THE SAMPLING THEOREM

THE SAMPLING THEOREM

ACHIEVEMENTS: *experimental verification of the sampling theorem; sampling and message reconstruction (interpolation)*

PREREQUISITES: *completion of the experiment entitled* **Modelling an equation**.

PREPARATION

A sample is part of something. How many samples of something does one need, in order to be able to deduce what the something is ? If the something was an electrical signal, say a message, then the samples could be obtained by looking at it for short periods on a regular basis. For how long must one look, and how often, in order to be able to work out the nature of the message whose samples we have - to be able to *reconstruct* the message from its samples ?

This could be considered as merely an academic question, but of course there are practical applications of sampling and reconstruction.

Suppose it was convenient to transmit these samples down a channel. If the samples were short, compared with the time between them, and made on a regular basis - *periodically* - there would be lots of time during which nothing was being sent. This time could be used for sending something else, including a set of samples taken of another message, at the same rate, but at slightly different times. And if the samples were narrow enough, further messages could be sampled, and sandwiched in between those already present. Just how many messages could be packed into the channel ?

The answers to many of these questions will be discovered during the course of this experiment. It is *first* necessary to show that sampling and reconstruction are, indeed, possible !

The *sampling theorem* defines the conditions for successful sampling, of particular interest being the minimum rate at which samples must be taken. You should be reading about it in a suitable text book. A simple analysis is presented in Appendix A to this experiment.

This experiment is designed to introduce you to some of the fundamentals, including determination of the minimum sampling rate for distortion-less reconstruction.

EXPERIMENT

taking samples

In the first part of the experiment you will set up the arrangement illustrated in Figure 1. Conditions will be such that the requirements of the Sampling Theorem, not yet given, are met. The message will be a single audio tone.

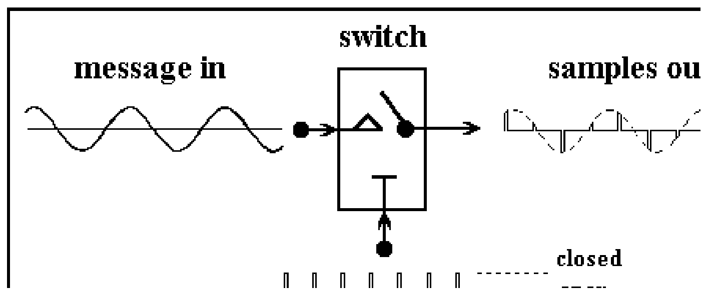

Figure 1: sampling a sine wave

To model the arrangement of Figure 1 with TIMS the modules required are a TWIN PULSE GENERATOR (only one pulse is used), to produce s(t) from a clock signal, and a DUAL ANALOG SWITCH (only one of the switches is used). The TIMS model is shown in Figure 2 below.

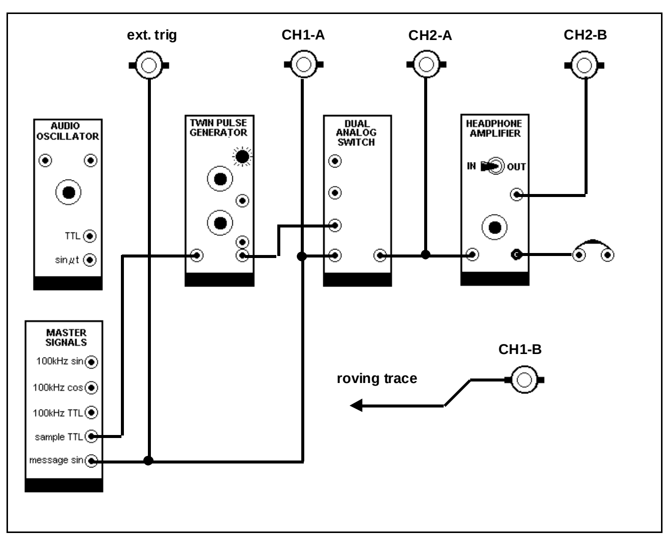

Figure 2: the TIMS model of Figure 1

T1 *patch up the model shown in Figure 2 above. Include the oscilloscope connections. Note the oscilloscope is externally triggered from the message.*

note: the oscilloscope is shown synchronized to the message. Since the message frequency is a sub-multiple of the sample clock, the sample clock could also have been used for this purpose. However, later in the experiment the message and clock are not so related. In that case the choice of synchronization signal will be determined by just what details of the displayed signals are of interest. *Check out this assertion as the experiment proceeds.*

T2 *view CH1-A and CH2-A, which are the message to be sampled, and the samples themselves. The sweep speed should be set to show two or three periods of the message on CH1-A*

T3 *adjust the width of the pulse from the TWIN PULSE GENERATOR with the pulse width control. The pulse is the switching function s(t), and its width is δt. You should be able to reproduce the sampled waveform of Figure 3.*

Your oscilloscope display will not show the message in dashed form (!), but you could use the oscilloscope shift controls to superimpose the two traces for comparison.

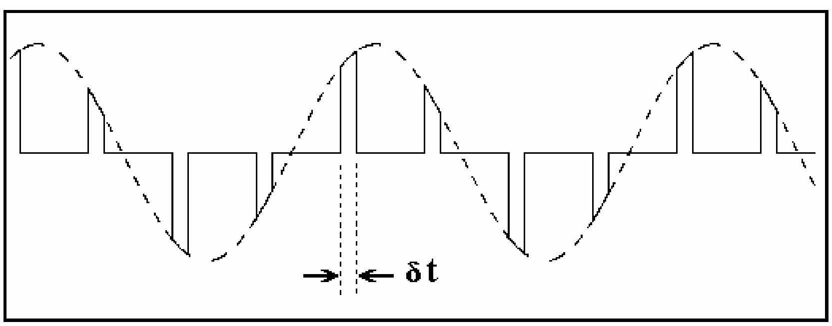

Figure 3: four samples per period of a sine wave.

Please remember that this oscilloscope display is that of a ***VERY SPECIAL CASE***, and is typical of that illustrated in text books.

The message and the samples are stationary on the screen

This is because the frequency of the message is an exact sub-multiple of the sampling frequency. This has been achieved with a message of (100/48) kHz, and a sampling rate of (100/12) kHz.

In general, if the oscilloscope is synchronized to the sample clock, successive views of the message samples would not overlap in amplitude. Individual samples would appear at the *same location* on the time axis, but samples from successive sweeps would be of *different amplitudes*. You will soon see this more general case.

Note that, for the sampling method being examined, the *shape* of the top of each sample is the same as that of the message. This is often called *natural sampling*.

reconstruction / interpolation

Having generated a train of samples, now observe that it is possible to recover, or *reconstruct (*or *interpolate)* the message from these samples.

From Fourier series analysis, and consideration of the nature of the sampled signal, you can already conclude that the spectrum of the sampled signal will contain components at and around harmonics of the switching signal, and hopefully the message itself. If this is so, then a lowpass filter would seem the obvious choice to extract the message. This can be checked by experiment.

Later in this experiment you will discover the properties this filter is required to have, but for the moment use the 3 kHz LPF from the HEADPHONE AMPLIFIER.

The reconstruction circuitry is illustrated in Figure 4.

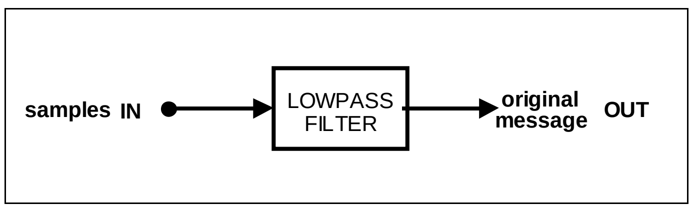

Figure 4: reconstruction circuit.

You can confirm that it recovers the message from the samples by connecting the output of the DUAL ANALOG SWITCH to the input of the 3 kHz LPF in the HEADPHONE AMPLIFIER module, and displaying the output on the oscilloscope.

T4 *connect the message samples, from the output of the DUAL ANALOG SWITCH, to the input of the 3 kHz LPF in the HEADPHONE AMPLIFIER module, as shown in the patching diagram of Figure 2.*

T5 *switch to CH2-B and there is the message. Its amplitude may be a little small, so use the oscilloscope CH2 gain control. If you choose to use a BUFFER AMPLIFIER, place it at the output of the LPF. Why not at the input ?*

The sample width selected for the above measurements was set arbitrarily at about 20% of the sampling period. What are the consequences of selecting a different width ?

sample width

Apart from varying the time interval between samples, what effect upon the message reconstruction does the sample width have ? This can be determined experimentally.

T6 *vary the width of the samples, and report the consequences as observed at the filter output*

reconstruction filter bandwidth

Demonstrating that reconstruction is possible by using the 3 kHz LPF within the HEADPHONE AMPLIFIER was perhaps cheating slightly ? Had the reconstructed message been distorted, the distortion components would have been removed by this filter, since the message frequency is not far below 3 kHz itself. Refer to the experiment entitled *Amplifier overload* (within *Volume A2 - Further & Advanced Analog Experiments*), and the precautions to be taken when measuring a narrow band system. The situation is similar here. As a check, you should lower the message frequency. This will also show some other effects. Carry out the next Task.

T7 *replace the 2 kHz message from the MASTER SIGNALS module with one from an AUDIO OSCILLATOR. In the first instance set the audio oscillator to about 2 kHz, and observe CH1-A and CH2-A simultaneously as you did in an earlier Task. You will see that the display is quite different.*

The individual samples are no longer visible - the display on CH2-A is not stationary.

T8 *change the oscilloscope triggering to the sample clock. Report results.*

T9 *return the oscilloscope triggering to the message source. Try fine adjustments to the message frequency (sub-multiples of the sampling rate).*

This time you have a different picture again - the message is stationary, but the samples are not. You can see how the text book display is just a snap shot over a few samples, and not a typical oscilloscope display *unless* there is a relationship between the message and sampling rate [1].

It is possible, as the message frequency is fine tuned, to achieve a stationary display, but only for a moment or two.

---

[1] or you have a special purpose oscilloscope

Now that you have a variable frequency message, it might be worthwhile to re-check the message reconstruction.

T10 *look again at the reconstructed message on CH2-B. Lower the message frequency, so that if any distortion products are present (harmonics of the message) they will pass via the 3 kHz LPF.*

pulse shape

You have been looking at a form of pulse amplitude modulated (PAM) signal. If this sampling is the first step in the conversion of the message to digital form, the next step would be to convert the pulse amplitude to a digital number. This would be pulse code modulation (PCM) [2].

The importance of the pulse shape will not be considered in this experiment. We will continue to consider the samples as retaining their shapes (as shown in the Figure 3, for example). Your measurements should show that the amplitude of the reconstituted message is *directly proportional* to the width of the samples.

to find the minimum sampling rate

Now that you have seen that an analog signal can be recovered from a train of periodic samples, you may be asking:

what is the slowest practical sampling rate for the recovery process to be successful ?

The sampling theorem was discovered in answer to this question. You are invited now to re-enact the discovery:

- use the 3 kHz LPF as the reconstruction filter. The highest frequency message that this will pass is determined by the filter passband edge f_c, nominally 3 kHz. You will need to measure this yourself. See Appendix B to this experiment.
- set the message frequency to f_c.
- use the VCO to provide a variable sampling rate, and reduce it until the message can no longer be reconstructed without visible distortion.
- use, in the first instance, a fixed sample width δt, say 20% of the sampling period.

The above procedure will be followed soon; but first there is a preparatory measurement to be performed.

[2] if the pulse is wide, with a sloping top, what is its amplitude ?

preparation

MDSDR

In the procedure to follow you are going to report when it is just visibly obvious, in the time domain, when a single sinewave has been corrupted by the presence of another. You will use frequencies which will approximate those present during a later part of the experiment.

The frequencies are:

- wanted component - 3 kHz
- unwanted component - 4 kHz

Suppose initially the amplitude of the unwanted signal is zero volt. While observing the wanted signal, in the time domain, how large an amplitude would the unwanted signal have to become for its presence to be (just) noticed ?

A knowledge of this phenomenon will be useful to you throughout your career. An estimate of this amplitude ratio will now be made with the model illustrated in Figure 5.

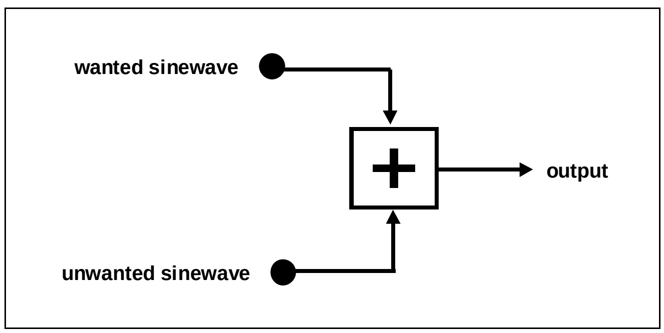

Figure 5: corruption measurement

T11 *obtain a VCO module. Set the 'FSK - VCO' switch, located on the circuit board, to 'VCO'. Set the front panel 'HI - LO' switch to 'LO'. Then plug the module into a convenient slot in the TIMS unit.*

T12 *model the block diagram of Figure 5. Use a VCO and an AUDIO OSCILLATOR for the two sinewaves. Reduce the unwanted signal to zero at the ADDER output. Set up the wanted signal output amplitude to say 4 volt peak-to-peak. Trigger the oscilloscope to the source of this signal. Increase the amplitude of the unwanted signal until its presence is just obvious on the oscilloscope. Measure the relative amplitudes of the two signals at the ADDER output. This is your MDSDR - the <u>maximum detectable signal-to-distortion ratio</u>. It would typically be quoted in decibels.*

use of MDSDR

Consider the spectrum of the signal samples. Refer to Appendix A of this experiment if necessary.

Components in the lower end of the spectrum of the sampled signal are shown in Figure 6 below. It is the job of the LPF to extract the very lowest component, which is the message (here represented by a single tone at frequency μ rad/s).

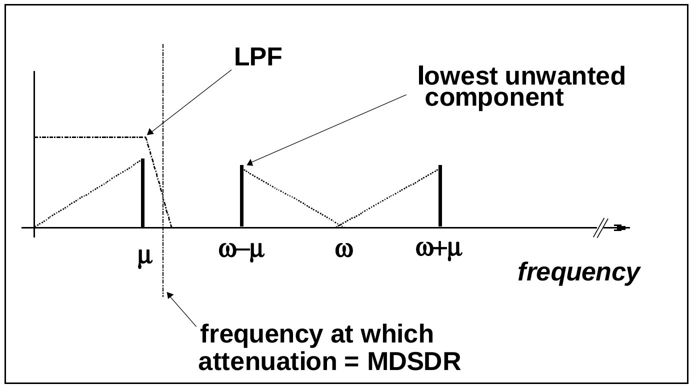

Figure 6: lower end of the spectrum of the sampled signal

During the measurement to follow, the frequency 'ω' will be gradually reduced, so that the unwanted components move lower in frequency towards the filter passband.

You will be observing the wanted component as it appears at the output of the LPF. The closest unwanted component is the one at frequency (ω - μ) rad/s.

Depending on the magnitude of 'ω', this component will be either:

1. outside the filter passband, and not visible in the LPF output (as in Figure 6)
2. in the transition band, and perhaps visible in the LPF output
3. within the filter passband, and certainly visible in the LPF output

Assuming both the wanted and unwanted components have the same amplitudes, the presence of the unwanted component will first be noticed when 'ω' falls to the frequency marked on the transition band of the LPF. This equals, in decibels, the MDSDR.

T13 *measure the frequency of your LPF at which the attenuation, relative to the passband attenuation, is equal to the MDSDR. Call this* f_{MDSDR}.

minimum sampling rate measurement

T14 *remove the patch lead from the 8.333 kHz SAMPLE CLOCK source on the MASTER SIGNALS module, and connect it instead to the VCO TTL OUTPUT socket. The VCO is now the sample clock source.*

T15 *use the FREQUENCY COUNTER to set the VCO to 10 kHz or above.*

T16 *use the FREQUENCY COUNTER to set the AUDIO OSCILLATOR to f_c, the edge of the 3 kHz LPF passband.*

T17 *synchronize the oscilloscope to the sample clock. Whilst observing the samples, set the sample width δt to about 20% of the sampling period.*

The sampling theorem states, inter alia, that the minimum sampling rate is twice the frequency of the message.

Under the above experimental conditions, the sampling rate is well above this minimum.

T18 *synchronize the oscilloscope to the message, direct from the AUDIO OSCILLATOR, and confirm that the message being sampled, and the reconstructed message, are identical in shape and frequency (the difference in amplitudes is of no consequence here).*

It is now time to determine the minimum sampling rate for undistorted message reconstruction.

T19 *whilst continuing to monitor both the message and the reconstructed message, slowly reduce the sampling rate (the VCO frequency). As soon as the message shows signs of distortion (aliasing distortion), increase the sampling rate until it just disappears. The sampling rate will now be the minimum possible.*

T20 *calculate the frequency of the unwanted component. It will be the just-measured minimum sampling rate, minus the message frequency. How does this compare with f_{MDSDR} measured in Task 13 ?*

T21 *compare your result with that declared by the sampling theorem. Explain discrepancies !*

further measurements

A good engineer would not stop here. Whilst agreeing that it is possible to sample and reconstruct a single sinewave, he would call for a more demanding test. Qualitatively he might try a speech message. Quantitatively he would probably try a two-tone test signal.

What ever method he tries, he would make sure he used a band-limited message. He will then know the highest frequency contained in the message, and adjust his sampling rate with respect to this.

If you have bandlimited speech available at TRUNKS, or a SPEECH MODULE, you should repeat the measurements of the previous section.

the two-tone test message

A two-tone test message consists of two audio tones added together.

The special properties of this test signal are discussed in the chapter entitled *Introduction to modelling with TIMS* (of this Volume) in the section headed *The two tone test signal*, to which you should refer. You should also refer to the experiment entitled *Amplifier overload* (within *Volume A2 - Further & Advanced Analog Experiments*).

You can make a two-tone test signal by adding the output of an AUDIO OSCILLATOR to the 2 kHz message from the MASTER SIGNALS module.

There may be a two-tone test signal at TRUNKS, or use a SPEECH Module.

summing up

You have been introduced to the principles of sampling and reconstruction.

The penalty for selecting too low a sampling rate was seen as distortion of the recovered message. This is known as *aliasing distortion*; the filter has allowed some of the unwanted components in the spectrum of the sampled signal to reach the output. Analysis of the spectrum can tell you where these have come from, and so how to re-configure the system - more appropriate filter, or faster sampling rate ? In the laboratory you can make some independent measurements to reach much the same conclusions.

In a practical situation it is necessary to:

1. select a filter with a passband edge at the highest message frequency, and a stopband attenuation to give the required signal to noise-plus-distortion ratio.
2. sample at a rate at least equal to the filter slot [3] band width *plus* the highest message frequency. This will be higher than the theoretical minimum rate. Can you see how this rate was arrived at ?

An application of sampling can be seen in the experiment entitled *Time division multiplexing - PAM* (within this Volume).

TUTORIAL QUESTIONS

Q1 *even if the signal to be sampled is already bandlimited, why is it good practice to include an anti-aliasing filter ?*

[3] the 'slot band' is defined in Appendix A at the rear of this Volume.

Q2 *in the experiment the patching diagram shows that the non-delayed pulse was taken from the TWIN PULSE GENERATOR to model the switching function s(t). What differences would there have been if the delayed pulse had been selected ? Explain.*

***note**: both pulses are of the same nominal width.*

Q3 *consider a sampling scheme as illustrated in Figure 1. The sampling rate is determined by the distance between the pulses of the switching function s(t). Assume the message was reconstructed using the scheme of Figure 4.*

Suppose the pulse rate was slowly increased, whilst keeping the pulse width fixed. Describe and explain what would be observed at the lowpass filter output.

APPENDIX A

analysis of sampling

sampling a cosine wave

Using elementary trigonometry it is possible to derive an expression for the spectrum of the sampled signal. Consider the simple case where the message is a single cosine wave, thus:

$$m(t) = V.\cos\mu t \quad \text{........ A-1}$$

Let this message be the input to a switch, which is opened and closed periodically. When closed, any input signal is passed on to the output.

The switch is controlled by a switching function s(t). When s(t) has the value '1' the switch is closed, and when '0' the switch is open. This is a periodic function, of period T, where:

$$T = (2.\pi) / \omega \text{ sec} \quad \text{........ A-2}$$

and is expressed analytically by the Fourier series expansion of eqn. A-3 below.

$$s(t) = a_o + a_1.\cos\omega t + a_2.\cos 2\omega t + a_3.\cos 3\omega t + ... \quad \text{........ A-3}$$

The coefficients a_i in this expression are a function of ($\delta t/T$) of the pulses in s(t), which is illustrated in Figure A-1 below.

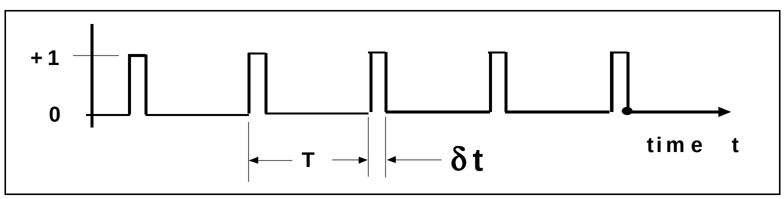

Figure A-1: the switching function s(t)

The sampled signal is given by:

$$\text{sampled signal } y(t) = m(t).\ s(t) \quad \text{........ A-4}$$

Expansion of y(t), using eqns. A-1 and A-3, shows it to be a series of DSBSC signals located on harmonics of the switching frequency ω, including the zeroeth harmonic, which is at DC, or baseband. The magnitude of each of the coefficients a_i will determine the amplitude of each DSBSC term.

The frequency spectrum of this signal is illustrated in graphical form in Figure A-2.

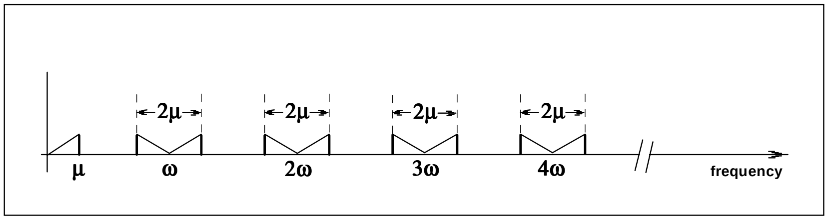

Figure A-2: the sampled signal in the frequency domain

Figure A-2 is representative of the case when the ratio (δt / T) is very small, making adjacent DSBSC amplitudes almost equal, as shown.

A special case occurs when (δt / T) = 0.5 which makes s(t) a square wave. It is well known for this case that the even a_i are all zero, and the odd terms are monotonically decreasing in amplitude.

The important thing to notice is that:

1. the DSBSC are spaced apart, in the frequency domain, by the sampling frequency ω rad/s.
2. the bandwidth of each DSBSC extends either side of its centre frequency by an amount equal to the message frequency μ rad/s.
3. the lowest frequency term - the baseband triangle - is the message itself.

Inspection of Figure A-2 reveals that, provided:

$$\omega \geq 2.\mu \qquad \text{........ A-5}$$

there will be no overlapping of the DSBSC, and, specifically, *the message can be separated from the remaining spectral components by a lowpass filter*.

That is what the sampling theorem says.

practical issues

When the sampling theorem says that the slowest useable sampling rate is twice the highest message frequency, it assumes that:

1. the message is truly bandlimited to the highest message frequency μ rad/s.
2. the lowpass filter which separates the message from the lowest DSBSC signal is *brick wall*.

Neither of these requirements can be met in practice.

If the message is bandlimited with a practical lowpass filter, account must be taken of the finite transition bandwidth in assessing that frequency beyond which there is no significant message energy.

The reconstruction filter will also have a finite transition bandwidth, and so account must be taken of its ability to suppress the low frequency component of the lowest frequency DSBSC signal.

aliasing distortion.

If the reconstruction filter does not remove all of the unwanted components - specifically the lower sideband of the nearest DSBSC, then these will be added to the message. Note that the unwanted DSBSC was derived from the original message. It will be a *frequency inverted* version of the message, shifted from its original position in the spectrum. The distortion introduced by these components, if present in the reconstructed message, is known as aliasing distortion.

anti-alias filter

No matter how good the reconstruction filter is, it cannot compensate for a non-bandlimited message. So as a first step to eliminate aliasing distortion the message must be bandlimited. The band limiting is performed by an *anti-aliasing filter*.

APPENDIX B

3 kHz LPF response

For this experiment it is necessary to know the frequency response of the 3 kHz LPF in your HEADPHONE AMPLIFIER.

If this is not available, then you must measure it yourself.

Take enough readings in order to plot the filter frequency response over the full range of the AUDIO OSCILLATOR. Voltage readings accurate to 10% will be adequate.

A measurement such as this is simplified if the generator acts as a pure voltage source; this means, in effect, that its amplitude should remain constant (say within a few percent) over the frequency range of interest. It is then only necessary to record the filter output voltage versus frequency. Check that the AUDIO OSCILLATOR meets this requirement.

Select an in-band frequency as reference - say 1 kHz. Call the output voltage at this frequency V_{ref}. Output voltage measurements over the full frequency range should then be recorded, and from them the normalized response, in dB, can be plotted.

Thus, for an output of V_o, the normalized response, in dB, is:

$$\text{response} = 20 \log_{10} (V_o / V_{ref}) \text{ dB}$$

Plot the response, in dB, versus log frequency. Prepare a table similar to that of Table B-1, and complete the entries.

Characteristic	Magnitude
passband width kHz	
transition band ratio	
stopband attenuation dB	
slot band width	

Table B-1: LPF filter characteristic

The *transition band* lies between the edge of the passband f_o and the start of the stop band f_s. The *transition band ratio* is (f_s / f_o). The *slot band* is defined as the sum of the passband and the transition band.

For comparison, the theoretical response of a 5th order elliptic filter is shown in Figure B-1. This has a passband edge at 3 kHz, passband ripple of 0.2 dB, and a stopband attenuation of 50 dB.

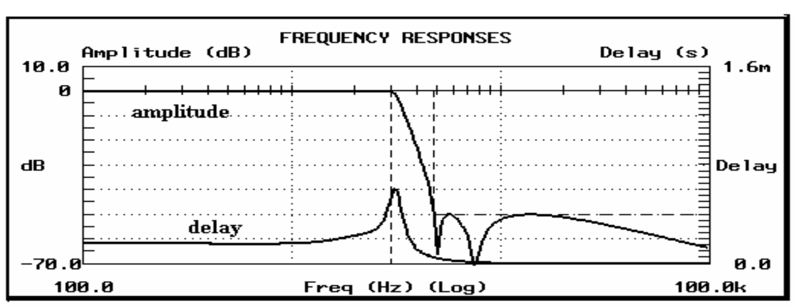

Figure B-1: theoretical amplitude response of the 5th order elliptic