# Signal recovery from noise in electronic instrumentation

T H Wilmshurst

# Signal recovery from noise in electronic instrumentation

T H Wilmshurst

*Reader in Electronics, University of Southampton*



Adam Hilger Ltd
Bristol and Boston

*British Library Cataloguing in Publication Data*

Wilmshurst, T. H.
Signal recovery from noise in electronic instrumentation.
1. Electronic instruments 2. Electronic noise
I. Title
621.38'0436 TK7878.4

ISBN 0-85274-783-7

Consultant Editor: **Mr A E Bailey**

Published by Adam Hilger Ltd
Techno House, Redcliffe Way, Bristol BS1 6NX
PO Box 230, Accord, MA 02018, USA

Typeset by Mid-County Press, London SW15
Printed in Great Britain by Page Brothers (Norwich) Ltd

# Preface

The subject of this book is the recovery of signals from noise in electronic instrumentation. The term 'electronic instrumentation' is sometimes reserved for instruments such as the oscilloscope and testmeter which are used to measure specifically electrical variables. It should therefore be made clear that the scope of the book goes much beyond this and covers instrumentation for the measurement of any variable, whether electrical or not. The term 'electronic' simply implies that electronic techniques are used in the process of measurement.

The term 'noise' also requires clarification. Until recently, it was used mainly to refer to random fluctuations such as white and $1/f$ noise. Now, however, the term is used to refer to almost any kind of unwanted signal in an electronic system. It is in the broader sense that the term is used in the title, since the recovery of the required signal from many kinds of unwanted signal is covered. These comprise offset, drift random noise, comprising white and $1/f$ noise, and interference. However, to avoid confusion, the term 'noise' will be reserved for random noise. The other types of unwanted signal will be referred to by their specific names.

There are, broadly, two ways of reducing the errors due to unwanted signals in an electronic system. The first is to prevent the unwanted signal from being introduced. This is a matter of low-noise amplifier design, screening, decoupling, etc, and is reasonably well covered by existing texts. The second approach, which is taken up when the first has been exploited as far as possible, is to devise 'signal recovery' techniques which distinguish the required signal as well as possible from whatever unwanted signals may remain. This second approach is less well covered at present and therefore represents the subject matter of the book.

The text is intended for anyone who is to be involved in the development of electronic instrumentation, at graduate level or beyond. In order to indicate better the intended readership it will be helpful to consider briefly the way in which electronic instruments are, or should be, developed. A number of different disciplines is nearly always involved. Consider, for example, such a system as the laser anemometer. This is a device which uses both electronic and optical techniques for the measurement of fluid velocity. Thus the development of this instrument would require experts in optics, electronics and fluid dynamics. When such a team is formed, it is important that each member should attempt to master all aspects of the system, not just those of his own discipline. This is as true for the electronic aspects of the system as for any other. Thus, the book has to be acceptable to two rather different classes of reader. The first is the undergraduate electronics engineer who intends to specialise in instrumentation. He will meet the subject probably in the third year of his undergraduate studies and will need to obtain a good grasp of the entire contents. The other class of reader will be the non-electronics specialist who will probably come to the subject after graduation and at the time of joining the team. He will usually need to study in depth a limited selection of topics from the book. In order to help this class of reader, attempts are made to cover the material in descriptive as well as analytical manner. Also, rather more extensive subsection titling than usual has been adopted, in order to facilitate 'random access'.

The book stems from two courses that are given in the Department of Electronics of the University of Southampton. The first has run for over a decade and is intended for graduates in disciplines other than electronics who are engaged in the development of electronic instrumentation, or sometimes just in its use. These may be students working for higher degrees, postdoctoral research fellows, technicians and, sometimes, academic staff.

The second course is more recent, and is intended primarily for third-year undergraduates in electronic engineering. It covers that material outside the 'core' electronics subjects that is needed for a graduate who is to specialise in electronic instrument development. The course is also followed by students taking 'with-electronics' courses such as 'physics-with-electronics', 'chemistry-with-electronics', etc.

**T H Wilmshurst**

# Contents

# 1 Low-pass filtering and visual averaging

## 1.1 OVERVIEW

Before discussing low-pass filtering and visual averaging, the two signal recovery methods to be discussed as the main topic of the present chapter, we devote the first section to an overview of the entire book. Here the classes of unwanted signal, and also all of the signal recovery methods to be described, will be briefly introduced.

**Resistor bridge strain gauge**

The first step will be to describe the operation of a simple example of an electronic instrument: the resistor bridge strain gauge of figure 1.1. Here the variable to be measured is the strain (extension) of the mechanical component shown, such strain usually resulting from a mechanical force (stress) applied to the component.

The principal component of the strain gauge is the resistor $R_g$. This is a mesh of fine wires which is attached to the mechanical component by an adhesive. As the mechanical component is strained, $R_g$ increases; this causes the bridge to become unbalanced, producing a voltage $v_t$ which is approximately proportional to strain. $v_t$ is then amplified, low-pass filtered and displayed on a chart recorder.

The variety of instruments of this kind is vast, the single common feature being that some variable, which is usually non-electrical, is converted to an electrical signal which is processed and displayed. The device which converts the variable of interest to the electrical signal is called the input 'transducer' and for the present system the transducer is the resistor bridge. This converts the variable of interest, the strain, into the electrical signal $v_t$. Although the range of input transducer types in use may constitute an important topic for

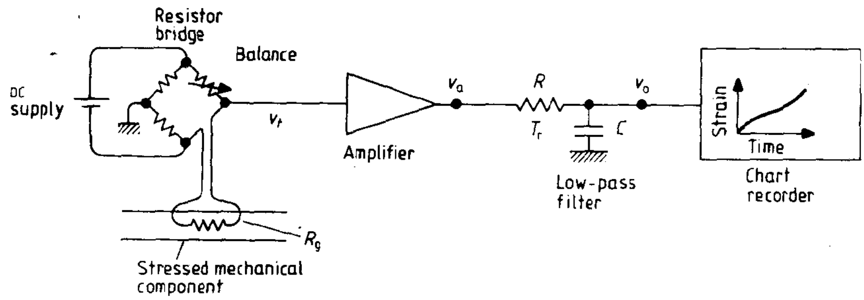


**Figure 1.1** Resistor bridge strain gauge.

the instrument engineer, this is not the subject of the present book, and is adequately covered elsewhere. Instead, our objective is to establish the principles whereby the electrical signal developed by such a transducer is recovered from the various kinds of unwanted signal.

This objective will be best served if, wherever possible in the course of introducing the signal recovery methods, the same simple example of the strain gauge is retained. This is in contrast to the alternative approach of constantly altering the example in order at the same time to try to give some idea of the differing transducer types. With the principles of signal recovery thus firmly established, using one type of transducer, it will be a simple matter to apply the principles to any other type of transducer that might be encountered.

**Low-pass filtering of white noise**

The purpose of the low-pass filter in the strain gauge of figure 1.1 is to reduce the amplitude of the unwanted signal components, of which the first to be considered is white noise. Figure 1.2 illustrates the way in which the low-pass filter reduces such noise. Here a fixed stress is applied to the mechanical component at time $t=0$ and the requirement is to measure the difference in recorded strain before and after application of the stress. Figure 1.2(*a*) shows what might be observed at the output of the amplifier in this situation. Here there is a moderately high level of white noise superimposed upon the signal component.

In due course it is shown that the effect of the filter is to produce the output voltage $v_o$, which is the 'running average' of the input voltage $v_a$. The function is illustrated in figure 1.2(*b*). Here, at time $t$, $v_o$ is the average of $v_a$ over the period shown from $t-T_r$ to $t$. Thus $T_r$ is termed the 'averaging time' for the filter.

In order to construct the output function $v_o$, the averaging box in figure 1.2(*b*) must be made to 'run' along the input function $v_a$ in (*a*), thereby describing the output function $v_o$ in (*c*). Clearly the effect is to reduce the noise. It is shown subsequently that the amplitude $\tilde{v}_n$ of the noise component of $v_o$ is proportional to $T_r^{-1/2}$, i.e. $\tilde{v}_n \propto T_r^{-1/2}$.

The other feature that is clear from figure 1.2 is that it takes the finite time $T_r$

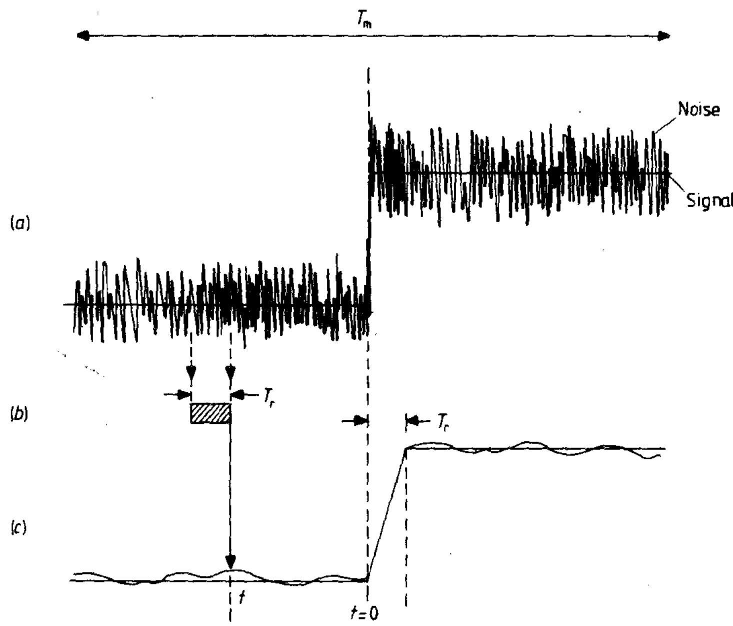


**Figure 1.2** Diagram showing how the low-pass filter of figure 1.1 reduces the white noise amplitude by forming a running average of the filter input. (*a*) Filter input ($v_a$ in figure 1.1), (*b*) running average process, (*c*) filter output $v_o$.

for $v_o$ to respond to the step in $v_{in}$. Thus $T_r$ is also known as the filter 'response time'.

**Visual averaging**

The second method of signal recovery to be considered is visual averaging. Looking at the filtered noisy signal step in figure 1.2, the basic requirement is to determine the difference between the signal before and after the step. Here the eye forms two averages, one before and one after the step, and subtracts one from the other. Now, because there are several noise fluctuations over each averaging period, the eye is able to determine the signal voltage to an accuracy greater than that given by the noise amplitude, even at the filter output.

The reason for this improvement is that basically it is the process of averaging the noise which reduces its effect and filter and eye are just two different methods of averaging. The reason that the noise amplitude $\tilde{v}_n$ at the filter output is proportional to $T_r^{-1/2}$ is that when white noise is averaged by any means over a period $T_{av}$ the value of the typical noise error in the average is proportional to $T_{av}^{-1/2}$. Then, since for the filter $T_{av} = T_r$, the noise error is proportional to $T_r^{-1/2}$. When, however, the eye extends the averaging period to the period $T_m/2$ of the half step the noise error is reduced from the value of say $kT_r^{-1/2}$ given by the amplitude to $k(T_m/2)^{-1/2}$.

It is the recovery of signals from white noise by low-pass filtering and by visual averaging that constitutes the main topic of the present chapter. However, before proceeding to this detailed study we complete the overview by briefly introducing the other types of unwanted signal and signal recovery methods.

**Drift and offset**

The next two types of unwanted signal to be covered are offset and drift. Of these, offset is defined as that component of the total of all unwanted signal components which does not vary with time. For the resistor bridge strain gauge of figure 1.1 the two main sources of offset are the transducer (the resistor bridge) and the signal amplifier. The transducer offset arises from initial imbalance in the bridge and is easily removed by adjusting the balance resistor shown. Similarly the DC coupled signal amplifier will also be provided with a suitable balancing potentiometer. Thus, for the present example, offset is not a major problem. However, for some systems the offset cannot be so simply removed. Then an acceptable alternative is the baseline correction method to be discussed in §1.4.

Unfortunately the offset, even if initially adjusted to zero, tends to 'drift' as time proceeds. This is usually because of slow changes in the temperature of the circuit.

Now, drift error tends to increase as the time $T_m$ spent making a measurement increases; this is directly opposed to the trend for white noise, for which the noise error varies as $T_m^{-1/2}$. Thus, drift tends to frustrate attempts to obtain a low white noise error by using a long measurement period, and we shall find that much of the strategy in signal recovery lies in devising methods of overcoming this problem.

**Multiple time averaging**

Two main methods are used to reduce the drift error and so allow the low white noise error associated with a long period of measurement to be realised. These are multiple time averaging (MTA) and the phase-sensitive detector (PSD) method. Of these MTA is a matter of carrying out the experiment rapidly rather than slowly, thus reducing the drift error, and then extending the averaging time to obtain the low white noise error by repeating the experiment many times and averaging the results. The use of MTA to avoid effects of drift in this way constitutes the main topic of chapter 2.

**Phase-sensitive detection**

Sometimes the experimental constraints do not allow the time scale of the experiment to be reduced. Here a suitable alternative is to modify the experiment in such a way that the transducer produces an AC rather than a DC signal. This allows an AC coupled amplifier to be used which does not respond to offset or drift.

Here some kind of rectifier is usually required in order to convert the AC amplifier output into a DC signal suitable for display. This, then, is the function of the phase-sensitive detector. It is the use of the PSD method to overcome drift that constitutes the main topic of chapter 3.

**$1/f$ noise**

The major remaining class of unwanted signal to be considered is $1/f$ noise. For both white and $1/f$ noise the names refers to the spectral distribution of the noise. For white noise the spectral density $G$ is independent of frequency $f$, while for $1/f$ noise $G \propto f^{-1}$. These terms belong to the frequency domain view and will be elaborated later. For the present, the waveforms of figure 1.3(*a*) and (*b*) will suffice to identify the difference between the two. These are specimens of white and $1/f$ noise as might be observed at the output of the signal amplifier in the strain gauge of figure 1.1. Clearly the $1/f$ noise has a greater tendency to 'walk away' from the mean. This means that the advantages of using a larger measurement period $T_m$ are less than for white noise. In fact, it will be shown that the $1/f$ noise error is independent of $T_m$. This contrasts with white noise, for which the error is proportional to $T_m^{-1/2}$, and drift, for which the error increases with $T_m$.

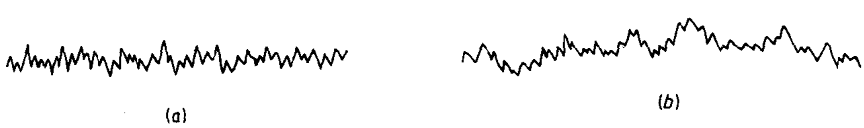


**Figure 1.3** Types of noise seen at the output of the filter in figure 1.1: (*a*) white noise, (*b*) $1/f$ noise.

It is thus clear that $1/f$ noise too frustrates any attempt to obtain a low white noise error by increasing $T_m$. For low values of $T_m$ the white noise may dominate, and then increasing $T_m$ will cause a reduction. However, a point will be reached at which the white noise error falls below the $1/f$ noise error, which does not vary with $T_m$. Then any further increase in $T_m$ will have no effect.

Fortunately, both MTA and PSD methods are effective in removing $1/f$ noise error but, in order to show how this is done, it is necessary to use the spectral view of both signal and noise. This is in contrast with the rather more direct 'time-domain' view used in the chapters listed so far.

**Frequency-domain view**

The frequency-domain view is a commonly adopted one and the present text is perhaps a little unusual in not adopting it from the start. However, I have preferred the directness of the time-domain approach for the initial treatment. Thus, in chapter 4 the frequency-domain or 'spectral' view will be introduced and the ground previously covered from the time-domain viewpoint re-covered from the frequency-domain viewpoint. Chapters 5 to 7 then use the frequency-domain view to show how the MTA and PSD methods remove the $1/f$ noise.

**Digitisation**

These seven chapters comprise the first part of the book, which deals with the recovery of a continuously varying analogue signal from the various kinds of unwanted signal. The remainder deals with one or two closely related topics. Frequently today an analogue signal must be digitised for computer storage and perhaps for further processing. In chapter 8 the problems of digitisation are covered. It transpires that the main factor of importance is that no data should be wasted.

**Pulsed signals**

Chapter 9 deals with pulsed signals or more general types of signal transient. Of particular interest is the case where the shape of a transient is known but the amplitude is not, being the factor to be measured. It is shown that the processing required to give minimum noise error is what is known as 'matched filtering'.

**Signal timing**

In chapter 10 the problem is a little different. Here again the shape of the signal transient is known but the factor to be measured is the time of occurrence, rather than the amplitude, of the signal. The necessary modifications to the matched filtering in this case are discussed.

The need to use a matched filter when determining the amplitude of a signal transient of known shape is well established. Also the modifications needed when signal timing is to be determined are reasonably well documented. However, such coverage is usually for white noise only. In instrumentation, as distinct from communications applications, it is highly likely that drift and $1/f$ noise will be the predominant unwanted signals. It is the distinctive contribution of the present text that the case of $1/f$ noise is covered also.

## 1.2 LOW-PASS FILTERING OF SHOT NOISE

The overview is now complete and we proceed to the main topic of the present chapter. This is the way in which low-pass filtering and visual averaging recover a required signal from white noise. There are two main types of white noise: thermal and shot noise. For all purposes of signal recovery these are indistinguishable. Thus in this section the effect of low-pass filtering on shot noise will be examined. It will be shown in particular that the amplitude $\tilde{v}_n$ of the noise at the filter output is proportional to $T_r^{-1/2}$, where $T_r$ is the filter averaging time. This relation also holds for thermal noise.

**Thermal noise**

It will, nevertheless, be useful to indicate briefly the origins of thermal noise. This originates mainly from the resistors in a circuit and arises from the fact that the mobile charge carriers in a resistor are in constant thermal motion. Thus, at any one time the charge distribution over the resistor is not quite uniform. This causes a potential difference across the ends of the device. As the thermal motion proceeds, the potential fluctuates in a random manner about a mean of zero. This class of noise is discussed further in §4.8, with quantitative precision.

**Shot noise**

Shot noise tends to predominate in the semiconductor components in a circuit, i.e. in the diodes and the transistors. It is actually a manifestation of the fact that an electric current is not a continuum but consists of discrete current carriers, i.e. electrons and 'holes'. Thus the commonly used analogy of a stream of flowing water to represent an electric current might more properly be replaced by a stream of grains of sand. It is then the 'grainy' nature of the flow that constitutes the shot noise.

**Shot noise model**

Figure 1.4 gives a simple model for the shot noise in the amplifier of figure 1.1. Here it is reasonable to assume that all of the shot noise originates from the first stage, because that from later stages is subject to less gain. The transistor

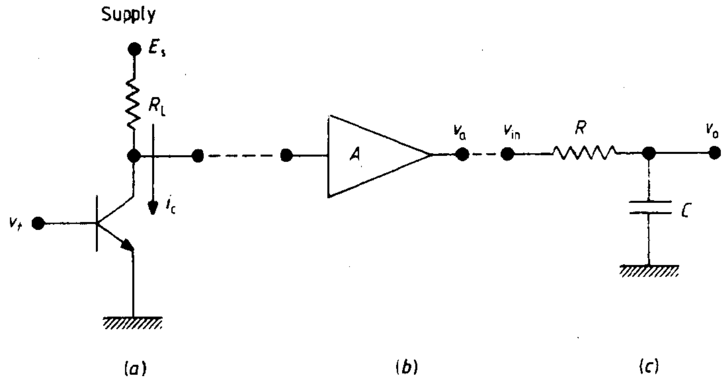


**Figure 1.4** Circuit used to show the effect of the low-pass filter in figure 1.1 upon the amplifier shot noise. (*a*) First stage, (*b*) further stages, (*c*) low-pass filter.

current $i_c$ actually consists of a random train of pulses, with each pulse corresponding to the passage of an electron from the emitter to the collector of the transistor. The area of each pulse will be $q_e$, the electron charge. Thus the area $q$ of the corresponding pulse at the output of the amplifier $A$ will be given by

$$q = q_e R_L A. \tag{1.1}$$

Figure 1.5(*a*) shows the sequence of these random pulses that make up the waveform for $v_a = v_{in}$ in figure 1.4. This is somewhat idealised, assuming that the amplifier contains no low-pass filter, but will do for our present purposes.

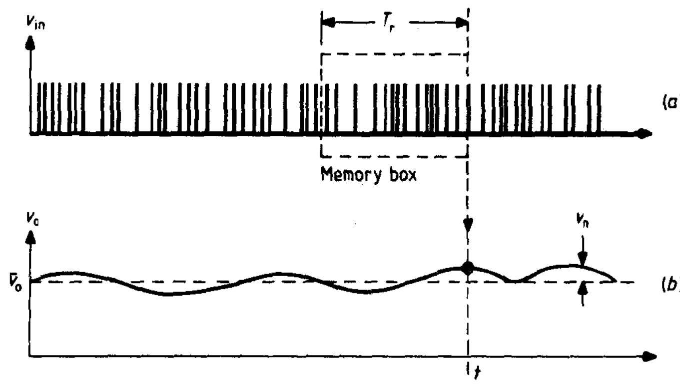


**Figure 1.5** Memory box running average approximation to filtering of shot noise pulse train for transistor in figure 1.4. (*a*) Shot noise pulse train $v_{in}$, (*b*) running average $v_o$.

**Running average**

It will next be shown that the essential feature of the filter, in the present view, is that it forms a 'running average' of the input. Figure 1.5 illustrates this function once more. Here the output $v_o$ at time $t$ is the average of $v_{in}$ over the period $T_r$ of the memory box. Then the output function $v_o$ is constructed as the box 'runs' along the input function. In mathematical terms

$$v_o(t) = T_r^{-1} \int_{t-T_r}^{t} v_{in}(t')\, dt'. \tag{1.2}$$

The fluctuation component $v_n$ of $v_o$ is the shot noise and arises because $v_o$ is proportional to the number of pulses $p$ in the box and $p$ fluctuates because of the random placing of the pulses. The fluctuation time is $T_r$ because this is the time taken for all of the pulses in the box to be replaced.

**Filter response**

The next step is to determine the exact relation between the input $v_{in}$ and output $v_o$ of the low-pass filter and thus to show that $v_o$ is essentially the running average of $v_{in}$.

Consider first the 'impulse response' $g_i$ of the filter, shown in figure 1.6(*f*).

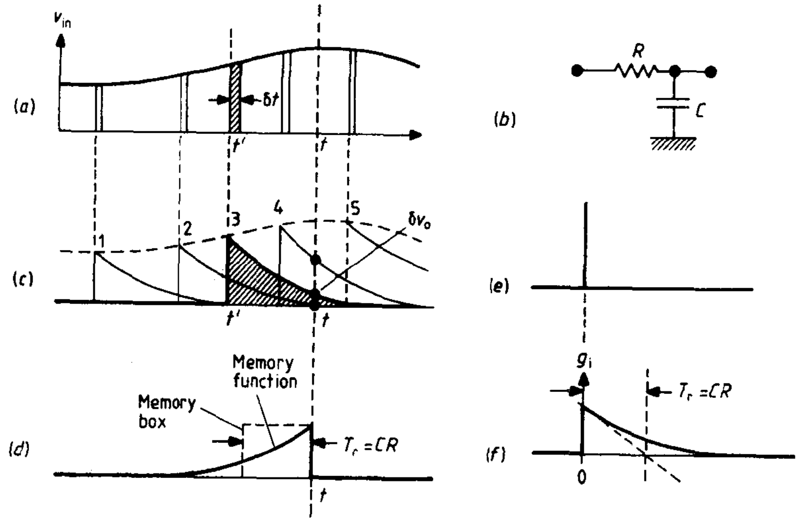


**Figure 1.6** Relation between output $v_o$ and input $v_{in}$ for a low-pass filter. (*a*) Input $v_{in}$, (*b*) filter, (*c*) components of output $v_o$ originating from elements in (*a*), (*d*) memory function, (*e*) input impulse, (*f*) impulse response.

This is the response of the filter to the unit impulse shown in figure 1.6(*e*), which is applied at time $t=0$. Here the unit impulse is a pulse of unit area and of width approaching zero.

The input impulse leaves the capacitor with a charge. This decays exponentially through the resistor $R$ as shown, with the decay time $CR$.

Next consider the shaded element at time $t'$ of the general input waveform $v_{in}$ shown in (*a*). For $\delta t$ small, this approximates to an impulse of area $v_{in}(t')\,\delta t$. This will produce at the filter output the shaded transient shown in (*c*). At the time $t$, the output component $\delta v_o$ will then equal the impulse area multiplied by $g_i(t-t')$, i.e. $v_{in}(t')g_i(t-t')\,\delta t$. Then summing $\delta v_o$ for all values of $v_{in}$,

$$v_o(t)=\int_{-\infty}^{t} v_{in}(t')g_i(t-t')\,\mathrm{d}t'. \tag{1.3}$$

Looking at the transients in (*c*) following the various elements of $v_{in}$ shown in (*a*), and numbered 1 to 5, it is clear that $v_o$ consists of contributions from $v_{in}$ extending back from time $t$ at which the filter output is observed for a period roughly equal to $CR$. This is highly reminiscent of the running average, with the averaging period $T_r = CR$. The main difference is that the 'memory function' $g_i(t-t')$ from equation (1.3) and shown in (*d*) has an exponential 'taper'. This contrasts with the abrupt cut-off of the 'memory-box' approximation corresponding to the running average. This property is sometimes referred to as a 'fading' memory of the input.