INTRODUCTION

The IBM 7770 and 7772 and elements of the New York Stock Exchange Market Data System form a family of IBM devices providing voice output facilities. The devices function similarly in that each gives a computer-generated voice response to a dialed inquiry. The audio generation principle incorporated in the 7770 is a derivative of the original Voice Answer Back principle—that of adjusting word length to fit machine time-slots. Then by access to these words from many input lines, sentences are formed into a voice response. The 7772, on the other hand, generates audio on the “vocoder” principle—that of energizing tone filters and combining the output result first to form words and then sentences in a manner similar to that of the 7770.

Input to the New York Stock Exchange Market Data System was accomplished through the use of the IBM 7750 program transmission control. In contrast, the 7770 and 7772 are self-contained input/output devices designed for widely diversified applications requiring various types and lengths of inquiries and responses in such industries as banking, insurance, manufacturing, and retailing.

They are, therefore, modular in increments of numbers of “lines” and “words.” A “line” is defined as a half-duplex communication channel to which more than one telephone may be connected, but where only one transaction takes place at one time. In the case of the 7770 and 7772, a line is used for transmission of digital information in the input direction and voice in the output direction. A “word” is a unit of vocabulary which, for speech-processing purposes, may be either in analog or digital form.

There are differences in the number of lines and words available on each device and in the method of generating voice output. There are three models of the 7770 and one of the 7772. The difference among the three models of the 7770 is in methods of attachment to host processors. Table 1 shows how many lines and words are available on each and the type of processors to which each device can be attached.

Both units are available in languages other than English, but because of the many variations in vocabulary and methods of attachment to communication equipment on a worldwide basis, this discussion is limited to that of the 7770 Model 3 and 7772 as attachable to IBM System/360 within the continental United States. The external appearance of the two machines is similar because the same
### Table 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of Lines</th>
<th>No. of Words</th>
<th>Processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>7770-1</td>
<td>4 to 48 (in increments of 4)</td>
<td>32 to 126 (in increments of 16)</td>
<td>IBM 1401, 1140, 1460</td>
</tr>
<tr>
<td>7770-2</td>
<td>4 to 48 (in increments of 4)</td>
<td>32 to 127 (in increments of 16)</td>
<td>IBM 1410, 7010</td>
</tr>
<tr>
<td>7770-3</td>
<td>4 to 48 (in increments of 4)</td>
<td>32 to 128 (in increments of 16)</td>
<td>IBM System/360 Mod. 30, 40, or 50</td>
</tr>
<tr>
<td>7772</td>
<td>2 to 8 (in increments of 2)</td>
<td>Any amount from available vocabulary list. Limited by available storage.</td>
<td>IBM System/360 Mod. 30, 40, or 50</td>
</tr>
</tbody>
</table>

frame with similar means of connection to communication facilities and host processors is used for both. (See Fig. 1.) In function, the 7770 and 7772 are the same; the difference between the two is in the method of generating the voice output as will be described later in this article.

There are three basic sections to each device—input, output, and control. The input and output sections connect to the common-carrier communications network. The control section connects the input and output sections to the host processor. A transaction takes place in the following way.

Using the telephone, an inquirer first dials the telephone number allotted to a 7770 or 7772. When the ringing stops, a tone will be heard, indicating that the call has been answered. The inquirer now dials his input message. The 7770 or 7772 forwards this message, a character at a time, to the attached computer which processes the input data and returns a digital output message. The message is converted to audio and heard by the inquirer.

**METHODS OF ATTACHING COMMUNICATIONS EQUIPMENT**

The design of the communications interface required consideration of the various uses of a telephone in a machine environment. The telephone, in this case, was the prime input and output device involving dialed digits as input and audio as output.

Some of the first design problems affecting the telephone as an inquiry terminal involved consideration of the human element. How does the inquirer react to a telephone that he knows is connected to a machine? If the inquirer receives no reply, does he hang up? If so, right away? In 20 seconds? In 2

![Figure 1. An audio response unit in background.](From the collection of the Computer History Museum (www.computerhistory.org))
It seemed that most human-factor problems fell in the category of "normal telephone practice" and that the real problem was not what the inquirer would do but how the 7770 or 7772 would react to questions like these:

- What does it do if the inquirer misdials?
- How will it recognize end-of-inquiry?
- What codes will be presented to it?
- How much error-checking can reasonably be done?
- Should it accept d-c dial pulses or tones?
- Should it accept serial and/or parallel data?

In addition, there were questions concerning Audio Output over the communications interface.

- Could any data sets be modified to transmit audio?
- Could this be done on balanced lines? On unbalanced lines?
- What should be the level at which audio is transmitted?
- What happens if an inquirer dials-in while audio is being transmitted?

The answer to most of these questions lay in designing the 7770/7772 "front end" to fit a parallel communications data set interface; namely, that of the AT&T 400 series data sets or their equivalent. These data sets are serial-by-character, parallel-by-bit data transmitters and receivers which are capable of handling numeric and/or alphanumeric data. In addition, one receiver type has since been modified to allow audio transmission in the output direction. One data set receiver (equivalent to the Western Electric 403A or 401J3) must be used per 7770 or 7772 line as shown in Figs. 2 and 3 respectively.

Common to all of the connections in Fig. 2 is the use of pushbuttons for inquiry. Entering digits in this manner is faster, more reliable and is gaining in popularity, but there still are two basic types of telephones that can be used—the rotary dial telephone and the pushbutton telephone. If a rotary dial telephone is used, a pushbutton attachment is usually added. An inquirer would first dial the system number with the rotary dial telephone and enter the inquiry with the pushbutton attachment. If the inquirer uses a pushbutton telephone as shown at Fig. 2D, he dials the number and enters the inquiry using the same telephone.

From the collection of the Computer History Museum (www.computerhistory.org)
AUDIO OUTPUT

On the 7770, vocabulary is stored on a drum similar in form to storage of words on a tape recorder. Words are stored around the circumference of the drum surface on tracks. The drum has 128 tracks; it is 4 inches in diameter, 10 inches long, and rotates at 120 revolutions per minute. Each drum track has an associated read head and amplifier for retrieving the recorded word impulses. The rotational speed of the drum dictates that the information per track must fall within a 500-millisecond time period. A process has been developed in IBM to compress words or segments of speech into 500-millisecond time-slots. Words having a time duration greater than 500 milliseconds are placed in 2, or more time-slots or tracks.

The vocabulary is first generated in the following way. An elocutionist speaks the vocabulary words onto a tape recorder. This tape is then digitized through an analog to digital encoder, the output of which is edited and processed by a computer program to fit the words into 500-millisecond time-slots; these digital time-sloted words are stored on tape. At such times as a specific vocabulary is required, words are converted to analog form and placed on the drum at the specific track locations required by the application. This is normally done once for each application. Vocabulary modification is accomplished by removal of the recorded drum cylinder and its replacement with another cylinder having a different vocabulary.

The way the vocabulary is accessed is depicted in Fig. 4, which shows a functional diagram of the drum and the associated analog circuits. For each application, the processor has a table of addresses corresponding to vocabulary words. Upon analyzing an input message, the processor formulates the required output message which is transmitted to the 7770. This digital output message consists of a series of drum track addresses preceded by a line address. Each track address conditions a specific word analog gate allowing a word to be gated onto the Pulse Amplitude Modulation Bus (PAM BUS). From this bus, the word is gated through any message analog gate conditioned by a specific line address. This allows each word to be transmitted to any line and simply represents time-division-multiplexing of the analog word signal. Since this leaves the audio in a rather chopped-up state, the signal passes through a reconstruction filter before being transmitted to the output line. As long as a relatively high sample frequency (e.g., 12KC) is maintained, no appreciable degradation in audio quality is noticed. It must be remembered that the audio output is in the 200 to 4000 cycle per second range.

For the 7772, the method of recording vocabulary is similar to that for the 7770; however, the processing phase is different. The speech on tape is converted to digital data through an analog-to-digital encoder. The processing of vocabulary is then accomplished by band-compressing each word, thereby limiting the numbers of digits per word. This results in providing a stream of digital data which is stored on cards or tape for later transfer to a disk file or similar random-access storage device within the system. This digital data, called Digitally Coded Voice (DCV), consists of sequential aggregate and excitation functions. An aggregate function is 45 bits in length and represents a portion of the analog signal. The excitation function is 8 bits in length and acts as a counter determining the length of time an aggregate function should be used for a specific segment of analog signal. The sequential combina-

*Registered trademark of McGraw-Edison Co.
tion of these aggregate and excitation functions makes the DCV representation of a particular word of vocabulary.

Approximately 300 bytes of data (DCV) are required per second of audio. The word “balance,” for example, would be stored on a disk file as approximately 75 bytes of digital data (600 bits).

Vocabulary in the 7772 is not stored on a drum in analog form but is stored within the system in bit form on a disk file. The CPU reads the required vocabulary words into its own core storage from the file. The information from the file is transferred through the multiplexer channel to the 7772 control unit which, in turn, transmits the data to a Voice Code Translator (VCT); the VCT then converts the data from digital to analog form. One Voice Code Translator is shared by every two input lines.

The Voice Code Translator consists basically of a set of 15 filters, each with suitable energizers and covering the voice frequency range of approximately 200 to 3,700 cycles per second. Each filter covers a specific portion of the voice band. For example, filter No. 1 covers the 200 to 300 cps range, whereas filter No. 15 covers the 3,150 to 3,700 cps range. An audio output is obtained when a combination of these filters is, in effect, energized by a pattern of input data.

A functional diagram of the VCT is shown in Fig. 5. An aggregate function representing a particular portion of a word is placed in the aggregate function register with its corresponding excitation function in the excitation function register. The DCV data (aggregate function) is converted to analog form and gated into the band filters, thereby energizing each band filter to a level and frequency determined by the format of the aggregate function. The output from the band filters is integrated, reconstructed, amplified, and transmitted to a telephone line as audio. The length of time that each aggregate function is used is determined by the excitation function which in turn is dependent on the dynamic range of the analog signal. For example, a constant tone would be signified by using an aggregate function having a corresponding excitation function with a large count.

**Vocabulary**

Since the 7770 vocabulary is stored within the device on a drum, the number of words per application is limited to the number of tracks on the drum. (See Table 1.) The words are chosen from an available vocabulary list. Special words, proper names, and dialects, etc., can be obtained on a charge basis. A user may change his vocabulary by requesting the drum rotor in the machine to be rerecorded with new words of his choice.
There is no storage of vocabulary within the 7772 unit; instead, the words are stored on available random-access devices within the system to which the 7772 is attached. Because the words are in digital form, they require about 2400 bits of storage for each second of speech. If an average speaking rate of 180 words per minute is considered, this means approximately three words per 2400 bits of storage. There is, therefore, only the system limitation of available storage restricting the size of vocabulary per application for the 7772. Changing the vocabulary is simple since it requires only the reading of new words from cards or tape into storage. A list of words is also available for the 7772 with special words or dialects again offered on a charge basis.

DATA FLOW

Both the 7770 Model 3 and the 7772 connect to the common-carrier data set receivers on one side and, on the other, to the multiplexer channel of an IBM System/360 Processor. Input information is entered via a telephone or similar terminal through a common-carrier data set to the 7770 or 7772 input section. It is then converted to data interpretable by the processor and is forwarded, a character at a time, to the processor. The multiplexer channel has the capability of interleaving operations from many low-speed input/output devices and thus provides a high degree of I/O efficiency and adaptability.

The multiplexer channel operates asynchronously with the central processing unit and contains several subchannels. One subchannel is used per I/O line. Data is transmitted to and from the subchannel and the line in parallel 8-bit-byte form. Data operations between the 7770 and the multiplexer channel are controlled by a sequence of commands which, in turn, are controlled by the Operating System/360 control program. The functional block diagrams of a basic 4-line 7770 Model 3 and a 2-line 7772 are shown in Fig. 6 and 7, respectively.

To understand the operation in more detail it is first necessary to define some basic commands used in IBM System/360 for communication devices.

- **ENABLE** — is used to condition a line for accepting or maintaining a call.
- **DISABLE** — is used to condition a line for terminating a call.
- **READ** — is used when information is being transmitted from the 770 to CPU storage.
- **WRITE** — is used when information is being transmitted from CPU storage to the 7770.

---

![Figure 6. Functional block diagram of a 4-line IBM 7770-3 audio response system.](image-url)
All of the above commands are under the control of the IBM System/360 Operating System Control Program. The flow chart (Fig. 8) describes the operation.

A transaction is initiated by removing the telephone receiver and dialing the machine number. If the number is not busy and the line has been previously ENABLED, an inquirer hears a tone that is 2 to 5 seconds in length, indicating that the call has been answered. The inquirer then dials an inquiry character and the 7770 or 7772 requests service from the multiplexer channel. If the caller does not hang up and the character is accepted, it is transferred to processor core storage. This operation continues until the caller stops entering characters (as is indicated by a time-out of 5 seconds or by the caller’s depression of an “end of message” key). If this time-out occurs, the device indicates “end of message” to the multiplexer channel which interrupts the processor. The processor will analyze the completed inquiry and formulate the response message in the form of a sequence of drum addresses (7770) or “DCV” (7772). This sequence of drum addresses or DCV is preceded by the address of the line requiring the response. The multiplexer channel uses this address to identify the line requiring the response. Each time the multiplexer channel sends an output message to the 7770 or 7772, an audio word or portion thereof is gated onto the respective line requiring it.

This is repeated until the channel control word (CCW) count is zero indicating the last word. If conversation mode is not indicated, the line is then disconnected by issuance of a DISABLE command and then re-ENABLED for a new inquiry. If conversation mode is indicated, it will either be a READ command indicating more information from input, or a WRITE command indicating additional response will be transmitted by the processor. Conversation mode is a means by which a caller can effectively conduct a conversation with the computer under control of the operating program.

APPLICATIONS

The choice of either a 7770 or 7772 Audio Response Unit depends entirely on the application. The number of calls per day, the length of the output message, and the size and variations in vocabulary are all factors that must be considered. The 7770 has greater throughput offering service to more lines than the 7772 but has the limitation of approximately 128 vocabulary words per application. The 7772, on the other hand, is limited in vocabulary only by available storage. The 7772, has a lesser number of input/output lines and has less
throughput imposed by the higher data transfer rate across the device/processor interface during vocabulary transfer from the processor.

Both Audio Response Units are suitable for use by any business having a centralized file system in which the telephone is useful as an inquiry terminal. Responsible individuals within an organization can dial the computer directly for any information

Figure 8. Flow chart of inquiry and response.
about an account and receive a voice answer within seconds. Only a small number of words in the vocabulary list take more than 1/2 second; hence, a 20-word message generally takes less than 10 seconds.

Because audio response is a recent addition to data processing, the application of these devices is left largely to the imagination of the system application engineer. A number of applications have been defined but are too lengthy to describe within the scope of this article. A brief description of a banking application follows:

Banking

Account, loan, credit or mortgage type inquiries can easily and quickly be handled by voice response. When a teller needs to determine the account status of a customer who desires to cash a check, he simply picks up this telephone receiver and dials the account number, and follows with some predetermined code; for example, dial 6 for account status information. The voice response received confirms the account number by repeating the number the teller dialed, and then supplies the account balance and any other information the teller requested. After the customer cashes his check on the teller’s approval, the teller, still holding the line, dials the code to inform the processor to update the customer’s account by debiting the amount of the check. In most cases, the transaction from initial inquiry to final response will take the teller less than one minute; there is no paper-work involved and the teller is using a terminal he is already familiar with—the telephone.

Other applications are:

- Finance . . . . . . . stock quotations
  - margin account balance
- Insurance . . . . . . policy status
  - premium information
- Retail . . . . . . . credit inquiry
  - inventory inquiry
- Manufacturing . . . inventory inquiry
  - job status
  - parts cost inquiry

The addition of Audio Response to the family of IBM System/360 input/output devices now provides a closer link between man and the equipment that stores and processes the data with which he is concerned.