VOICE ANNOUNCEMENTS OF TIME: A NEW APPROACH*

J. Jespersen, G. Kamas, and M. Weiss
National Bureau of Standards
Boulder, Colorado

ABSTRACT

A recent survey by NBS reveals that the voice time announcements provided by radio stations WWV and WWVH are used more often than any other features of the time signals. It is the purpose of this paper to describe some recent NBS work aimed at exploring a different technique for generating voice time announcements. The idea is simply this--a time code from any source, such as those broadcast by WWV, WWVH, WWVB, CHU, or the GOES satellite is translated electronically into a voice announcement. This approach is attractive for several reasons. (1) In many areas voice time announcements are weak and noisy and it is difficult to understand them. In addition, there may be interference from other "standard time broadcast" stations. It is often easy, under such conditions, to detect and error correct a time code. The "cleaned-up" time code is then electronically converted into a noisefree voice announcement. (2) Normal time broadcasts provide voice announcements only at regular intervals of timesuch as every minute. With the code-voice conversion technique, a voice announcement is available on demand. (3) Any time code signal may be used. Thus, the GOES satellite which

broadcasts only a time code, can be made to "appear" to provide voice time announcements. (4) The same time code may be converted into one or more languages at the receiver. This may be important for solving "the problem" of what language or languages should be broadcast from a standard time satellite broadcast. That is, it may not be necessary to broadcast any voice announcement in any language from the satellite--only a code--and the receiver would contain the option to select the language desired.

NBS has developed equipment to convert time codes from several different sources into voice announcements. Although the emphasis in this development work was intended to demonstrate technical feasibility, rapid progress in the production of commercial electronic voice generation units will, no doubt, make the approach suggested here feasible both technically and economically in the near future.

INTRODUCTION

One of the most useful and popular services offered by the NBS is the shortwave time and frequency broadcasts from radio stations WWV (Fort Collins, Colorado) and WWVH (Kauai, Hawaii). Over the years these broadcasts have evolved from a simple service used primarily for frequency calibration, to today's signals which provide time and frequency information by codes and voice. A recent survey [1] reveals that the voice time announcements are used more than any other feature of the signal. It is the purpose of this paper to describe some recent work in the NBS Time and Frequency Division which is aimed at exploring a different technique for providing voice time announcements. This approach, made more attractive by recent advances in solid state electronics, seems ideally suited for time broadcasts. The idea is simply this--a time code from any source, such as those broadcast by WWV and WWVH is translated, by electronic means, into a voice time announcement (See Fig. 1). Thus, although the input time information is in the form of a code, the output is a voice announcement. One advantage of this approach is that any coded time signal can be converted into voice. There are other implications, as we shall see in the next section.

PROBLEMS WITH PRESENT VOICE TIME BROADCASTS

Many time services, particularly those in the shortwave ranges, have difficulties. First, in many areas, signals are weak and noisy and it is hard to understand the announcement. Second, there are many world wide standard time broadcast stations operating at the same frequency, e.g., 10 MHz, so that a particular location may experience interference. Third, because of competition for radio spectrum space, pressures to reduce the bandwidth now allocated for standard time broadcasts may develop.[2] Any such bandwidth reduction will tend to make the voice announcements even less clear, particularly if the present double-side-band, AM modulation (DSBAM) techniques are retained. Fourth, the need for voice time announcements is world wide, so announcements must be provided in many different languages, and in some cases in more than one language by a single stations. For example, the Canadian Standard Time signal station, CHU, broadcasts both English and French. Fifth, time voice announcements are available only at regular intervals of time. For example, WWV and WWVH provide voice announcements only on the minute. This means the user cannot obtain voice announcements on demand. The time-code-to-voice conversion approach discussed in this paper addresses these five problem areas.

ADVANTAGES OF TIME-CODE-TO-VOICE CONVERSION APPROACH

In regions where voice reception is difficult, the human mind is usually the best and only "integrator" availabe to extract the voice announcement from noisy, fading signals. Although it is possible, in principle, to construct electronic devices to integrate voice signals it is not easy or cheap. Although many standard time stations broadcast at several frequencies simultaneously, there are many times when no static-free signal is available.

Compared to integrating voice signals, integrating a coded signal is relatively easy. Most time signals are coded in a straightforward way. As an example, the time codes carried by WWV and WWVH are a modified version of the IRIG-H format. Data is broadcast on a 100 Hz subcarrier at a one-pulse-per-second rate. The time frame is one minute long and each frame containes minute, hour and day information as well as the "UT1 correction", which is used primarily by navigators for celestial navigation. In the presence of noise, errors in this code can be detected by simply comparing several successive "decodes" of the signal. If, for example, three "decodes" step along in the proper time sequence, it can be assumed that the time code is being correctly decoded. If, on the other hand, there is a "misstep", the time code signal output can be made to "flywheel", on the receiver's internal clock, until there are successive, correct decodes.

INTERSTATION INTERFERENCE

The problem of interference between standard time broadcast stations operating at the same carrier frequency is serious and continues to grow as more stations start up in different parts of the world.[3] Various solutions have been considered. One is to stagger the carrier frequencies at 4 kHz intervals within the allocated band The bandwidth for these stations is ± 10 kHz at the standard broadcast frequencies of 15, 20, and 25 MHz and ± 5 kHz at frequencies 2.5, 5 and 10 MHz. The difficulty with this suggestion is that those stations assigned frequencies near either end of the allocated band would have to change from DSBAM to some other form of modulation to avoid "spilling" over into adjacent radio spectrum reserved for other purposes. In addition, some experiments have been conducted with these staggered allocations and there is still serious distortion due to interstation interference in commonly used shortwave receivers.

An alternative would be to have each standard time station broadcast a time code signal which is located in the time-frequency domain so that it can be selected without interference from other stations. As a simple illustration, suppose there are ten standard time broadcasts which might potentially interfere with each other, all operating at a nominal 10 MHz carrier frequency. We could imagine a 10 second segment, say, out of each minute during which each station broadcasts a time code for 1 second, or less, while the other nine stations are off the air. Each station broadcasts in turn throughout the 10 second sequence and the process is repeated throughout each minute. Such a routine would require time coordination among the participating stations, but this should not be difficult in view of the fact that these stations maintain time to within at least 1 ms of UTC.

A specific example of a code that could be used is one that is now being broadcast over the Canadian standard time station, CHU. A complete message is 0.365 seconds long and contains day, hour, minute, and second information--repeated twice for error checking. The code employs the standard commercial 300 baud FSK system with tones at 2025 and 2225 Hz. Since the message is short it does not interfere with the "seconds" ticks provided by CHU.

The time sharing scheme suggested here (or some variation on it) would provide interference-free time code signals part of the time out of each minute, and thus has some utility. But such a time-shared code coupled with a code-voice conversion unit could provide interference-free voice time announcements.

The time sharing scheme discussed here is not the only possibility. Frequency division multiplexing could also be employed. That is, each standard time broadcast station would be assigned a unique frequency in which to broadcast its time code, and other stations would be excluded from this spectral region. One of the difficulties with this approach is that if any aspect of the communication channel is non-linear then new frequencies will be generated which may spill over into adjacent channels. Thus, the spacing between adjacent channels would have to be increased to avoid overlapping signals.

A final example is spread spectrum techniques, which have the advantage that they minimize effects of nonlinear elements in the transmission channel (along with some other advantages) at the cost of a more complex receiver. While the time and frequency sharing schemes discussed in the previous paragraphs are widely employed and are essentially self explanatory, spread spectrum techniques may be less well known to the reader. Appendix I is a rather elementary treatment of spread spectrum as it might apply to several standard time broadcast stations operating at the same frequency, f_c.

SPECTRUM CONSERVATION

The radio spectrum is a scarce natural resource and pressures to use it efficiently mount steadily. As pointed out, a voice time announcement uses considerably more bandwidth than is required to transmit the actual information content of the message. Thus an obvious advantage of a time-code-to-voice conversion approach is that the communication channel requirements are more in accord with actual information content of the message. Note there is presently an international allocation for the broadcast of time from a satellite. This allocation is ± 50 kHz wide and is centered at 400.1 MHz. Becuase of the worldwide need for voice time announcements and the large area covered by satellites a difficult question is raised: "What language or languages should be broadcast from the satellite?" In addition, there is a continuing worldwide need for improved time accuracy, beyond the accuracies shortwave broadcasts can provide. This means that a certain part of the allocated satellite band must be reserved for providing a signal whose time of arrival at the receiver can be measured precisely--the more precise this measurement the greater the signal bandwidth required for a given measurement interval. If the allocated band is used up by numerous voice announcements in a number of different languages, the ability to provide accurate time from a satellite is severely compromised. Aside from this technical difficulty is the political problem of deciding which languages to broadcast. If a code-voice conversion approach is adopted, the satellite decoder could be made to speak any language. The political problem is avoided and maximum bandwidth is available to design a signal which can provide high accuracy time signals.

TIME ON DEMAND

Finally, the code-voice conversion approach allows the user to obtain a voice announcement on demand. He does not have to wait for the next voice announcement to "come up" in the time signal format. Although we have been discussing this feature with primary reference to the standard time broadcasts, it has potential application in other areas. For example, it is technically possible to insert, unobtrusively, time coded information into almost any kind of broadcast signal, e.g., AM, FM, and TV. In principle, all existing broadcast facilities are potential candidates to provide voice time announcements on demand.

One might wonder, why go to the trouble to convert the time code to a voice announcement? Why not simply display the time visually with LED's or perhaps use the code to keep an analog wall clock on time. Certainly in many cases this is desirable. On the other hand, there are many instances when people are processing several different information inputs at once. An airplane pilot may be scanning his instrument panel while he is listening to a voice announcement of the time, or perhaps watching his altimeter reading. Also, voice signals can be heard around corners while visual displays must be within line of sight of the viewer. In any case, for whatever reasons, the NBS survey clearly shows that the most desired feature of the standard time broadcasts is the voice announcements.

DESCRIPTION OF NBS TESTS

Basically, two different speech compression schemes have been investigated for time-code-to-voice conversion: waveform coding and source coding. In waveform coding a facsimile of the original acoustical signal form is retained. An example of waveform coding is to simply sample the acoustic signal at some specified rate, such as 32 k bits/sec, and store the bits in some memory device. To reproduce speech the process is reversed. The bits are read out of memory into a D/A converter which produces a replica of the original waveform. The degree to which the replicated waveform resembles the original waveform depends, of course, upon the sampling rate: the higher the sampling rate, the better the resemblence. Although this is conceptually a straightforward process, there are other related techniques which are easier to implement and which retain the same degree of fidelity, but use lower sampling rates. One such technique, called continuously variable slope delta (CVSD) encoding, was employed in the present tests. We shall describe this technique more thoroughly a little later.

The other technique, source coding, does not reproduce a replica of the original acoustical signal. When this technique is applied to speech it is called voice coding and the devices which perform the codings are generically termed "vocoders". The essential idea behind vocoding is that a model of the human voice system is created electronically or mechanically. A mechanical model might consist of a bellows, a vibrating reed, accoustical resonators resembling the mouth cavity, and so forth. This machine could be operated somewhat along the lines of a player piano to produce speech. The actual data required to generate speech with this machine is simply whatever data needs to be stored on the "piano roll". Sophisticated electronic versions of this machine can produce intelligible speech with data rates as low as a few kilobits or less per second, although the speech sounds artificial. Sophisticated versions of the waveform coding technique, discussed earlier, succeed with rates as low as 5 k bits/sec or so, but speaker recognition is difficult at these rates. For time annoucements speaker recognition is not important, but of course, intelligibility is essential.

A primary goal of waveform or source coding is, of course, to remove redundancies in the acoustic waveform so that channel communication requirements can be reduced, or in our application, to keep memory requirements in the time-code-to-voice conversion unit to a minumum. In the general case one would like to be able to send any arbitrary message. Here some arbitrary acoustic signal is coded (source or waveform) to remove redundancies. Since most real communication channels are subject to noise and other problems which introduce errors, the output of the acoustic coder will usually be recoded (channel encoded) to minimize errors introduced in the communication channel. That is, extra bits of information will be added to the data stream for error correction and detection. Thus, the acoustic coder removes redundancies and the channel coder introduces them again, but in a way that is designed to minimize errors introduced by the channel.

In the case of time signals it is not necessary to send arbitrary messages. A time signal message can be assembled in English from about 30 words: one, two, three, four, . . . minutes, hours, seconds, etc. Thus there is no need to initially code a voice announcement of time. The signal can originate as a code. We could at this point, if we wished, introduce extra bits of information for error detection and correction. But, because time codes are by their very nature quite redundant, this redundancy in itself may be sufficient to overcome channel distortion.

The limited vocabulary also simplifies requirements at the receiver. The memory necessary at the time-code-to-voice conversion receiver need only be sufficient to store 30 words or so of vocabulary. In the next two sections we describe the use of both source and waveform coding to minimize memory requirements at the receiver.

DESCRIPTION OF NBS WAVEFORM CODER TESTS

This system consists of a receiver (WWVB, GOES Satellite, etc.), an A/D converter and storage system, an internal crystal oscillator clock and a processor. The processor performs several functions after receiving a time code from some source. First the processor sets its internal clock only after receiving three consecutive successful decodes. At this juncture the processor turns on the "Time Valid" light. The internal clock is accurate to 1/2 second in 30 minutes so that if 3 consecutive successful decodes are not obtained at least once every 30 minutes the "Time Valid light" goes off and the system will not output a voice announcement.

When time is requested, the processor decides which words, stored in the Read-Only-Memory (ROM), need to be assembled to provide the correct voice time announcement. These words are then converted to analog by the A/D converter. The voice announces time for begins 10 seconds after the initial request concluding with an "on-time" audio tone.

The A/D converter is a continuously variable slope delta (CVSD) modulator which samples the analog signal at 57.6 kbits/second.

The output of the converter is a string of 0's and 1's which indicate increases or decreases in the analog voltage. These bits are stored in memory. The amount of increase or decrease, of the slope, reflected by these numbers, depends upon the past history. At the beginning of a waveform the slope is always some fixed value. After 3 consecutive increases or decreases the slope is increased until the coded digitized output steps over the top, or under the bottom of the analog waveform (see Fig. 2).

When a voice announcement is requested, the CVSD turns the 0's and 1's back into an analog signal using exactly the same algorithm. At 57.6 kilo-bits/sec individual speakers are recognizable.

The necessary vocabulary and the tone have been digitized previously with the same CVSD modulator and stored in ROM. The vocabulary is: the digits "one", "two", . . . , "nine"; the words "twenty", "thirty", "forty", "fifty"; the word "teen", with special preambles, "fif", "thir", "fort"(for 14), and "ninet"(for 19); the single-use words, "zero", "ten", "eleven", and "twelve"; and, finally, the words for the beginning message, "National", "Bureau", "of Standards' Time". The memory required for the words and tone is 91,968 bytes or 735,744 bits. The digits are used both individually and as post-ambles for "twenty" through "fifty". In addition, "six", "seven", and "eight" are used as pre-ambles for "teen". Recording the words requires great care in maintaining a constant level, a continuous cadence, and a flat intonation so the different word "pieces" meld smoothly. It is important to carefully determine pause-time between words, and to notice that the "four" in "fourteen" and the "nine" in "nineteen" are different from the single digit sounds in creating natural sounding speech.

DESCRIPTION OF NBS VOCODER TESTS

As explained, the vocoder type of voice recording and reproduction depends on having prerecorded sound segments in the computer memory. Usually these sounds are simply the necessary parts of words that can be selectively joined by software to make a word. An example illustrates the problem of trying to form words in this manner. Using the common word "you", the programmer looks through the recorded options available in his set of sounds. After trying a number of combinations with various durations (long, short, etc) the programmer settles on EEUUU as best representing the spoken word "you".

Experience with voice generation in this manner soon leads to a rather extensive list of sounds with slight variations. A large list is necessary so that the final word sound is acceptable. Since pauses and sound duration are usually programmable, the number of possible combinations is very great.

This leads to the problems of using a Vocoder technique. The final sound is only as good as the programmer's ability and patience allows. Even if a single word is acceptable in quality, the following word may not sound acceptable when used in conjunction with the first. This problem plus the general one of being unable to completely overcome the machine-like sound of vocoders helped in making the decision to use a straightforward waveform digitizer for natural sounding words.

The NBS tests of a vocoder type of word generator used a commercial voice synthesizer. This was a proprietary instrument that had many variations of a single sound. The results did improve as the programming was altered after listening tests, but the final output was felt to be unacceptable due to its machine-like sound qualities. Some thought was given to finding a very experienced programmer. Several companies that deal with such devices do offer to provide programs. This is apparently based on their experience and on having gained a repertory of words. This approach was considered, but rejected on the basis of cost and quality of the finished work.

Worth mentioning is the ability of listeners to adapt to slightly unfamiliar sounds. Listeners are almost always able to understand the spoken text after only a few practice sessions. The first time is hardest and it was felt that many users of a talking clock would be first time users, i.e. for a telephone time of day service.

Given such a comprehensive list of available sounds, any language can be output from the vocoder. Russian and French words were easily synthesized.

EXISTING CODED TIME SIGNALS

Throughout the world there are a number of coded time signals. We have already mentioned WWV, WWVH, CHU, and WWVB which was the time signal source for the NBS tests described in the previous two sections. Perhaps one of the best candidates for the time-code-to-voice conversion approach is the NBS satellite time code. This code is provided by two geostationary, meteorological data collection satellites (the GOES satellites) operated by NOAA. A time code is included as part of a satellite interrogation signal which is used to communicate with remote data collection devices. The time signal, which contains day, hour, minute, and second information allows the meteorological data to be tagged in time. But anyone within the coverage area can obtain the time signals with receivers now commercially available. As shown in Fig. 3, the signals cover essentially the entire western hemisphere. The signal frequencies are near 469 MHz so they are not subject to the kinds of problems that plague broadcasts in the shortwave band: fading, interstation interference, etc. Thus a voice conversion unit driven by a GOES time signal would provide voice time announcements which are at least an order of magnitude more reliable than those now available from standard broadcasts in the shortwave band.

CONCLUSIONS

The tests discussed in this paper were primarily undertaken to demonstrate the utility of a code-voice conversion approach to time announcements. This approach leads to gains in a number of areas such as spectrum conservation, clear voice announcements in the presence of noise, and voice announcements on command and in any language. It was not our intent to design a device with minimum complexity, although this is obviously an important goal if the system is to be economically feasible. When the tests were first initiated components costs were in the hundreds of dollars. But devices are now coming on the market, in the $10 or less range, providing several hundred words of vocabulary. This is more than adequate for time announcements which typically require a 30 word vocabulary. It seems apparent then that in the not too distant future a time-code-to-voice conversion approach to time announcements is entirely feasible.

REFERENCES

1. "Report On the 1975 Survey of Users of the Services of Radio Stations WWV and WWVH", NBS Technical Note 674, J.A. Barnes and R.E. Beehler.

2. Private Communication R.E. Beehler.

3. Ibid.

4. R. C. Dixon, Spread Spectrum Systems, John Wiley and Sons, 1976.

APPENDIX I

Consider stations broadcasting signals $s_1(t)$, $s_2(t)$, $s_3(t)$. . . $s_n(t)$. Quite generally, the desired signal from the jth station is:

$$s_j(t) = A_j(t)\cos(2\ f_c t + \phi_j t)$$

where $A_j(t)$ represents amplitude and $\phi_j(t)$, angle modulation. Before broadcasting, $s_j(t)$, it is multiplied by a time function $g_j(t)$ which has the property of spreading the original signal $s_j(t)$ over a bandwidth considerably greater than the original. Each standard time broadcast station would be assigned a unique $g_j(t)$. Now suppose we are located in an area where all *n* signals can be detected and we wish to extract, say, $s_1(t)$ from the others. If we multiply the incoming composite signal by $g_1(t)$ we obtain:

$$\begin{aligned}&g_1(t)^2 s_1(t) + \\ &g_1(t)g_2(t)s_2(t) + g_1 g_3 s_2(t) + g_1(t)g_3(t)s_3(t) + \ldots \\ &g_1(t)g_n(t)s_n(t).\end{aligned}$$

If the $g_j(t)$ signals have the property that $g_j(t)g_j(t) = 1$ and $g_j(t)g_k(t) = 0$, then the only output of the multiplier is $s_1(t)$, the wanted signal, while all others are supressed.

As a final point, we have assumed that all n signals entering the multiplier are spread at the transmitter by some function $g_j(t)$, which in actual practice will probably not be the case. But even here, unspread or cw type signals will be spread by the multiplier $g_j(t)$ at the receiver, so that spread spectrum technique still yields an advantage.[4]

Figure 1. Code to Voice Converter

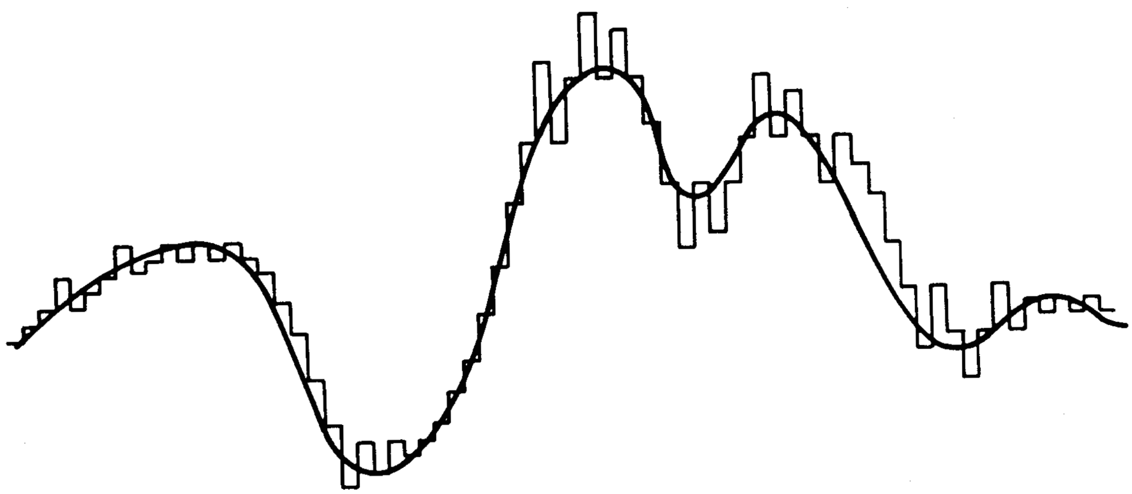

Figure 2. This Figure Demonstrates How the Magnitude of the Converter Output is Related to the Slope of the Waveform

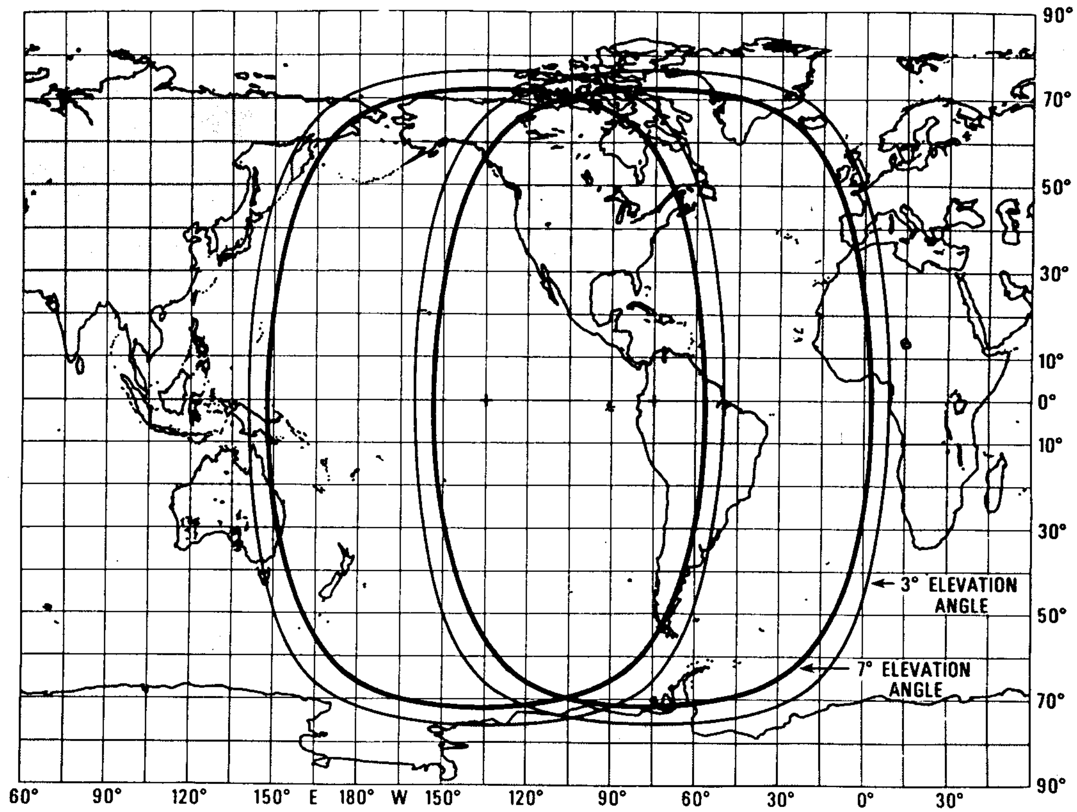

FIG. 3. Coverage of the GOES Satellite.