DR. BOYLE: Ladies and gentlemen, we will get cracking on the last session of the hardware panels. We are going to be talking about a very interesting aspect of small systems. This first part is particularly concerned with audio input, and, as a corollary to this, audio output.

We have many times regretted the number of errors we get from keying in information. There seems to be a considerable loss of information between our thinking and our fingers, unless we are very experienced typists, and most of us doing editing or digitizing work are not in this category. We do, however, have a very good communication link between our thinking and our speaking. Most of the speech input units that have been developed are only good at working on single words, but have quite a good vocabulary, and can, in fact, have multiple vocabularies. So you can have one vocabulary for one type of data and a vocabulary for another. It is quite another aspect if you want them to understand complete sentences the syntax of sentences and so forth.

The audio output side, which will be discussed by a small company from Palo Alto, Telesensory, is also very effective as a feedback. In other words, if you are digitizing the numbers on a hydrographic chart, and say "393," the unit immediately answers back after the analysis, and in its own voice, "393." Without conscious thinking an automatic loop is formed in your mind. If it misrecognizes your words because you were out drinking too much beer the night before and it comes back as "494," you say, "No, 393." It if keeps answering back "494," well, either you go and have a drink of water, or you retrain the program, because your voice has changed over the period. I find it is very important to have audio feedback whenever I have audio input.

I have always used speech input -- as a single word input with a vocabulary of about 50 words -- as one of the exercises I give to fourth year electrical engineering students at the university. We give them a microphone, a bunch of transistors, a mini-computer, and say, "Go to it." They have about 20 hours altogether to do this work. Some of them only get about 50 percent recognition, but others obtain 95 percent readability from the results. It is not a terribly difficult thing to do if you keep it simple, and, particularly, if you ask them to only do the digits and maybe you say, "Well, don't call it five, call it fifer," so that it is distinguished from "four" quite clearly. It is a very good exercise for students. I should think we have done this project about 25 times at our university with different students. All have made it work.

Against the idea of expensive military systems, you might be interested to know there is a small company in this area -- actually, it is in Los Altos on San Antonio Road, called Heuristics -- making one of these units for the computer hobby market. It is supposed to be on the market, in fact, this month. I saw it about six months ago in their shops; it works extremely effectively, costing about a thousand dollars. It has a good vocabulary and full learning routines. It is done on micro processors with the program in Basic. It is a delightful little unit.

I also visited another small company, Telesensory, and we have Mr. Walko on this panel. Unfortunately, Mr. Obester was called away to Washington and Mr. Walko, at the last minute, has offered to tell you a few of the things that are going on in audio output.

I was very impressed with their pocket calculator for the blind, which has audio output. It had a switch on the side and you can change it from English to German and to Arabic. Guess why it's in Arabic? I took a recording of this; I do not know how it is going to come out over the microphones and the loud speakers, but I will later play a second or two for you.

I am then going to ask Major Broglie from Rome Air Development Center, who has been involved in the audio input side and knows what different people are doing. There are many different systems. I know that the USGS at Reston are experimenting for geographic name input. Other people are using them for audio input of hydrographic soundings.

This is a small panel, but I think we will be able to cover most of the areas and enable you to ask questions. I feel audio is a very simple, powerful addition to both digitizing and interactive editing. Now the recording. (Whereupon, a tape recording was played by Dr. Boyle). That is audio output from a micro processor. They are just bits stuck together to make them sound as if they are words. All right, Major Broglie.

MAJOR JAMES BROGLIE: Thank you, Dr. Boyle. I would like to present, from the RADC point of view, what has gotten us into the area of voice input for cartographic applications. Rome Air Development Center, as has been mentioned by Mr. Jamberdino earlier, is a full spectrum laboratory dealing with basic research through experimental development, advanced development, and, finally, engineering development. I would like to present this paper in four parts, starting with the history of automatic speech processing, which led us to the application of voice input for cartographic applications; some experimental work that has been accomplished and published; a brief description of the ongoing current work at RADC; and some plans for the future.

Automatic speech processing is actually a three-pronged development undertaken by RADC: message monitoring, voice control systems, and speaker verification. These three lent themselves to the idea that we could take this technology and apply it to cartography. With the advent of voluminous cartographic data becoming available and the fact that we had a technology available, we decided to marry the need with the technology.

Many of you are familiar with bathymetric data. There seems to be, for a simple system using a very limited vocabulary, an area that would benefit mostly by this, and that is the area of multiple, limited vocabulary, sounding information that needed to be put into digital form such that software could operate on it to produce a series of bottom contours.

The existing cartographic or bathymetric input that we decided to operate on had to do with a digitizing table and a keyboard entry system that generated a set of cards that were then run to provide the information in a digital form. We decided to introduce the system at the Hydrographic Center of the Defense Mapping Agency. These are the experimental hardware units that were involved with the system that we evolved. I will describe each of them in detail but I would just like to give them all a name right now. We have the word recognition module, voice input headset, and an IBM card punch. There we have a Sony tape recorder that we are using to pick up the spoken words that we can do a statistical analysis. Here we have a Bendix digitizing table.

Basically we have here a voice input remote control unit, an alphanumeric display, pre-processor, Nova 1200 mini-computer, a tape unit, and a digitizer interface. In the voice input remote control unit shown here in detail we have the capability of selecting the operator, the word we wish to train, (and a gain control for various types of voice), a training mode, and a prompting or displaying unit, which also allows for viewing several of the past entries. The training procedures that are used are speaker dependent using unconnected speech. We use a limited vocabulary, in this case, the digits zero through nine, and five control words. The unique part about the system is that it can be re-trained instantly such that if there are speech difficulties due to colds, or tonal differences, you will get a re-training indication through the error button. During the training mode the word that we wish to train is displayed on the console, and then, through a preset number of repetitions, is entered into the pre-processor where a template is electronically formed, the template will be stored and used to match the speaker's voice with the stored template during an operational mode.

Here is the cursor that was used in the experimental work. You will notice one of the important things we did here (to provide ourselves an instant edit capability) is that we employed a visible feedback on the cursor, a light emitting diode unit. As the speaker speaks the words into the boom microphone, this system automatically, visually feeds it back to the operator, thus allowing him a complete concentration on the work area. Errors or misrecognitions are easily backstepped through a single word voice control, thus allowing proper numerical entry to occur.

This is a slide showing one of the operators that took part in our experimental work and the method of data entry. In the foreground you see, again, the keyboard that used to be used in the manual mode of digitizing. Here we move into our current outgoing work. This is a slide showing the hardware components of the system that is called the bathymetric data reduction system. The experimental work that was done was statistically analyzed and gave forth results that, while not absolutely conclusive as to certain size, numerical groupings, (as far as whether they are more rapid by input or not than the manual input), led us to believe that the results merited further investigation, particularly

from a production environment point of view. We therefore went under contract to provide this bathymetric data reduction system in two phases, using both manual and voice input. The first phase hopefully should be concluding within the next couple of months and is called the basic operating capability, which would: create for us a data base in our data management system for both voice and manual input capabilities and yeild several algorithms operating on various types of input. The final phase of the program would then be to convert several other existing algorithms that are running on a large computer to a smaller system. For the experimental work we discovered that certain modifications to the cursor would prove to be useful from an operator point of view. Consequently, the cursor was modified. It is lighter weight, somewhat smaller, and still gives us the capability for a rapid visual feedback of the spoken word. Where do we go from here? Plans for the future include connected speech. Experimental systems under development now at Rome Air Development Center have provided 600-word vocabularies. These are node structured, and allow very rapid entry through high speed techniques that have recently been developed. We will be addressing even larger word vocabularies. A thousand words does not seem unreal at this point in time. Areas that are being in our advanced development model include digital radar landmass simulation, where highly structured feature analysis data tables will be used to describe certain characteristics of cultural features that are then fed verbally through the system to be stored on tape; and in flight information production program, again, where we have a very highly structured data input format. Let me just summarize where we have been, why we got there, and where we are going. The technology that we are applying here for voice and cartographic entry was developed from ongoing programs at Rome Air Development Center in automatic speech processing. This work was done under the auspices of Dr. Bruno Beek, Captain Bob Curtis, and Mr. Dick Vonusa. They are taking it from the very basic stages to the areas where the applications could be transferred to our systems. I mentioned some experimental work that we did. The results of that work were published in the fall 1977 proceedings of the ACSM. Currently, we are under development with a bathymetric data reduction system. This system will provide an on-line operational use of a voice technology. Our plans for the future include larger vocabularies, connected speech, and possible applications areas that use highly structured data format for input. Thank you for your attention. (Applause.)

DR. BOYLE: Thank you very much indeed. I will now ask Mr. Walko to tell us a little bit about the speech output and some of the interesting new developments are being done. Incidentally, I would say that this area around Palo Alto, Menlo Park and these places is a very exciting area for me to go around and to see many small companies doing a great deal of work on optical character recognition, audio input, audio output and so on. The excitement of the development in these areas is very real. These are companies you may not have heard about, but if we serve as a little bit of an introduction to their names I think it is serving a useful purpose. We always like at these meetings to bring in local companies around an area of a meeting, and I think that the optical and audio input we have not at this conference on hardware talked about optical character recognition at all. It is one of the things we ought to have done, really, but it was too much to try to get in. But the audio one is quite an exciting part of this area. So, we have given you a few introductions. If you want to hear more about them, well, get in touch with them yourselves. Mr. Walko, will you please tell us a little bit about your work.

MR. WALKO: Thank you Dr. Boyle. This should be rather interesting in two respects. First of all, Telesensory Systems has nothing to do with cartography at all, not anywhere close to it. Second, I was not supposed to speak today, as Dr. Boyle mentioned. Dr. Obester was to speak. I will give you some of the highlights the way I know them. My position at Telesensory Systems is Speech Products Manager. I have been developing marketing strategies for new developments that we have in devices that are of interest to original equipment manufacturers. I would just like to give you a little description of the company and the background.

The company started around a product called the Optacon, which stands for optical to tactile converter. All the products that Telesensory makes are for the handicapped, and especially the blind. Now the Optacon was developed with the help of OE grant to Stanford University and SRI to develop a reading system for the blind. What the Optacon does is to take a printed material and convert it to a tactile reproduction. A blind person would have in his hand a small camera bout the size of a pocketknife. He would then scan print with this camera. As he does so, he has a device with him about the size of a cassette player with a small array of 144 pins in it. As each character is scanned, there is no character recognition but simply a reproduction in vibrating pins. So under his index finger, as he scans the page, if he covers the "C," a "C" will be reproduced in the pins. In reading

letter to letter he is able to develop words and develop reading rates of between 40 and 60 words per minute. That was the device that started the company. The whole reason behind Telesensory was to take the Optacon from the university atmosphere out into the marketplace so that the maximum number of people could benefit from the device.

With the profits made from the Optacon, we began to look into other areas, other products for the handicapped. One thing that we discovered that was a real need for the blind was some type of calculator. Up until the point that we developed our calculator, most blind people would be using an abacus to do their mathematics. We searched the market and found there were a number of calculators already available that had Braille outputs and some printed outputs but nothing was really useful for the blind. So we began to see that the only thing that would really be useful would be if the calculator could talk. We began doing research into speech technology and the development of our own speech technology. At that time a professor at Berkeley, Dr. Mozer, was doing some work, and we joined forces with him and developed the first use of custom LSI, or large scale integration, for generation of speech. So, the calculator, when it came out in 1976, was a first in two respects. It was the first use of a microprocessor in a calculator, and it was also the first use of a customer LSI to develop.

I have the SPEECH PLUS Calculator with me right now. This is a small, hand-held unit. It works on rechargeable betteries, and it is a six-function calculator, similar to any standard electronic calculator that you can find on the market right now. The difference between this one and the one which you might buy is the fact that this one is able to talk. I will give you an example of how it sounds when it is turned on. (Demonstrating). This is a sample of a calculation. So, in fact, the calculator speaks on every key entry and every function entry, and then will repeat the answer at any time. Whatever is in the display you can read out again. (Demonstrating.) After the development of the English vocabulary, we went into a second language, which was German. (Demonstrating.)

When the calculator was first developed we thought there would probably be a sighted market for it also, and it was on display in Neiman-Marcus in New York and in a couple of other large department stores. While it was on display at Neiman-Marcus -- it was kind of funny the way this happened -- an Arab sheik and his entourage came walking through the store and saw the calculator and just fell for it as a novelty item, and decided that they had to have some themselves. So they bought a few English models. The sheik

decided it would be really something if he had an Arabic model. (Laughter.) So he assisted Telesensory by putting up some of the development money for a new vocabulary, which runs somewhere around $10,000, and bought quite a large number of these calculators to disseminate in his country. Later, I should explain, he did find out it did have a good use in the blind community, and it turns out he was quite a benefactor in his country. The name of the family is Binladen. They bought these calculators to disseminate to the blind, so it was not just a frivolous type of thing. But it is kind of interesting the way it came about. That was the calculator development. By the way, the calculator sells for $395, if you might be interested in the price of something like that. When we came out with the calculator, and since it was a first in speech technology, many companies were interested in the use of the technology in their own products to make their products speak also. At that time our company charter was such that since we were interested only in products for the handicapped and especially the blind, we really did not want anything to do with any outside manufacturers, even though we had something that was quite unique and something other people were interested in. Since then our attitude has changed greatly, basically because we need more and more money for development of new products for the blind. One way to generate profits would be to sell this technology to other manufacturers for use in their product.

This is a module right now that is available in two 64-word vocabularies. This sells for around $179. Two modules are available with fixed vocabularies. One of them is an ASCII vocabulary, so it is capable of doing the alphanumerics of the keyboard. The other one is a standard vocabulary of 64 words. It can count up to a million, and also has the capability of saying some measurement terms like "second," "degrees," "pound," "ounces," things like that. We also have the capability of custom programming any vocabulary that is necessary. Just a short explanation of what is on here: There is a microprocessor we developed, a ROM, read only memory. The microprocessor takes information from the memory as to what full names to string together, how long to say them, the duration of the word, and then send signals through the audio filtering circuit to the audio circuit for the speech.

When the Optacon was first conceived, the idea even at that time was that the ideal thing for it to do, instead of just reproducing a letter, would be to actually speak. This system is now under development at our company, and it is pretty well developed at this point in a PDP 11 computer. Work is being done to reduce it to micro circuits to produce what we term "the spoken word output accessory to the Optacon." This right now is being billed as

probably the most sophisticated consumer product ever developed. What it will do is, as you scan the page of printed material or numbers, use a text to speech machine to actually read to you. I have an example of the speech quality. We are also developing a new method of speech synthesis in conjunction with MIT. Professor Allen is doing a lot of work there in speech synthesis. I have an example of it. It is a more natural sounding speech, and puts intonation in, and it has a little bit of an Irish brogue to it. It is very clear. Let me play it for you. (Demonstrating.) This was done by just scanning a typed page with those words on it, and the computer spoke as it was hand-scanned.

There is one last tape I would like to play for you, and it will give you an idea of how speech is built up. It is a very interesting presentation that was done recently at a news conference that we had. This is the voice of Jim Caldwell, one of our scientists who is doing a lot of work in speech synthesis. It is a very good explanation, and I wanted to play that for you. (Demonstrating.) That is about all I have to tell you. If you have any questions you may ask me later. I would like to thank you for the opportunity to speak, and I hope that you might see some use for our speech generation in the type of products you are interested in. I thank you for your attention. (Applause.)

DR. BOYLE: Thank you very much. I find these sorts of developments very exciting ones. You just try to think how you apply these to the sorts of problems you have.

Are there any questions on either of these audio output aspects? No questions? No comments about any other audio systems that may be in use which might be helpful to people?