# PERCEPTION OF PITCH IN PULSE WAVE FORMS WHOSE POWER SPECTRA ARE FLAT

Thesis by

Craig McClain Cheetham

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1977

(Submitted May 19, 1977)

ACKNOWLEDGEMENTS

Dr. John R. Pierce provided the inspiration and encouragement for the work. Dr. Hardy C. Martel also provided very helpful consultation on the electronic design. Thanks are also due to my wife, Karen, for her editorial assistance and typing. Marlys Ricards also contributed to the typing effort and Betty Wood provided many of the drawings.

A particular debt of gratitude is owed to the subjects who contributed many hours in the sound booth. Among these are John Gustafson, without whom this work may well have been very much shorter and completely wrong, James Boyk and Ruth Morris. The generous financial support of Bell Telephone Laboratories is gratefully acknowledged.

## ABSTRACT

These experiments were performed to explore the upper limits of pitch perception based on the time waveform of the stimulus. Flanagan and Guttman, using pulses with various polarity patterns, showed that pitch may be assigned to a sound based on either the time waveform or the Fourier spectrum. Their results show these two mechanisms in conflict in the region between 100 and 200 hertz. The stimulus used here consisted of a train of uniformly spaced pulses of equal amplitude whose polarities were randomly chosen. Such a signal is shown to have a flat power spectrum and as such should offer no competition to the time-based pitch mechanism. Subjects listened alternately to the stimulus and to a unipolar pulse generator whose pulse rate was under their control. The subjects were instructed to match the two pitches. Some musically talented subjects were found who could match the pitch even when the random polarity pulse rate was as high as 9.5 kHz. Matching is defined to occur when an integral relation exists between the pulse rates as set by the experimenter and the subject Matching at high rates was found to occur at stimulus levels as low as 10 db or when both the stimulus and matching signals were high-pass filtered at 8 kHz.

Investigation of the short term spectrum of the random polarity pulse train shows that there are clues to the pulse rate even though the long term power spectrum is flat. One of these clues is the different probability distribution of the amplitude peaks at frequencies of nfp/2 where fp is the pulse rate. This variability of amplitude distribution can be reduced, and the frequencies which then occur

changed, by using polarity patterns randomly chosen from a certain set of patterns. However, experiments performed using pulse patterns showed little or no change in subjects' ability to match.

# TABLE OF CONTENTS


| Chapter | Title | Page |
|---|---|---|
| | Acknowledgements | ii |
| | Abstract | iii |
| | Table of Contents | v |
| I | Theory of Pitch Perception: Past to Present | 1 |
| | 1.1 Anatomy of the ear | 1 |
| | 1.2 Place theory of pitch | 3 |
| | 1.2.1 Helmholtz' work | 3 |
| | 1.2.2 Von Bekesy | 6 |
| | 1.2.3 Neuron timing curves | 7 |
| | 1.2.4 Difficulties in place theory of pitch | 9 |
| | 1.3 Time theory of pitch | 10 |
| | 1.3.1 Periodicity pitch | 10 |
| | 1.3.2 Synchrony in neuron responses | 13 |
| | 1.3.3 Experiments with bipolar periodic pulse waveforms | 15 |
| | 1.3.3a Flanagan and Guttman -- Initial experiment | 15 |
| | 1.3.3b Basilar membrane model | 18 |
| | 1.3.3c Effect of removal of low-order harmonics | 19 |
| | 1.3.3d Effect of fixed high-pass filter | 20 |
| | 1.3.3e Effect of selective masking | 21 |
| | 1.3.4 Other evidence for use of time information in the auditory system | 22 |
| II | The Experiment: Apparatus and Procedures | 24 |
| | 2.1 Introduction | 24 |
| | 2.2 Initial experiment | 25 |
| | 2.2.1 Objective | 25 |
| | 2.2.2 Apparatus | 25 |
| | 2.2.3 Procedure | 25 |

TABLE OF CONTENTS (CONT'D)


| Chapter | Title | Page |
|---|---|---|
| | 2.3 Problems encountered | 27 |
| | 2.4 Use of fixed high-pass filter | 32 |
| III | Discussion of Data | 34 |
| | 3.1 Definition of terms | 34 |
| | 3.2 Choice of clock frequencies | 35 |
| | 3.3 Discussion of raw data | 35 |
| | 3.4 Error measures | 57 |
| | 3.4.1 Logarithmic error | 57 |
| | 3.4.2 Arithmetic error | 58 |
| | 3.5 Use of critical bands | 59 |
| | 3.6 Discussion of error plots | 61 |
| | 3.6.1 Figures | 61 |
| | 3.6.2 Type 1 experiments | 62 |
| | 3.6.3 High-pass filter experiments | 63 |
| | 3.6.4 Pattern experiments | 64 |
| IV | Discussion of Results | 86 |
| | 4.1 Initial expectations | 86 |
| | 4.2 Flaws in experimental design removed | 86 |
| | 4.2.1 Equipment flaws | 87 |
| | 4.2.2 Loudness disparity | 87 |
| | 4.3 Possible explanations of results | 88 |
| | 4.3.1 Pulse rate match | 88 |
| | 4.3.2 Short term spectrum anomalies | 89 |
| | 4.3.2a Probability distribution anomalies | 89 |
| | 4.3.2b Zero crossing periodicities | 90 |
| | 4.3.2c Spectral periodicities | 91 |

TABLE OF CONTENTS (CONT'D)


| Chapter | Title | Page |
|---|---|---|
| | 4.3.3 Internal nonlinearities | 92 |
| | 4.4 Summary | 93 |
| Appendices | | 95 |
| | A: Derivation of spectrum of pulse trains of random polarity | 95 |
| | B: Feedback shift registers | 98 |
| | C: Spectrum of random pulse patterns | 105 |
| | D: Amplitude distributions of filtered pulse trains | 111 |
| | E: Custom experimental equipment | 117 |
| References | | 130 |

# CHAPTER I THEORY OF PITCH PERCEPTION: PAST TO PRESENT

The research to be described here has been undertaken to shed light on the mechanisms used by the brain in the perception of pitch in auditory signals. There is a long-standing controversy in auditory theory concerning pitch perception, usually referred to as "the place theory versus the time theory." This chapter will review these two theories and the main experimental results that lead to them.

## 1.1 Anatomy of the ear

Figure 1 shows the human ear in cross-section with the cochlea uncoiled. The cochlea is in fact coiled like a snail shell

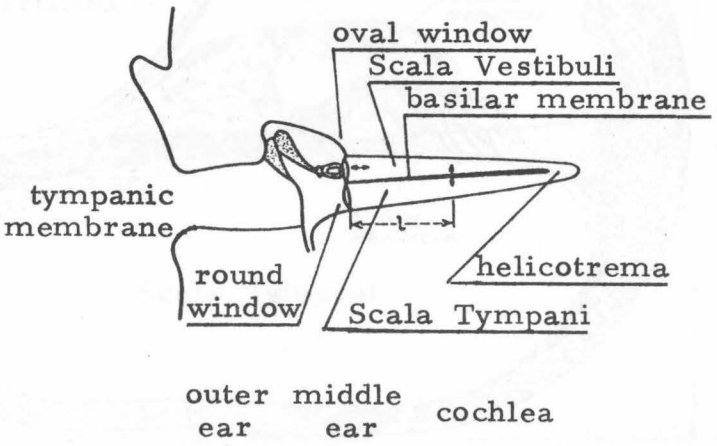



FIGURE 1 (taken from Ref. 1)

but that is inconvenient for the purposes of discussion. An external sound wave enters the outer ear and causes the tympanic membrane to vibrate. These vibrations are transmitted by the ossicles of the middle ear to the oval window of the cochlea. The

middle ear is a complex mechanical system, and as such it has a frequency-dependent transmission coefficient. The middle ear can be modelled reasonably well by a low-pass filter with a cut-off around 1000 Hz.[1,2] The middle ear serves as an impedance converter from the low acoustic impedance of the air to the high impedance of the cochlear fluid.

The cochlea is shown in cross section in figure 2. The cochlea is a fluid filled tube divided lengthwise by the basilar

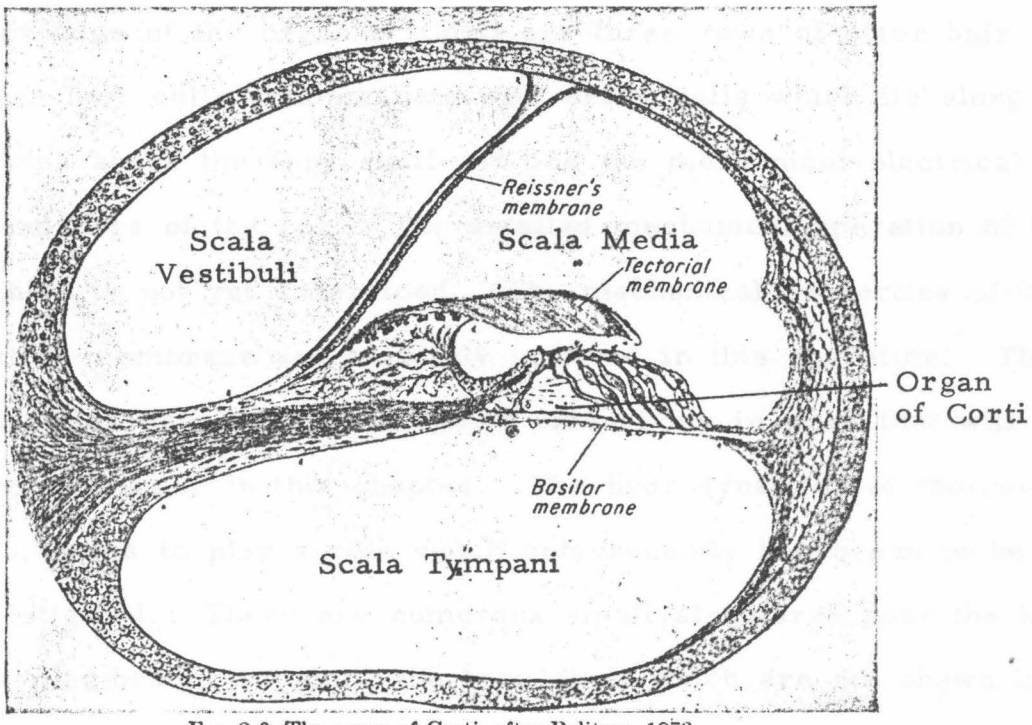



FIG. 2-6. The organ of Corti, after Politzer, 1873.

FIGURE 2 (taken from Ref. 4)

membrane. Reissner's membrane separates the different fluids of the Scala Vestibuli and Scala Media but apparently has little effect on the mechanics of the cochlea. The Scala Vestibuli and Scala Tympani communicate at the far (apical) end through the

helicotrema. Thus an inward deflection of the oval window causes a downward pressure on the basilar membrane and a fluid flow through the helicotrema and results in an outward deflection of the round window. It also should be noted that the basilar membrane steadily widens toward the helicotrema. Vibrations of the middle ear are transmitted to vertical vibrations of the basilar membrane. Alongside the organ of Corti are two different types of hair cells. There is a single row of inner hair cells on the side toward the bony shelf (to the left in figure 2). Along the other edge of the organ of Corti are three rows of outer hair cells. These hair cells communicate with nerve cells which lie along the cochlea above the bony shelf and are the mechanical-electrical transducers of the ear. The detailed mechanical operation of the cochlea is not yet understood. The mechanical properties of the basilar membrane are certainly involved in this operation. These properties have been investigated extensively in work that will be discussed later in this chapter. The hydrodynamics of the cochlea also seems to play a role which only recently has begun to be investigated. There are numerous small structures near the basilar membrane and along the bony shelf which are not shown in figures 1 or 2 and whose role in the function of the cochlea are not understood.

## 1.2 Place theory of pitch

### 1.2.1 Helmholtz' work

In his book, On the Sensations of Tone[3] (published in 1863), Helmholtz offered an explanation of the operation of the ear. His

work became the basis of what is known today as the place theory of pitch. Helmholtz' theory was based on musical experience, psychophysical experiments and the anatomy of the ear. He had glass chambers constructed with a tube at one end for insertion into the ear and an opening at the other end to allow sound to enter. These chambers were called resonators and acted as band-pass filters. Such a resonator has several resonant frequencies which are not harmonically related. The main psychoacoustic experiment showed that people are able to hear the separate harmonic components in a periodic sound. In everyday listening one does not do this, musical notes for instance are perceived as single sounds. Helmholtz was able to hear the harmonics, and enabled other subjects to, by using his resonators to direct the attention to a particular harmonic. His procedure was to sound a tone, listen to it through a resonator tuned to a particular harmonic of the tone and then, with the pitch of the harmonic firmly in mind, remove the resonator. It is important to emphasize that the sole purpose of the resonator is to direct the attention to the correct pitch, the ear is capable of singling out the harmonic unaided. With practice it is possible to hear the upper harmonics without the resonator at all. Objections that this is all in the imagination were answered by producing tones with no harmonics at a given frequency and observing that no tone was heard at that frequency. Helmholtz accomplished this by plucking strings at the nodes of the harmonics he intended to exclude.

Helmholtz' next experiment was designed to show the effect

of the phase relations of the various harmonics of a tone. He built what would today be called a wave synthesizer, i.e., he could generate a sine-wave and its harmonics and separately control their amplitude and phase. To do this, he used a bank of tuning forks, each tuned to a separate harmonic and driven by a coil. A separate, electrically driven, tuning fork, which dipped a contact in and out of a mercury cup, provided the driving current for the coils. A resonant chamber was attached near each tuning fork to make it audible. The amplitude of each harmonic was controlled by changing the distance between a fork and its resonator, while the phase was controlled by detuning the resonator slightly. With this apparatus, he succeeded in synthesizing the vowels (except E and I, he apparently could not get enough sound out of the high frequency forks). Helmholtz then demonstrated that the relative phases of the harmonics did not affect the quality of the tone produced; that is, he could detect no difference when changing the phase of a harmonic. Thus Helmholtz provided psychophysical evidence for the theory which Ohm had previously proposed, namely that the ear performs a Fourier analysis of the incoming sound signal and responds only to the power and not to the phases of the various spectral components.

Helmholtz also offered a physiological basis for Ohm's law, a basis that has become known as the resonator theory. Based on the observation that the basilar membrane split easily in the direction perpendicular to the length of the cochlea, he suggested that the basilar membrane behaved as a series of stretched strings,

each with a different resonant frequency. Because the basilar membrane gets wider toward the apical end, the hypothetical string length increases, thereby lowering the resonant frequency. Therefore, the spectral components of the incoming sound signal would excite the appropriate string, whose motion would be converted to a nerve response by the adjacent hair cells. Helmholtz' psychoacoustical results show that we are able to perceive the separate spectral components, and further, that only their absolute power has any effect.

### 1.2.2 Von Bekesy

The next major contributor to the place theory was Georg von Bekesy.[4] He began his long career in hearing in 1924 when he set out to provide numerical mechanical data about the cochlea. His earliest results destroyed the resonator theory by showing it to be inconsistent with the mechanical properties of the basilar membrane. The most quoted discovery is that when the basilar membrane is cut longitudinally, the sides of the cut do not pull away from each other as they would in an elastic membrane under tension. The resonator theory also required a large difference in the elasticity of the membrane in the transverse and longitudinal directions. Von Bekesy tested for this difference by pressing a thin fiber against the membrane and observing the shape of the resulting deformation. At the apical end, the deformation was nearly circular, and at the basal end, the longitudinal axis of the deformation was longer by only a factor of two. The resonator theory requires the transverse axis to be much longer than the

longitudinal axis.

Von Bekesy devoted a great deal of effort to the development of mechanical models of the cochlea. First he measured the relevant mechanical properties of the cochlea and then built a scaled-up version which could be more easily observed. With the understanding gained from the model, the effects could be observed more easily in the cochlea. He discovered that the dominant mechanical property of the cochlea is the exponentially increasing elasticity of the basilar membrane. Von Bekesy demonstrated that the basilar membrane does indeed respond at a particular place for any given frequency but that the mechanism was that of a wave traveling down the membrane. For a sinusoidal excitation, the wave starts at the basal end of the membrane. As it travels, it slows down and the amplitude increases until at a certain distance along the membrane it quickly dies off. The distance of peak amplitude increases as the frequency of excitation decreases. Thus von Bekesy demonstrated a great deal about the workings of the cochlea but left the place theory of pitch perception largely unchanged.

### 1.2.3 Neuron timing curves

More recently another kind of experimental data has been obtained by numerous investigators to elucidate the operation of the cochlea. These data are the single unit neuron response curves. They are obtained by inserting a microelectrode into a single neuron of the auditory system of an experimental animal. Cats, guinea pigs and squirrel monkeys are the most frequently

used experimental animals. It is well known that the voltage waveform from a single neuron is essentially binary and consists of a train of short pulses followed by a longer resting time. The appropriate measure of the neuron response is the average number of pulses per unit time.[5] When an electrode is inserted into a neuron and no acoustic stimulus is present, a background firing rate may or may not be observed. A stimulus usually causes an increase in the firing rate, termed an excitory response; but frequently a stimulus causes a decrease of inhibitory response. Neurons with no background rate obviously cannot be inhibited. It should be clearly understood that inserting an electrode into a nerve more or less randomly chooses a neuron. Therefore it is not possible to return to the same neuron from animal to animal and no detailed "wiring diagram" exists.

Figure 3 shows the neuron tuning curve for a typical primary neuron,[6] i.e., one that is triggered directly by the hair cells of the cochlea. The acoustic stimulus is usually a gated sine-wave whose amplitude in db is shown as a parameter. The

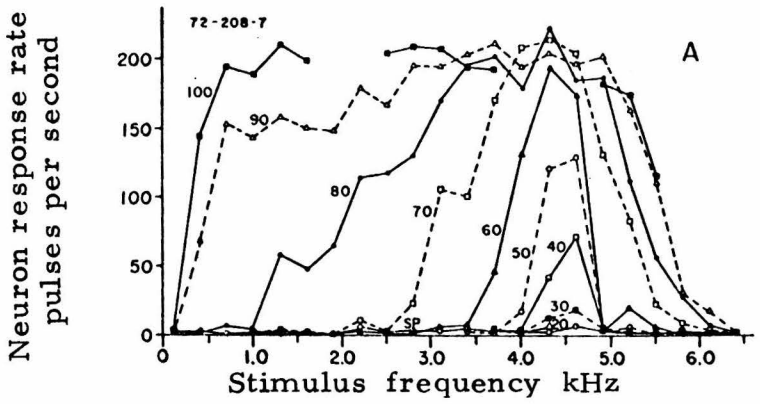



FIGURE 3 (taken from Ref. 6)

primary neuron has a best response frequency (4.5 kHz in figure 3), called the characteristic frequency, at which it responds at low sound levels and at which it reaches its maximum firing rate rapidly with increasing stimulus level. Frequencies below the characteristic frequency also cause the neuron to respond, but the neuron requires higher sound levels. Above the characteristic frequency, the firing rate falls off very rapidly at any given sound level. These curves are consistent with the amplitude of the traveling waves on the basilar membrane. For any fixed point on the membrane, the amplitude of vibration increases with increasing frequency up to some critical frequency beyond which the amplitude dies off rapidly. These results are also in tune with the place theory. However they also point to one of the major difficulties of the place theory -- the very small distinguishable differences in frequency.

### 1.2.4 Difficulties in place theory of pitch

Human subjects can distinguish between two frequencies differing by 0.1% over much of the audio spectrum. Such a change in stimulus has very little impact on the envelope of the amplitude of vibration of the basilar membrane. Von Bekesy met this objection by proposing some form of neural sharpening. He built models of the cochlea with the basilar membrane situated so that the forearm could be placed along it.[4] When the model was driven by a sinusoid, a broad pattern of vibration, as one would find on the basilar membrane, resulted. However, this pattern of vibration is perceived through the tactile sensors of the skin as occurring

over a much narrower extent. The neural tuning curves indicate some sharpening of the response but not enough to account for the fine difference limens found psychophysically. This fact destroys the beautiful simplicity of Helmholtz' original theory that each nerve fiber corresponded to a different pitch. Other psychophysical evidence that does not fit with the place theory has been known for a long time. But before modern electronics, stimuli were hard to generate and their spectra difficult to measure. Thus these experiments lacked impact.

## 1.3 Time theory of pitch

### 1.3.1 Periodicity pitch

In 1940 J. F. Schouten[8] first published his theory of residue pitch. He said that individual spectral components of a sound could be heard if they were sufficiently spaced (Ohm's law), but that closely spaced components that are harmonics of a single frequency are heard as a single percept, a residue, and this residue has a pitch corresponding to the fundamental. Signals with harmonically related components but no fundamental can be synthesized in a variety of ways: 1) A periodic signal such as a pulse train can be high-pass filtered[9], 2) a sine-wave of the frequency, amplitude and phase of the fundamental can be subtracted from the signal[10], or 3) amplitude modulating a high frequency sine-wave (say 2000 Hz) with an integer submultiple (say 200 Hz) sine-wave produces a signal with three components[8],

$$(1+m\cos\omega t)\sin k\omega t = \tfrac{1}{2}m\sin(k-1)\omega t + \sin k\omega t + \tfrac{1}{2}m\sin(k+1)\omega t$$

The objection offered to the residue phenomenon is the

possibility of the fundamental being generated by nonlinearities in the ear. If a periodic signal undergoes a nonlinear transformation, the output will still be periodic; but with the waveform altered a different Fourier decomposition will result which may contain any harmonic of the fundamental depending on the form of the nonlinearity. The ear is known to be nonlinear, and the production of difference tones and combination tones has been explored psychoacoustically.[11, 12]

When spurious harmonics are generated in the ear, they behave just as if they had been present in the sound field. Thus beats can be obtained by introducing a probe tone of a slightly different frequency from the harmonic being investigated. The intensity of the harmonic can be estimated from the intensity of the probe tone that produces the most prominent beats. These studies fail to show any spurious harmonics when the primary tones are less than about 55 to 60 db SPL. Furthermore at these levels, the harmonics are about 40 db below the primary. The level of the spurious harmonic apparently increases faster than the level of the primary, for Moe[11] found harmonics 20 db down with 80 db SPL primaries. Most experiments on periodicity pitch are run at 30 to 40 db above sensation level, and Thurlow and Small[9] found residue effects as low as 10 db.

In another experiment Schouten[8] used a 200 Hz sine-wave of variable amplitude and phase added to a pulse train of 200 Hz. By varying the phase and amplitude, he was able to cancel the sum of the fundamental and any subjective fundamental that might

be present. When this was done, a sharp 200 Hz residue remained. Finally, filtered noise has been used in an attempt to mask the residue pitch.[9,13] This has not succeeded and leaves the conclusion that it is possible to elicit a low pitch using only the high frequency portion of the basilar membrane, in direct contradiction to Ohm's law.

Various other experimenters have used sine-wave generators tuned to harmonic frequencies to study periodicity pitch.[14,15] The general result is that if the phases of the components are adjusted to produce a signal envelope which tends to be flat, no periodicity pitch is heard. If the signal envelope shows marked variations over the period, then the fundamental pitch is heard. However the envelope does not directly determine the pitch. This was shown by work of Schouten[16] in which he shifted the frequency components by a constant frequency while keeping the component separation and thus the envelope period constant. His subjects reported slight pitch shifts and pitch ambiguities in some cases. Schouten

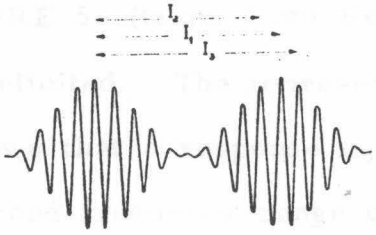


FIGURE 4 (taken from Ref. 16)

concluded that they were responding to peaks (or possibly zero crossings) of the carrier signal which occurred near the envelope peaks. It seems that the pitch extracting mechanism of the ear does not work on the spectral spacing of the signal components,

as has been suggested, but operates directly in the time domain.

1.3.2 Synchrony in neuron responses

Evidence for the direct use of the time domain by the brain also comes from recent neurophysiological experiments. Rose et al.[17] noted the time of occurrence of neural pulses in single units of the squirrel monkey. With pure tone stimuli of 10 to 20 second duration, they found that the length of time between neural pulses strongly tended toward integer multiples of the stimulus period, see figure 5. This held true for any frequency for which

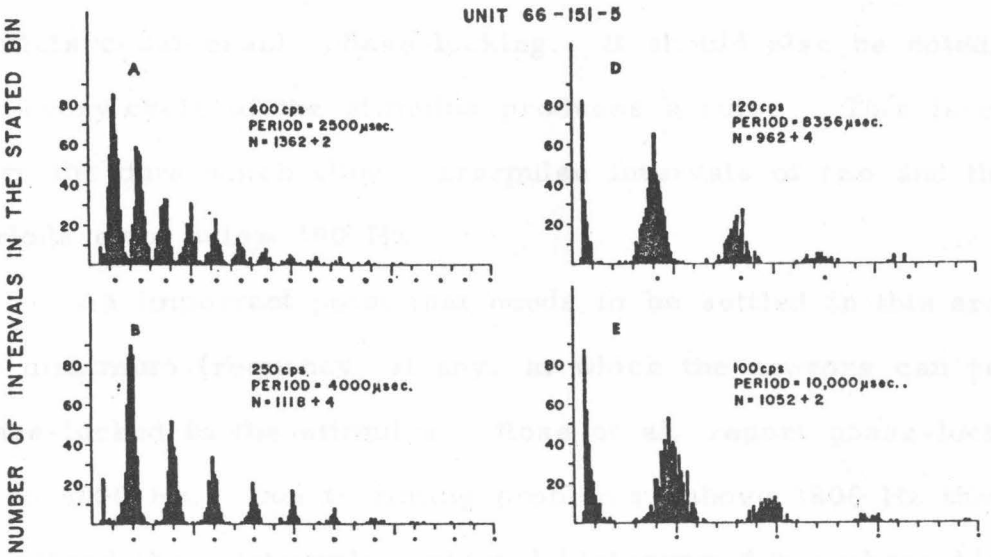



FIGURE 5 (taken from Ref. 17)

a response could be elicited. The representative neural tuning curve of figure 3 shows that a response occurs, for substantial intensities, over a broad frequency range of usually a few hundred hertz to just above the unit's characteristic frequency. The authors note that when the stimulus period becomes less than the refractory period, which they state to be in the range of 800 to 900 microseconds, the peak in the histogram at $T_0$ disappears, and the next few peaks may be delayed in time by a small fraction of

a period. The histogram of figure 5 also has an exponential envelope. The area under the curve is the total number of pulses in the neuron occurring during the stimulus. Thus the total area varies with stimulus frequency and intensity as determined by the unit's tuning curve. This can be seen in the data presented by the authors. However, as the total number of pulses decreases, the exponential envelope tends to flatten. Even at intensities at which the tuning curve implies there is no response (at least for frequencies around the characteristic frequency), the spontaneous response exhibits considerable phase-locking. It should also be noted that not every cycle of the stimulus produces a pulse. This is evident from the data which show interpulse intervals of two and three periods even below 100 Hz.

An important point that needs to be settled in this area is the maximum frequency, if any, at which the neurons can remain phase-locked to the stimulus. Rose et al. report phase-locking up to 5000 Hz. Due to timing problems, above 1800 Hz they abandoned their interpulse interval histogram for a phase histogram. That is, they accumulated the times of occurrence of pulses not from the proceeding pulse but from a reference point of the stimulus. Figure 6 shows a representative histogram. They speculate that lack of evidence of phase-locking above 5000 Hz may be due to measurement errors, and they state that they found statistically significant phase-locking as high as 12 kHz. A later paper[18] by this group of authors does not allude to such high frequencies and flatly states that the upper limit of phase-locking

is about 4 kHz.

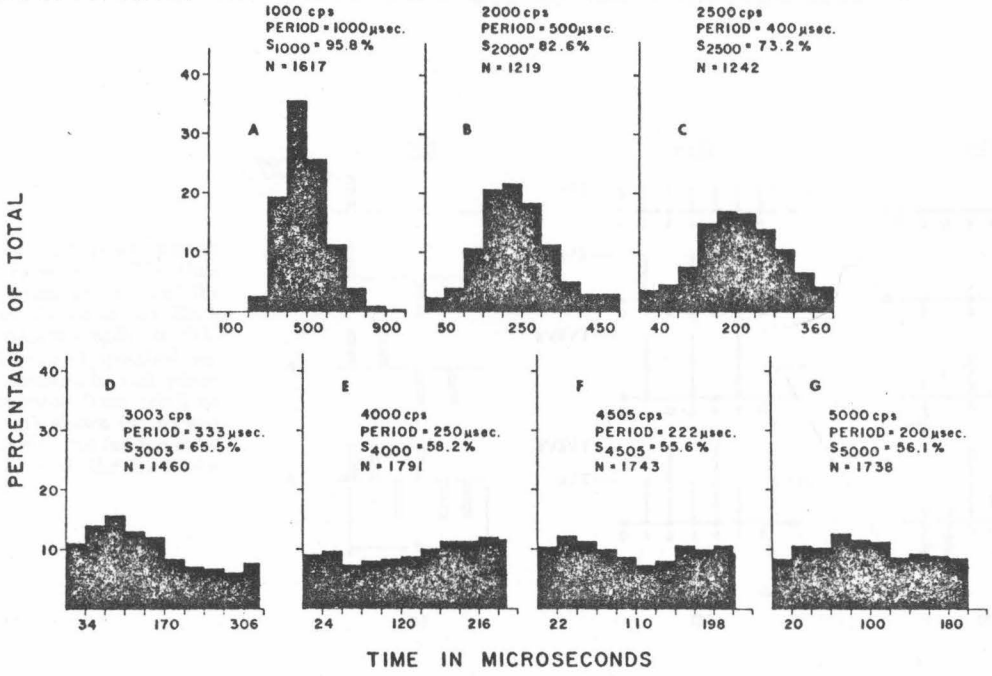



FIGURE 6 (taken from Ref. 17)

## 1.3.3 Experiments with bipolar periodic pulse waveforms

### 1.3.3a Flanagan and Guttman -- initial experiment

In 1960 Flanagan and Guttman[19] reported a unique psychoacoustic experiment which gives evidence for both the time and place theories of pitch. They generated pulse trains which consisted of periodic repetitions of certain patterns, shown in figure 7. The pulses were equally spaced in time, of equal amplitude, and of 100 microseconds duration. The polarity of each pulse was determined by the particular pattern. Subjects were asked to compare the pitches of two pulse trains. The pulse repetition rate of one of the pulse trains was set by the experimenter, the other was under control of the subject. The subject was instructed to match the pitches of the two signals. All six combinations of pairs of the four pulse patterns were used, and data were taken

as a function of the pulse rate set by the experimenter.

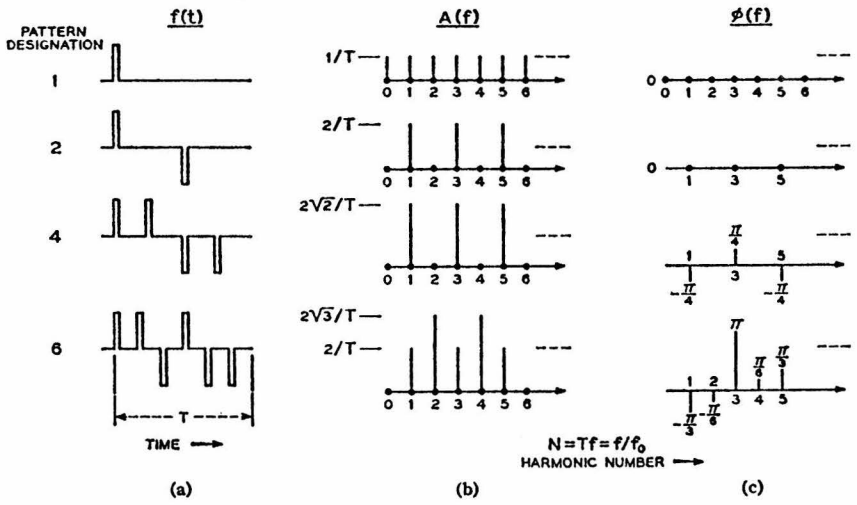



FIG. 1. Periodic pulse stimuli used in the experiment. The time waveforms produced across the earphones are shown in (a). Each pulse has uniform amplitude, 0.1-msec duration and specified polarity. The amplitude and phase spectra, computed from relations in the text, are shown in (b) and (c), respectively. The fundamental period is $T$, and the harmonic number is $n$.

FIGURE 7 (taken from Ref. 19)

For single pulses, positive and negative pulses cannot be distinguished by the ear. Thus at low pulse rates (in the limit, pulses are heard singly) subjects match, as expected, on the pulse rates. If, for instance, the stimulus pattern (rate set by experimenter) is pattern 4 and the response (rate set by subject) is pattern 2, the subject would halve the pulse repetition time, as shown in figure 7, for pattern 2. This would however double the frequency of pattern 2 which is plainly a mismatch in the frequency domain. Flanagan and Guttman found that for high pulse rates subjects matched the signals in the frequency domain (figure 7 shows the patterns matched that way).

The ranges of pulse rates where these two modes occur is of primary interest. In general if the pulse rate was less than

100 pulses per second, the subjects matched on pulse rate. In this case, the fundamental frequency is less than 100 Hz for all patterns and less than 16.7 Hz for pattern 6. Frequency mode matches occurred reliably for fundamental frequencies greater than 200 Hz. This frequency range corresponds to pulse rates of over 200 per second for pattern 1 and 1200 per second for pattern 6. Figures 8a and b show the data for matches between pattern 1 and each of the other three. Figure 8a plots the response fundamental frequency (ordinate) versus the stimulus fundamental. Figure 8b shows the same data but with the pulse repetition rates plotted. Each

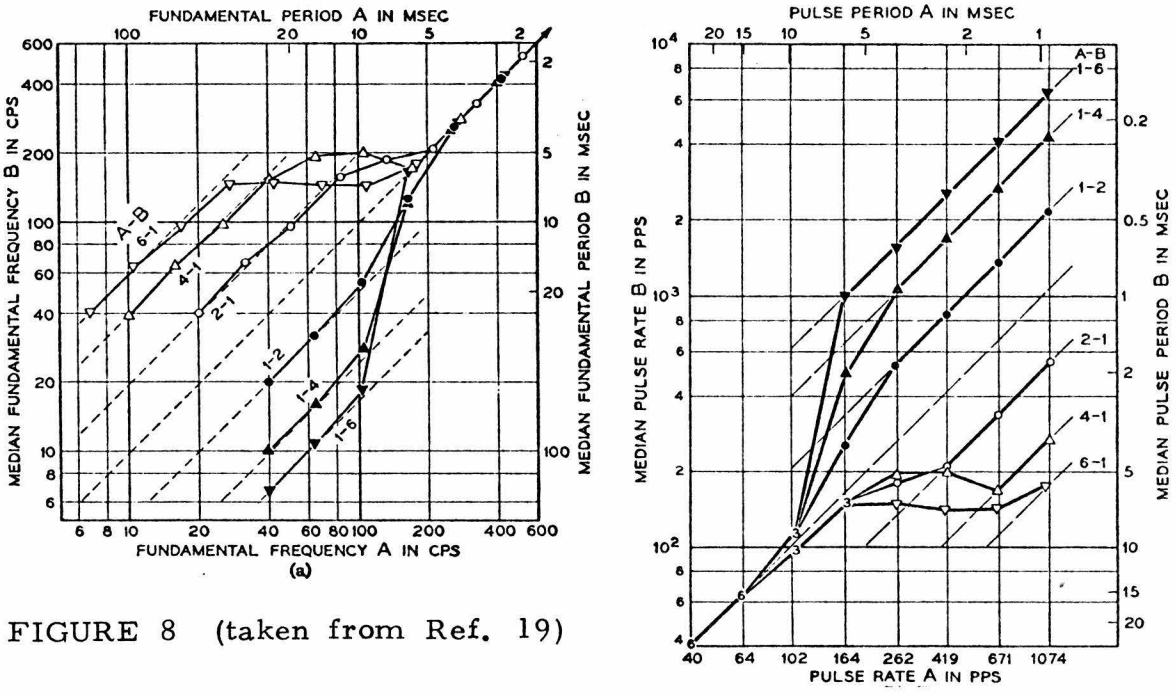



FIGURE 8 (taken from Ref. 19)

point is the median calculated from ten responses, two each from five subjects. The parameter accompanying each curve is the stimulus pattern number followed by the response pattern number. Each curve shows a transition from pulse rate match to frequency match as the stimulus rate is increased. In the transition region,

the subjects' responses are primarily divided between these two modes. Thus the medians calculated from these responses are not entirely meaningful.

1.3.3b Basilar membrane model

In an effort to explain the results of their initial experiment, Flanagan and Guttman applied their pulse patterns to an electrical analog of the basilar membrane. They suggest that pulse rate matches occur when the pulsed nature of the signal is maintained in the time waveform of the membrane displacement along the entire length of the membrane. Their model shows this occurring for pattern 2 at 100 pulses per second or less. At 150 pulses per second, the response waveform at the apical end of the simulated basilar membrane is almost sinusoidal. According to this theory, frequency matches occur if the fundamental frequency component of the signal causes a large displacement at its characteristic place. For pattern 2 at 2000 pulses per second, the analog waveforms show this happening clearly. The response at the 1000 Hz place is sinusoidal, and its magnitude clearly overshadows that at other places. However at pulse rates between these two, the waveforms presented by Flanagan and Guttman show fairly large sinusoidal displacements at the appropriate place along with pulse-like displacements, nearly equal in amplitude, at the basal end. It would seem from their psychoacoustic results and their basilar membrane model that even a small indication of place localization on the basilar membrane overshadows the conflicting time information at the basal end.

1.3.3c Effect of removal of low-order harmonics

In a further experiment reported in a companion paper,[20] Flanagan and Guttman filtered the lower spectral components out. Schouten's residue pitch theory predicts that this should have no effect, i.e., the higher components should produce a residue pitch at the fundamental. In fact, filtering produced a third mode of pitch match where the lowest spectral components of the two signals were equated.

To perform these experiments, a high-pass filter was used on the stimulus signal. The filter was adjusted to pass either the second or third spectral component and, of course, all higher components. At pulse rates below 100 per second, subjects matched according to pulse rates as in the previous experiment. Of course, in this region, the filter was set quite low, and one would expect little effect. Only patterns 1, 2 and 4 were used in these experiments, and pulse rates below 100 per second resulted in the following filter settings: 1) pattern 1 filtered at second harmonic, less than 200 Hz; 2) pattern 1 third harmonic, less than 300 Hz; 3) pattern 2 third harmonic, less than 150 Hz; and 4) pattern 4 third harmonic, less than 75 Hz. Recall that patterns 2 and 4 have no energy at the second harmonic. As before, fundamental frequencies above 200 Hz brought frequency mode matches, and the same sort of transition region exists between these two regions. The filtering manifests itself at fundamental frequencies greater than about 500 Hz. Above 1000 Hz, the signals are matched on the basis of the lowest spectral component present. Between these

two fundamental frequencies is a transition region similar to the one between pulse rate and fundamental mode matches. That is, the data show a bimodal characteristic where subjects seem to match either on the fundamental or the lowest component, but choosing randomly between these from trial to trial.

1.3.3d Effect of fixed high-pass filter

In 1964 Flanagan and Guttman reported on an extension[21] of their previous experiments. They high-pass filtered the stimulus signal at fixed high frequencies. Two and four kilohertz were the frequencies used. They reasoned that this procedure would eliminate the response at the low frequency end of the basilar membrane where the place information is strongest and most easily separable. This would result in pulse rate matches for higher stimulus pulse rates than before. Their data indicate that pulse rate matches occurred up to about 500 Hz as opposed to 100 Hz previously. Their data were somewhat obscured by another mode of pitch match which they ascribed to the difference frequency. The difference frequency is the spacing of the spectral components of a signal. For pulse pattern 1, the difference frequency is the fundamental frequency; for patterns 2 and 4, it is twice the fundamental. The high-pass filter combines with the low-pass frequency characteristic of the middle ear to form a bandpass filter. When the signal pulse rate is such that a small number of spectral components fall in the passband of this filter, the resulting signal into the cochlea develops a periodic envelope whose frequency is that of the difference frequency. This may then elicit a pitch

corresponding to the envelope periodicity as found by Schouten.

1.3.3e Effect of selective masking

Another experimenter, Rosenberg, who worked with Flanagan and Guttman, used filtered white noise to mask various places on the basilar membrane.[22] For signals in the transition region between pulse rate and fundamental mode pitch, high- or low-pass noise can mask one mode and force the pitch judgment the other way. This also creates a new transition region either higher or lower than the one occurring without masking. As expected, low-pass noise masks the fundamental mode, while high-pass noise masks the pulse rate mode. Rosenberg found that the cutoff frequency of 1000 Hz was rather critical for this masking effect. Increasing the bandwidth of the noise (that is, increasing the cut-off for low-pass noise or decreasing it for high-pass noise) tended to mask the entire signal. Decreasing the noise bandwidth reduced the selective mode effect.

In informal listening experiments, Rosenberg found that the places on the basilar membrane corresponding to 500 Hz and 5000 Hz were the most critical with respect to fundamental-mode pitch and pulse-rate-mode pitch, respectively. He determined this by using narrow band noise centered at various frequencies. This is of course only a rule of thumb and does not mean that these two places are the only ones capable of supporting the two pitch modes. The upper limit Rosenberg found for pulse mode pitch perception was about 300 pulses per second.

### 1.3.4 Other evidence for use of time information in the auditory system

There have been many other psychoacoustic experiments which have attempted to elicit a sensation of pitch from the time domain properties of a signal.[23, 24] When these signals have no obvious spectral characteristics corresponding to the pitch frequency, this is regarded as proof that the brain can assign a pitch to a sound on the basis of time information alone. Of course for the set of physically realizable signals, there is a one-to-one correspondence between the time domain and frequency domain. The question in pitch perception is which of these two representations, or what combination, does the brain use in assigning a pitch to a sound. When considering the frequency domain, the resolution properties of the basilar membrane must be kept in mind. The ear is generally insensitive to phase variations over the audio spectrum, and it is reasonable that the brain would ignore them since the basilar membrane induces larger time delays for lower frequencies. However, as some of Schouten's experiments have shown, when the phase of a periodic signal with only a few components is varied, the time domain signal becomes more or less impulsive, and these variations are clearly audible.

The auditory system uses time information in other ways as demonstrated by binaural experiments. When different signals are presented to the two ears, several different effects have been noted. Sound sources in a natural environment can be localized. When a sound source is to the right, the sound arrives in the right ear sooner and is louder than that in the left ear. For sine-waves

heard through stereo headphones, the time or phase difference predominates for frequencies below about 1500 Hz, while the amplitude difference predominates above 1500 Hz.[25] In the phenomena of binaural unmasking, if a sine-wave is masked by white thermal noise in one ear and then the same noise alone is presented to the other ear, the sine-wave will become clearly audible.[26]

The preceding material should give the reader an understanding of the history and present state of the theory of pitch perception. The following chapters describe the current experiment undertaken to explore the upper limits of pulse rate mode pitch perception.

# CHAPTER II THE EXPERIMENT: APPARATUS AND PROCEDURES

## 2.1 Introduction

The Flanagan-Guttman experiments described previously show clearly that the auditory system is capable of using time information to assign pitch. When many neurons are firing periodically, pitch can be assigned to a frequency corresponding to the inverse period; this assignment is independent of the place on the basilar membrane where the activity originates. Their first experiments gave an upper limit of about 100 pulses per second for this mode. Subsequent experiments that included high-pass filtering or filtered masking noise pushed this limit up to about 300 pulses per second. Above the limit, the pitch of the signal becomes ambiguous and subjects seem to have two choices for the pitch of the signal. This ambiguity appears to be related to the emergence of spatially localized activity on the basilar membrane. Thus a place mechanism of pitch begins to compete with the time mechanism.

The experiments described here have been designed to explore the upper frequency limits of pitch perception derived from time information. The single unit neurophysiological experiments indicate that time synchrony is present up to about 4000 Hz, at least in animals. Furthermore, localization in humans shows that time information is present in the auditory system up to about 1500 Hz.

## 2.2 Initial experiment

### 2.2.1 Objective

To test for a time based pitch, pulse trains with a uniform spacing in time have been used. It is necessary to avoid competition with a frequency or place-based pitch mechanism. This has been done by choosing the pulse polarity randomly. Such a signal has a flat power spectrum as shown in Appendix A. This random choice was obtained from a feedback shift register and is not truly random, hence the term pseudo-random (see Appendix B).

### 2.2.2 Apparatus

A signal generator was constructed with a feedback shift resister which implemented a fixed maximum length polynomial of length 35. The shift register drove a positive/negative pulse generator which produced 10 microsecond pulses. At that time, the length 35 polynomial was the longest known and provided a ridiculously long repetition time at the pulse rates used. Appendix E describes the operation of the circuit. The positive and negative pulse durations were controlled by two different one-shot multivibrators whose durations were adjusted to be equal by viewing the waveforms on an oscilloscope. Thus the adjustment could be made with an accuracy of about 2%. The amplitude of the negative going pulse was fixed by a resistor network, while the amplitude of the positive pulse was adjustable. The pulse amplitudes were balanced by oscilloscope to an accuracy of about 5%.

### 2.2.3 Procedure

The experiments were performed using the following

experimental procedure. The subject sat in a double wall sound booth (Industrial Acoustics Company Model 1202-A, Sound Controlled Research Room Type 5) wearing a pair of headphones (Koss PRO-4A) driven by one channel of a stereo audio amplifier (Sony TA-1055); the other channel was unused. All signals were presented dioticially with both headphones connected in parallel at the amplifier output. The subject controlled the input to the amplifier by means of two spring return switches: one connected to the noise generator, the other connected to a unipolar pulse generator. Outside the booth, the experimenter controlled the clock oscillator which drove the noise generator and could also measure the frequency of either the oscillator or the pulse generator inside the booth. The amplitude and duration of the unipolar pulses of the subject's pulse generator were the same as the amplitude and duration of the pulses of the noise generator. Prior to each experimental session, the clock frequency of the noise generator was set to 1 kHz and a 30 db attenuator was inserted between the noise generator and the amplifier. Then the subject was instructed to adjust the gain of the amplifier so that the noise was just audible. To assist this judgment the subject was allowed to switch the noise on and off as he pleased. When the attenuator was removed for the experiment, both the noise generator and the pulse generator were nominally at 30 db above sensation level.

The subject was instructed to listen to the noise generator, to listen to the pulse generator, and to adjust the frequency of the pulse generator to match the pitch of the noise. The subject

was permitted to switch back and forth between the two signals as often as necessary. The experimenter recorded the frequencies of both the clock oscillator and the subject's pulse generator when the subject signalled that he had finished. In the early experiments, the subject was also asked to adjust the amplitude of the pulse generator in the booth to match the loudness of the "tonal part" of the noise signal. It was expected that the perceptual quality of the signal would shift from a tonal perception to a (white) noise-like perception as the pulse rate increased, just as the Flanagan-Guttman signals had done between tonal perceptions based on different pitches. It was also expected that the amplitude data would quantify the disappearance of the time-based pitch perception as the clock rate was increased. Unfortunately, the subjects found it difficult to identify the loudness of the vaguely defined "tonal part" of the signal. In fact, many subjects never touched the amplitude control, and therefore the amplitude match was dropped.

Using this apparatus, 14 experiments were performed in September and October of 1974. Some subjects, JLG in particular, were able to match the pitch exceedingly well. JLG was consistently able to match at the highest clock rate used, 8 kHz.

2.3 Problems encountered

This consistent matching raised questions about the use of pseudo-random noise as opposed to genuinely random pulses. Because the noise generator can produce runs of pulses of the same polarity of any length up to the length of the shift register, i.e., 35, it was thought that these runs produced short bursts of

apparently pure tone imbedded in a background of white noise.

An additional and also related problem occurs when a train of random polarity pulses is applied to the input of a narrow bandpass filter. The probability distribution of the amplitude of the output varies with the ratio of pulse repetition rate to filter center frequency (see Appendix D). This variation may provide a clue to the pitch of the noise signal. If we model the ear as a bank of bandpass filters of varying center frequency whose outputs are processed by later neural systems, it may be possible for the central nervous system to pick out those filters whose amplitude distributions are peculiar, i.e., whose center frequency is integrally related to the pulse repetition rate.

It is possible to reduce and change the variability of the amplitude distribution by replacing each random polarity pulse with a pattern of N pulses (see Appendices B and D). Instead of choosing plus or minus every T seconds, we choose one of K patterns every T/N seconds, thereby maintaining a pulse repetition time of T in both cases. With the appropriate set of patterns and k-ary probability space, such a signal still has a flat power spectrum. Note also that if the set of patterns does not include a pattern of unipolar pulses, then this procedure eliminates long runs. A maximal length run of plusses occurs when the pattern with the greatest number of trailing plusses is followed immediately by the pattern with the greatest number of leading plusses. This argument is also true for runs of minuses.

This pulse pattern scheme was implemented in November

of 1974 with specially built hardware, which used four patterns of length four. The results of the experiment on JLG were substantially the same; he consistently matched at very high pulse repetition rates. Therefore, patterns of greater length were used. To accomplish this, an interface was built and attached to a PDP 11/45 minicomputer (see Appendix E). With the construction of the computer interface, the pulse generation circuit was improved.

Note that the analysis in Appendix A assumes pulses, h(t), occurring exactly T seconds apart and, implicitly, positive and negative pulses having the same shapes. If h(t) is the waveform of the positive pulses, and h'(t) is that of the negative pulses, then let the difference pulse be g(t), such that

$$g(t) = h(t) - h'(t)$$

where we assume all waveforms are time limited. The true output, $\underset{\sim}{r}(t)$, can be regarded as the sum of the ideal random process and a train of difference pulses, all of the same polarity but occurring at random times.

$$\underset{\sim}{r}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{a}_n h(t-nT) + \sum_{n=-\infty}^{\infty} \underset{\sim}{b}_n g(t-nT) \quad \text{where } a_n = 1 \rightarrow b_n = 0;\ a_n = -1 \rightarrow b_n = 1$$

$$= \underset{\sim}{y}(t) + \underset{\sim}{y}'(t)$$

If we apply the analysis in Appendix A to the function y'(t), we find that its autocorrelation function is

$$R'_{y'y'}(t+\tau, t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} E[\underset{\sim}{b}_n \underset{\sim}{b}_m]\, g(t+\tau - nT)\, g(t-mT)$$

However,

$$E[\underset{\sim}{b}_n \underset{\sim}{b}_m] = \frac{1}{2} \text{ if } m = n \text{ or } \frac{1}{4} \text{ if } m \neq n$$

$$E[\underset{\sim}{b}_n \underset{\sim}{b}_m] = \frac{1}{4}(1 + \delta_{mn})$$

Thus $R_{y'y'}$ is not limited in the $\tau$ dimension as is $R_{yy}$ of Appendix A. So when $R_{y'y'}$ is integrated over t,

$$R'(\tau) = \lim_{L \to \infty} \frac{1}{2L} \int_{-L}^{L} R_{y'y'}(t+\tau, t)dt = \frac{1}{4} A(\tau) + \frac{1}{4} \sum_{n=-\infty}^{\infty} A(\tau - nT)$$

The result is a periodic function of $\tau$. The Fourier transform of $R'(\tau)$ gives the power spectrum of y'(t). The result is that the spectrum of r(t) is the sum of a flat spectrum, due to y(t), and a line spectrum, due to y'(t).

To estimate the power in the line spectrum part of the signal, we let h(t) be a pulse of height V and duration D. Let h'(t) have height $(1-\Delta_D)V$ and duration $(1-\Delta_D)D$, then g(t) is as shown in figure 9. The energy in g(t) is

$$E_g = V^2\Delta_V^2(1-\Delta_D)D + V^2\Delta_D = V^2D(\Delta_V^2+\Delta_D-\Delta_V^2\Delta_D)$$

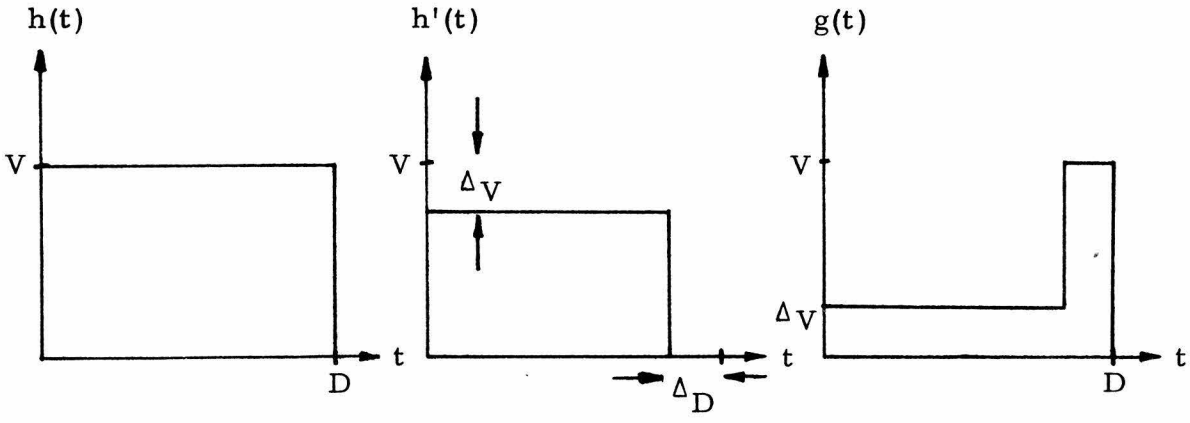



FIGURE 9

The power in y'(t) is 1/2 $E_g/T$, because a pulse g(t) occurs with probability 1/2 every T seconds. As noted above, $\Delta_V$ was approximately 2.25% of that in y(t), or 16 db down. Further asymmetry

between the positive and negative pulses is produced in the circuitry used in the first experiments. The pulses were produced by a 741 operational amplifier driven by TTL logic integrated circuits. The op amp cannot respond to the very fast rise and fall times of the TTL signals, and therefore the rise and fall times of the pulses produced are determined by the slew rate of the op amp. However, solid state amplifiers have inherently different slew rates for positive going and negative going signals.

The pulse generator used in the experiments is described in detail in Appendix E. It was designed with a common time base for both positive and negative pulses, and the differential delays were held to less than 10 nanoseconds for a $\Delta_D$ of .001. Amplitude balancing was accomplished by turning both positive and negative current generators on simultaneously and adjusting for a null output. This resulted in a $\Delta_V$ of about .005. Thus the power in y'(t) was about 30 db below the total power in the noise signal. The slew rate problem of the 741 operational amplifier also existed with the audio amplifier used to drive the headphones. It was therefore unwise to use pulses with sharp rise and fall times. To solve this problem, a passive filter in the form of an RLC low-pass circuit was used to smooth the pulses. An output attenuator was included in the interface to allow the output level to be accurately changed in 5 db steps.

In July and August 1975, a number of experiments were performed with length 12 patterns. The experimental setup and procedure were the same as in the previous experiments. The

results showed an increased tendency to match to lower harmonics but still did not show a fall-off of ability to match at high pulse repetition rates.

2.4 Use of fixed high-pass filter

As was mentioned in Chapter I, follow on work to the Flanagan-Guttman experiments[21] used a high-pass filter to confine the stimulus to the basal end of the basilar membrane. The same idea was tried in these experiments. The flatness of the noise spectrum is due to the time integration of the signal over very long time intervals. Plainly the ear does not integrate over such long time intervals. Short runs of pulses, either unipolar or perhaps patterns with periods of two or three pulses, might produce temporary displacement patterns on the basilar membrane which would give localized peaks. Thus a place mechanism might be operating to give a pitch to the pseudo-random noise signals. Therefore elimination of the response at the apical end should reduce this localization effect.

Toward this end, a high-pass filter was constructed with selectable cutoff frequencies of 4, 8, 12, and 16 kHz. A 4 pole Chebyshev design was used. In order to mask any low level leakage at low frequencies, a Gaussian noise signal from a General Radio 1390 noise generator was added directly at the headphones.

When the high-pass filter was used, the experimental procedure was modified as follows. First the 30 db attenuator was switched in after the high-pass filter. With the pseudo-random noise at a pulse repetition rate of 1 kHz, the subject was asked to

adjust the audio amplifier gain so that the signal was just audible. Later, when the attenuator was switched out, the signal was left at 30 db sensation level. Then with no signal into the amplifier, the subject was asked to adjust the level of the noise generator so that its output was just audible. The voltage at the headphones was recorded, and the noise generator level was increased 20 db. The pseudo-noise signal and the matching signal were observed to be clearly audible over the noise, even when filtered. The experimental procedure then continued as before.

The 4-kHz filter seemed to have little effect on any subject's ability to match. The 12- and 16-kHz filters so reduced the level of the signal as to make it difficult to proceed. Thus the bulk of the data was taken with the 8 kHz filter setting.

Experiments were performed with this final setup from October 1975 to May 1976 with five subjects participating regularly. Chapter III will present the data taken from all experiments in detail.

# CHAPTER III: DISCUSSION OF DATA

## 3.1 Definition of terms

The principal data taken in these experiments consist of pairs of frequencies: the pulse repetition rate of the pseudo-random noise (hereinafter called the clock), set by the experimenter; and the pulse repetition rate of the unipolar pulse generator (called the response), set by the subject. An experiment is defined as the data taken in one session, usually lasting about an hour. A point is defined as a single pair of pulse repetition rates (clock, response). The number of points taken in an experiment ranged from about 10 to 35, depending on the ability and enthusiasm of the subject. A type number was assigned to each experiment, as shown in table 1.

| TYPE | SIGNAL LEVEL | FILTER | PN LENGTH | PATTERN LENGTH |
|---|---|---|---|---|
| 0 | 30 db | 8 kHz | 16 | 1 bit |
| 1 | 30 | none | 16 or 35 | 1 |
| 2 | 30 | none | 35 | 4 |
| 3 | 30 | none | 35 | 12 |
| 4 | 20 | none | 16 | 1 |
| 5 | 10 | none | 16 | 1 |

TABLE 1: EXPERIMENT TYPE NUMBERS

The most natural way to present the data is as Flanagan and Guttman did (see Chapter I, figure 8). That is, as points in the plane with clock as abscissa and response as ordinate. In this way, successful matches would lie on the line $x = y$. However, in order to increase the resolution for a given size of graph paper, clock/response versus clock rate is plotted. Thus successful matches lie on the line $y = 1$. Figures 10 through 30 show the raw data on log-log axes. Each plot shows all of the points (from all

of the experiments) for a particular subject and experiment type.

3.2 Choice of clock frequencies

The early experiments used a fixed list of 13 clock frequencies logarithmically spaced. There were three different random orders for the frequencies. The experimenter randomly selected a list before each experiment. This procedure prevented the subject from guessing the next clock frequency. In subsequent experiments, the experimenter set the clock generator at (mentally) randomly chosen frequencies. The experimenter attempted to avoid constantly increasing or decreasing the frequencies regularly and also to "fill-in" the accumulated data for that subject type, thereby making each point an independent trial. The author believes this to be the case, both from his experience as a subject and from discussions with other subjects.

3.3 Discussion of raw data

It was expected that subjects would be able to match response to clock for frequencies up to at least 400 Hz and perhaps somewhat beyond. However for higher frequencies, say 2 to 3 kHz, it was expected that the auditory system would be unable to maintain a pitch perception based on the impulsive nature of the time waveform. In the frequency domain, there is no clue to the pulse repetition rate; therefore, we expected random guesses as responses and expected the plots to scatter. This did not happen, at least for some subjects, as can be seen from the raw data. For instance JLG type 1 (figure 17) matched consistently up to about 5.5 kHz. (It was his data that first indicated an ability to match

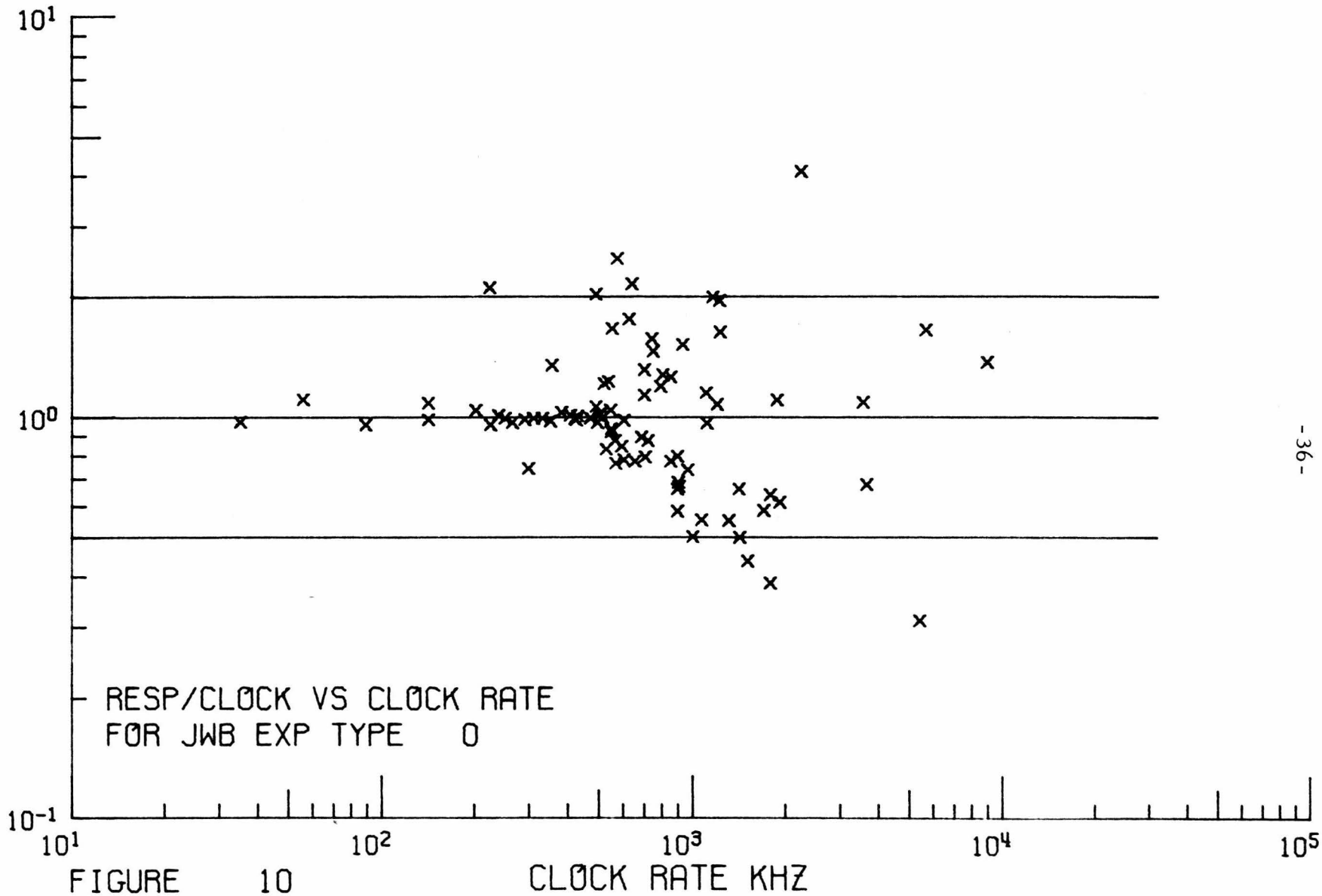

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JWB EXP TYPE 0
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 10

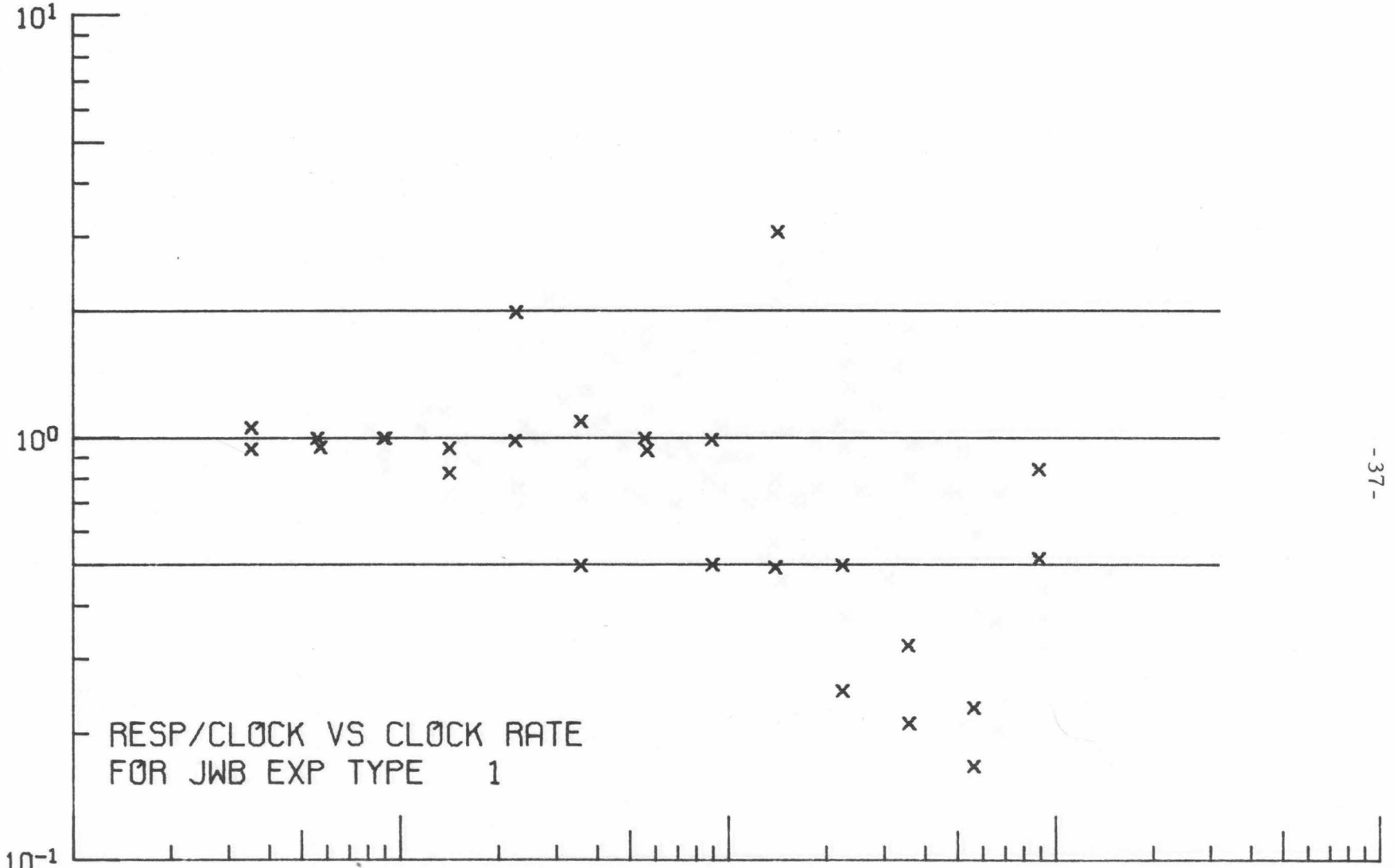

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JWB EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 11

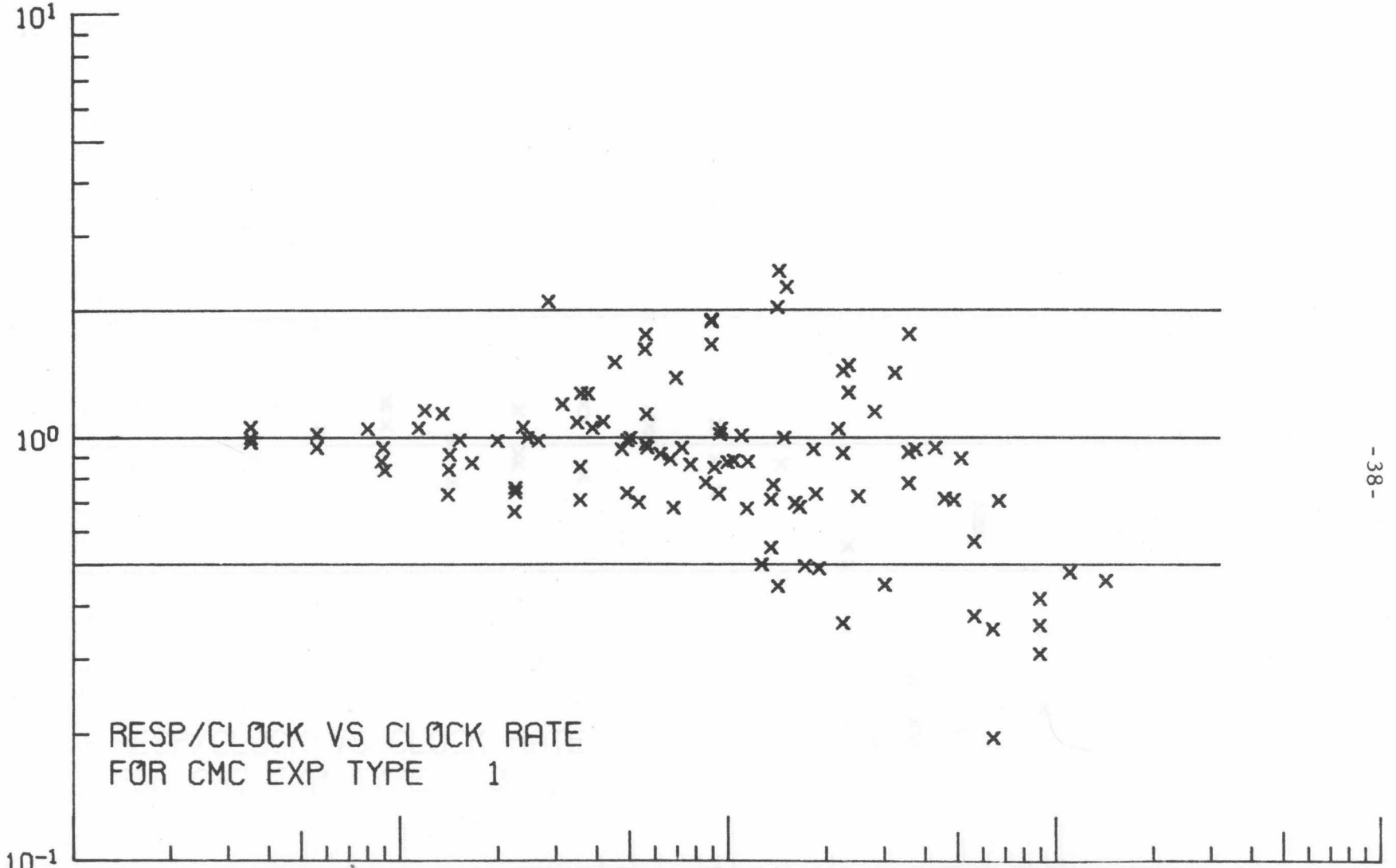

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR CMC EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 12

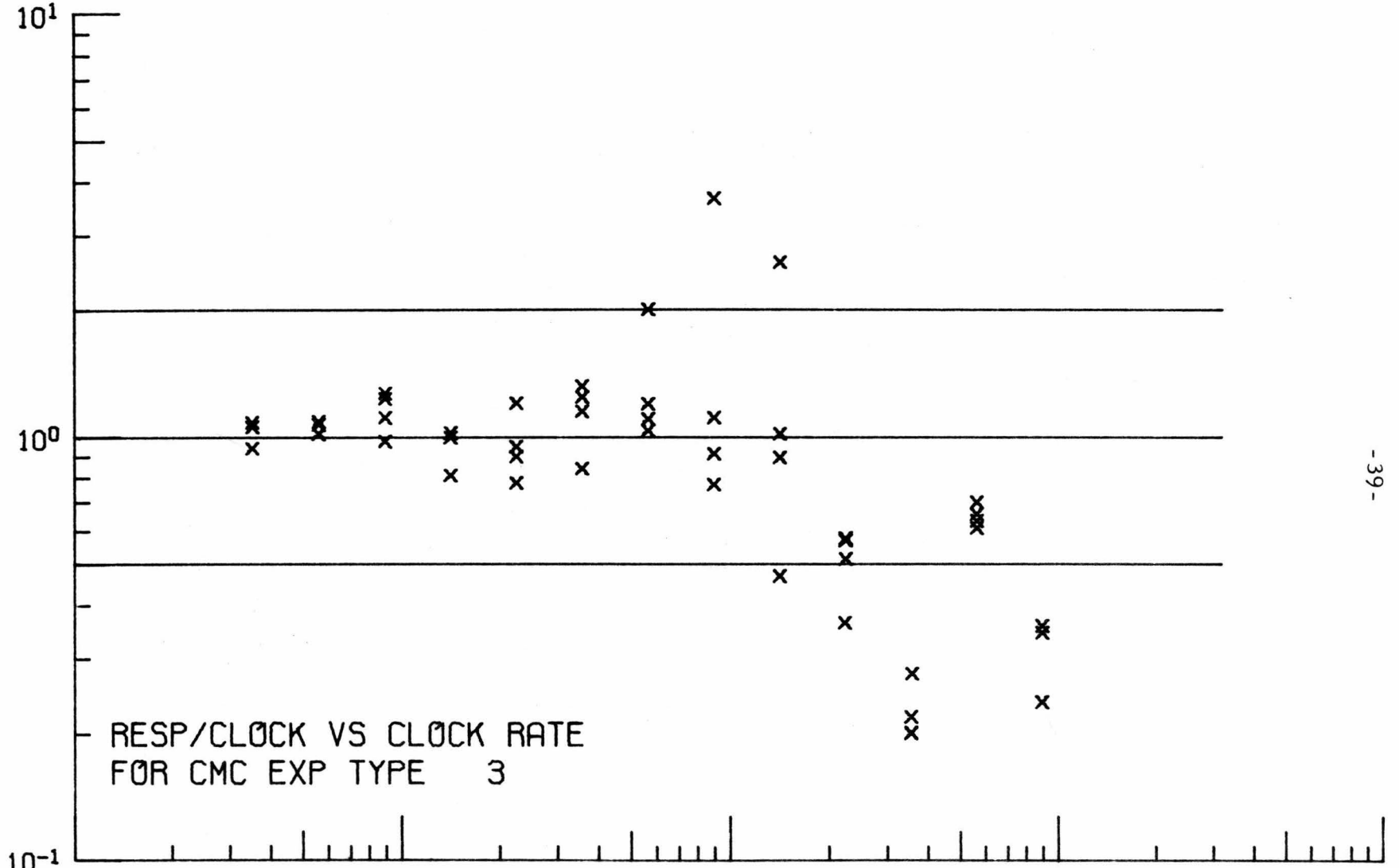

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR CMC EXP TYPE 3
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 13

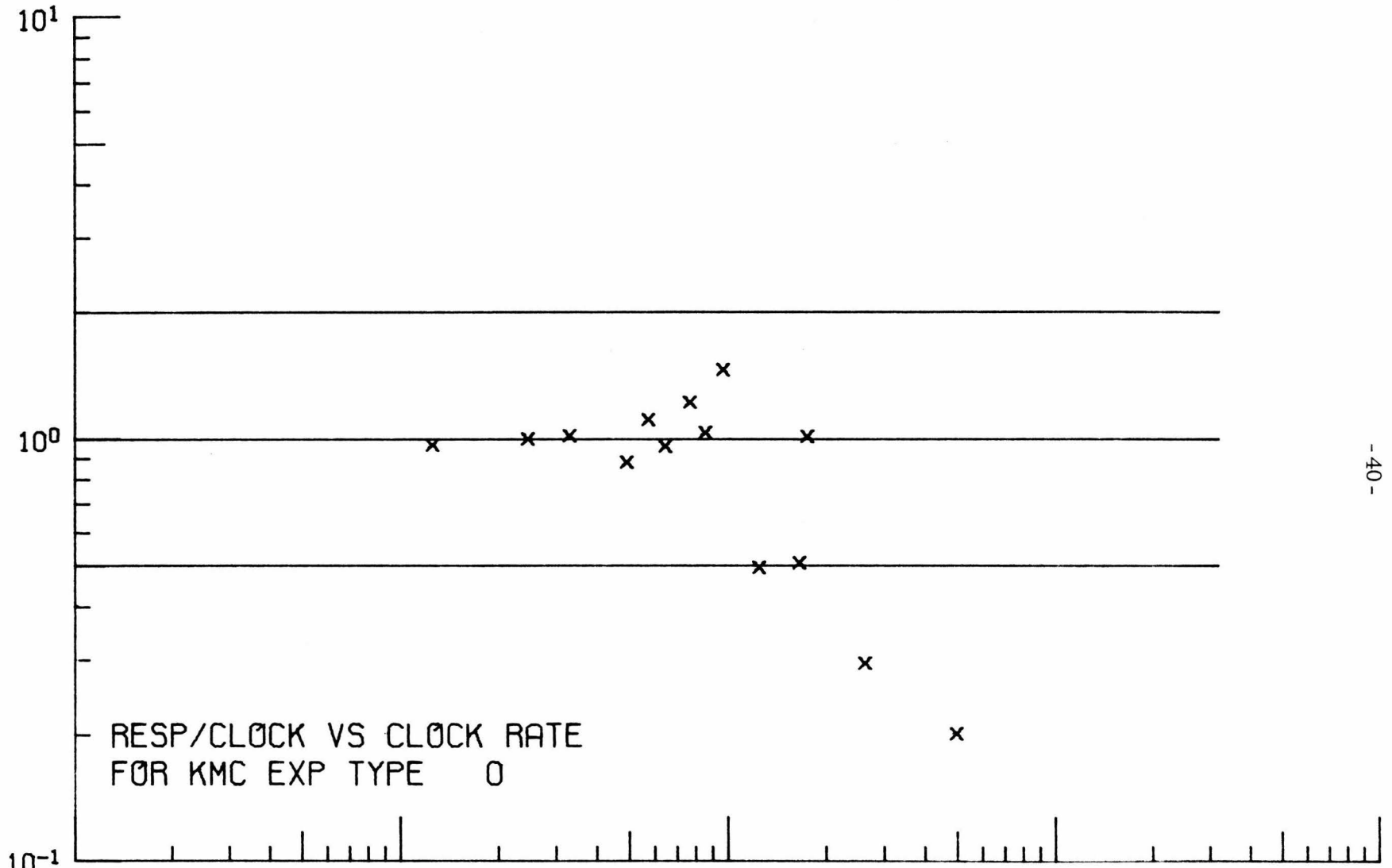

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR KMC EXP TYPE 0
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 14

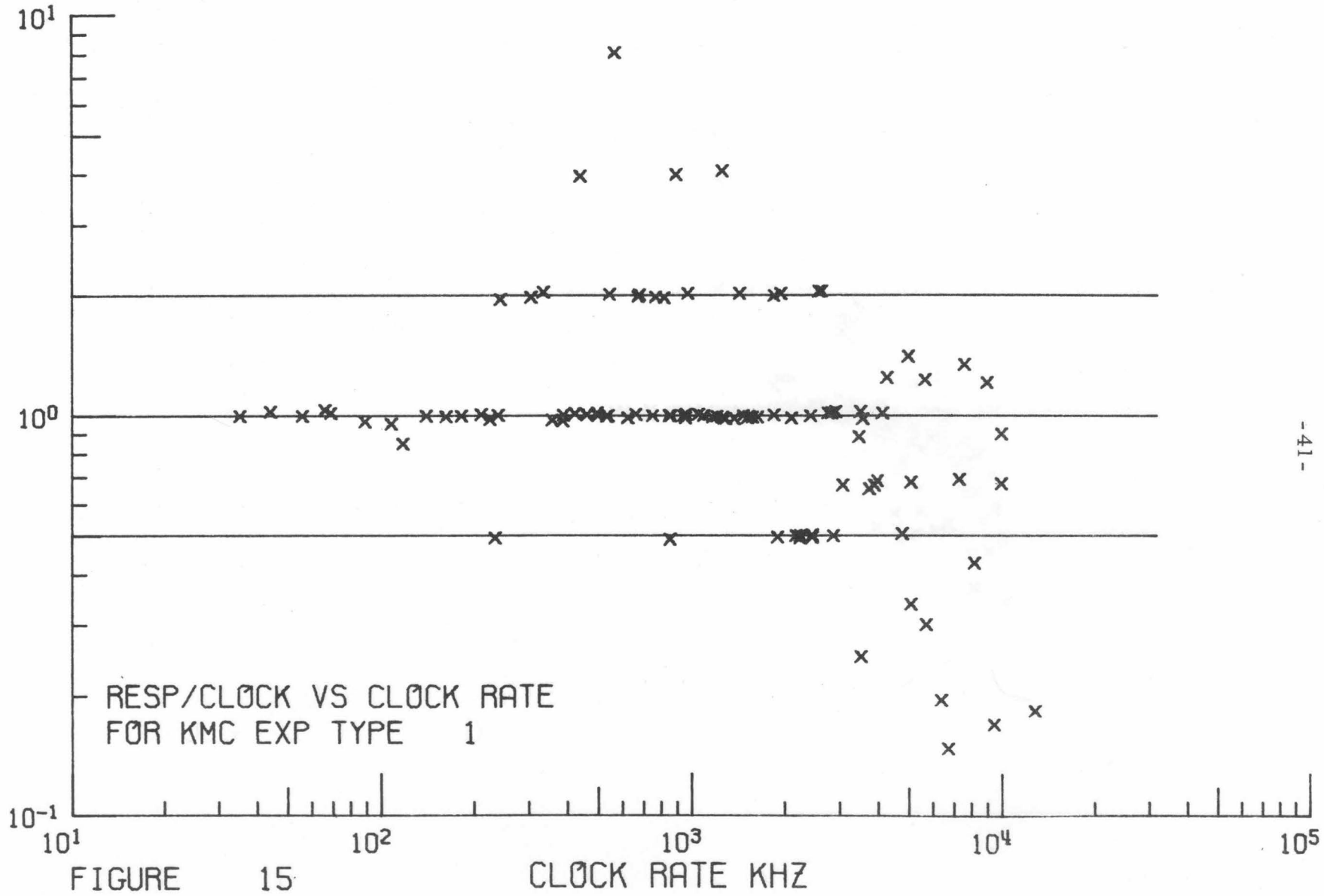

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR KMC EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 15

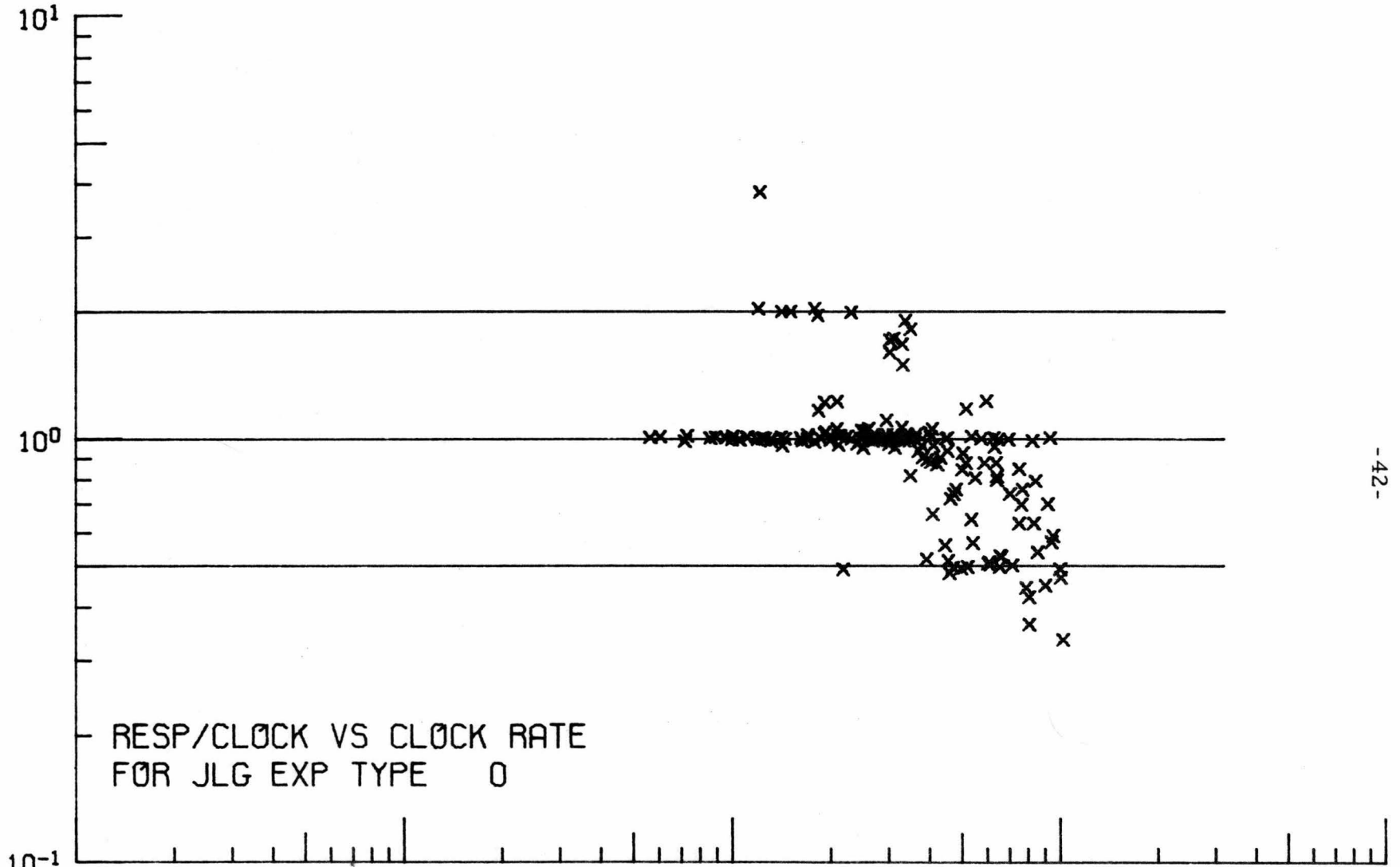

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 0
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 16

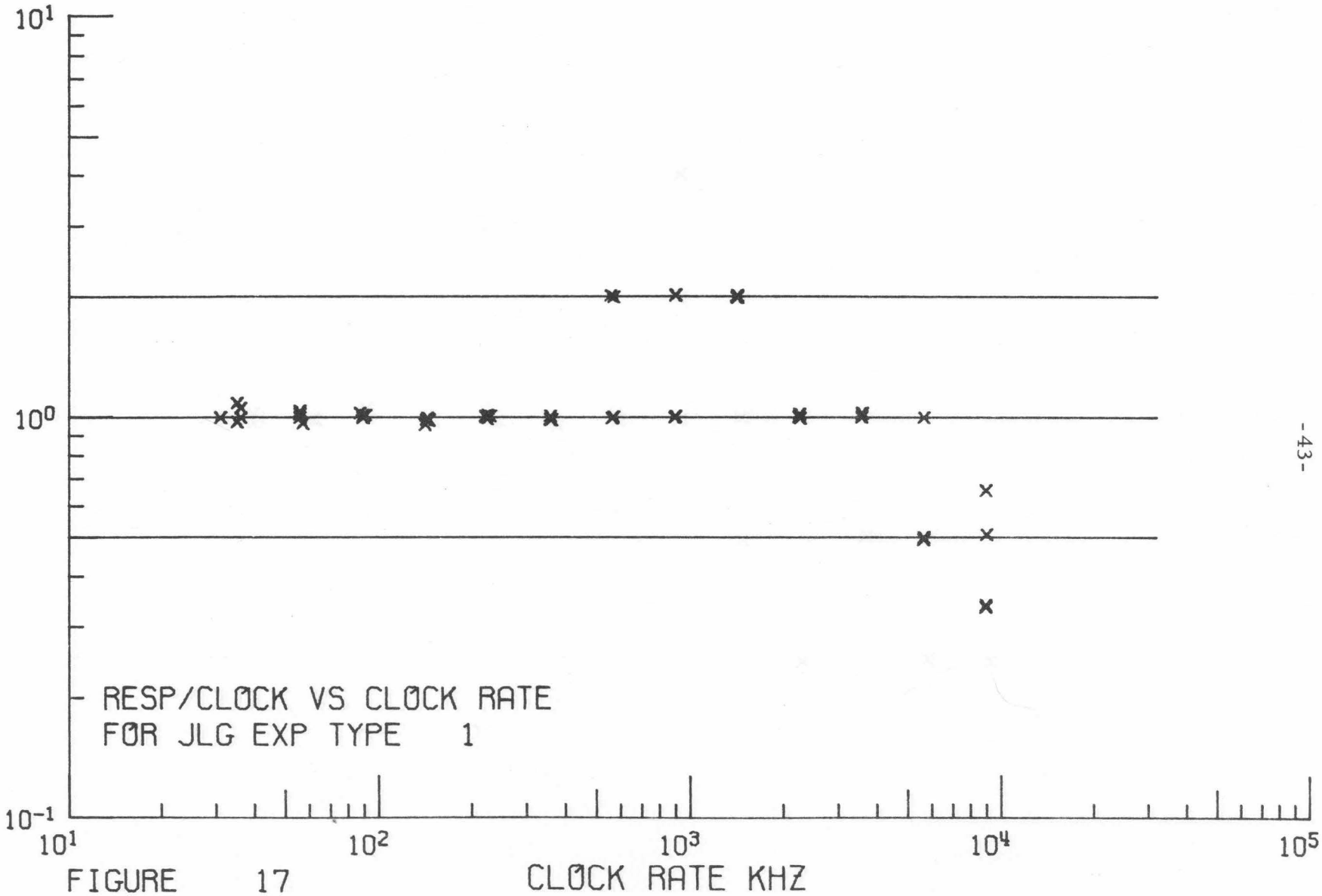

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 17

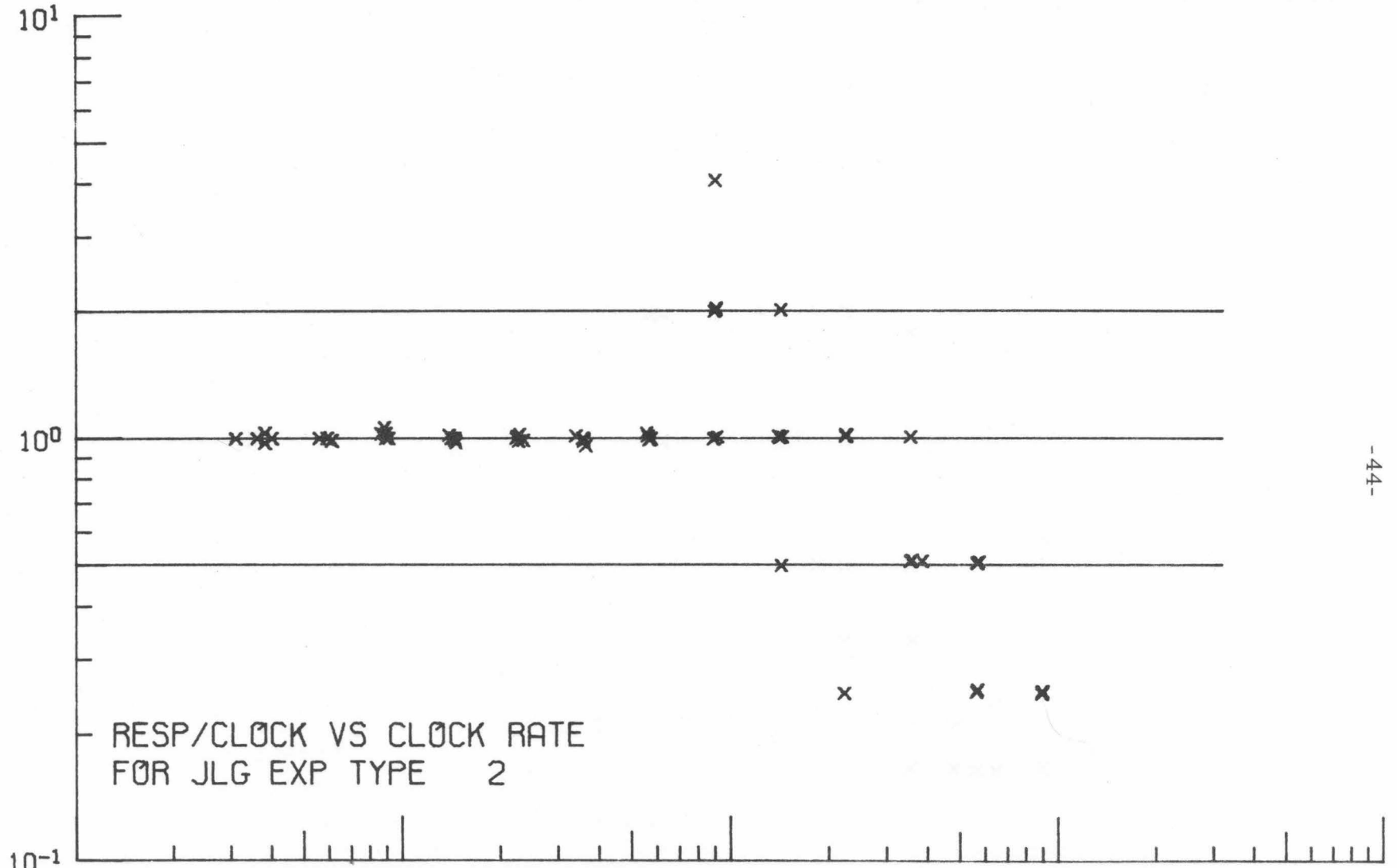

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 2
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 18

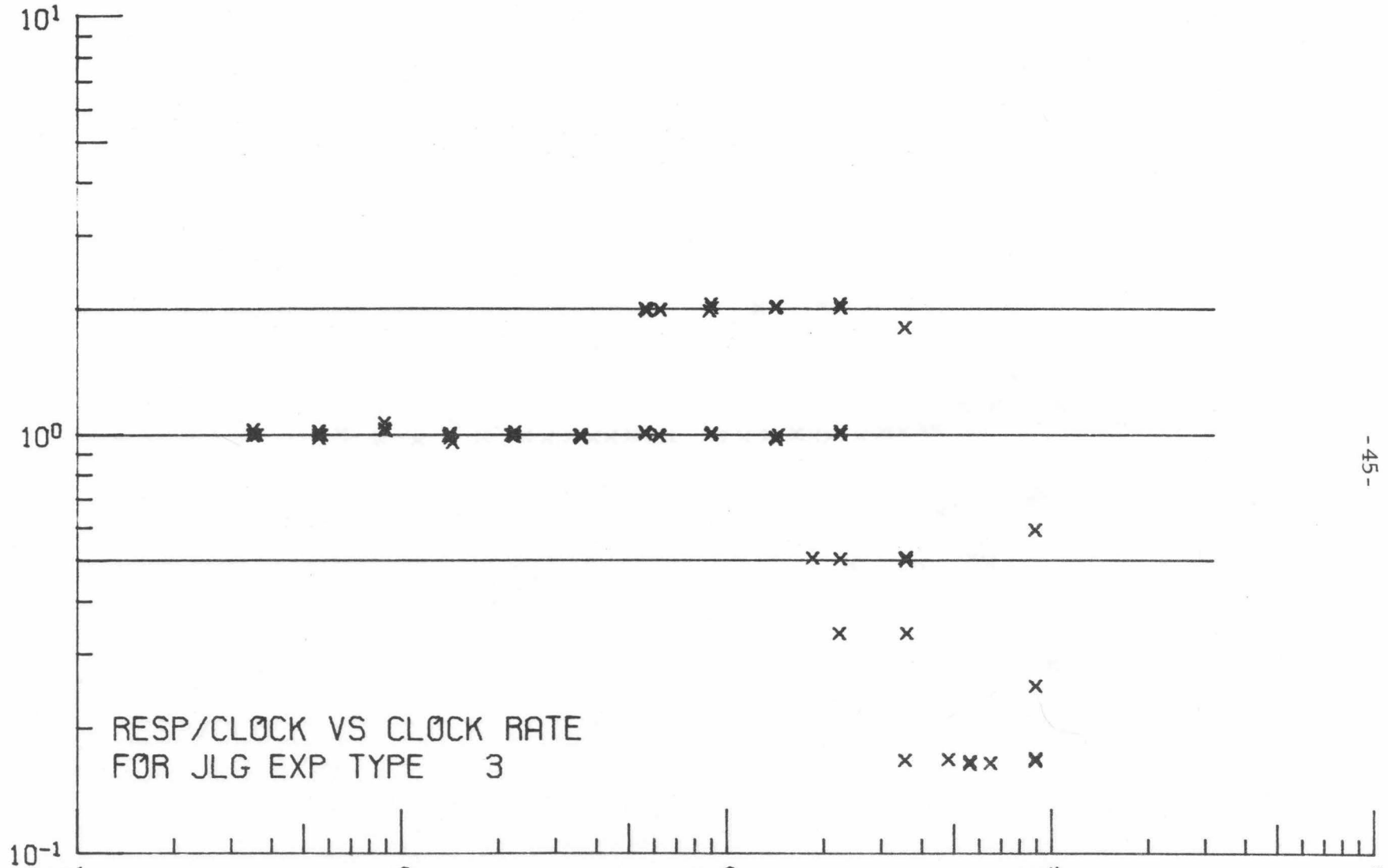

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 3
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 19

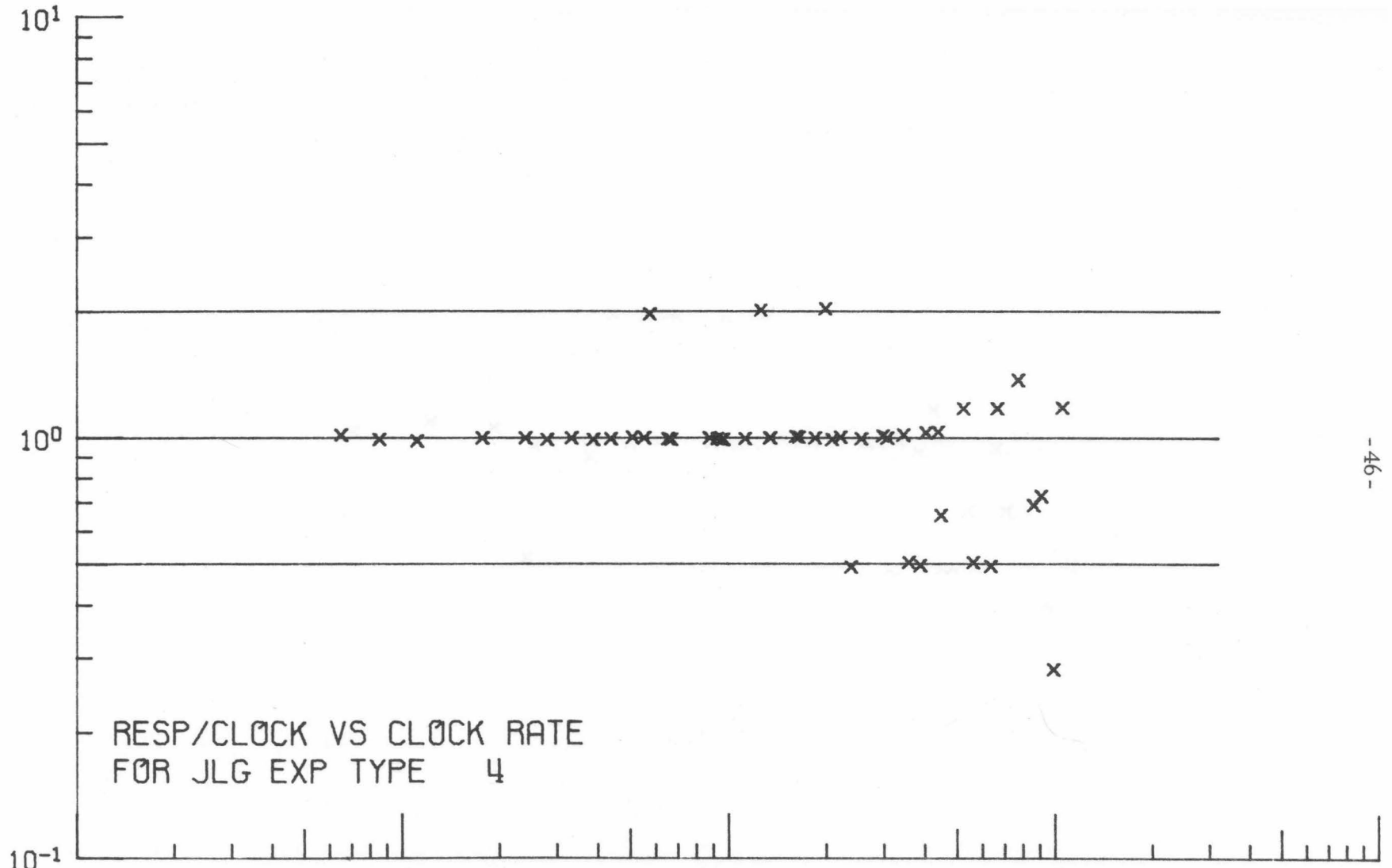

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 4
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 20

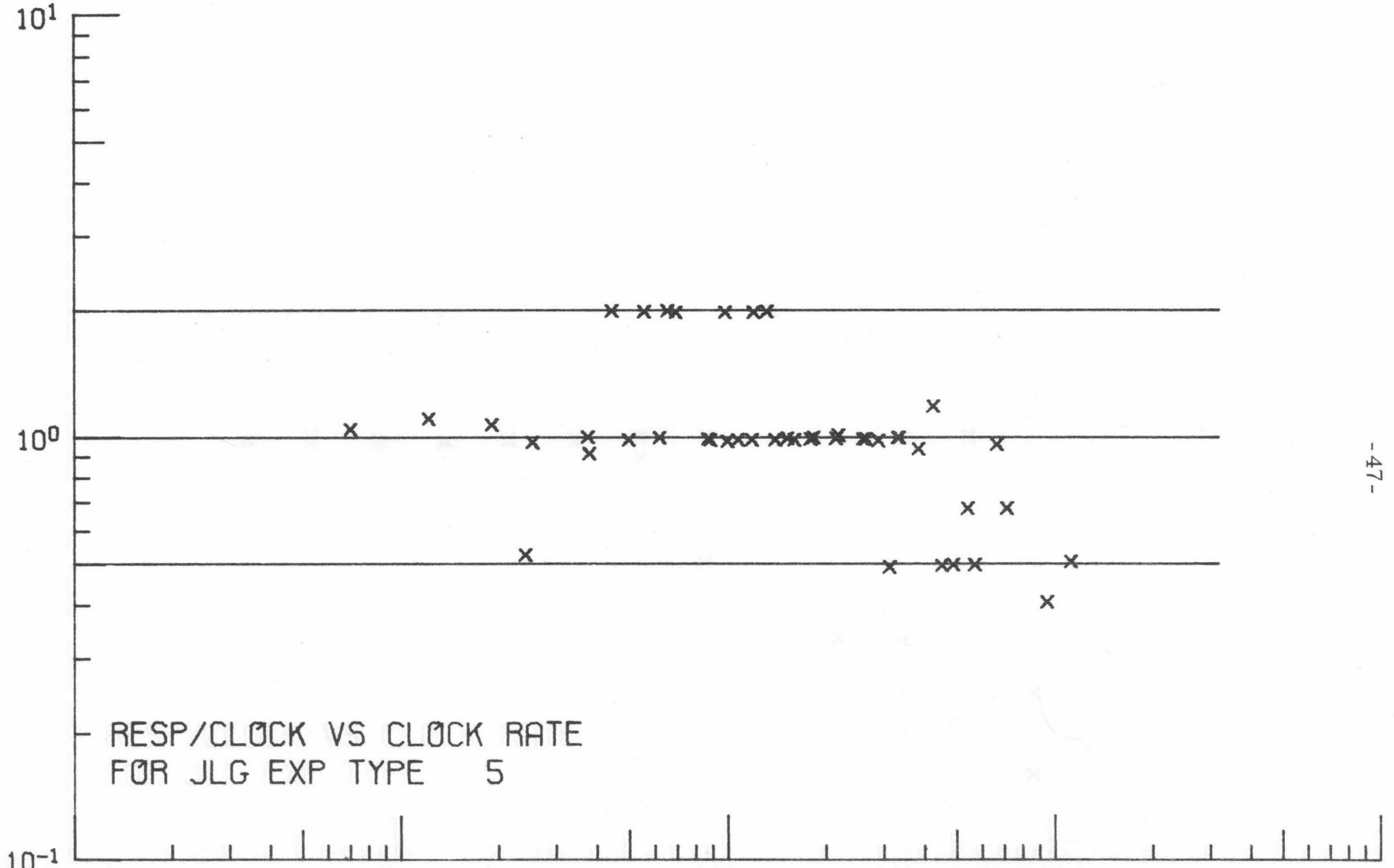

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLG EXP TYPE 5
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 21

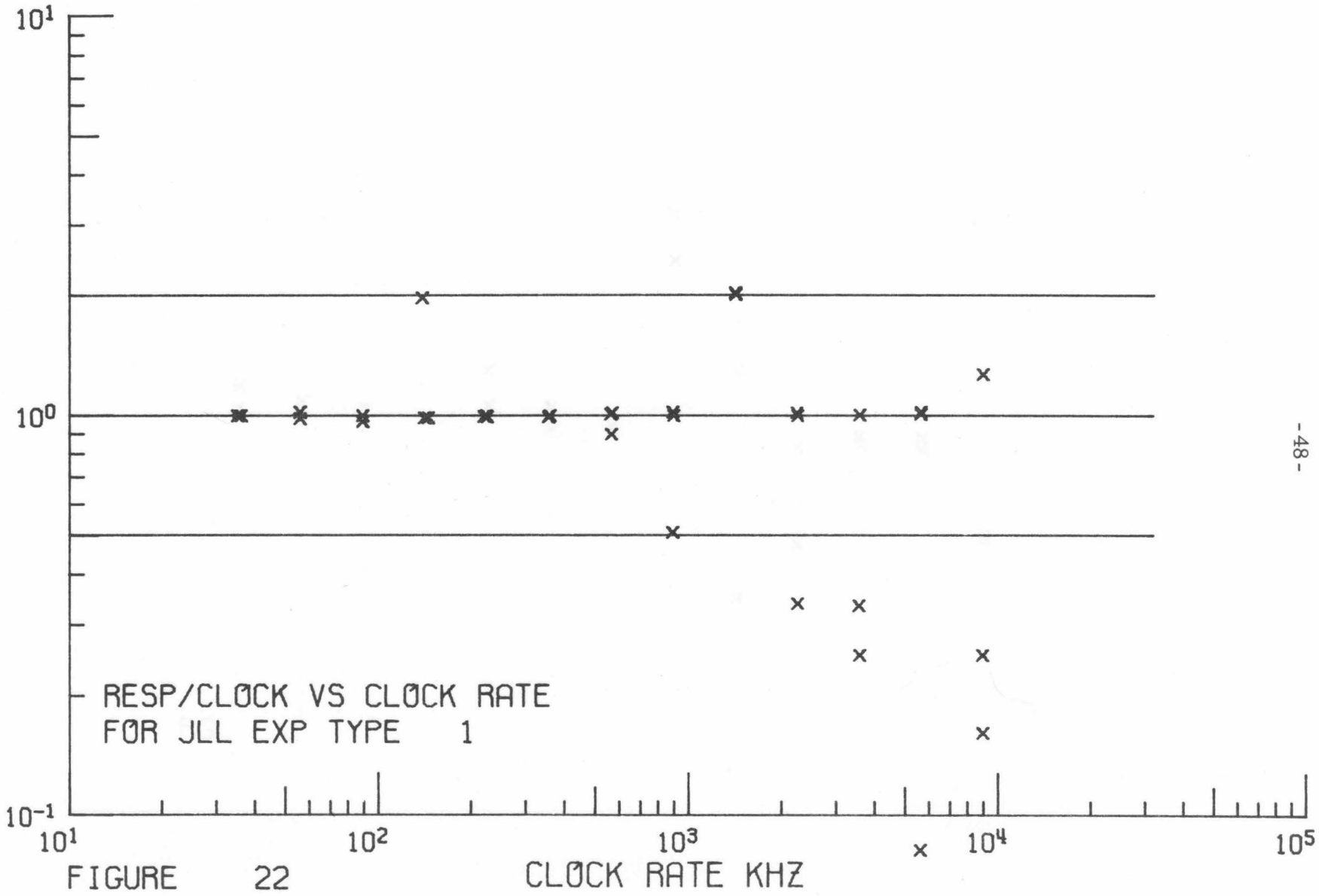

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JLL EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 22

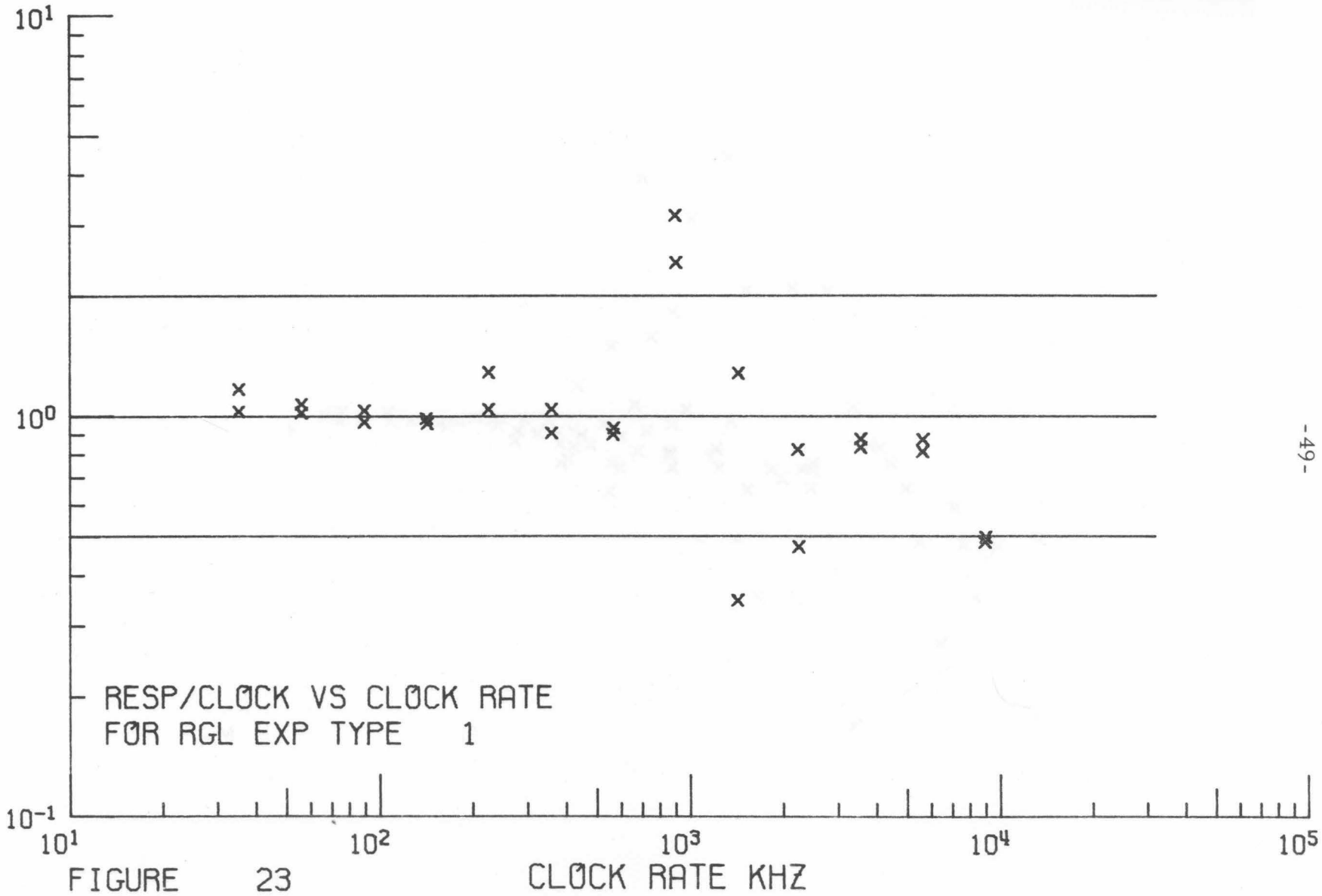

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR RGL EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 23

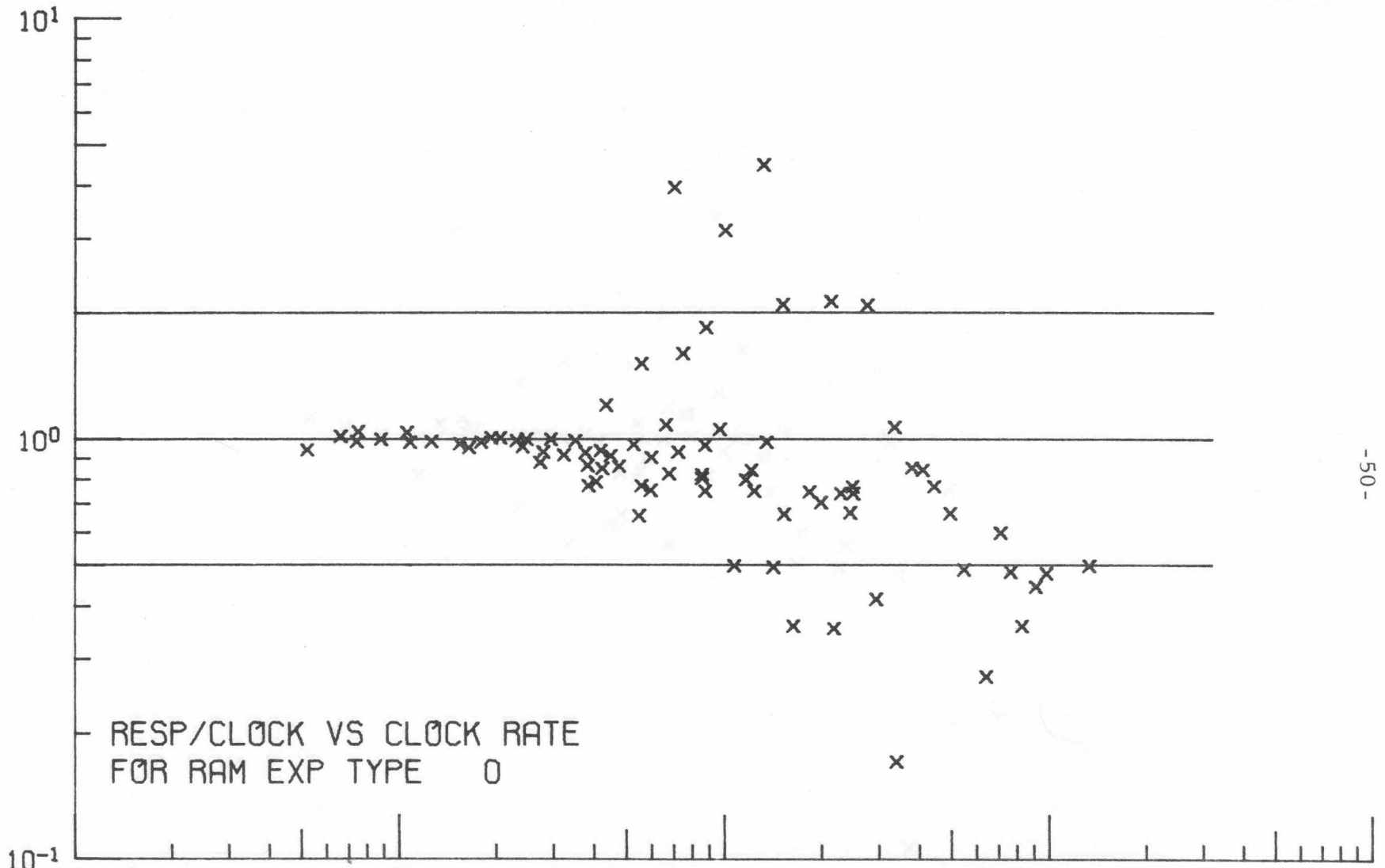

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR RAM EXP TYPE 0
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 24

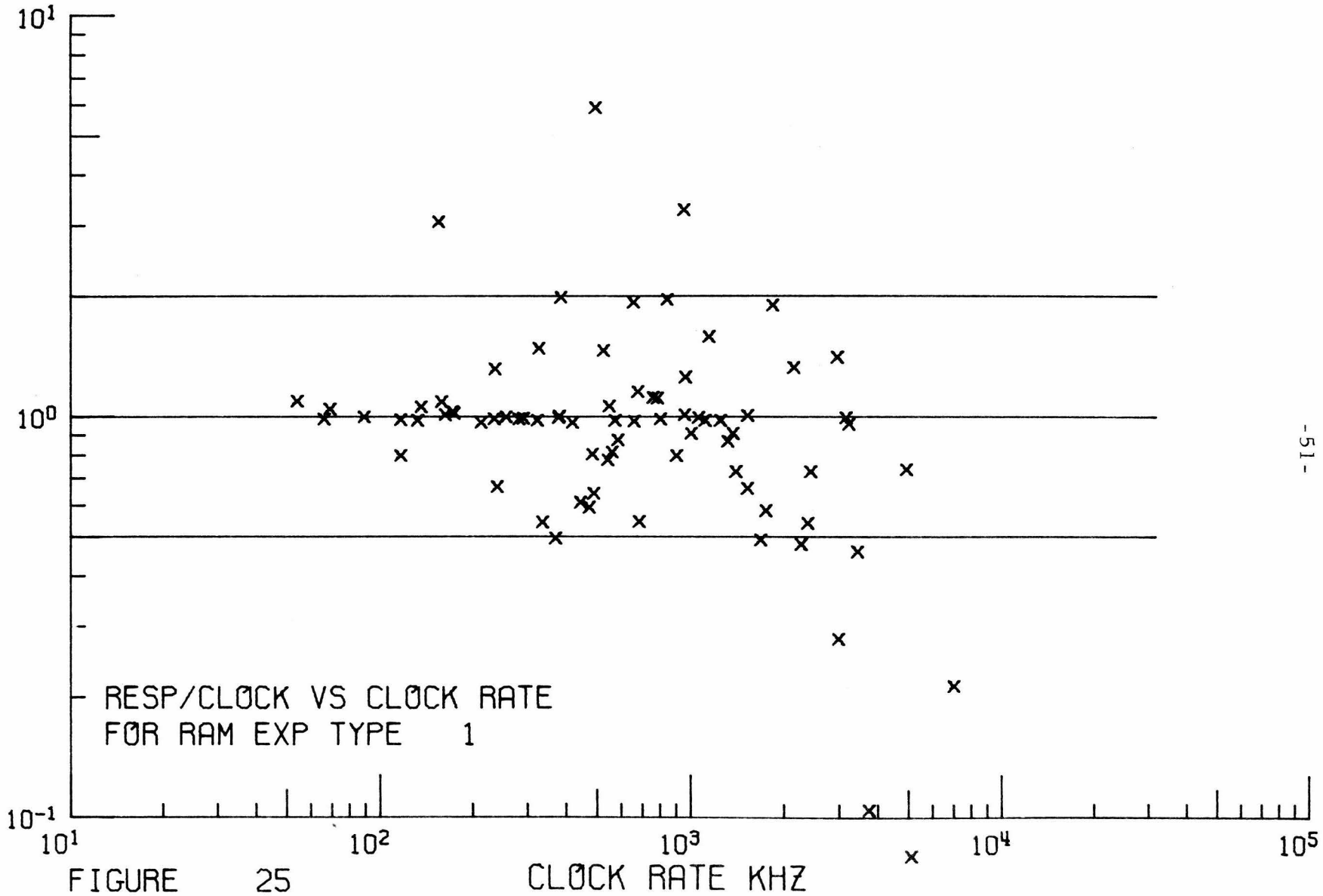

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR RAM EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 25

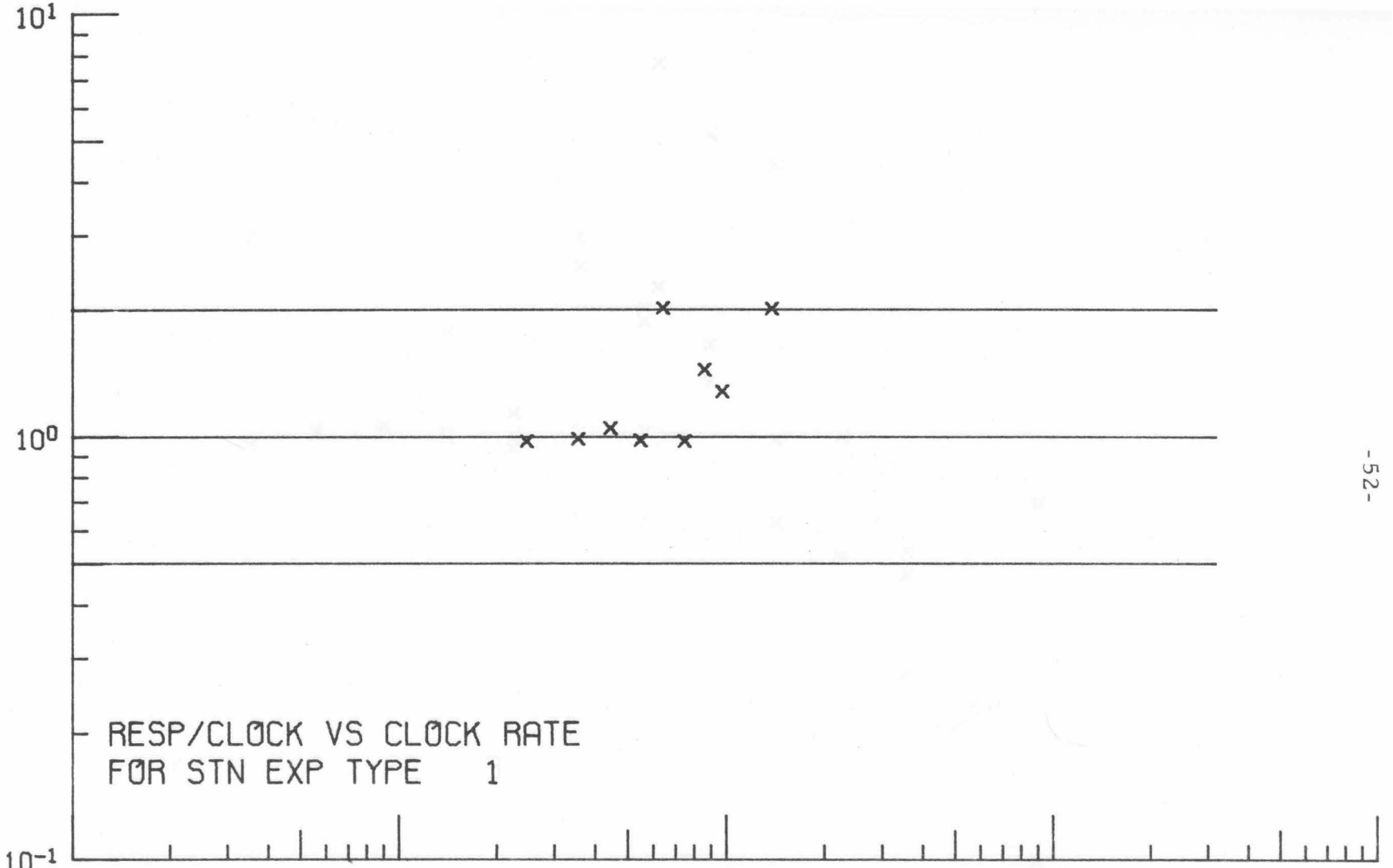

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR STN EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 26

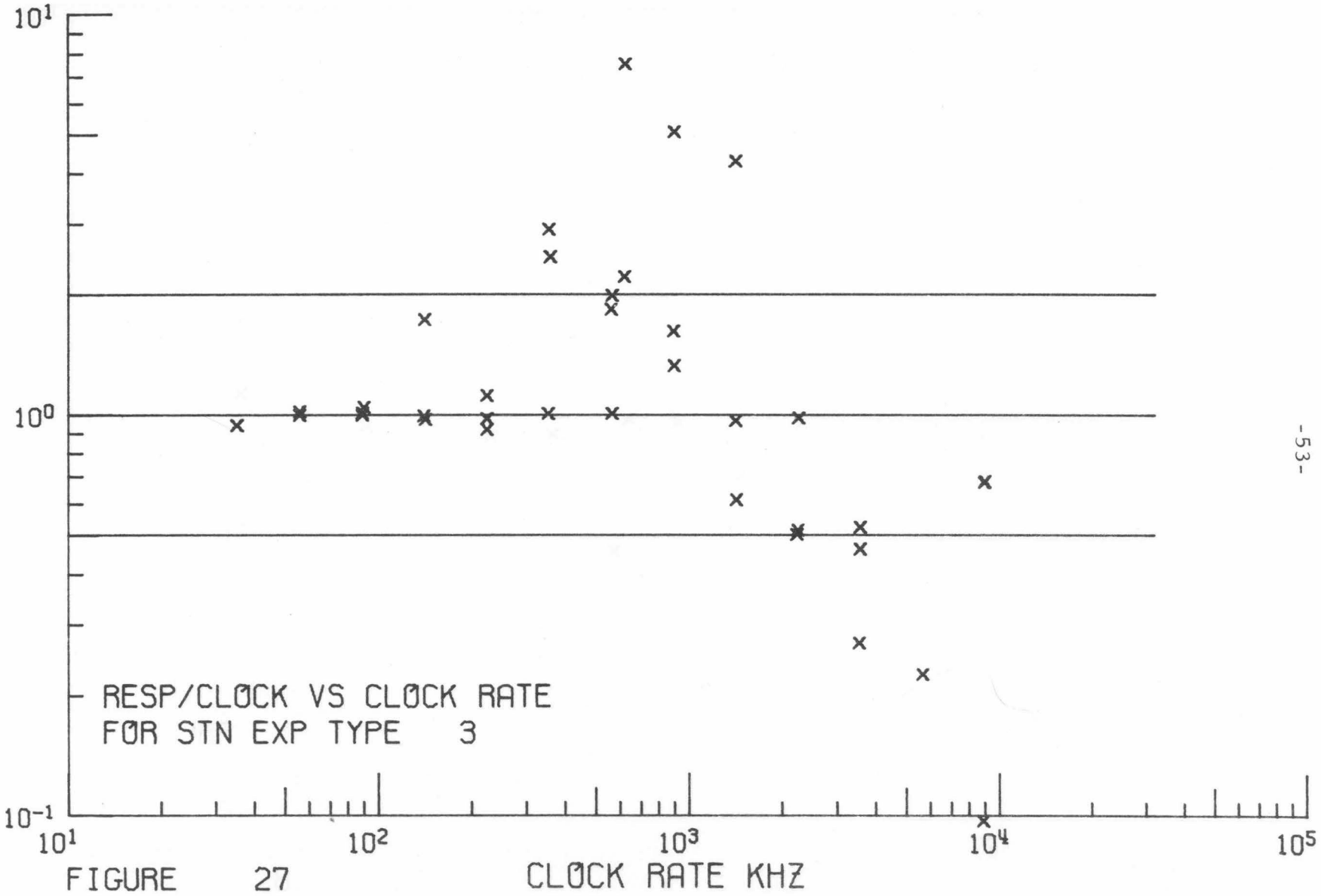

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR STN EXP TYPE 3
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 27

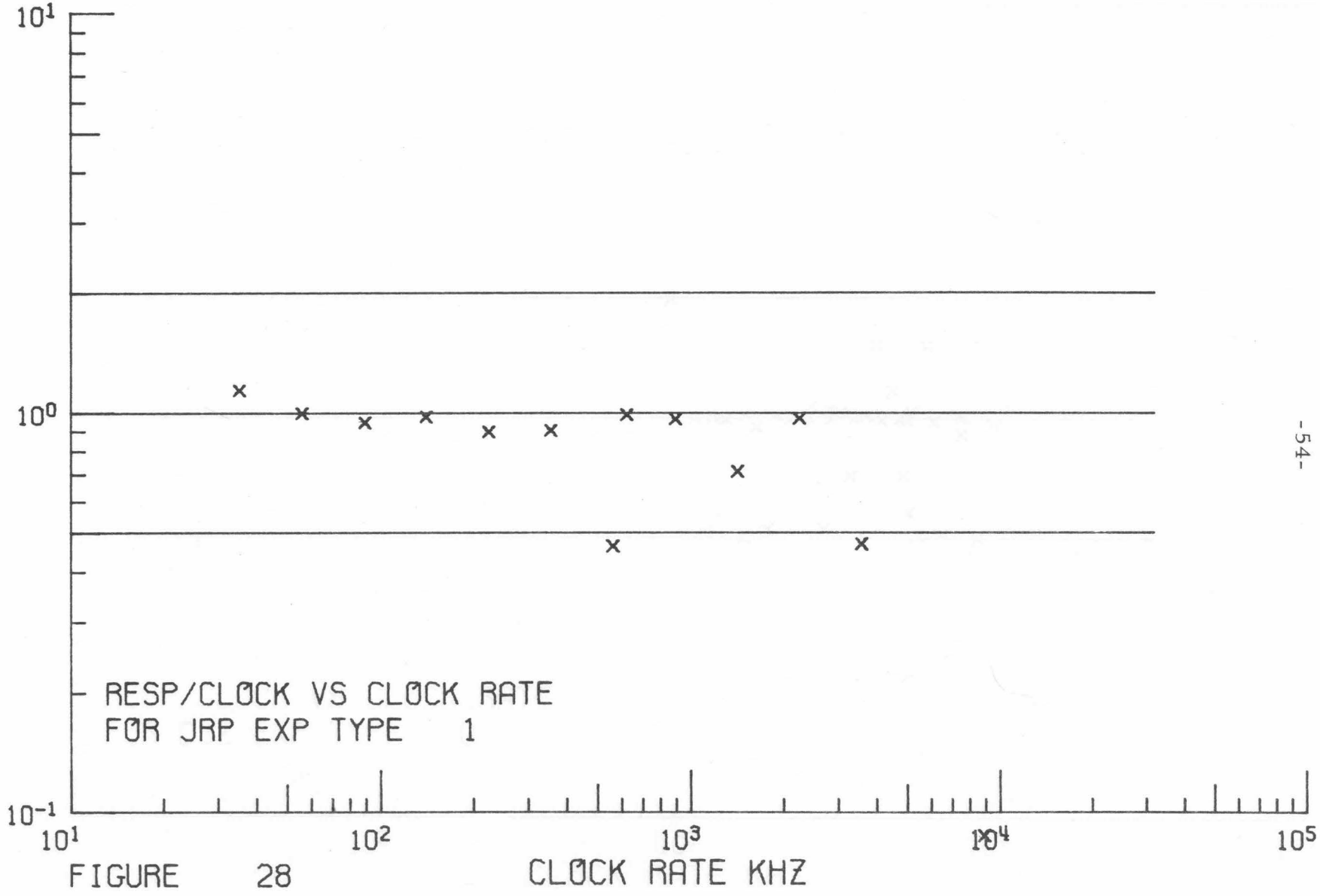

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR JRP EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 28

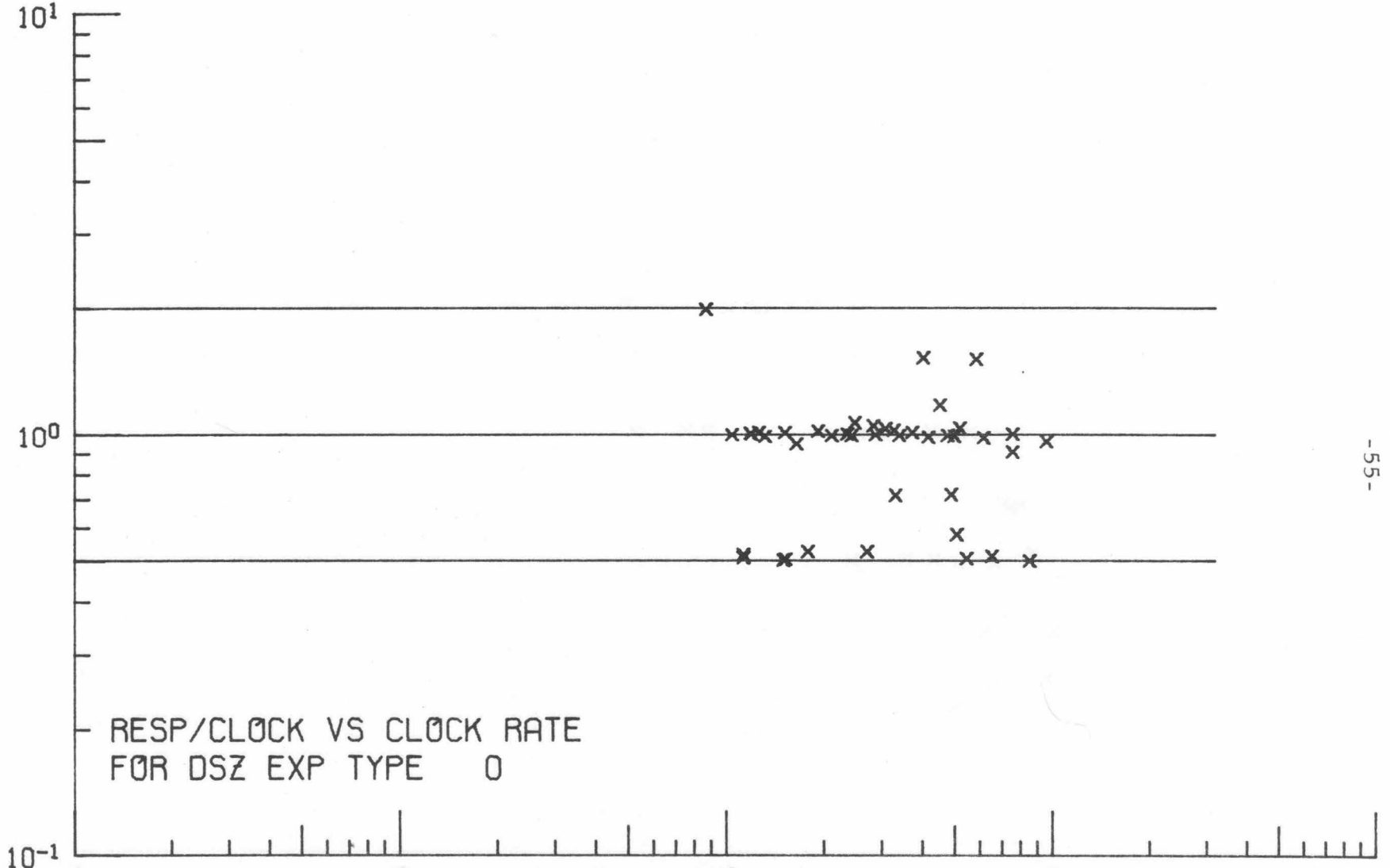

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR DSZ EXP TYPE 0
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 29

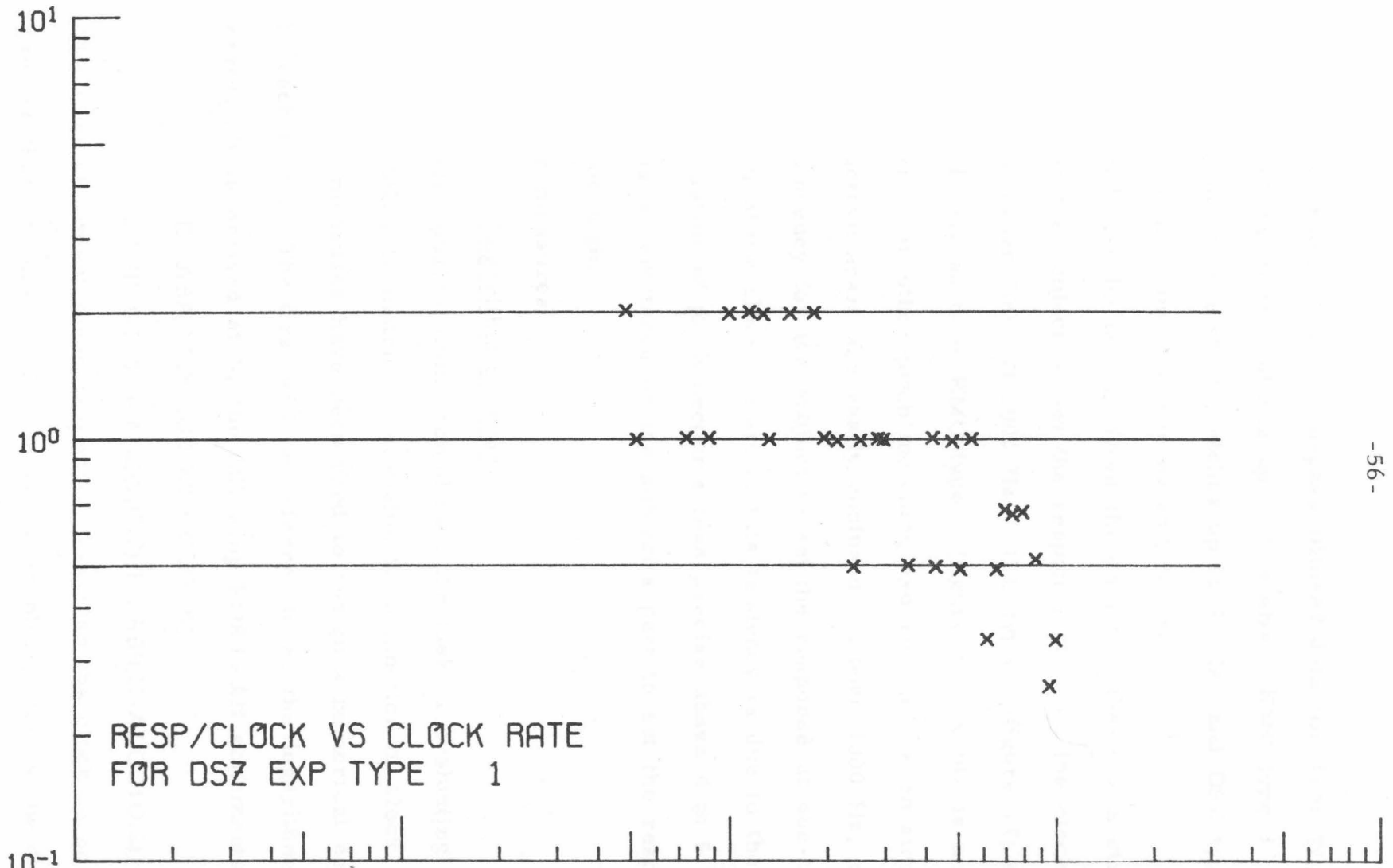

$10^1$
$10^0$
$10^{-1}$
RESP/CLOCK VS CLOCK RATE
FOR DSZ EXP TYPE 1
$10^1$
$10^2$
$10^3$
$10^4$
$10^5$
CLOCK RATE KHZ

FIGURE 30

at high clock rates.) The high-pass filtered data for JLG (type 0, figure 16) shows many matches up to 9.5 kHz. KMC type 1 (figure 15) shows consistent matches up to 4 kHz, and DSZ type 0 (figure 29) shows many matches up to 9.5 kHz.

It will also be noticed from the data that there is a strong tendency for the subject to set the response at twice the clock, especially between 500 and 1000 Hz. JLG type 1 (figure 17) shows this quite clearly as does KMC type 1 (figure 15). This is understandable since in other pitch matching experiments, even sine waves an octave apart are easily confused. Above 1000 Hz, there is also a tendency for the subject to set the response at one-half or one-third of the clock. Part of this tendency is due to the fact that the sensation of pitch becomes less precise above 4 to 5 kHz, and there is a reluctance on the subject's part to set the response generator too high.

## 3.4 Error measures

### 3.4.1 Logarithmic error

These considerations complicate the task of evaluating the subjects' ability to match the stimulus as a function of clock rate. Two error measures have been used to assign a numerical error to each point. The first will be referred to as the logarithmic error. It is arrived at by the following FORTRAN statements,

```
E = ABS(ALOG10(RESP/CLOCK))
IF (E.GT.0.5*ALOG10(2))E = ABS(E-ALOG10(2))
```

The first statement maps responses less than the clock to ones greater than the clock. On the raw data plots, this is the distance

measure to one from the line y = 2 (or y = 1/2). The log error increases as the response increases above the clock until the response is half an octave up. Figure 31 is a graph of the log error as a function of log (response/clock).

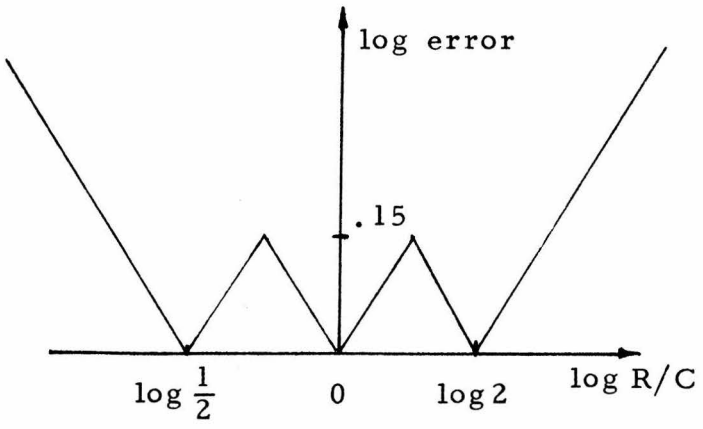



FIGURE 31: LOG ERROR

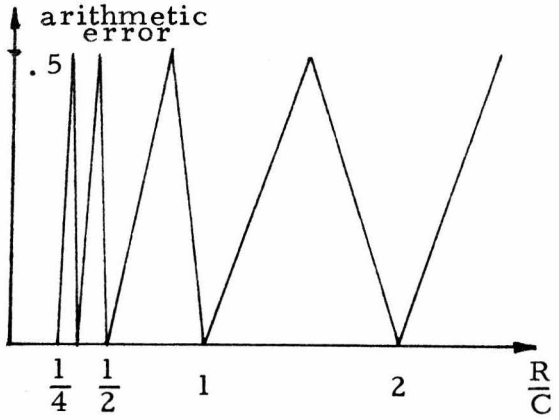



FIGURE 32: ARITHMETIC ERROR

### 3.4.2 Arithmetic error

The second error criterion, referred to as the arithmetic error, was prompted by the large number of responses at 1/3, 1/4, and even 1/5 and 1/6 of the clock rate. These occur principally in the experiments with KMC. The arithmetic error is arrived at by the following FORTRAN statements,

```
E = RESP/CLOCK
IF (E.LT.1.0)  E = 1./E
E = E-AINT(E)
IF (E.GT.0.5)  E = 1.0-E
```

The first two statements again map all responses less than the clock to ones greater than the clock. In the third statement, the

AINT function returns the greatest integer in E. This statement then forms the fractional part of E, limiting E to between 0 and 1.0. If RESP/CLOCK is closer to the integer above rather than the one below, the last statement changes E to a measure from the one above. Thus E is less than 0.5. The arithmetic error allows the subject to match with any integral ratio of response to clock. Of course these two error criterion give different numerical results, since their maximas between RESP/CLOCK = 1 and 2 are different, i.e., .1505 and .5, respectively. Thus they must be scaled if they are to be compared. Also the exact shapes of the curves are slightly different when they are plotted on the same scale of RESP/CLOCK, but this has very little effect.

3.5 Use of critical bands

Another problem in determining how well subjects match as a function of clock rate is the variability of the data. In psychophysical experiments in general, subjects rarely perform optimally at every trial. No single point can be regarded as significant. A single response/clock ratio of 1.0 might be due to luck, a single bad response may be due to the subject's not feeling well or a temporary feeling of frustration (particularly in the author's case). It is therefore necessary to average over several points to get any real indication of ability to match.

It is very dangerous to average data between subjects because of the great differences in subjects' ability. For instance CMC type 1 (figure 3) shows considerable spread of response/clock for frequencies below 150 Hz. RAM type 1 (figure 16) shows a

substantial portion of responses below 500 Hz more tightly clustered around 1.0. DSZ type 1 (figure 20) shows a group of responses almost exactly 1 up to 5 kHz. As previously mentioned, the early experiments used a fixed set of 13 clock rates. The original intention was to average the points from each subject for each of these rates. This scheme was abandoned partially in a desire to get a finer picture of matching ability as a function of frequency. Also it was felt that some subjects (particularly JLG, JWB and DSZ), due to their outstanding ability, might memorize the frequencies used and (subconsciously) limit their responses to the proscribed list. But in using more clock rates, it becomes necessary to average together points over some range of clock rates, since no given rate that is used necessarily reappears.

The auditory critical bands[27] were chosen as appropriate frequency bands over which to average. In experiments involving bandwidth such as loudness of bands of Gaussian noise or the masking of tones with bands of noise, there is a certain bandwidth at which the character of subjects' responses changes abruptly. This bandwidth is a function of frequency and is termed the critical band. It is generally thought to be the effective bandwidth of the basilar membrane filter function at a particular place. Critical bands are widely used because radically different experiments show amazing agreement on the value of bandwidth as a function of frequency. For frequencies below 1000 Hz, the critical bandwidth is about 100 Hz. Above 1000 Hz, the critical bandwidth increases monotonically with frequency, reaching about 2 kHz at

10 kHz. For the purposes of analyzing the data here, the audio spectrum is divided into bands whose width is the critical bandwidth. Table 2 lists the bands used. The band edges are completely arbitrary; however any two clock frequencies in the same band will produce similar responses.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0-100 | 7 | 630- 770 | 13 | 1720-2000 | 19 | 4400- 5300 |
| 2 | 100-200 | 8 | 770- 920 | 14 | 2000-2320 | 20 | 5300- 6400 |
| 3 | 200-300 | 9 | 920-1080 | 15 | 2320-2700 | 21 | 6400- 7700 |
| 4 | 300-400 | 10 | 1080-1270 | 16 | 2700-3150 | 22 | 7700- 9500 |
| 5 | 400-510 | 11 | 1270-1480 | 17 | 3150-3700 | 23 | 9500-12000 |
| 6 | 510-630 | 12 | 1480-1720 | 18 | 3700-4400 | 24 | 12000-15500 |

TABLE 2: CRITICAL BANDS

## 3.6 Discussion of error plots

### 3.6.1 Figures

Figures 33 through 53 are computer generated plots of averages over a critical band versus critical band number. Each figure is for a different subject and experiment type, and contains four plots: the response to clock ratio, the absolute value of the logarithm of the ratio, the logarithmic error, and the arithmetic error. The abscissa of each plot is the critical band number, and therefore corresponds roughly to log frequency.

A line is drawn on the logarithmic and arithmetic error plots at the value which results if the subject matched randomly. This value is arrived at by assuming that the ratio, r = (response/clock) is uniformly distributed between two ratios, $r_1$ and $r_2$. Since the logarithmic error is unbounded, fixed values of $r_1 = 1/4$ and $r_2 = 4$ have been chosen as reasonable bounds on a subject's

ratio. This results in a random error of 0.4257 for the logarithmic error and 0.2478 for the arithmetic error.

3.6.2 Type 1 experiments

The error plots are intended as aids to the evaluation of the raw data. JLG was clearly able to match all clock rates used in the type 1 experiments. Of the 65 points in figure 17, only at 9 kHz did he fail to match with ratios of 1/2, 1 or 2. At 9 kHz, he has 3 points at 1/4, 1 at 1/2, and 1 at 2/3 which could also be argued to be a match. In the error plots for JLG type 1 (figure 40), the log error and the arithmetic error reflect his ability to match, staying quite close to zero up to critical band 22. The deviation in the 2 ratio plots of figure 40 are the result of averaging ratios of 1.0 and 2.0, thus the average in bands 6 and 8 is 1.4.

All of the JLG type 1 data were taken with the first pseudo-noise generator which has been shown elsewhere to be suspect. However this objection does not apply to DSZ, whose type 1 experiments were run with the improved pulse generator. His raw data (figure 30) are quite similar to JLG's. Indeed, the DSZ type 1 arithmetic error (figure 53) is nearly zero up to critical band 21. (DSZ's match at 1/3 at 6 kHz throws the log error off in critical band 20.)

Other subjects for whom the error plots indicate an ability to match are JWB type 1 (figure 34), KMC type 1 (figure 38), and JLL type 1 (figure 45). JWB is able to match in band 14, 2.1 kHz, but not in band 16, 2.9 kHz. The abrupt change shown by

his arithmetic error plot is not obvious from his raw data, figure 11. KMC's type 1 arithmetic error rises abruptly in band 18, about 4 kHz. From the raw data of figure 15, one would be inclined to put her limit at 3 kHz, with the first appearance of a match at 2/3.

Five subjects have been cited as being able to match: JLG, DSZ, JWB, JLL, and KMC. Concerning the other five subjects, there are very little data for JRP and RGL and not much for STN. CMC type 1 (figure 37) is based on a considerable number of points and shows very little ability, even at low clock rates. RAM type 1 is difficult to evaluate, apparently having trouble in bands 5 and 6, 400 to 600 Hz. RAM's arithmetic error (figure 48) then is fairly low in bands 7 and 8 and then climbs slowly until about band 15.

### 3.6.3 High-pass filter experiments

Five subjects participated in the high-pass filter experiments, type 0. Examining figures 39 and 40, JLG's arithmetic error shows his sharp change in band 22 for type 1 becoming more gradual over bands 16 to 18. The corresponding plots for DSZ, figures 52 and 53, show a very similar effect. The error plots for JWB, figures 33 and 34, again show an abrupt increase in error moving to lower frequency with the addition of the high-pass filter. However, in this case everything occurs lower in frequency and with larger errors. His cutoff moves from band 16, 2.9 kHz, to band 6, 550 Hz. There are very little data for KMC type 0, figures 37 and 38, the points of the error curves each being based on only 1 data

point. One is tempted to speculate that her cutoff moved from band 19 to band 15. RAM's data (figure 47), while based on many points, present an unclear picture of her ability, showing matches in some bands and not in others.

### 3.6.4 Pattern experiments

Three subjects participated in the computer generated pattern experiments, type 3. No conclusions can be drawn from CMC (figure 36). STN's error plots, figures 49 and 50, seem to indicate a destruction of ability to match when using patterns, however these curves are based on very few points. JLG is the most important case for the pattern data, which also includes experiment type 2. It was his ability to match that we were trying to destroy with the patterns. Examining figures 40 through 42, we see that the length 4 patterns of type 2 had very little effect on the arithmetic error. The length patterns of type 3 did increase his errors substantially in bands 18 to 22, but they are still quite small in comparison to other subjects. Also note that the log error plots for the pattern experiments are substantially larger as a result of his matching at ratios of 1/3, 1/4, 1/5, and 1/6 (see figures 18 and 19).

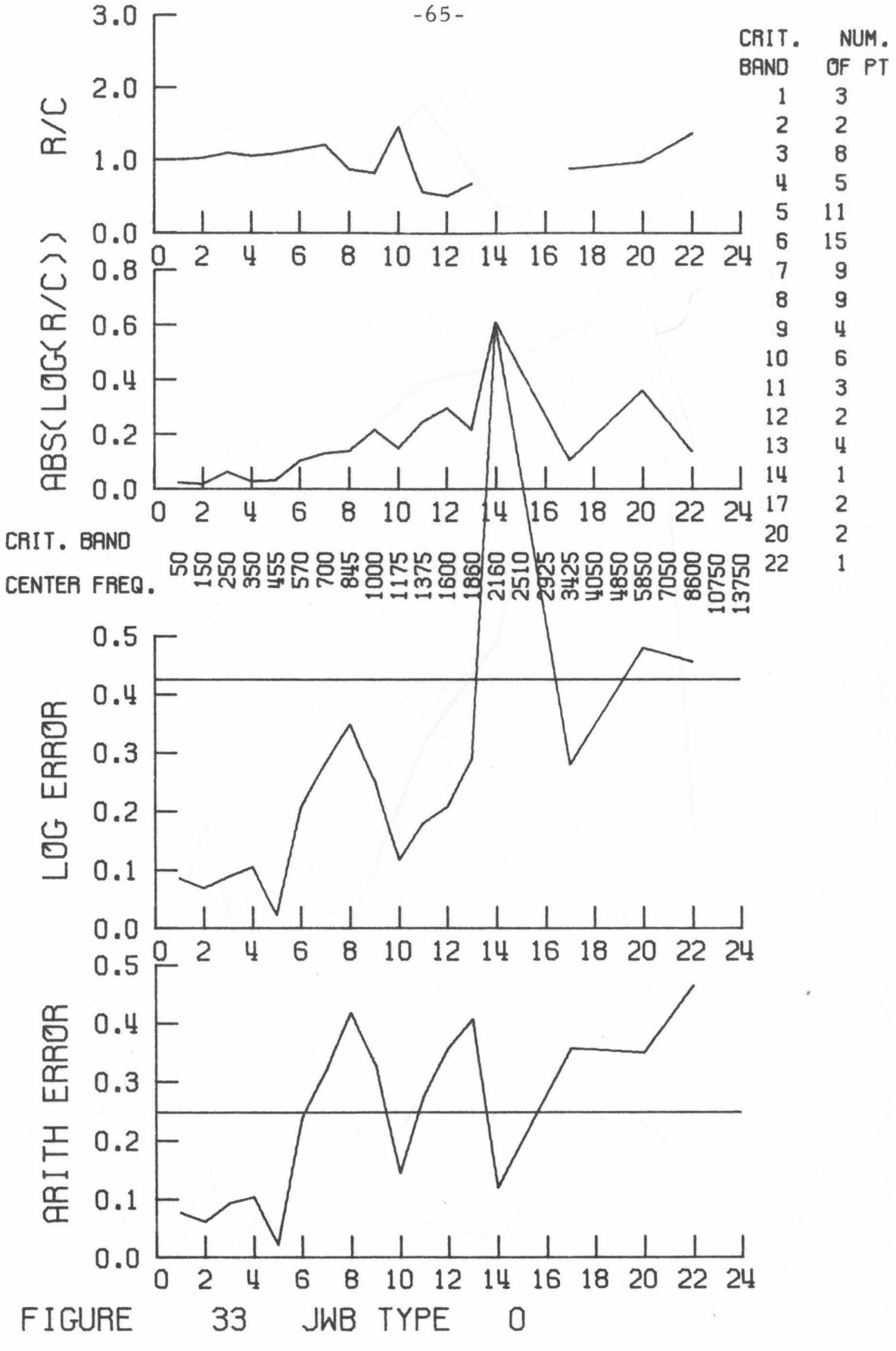



FIGURE 33 JWB TYPE 0

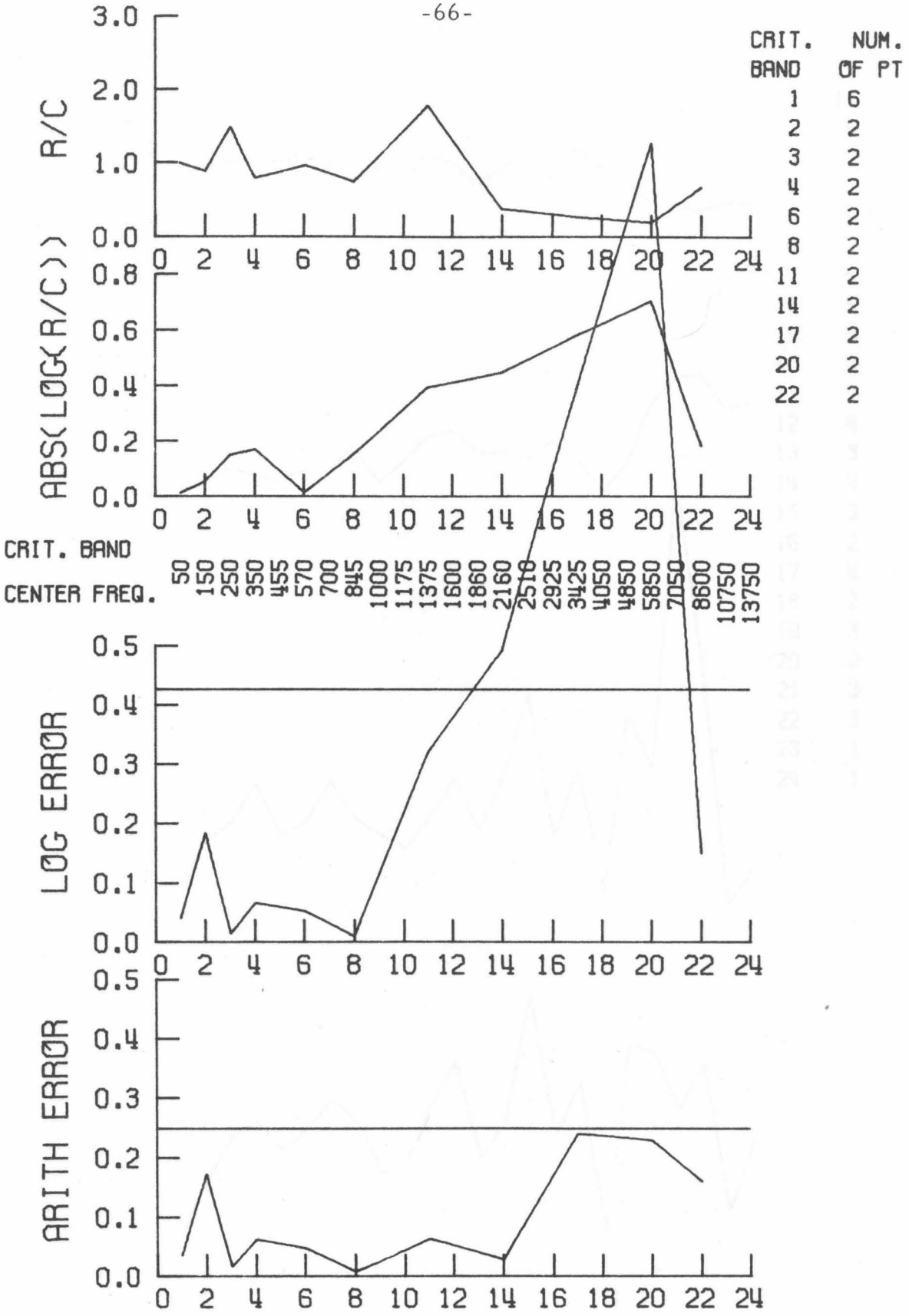



FIGURE 34 JWB TYPE 1

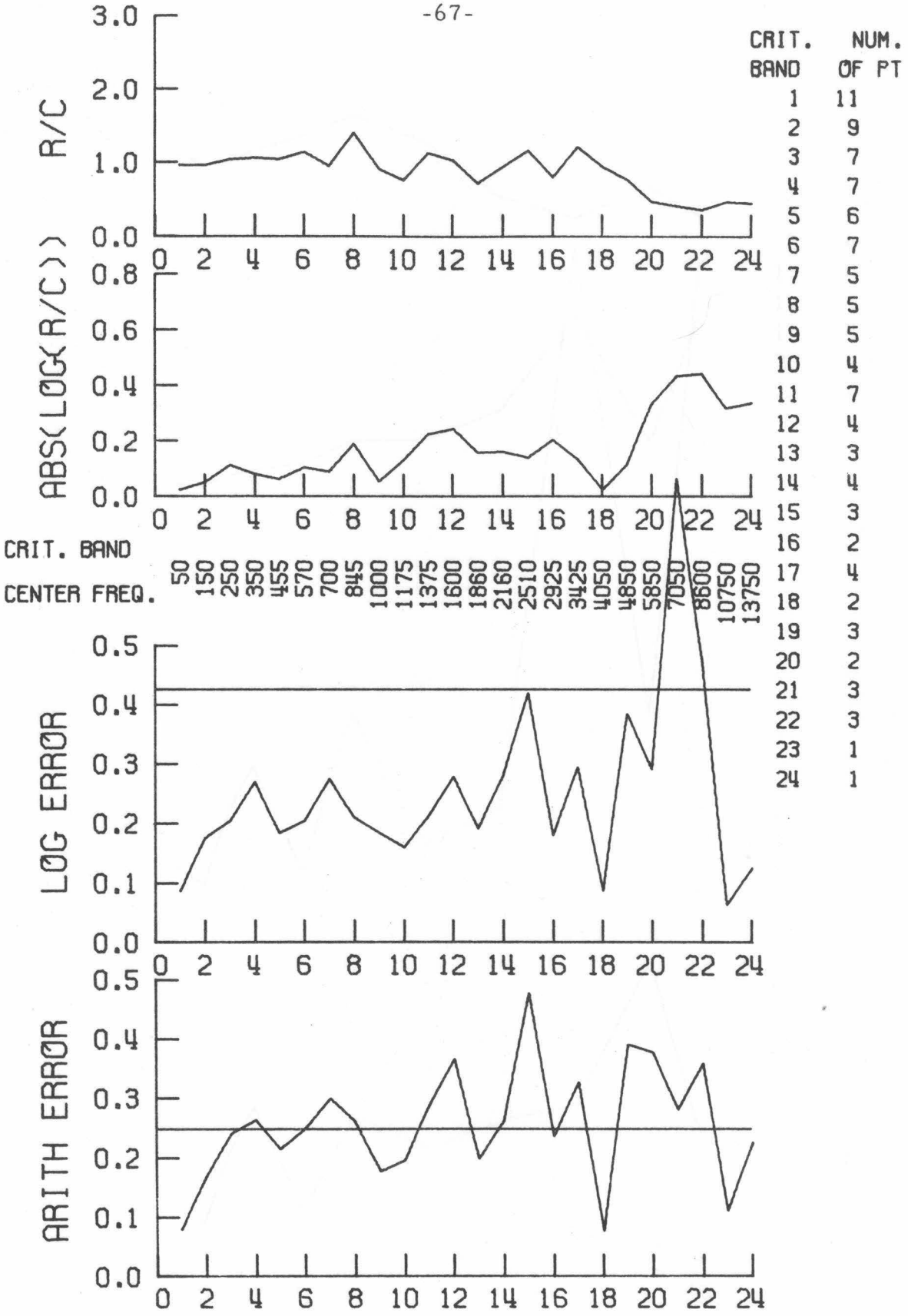



FIGURE 35 CMC TYPE 1

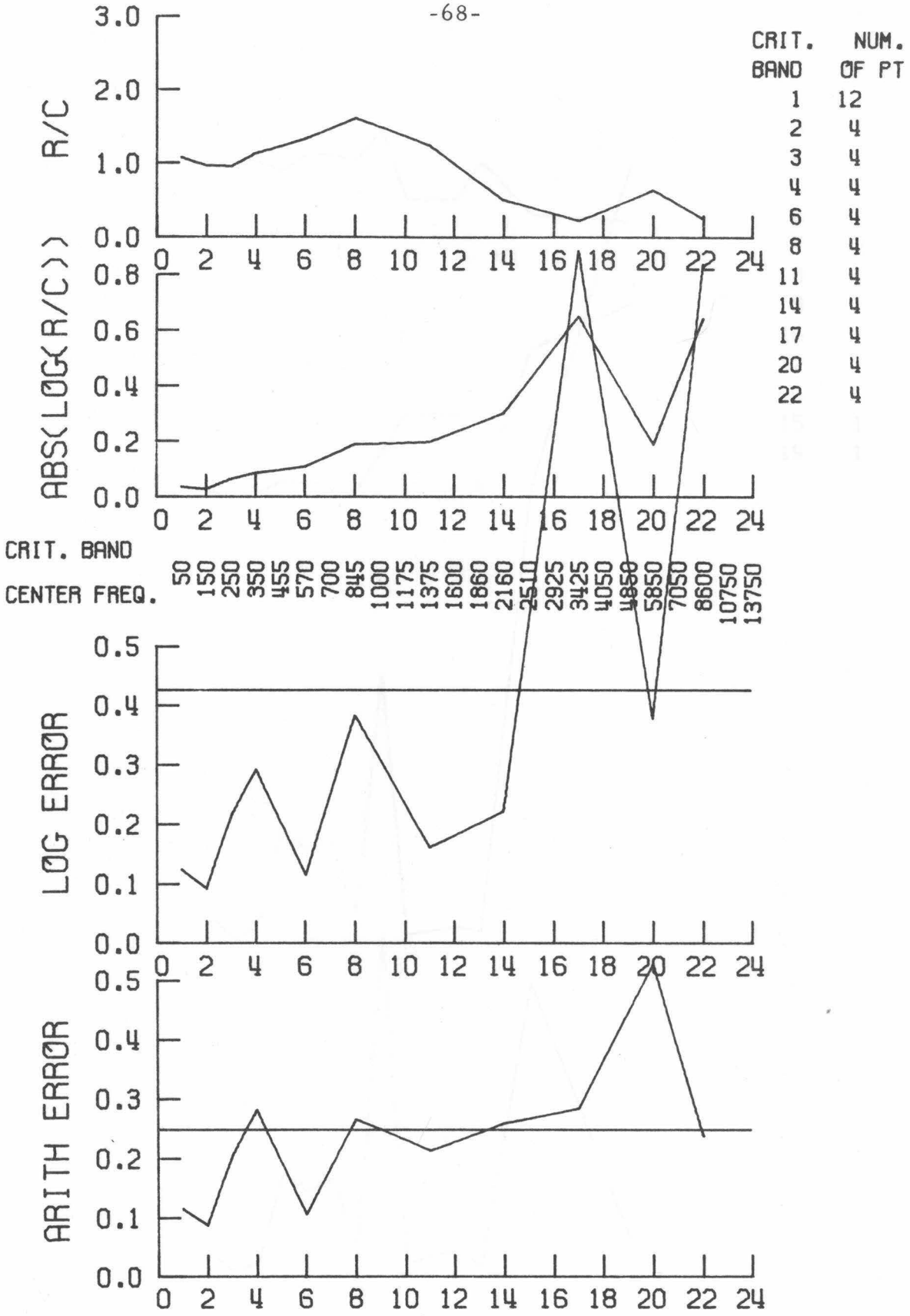



FIGURE 36 CMC TYPE 3

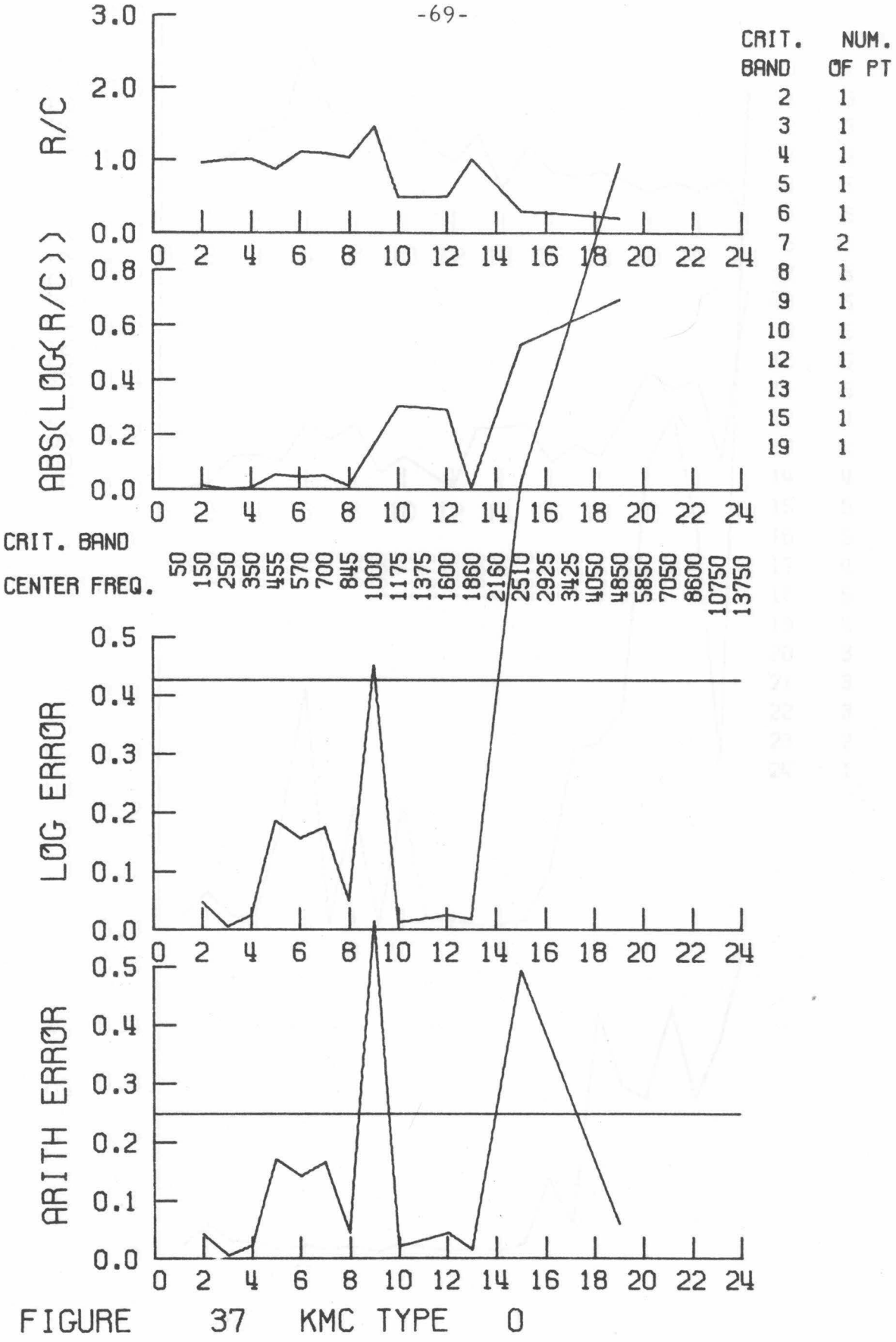



FIGURE 37 KMC TYPE 0

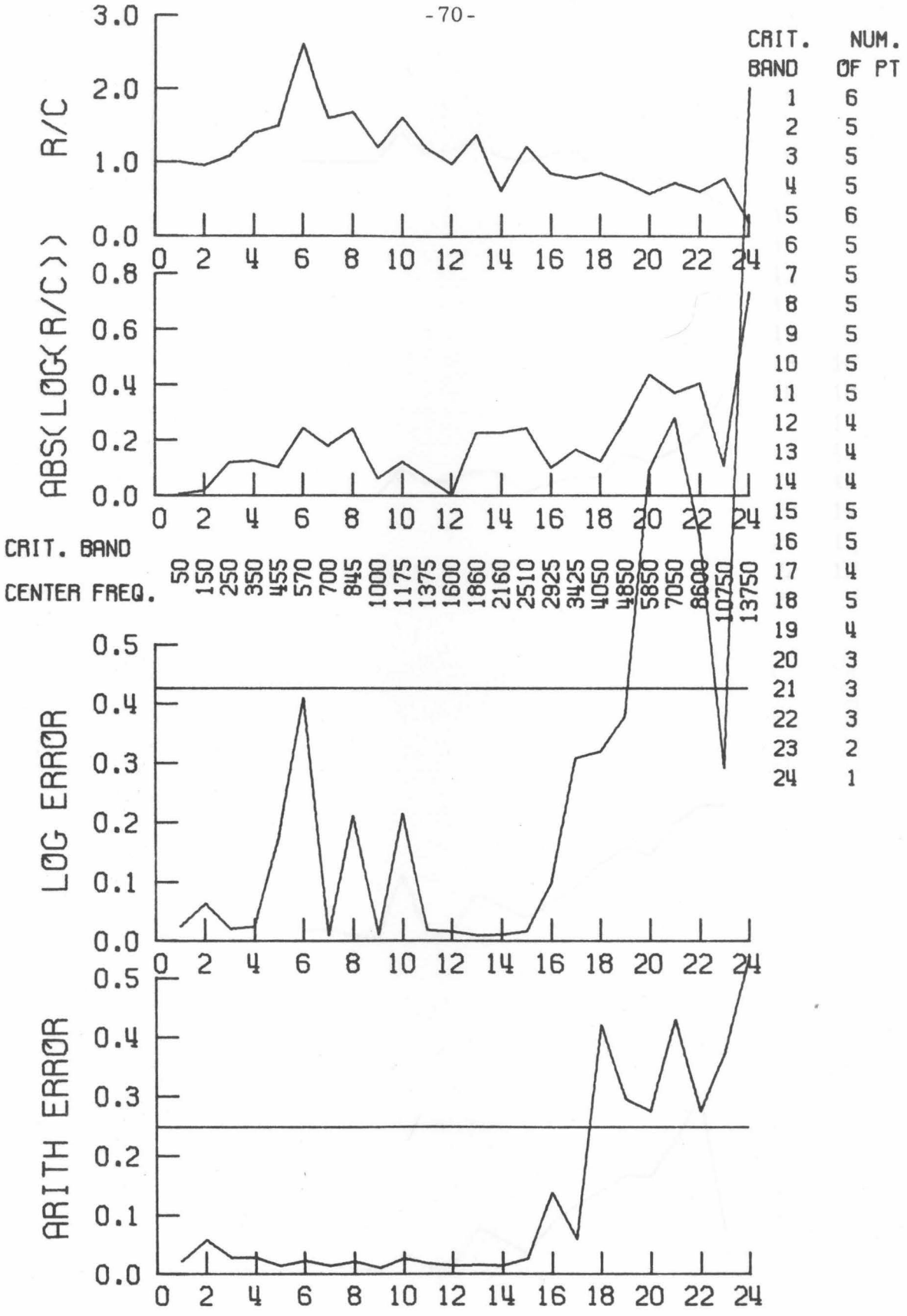



FIGURE 38 KMC TYPE 1

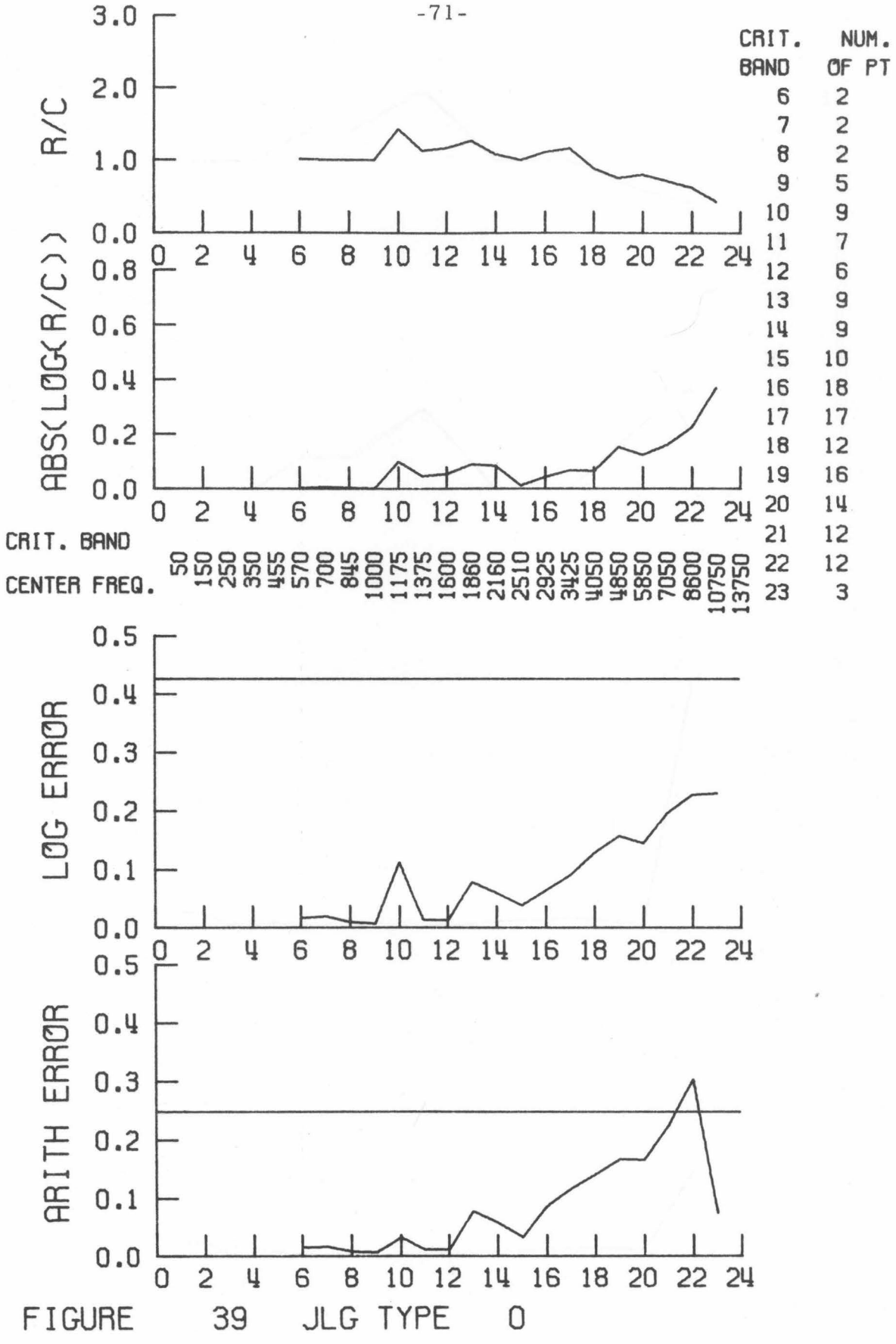
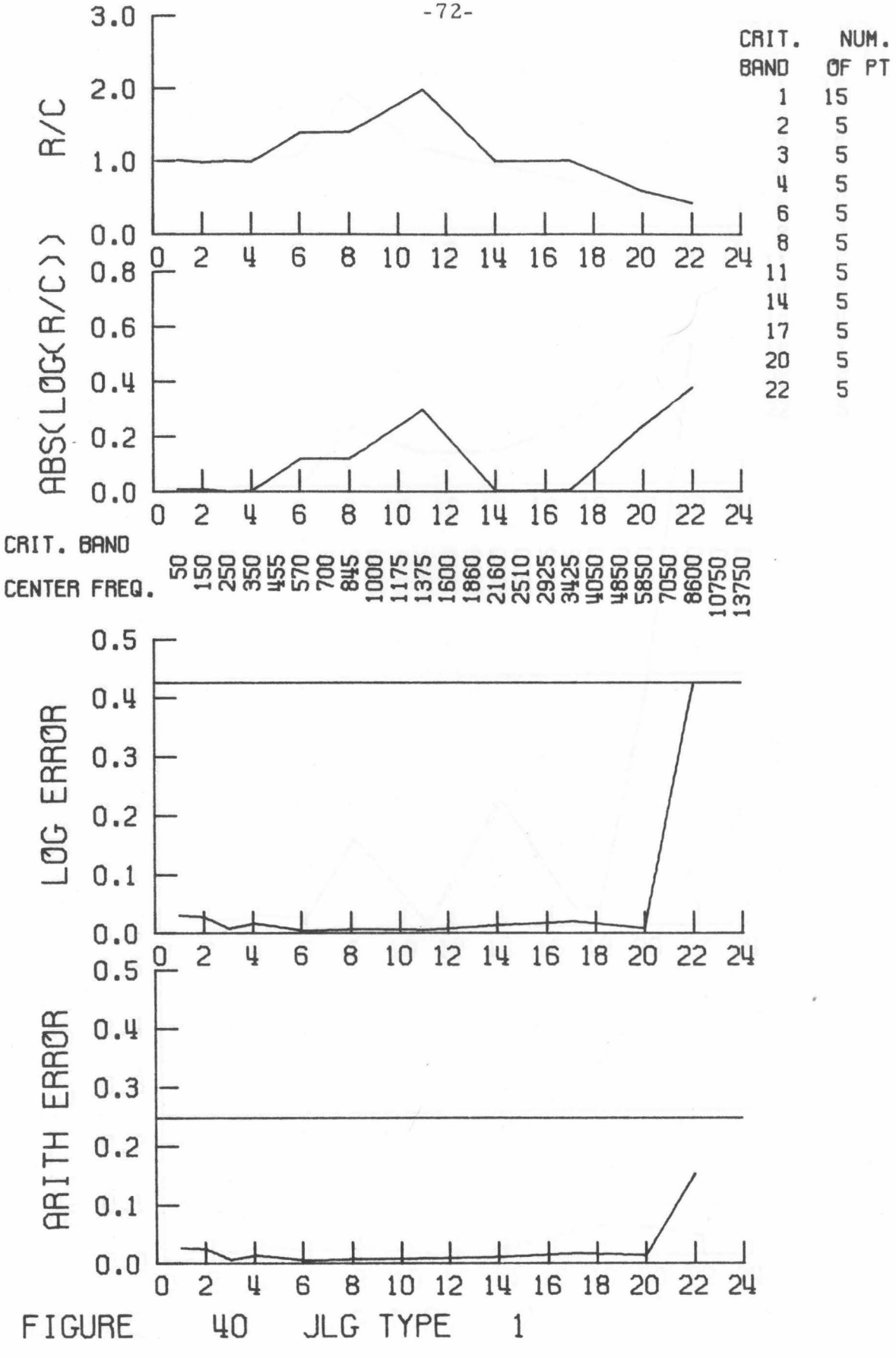



FIGURE 39 JLG TYPE 0




| CRIT. BAND | NUM. OF PT |
|---|---|
| 1 | 15 |
| 2 | 5 |
| 3 | 5 |
| 4 | 5 |
| 6 | 5 |
| 8 | 5 |
| 11 | 5 |
| 14 | 5 |
| 17 | 5 |
| 20 | 5 |
| 22 | 5 |

FIGURE 40 JLG TYPE 1

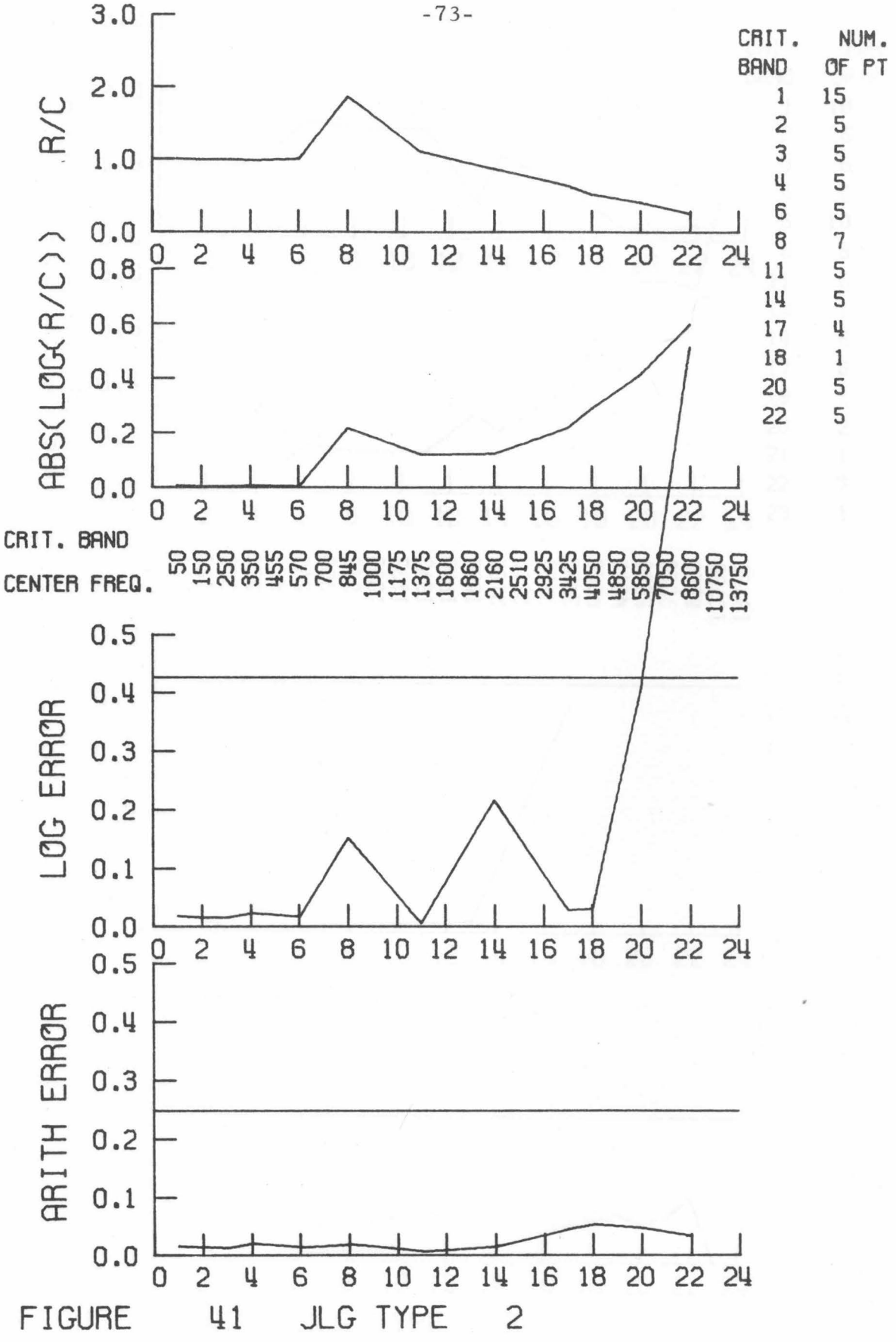



FIGURE 41 JLG TYPE 2

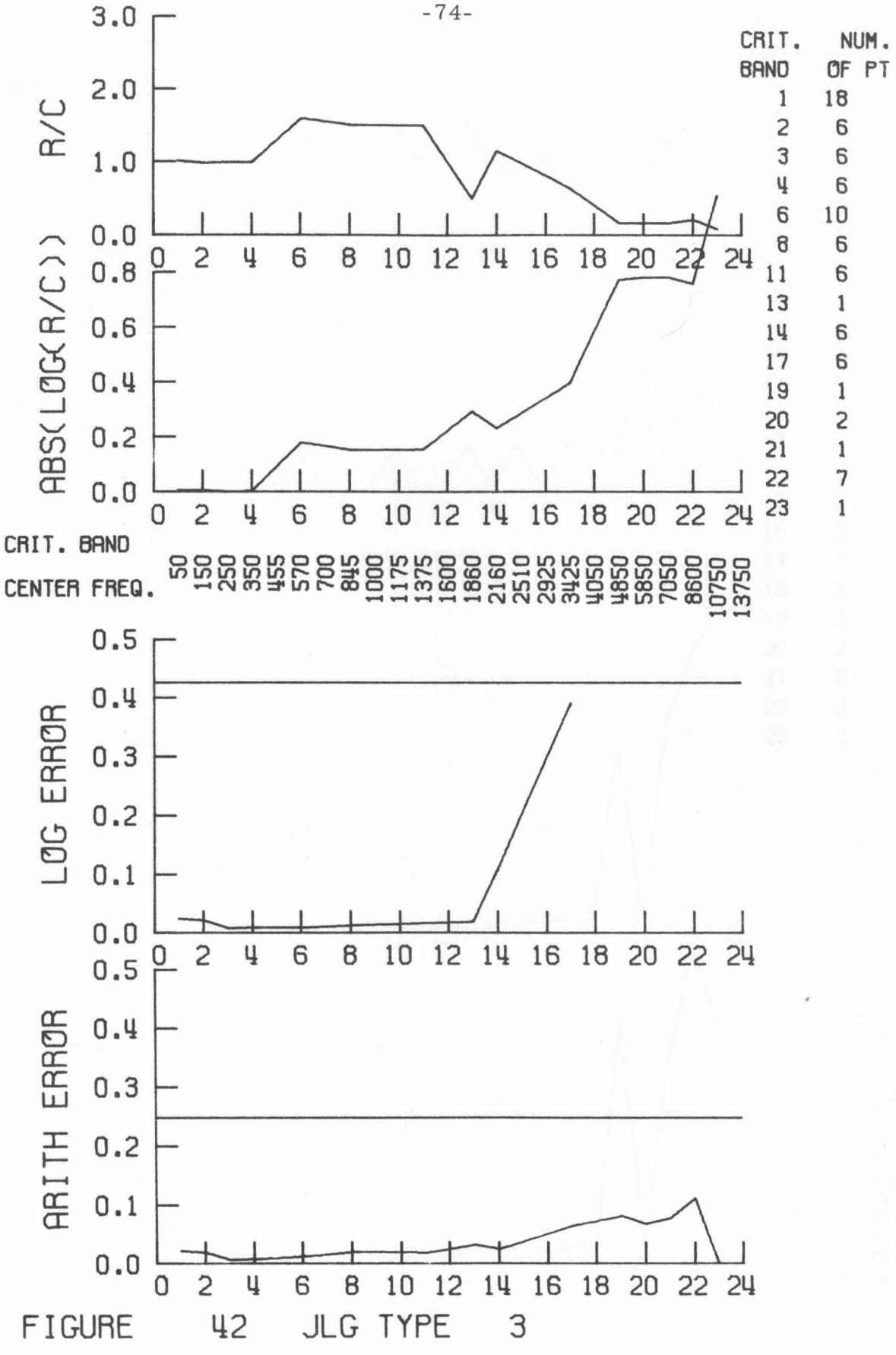



FIGURE 42 JLG TYPE 3

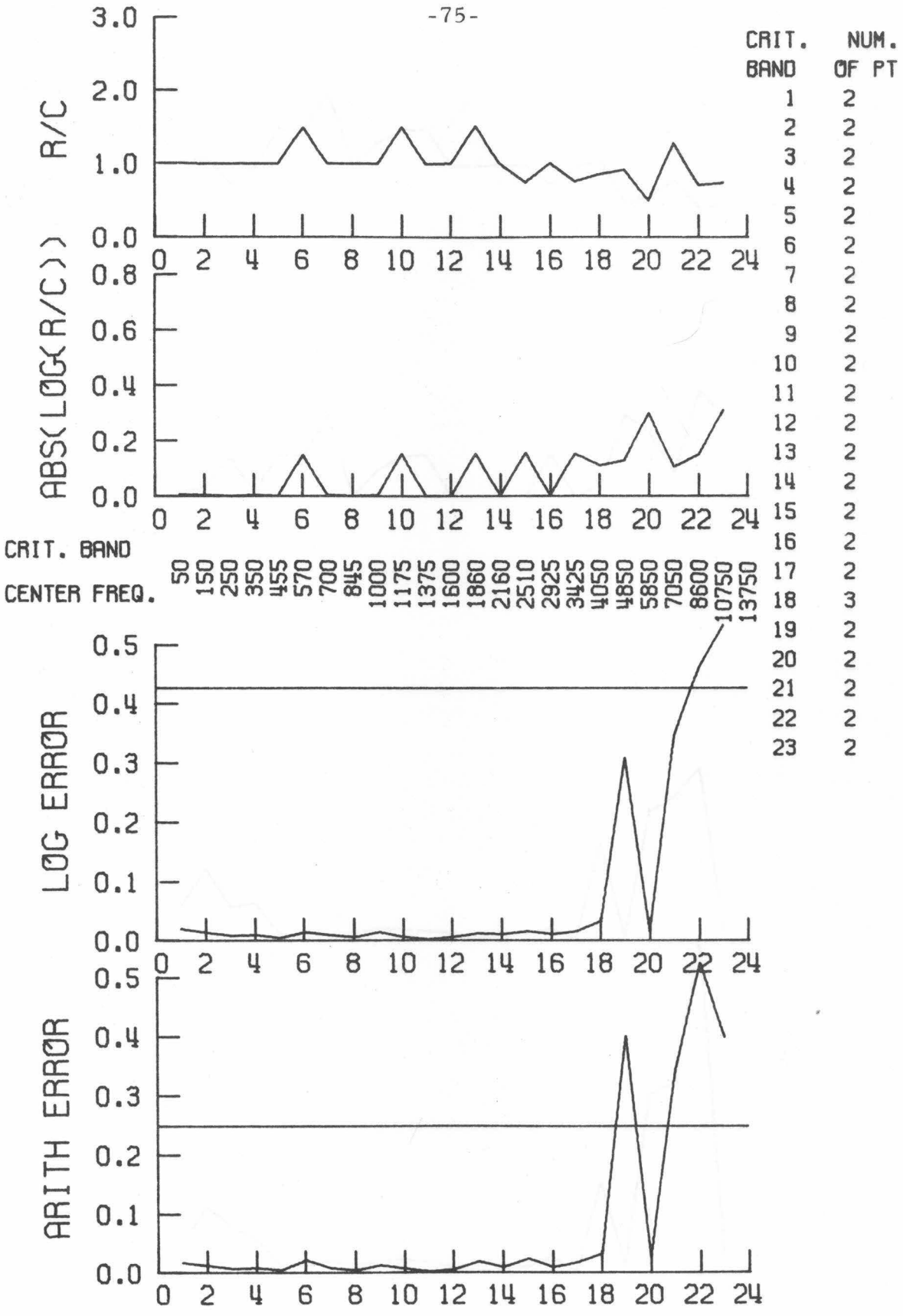



FIGURE 43 JLG TYPE 4

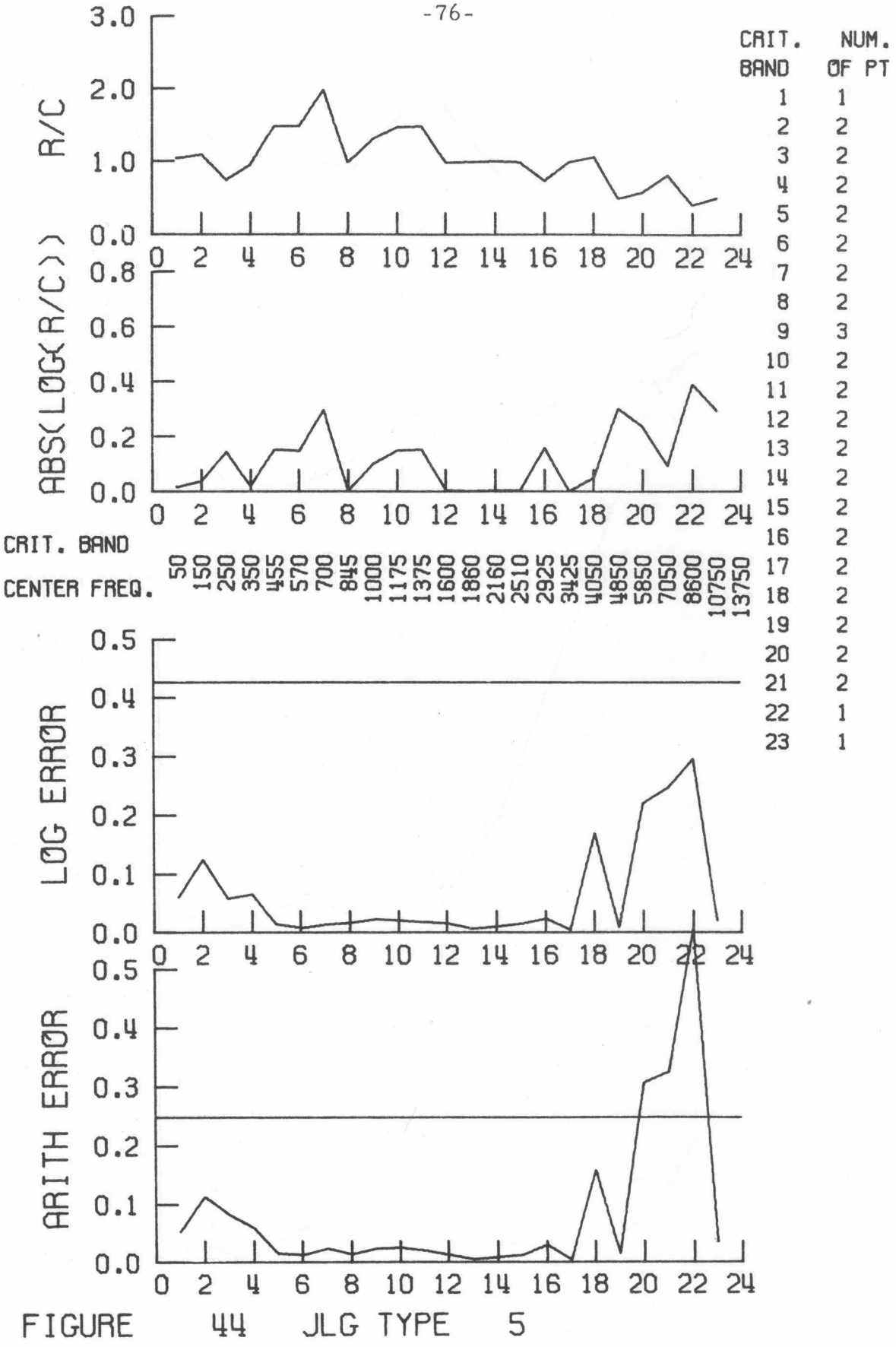



FIGURE 44 JLG TYPE 5

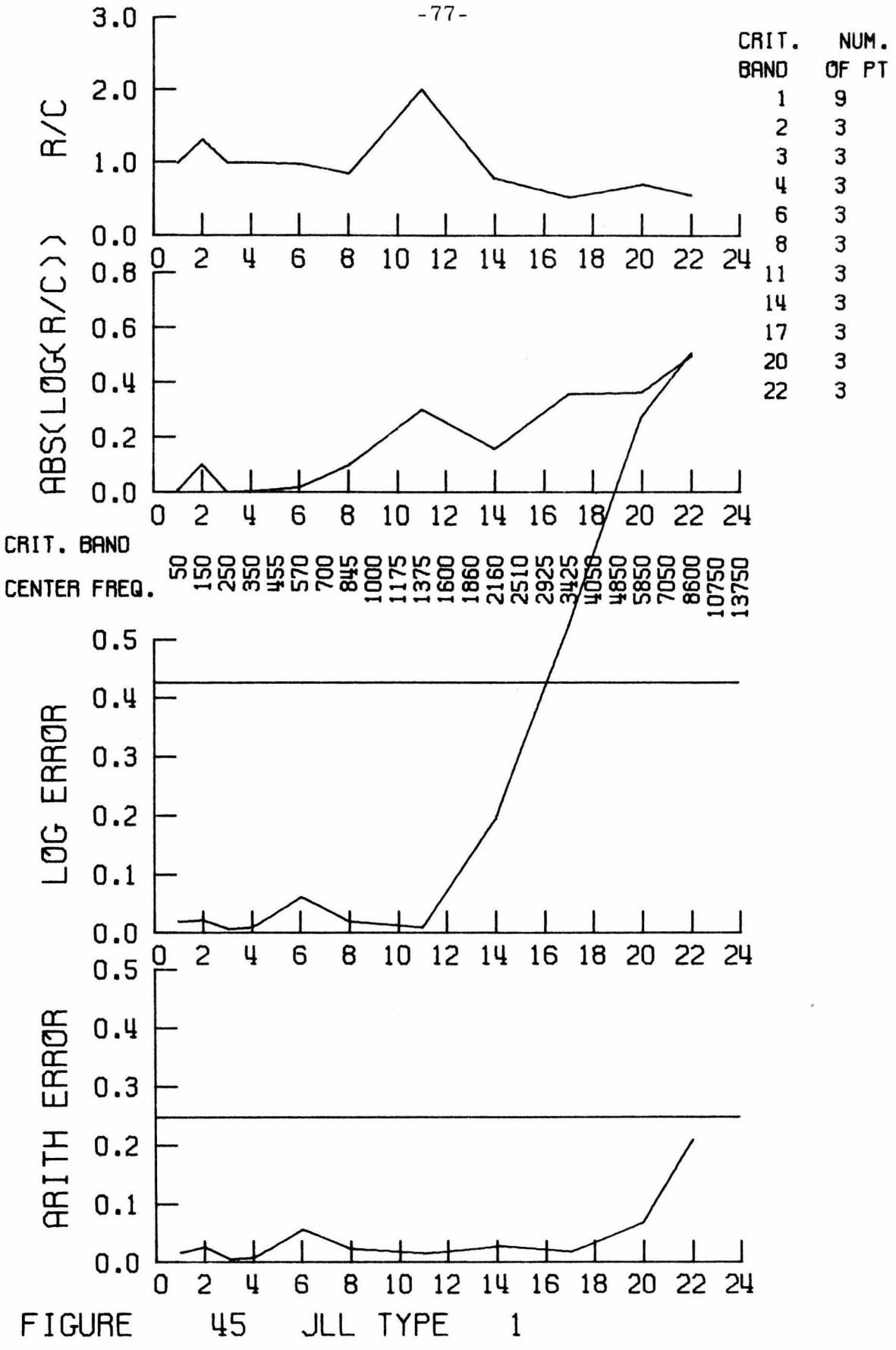



FIGURE 45 JLL TYPE 1

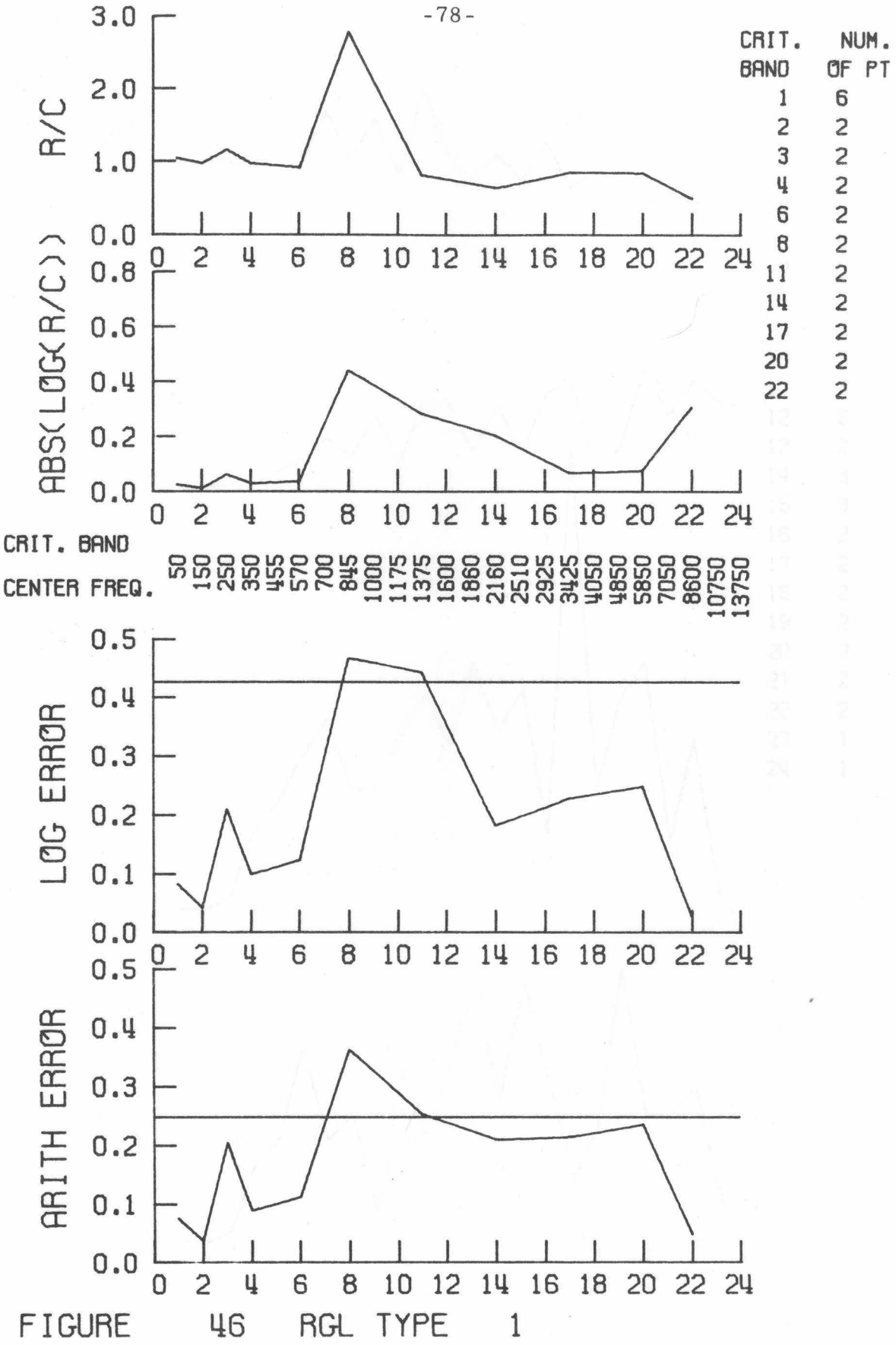



FIGURE 46 RGL TYPE 1

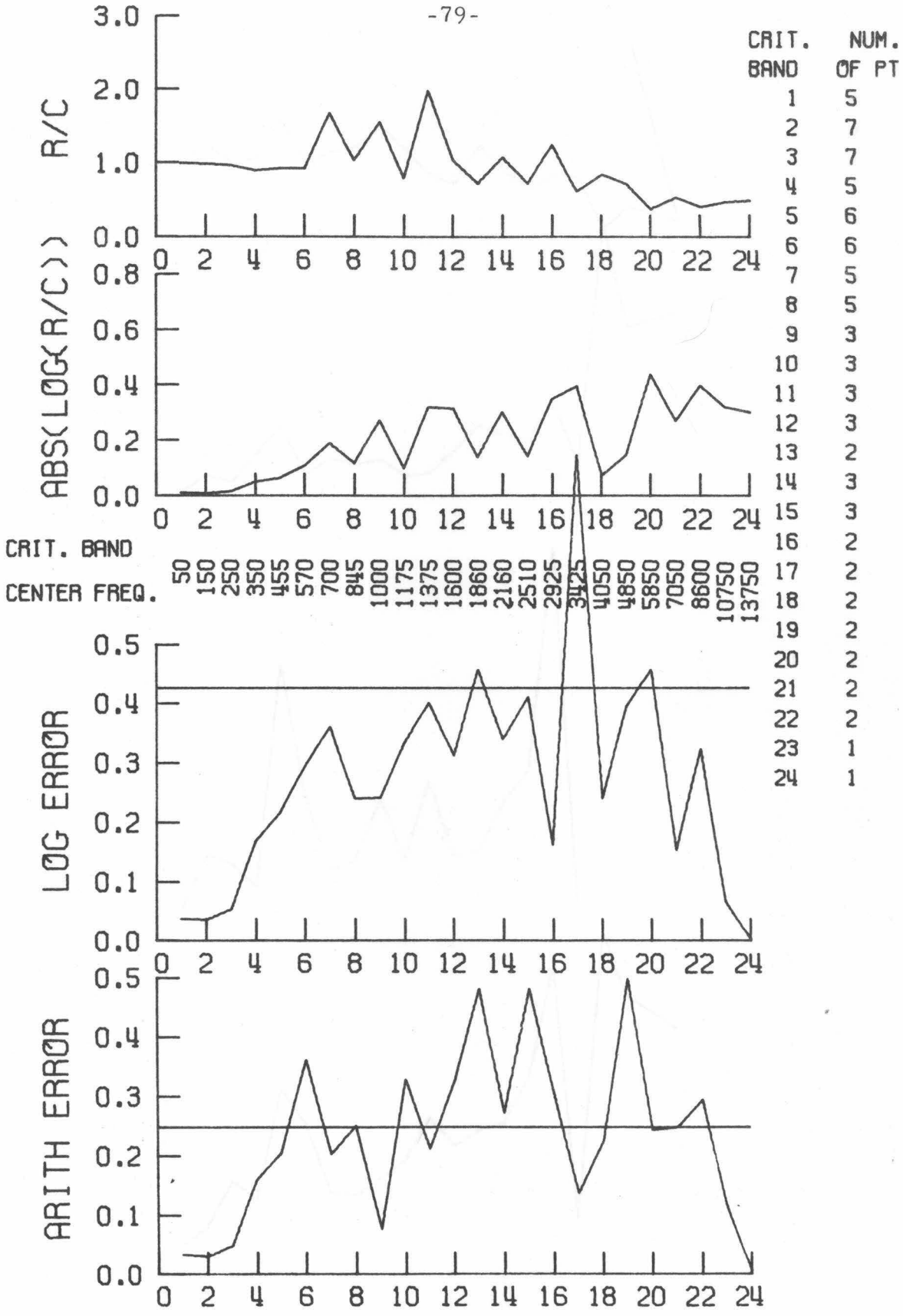



FIGURE 47 RAM TYPE 0

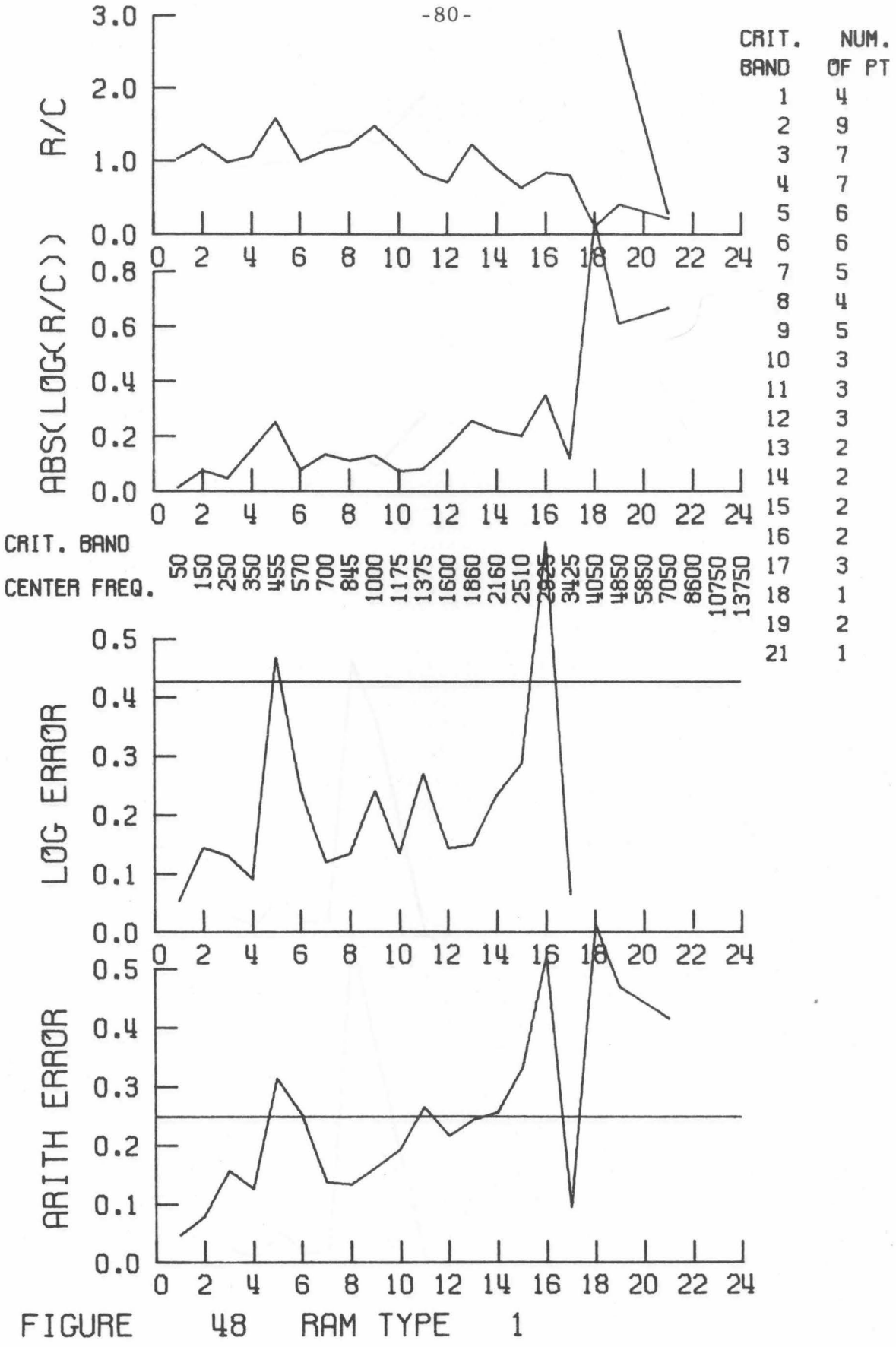
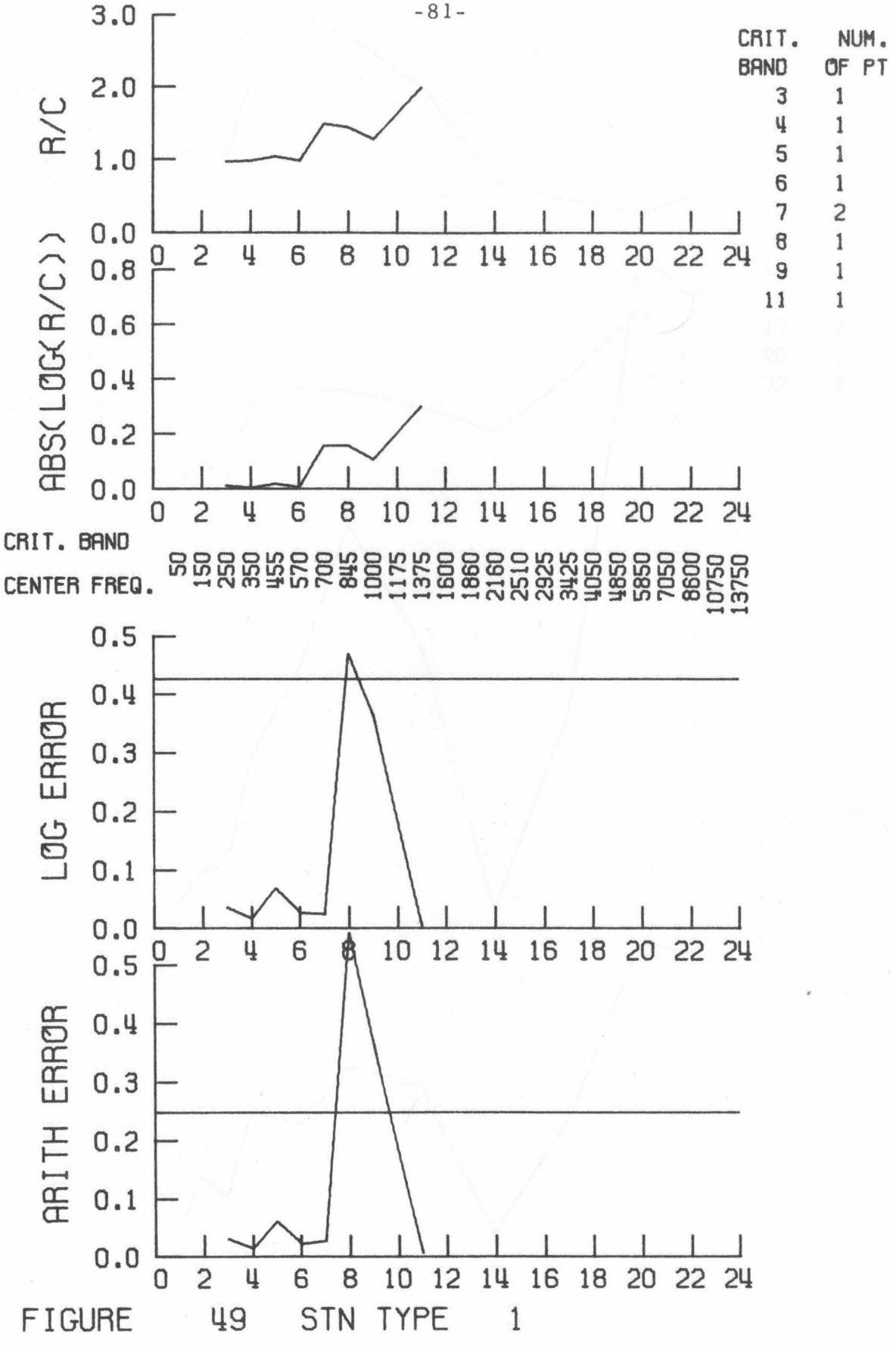



FIGURE 48 RAM TYPE 1




| CRIT. BAND | NUM. OF PT |
|---|---|
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 2 |
| 8 | 1 |
| 9 | 1 |
| 11 | 1 |

FIGURE 49 STN TYPE 1

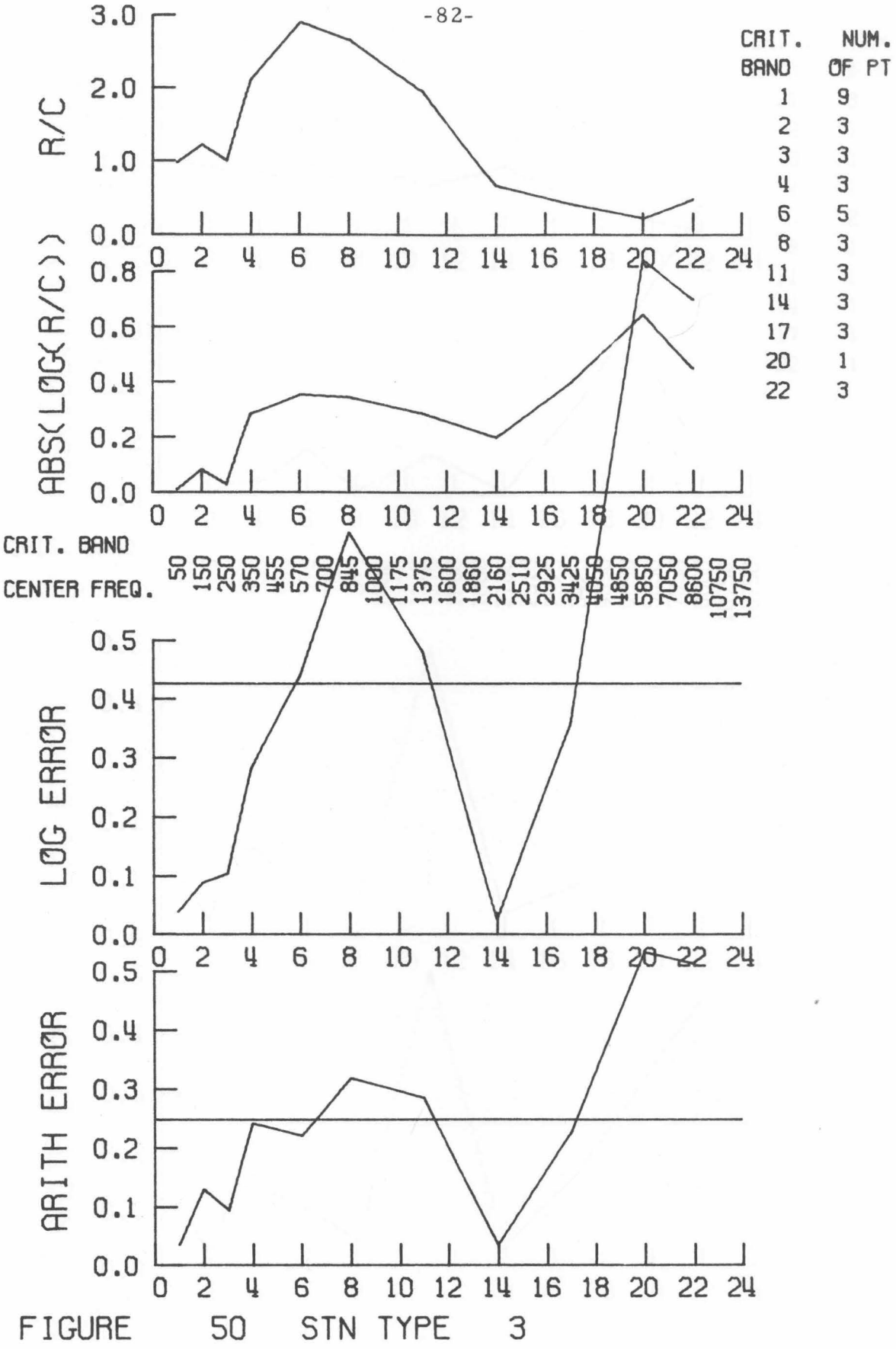



FIGURE 50 STN TYPE 3

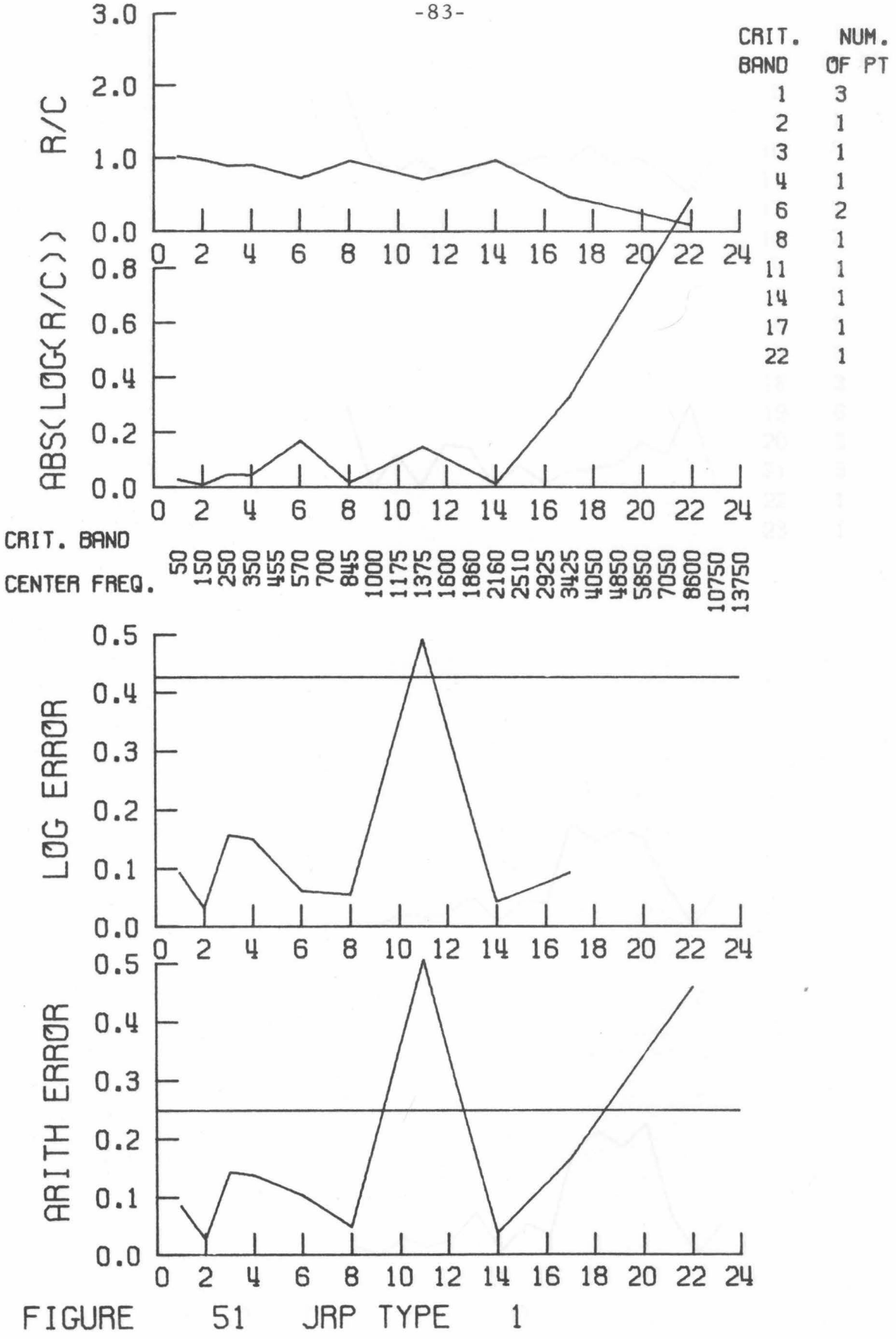



FIGURE 51 JRP TYPE 1

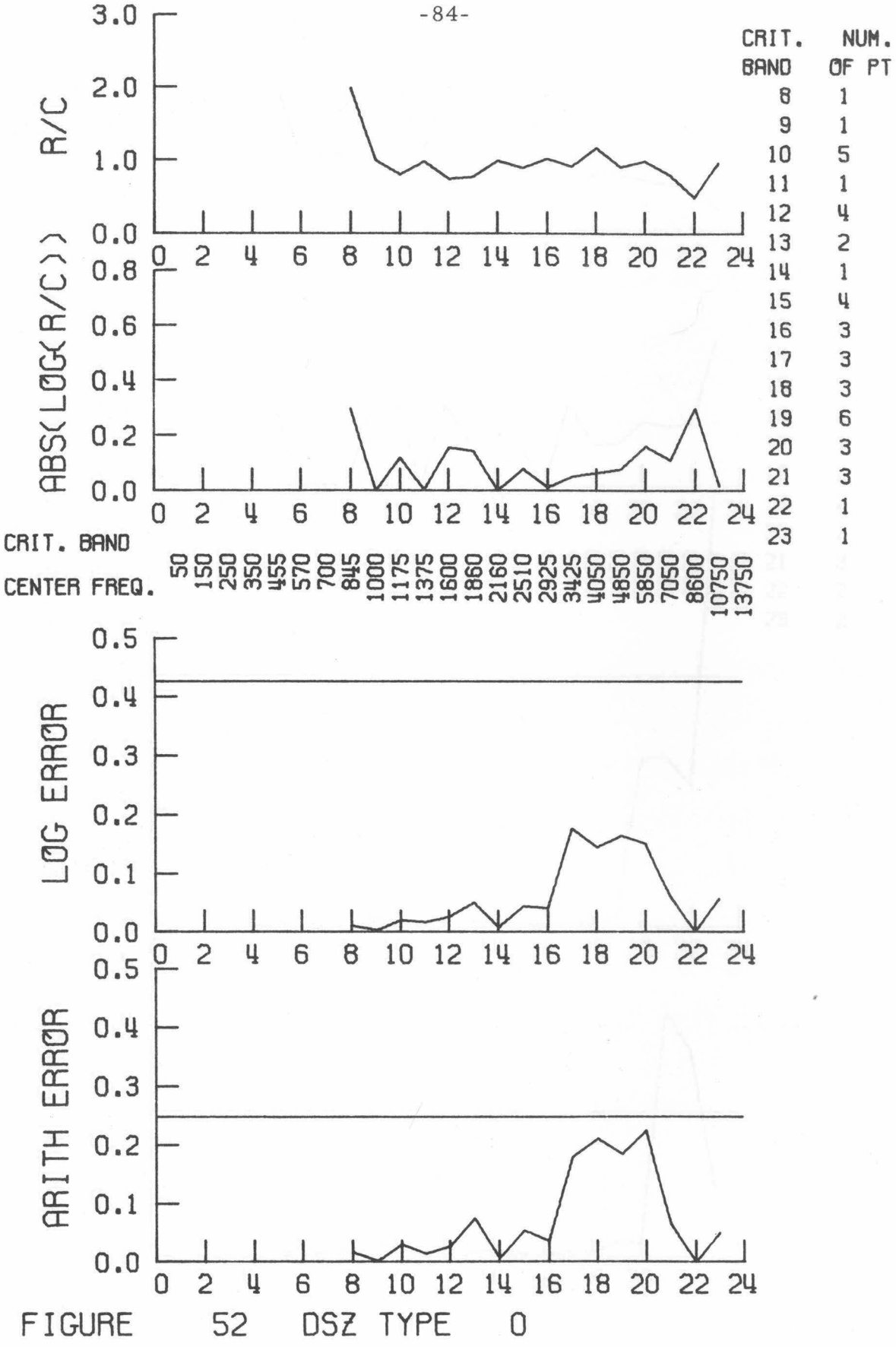



FIGURE 52 DSZ TYPE 0

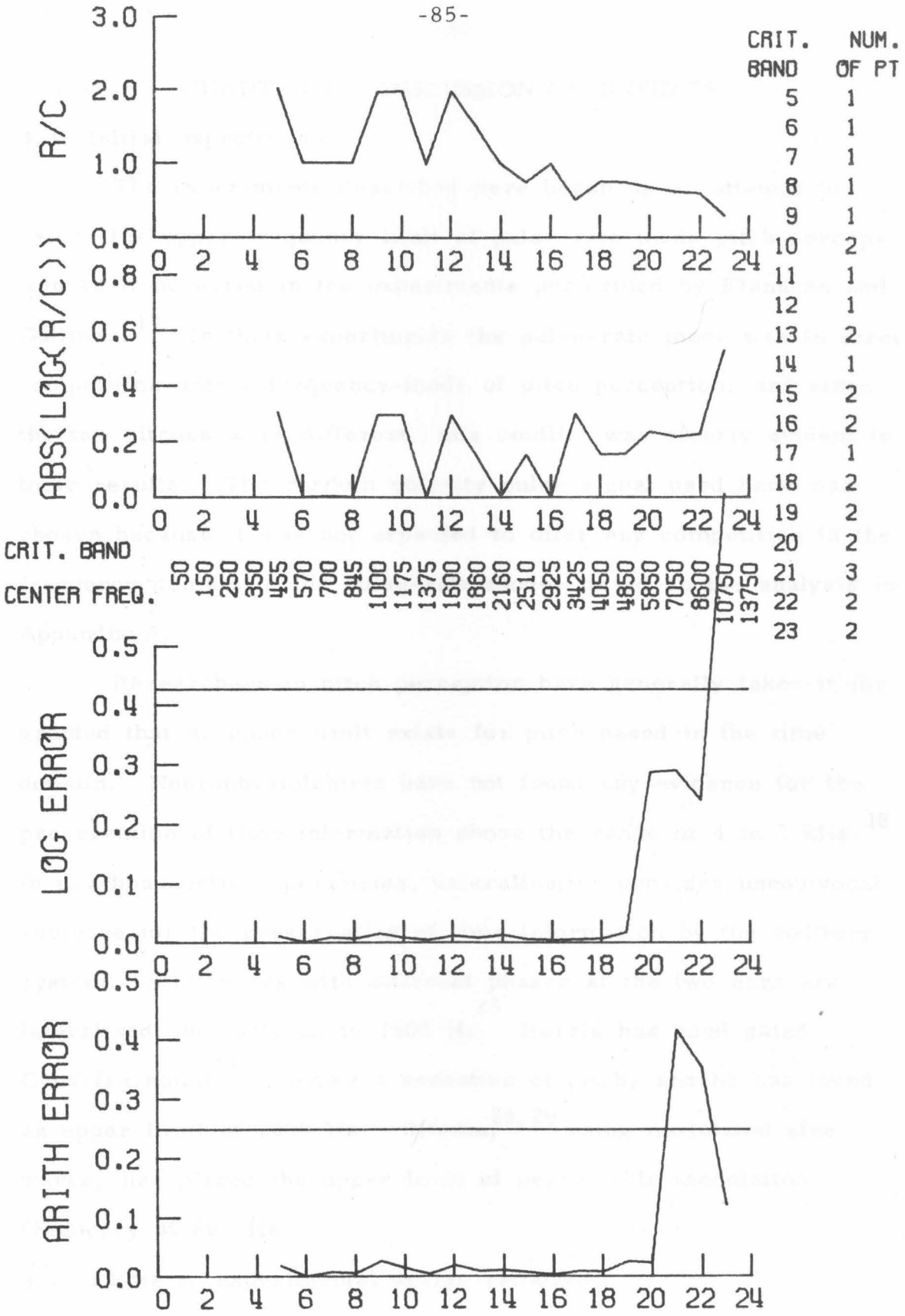



FIGURE 53 DSZ TYPE 1

# CHAPTER IV DISCUSSION OF RESULTS

## 4.1 Initial expectations

The experiments described here began as an attempt to extend the upper frequency limit of pulse-rate mode pitch perception as it occurred in the experiments performed by Flanagan and Guttman.[19] In their experiments the pulse-rate mode was in direct competition with a frequency-mode of pitch perception, and since the two pitches were different, this conflict was clearly evident in their results. The random polarity pulse signal used here was chosen because it was not expected to offer any competition in the frequency domain to the pulse-rate mode, based on the analysis in Appendix A.

Researchers in pitch perception have generally taken it for granted that an upper limit exists for pitch based in the time domain. Neurophysiologists have not found any evidence for the preservation of time information above the range of 4 to 5 kHz.[18] In psychoacoustic experiments, lateralization provides unequivocal evidence for the preservation of time information by the auditory system. Sine waves with different phases at the two ears are lateralized, but only up to 1500 Hz.[25] Harris has used gated Gaussian noise[24] to evoke a sensation of pitch, and he has found an upper limit of 2000 Hz. Ritsma,[28,29] using modulated sine waves, has placed the upper limit of perceivable modulation frequency at 800 Hz.

## 4.2 Flaws in experimental design removed

The results of the type 1 experiments show that some

subjects can hear a pitch in the random polarity pulse signal at pulse rates as high as 9 kHz. Since this frequency is so much higher than the limits found by other researchers, we must ask if something else is occurring in these experiments. The first place to look is the equipment used.

4.2.1 Equipment flaws

As indicated in Chapter II (see also Appendix E), the generation of positive/negative pulses must be handled with great care to insure identical pulse shapes and accurate times of occurrence. In addition, the low duty cycle of the signals used presented a problem in the aduio amplifier. Note that the power in a sine wave of peak voltage V is $V^2/2$. In a pulse train of 20 microsecond pulses at 100 Hz, if the pulse height is V, the power is $V^2(20/10,000)$ or $V^2/500$. Thus considerable dynamic range is required in the audio amplifier for pulse trains of even modest loudness. Even after these problems were solved, subjects were still able to match at the highest pulse repetition rates used.

4.2.2 Loudness disparity

Flanagan and Guttman[19] showed that the audibility of their periodic pulse trains was proportional to the RMS voltage for pulse rates above approximately 100 pulses per second, and approached a constant as the rate was lowered. They also found that the constant audibility at low rates was proportional to pulse energy. These results support the expected conclusion that the loudness of pulse trains is proportional to RMS voltage when the pulse rate is not very low.

The RMS voltage of the pulse train used in both our experiments and their experiments varies as the square root of the pulse rate. The possibility is then raised that subjects may have used the loudness of the signals as a clue to the pulse rate. The first objection in the use of loudness as a clue comes from anyone who has listened to these signals. The variations in loudness is too slight to allow the accuracy of matches we have found.

In both our experiments and Flanagan and Guttman's, some experiments were run with the subject in control of the amplitude of the matching signal. Flanagan and Guttman's intent was to investigate the loudness of their signals and they confirmed that loudness was proportional to RMS voltage. Our intent was different, as explained in Chapter II, but since changing the level of unipolar pulses had no effect on matching ability we may conclude that subjects do not use loudness as a clue to pulse rate. In addition, Flanagan and Guttman ran some experiments in which subjects were asked to specify which signal was louder. Their results again confirm the proportionality of loudness to RMS voltage and further demonstrate subjects' ability to match despite a loudness disparity.

## 4.3 Possible explanations of results

### 4.3.1 Pulse rate match

The possibility exists that the auditory system can assign pitch on the basis of time periodicities up to 9 kHz. Neurons from all places higher than the pulse rate would have to maintain phase-lock to the signal. Based on the animal data from Rose

et al.,[17] we would expect each neuron in the group to have pulse separation times of some varying multiple of the clock period, for example 100 microseconds, but greater than the recovery time, say 800 microseconds. The firings of the neurons would not be in phase due to the time delay of the traveling wave on the basilar membrane. Due to the filter function of the basilar membrane, neurons from places lower in frequency than the pulse rate would be firing with multiples of their own characteristic period. The central auditory system would have to infer the common period of the high frequency neurons and assign a pitch on that basis.

### 4.3.2 Short term spectrum anomalies

#### 4.3.2a Probability distribution anomalies

Another possible method which may be used by the auditory system in identifying the pulse rate is based on the differences in probability distribution of the amplitudes of different frequency components. As shown in Appendix D, if a bank of bandpass filters is excited by the random polarity pulse train, the peak amplitudes on the outputs of those filters tuned to multiples of 1/2 the pulse rate have a Gaussian distribution. All other filters have a Rayleigh distribution of peaks. It is conceivable that the central auditory system is capable of picking out those places on the basilar membrane which have the unusual distribution. This set of distinguished places would be identical to those that would be excited by a periodic signal whose fundamental frequency is half the pulse rate.

An informal experiment was conducted in the Spring of 1976 to investigate this possibility. A bandpass filter was constructed

with a center frequency of 3 kHz and an effective Q of approximately 100, corresponding to a 3 db bandwidth of 30 Hz. The filter was made by connecting three 2-pole active filter sections in series, each with a center frequency of 3 kHz and a Q of 50. The filter was driven by the random polarity pulse train, and the subject listened to the output through headphones. The subject attempted to tune the pulse rate to a frequency, f, such that $k(f/2) = 3$ kHz, by listening for a change in the quality of the output. All of the subjects, CMC, JRP and JLG, who tried the experiment succeeded easily. The tuning was quite precise for a wide range of k's satisfying the above equation. Different values of k were not noticeably distinguishable. The output had, of course, a pitch of 3 kHz, but the author had an impression of a "tinkling" in the background which disappeared when the pulse rate was properly tuned.

4.3.2b Zero crossing periodicities

However, this experiment raised questions of its own. It was quickly noticed during the course of the experiment that when the input was correctly tuned, the zero crossing on the oscilloscope were exactly periodic. This fact is derived analytically in Appendix D. The sound presented by the filter stimulates only a very narrow region of the basilar membrane. If the auditory system is capable of maintaining the phase information at 3 kHz, it may be that the exact periodicity of zero crossings is what enables subjects to tune the pulse rate.

A related experiment has been performed by Pollack[30] in

which he presented subjects with unipolar pulse trains whose inter-pulse spacing was randomly jittered. He tested under a wide variety of conditions, but the one most relevant to this discussion was the most sensitive. For signals which consisted of 128 pulses with an average pulse spacing of 0.3 milliseconds his subjects were able to detect the presence of jitters around 0.1% of the pulse spacing. From his data it appears that in choosing 3 kHz for the filter center frequency we chanced upon the most sensitive region.

Thus a possible explanation for subjects matching at high pulse rates is that the central auditory system can select those places of the basilar membrane for which the membrane zero crossings are exactly periodic.

The highest clock rate at which JLG matched near 1 was 9316 Hz. His response was 9365 Hz for a ratio of 1.00526. This explanation requires him to detect the lack of jitter in zero crossings at the 4658 Hz place, which is within the region for which Pollack reported detection of jitter on the order of .1%. If the auditory system could accomplish this, the pitch assignment could be done on the basis of place. However this performance requires the auditory system to achieve a precision in locating neural pulses on the order of .2 microseconds, a figure Pollack seemed reluctant to claim. He suggested that his subjects may have performed on the basis of spectral spread of the signals.

4.3.2c Spectral periodicities

Another possible spectral basis for the identification of pulse

rate in the random polarity pulse signal concerns the periodicity of the short term spectrum. As shown in Appendix D, the short term spectrum is periodic with the pulse rate. Because of random fluctuations in the signal, over a short term there will occur occasionally a peak at the pulse rate. When this occurs, there will also be peaks at all multiples of the pulse rate. The residue effect might then give a strong pitch at the pulse rate at that time. Because the spectrum is flat, other frequencies will also have peaks at other times, such that the time integrated power is the same for all frequencies. However for a short term peak at the clock rate minus e, C-e, peaks will also occur at nC-e. Since these are inharmonic the residue effect will not enforce the pitch at C-e and may in fact contribute to the pitch at C, due to the spectral spacing. Frequencies at submultiples of the clock, C/k, will also be reinforced by the residue effect. This may explain the tendency of subjects to match at submultiples at high pulse rates. This explanation also fits the subjective responses of the subjects, that the random polarity pulse train sounds like a series of short bursts of tones at various frequencies.

### 4.3.3 Internal nonlinearities

Finally we are left with the possibility that nonlinearities in the ear extract the pulse rate and provide a spectral peak for a place mechanism to find. The experiments were run at a sensation level of 30 db, which is standard in periodicity pitch experiments. As has been shown in Chapter I, this level produces no nonlinearities for sinusoidal signals. As previously mentioned,

pulsed signals are more severe with respect to level dependent nonlinearities. It should also be noted that the level of the signal was set with the pulse rate at 1 kHz. During the experiment it was varied over a range of approximately 100 Hz to 10 kHz. Since the power in the signal is proportional to the pulse rate, this results in the signal level varying between 25 db and 35 db. Of course, again due to the pulse nature of the signal, the peak voltage does not change. In the reduced level experiments, types 4 and 5, figures 43 and 44, JLG's performance did deteriorate at least above about 6 kHz.

4.4 Summary

In conclusion, we must admit that no single satisfying explanation for the results of the experiments exists. Five possible explanations for the ability to perform in these experiments have been offered: 1) pitch assignment on the basis of time periodicities, 2) selection of harmonically related places due to differing amplitude probability distribution, 3) selection of place due to lack of jitter, 4) pitch assignment based on periodicities in the spectrum, and 5) nonlinearities in the ear. The difficulties encountered here point up the general problems in separating the time and frequency domains, as so many researchers in pitch perception attempt to do. As has been shown, even in the case of a flat power spectrum, many subtleties exist in the frequency domain. These must be carefully dealt with whenever an experiment is designed to shed light on the place vs. time controversy. Previous workers in psychoacoustics appear to have completely

ignored these short-term effects when using "white spectrum" signals.

# APPENDIX A: DERIVATION OF SPECTRUM OF PULSE TRAINS OF RANDOM POLARITY

Define the random process $\underset{\sim}{x}(t)$ by

$$\underset{\sim}{x}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{a}_n \delta(t-nT) \tag{A1}$$

where $Pr[\underset{\sim}{a}_n = -1] = \frac{1}{2} = Pr[\underset{\sim}{a}_n = 1]$, and the $\underset{\sim}{a}_n$'s are statistically independent. If we apply $\underset{\sim}{x}(t)$ to the input of a linear system with impulse response h(t), the output is

$$\underset{\sim}{y}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{a}_n h(t-nT) \tag{A2}$$

The autocorrelation of the random process $\underset{\sim}{y}(t)$ is

$$R_{yy}(t+\tau, t) = E[\underset{\sim}{y}(t+\tau)\underset{\sim}{y}(t)]$$

$$= E \sum_{n=-\infty}^{\infty} \underset{\sim}{a}_n h(t+\tau-nT) \sum_{m=-\infty}^{\infty} \underset{\sim}{a}_m h(t-mT) \tag{A3}$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} E[\underset{\sim}{a}_n \underset{\sim}{a}_m] h(t+\tau-nT)\, h(t-mT)$$

where

$$E[\underset{\sim}{a}_n \underset{\sim}{a}_m] = E[\underset{\sim}{a}_n]\, E[\underset{\sim}{a}_m] = 0 \quad \text{for } n \neq m \tag{A4a}$$

since the $\underset{\sim}{a}_n$'s are independent, and

$$E[\underset{\sim}{a}_n \underset{\sim}{a}_m] = E[\underset{\sim}{a}_n^2] = 1 \qquad \text{for } n = m \tag{A4b}$$

Thus,

$$R_{yy}(t+\tau, t) = \sum_{n=-\infty}^{\infty} h(t+\tau-nT)\, h(t-nT) \tag{A5}$$

If h(t) is time-limited such that $h(t) = 0$ for $t < 0$ or $t > \Delta < T$, then $R_{yy}(t+\tau, t) = 0$ for $|\tau| > \Delta$, since for each n, one of the two

arguments in $h(t+\tau-nT)\,h(t-nT)$ must be outside the range where $h(x)$ is nonzero. For $n = 0$, the second function limits $R_{yy}$ to $0 \le t \le \Delta$, and for each value of t in this range, there is a range of $\tau$ of length $\Delta$ for which $R_{yy}$ may be nonzero. This yields the parallelogram to the right of the origin as shown in figure A1. Different

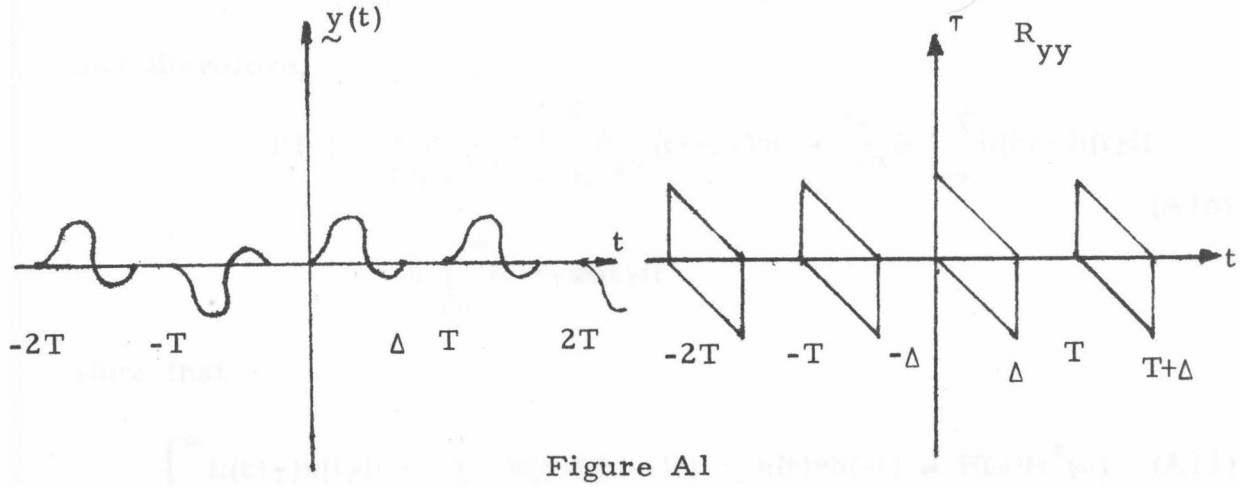



Figure A1

values of n translate the parallelogram by nT. The requirement that $\Delta < T$ separates these regions as shown in figure A1. Thus $R_{yy}$ is bounded if $h(t)$ is bounded. But $R_{yy}$ is not integrable since it extends infinitely far in each direction along the t axis. If we allow $\Delta$ to be greater than T, then the parallelograms will overlap. However, $R_{yy}$ will remain bounded as $h(t) \to 0$ as $t \to \infty$ sufficiently rapidly.

We now form the integral

$$R(\tau) = \lim_{L\to\infty} \frac{1}{2L}\int_{-L}^{L} R_{yy}(t+\tau, t)\, dt \qquad (A6)$$

For $\Delta < L < T$,

$$\int_{-L}^{L} R_{yy}(t+\tau, t)dt = \int_{-\infty}^{\infty} h(t+\tau)\,h(t)dt = \int_{0}^{\Delta} h(t+\tau)\,h(t)dt \qquad (A7)$$

and for $\Delta+T < L < 2T$,

$$\int_{-L}^{L} R_{yy}(t+\tau, t)dt = 3\int_{-\infty}^{\infty} h(t+\tau)\,h(t)dt \tag{A8}$$

In general, for $(n-1)T+\Delta < L < nT$,

$$\int_{-L}^{L} R_{yy}(t+\tau, t)dt = (2n-1)\int_{-\infty}^{\infty} h(t+\tau)\,h(t)dt \tag{A9}$$

and therefore,

$$\begin{aligned} R(\tau) &= \lim_{L\to\infty} \frac{1}{2L}\int_{-L}^{L} R_{yy}(t+\tau, t)dt \to \frac{2n-1}{2nT}\int_{-\infty}^{\infty} h(t+\tau)h(t)dt \\ &= \frac{1}{T}\int_{-\infty}^{\infty} h(t+\tau)h(t)dt \end{aligned} \tag{A10}$$

Note that

$$\int_{-\infty}^{\infty} h(t+\tau)h(t)dt = -\int_{\infty}^{-\infty} h(\tau-t)h(-t)dt = h(t)*h(-t) \leftrightarrow H(\omega)H^{*}(\omega) \tag{A11}$$

Thus,

$$R(\tau) \leftrightarrow \frac{1}{T}\left|H(\omega)\right|^{2}$$

The average power spectrum of the signal $\underset{\sim}{y}(t)$ is of the same shape as the linear system. In this sense, we may regard the input process, $\underset{\sim}{x}(t)$, as having a flat power spectrum.

# APPENDIX B: FEEDBACK SHIFT REGISTERS

The integers modulo 2 form a field GF(2). Addition and multiplication are shown in Table B1. We may form a polynomial,

| + | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

| X | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

Table B1

P(x), of degree n, over GF(2) by

$$P(x) = c_n x^n + c_{n-1} x^{n-1} + \ldots + c_0 \qquad (B1)$$

where each $c_i$ is either 0 or 1. The polynomials over GF(2) form a ring where the powers of the variables are handled in the usual way and the arithmetic of the coefficients of each power is that of GF(2). (See Peterson, Ref. 31)

A polynomial, P(x), is defined as irreducible if and only if it cannot be expressed as the product of two polynomials of lesser degree. Given an irreducible polynomial of degree n, P(x), the set of polynomials modulo P(x) forms a field. This field is denoted $GF(2^n)$. Any polynomial of degree greater than or equal to n is congruent to one of degree less than n. This can be seen by

$$P(x) = x^n + c_{n-1} x^{n-1} + \ldots + c_0 \equiv 0 \mod P(x)$$

$$x^n \equiv \left[ c_{n-1} x^{n-1} + \ldots + c_0 \right] \equiv c_{n-1} x^{n-1} + \ldots + c_0 \qquad (B2)$$

Note that addition and subtraction are the same in GF(2). Thus for m > n,

$$Q(x) = d_m x^m + \ldots + d_0 \equiv d_m x^{m-n}(c_{n-1}x^{n-1} + \ldots + c_0) + d_{m-1}x^{m-1} + \ldots + d_0$$

$$\equiv (d_m c_{n-1} + d_{m-1})x^{m-1} + (d_m c_{n-2} + d_{m-2})x^{m-2} + \ldots + d_0 \qquad \text{(B3)}$$

A feedback shift register as shown in figure B1 performs multiplication by x in $GF(2^n)$. The content of each flip-flop is

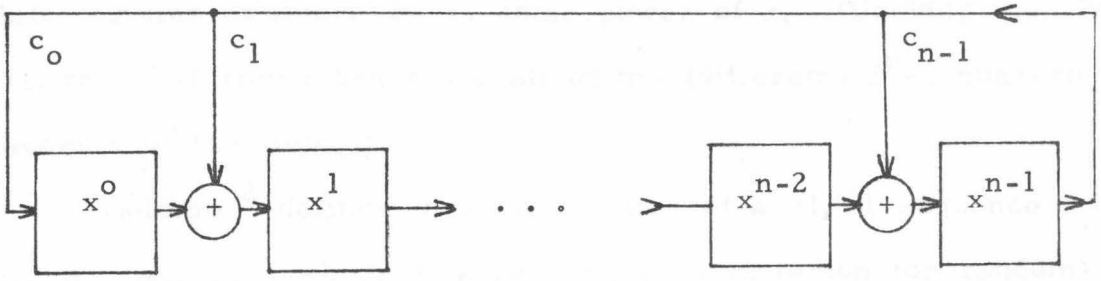



Figure B1

regarded as the coefficient of the indicated power of x of a polynomial of degree n-1 or less. Thus, the shift register may contain any member of $GF(2^n)$. If the coefficient of $x^{n-1}$ is 0, clocking the shift register moves each bit to the right, effectively multiplying by $x^1$. If the coefficient of $x^{n-1}$ is 1, the $x^n$ result is added in as $c_{n-1}x^{n-1} + \ldots + c_0$ according to the above reduction process.

Let $\alpha$ be a member of $GF(2^n)$ and consider the sequence $\alpha^0, \alpha^1, \alpha^2, \ldots$ . Since $GF(2^n)$ is a field, each member of the sequence is in $GF(2^n)$, and since $GF(2^n)$ is finite, there must be some k such that $\alpha^k = \alpha^0 = 1$. $\alpha^j$ cannot be 0 unless $\alpha$ is 0, and since $GF(2^n)$ contains $2^n$ members, the maximum possible k is

$2^{n-1}$. A member of $GF(2^n)$, $\alpha$, for which the least value of k such that $\alpha^k = 1$ is $k = 2^n-1$ is called primitive. Thus the powers of a primitive element range through all of the nonzero members of $GF(2^n)$.

An irreducible polynomial, P(x), of degree n is called primitive if x is a primitive element of the field modulo P(x). Since every element of the field is a power of x, any state of the shift register of figure B1 is some power of x. Clocking the shift register $2^n-1$ times generates all of the (different) $2^n-1$ nonzero elements of the field.

Golomb[32] defines three properties of a +1, -1 sequence $(a_0, a_1, a_2, \ldots)$ which may be used as a criterion for randomness: R1, R2, and R3. For a periodic sequence, R1 states that in any period the number of +1's is nearly equal to the number of -1's. (This may be taken to mean $\left|\sum_{n=0}^{p} a_n\right| \leq 1$.) R2 states that in every period, half of the runs have length one, one-fourth have length two, etc. Also, of the runs of each length, half are +1's and half are -1's. R3 is a condition on the autocorrelation,

$$pC(\tau) = \sum_{n=1}^{p} a_n a_{n+\tau} = \begin{cases} p, & \text{if } \tau = 0 \\ k, & \text{if } 0 < \tau < p \end{cases} \qquad \text{(B4)}$$

Golomb uses a different shift register configuration from the one used in this research. However, the two can be shown equivalent as follows. Golomb's shift register is shown in figure B2. Clearly each stage is a delayed copy of the first stage. If $(a_0, a_1, a_2, \ldots)$ is the history of the first stage,

$$a_n = \sum_{i=1}^{r} b_i a_{n-1} \qquad \text{(B5)}$$

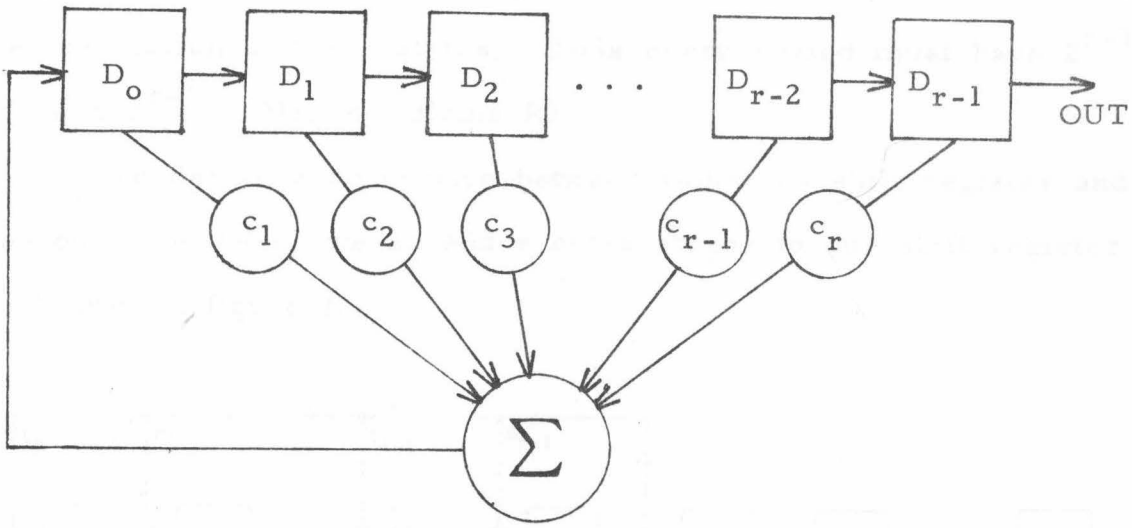



Figure B2

The characteristic polynomial of the sequence $(a_n)$ is defined by

$$f(x) = 1 - \sum_{i=1}^{r} b_i x^i \qquad \text{(B6)}$$

Golomb proves that the period, p, of the sequence $(a_n)$ is the smallest positive integer such that f(x) divides $x^p+1$ (Theorem 2.3, p. 32). Peterson (p. 103) shows that an irreducible polynomial, f(x) of degree m is primitive if and only if it divides $x^p-1$ for no p less than $2^m-1$. Thus a maximal length sequence is produced if and only if f(x) is primitive.

Golomb shows that a shift register which produces a maximum length sequence satisfies the three randomness criteria. R1 will be shown here to give the flavor of the arguments. R2 and R3 are omitted since they involve longer combinatorial

arguments. For R1, we note that if the shift register output has period $2^r-1$, the register must go through all $2^r-1$ possible nonzero states. The output may be regarded as the 1's digit of the binary representation of these states. Thus every period must have $2^{r-1}$ 1's and $2^{r-1}-1$ 0's, satisfying R1.

To derive a connection between Golomb's shift register and the one used here, we append r extra stages to our shift register as shown in figure B3.

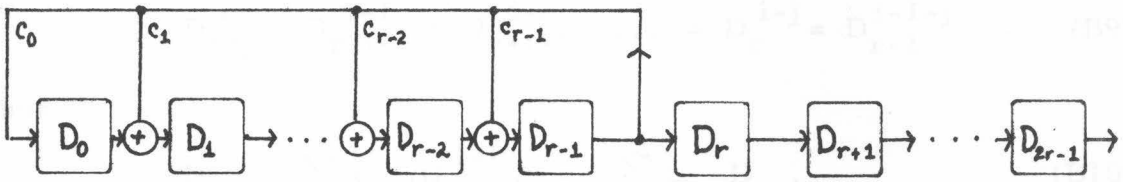



Figure B3

$D_j^i$ will denote the contents of stage $D_j$ at time i. Then,

$$D_j^i = c_j D_{r-1}^{i-1} + D_{j-1}^{i-1} \quad \text{for } 0 \le j \le 2r-1 \tag{B7}$$

where

$$D_{-1} = 0 \quad \forall i;\ c_j = 0\ ;\ r \le j \le 2r-1$$

Furthermore,

$$P(x) = x^r + \sum_{i=0}^{r-1} c_i x^i$$

is the primitive polynomial used in our shift register. It can be shown that $D_r, \ldots, D_{2r-1}$ forms a maximum length shift register of Golomb's type as follows:

$$D_r^i = D_{r-1}^{i-1} = c_{r-1}D_{r-1}^{i-2} + D_{r-2}^{i-2}$$

$$= c_{r-1}D_{r-1}^{i-2} + c_{r-2}D_{r-1}^{i-3} + D_{r-3}^{i-3}$$

$$= c_{r-1}D_{r-1}^{i-2} + c_{r-2}D_{r-1}^{i-3} + c_{r-3}D_{r-1}^{i-4} + D_{r-4}^{i-4}$$

$$= c_{r-1}D_{r-1}^{i-2} + \ldots + c_{r-1}D_{r-1}^{r-(r+1)}$$

$$= \sum_{j=1}^{r} c_{r-j}D_{r-1}^{i-1-j} + 0 \qquad \text{(B8)}$$

Also,

$$D_{r+j}^i = D_{r+j-1}^{i-1} = D_{r+j-2}^{i-2} = \ldots = D_r^{i-j} = D_{r-1}^{i-1-j} \qquad \text{(B9)}$$

Thus

$$D_n^i = \sum_{j=1}^{r} c_{r-j}D_{r+j-1}^{i-1} = \sum_{j=1}^{r} c_{r-j}D_{r+j} \qquad \text{(B10)}$$

which is a linear function of the following stages. To convert to Golomb's notation, $a_n$ is the output of the first ($D_r$) stage at time n. Then,

$$a_n = D_r^n = \sum_{j=1}^{r} c_{r-j}D_r^{n-j} = \sum_{j=1}^{r} c_{r-j}a_{n-j}$$

$$= \sum_{j=1}^{r} b_j a_{n-j} \quad , \text{ where } b_j = c_{r-j} \qquad \text{(B11)}$$

Peterson (problem 6.7, p. 106) defines the reciprocal polynomial, $P^*(x) = x^r P(x^{-1})$. He states that P(x) is primitive if and only if $P^*(x)$ is primitive. Finally, we note

$$P^*(x) = x^r(x^{-r} + \sum_{j=0}^{r-1} c_j x^{-j}) = 1 + \sum_{j=0}^{r-1} c_j x^{r-j}$$

$$= 1 + \sum_{i=r}^{1} c_{r-i}x^{r-(r-i)} = 1 + \sum_{i=1}^{r} b_i x^i = f(x) \qquad \text{(B12)}$$

Thus we see that a primitive polynomial P(x) in our shift register produces the same sequence that would be produced if its reciprocal polynomial f(x) = P*(x) were used as the characteristic polynomial in Golomb's shift register. Furthermore, any primitive polynomial, P(x), in our shift register produces an output sequence with the three above mentioned randomness properties.

## APPENDIX C: SPECTRUM OF RANDOM PULSE PATTERNS

Let $Y = \left[y_{ij}\right]$ be a $\pm 1$ matrix of size 2M by k. We form a random pulse process by selecting a k-bit pattern from the 2M patterns specified by Y. Let

$$(n) = n \bmod k\ , \quad \left[n\right] = \frac{n-(n)}{k} \tag{C1}$$

Then

$$\underset{\sim}{y}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{y}_n h(t-nT) \tag{C2}$$

where

$$\underset{\sim}{y}_n = y_{\underset{\sim}{i}_{[n]},\,(n)}$$

and

$$\Pr\left[\underset{\sim}{i}_n = i\right] = p_i \quad \text{for } 0 \le i \le 2M-1 \tag{C3}$$

where $\underset{\sim}{i}_n$ is independent from $\underset{\sim}{i}_m$ for $m \neq n$. For reasons that will appear shortly, we require that for each pattern in Y, the negative is also in Y and with the same probability. That is,

$$y_{i,\,j} = -y_{i+M,\,j} \quad \text{and} \quad p_i = p_{i+M} \tag{C4}$$

for $0 \le i \le M-1$ and $0 \le j \le k-1$ .

As in Appendix A, we form the autocorrelation function

$$R_{yy}(t+\tau,\,t) = E\left[\underset{\sim}{y}(t)\underset{\sim}{y}(t+\tau)\right] = E\left[\sum_{i=-\infty}^{\infty} \underset{\sim}{y}_i h(t-iT) \sum_{j=-\infty}^{\infty} \underset{\sim}{y}_j h(t+\tau-jT)\right]$$

$$= E\left[\sum_{p=-\infty}^{\infty} \sum_{m=0}^{k-1} \underset{\sim}{y}_{pk+m} h(t-pkT-mT) \sum_{q=-\infty}^{\infty} \sum_{n=0}^{k-1} \underset{\sim}{y}_{qk+n} h(t+\tau-qkT-nT)\right]$$

$$= \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} E\left[\underset{\sim}{y}_{pk+m} \underset{\sim}{y}_{qk+n}\right] h(t-pkT-mT)h(t+\tau-qkT-nT) \tag{C5}$$

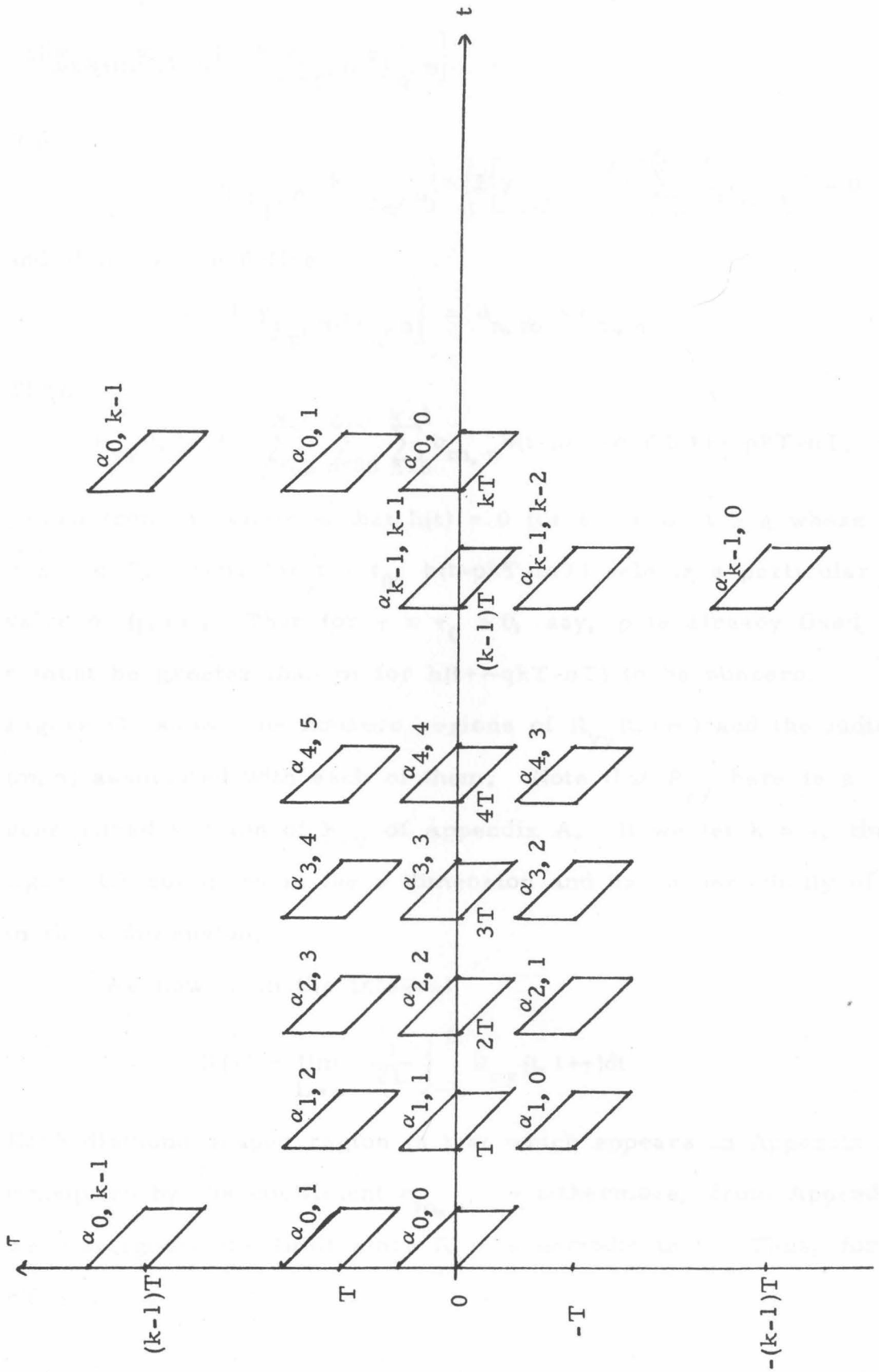



FIGURE C1: $R_{yy}(t, t+\tau)$

$$E\left[\underset{\sim}{y}_{pk+m}\underset{\sim}{y}_{qk+n}\right] = E\left[\underset{\sim}{y}_{i_p,\,m}\underset{\sim}{y}_{i_q,\,n}\right]$$

If $p \neq q$,

$$= E\left[\underset{\sim}{y}_{i_p,\,m}\right]E\left[\underset{\sim}{y}_{i_q,\,n}\right] = \left(E\left[\underset{\sim}{y}_{i_p,\,m}\right]\right)^2 = \left(\sum_{i=0}^{2m-1} p_i y_{i,\,m}\right)^2 = 0 \quad \text{(C6)}$$

and if $p = q$, we define

$$E\left[\underset{\sim}{y}_{i_p,\,m}\underset{\sim}{y}_{i_q,\,n}\right] = \alpha_{n,\,m} = \alpha_{m,\,n}$$

Thus,

$$R_{yy}(t,\, t+\tau) = \sum_{p=-\infty}^{\infty}\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}\alpha_{m,\,n}h(t-pkT-mT)h(t+\tau-pkT-nT) \quad \text{(C7)}$$

Recall from Appendix A that $h(t) = 0$ for $t < 0$ or $t \geq \Delta$ where $0 \leq \Delta \leq T$. Thus for $t = t_0$, $h(t-pkT-mT)$ selects a particular value of $(p, m)$. Then for $\tau = \tau_0 \geq 0$, say, $p$ is already fixed, so $n$ must be greater than $m$ for $h(t+\tau-qkT-nT)$ to be nonzero. Figure C1 shows the nonzero regions of $R_{yy}(t, t+\tau)$ and the indices $(m, n)$ associated with each of them. Note that $R_{yy}$ here is a generalized version of $R_{yy}$ of Appendix A. If we let $k = 1$, then figure C1 collapses in the $\tau$ dimension and has a periodicity of 1 in the t dimension.

We now form the integral

$$R(\tau) = \lim_{L\to\infty}\frac{1}{2L}\int_{-L}^{L} R_{yy}(t,\, t+\tau)dt \quad \text{(C8)}$$

Each diamond shaped region is that which appears in Appendix A, multiplied by the coefficient $\alpha_{m,\,n}$. Furthermore, from Appendix A, we may ignore the limit since $R_{yy}$ is periodic in t. Thus, for $nT-\Delta \leq \tau \leq nT+\Delta$,

$$R(\tau) = \frac{1}{kT}\sum_{m=0}^{k-1-n} \alpha_{m,\,m+n} \int_{-\infty}^{\infty} h(t+\tau-nT)\,h(t)\,dt$$

$$= E\left[\frac{1}{k}\sum_{m=0}^{k-1-n} y_{\underset{\sim}{i}_p,\,m}\, y_{\underset{\sim}{i}_p,\,m+n}\right]\frac{1}{T}\int_{-\infty}^{\infty} h(t+\tau-nT)\,h(t)\,dt \qquad \text{(C9)}$$

We now define the correlation of the sequence $\{y_n\}$ to be

$$C(\ell) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} E\left[\underset{\sim}{y}_n \underset{\sim}{y}_{n+\ell}\right] \qquad \text{(C10)}$$

and $R_i(\ell)$

$$R_i(\ell) = \sum_{m=0}^{k-1-\ell} y_{i,\,m}\, y_{i,\,m+\ell} \qquad \text{(C11)}$$

is the correlation of pattern i.

$$C(\ell) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} E\left[\underset{\sim}{y}_n \underset{\sim}{y}_{n+\ell}\right]$$

$$= \lim_{h\to\infty}\frac{1}{2hk+1}\sum_{p=-h}^{h}\sum_{m=0}^{k-1} E\left[\underset{\sim}{y}_{pk+m}\underset{\sim}{y}_{pk+m+\ell}\right] \qquad \text{(C12)}$$

$$= \lim_{h\to\infty}\frac{1}{2hk+1}\sum_{p=-h}^{h}\sum_{m=0}^{k-1} E\left[y_{\underset{\sim}{i}_p,\,m}\, y_{\underset{\sim}{i}_{pk+m+\ell}\,,\,(m+\ell)}\right]$$

For $\ell \geq k$, $[pk+m+\ell] > p$, thus

$$E\left[y_{\underset{\sim}{i}_p,\,m} y_{\underset{\sim}{i}_{pk+m+\ell}\,,\,(m+\ell)}\right] = E\left[y_{\underset{\sim}{i}_p,\,m}\right] E\left[y_{\underset{\sim}{i}_{p'},\,(m+\ell)}\right] = 0$$

thus $$C(\ell) = 0$$

and for $\ell < k$,

$$C(\ell) = \lim_{h\to\infty}\frac{1}{2hk+1}\sum_{p=-h}^{h}\left(\sum_{m=0}^{k-1-\ell} E\left[y_{\underset{\sim}{i}_p,\,m} y_{\underset{\sim}{i}_p,\,m+\ell}\right] + \sum_{m=k-\ell}^{k-1} E\left[y_{\underset{\sim}{i}_p,\,m} y_{\underset{\sim}{i}_{p+1},\,m+\ell}\right]\right)$$

The last summation is zero again due to the independence of $\underset{\sim}{i}_p$ and $\underset{\sim}{i}_{p+1}$.

$$C(\ell) = \lim_{h\to\infty} \frac{1}{2hk+1} \sum_{p=-h}^{h} \sum_{m=0}^{k-1-\ell} \sum_{i=0}^{2M-1} p_i y_{i,m} y_{i,m+\ell}$$

$$= \lim_{h\to\infty} \frac{1}{2h} \; \frac{1}{k+\frac{1}{2h}} (2k+1) \sum_{i=0}^{2M-1} p_i \sum_{m=0}^{k-1-\ell} y_{i,m} y_{i,m+\ell} \qquad \text{(C1 )}$$

$$= \sum_{i=0}^{2M-1} p_i \frac{1}{k} \sum_{m=0}^{k-1-\ell} y_{i,m} y_{i,m+\ell} = E\left[\frac{1}{k} R_i(\ell)\right]$$

Thus,

$$R(\tau) = C(\ell) \frac{1}{T} \int_{-\infty}^{\infty} h(t+\tau-\ell T)\, h(t)\, dt$$

where $\ell T - \Delta \le \tau \le \ell T + \Delta$.

If we arrange Y so that $C(\ell) = 0$ for $\ell \neq 0$, then $R(\tau) = h(t)*h(-t)$ as in Appendix A, and we have that the spectrum of random pulse patterns is flat in the same sense as the spectrum of random polarity pulses.

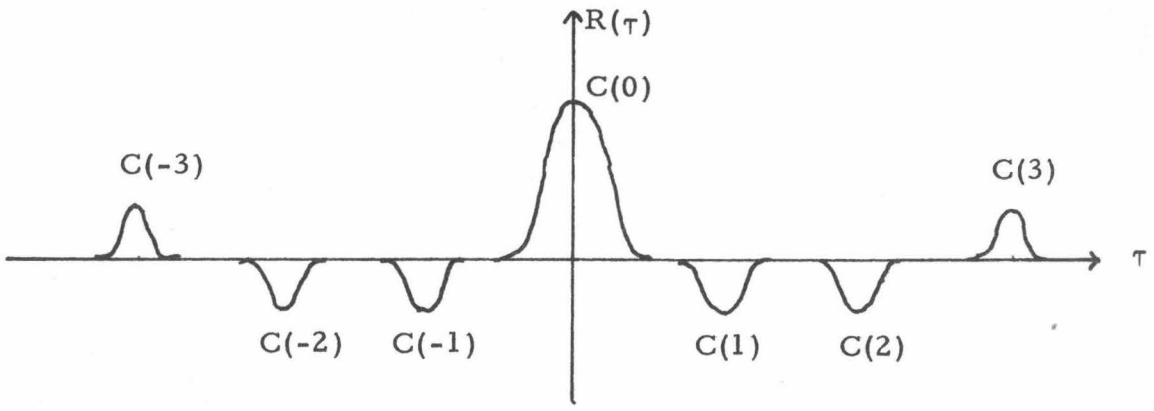



Figure C2

For the sake of completeness, we continue the development with an arbitrary $C(\ell)$. Define

$$f(\tau) = h(t)*h(-t) \qquad \text{(C1 )}$$

then,

$$R(\tau) = \sum_{\ell=-k}^{k} C(\ell) f(\tau - \ell T) \qquad (C1\ )$$

and

$$R(\tau) \leftrightarrow \sum_{\ell=-k}^{k} C(\ell) e^{-j\omega \ell T} |H(j\omega)|^2$$

$$= |H(j\omega)|^2 \left[ 1 + \sum_{\ell=1}^{k} C(\ell)\ e^{j\omega \ell T} + e^{-j\omega \ell T} \right] \qquad (C1\ )$$

since $C(\ell) = C(-\ell)$

$$= |H(j\omega)|^2 \left[ 1 + 2 \sum_{\ell=1}^{k} C(\ell) \cos \omega \ell T \right]$$

# APPENDIX D: AMPLITUDE DISTRIBUTIONS OF FILTERED PULSE TRAINS

## D.1 Periodicity of short-term spectrum

The short-term spectrum of a signal f(t) may be defined by

$$X(\omega, t_0) = \int_{-\infty}^{\infty} w(t-t_0)\, f(t)\, e^{-j\omega t}\, dt \qquad (D1)$$

where w(t) is a window function. The window function is specified to reflect two properties of a realizable spectrum analyzer. The first property is causality, i.e., at time $t_0$, $X(\omega, t_0)$ can depend only on values of f(t) for $t \leq t_0$. The second property is that values of f(t) in the remote past should have little or no effect on the present values of $X(\omega, t_0)$, which requires $w(t) \to 0$ as $t \to -\infty$.

If f(t) is a train of impulses, then f(t) = 0 for $t \neq kT$. If we replace $\omega$ by $\omega + 2\pi/T$ in equation D1, the right side is not affected since f(t) is nonzero only for t = kT and

$$e^{-j(\omega + 2\pi/T)kT} = e^{-j\omega kT}\, e^{-j2\pi k} = e^{-j\omega kT} \qquad (D2)$$

Therefore $X(\omega, t_0)$ is periodic in $\omega$ with period $2\pi/T$.

Furthermore, if we let $\omega = 2\pi/T + \omega_1$,

$$e^{-j(2\pi/T + \omega_1)kT} = e^{j(-2\pi k - \omega_1 kT) + j4\pi k} = e^{-j(2\pi/T - \omega_1)kT} \qquad (D3)$$

Thus

$$X(2\pi/T + \omega_1, t_0) = X^*(2\pi/T - \omega_1) \qquad (D4)$$

and since the power spectrum is given by $X(\omega, t_0)X^*(\omega, t_0)$, we see that the power spectrum is symmetric about $\omega = 2\pi/T$.

D.2 Probability distribution of peak amplitude as a function of pulse rate

We now consider the output of a filter whose impulse response is N periods of a cosine wave,

$$h(t) = p_{T_o}(t-T_o)\cos\omega_o t \leftrightarrow \frac{\sin(\omega-\omega_o)T_o}{\omega-\omega_o} e^{-j(\omega-\omega_o)T_o} + \frac{\sin(\omega+\omega_o)}{\omega+\omega_o} e^{-j(\omega+\omega_o)T_o} = H(\omega) \tag{D5}$$

where

$$T_o = \frac{N\pi}{\omega_o} \quad \text{and} \quad p_{T_o}(t) = \begin{cases} 1, & -T_o \le t \le T_o \\ 0, & \text{otherwise} \end{cases}$$

For N large, $H(\omega)$ is a narrow band filter centered about $\omega_o$.

Let the input of the filter be the random polarity pulse train of Appendix A,

$$\underset{\sim}{x}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{a}_n \delta(t-nT) \tag{D6}$$

If $T = \pi/\omega_o$, i.e., the impulse spacing is one-half of the filter period, then the output is a cosine function whose amplitude each half period is determined by the sum of the amplitudes of the preceding N input impulses. Note that the zero crossings of the output are periodic. The peak amplitude can change by only +2, 0, or -2 from half-period to half-period, and it is binomially distributed. As N becomes large the distribution approaches the Gaussian distribution.

The above analysis is easily extended to the case of $T = k\pi/\omega_o$. If $N = nk$, the amplitude of the cosine wave changes by only +2, 0, or -2 every k half-periods. If k does not divide N,

then amplitude changes of ±1 occur because new impulse responses no longer begin when old ones terminate. This effect is negligible for large N.

If T is not a multiple of the half-period of the filter, the output is considerably more complicated since the impulse responses no longer add exactly in phase. The zero crossings are no longer exactly in phase. The output is the sum of N cosine waves whose amplitudes are ±1 and whose phases are $\omega_o nT$, for $0 \le n \le N-1$. If N is large and the phases are distributed around the unit circle (i.e., T is not close to a multiple of a half-period), we may regard the phases as being randomly distributed. The distribution of the peak amplitudes thus approaches the Rayleigh distribution. Note that the amplitude changes slowly, and thus the filter output closely approximates a cosine wave of slowly changing amplitude and phase.

D.3 Maximum output as a function of frequency

Consider a spectrum analyzer made up of a bank of filters of the type described in D.2. If we apply a random polarity pulse train to the input of one of these filters, the output energy is given by (see Appendices A, C)

$$\int_{-\infty}^{\infty} |H(\omega)|^2 d\omega = \int_{-\infty}^{\infty} h^2(t)dt = \int_0^{2\pi N/\omega_o} \cos^2 \omega_o t dt$$
$$= \frac{N\pi}{\omega_o} = T_o \qquad \text{(D7)}$$

In order that the outputs of filters of different $\omega_o$ be comparable, N must be varied to hold $T_o$ constant. Thus each filter of our spectrum analyzer has an impulse response of the same duration,

NT, and is affected by only the N preceding input impulses.

We now consider the maximum outputs of the different filters for an input pulse rate of 1/T. The time average power output will be the same for each filter as we have seen, but for particular input patterns the maximum or peak output differs as shown in figure D2 for N = 4. The contribution of each impulse response is represented by a phasor of unit amplitude. In figures D2a and D2b, the amplitude is 4 since the responses all add in phase. In D2c, $\omega_o$ is $\pi/2T$, and two of the responses are in quadrature resulting in a maximum output of $2\sqrt{2} = 2.828$. Figure D2d shows that the peak output at $\omega_o = \pi/4T$ is approximately 2.613.

The variability of the maximum can be changed by the use of pulse patterns as described in Appendix C. If we choose the patterns specified by

$$y = \begin{bmatrix} +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & +1 \\ -1 & -1 & -1 & +1 \\ -1 & -1 & +1 & -1 \end{bmatrix}$$

and let N = 8, the maximum outputs at various frequencies are shown in figure D3. The ratio of maximums in figure D2 reaches 1.414 for $\omega_o = \pi/T$ versus $\omega_o = \pi/2T$, and the greatest maximum occurs at $\omega_o = \pi/T$ (or $\omega_o = 2\pi/T$). For the patterns of figure D3, this ratio is reduced to 1.307, and the greatest maximum occurs at $\omega = \pi/4T$. Thus patterns can be used to shift short-term spectrum peaks and reduce their variations.

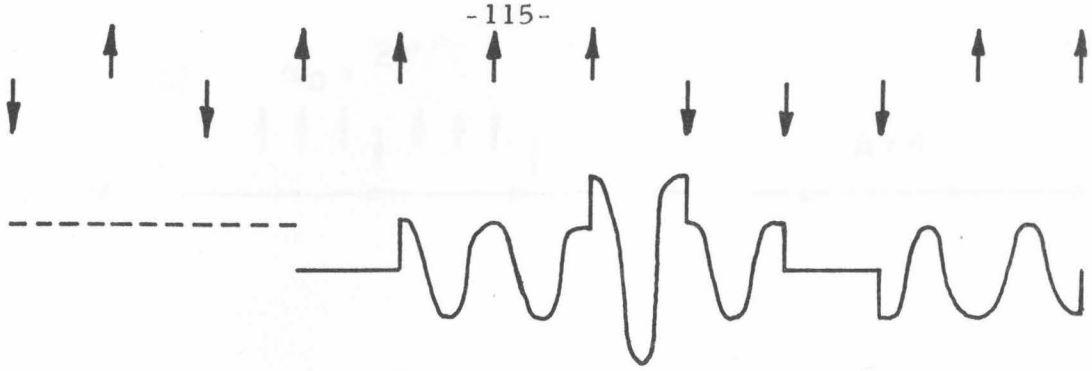


FIGURE D1 $N=2 \quad \omega_0 = \pi/T$

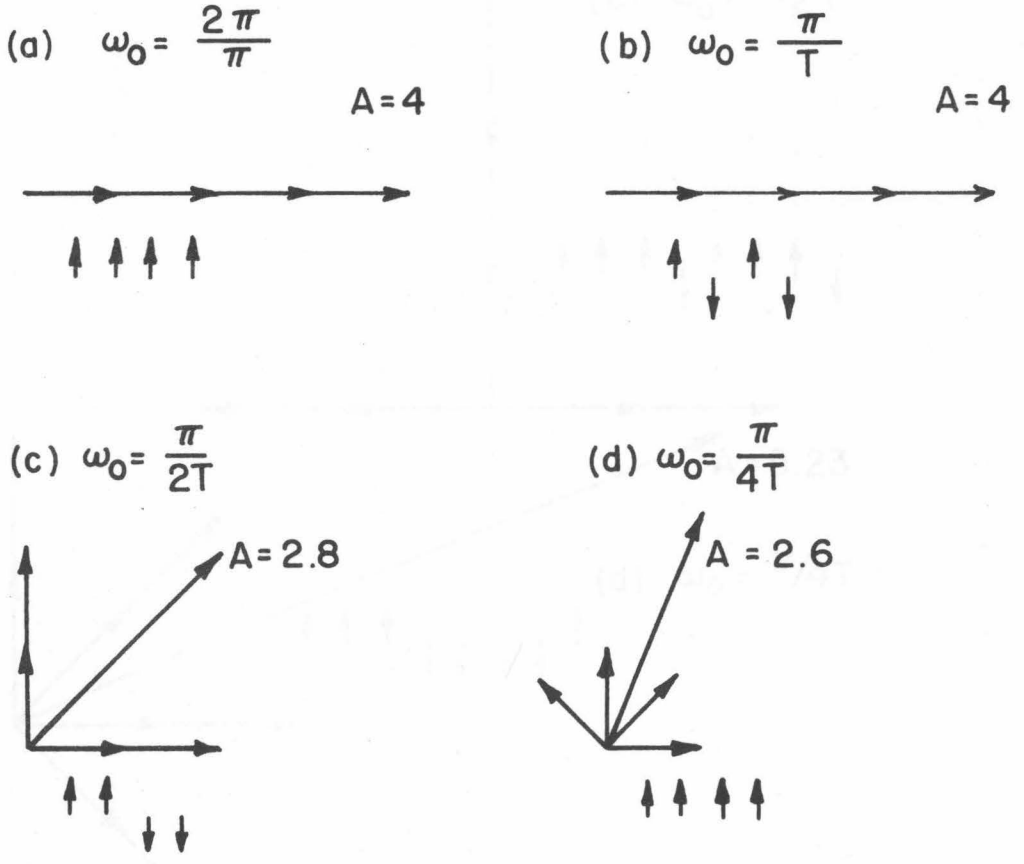



FIGURE D2

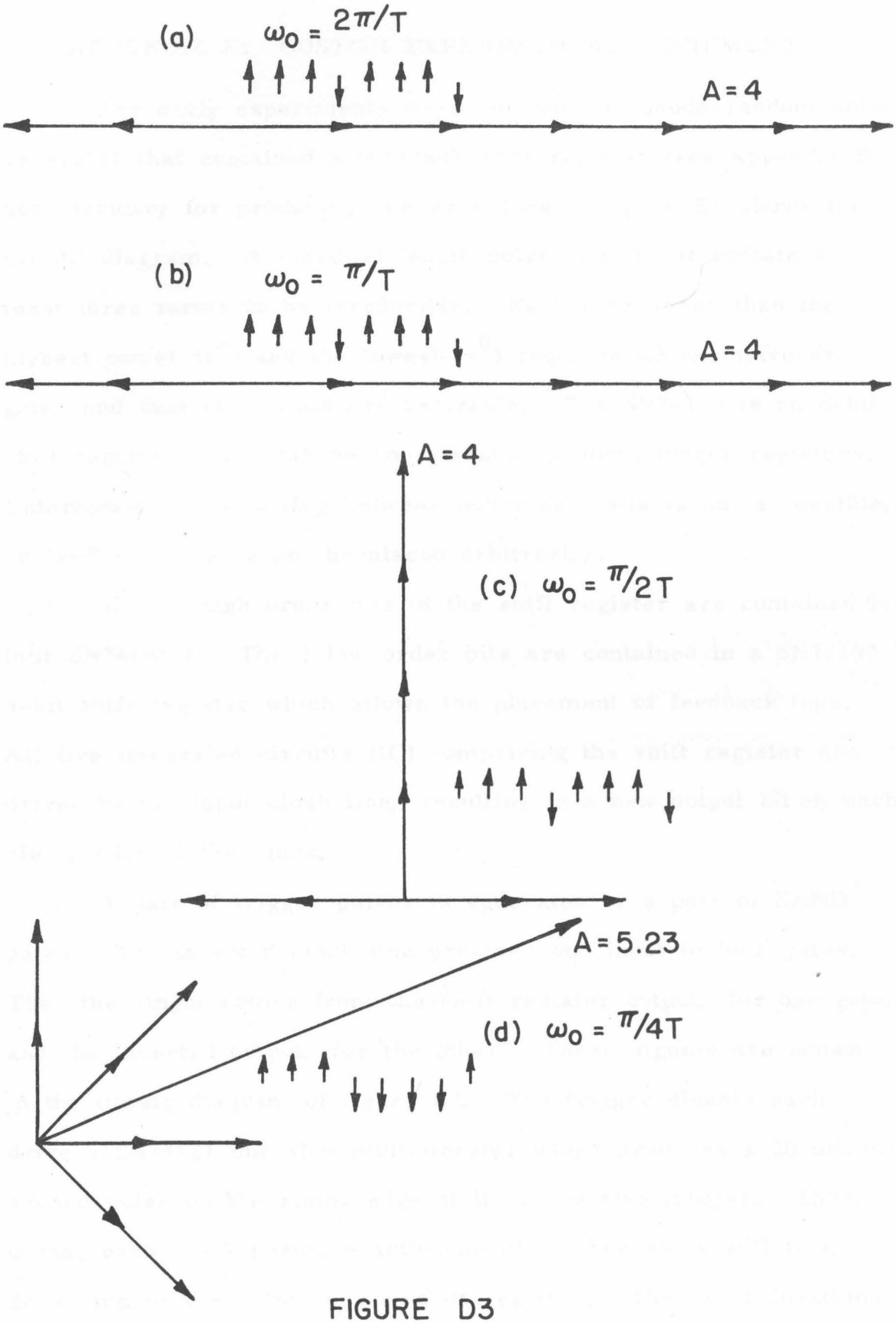

(a) $\omega_0 = 2\pi/T$
A = 4
(b) $\omega_0 = \pi/T$
A = 4
A = 4
(c) $\omega_0 = \pi/2T$
A = 5.23
(d) $\omega_0 = \pi/4T$

FIGURE D3

# APPENDIX E: CUSTOM EXPERIMENTAL EQUIPMENT

The early experiments were run with a pseudo-random noise generator that contained a feedback shift register (see Appendix B) and circuitry for producing bipolar pulses. Figure E1 shows the circuit diagram. A maximal length polynomial must contain at least three terms to be irreducible. Each term other than the highest power ($x^n$) and the lowest ($x^0$) requires an exclusive-or gate, and thus trinomials are desirable. The SN74165 is an 8-bit shift register which can be concatenated to form longer registers. Unfortunately, the wiring between individual cells is not accessible, so feedback taps cannot be placed arbitrarily.

The 32 high order bits of the shift register are contained in four SN74165's. The 3 low order bits are contained in a SN74194 4-bit shift register which allows the placement of feedback taps. All five integrated circuits (IC) comprising the shift register are driven by the input clock line, resulting in a new output bit on each rising edge of the clock.

A pair of trigger pulses is generated by a pair of NAND gates. The inverted clock line provides one input to both gates. The other input comes from the shift register output, for one gate, and the inverted output, for the other. These signals are shown in the timing diagram of figure E2. The trigger signals each drive a SN74121 one-shot multivibrator which produces a 20 microsecond pulse on the rising edge of its respective trigger. Thus, during each clock period exactly one of the one-shots will fire, depending on the output of the shift register. The exact durations

of the one-shots are controlled by trim-pots.

The outputs of the one-shots, transistor-transistor logic (TTL) signals which are either nominally 0 volts or 5 volts, are summed by a 741 operational amplifier. The top one-shot of figure E1, triggered by a "1" output of the shift register, is connected to the negative input of the summing circuit, thereby producing a negative going pulse out of the op amp. The other one-shot, triggered by a "0" output, produces a positive going pulse. The gain of the negative input is controlled by a resistance ratio and is fixed at 1. The positive input gain is controlled by a trim-pot and can be adjusted to match the negative input gain.

The circuit was adjusted by simultaneously observing the waveforms of the two one-shots on an oscilloscope and then adjusting the durations to be the same. By setting the time scale so that the pulses occupy the full screen, this adjustment can be made to approximately 2%. The amplitude balancing was done by viewing the op amp output on the oscilloscope with the height of each pulse occupying one-half of the vertical dimension of the screen, so as to present both positive and negative pulses. The amplitude adjustment could be made to approximately 5%. This balancing is somewhat complicated by the difference in pulse shape of the two pulses and slight ringing that occurs. We attempted to measure the DC output of the generator and adjust the amplitude so that it was zero. However, in any short term measure of the output, there is a varying DC component due to the random fluctuations of the number of positive versus negative pulses in the

time window of integration. It was observed that the waveform adjustment was more sensitive than the DC measurement.

The second set of experiments were run with pulse patterns produced by a Digital Equipment Corporation PDP 11/45 minicomputer. The fact that the computer and the sound booth were in different buildings complicated the task. Figure E3 is a block diagram of the computer interface; the detailed circuit diagram will not be presented. From the viewpoint of the sound booth site, the function of the computer was exactly the same as that of the feedback shift register described above. That is, a rising edge of the clock produces a new output bit. The shift register in the computer interface performed this function. The control logic counted the bits going out of the shift register and loaded a new set of bits from the buffer register when it was empty. After the buffer register was loaded into the shift register, a flag was set which could be interrogated by the computer under program control. The program would then load a new set of bits into the buffer register and reset the flag.

Figure E4 shows the circuit used at the experimental site (outside the sound booth). The clock signal is applied to ICs 2 and 5. IC 2 is a SN75109 line driver which transmits the clock signal to the computer over a twisted pair transmission line. IC 1 is a SN75107 line receiver which receives binary data from the computer over another twisted pair and produces a TTL output. A rising clock edge produces a new bit from the computer approximately 15 microseconds later. A falling clock edge triggers IC 5,

a SN74121 one-shot, which produces a 20 microsecond pulse. Since both positive and negative pulses are produced by the same one-shot, problems of adjusting two circuits to the same duration and then worrying about the differential effects of age and temperature are eliminated. Using both edges of the clock allows the output of IC 1 to settle before the one-shot is fired. A clock frequency of 20 kHz allows 25 microseconds between edges.

Transistors T1 and T2 act as constant current sources (about 10 milliamps). When the one-shot fires, one of the AND gates, IC 7 or IC 3, selected by the bit from the computer through IC 1, goes to logic 1. With the one-shot off, T3 and T4 are both on, shunting the current sources. T5 and T6 interface the TTL level signals to the current shunts (T4, T5) turning one of them off. Schottky diodes (1N5711) were placed in the circuit to prevent the current shunts from going into saturation, thus making the turn-off very rapid. When the shunt turns off, its current source flows through the 180 ohm resistor, producing a voltage pulse of appropriate polarity. The 1N914 diodes in series with the current sources isolate the unused source.

After the experiments with the pulse patterns were completed, it was necessary to provide a feedback shift register to drive the pulse generator described above. A switch was installed in the pulse generator so that the bit stream which had been coming from the twisted pair line receiver could be supplied by an external TTL signal.

The feedback shift register was designed to be a generally useful instrument and thus contained some features beyond those called for in the experiment. Figure E5 shows the circuit diagram. ICs 9 through 12 are SN74174 hex D type flip-flops which form the register. A 24 stage shift register is formed by external wiring. ICs 2 through 8 contain a total of 24 exclusive-or gates (i.e., mod-2 adders), each one providing the input to each D flip-flop. One adder input was connected to the output of the preceding flip-flop. The first stage is driven by an adder simply because there was one left over. Since there is no preceding stage, its input was connected to a switch which normally provides a logic 0. Throwing the switch allows 1's to enter the register, a useful feature for getting out of the all-zero state. The other adder input is connected to a switch which provides either logic 0, making the adder transparent to the signal flow, or the logic value of the feedback line. The feedback line is obtained through IC 22, an open collector high current driver which is driven by the last stage of the register.

This connection of adders and flip-flops forms the feedback shift register described in Appendix B. Twenty-four slide switches allow one to use any polynomial of length 24 or less. ICs 17 through 20 synthesize a 24-input AND gate. Twenty-four additional slide switches allow each gate input to be switched between the output of the corresponding register stage or its complement, provided by the inverters in ICs 13 through 16. The output of this gate provides a synchronization pulse when the register is in

the state specified by the switches.

The feedback shift register box also contains a positive/negative pulse generator circuit using two one-shots and a 741 operational amplifier. This circuit was built in before the pulse generator described above was designed. It was not used in any of the formal experiments.

Figure E6 shows the circuit diagram of the high-pass filter used in the later experiments. A four-pole Chebyshev design was chosen to provide the sharpest cutoff in the frequency response. Two Burr-Brown UAF/31 universal filter modules were used. A five-pole, four position rotary switch connected different resistors in the feedback path of the circuit to change the cutoff frequency of the filter. Figure E7 shows the characteristics of the filter as a function of frequency for the four switch positions.

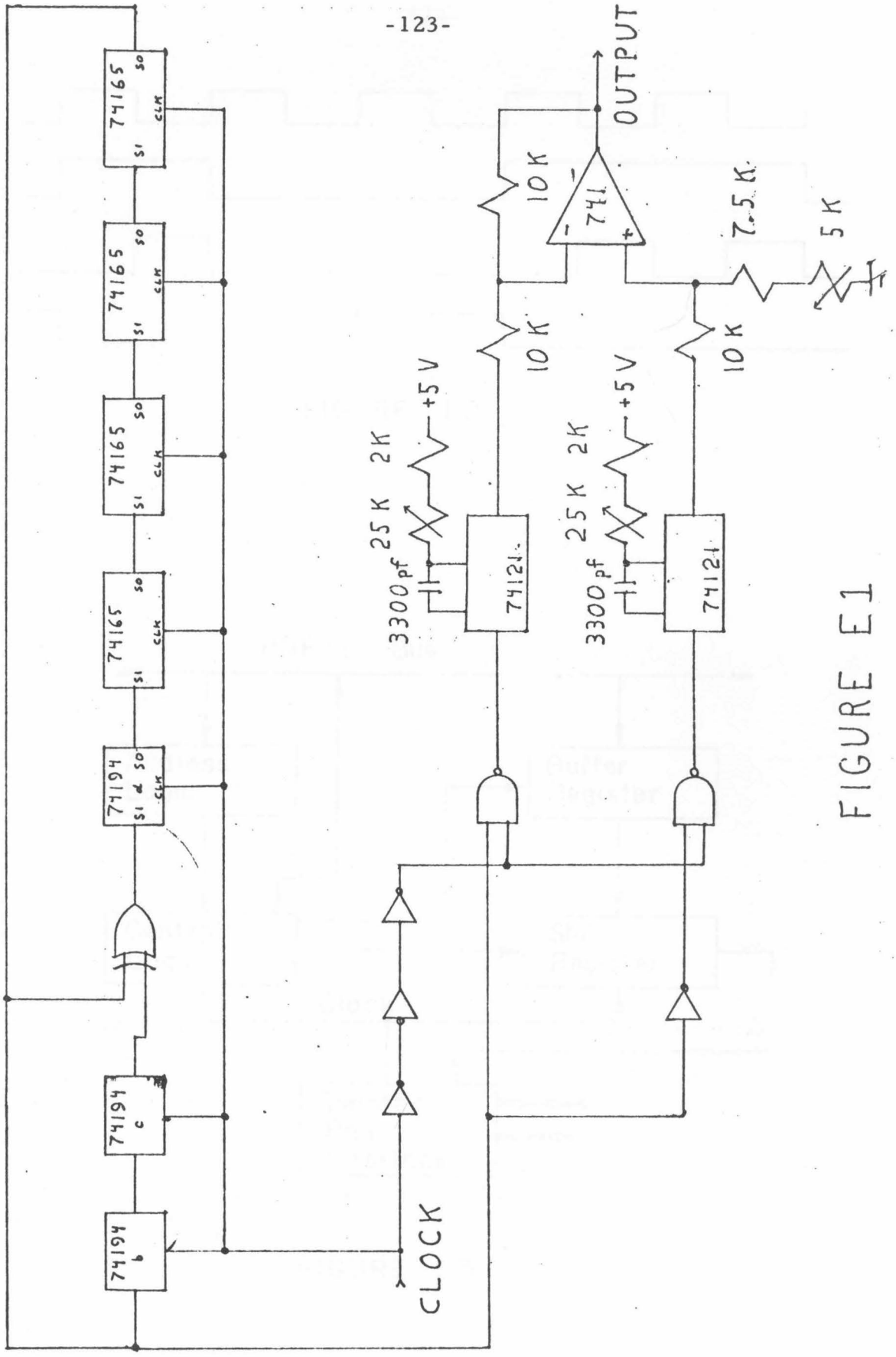

74165
74165
74165
74165
74194
74194
74194
CLOCK
74121
74121
3300pf
25K
2K
+5V
3300pf
25K
2K
+5V
10K
10K
10K
741
OUTPUT
7.5K
5K

FIGURE E1

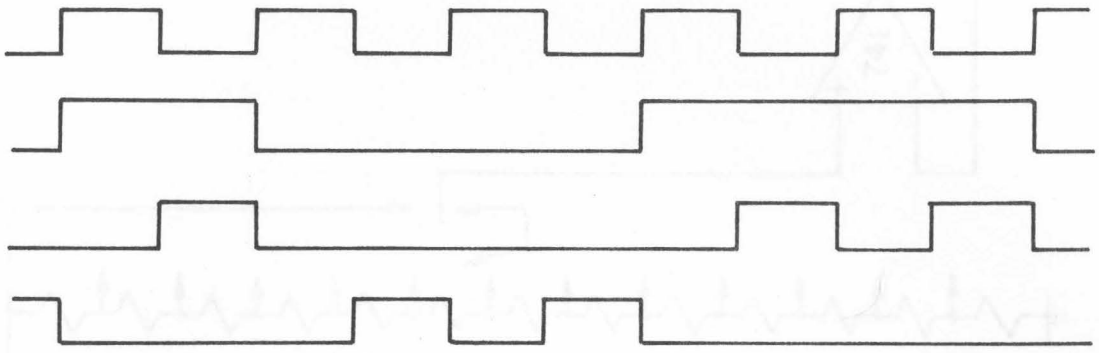


FIGURE E2

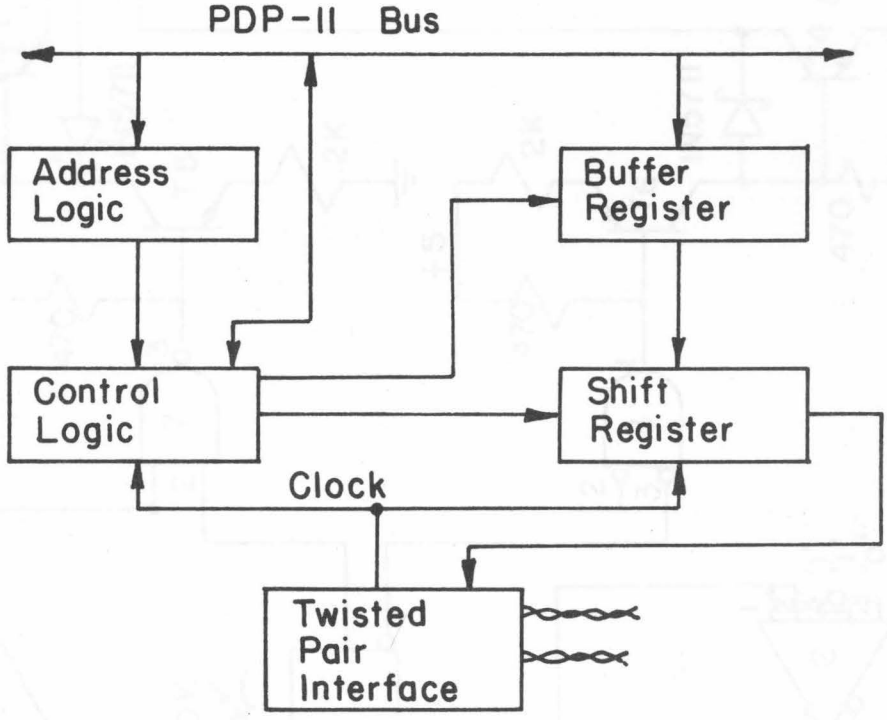



FIGURE E3

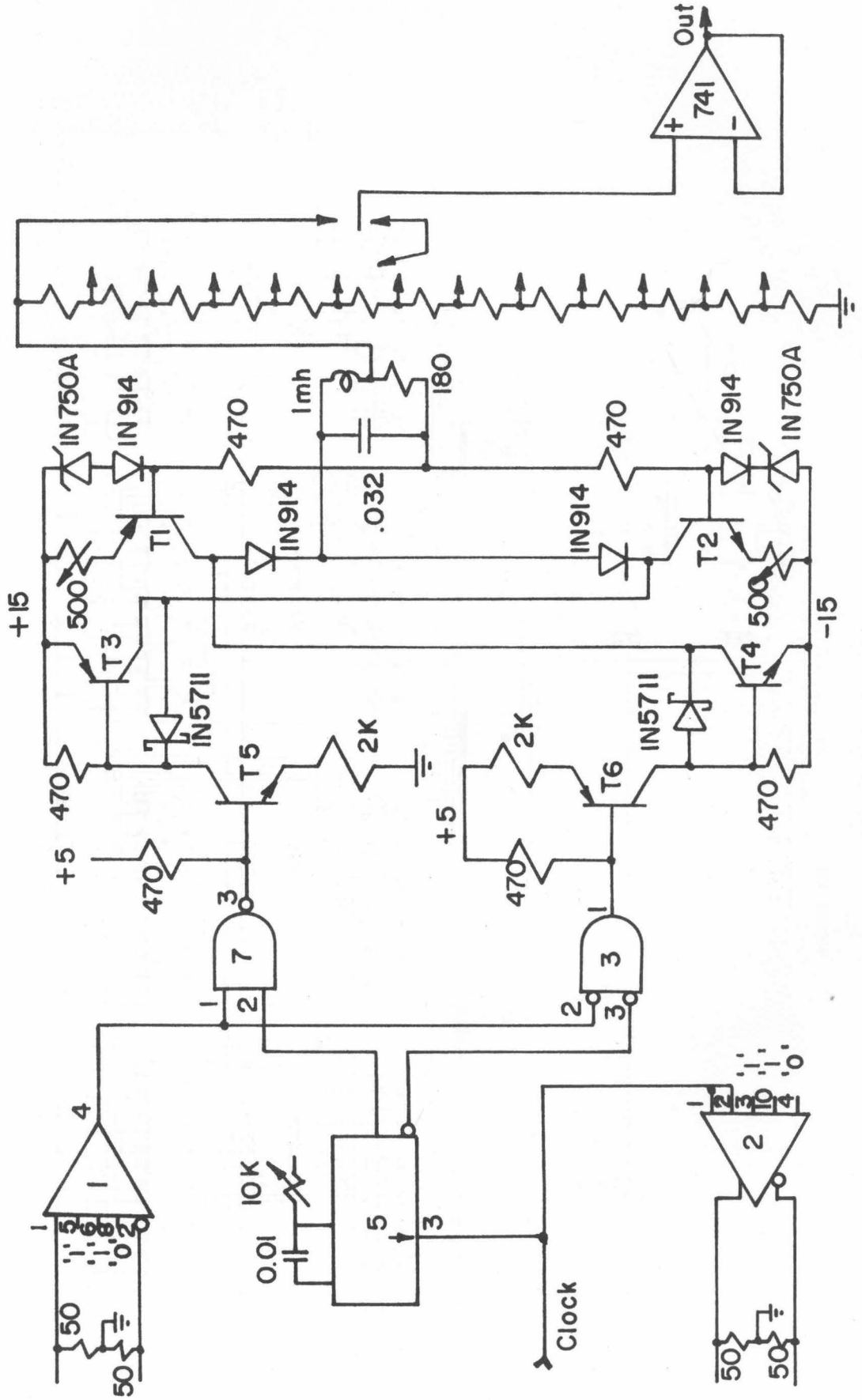

Out
741
IN750A
IN914
470
1mh
180
.032
IN914
470
IN914
IN750A
+15
500
T1
T2
500
-15
T3
T4
IN5711
IN5711
2K
2K
T5
T6
470
470
+5
+5
470
470
7
3
10K
0.01
5
3
Clock
4
1
2
50
50
50
50

FIGURE E4

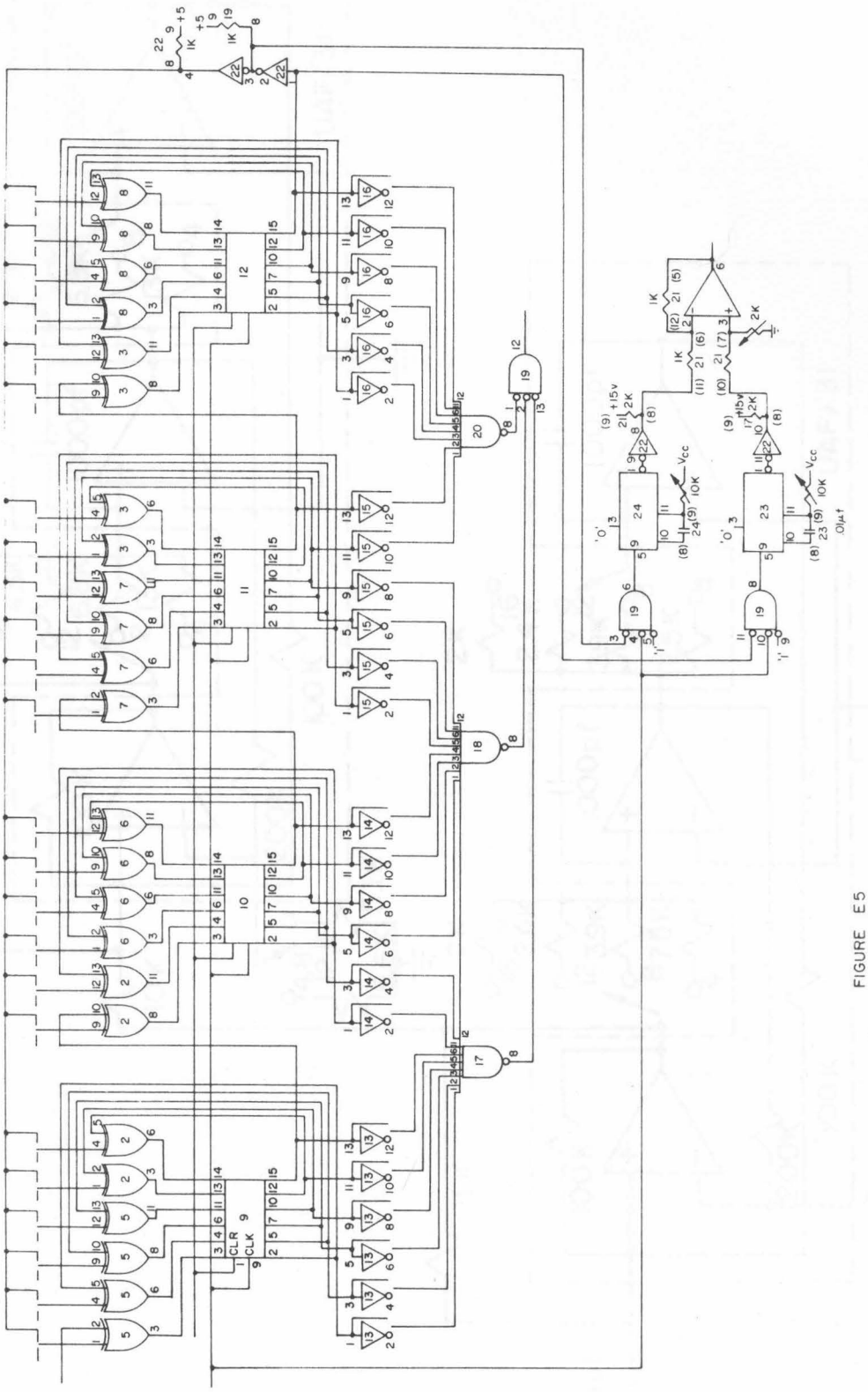



FIGURE E5

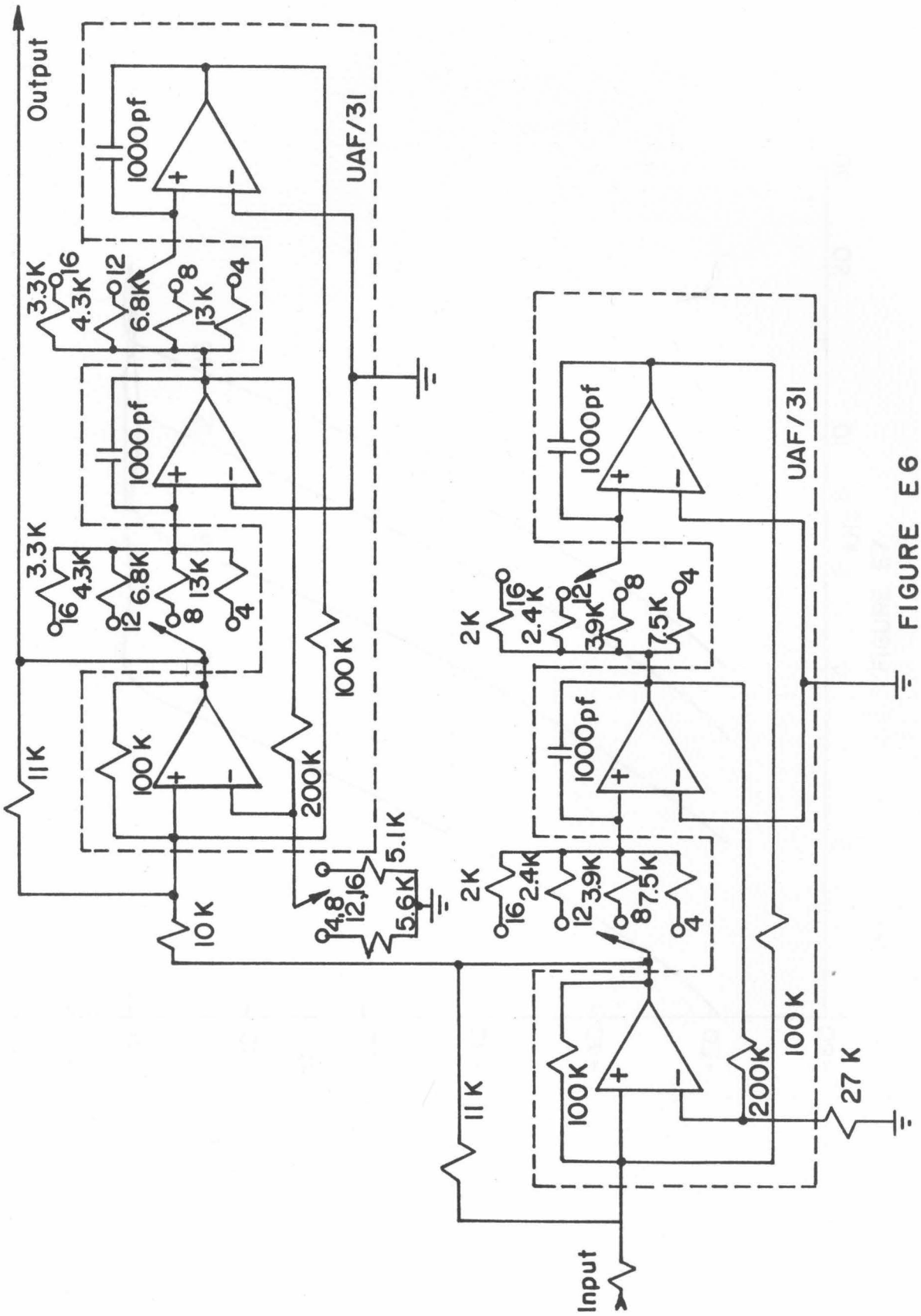

Output
UAF/31
1000pf
3.3K
4.3K
6.8K
13K
100K
200K
11K
10K
5.1K
5.6K
2K
2.4K
3.9K
7.5K
27K
Input

FIGURE E6

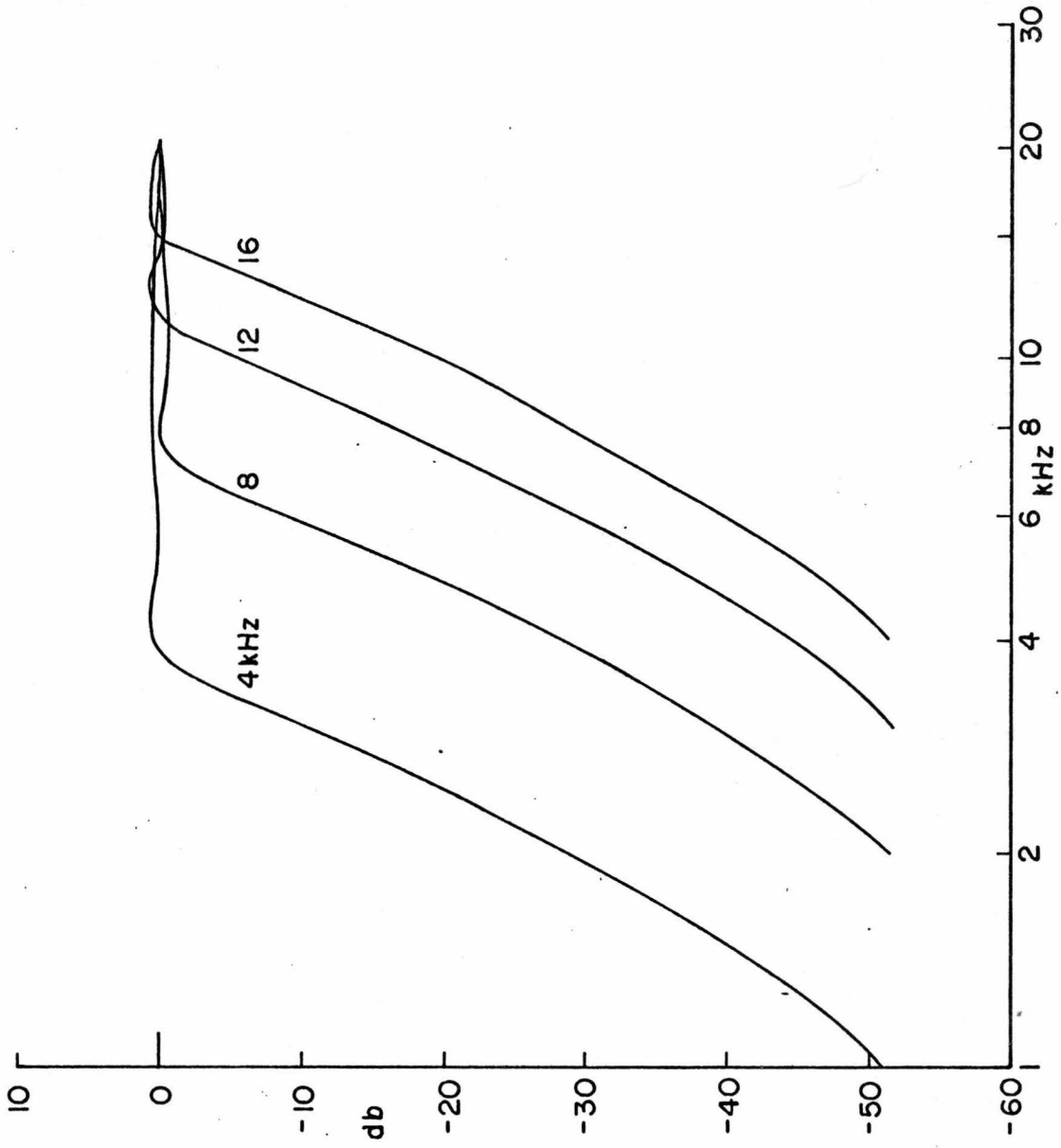



FIGURE E7

## REFERENCES

1. Flanagan, James L. 1962 Models for Approximating Basilar Membrane Displacement - Part II. Effects of Middle-Ear Transmission and Some Relations between Subjective and Physiological Behavior. Bell System Technical Journal, 41, no. 3, p. 959.

2. Møller, Aage R. 1972 "The Middle Ear", Chapter 4, Foundations of Modern Auditory Theory, Vol. II, Jerry V. Tobias, Ed., Academic Press.

3. Helmholtz, Hermann 1863 On the Sensations of Tone, translated by Alexander J. Ellis, 1885, republished by Dover Publications, 1954.

4. Von Bekesy, Georg 1960 Experiments in Hearing, McGraw-Hill.

5. Teas, Donald C. 1970 "Cochlear Processes", Chapter 7, Foundations of Modern Auditory Theory, Vol. I, Jerry V. Tobias, ed., Academic Press.

6. Geisler, C. Daniel, William S. Rhode and Duncan T. Kennedy 1974 Responses to Tonal Stimuli of Single Auditory Nerve Fibers and Their Relationship to Basilar Membrane Motion in the Squirrel Monkey. J. Neurophysiol., 37, p. 1156-1172.

7. König, E. 1957 Effect of time on pitch discrimination thresholds under several psychophysical procedures; comparison with intensity discrimination thresholds. J. Acous. Soc. of America, 29, no. 5, p. 606-612.

REFERENCES (CONT'D)

8. Schouten, J. F. 1940 The Residue, a New Component in Subjective Sound Analysis. Proc. Koninkl. Ned. Akad. Wetenschap. 43, p. 356-365.

9. Thurlow, W. R. and A. M. Small, Jr. 1955 Pitch Perception for Certain Periodic Stimuli. J. Acous. Soc. of America, 27, no. 1, p. 132-137.

10. Schouten, J. F. 1938 The Perception of Subjective Tones. Proc. Koninkl. Ned. Akad. Wetenschap. 41, p. 1086-1093.

11. Moe, C. R. 1942 An Experimental Study of Subjective Tones Produced within the Human Ear. J. Acous. Soc. of America, 14 p. 159-166.

12. Egan, J. P. and R. G. Klumpp 1951 The error due to masking in the measurement of aural harmonics by the method of best beats. J. Acous. Soc. of America, 23, p. 275-286.

13. Small, A. M., Jr. and R. A. Campbell 1961 Masking of pulsed tones by bands of noise. J. Acous. Soc. of America, 33, p. 1570-1576.

14. Mathes, R. C. and R. L. Miller 1947 Phase effects in monaural perception. J. Acous. Soc. of America, 19, p. 780-797.

15. Licklider, J. C. R. 1955 Influence of phase coherence upon the pitch of complex, periodic sounds. J. Acous. Soc. of America, 27, p. 996(A).

16. Schouten, J. F., R. J. Ritsma and B. Lopes Cardozo 1962 Pitch of the Residue. J. Acous. Soc. of America, 34, no. 8, part 2, p. 1418-1424.

REFERENCES (CONT'D)

17. Rose, Jerzy E., John F. Brugge, David J. Anderson and Joseph E. Hind 1967 Phase-locked Response to Low-Frequency Tones in Single Auditory Nerve Fibers of the Squirrel Monkey. J. Neurophysiol. 30, p. 769-793.

18. Anderson, David J., Jerzy E. Rose, Joseph E. Hind and John F. Brugge 1971 Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects. J. Acous. Soc. of America, 49, p. 1131.

19. Flanagan, James L. and Newman Guttman 1960 On the Pitch of Periodic Pulses. J. Acous. Soc. of America, 32, no. 10, p. 1308-1319.

20. Flanagan, James L. and Newman Guttman 1960 Pitch of Periodic Pulses without Fundamental Component. J. Acous. Soc. of America, 32, no. 10, p. 1319-1328.

21. Guttman, Newman and J. L. Flanagan 1964 Pitch of High-Pass Filtered Pulse Trains. J. Acous. Soc. of America, 36, no. 4, p. 757-765.

22. Rosenberg, Aaron E. 1965 Effect of Masking on the Pitch of Periodic Pulses. J. Acous. Soc. of America, 38, no. 5, p. 747-758.

23. Small, Arnold M., Jr. and Max E. McClellan 1963 Pitch Associated with Time Delay between Two Pulse Trains. J. Acous. Soc. of America, 35, no. 8, p. 1246-1255.

REFERENCES (CONT'D)

24. Harris, Gerard G. 1963 Periodicity Perception by Using Gated Noise. J. Acous. Soc. of America, 35, no. 8, p. 1229-1233.

25. Mills, A. W. 1958 On the Minimum Audible Angle. J. Acous. Soc. of America, 30, no. 4, p. 237-246.

26. Jeffress, Lloyd A. 1972 "Binaural Signal Detection: Vector Theory", Chapter 9, Foundations of Modern Auditory Theory, Vol. II Jerry V. Tobias, ed., Academic Press.

27. Scharf, Bertram 1970 "Critical Bands", Chapter 5, Foundations of Modern Auditory Theory, Vol. I, Jerry V. Tobias, ed., Academic Press.

28. Ritsma, R. J. 1962 Existence region of the tonal residue. I. J. Acous. Soc. of America, 34, p. 1224-1229.

29. Ritsma, R. J. 1963 Existence region of the tonal residue. II. J. Acous. Soc. of America, 35, p. 1241-1245.

30. Pollack, Irwin 1968 Detection and Relative Discrimination of Auditory 'Jitter'. J. Acous. Soc. of America, 43, no. 2, p. 308-315.

31. Peterson, W. W. 1961 Error-Correcting Codes. MIT Press, Cambridge, Mass.

32. Golomb, Solomon W. 1967 Shift Register Sequences. Holden-Day, Inc.