(2)

AD A122390

DTIC FILE COPY

LIMITATIONS ON TOA ACCURACY

T. Finley Burke

September 1979

DTIC
ELECTE
DEC 14 1982
S D
E

P-6386

LIMITATIONS ON TOA ACCURACY*

T. Finley Burke
The Rand Corporation
Santa Monica, California 90406

ABSTRACT

A directional antenna is a filter whose response varies with direction. A modulated signal transmitted from such an antenna produces different far-field waveforms in all directions (except for possible symmetry). The response of a receiver to the transmission will vary with position in the antenna pattern of the emitter. This effect is ordinarily negligible, but it could limit the ultimate performance of any system that accepts off-axis signals and relies upon the details of waveform. In a leading-edge TOA system used to determine the location of a pulsed emitter, the several receivers necessarily lie at different angular positions from that emitter. This analysis indicates that in a simple idealized TOA system that would otherwise be entirely free from error, the waveform effect arising from a directional emitter can lead to errors of several dozen meters in the computed location.

INTRODUCTION

A directional antenna is a frequency-selective filter whose response varies with angle (the directivity pattern varies with frequency). If a modulated signal is transmitted, the various spectral components experience different phase and amplitude changes at any one angle, and each spectral component experiences different shifts at different angles. Consequently, the far-field waveform is different at every angle (except for possible symmetry).

Most analyses of system behavior assume that the far-field waveforms differ only by an amplitude coefficient. Although this approximation is ordinarily acceptable, it overlooks effects that might be significant in systems that accept signals over a range of angles and that rely on waveform details. This paper indicates that the effects that arise are associated with each pattern null of the directional antenna and are found even at the edge of the main beam. Thus, this minor effect could even limit the ultimate performance of systems that do not employ wide-angle signals.

TOA SYSTEM

In a so-called leading-edge TOA (time of arrival) system, several receivers at diverse locations record the time at which the leading edge of a pulse from an emitter arrives. If the locations of the receivers are known, the differences among the arrival times can be used to calculate the location of the emitter. To achieve useful accuracy, the receiver array must subtend a substantial angle at the emitter; this need can be understood to resemble the need for a long baseline in an optical rangefinder. Consequently, such a system necessarily relies on signals received at appreciably different angles from the axis of the emitter. Moreover, it is necessary to make highly accurate measurements on a strongly modulated signal--the leading edge of a pulse. For these reasons, this kind of system could be particularly vulnerable to the waveform effect described above.

If the TOA receivers are identical, and if the emitter waveform could be the same at all angles, the exact response of the receiver would not be especially critical. The outputs from two receivers, suitably adjusted in amplitude, could, in principle, be made exactly coincident by a suitable time displacement. That displacement would then be the difference in arrival times. When the receivers receive different input waveforms, their outputs necessarily differ, and no combination of amplitude adjustment and time displacement will make them coincide. Thus the difference in arrival times is no longer defined uniquely, and it is necessary to adopt some arbitrary definition of the event that will constitute "arrival." The numerical value of the difference in arrival times will depend upon the definition chosen. Various possible definitions will lead to different solutions for the emitter location, and this arbitrariness indicates that some system error will occur even in an idealized system that might be expected to be perfect.

To estimate the magnitude of system error that might be expected, it is necessary to analyze in detail how the transient response of a single receiver varies with position in the antenna pattern of the emitter. Once this is done, it is possible to examine the system performance of a rudimentary TOA system consisting of three identical receivers. Analytical difficulties dictate that a simplified and idealized receiver design be considered, that we assume the system to be free of noise and interfering signals, free of multipath and propagation anomalies, and that time can be measured perfectly. It is intended that the system would be free of error were it not for the effect of the emitter antenna on waveform.

EMITTER

Our inability to solve the wave equation subject to the boundary conditions of a directional antenna necessitates the adoption of the customary model wherein the antenna is regarded as an aperture in an infinite screen. Although this model works well enough for ordinary purposes, it is well known to fail at large off-axis angles; use of this model is undoubtedly the most serious defect in this analysis. It is convenient, and less harmful, also to adopt the Kirchoff approximate boundary conditions. The emitter will be taken to be a uniformly illuminated line or rectangle of length L, radiating simple long pulses at angular carrier frequency ω_0. The directivity pattern at the carrier frequency is taken to be sinZ/Z where

$$Z = \frac{\omega_0 L}{2c} \sin\theta$$

and θ = angle from the emitter axis.

AXIAL WAVEFORM

The analysis starts with the adoption of a suitable pulse waveform that is found on axis in the far field. The aperture model of the antenna, together with the approximate boundary conditions, allows us to suppose that this is also the illumination incident upon the aperture. (The far-field axial waveform could probably not be found in the near field of a real antenna nor in the feed structure. This assumption is merely a stratagem whereby we seek to estimate the off-axis radiation without really solving the boundary value problem.)

It will be supposed that the pulses are sufficiently long that we can ignore events associated with the end of the pulse; we seek a satisfactory description of the front end. It would be convenient to assume the axial waveform to be

$$F(t) = \left[1 - \exp(-K\omega_0 t)\right] \sin(\omega_0 t + \psi);\ t \geq 0$$

That is, an amplitude envelope multiplying a fixed carrier.

Carrier phase ψ, defined with respect to the pulse gate and regarded as constant during a pulse, should be arbitrary and possibly variable from pulse to pulse. However, the average value of the integral over time of F(t) is not zero unless $\tan\psi = K$. Moreover, this average accumulates rapidly during the early portion of the pulse. A non-zero average would mean that the emitter delivers a net average electric charge to the far field, and that is impossible. Thus, this convenient envelope form, with arbitrary phase, is physically impossible. For most purposes, this shortcoming is harmless, but in this case the non-zero average would lead to serious difficulty in treating the heterodyne stage of the receiver. To avoid that, it is necessary to adopt a waveform whose average is zero.

If a step-modulated carrier

$$u(0)\frac{\sqrt{1 + 3\Delta_0^2}}{2\Delta_0} \sin(\omega_0 t + \psi)$$

is passed through a single-stage bandpass filter whose Q is Q_0, and if the filter is tuned to ω_0/Δ_0, where

$$\Delta_0 = \sqrt{1 - (1/2Q_0)^2}$$

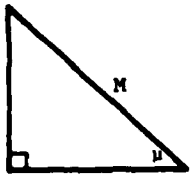
then the output is

$$F_0(t) = \sin(\omega_0 t + \psi + \phi)$$

$$- \frac{M}{\Delta_0} \exp\left[-\omega_0 t/2\Delta_0 Q_0\right] \sin(\omega_0 t + \psi + \mu - \phi)$$

where $\tan\phi = 1/4\Delta_0 Q_0$

and



The tuned frequency of the filter is offset slightly from ω_0 to avoid different frequencies in the two sinusoids. That mistuning introduces a small insertion loss at ω_0, and the amplitude of the stepped carrier is adjusted to normalize the output from the filter.

This waveform, $F_0(t)$, is taken to be the far-field waveform on the emitter axis. As in F(t), the expression contains a steady-state term and an exponentially decaying transient, both at frequency ω_0. However, $F_0(t)$ is phase-modulated and does not have evenly spaced zero crossings until the transient dies out. Moreover, the amplitude of the transient term in F_0 depends upon ψ. The local phase adjustments avoid the accumulation of a long-term average, and F_0 does not deliver a net charge. F_0 cannot be represented as an amplitude envelope multiplying a fixed carrier. For that reason this form necessitates that the transient analysis of the TOA receiver be carried out exactly.

In all numerical examples below the value of Q_0 is chosen so

$$2\Delta_0 Q_0 = 6\pi$$

$$Q_0 = 9.438$$

This attributes to the emitter a fairly typical half-power bandwidth of 10.6 percent. It will be apparent below that results of the analysis would be altered only in minor detail by other similar choices of Q_0; the pulse rise time has very little

influence on the receiver response if it is short compared to the rise time of the receiver, as it usually will be if the emitter is a radar.

OFF-AXIS WAVEFORMS

When viewing the emitter aperture off axis, one end is closer to the field point than is the center of the aperture. If the signal arriving from the center commences at t = 0, then the signal starts to arrive from the near end when

$$t = -\frac{L\sin\theta}{2c} = -\frac{Z}{\omega_0}$$

During the time interval $-Z/\omega_0 \le t \le Z/\omega_0$, the portion of the aperture that contributes to the arriving signal increases. The incremental contributions from different portions of the aperture experience different transit time delays, and the resultant waveform must be calculated by what amounts to a convolution with the impulse response of the aperture.

The signal arriving during the early time interval is given by

$$F_1(t) = \frac{1}{2Z}\int_0^{\omega_0 t+Z} F_0(z)\,dz \quad ; \quad -Z \le \omega_0 t \le Z$$

whereas in the subsequent interval, when the entire aperture contributes, the signal is

$$F_2(t) = \frac{1}{2Z}\int_{\omega_0 t-Z}^{\omega_0 t+Z} F_0(z)\,dz \quad ; \quad \omega_0 t \ge Z$$

Here the variable z has been substituted for the retarded time function

$$z = \omega_0\left(t + \frac{L\sin\theta}{c}\right)$$

The representation of the off-axis waveform as two successive segments is mathematically convenient, but the two segments are smoothly joined at $t = Z/\omega_0$ (at that time, $F_1 = F_2$ and $\dot{F}_1 = \dot{F}_2$) and they comprise a single waveform. The waveform as a whole will be designated $F_{1,2}$; that dual subscription will be used from here on to represent two smoothly joined successive segments.

A different aperture shape or non-uniform illumination would introduce into the integrands above an additional factor that is a function of z. That would make the integration more difficult, and would yield different waveforms, but two segments would still occur, and the duration of F_1 would remain $2Z/\omega_0$. For the form of F_0 adopted here,

$$F_1 = \frac{1}{2Z}\left\{-\cos(\omega_0 t + \psi + Z + \phi) + Me^{-\frac{\omega_0 t+Z}{2\Delta_0 Q_0}}\cos(\omega_0 t + \psi + Z + \mu - \phi)\right\}$$

$$F_2 = \frac{\sin Z}{Z}\sin(\omega_0 t + \psi + \phi) + \frac{M}{2Z}\left\{e^{-\frac{\omega_0 t+Z}{2\Delta_0 Q_0}}\cos(\omega_0 t + \psi + Z + \mu - \phi) - e^{-\frac{\omega_0 t-Z}{2\Delta_0 Q_0}}\cos(\omega_0 t + \psi - Z + \mu - \phi)\right\}$$

Two important points arise: (1) F_2 contains terms that decay exponentially; thus it should not be thought that F_1 is the transient and F_2 the steady state. (2) Only the steady-state term in F_2 carries the coefficient $\sin Z/Z$. Thus, it is only this steady-state term that exhibits the directivity pattern of the aperture. None of the other terms go to zero when $\sin Z = 0$; there is a signal present in all the nulls of the $\sin Z/Z$ pattern. Moreover, the terms that comprise that signal are present at all Z, not merely in the nulls.

THE "EXTRA" SIGNAL

It was noted above that most discussions of the far field regard the off-axis waveform as $F_0(t)$ multiplied by an amplitude coefficient. Thus, it is not inappropriate to call

$$\frac{\sin Z}{Z}F_0(t) \quad ; \quad t \ge 0$$

the waveform that is ordinarily "expected." It is convenient here to call

$$F_{1,2} - \frac{\sin Z}{Z}F_0(t)$$

the "extra" signal that shows up because of the antenna effect. The two terms that comprise the "expected" signal do not occur in $F_{1,2}$; the expression for the "extra" signal is complicated and will not be displayed here because it is not employed in that form in the subsequent analysis. However, the view that $F_{1,2}$ can be regarded as the sum of these two signals is useful in interpreting the results of the analysis.

Figure 1 shows $F_{1,2}$ on the source axis, in the first pattern null, near the top of the first lobe, and in the second null. Figure 2 shows $F_{1,2}$ in the 12th null, the 12th lobe, the 50th null, and the 50th lobe ($Z/\pi = 50$ occurs about 90° from the axis of a source whose beamwidth is one degree). In all pattern nulls, the "expected" signal is absent and the signal is the "extra" signal alone. At other angular locations both are present, and it is difficult to visualize the appearance of the "extra" signal.

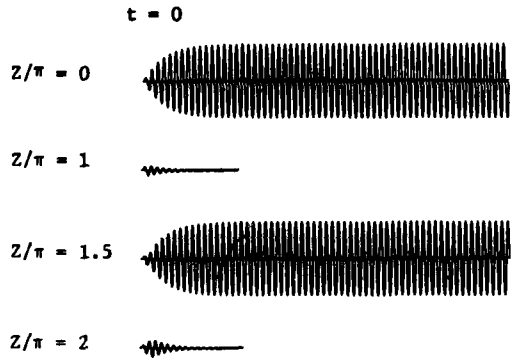


Figure 1

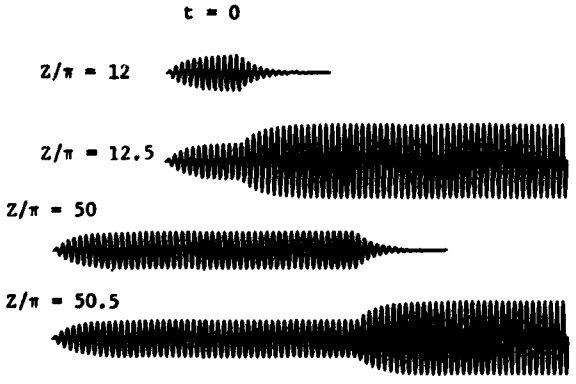


Figure 2

Figure 3 shows the "extra" signal in two nulls and two lobes at large off-axis angles.

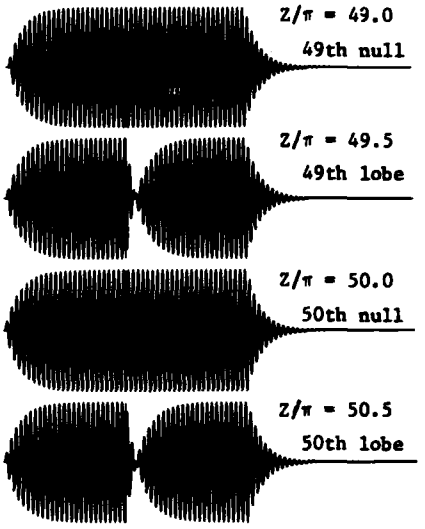


Figure 3

Figure 4 shows the "extra" signal at intermediate angles. (In these and all subsequent figures, the waveforms are shown for $\psi = 0$. Other choices of ψ lead to minor changes that can scarcely be observed by eye.)

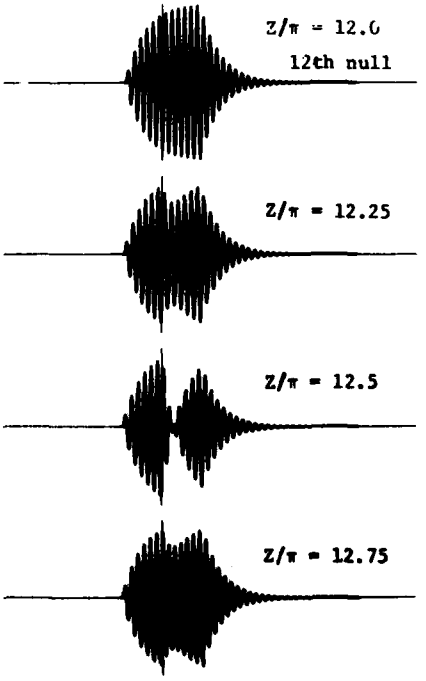


Figure 4

In all cases, the duration of the "extra" signal is somewhat longer than $2Z/\omega_0$ because the transient terms of F_2 endure after $t = Z/\omega_0$. Nevertheless, the duration of the extra signal is brief even at large Z. In particular, the duration is likely to be small compared to the rise time of the receiver. It should be noted that, at any given value of Z, the amplitude of the "extra" signal is not generally small compared to the amplitude of the "expected" signal. Thus, the "extra" signal usually delivers much less energy to the receiver than does the "expected" signal because of the difference in their duration, not their amplitude.

TRANSIENT RESPONSE OF THE RECEIVER

To minimize the analytic burden, the design of the receiver is simplified and idealized. Thus, it is supposed that the receiver employs a wideband non-directional antenna so the waveform delivered to the RF amplifier is exactly the same as the free-space waveform (effects of the receiver itself on the radiation field are ignored). A directional receiving antenna would add further effects counterpart to those that lead to $F_{1,2}$, if the receiver accepted signals off its own axis. Such a situation would increase the analytic task without adding anything new and is not worthwhile. The second antenna would not, however, offset the effects of the first; rather, the effects would be compounded.

The receiver will be taken to contain only an RF front end, a heterodyne detector, an IF strip whose angular center frequency is ω_1, and an envelope detector. The arrival time of the incoming signal is measured at the output of the envelope detector by a circuit that implements the selected definition of "arrival." Although rudimentary in all aspects, this design contains the major elements of a heterodyne receiver and it suffices to illustrate the phenomena that occur.

It is not uncommon, in analyzing a heterodyne receiver, to ignore the selectivity of the RF front end because most of the selectivity is in the IF strip. We commence with that approximation, but this choice is reconsidered below. At this point it is assumed that signal $F_{1,2}$ (or F_0 on axis) is delivered to the heterodyne detector.

The local oscillator signal is taken to be

$$\cos\left[(\omega_0 + \omega_1)t + \sigma\right]$$

at an amplitude much higher than that of $F_{1,2}$. Phase σ should be regarded as random from one receiver to another on any one pulse and random from pulse to pulse in any one receiver. (A change of one half of an RF period in the transit time from the emitter to a receiver would shift σ over the entire significant range.) If σ is regarded as constant over the pulse duration, then this represents the receiver to be perfectly tuned to the source carrier. Slight mistuning could be described by σ changing during the pulse; that would lead to more complicated effects than are discussed here.

It is impractical to undertake an exact analysis of the nonlinear heterodyne detector. Instead, we adopt the conventional approximation that the detector is a square-law device. The customary arguments are invoked to justify neglect of the squares of $F_{1,2}$ and the local oscillator signal, and the output from the detector is taken to be

$$F_{3,4} = 2F_{1,2}\cos\left[(\omega_0 + \omega_1)t + \sigma\right]$$

This output, represented as two successive segments, is the input to the IF strip.

At this point, $F_{3,4}$ can be decomposed into terms at the sum frequency, $2\omega_0 + \omega_1$, and the difference frequency, ω_1. It is customary to discard the sum frequency terms because the steady-state response of the IF strip is down many dB at that frequency. The argument is adequate under most circumstances, especially if the signal is not strongly modulated. However, the transient response of the IF strip is not down nearly so many dB, and neglect of the sum terms would inject a troublesome formal difficulty here. Certainly, F_1 equals zero when $t = -Z/\omega_0$, so F_3 must also equal zero at that time. However, if the sum frequency terms are discarded, it will be seen that F_3 is not zero then; instead, F_3 undergoes a discontinuous jump. Such a step, introduced as an analytic artifact, would cast doubt on the subsequent analysis. Thus, the sum frequency terms are retained. However, it is found that these terms yield, in the output of the IF strip, components that have relatively weak initial amplitudes and fast-falling exponential factors. Their contribution is negligible except during the very early portion of the IF output. Although it was formally necessary that they be retained here, they could usually be neglected in future analyses.

F_3 contains four terms and F_4 contains six; half the terms run at the sum frequency and half at the difference frequency. Only the two steady-state terms in F_4 carry the directivity pattern coefficient $\sin Z/Z$; the others carry coefficient $1/2Z$. Phases σ and ψ appear in the combination $(\sigma - \psi)$ in the dominant difference frequency terms and $(\sigma + \psi)$ in the weaker sum frequency terms. Thus, although the influences of the two phases are not precisely interchangeable, they are very nearly so; if either one is effectively random, it is immaterial whether the other is controlled.

Figures 5 and 6 show $F_{3,4}$ corresponding to signals $F_{1,2}$ shown in Figs. 1 and 2. The sinuous shapes are seen to reflect the sum and difference terms added (not modulating each other). The $F_{3,4}$ waveform at $Z = \pi$, shown in Fig. 5, will be discussed in detail below. (In these figures, the choice $\sigma = 0$ is shown, and $\omega_0 = 40\omega_1$.)

The IF strip is taken to be a four-stage Butterworth (maximally flat) bandpass filter whose half-power bandwidth is ω_1/Q_1. Such a filter is perhaps rudimentary compared with common practice, but this was the most complicated filter that it seemed feasible to analyze. The analysis requires that the correct bandpass impulse response of the

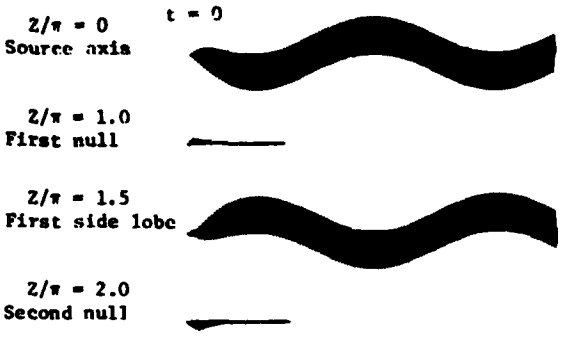


Figure 5

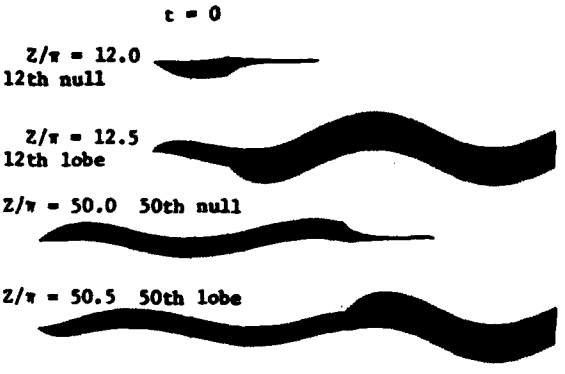


Figure 6

filter be known, and the calculation of that response was difficult; it is given in the appendix.

The output of the IF strip is obtained by convolution of $F_{3,4}$ with the impulse response of the strip. The first segment of the output, F_5, contains 64 terms when evaluated at the limits of integration. The second segment, F_6, contains 96. These were written out, but it was not found to be possible to collect them down to a manageable number of terms. During the study, these were expressed in generic form, the numerous constants were listed in a computer, and the computer program was looped to run through all the terms; that loop must be rerun for each value of t to obtain the output going to the envelope detector.

The envelope detector was taken to be a perfect full-wave rectifier followed by four identical mutually isolated lowpass RC stages. The impulse response of this lowpass filter is

$$h(t) = \omega_c \frac{(\omega_c t)^3}{3!} e^{-\omega_c t}$$

where ω_c = angular cutoff frequency. To obtain the receiver output waveform, F_7, it is necessary to convolve $|F_{5,6}|$ with this impulse response. Inasmuch as the zero crossings of $F_{5,6}$ are not regularly spaced, it is impractical to perform the convolution analytically. The convolution was performed by numerical integration as the computer produced $F_{5,6}$.

The numerical examples shown here were obtained for a source carrier frequency of 3000 MHz, an IF center frequency of 80 MHz, and IF half-power bandwidth of 10 MHz ($Q_1 = 8$). The cutoff frequency of the envelope detector was set at 30 MHz ($\omega_c = 3\omega_1/8$). This choice reflects detailed consideration of the tradeoff between the rise time of the detector and acceptable peak-to-peak ripple. The rise steepness of the output signal, F_7, is controlled almost entirely by that of the IF strip.

DEPENDENCE OF AMPLITUDE ON PHASE

In Fig. 5, the $F_{3,4}$ waveform at $Z = \pi$ is seen to lie almost entirely on one side of the axis. Despite the care taken to eliminate a long-term average value from the source waveform F_0, it is evident that $F_{3,4}$ can deliver a net charge to the IF strip. Unlike the radiation field of the emitter, the heterodyne detector is capable of delivering such a charge. Although not so evident to the eye, every waveform shown in Figs. 5 and 6 has a non-zero average value and delivers a net input charge to the IF strip.

The average values of these waveforms depend on phases σ and ψ. Figure 7 shows $F_{3,4}$ at $Z = \pi$ for two values of σ (90° apart) that maximize and eliminate the average value. If it is supposed that $F_{3,4}$ is expressed in volts, then at $Z = \pi$, if $\psi = 0$, the average value of $F_{3,4}$ is

$$-4.383 \times 10^{-11} \cos(\sigma - 2.7274) \text{ volt-seconds}$$

Moreover, this integral accumulates very quickly and is very nearly equivalent to an impulse that is, for $\sigma = 2.7274$, 207 dB below a unit impulse.

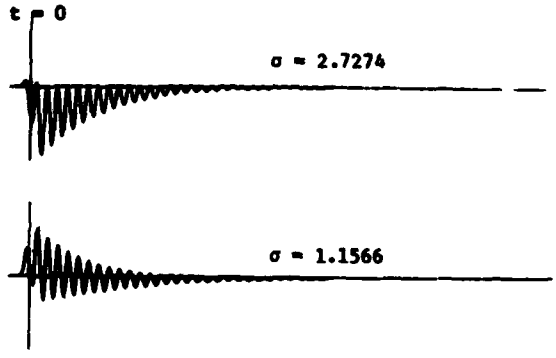


Figure 7

This situation is not restricted to the pattern nulls. At any $Z > 0$ the average value of $F_{3,4}$, in volt-seconds, is (for $\psi = 0$)

$$\frac{\sin\left(\frac{41}{40}Z\right)}{\left(\frac{41}{40}Z\right)}\left[1.799 \times 10^{-9}\cos(\sigma - 2.7274)\right]$$

The ratio 41/40 reflects the choice here of $\omega_0 = 40\omega_1$. It is remarkable that the phase dependence is independent of the off-axis angle variable, Z. Phase ψ plays a role that is essentially equivalent to that of σ; changes of either one will cause the impulse delivered to the IF strip to vary. Such an impulse will generally be delivered unless sin(41/40Z) is zero. The fact that this is "out of step" with the sinZ/Z pattern leads to an intricate dependence of the receiver response on Z.

When this net charge is delivered to the IF strip it excites, essentially, the impulse response of the strip at an amplitude that depends on phases σ and ψ. The equivalent insertion time of this "impulse" varies slightly with Z and with the phases. Moreover, the response that is excited is not exactly the impulse response, and the response waveform varies slightly with Z because the rapidity of insertion depends on the shape and duration of the transient terms of $F_{3,4}$.

Finally, it should be noted that, although the lower waveform in Fig. 7 does not deliver a net charge, it does deliver energy to the IF strip and it does so quickly. It too excites very nearly the impulse response of the strip, but the excitation amplitude is lower because of the absence of the net charge.

RECEIVER RESPONSE TO TWO INPUT SIGNALS

The foregoing discussion underlies a qualitative understanding of the receiver response that is more illuminating than the detailed analysis, in part because the qualitative description is comparatively insensitive to particular values of the several parameters such as ω_0/ω_1, Q_0, Q_1, and so on.

The rise time of the "expected" signal is much shorter than the rise time of the IF strip. Consequently it will excite in the receiver a response that is very nearly the response to a stepped carrier (not the step response of the receiver). The upper portion of Fig. 8 shows that response waveform at the output of the IF strip (oscillatory curve) and at the output of the envelope detector (smooth curve). The "extra" signal is much briefer than the IF rise time, and it excites very nearly the impulse response of the receiver. That response waveform is shown, in the lower portion of Fig. 8, at the IF output and at the output of the envelope detector. These curves are marked off in time intervals of IF periods (12.5 nanoseconds) and in equivalent distances, starting at t = 0 as defined earlier.

Over most of the width of any source side lobe, sinZ is sufficiently large so the response excited by the "expected" signal swamps that excited by the "extra" signal. The receiver output is scarcely different from that found on the

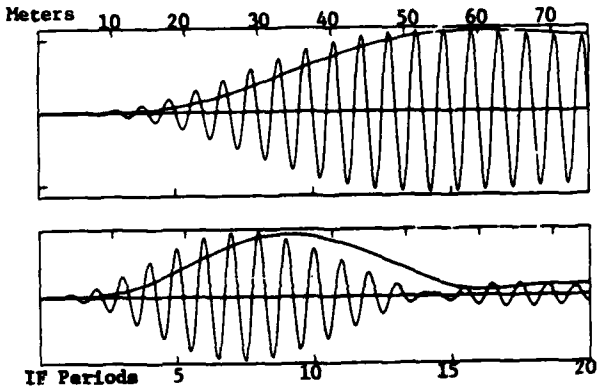


Figure 8

source axis and is essentially the response to a stepped carrier. On the other hand, in any pattern null, the "expected" signal is absent and the receiver response is essentially the impulse response of the receiver. It is evident in Fig. 8 that the latter "arrives" earlier than the stepped-carrier response. (Both responses commence at $t = -Z/\omega_0$, but that is unobservable.) The difference in the two arrival times depends both on the definition of "arrival" that is chosen and on the receiver design; for the definition discussed and the design considered here the difference is about 14.2 meters, depending slightly on the phases.

The difference between the arrival times has nothing to do with the source aperture width or any other feature of the source. It is found below that this quantity controls the magnitudes of the errors made by a TOA system, just as it is seen here to control the disparities of arrival time in a single receiver. For that reason it will be termed the scale length of the receiver, designated by symbol S. It might be supposed that more selective IF designs, with more poles, will exhibit longer scale lengths than the 14.2 meters seen here. If so, then the examples of system error shown below understate the magnitude of the errors to be expected in real hardware.

Close to, but not exactly in, each pattern null sinZ is so small that the "expected" and "extra" signals excite their respective receiver responses simultaneously at comparable amplitudes. Ahead of the envelope detector these responses are oscillatory, their zero crossings are not evenly spaced, and their amplitudes vary differently in time. They interfere, and the resultant is acutely sensitive to their relative timing and amplitude. Moreover, the "extra" signal does not quite excite the true impulse response, and the changing duration with Z leads to interference that is different in detail around each pattern null. Further, the amplitude of the impulse-like excitation changes with phases σ and ψ because of the changing average value of $F_{3,4}$. Although phase changes scarcely influence the receiver response in the pattern nulls or well up in the lobes, phase alters the response considerably at the lobe edges where significant interference occurs. In some instances, the shift of arrival time with phase exceeds S;

thus, although S sets the scale of arrival time shifts, some shifts larger than S occur.

OUTPUT WAVEFORMS

Figures 9 through 16 show examples of the receiver output, F_7. They are selected for illustrative purposes and are not a representative set. In each figure two curves are drawn for phases 90° apart chosen to show the approximate extrema at that value of Z. (In all cases the shapes of the curves change with phase and no pair can be said to show precise extrema.)

Figure 9 shows the response to F_0 on the source axis. This is essentially the stepped-carrier response and is nearly insensitive to phase. For visual reference this response is shown dotted in Figures 10 through 16, but at amplitude sinZ/Z. (In all these curves, zero dB is the steady-state level on the source axis.)

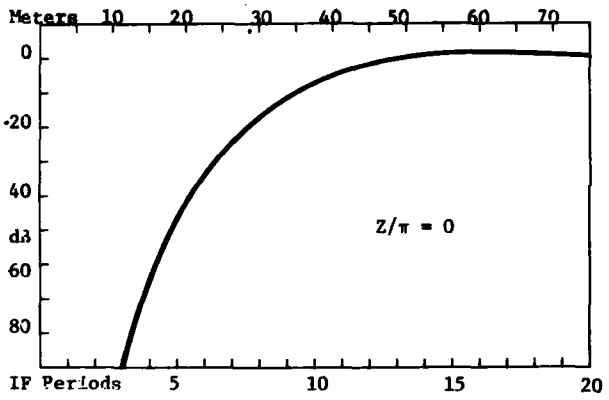


Figure 9

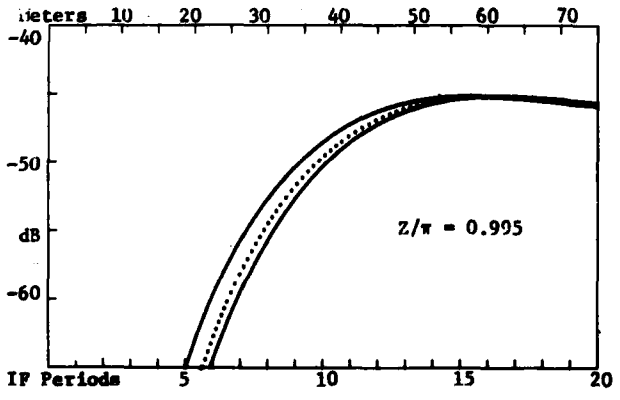


Figure 10

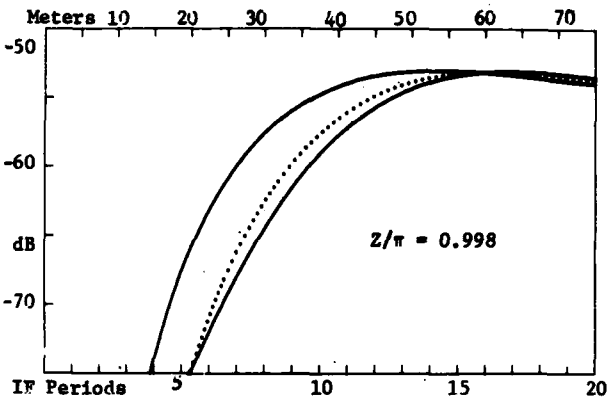


Figure 11

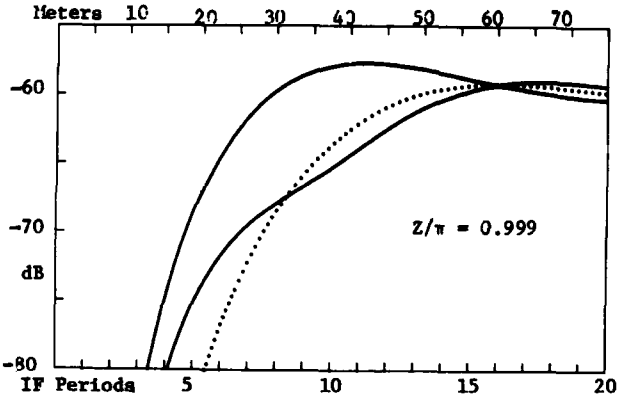


Figure 12

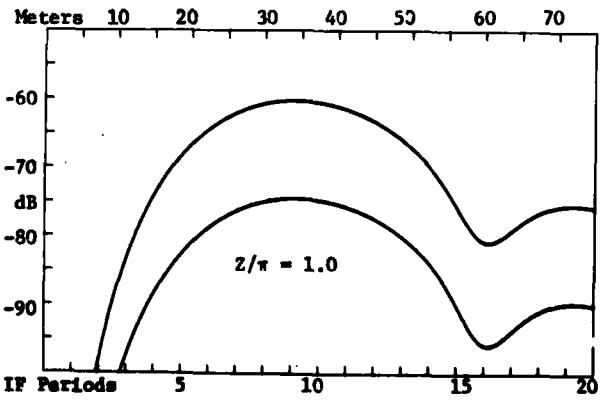


Figure 13

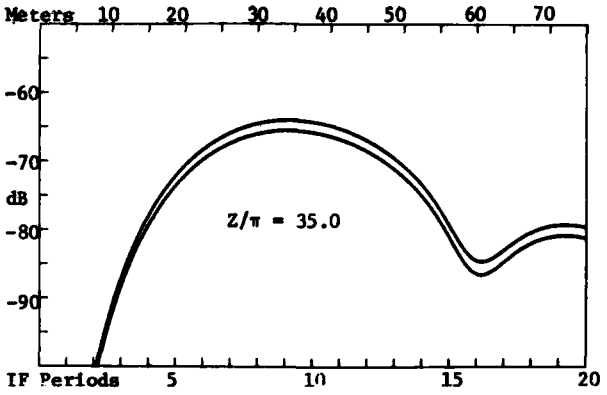


Figure 14

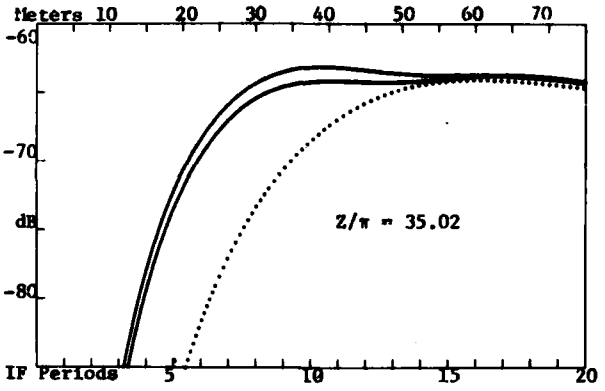


Figure 15

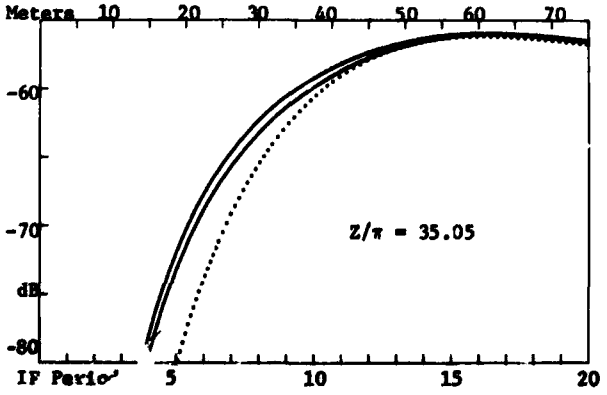


Figure 16

Figures 10 through 13 show the rapid evolution of waveforms over a very narrow angular interval as the first pattern null is approached. In the null itself the two curves are displaced vertically about 14 dB. The two curves in the null are almost exactly the same, and they are essentially the impulse response of the receiver; the two have the same arrival time within 0.2 meters, and that time is about 14 meters earlier than on the source axis. The 14-dB shift in level reflects the presence or absence of a net charge injected into the IF strip by $F_{3,4}$. The upper curve is within a fraction of a dB of the theoretical response to an input impulse 207 dB below 1 volt-second (see above). The level of the lower curve reflects the energy delivered by $F_{3,4}$ even when the net charge injection is zero. No dotted curve appears in Figs. 13 and 14 because sinZ = 0; in Figs. 10 through 12 it should be noted that the dotted curve tends to be bracketed by the solid curves; arrival times either later or earlier than the axial time may occur, depending on phase.

Figures 14 through 16 illustrate the situation in the vicinity of the 35th pattern null. The swath embraced by the solid curves is narrower here, but the waveforms are about the same in the first and 35th nulls. Just outside the 35th null the entire swatch is seen to arrive considerably earlier than the axial signal. Further, the interval of Z about each null over which the curves differ substantially from the axial response increases with increasing Z. $Z/\pi = 35.05$ yields curves that are displaced from the axial response roughly as much as those at $Z/\pi = .995$.

Evidently the troubles caused by the antenna effect are localized near each null of the source pattern; although there are progressive changes with Z, it cannot be said that there is a clear trend to more trouble at large Z than at small. One might feel that the troublesome angular regions are so narrow as to be inconsequential for a system. However, a highly directional emitter offers many pattern nulls, and the analysis suggests that troublesome locations, where appreciable shifts of arrival time occur, occupy roughly 10 percent of the interval $0 \leq \theta \leq 90°$. That, in turn, suggests that about 25 percent of the time at least one of three TOA receivers will be in a troubled position. That is, of course, a rough estimate that depends not least on what constitutes "trouble;" nevertheless it appears unwise to suppose that the narrowness of the angular intervals will make these effects unimportant.

DEFINITION OF "ARRIVAL"

The discussion above deferred making an explicit choice of the event that constitutes "arrival" (although some numerical values mentioned depend on the choice). To discuss system errors it is necessary to make such a choice, and a couple of candidates deserve mention.

"Arrival" might be said to occur at the instant that F_7 first rises past a predetermined fixed threshold level. (In practice, the several receivers could probably make adequate allowance, if necessary, for their different distances from the emitter; this would generally be a small number of dB. Because arrival time would then depend on incoming signal level, this definition might seem inherently disadvantageous. However, it is seen in the figures that the early portion of F_7 rises very steeply; if the threshold is set low enough the dependence on level will be moderate and possibly acceptable. During this study, this method was examined for thresholds set at -80 dB and lower (depending upon operational circumstances, including background noise and interference, such levels might be feasible in practice). Although this fixed threshold method was not used in the balance of the study, it might warrant consideration in some cases.

In this study, "arrival" is said to occur at the time that F_7 first reaches one-half the value reached in the highest subsequent peak. To implement this definition, it would be necessary to delay the signal to allow a retrospective determination, but the needed delay is short inasmuch as the first peak is the highest and it comes only a few IF periods after "arrival."

To minimize measurement errors, it is desirable that F_7 rise as steeply as possible through this 1/2 peak level. The design of the envelope detector was chosen so the steepness is nearly that of the IF strip itself; this steepness criterion is not quite equivalent to the throughput time, which is immaterial.

FIRST COMPUTER EXPERIMENT

The error made by a TOA system in computing the location of the emitter depends on the various arrival time measurements made when the receivers are at different locations with respect to the emitter axis and all, therefore, experience different values of Z. Further, if the source rotates only a little, all the values of Z are changed and an entirely different set of arrival times may be reported. The consequences were examined in a computer experiment that modeled a rudimentary TOA system.

Three identical receivers were arranged in a symmetric array that subtended 90° (two 45° angles) at the source. The receiver array was stationary, as was the source, but the source was allowed to rotate through 45°, starting with the axis aimed at the middle receiver and ending with it aimed at one of the outer receivers. (The representation of the emitter as an aperture in a screen is meaningless for $\theta > 90°$, and that restriction prohibits consideration of rotation beyond 45° if the total array subtense is 90°. Regretably, available antenna theory does not permit a meaningful treatment of signals in the rear hemisphere; indeed, the theory used here is highly questionable at large values of Z in the forward hemisphere.)

The computer determined the three arrival times at each of 10,000 equally spaced source rotation angles over the 45° range; such frequent sampling probed the lobe structure in detail. An approximation was used to find the average arrival time for each receiver if, for each source rotation angle, a large number of pulses was observed so as to average out the effects of phase. This somewhat advantageous averaging over phase is probably unrealistic; it is unlikely that a real system would be able to observe a great many pulses at each source orientation.

Averaging over phase removes the only random element, and the situation is right-left symmetric. Thus, a source rotation over 45° in the other direction was unnecessary. At each of the 10,000 orientations the three arrival times were used to calculate the radial distance from the computed to the correct source location. Those 10,000 errors were arranged to form a cumulative error distribution. The experiment was carried out four times to examine four values of source directionality: L/λ = 30.5, 40.5, 50.5, and 60.5. (Half-integer values were used to avoid placing a null at $\theta = 90°$.)

The four distributions of error are shown in Fig. 17. The curves show the fraction, p, of 10,000 errors that exceeded the distance shown. It is seen that errors larger than 5 meters occurred some 15 to 30 percent of the time, and errors larger than 10 meters some 10 to 20 percent of the time; there were a few errors as large as 48 meters. The four curves have the same shape, and that shape is determined by the arrangement and exact shape of the source lobe structure, the definition of "arrival," and the transient response of the receiver. There is nothing random here.

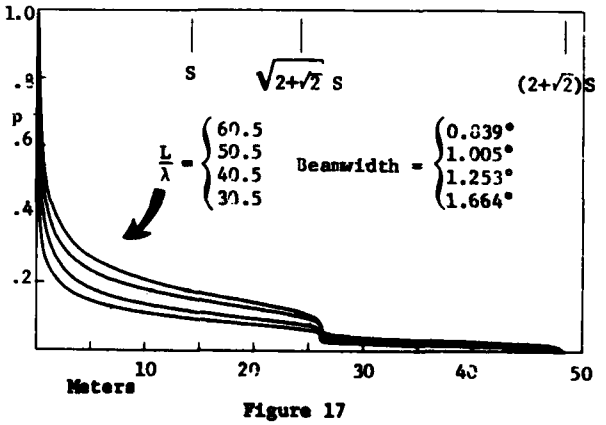


Figure 17

There are three tick marks at the top of Fig. 17. The first, on the left, shows scale length S = 14.2 meters. The curves do not exhibit any feature at this distance because S does not embody the angular geometry of the receiver array. The other two ticks show S multiplied by coefficients that arise from the array geometry; the two values reflect whether one or two receivers are troubled,

and whether it is the middle receiver that is troubled. The curves show a sharp break at the middle tick and end at the third tick. Clearly, S establishes the distance scale for any given array geometry.

This is an unusual kind of error distribution, even aside from the fact that it does not arise from a random process. Although the scale of error is controlled by receiver design, irrespective of the emitter, fraction p--the likelihood of experiencing any given amount of error--is governed by the lobe pattern of the emitter. The more nulls, the more chance of trouble. The number of nulls in the range $0 \leq \theta \leq 90°$ is equal to the integer part of L/λ, and it is seen here that the four curves are arranged in order of L/λ.

TUNED FRONT END

Because scale length S is the difference between the arrival times of the impulse response and the stepped-carrier response of the receiver, the results shown in Fig. 17 depend upon the fact that the "extra" signal excites the impulse response. This brings into question the assumption made earlier that the selectivity of the RF front end could be neglected. Perhaps the "ringing" of a tuned RF stage would "stretch" the "extra" signal enough so the response would no longer be the impulse response. If so, most of the effect might disappear.

To investigate this question, the entire analysis of the transient response of the receiver was repeated, this time with one bandpass stage between $F_{1,2}$ and the heterodyne detector. (More than one such stage would be too much analytic burden. Even one stage adds many more terms in the expressions for $F_{5,6}$.)

Front-end bandwidths of 5, 10, and 20 percent were examined, and 10 percent bandwidth was explored extensively. For the 3000 GHz source carrier considered in numerical work, these correspond to bandwidths of 150, 300, and 600 MHz. In view of the 10-MHz IF bandwidth, it seems unlikely that a designer would use a front-end bandwidth much narrower than 5 percent.

Front-end tuning complicates the results greatly, but no systematic trends with the front-end Q were evident. On the whole, the qualitative explanation in terms of the impulse response remains valid. An approximate subsidiary argument suggests that the front-end Q would have to be about 100 in order to "stretch" the "extra" signal enough so the receiver response would more nearly resemble the stepped-carrier response.

The important changes produced by 10 percent front-end bandwidth appear to be:

1. The troublesome range of Z about each pattern null is narrowed, reducing the likelihood of a receiver being troubled.

2. In the troublesome regions the shift of arrival time with phases σ and ψ is exaggerated, tending to reduce the benefit from 1.

3. The arrival times on the lobe tops and exactly in the nulls are no longer the same across all tops and nulls. Instead, these values remain about 14.2 meters apart and both drift upward slowly as Z increases. Around $Z = 50\pi$, the arrival times of the "expected" and "extra" signals are about 2 meters later than they are at $Z = 0$ and $Z = \pi$.

Although 1 and 2 above conflict, the two together are mildly helpful. However, 3 introduces an entirely new harmful effect. Even if all the TOA receivers are at lobe tops, they no longer report the same arrival times; those situated at larger Z report slightly later arrival. Thus, small system errors become nearly inescapable, and the system makes a great many small errors. The three items above, when taken together, produce a tradeoff wherein the likelihood of large errors is somewhat reduced, but at the expense of many more small errors.

SECOND COMPUTER EXPERIMENT

The first computer experiment was repeated with the following changes:

- The receivers had 10 percent front-end bandwidth.
- The emitter rotated 90°, from one outboard receiver to the other.
- Only 100 receiver orientation angles, equally spaced over 90°, were considered.
- No phase averaging was done and only one source pulse was observed (by all three receivers) at each source orientation.

A new random number was drawn to represent source phase ψ, and three random numbers were drawn to represent the three receiver phases σ, for each pulse. The computer used the exact routine for F_7 to compute the time of arrival of each pulse in each receiver. Thus this experiment more nearly represents what might be expected when the system observes a rotating emitter. This second experiment was carried out only once for source directivity $L/\lambda = 50.5$. The distribution of the 100 system errors is shown in Fig. 18.

The curve in Fig. 18 is not as smooth as those in Fig. 17 because it has only 1/100 as many data. Nevertheless, the shape is quite similar, and a break seems to occur at the middle tick mark. In the absence of phase averaging, errors even larger

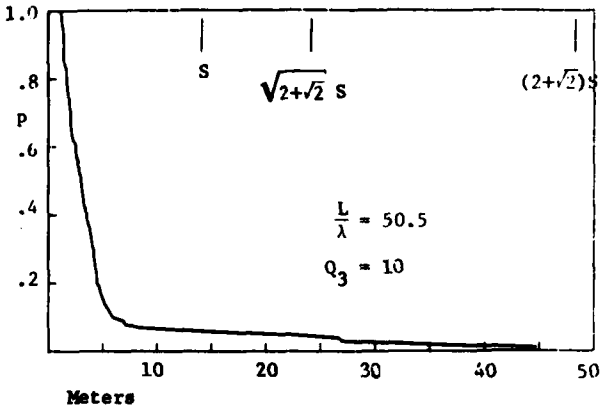


Figure 18

than the right-hand tick are possible, but none happened to occur in this trial. It will be seen that the probability of large errors is reduced considerably, but errors of a few meters are prevalent; the two curves for $L/\lambda = 50.5$ cross at about 5 meters.

CONCLUSION

The "expected" signal excites the stepped-carrier response of the receiver, whereas the "extra" signal excites the impulse response. The arrival times of these two responses differ by an amount, S, which is a feature of the receiver design. S controls the magnitude of the system errors. System errors arise mostly from one or more of the receivers being in or near a pattern null of the source. Consequently, it is the lobe structure of the source that controls the likelihood of error.

The far-field emitter waveforms that underlie this process should be calculated by convolving the excitation going to the antenna feed with the true impulse response of the whole antenna (the response varies with angle). Our inability to solve the wave equation subject to real boundary conditions forces us to use the highly fictitious model that represents the antenna as a hole in a screen. (The shortcomings of the Kirchoff approximate boundary conditions are far less serious than the aperture model itself.) The true impulse response undoubtedly contains, among other things, an oscillatory tail arising from the parasitic oscillations excited through the entire structure. Thus, the far-field waveforms undoubtedly differ--perhaps considerably--from those examined here. For that reason these results must be regarded as tentative and no more than approximate.

Nevertheless, the existence of the "extra" signal is a necessary consequence of the same features that produce directionality. It appears that this aspect of antenna behavior could be troublesome in TOA systems (and possibly in any system that accepts off-axis waveforms and relies on waveform details). Whether or not the system performance is affected significantly depends, of course, on the particular application.

APPENDIX

An N-pole Butterworth bandpass filter contains N mutually isolated RLC bandpass stages. Let

$$j = 2, 4, 6, \ldots \begin{cases} N \\ \text{or} \\ N-1 \end{cases}$$

$$\beta_j = \frac{j-1}{N}\frac{\pi}{2}$$

$$c_j = 2\cos(2\beta_j)$$

ω_1 = angular center frequency

ω_1/Q_1 = angular half-power bandwidth

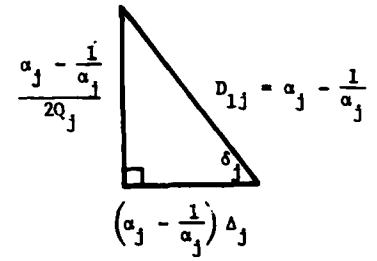
If N is even, there are N/2 conjugate pairs of stages wherein both members have Q equal to Q_1. One number is tuned to $\alpha_j\omega_1$, the other to ω_1/α_j. These quantities are determined by

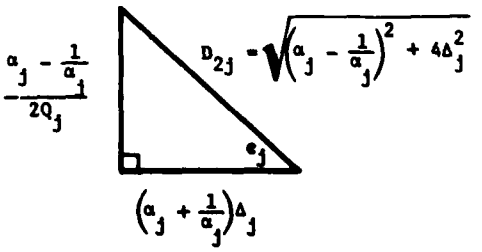
$$\frac{1}{Q_j^2} = \frac{1}{2Q_1^2} - 2\left[\sqrt{1 + c_j\left(\frac{1}{2Q_1}\right)^2 + \left(\frac{1}{2Q_1}\right)^4} - 1\right]$$

$$\left(\alpha_j - \frac{1}{\alpha_j}\right)^2 = \frac{1}{Q_1^2} - \frac{1}{Q_j^2}$$

If N is odd, there is also one stage tuned to ω_1 with Q equal to Q_1.

Define





For N = 4, the impulse response is

$$h(t) = \frac{\omega_1(1+\sqrt{2})}{Q_1 D_{24}}$$

$$\times \left\{ \alpha_4 e^{-\frac{\alpha_4 \omega_1 t}{2Q_4}} \sin\left(\alpha_4 \Delta_4 \omega_1 t + \frac{\pi}{8} - e_4 + \delta_4\right) \right.$$

$$\left. - \frac{e^{-\frac{\omega_1 t}{2\alpha_4 Q_4}}}{\alpha_4} \sin\left(\Delta_4 \omega_1 t/\alpha_4 - \frac{\pi}{8} + e_4 + \delta_4\right) \right\}$$

$$- \frac{\omega_1}{Q_1 D_{22}}$$

$$\times \left\{ \alpha_2 e^{-\frac{\alpha_2 \omega_1 t}{2Q_2}} \sin\left(\alpha_2 \Delta_2 \omega_1 t + \frac{3\pi}{8} - e_2 + \delta_2\right) \right.$$

$$\left. - \frac{e^{-\frac{\omega_1 t}{2\alpha_2 Q_2}}}{\alpha_2} \sin\left(\Delta_2 \omega_1 t/\alpha_2 - \frac{3\pi}{8} + e_2 + \delta_2\right) \right\}$$

---

*The work reported here was supported by the U.S. Air Force under contract DoD F49620-77-C-0023 with The Rand Corporation.

The influence of antenna directionality on far-field waveforms was discussed in Rand report R-1819-PR, August 1975. The entire study is reported in detail in Rand report R-2418-AF, forthcoming.

The author appreciates the helpful advice of Dr. Edward Bedrosian throughout the entire study. He also appreciates the assistance of Marianne Lakatos, who wrote and executed all the computer programs, including those to generate the illustrations, and who helped in their interpretation.