Jacob Gurin

Up to Our Ears in R/T Intercept

Mechanical improvements promise more efficient handling, but keeping afloat in the future floods of voice traffic remains a problem

Our prospects in the realm of voice intercept are for increasing embarrassment accompanying floods of riches. While no accurate estimates can be given of the volumes to be acquired, there is every reason to believe that our efforts to process today's influx, efforts which severely strain our resources, will be dwarfed by the struggle to come. The mind reels at some of the warnings of tidal waves of multichannel intercept, but even after applying the healthy discount usually granted to statements from prophets of doom, we should allow ourselves a few shivers of apprehension. After all, a 600-channel signal, intercepted only eight hours a day, could require about 600 linguists to listen to all the conversations *just once*. If it takes only two hours to transcribe one hour of intercept (and we know that it often takes longer), 120 linguists would be required to transcribe the best 10 percent of the take. That would take care of only one of the multichannel signals we know to be around, and not the largest one at that. And we could just as well contemplate the increases required in processors for 1200- and 2400-channel signals.

Voice intercept shares with other Sigint modes the problem of keeping processing capabilities abreast of improvements in collection effectiveness. A quick examination of the outlook for the next few years reveals a number of giant steps to be taken in the collection area, steps which must result in enormous increases in raw voice intercept. Are we inviting disaster? Are we starting something we won't be able to finish? Are we asking a guest into our home who will quickly fill all the available space and will regretfully but inexorably force us out into the street? In the light of our mission and the already demonstrated value of voice intercept as a source of intelligence, the only reasonable and responsible answer to such questions is: Not if we keep our cool and prepare ourselves to deal effectively with this lucrative source.

Right from the beginning it is important to recognize that consideration of the awesome prospects for the future should not result in paralysis in current activities. It is just too easy to dismiss the need to work strenuously on a 24-channel signal because the future promises us a 600-channel version. It is morally wrong to justify a mounting backlog in current high-priority material because the backlog is so small compared to what is expected for the future. The problem of coping with a large volume of voice intercept faces us *now*, although it is expected to gain in intensity, perhaps exponentially, in the near future.

The only sensible way to begin consideration of how to deal with significantly large increases in the voice area is to apply the same common sense doctrine as should apply to other Sigint modes: Select only that which is needed—discard the remainder. This may seem so obvious as to need no amplification, yet it is in attempting to carry out this simple principle that most of the trouble occurs.

The sooner we discard the unwanted, the better for the whole system. Logistical problems in supplying and transporting reels of magnetic tape constitute a major headache in the cryptologic community. In addition, it is terribly important that the transcriber's time be directed as much as possible to that intercept which contains information that justifies his efforts. One of the weaknesses in present practices is that intercept that is not worth listening to often manages to take up the time of several linguists in the processing chain. Here we must make a distinction between single-channel exploitable voice (or the narrowband equivalent from a multichannel signal) and any other kind, since it is only the former which can be understood while it is being intercepted. Ideally, only that traffic which is deemed by the intercept operator to justify further processing should be recorded and passed into the processing chain; all the remainder should be ticketed for oblivion. It has not been the general practice to assign to the intercept function the most experienced and skillful linguist-analysts. Yet that is the level of competence which would be required for the decision to retain the conversation or lose it forever.

Discriminate—Or Else

Of course there are sizable boulders in the path of early discard of unwanted intercept. First of all, it is not always easy to recognize what is valuable, even when there are clearly defined criteria for assigning values to information—something which is not characteristic of intelligence activities. Then, more frequently than not the quality of the incoming signal is just too poor to permit full comprehension on first listening. There is also an understandable reluctance on the part of the intercept operator to discard material which has been acquired after so much labor and expense, and which could conceivably be of some value.

In spite of these difficulties, it is in the discrimination between wanted and unwanted signals at the intercept stage, or at least before full linguistic treatment is applied, that our chances for survival lie. If we know which channels in a multichannel system carry the most valuable information, it may be possible to discard with confidence the remainder of the signal without even a cursory listening. But if we lack that convenience, we will have to develop the next best thing—automated processes that will enable us to select, from among an embarrassingly large number of intercepted channels, the few that are worth narrowbanding for further treatment.

We must assume that the number of transcribers to be employed in Sigint will be limited, no matter what the volume of traffic, and that we are close to employing that limit now. In addition, we must assume that the level of output of the transcriber will not be significantly higher than it is now. Such training as is available to our linguists is far more effective in upgrading the quality of their work than it is in increasing the quantity of their output.

So we must think in terms of automatic devices that will permit us to deal effectively with voice intercept. A few mechanical improvements have been introduced into voice processing in the past, but none very startling. We take the *pedal* for granted now, but it represented a real accomplishment when it freed the transcriber's hands from the task of stopping the tape, backing it up, and sending it forward once more. A *variable speed control* was introduced in the AN/TNH–11, the current standard recorder-reproducer, to permit a slow-down or speed-up of the tape as an aid to transcription or as an adjustment for recordings made at incorrect speeds. *Voice-operated relay*, or VOR, recognizes and records voice signals (and non-signals which look like voice) and has been available for tasks where the operator cannot always stay with the recorder, or is tending more than one. With VOR we can avoid long stretches of blank tape and reduce the amount of tape used.

There have been some attempts to design a console that would make the voice interceptor-transcriber's task easier or permit him to function more effectively and accomplish more while he is at his task. But in retrospect these attempts seem half-hearted and less than successful, and a quick look at how voice processing is carried out at NSA or in the field shows that little real progress has been made. If we face the issue squarely, it becomes clear that a good deal more thought and action must be devoted to the development of mechanical aids to the voice problem. Although this is by no means a new task for the R&D organizations, the scale in the past has been small compared to the size of the problem. Without any illusions about the possibility of automating speech processing now or in the immediate future, investments are being made by NSA in research into procedures, techniques and equipment in the hope that present practices can be improved and benefits realized while lessons are being learned for the solution of the problems of the future. Some areas of development activity are examined below.

In-House Efforts

Automatic speaker identification. For a number of years there has been an effort in major firms such as Bell Laboratories, IBM and RCA to develop a device or process which would recognize a speaker without resorting to human ears. A number of claims have been made for success of varying degrees, but nothing has yet been developed that is recognized as useful for Sigint. The most promising activity seems to be the local one on SPIDR (Speaker Identification Routine), which is being pursued with vigor in NSA.* While results of tests with laboratory tapes have been most encouraging, work with real signals has shown that not all possible conditions have been anticipated, and additional analytic work is being done in the attempt to make SPIDR a practical device for voice operator analysis.

Language discrimination. While there is at present no urgent need for automatic recognition of spoken languages, this is an intriguing problem with probable application to future intercept, especially tapes from international common-carrier (ILC) signals containing different languages which appear irregularly and unpredictably in a variety of channels. The object of the study is to develop a process that will automatically indicate the language spoken, given some sample of speech. This implies the discovery of some spectrum measurement which is large between any two samples of the same language but small between samples of different languages, or vice versa. Frequency spectra of speech samples in Russian, German, Vietnamese and English have been measured. This task is of relatively low priority at present.

Vietnamese digit recognition. In some ways this project can be looked on as word recognition at the most elementary level. Since the Vietnamese language is monosyllabic, each number consists of a single syllable with its pitch inflection. In addition, the context which ordinarily provides so much help in speech intercept is missing in the bald recitation of strings of numbers. Live intercept of typical quality is being studied, and both spectral measurements and pitch information are being used to establish differences. At present it seems that one digit will be easy to recognize, one very hard, and the others in between.

NSA is not alone in working on the recognition of Vietnamese spoken digits. Efforts are also under way at RCA

*For a brief description of SPIDR see the present writer's article in the Fall 1969 *NSA Technical Journal*. That article also summarizes other efforts at speaker identification by machine.

Camden using slope features, at Rome Air Development Center using speaker identification equipment, at Litton Industries using Rome ADC equipment, and at Federal Scientific Corporation using pitch information. It is still too early to predict the outcome of all these efforts in terms of a practical device for Sigint use.

New recorder-reproducer. A new tape recorder, the AN/USH-13/14, has been developed in NSA especially for the voice problem, both for intercept recording and for the playback operation. It represents a considerable advance over both the commercial models and the present standard recorder, the AN/TNH-11, and has a number of attractive features, some of which are:

Simultaneous recording and playback of four information tracks, permitting recording of time code and reference frequency where required in addition to the usual two tracks;

Ease in repeating a time-interval on the tape;

Speed range from $1\frac{5}{16}$ to $3\frac{3}{4}$ ips, providing a recording bandwidth of 4 to 16 kHz;

Use of half-mil thickness tapes as well as the standard one-mil tape;

Ease in loading;

VOR capability;

Minimum operator attention;

Remote control capability;

High reliability;

Ease of logistics and maintenance;

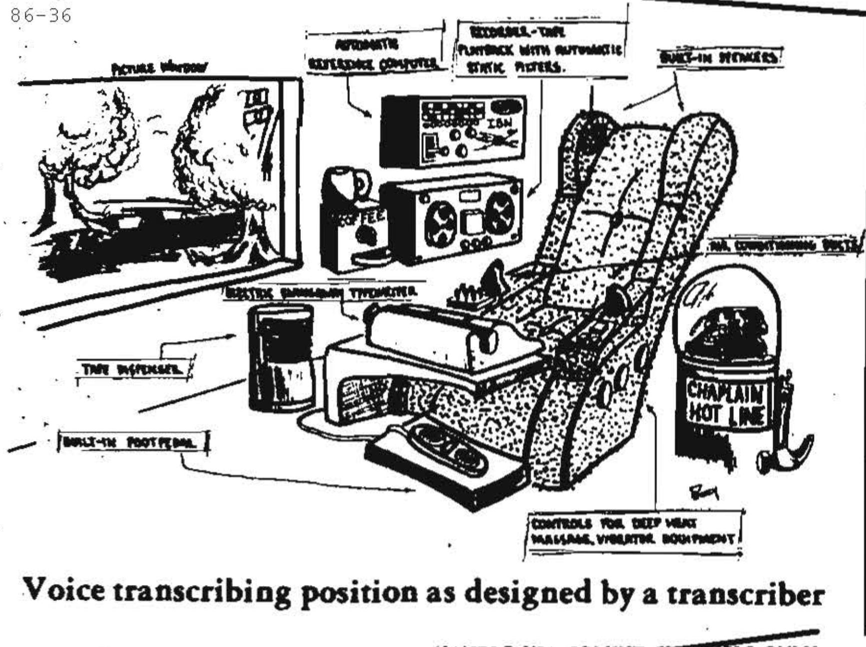
Small size.

[redacted] system, developed at NSA, plays back recorded audio information at speeds faster or slower than the original recording speed without changing the pitch. The speech is sampled at a rate that varies with the playback speed, and a portion of each sample is removed in the case of speech compression, or repeated in the expansion operation. In the compression, the compressed speech is fed into one ear and the discarded portion is picked up and fed into the other, creating a stereo effect. It has been found that even though the signals presented to the two ears are not in their proper time relationship, using both ears and giving the listener as much of the signal as possible improves the intelligibility of compressed speech. It seems that fusion of the two distinct signals takes place in the central nervous system, where their temporal order is restored.

[redacted] is now being tested in several operational areas. It should prove most useful in transcription, either in playing back a difficult passage slowly (down to one-half speed) or allowing the transcriber to rush through other material (up to three times normal speed).

[redacted]

(b)(3)-P.L. 86-36



Voice transcribing position as designed by a transcriber

(b)(1)
(b)(3)-50 USC 403
(b)(3)-P.L. 86-36

Contracts with institutions of higher learning and research centers encourage the kind of general speech research which, it is hoped, will provide the basis for practical solutions to Sigint problems. Work at U.S.C. is characterized by painstaking measurements of the components of several Asian languages. Speech synthesis is pursued at the Haskins Laboratory, and at M.I.T. attempts are being made to develop analytic expressions for spoken utterances. A contract with Purdue involves the attempt to recognize key words in continuous spoken text.

In quite another area, NSA has let a contract to develop recommendations for improving the environment in which the transcriber works. This investigation is concerned with equipment (headphones, tape playback, typewriters, loudspeakers), physical environment (distractions, ventilation, lighting, noise), and such things as the proportion of the transcriber's working day that is spent on various tasks. As a result of this investigation, measures for the improvement of the environment are expected to be identified and steps taken to improve transcriber effectiveness. Even a small improvement at each position could result in a respectable increase in the effectiveness of the total transcriber element in the cryptologic community.

The Future

It is not difficult to see that actions now under way in the R&D community are but a modest beginning in the task of mobilizing electronic and other non-linguistic assistance for the Agency's burgeoning voice problem. Put in perspective, however, that problem can be seen as utilizing a significant part of the total speech research talent in the United States. While the roll call of firms and research centers involved in some aspect of research into speech may seem impressive, the number of principal or major researchers is really quite small, and the departure of just one key scientist is often enough to doom a large project to inactivity and inconclusive results.

Are we doing enough and are we doing the right things? Do the prospects for the future justify a re-ordering of present efforts and a greater investment in research programs? With resources so limited, how can we be sure that we are really doing first things first?

There are many directions in which research activities might be pursued. Short of attempting to automate judgment, which at present seems both irresponsible and irrational, there should be no constraints on our thinking of possible solutions to the problem of coping with increased volumes of voice traffic.

Are we enjoying maximum hearability on our tapes? Even the native linguist must listen repeatedly to key passages because of noise, fading and other impediments to audibility. There have been instances of transcribers electing to work in cramped conditions in a van parked under the antenna rather than endure the hearability loss which was their lot at the end of a long cable carrying the signal to the operations building. There are many other sources of noise, and even small improvements sometimes make the crucial difference in hearing or not hearing what was said.

Are we tied down too tightly by traditional procedures in our processing methods? As far back as we care to go, there are the same tried and true methods: Get the signal on the receiver; start the tape recorder; slip the completed tape into the proforma envelope; mark the envelope and send it to the transcriber; re-listen to the tape; make gists, extracts or full transcriptions as desired by the customer. Should we be devising quicker ways to get at the money veins?

Are there transcriber functions which could be assumed by machines, freeing the man to perform judgmental rather than routine actions? How about time-consuming tasks such as looking up words in dictionaries, consulting callsign books, atlases, organization and personality listings? Should reference materials be available at the transcriber console in response to the touch of a button? Should they be supplied in oral form on request?

Are we delving as deeply into the nature of oral communication as we might in our attempts to find determinants for selection of desired portions of the take? Should we put more effort on discovering the nature of aural recognition processes in the human brain? What should we be doing about prosodic features in spoken communication—accent, stress, emphasis? How about inflection? Should we try to get a device to recognize heavy sarcasm, or a verbal sneer? What about the statement "It ain't what he said, it's the way he said it"?

Can we place a larger proportion of our linguistic and analytic experts at the intercept sites so that retain-or-discard decisions can be made before the take overwhelms processing capabilities? Can we overcome the very natural inclination to send the material on for decisions rather than accept the responsibility for an irreversible action?

There have to be answers to these questions, and we must work to get them. But one of our biggest problems is that we have gone on so long before beginning to ask the right questions. The questions above are offered as *some* of the right ones. If you can reshape and improve them, or offer new ones, please do so—you can help define the problem. For that is the stage of development in which R/T intercept now finds itself.