them by the listener. The ideas communicated may have long time correlations even though the pitches of successive sounds are unrelated.

The observation of 1/f-like power spectra for various musical quantities also has implications for stochastic music composition. Most stochastic compositions are based on a random number generator (white noise source), which produces unrelated notes, or on a low level Markov process, in which there is correlation over only a few successive notes. Neither of these techniques, however, approximate the 1/f spectrum and the long time correlations reported here in music. We have used independent 1/f noise sources in a simple algorithm to determine the duration (quantized as half, quarter, or eighth notes) and pitch (quantized in various standard scales) of successive notes of a melody. The music obtained by this method was judged by most listeners to be much more pleasing than that obtained using either a white noise source (which produced music that was "too random") or a $1/f^2$ noise source (which produced music that was "too correlated"). Indeed, the sophistication of this "1/f music" (which was "just right") extends far beyond what one might expect from such a simple algorithm, suggesting that a "1/f noise" (perhaps that in nerve membranes?) may play an essential role in the creative process.

References and Notes

1. J. B. Johnson, Phys. Rev. 26, 71 (1925).

2. C. J. Christenson and G. L. Pearson, Bell System Tech. J. 15, 197 (1936).

3. For a review, see Aldert van der Ziel, Noise: Sources, Characterization, Measurement (Prentice-Hall Inc., Englewood Cliffs, NJ, 1970) p. 106.

4. F. N. Hooge and A. M. H. Hoppenbrouwers, Physica 45, 386 (1969).

5. John Clarke and Richard F. Voss, Phys. Rev. Letters 33, 24 (1974).

6. J. L. Williams and R. K. Burdett, J. Phys. C 2, 298 (1969).

7. F. N. Hooge, Phys. Lett. 33, 169 (1970).

8. John Clarke and Thomas Y. Hsiang, Phys. Rev. Letters 34, 1217 (1975).

9. John Clarke and Gilbert Hawkins, I.E.E.E. Transactions on Magnetics MAG -11, 841 (1975).

10. A. A. Verveen and H. E. Derkson, Proc. I.E.E.E. 56, 906 (1968).

11. B. B. Mandelbrot and J. R. Wallis, Water Resources Research 5, 321 (1969).

12. $S_V(f)$ has been measured down to 5×10^{-7}Hz in semiconductors [M. A. Caloyannides, J. Appl. Phys. 45, 307 (1974)]; in the case of the river Nile, $S(f)$ extends down to 3×10^{-11}Hz.

13. F. Reif, Fundamentals of Statistical and Thermal Physics, (McGraw-Hill 1965) p. 585.

14. The fact that 1/f noise contains a distribution of correlation times is discussed by Aldert van der Ziel, Physica 16, 359 (1950).

15. A. L. McWhorter, Semiconductor Surface Physics, R. H. Kingston, Ed. (Univ. of PA. Press Philadelphia, 1957) p. 207.

16. This research was supported by the U.S.E.R.D.A.

Fig. 1. Bach's 1st Brandenburg Concerto (linear scales):

a) Power spectrum of audio signal, $S_V(f)$ versus f;

b) Power spectrum of loudness fluctuations, $S_{V^2}(f)$ versus f.

Fig. 2. Bach's 1st Brandenburg Concerto (log scales):

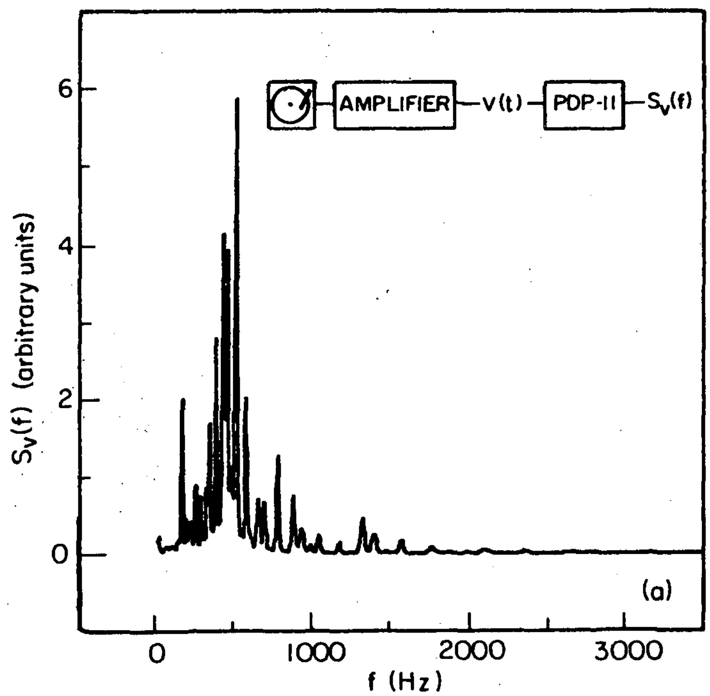
a) $S_V(f)$ versus f; (b) $S_{V^2}(f)$ versus f.

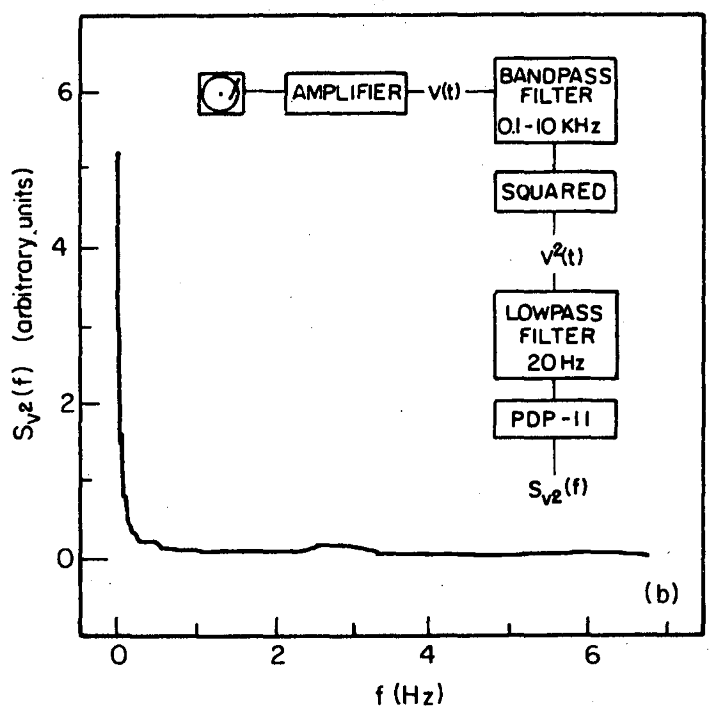
Fig. 3. Loudness fluctuation spectra, $S_{V^2}(f)$ versus f for:

a) Scott Joplin piano rags; (b) classical radio station;

c) rock station; (d) news and talk station.

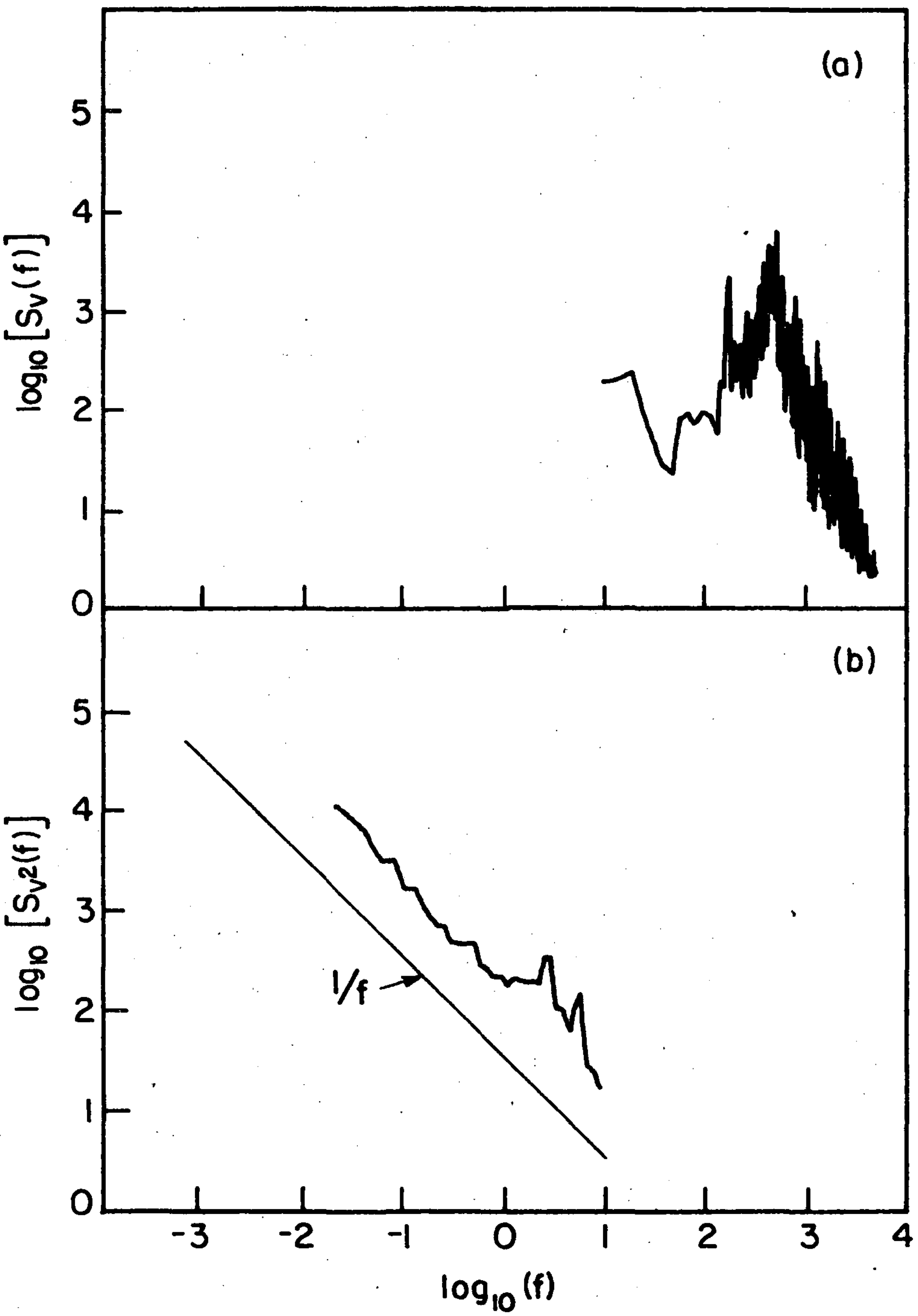
Fig. 4. Power spectra of pitch fluctuations, $S_Z(f)$ versus f, for three radio stations: a) Classical; (b) jazz and blues; (c) news and talk.





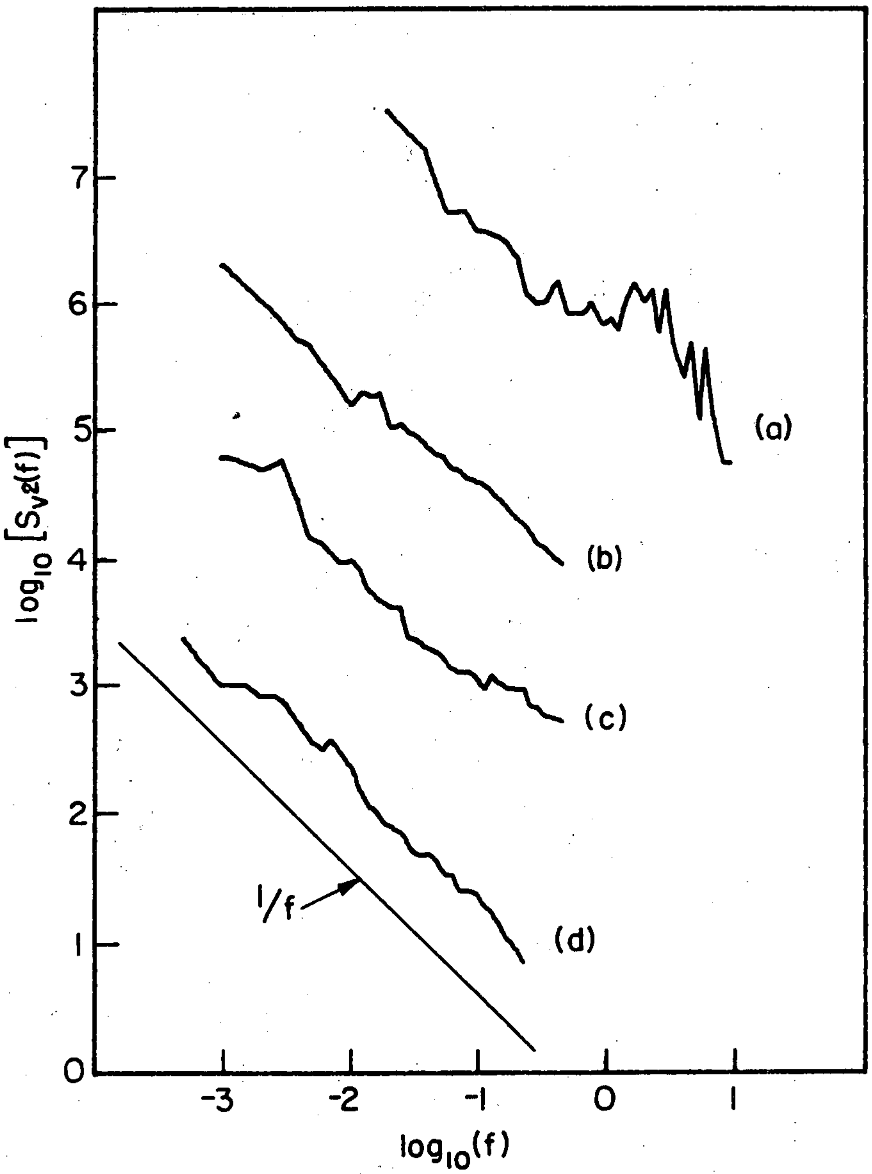
XBL 752-5846

Fig. 1



XBL752-5845

Fig. 2



XBL 752-5844

Fig. 3

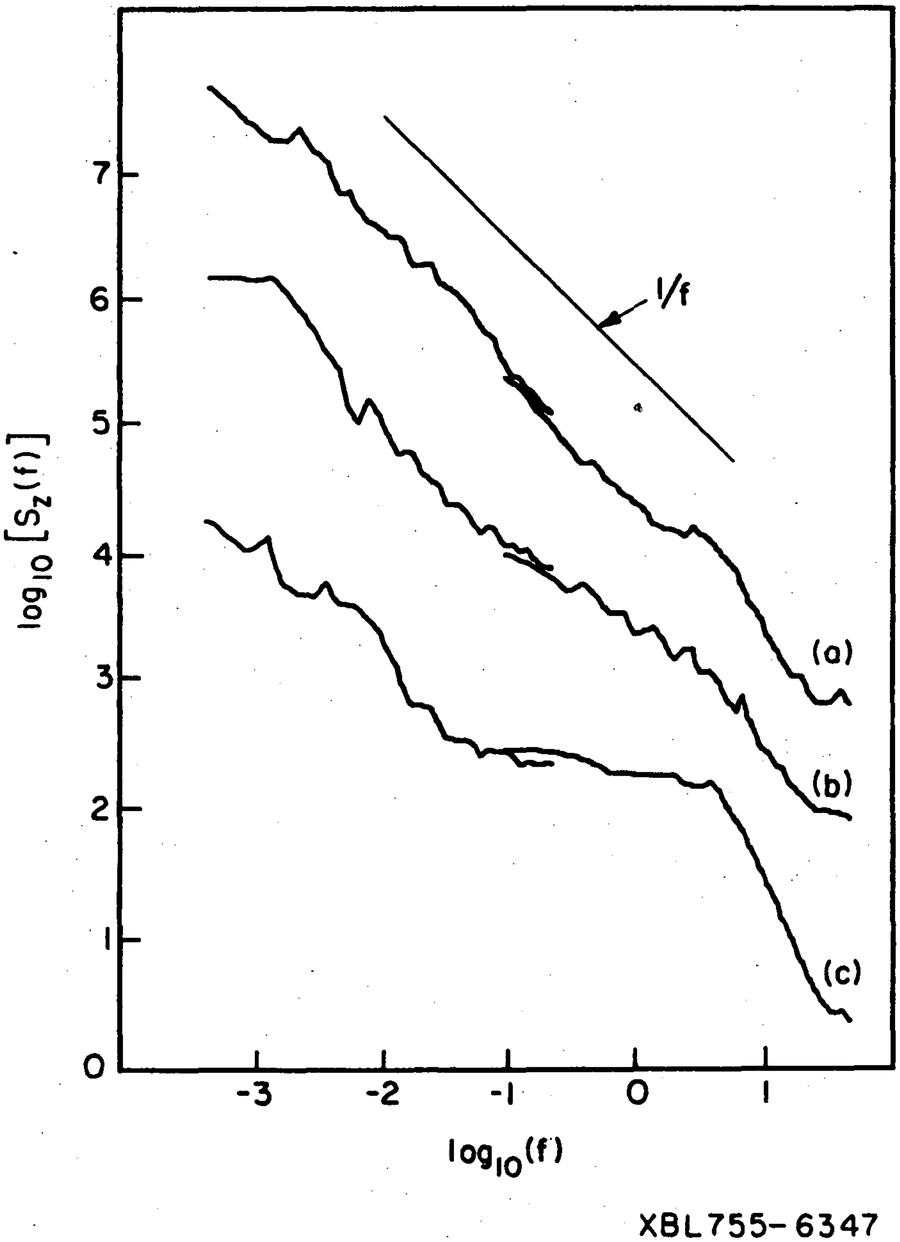


Fig. 4

0 0 0 0 4 3 0 6 0 1 2

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720